frontiers | Research Topics

Women in virology 2022

Edited by
Ana Grande-Pérez, Christine A. King and Antoinette van der Kuyl

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-3174-7
DOI 10.3389/978-2-8325-3174-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Women in virology: 2022

Topic editors

Ana Grande-Pérez — Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Spain

Christine A. King — Upstate Medical University, United States

Antoinette van der Kuyl — University of Amsterdam, Netherlands

Citation

Grande-Pérez, A., King, C. A., van der Kuyl, A., eds. (2023). *Women in virology: 2022*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3174-7

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 24 July 2023
DOI 10.3389/fmicb.2023.1244987

OPEN ACCESS

EDITED AND REVIEWED BY
Anna Kramvis,
University of the Witwatersrand, South Africa

*CORRESPONDENCE
Antoinette Cornelia van der Kuyl
✉ a.c.vanderkuyl@amsterdamumc.nl

RECEIVED 23 June 2023
ACCEPTED 13 July 2023
PUBLISHED 24 July 2023

CITATION
van der Kuyl AC, King CA and Grande-Pérez A (2023) Editorial: Women in virology: 2022. *Front. Microbiol.* 14:1244987. doi: 10.3389/fmicb.2023.1244987

Editorial: Women in virology: 2022

Antoinette Cornelia van der Kuyl[1]*, Christine A. King[2] and Ana Grande-Pérez[3]

[1]Medical Microbiology and Infection Prevention, University of Amsterdam, Amsterdam, Netherlands, [2]Department of Microbiology and Immunology, Upstate Medical University, Syracuse, NY, United States, [3]Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC) Málaga, Málaga, Spain

KEYWORDS

female scientists, STEM research, gender disparity, women, virology

Editorial on the Research Topic
Women in virology: 2022

Viruses have no gender, but virologists do. And, unfortunately, gender disparity in the STEM field continues, with, according to the UNESCO Institute for Statistics, only 30% of the world's researchers being women, and women being underrepresented at the highest levels of academia. As emphasized by UNESCO, science and gender equality are essential for a balanced future. To contribute to the laudable goal of creating a fairer world, and in celebration of International Women's Day 2022, Frontiers in Microbiology launched the Women in Virology 2022 Research Topic to highlight women's achievements in the field, and to act as a platform promoting the work of female scientists.

"*Women in virology: 2022*" was well received by the scientific community, with 13 manuscripts accepted for publication. In line with the broad call for submissions, the publications covered a variety of topics, describing both fundamental and clinical research on viruses infecting plants, oomycetes, and humans. An excellent overview by de la Higuera and Lázaro, is a valuable encounter with the world of viruses. The authors recap that viruses constitute the largest biomass on earth. With an estimated number of 10^{31} virus particles present, they greatly outnumber cellular organisms. Hundreds of millions of virions alone can be present in, for instance, a milliliter of seawater, and ≈20% of aquatic bacteria are lysed every day due to the actions of bacteriophages. Moreover, although it is impossible to draw a single phylogenetic tree of the Virosphere, the omnipresence of ancestral RNA-recognition motifs in viral replication proteins, which likely originate from the RNA world, viruses can be employed in studying the origin of life. The omnipresence and enormous diversity of viruses and their interaction with all organisms, further demonstrates their pronounced influence on the evolution of cellular life. For instance, viruses facilitate horizontal gene transfer in prokaryotes, while virus endogenization in eukaryotic genomes supports a role in evolutionary transitions. As viruses are found everywhere, are highly adaptable, easily dispersed by wind and water, and have a substantial influence on Earth's biochemical cycles and living organisms, the authors argue that, in the field of astrobiology, virus-like agents should be considered as potential carriers of life signatures, so-called "biosignatures," when searching for extraterrestrial life.

Turning to viruses infecting humans, it is remarkable that nine out of ten contributors report on viruses that infect chronically or have the capacity to do so. Kaczorowska et al. report on mother-to-child transmission of the enigmatic anelloviruses (AVs), ubiquitous, persistent single-stranded DNA viruses with no known associated pathology. Sequencing

the so-called anellome of five HIV-1 infected mother and child pairs suggested selected genotype transmission, as the majority of maternal lineages was not shared with the child. Some children harbored unique AV genotypes, likely acquired from other sources. In line with this finding, AV transmission of shared genotypes was not related to delivery or breastfeeding, as the HIV-positive mothers delivered by Caesarian section and did not breastfeed.

Five contributions investigated aspects of human papillomavirus (HPV) infection. In two papers, Yang et al., Yang et al. studied the epidemiology and risk factors associated with HPV infection and associated malignancies in women from Liaoning province in north-eastern China during 2018–2021, as in recent years, cervical cancer rates have been increasing in China, especially in younger women, and routine screening or vaccination is not standard practice. From a total of 16,589 participants, ±12.5% of women were found to be HPV+, with ±10% presenting abnormal cytology. For primary cervical cancer screening and ASCUS (atypical squamous cells of undetermined significance) triage, Li et al. compared the performance of a HPV DNA Chip test (Sample-to-Answer HPV Genotyping System from Beijing Bohui Innovation Biotechnology Co.) and an HPV PCR test (5 + 9 High Risk HPV Detection kit from Tellgen) on cervical samples from 7,241 women from rural areas in Shanxi, China during 2017–2018. Both tests performed better than standard liquid biopsy screening, while the HPV DNA Chip test (14.9% HPV+) outcompeted the HPV PCR test (21.1% HPV+) in the study. As the HPV situation in many parts of the world is critical, and prophylactic vaccination is not universal, Ren et al. performed a meta-analysis on the efficacy of prophylactic HPV vaccines among Asian females, and confirmed that these vaccines significantly reduce the incidence of persistent HPV infection, cytological abnormalities and cervical neoplasia in Asian populations. Unsurprisingly, the authors advocate raising awareness amongst women, and accelerated implementation of HPV vaccination. In a next study, Dai et al. found that stable expression of the HPV oncoproteins E6 and E7 in cell culture induces the differential expression of genes encoding chemokines and cell adhesion proteins. The observation may stimulate novel research into targeting E6E7 for HPV-related cancer treatment.

HBV is another example of a virus with an ability to persist and cause cancer in humans. Belaiba et al. investigated the genomic variation of hepatitis B virus (HBV) in five ETV (entecavir) non-responders in Tunisia. Putative drug-resistance mutations were present in the genotype D genomes, often predating treatment, as were a plethora of other mutations, some associated with disease progression and increased replication. Hepatitis C virus (HCV) is another potentially chronic hepatitis infection able to induce liver malignancy. García-Crespo et al. report a decrease in the effectiveness of antivirals when added, later rather than earlier, to an ongoing infection in cell culture. These results are important, as most patients begin treatment only when the infection has been established for some time, something which is not assessed in preclinical trials.

Human cytomegalovirus (HCMV) is another persistent virus infection. Herpesviruses are capable of alternating between lytic and latent cycles, whereby the induction of latency is an active process, in which major immediate early gene (IE) expression from the major immediate early promoter (MIEP) is suppressed, a process studied by Poole and Sinclair. They report that the cellular protein SERBP1 is upregulated during HCMV latency, is required for MIEP suppression and interacts with the transcriptional repressor CHD3 as well as with KAP1 to recruit SET1B, another repressor, to the viral genomic DNA, thus acting as a scaffold protein to maintain transcriptional silencing.

A non-persistent virus, the omicron-variant of SARS-CoV-2, was studied by Liu et al. who evaluated the presence of viral RNA in vaginal and anal swabs in 63 women with no, or only mild disease symptoms. All vaginal swabs were qRT-PCR negative for SARS-CoV-2 RNA. Four anal swabs were positive, one from a case with gastro-intestinal symptoms, suggesting that the omicron variant does not readily invade the gut.

A different section of the Virosphere was studied by Dye et al. who investigated the viral variation in cassava mosaic disease, an infection caused by a complex of single-stranded DNA viruses of the family *Geminiviridae*, genus *Begomovirus*. In the Kenyan Lake Victoria region, virus mixtures were found to be distinct from those in coastal samples. Experiments suggested that transmission dynamics of the viruses also differed, with African cassava mosaic virus (ACMV) being transmitted through vegetative propagation, while East African cassava mosaic virus (EACMV) relied on a whitefly vector. These findings could explain the divergent virus mixtures observed.

Raco et al. studied yet another realm of the virus world when searching for novel RNA viruses in the Vietnamese oomycete species, *Phytophthora castaneae*. Oomycetes are a group of fungus-like organisms with phylogenetic similarity to algae and diatoms, which include some of the most devastating plant pathogens. Five putative RNA virus genomes detected showed similarity to members of the order *Bunyavirales* and families *Endornaviridae*, *Megabirnaviridae*, *Narnaviridae*, *Totiviridae,* and the proposed family "Fusagraviridae"; a partial genome showed homology to the *Endornaviridae*. As initial virus identification was through small RNA sequencing, it was proposed that a RNAi antiviral defense mechanism had been targeting the viral RNA.

In conclusion, female virologists are as versatile as their subjects of study and work in all areas of the virosphere. Frontiers Research Topic "*Women in Virology*" helps showcasing their contributions to the field.

Author contributions

AvdK drafted the manuscript. All three authors edited the Frontier's Research Topic "*Women in Virology 2022*". All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fmicb.2022.911474

High Diversity of Novel Viruses in the Tree Pathogen *Phytophthora castaneae* Revealed by High-Throughput Sequencing of Total and Small RNA

Milica Raco[1], Eeva J. Vainio[2], Suvi Sutela[2], Aleš Eichmeier[3], Eliška Hakalová[3], Thomas Jung[1] and Leticia Botella[1]*

[1]Phytophthora Research Centre, Department of Forest Protection and Wildlife Management, Faculty of Forestry and Wood Technology, Mendel University in Brno, Brno, Czechia, [2]Natural Resources Institute Finland (Luke), Helsinki, Finland, [3]Mendeleum-Institute of Genetics, Faculty of Horticulture, Mendel University in Brno, Brno, Czechia

OPEN ACCESS

Edited by:
Ana Grande-Pérez,
Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Spain

Reviewed by:
Robert Henry Arnold Coutts,
University of Hertfordshire, United Kingdom
Robin Marion MacDiarmid,
The New Zealand Institute for Plant and Food Research Ltd, New Zealand

***Correspondence:**
Milica Raco
milica.raco@mendelu.cz

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 02 April 2022
Accepted: 21 April 2022
Published: 16 June 2022

Citation:
Raco M, Vainio EJ, Sutela S, Eichmeier A, Hakalová E, Jung T and Botella L (2022) High Diversity of Novel Viruses in the Tree Pathogen Phytophthora castaneae Revealed by High-Throughput Sequencing of Total and Small RNA. Front. Microbiol. 13:911474. doi: 10.3389/fmicb.2022.911474

Phytophthora castaneae, an oomycete pathogen causing root and trunk rot of different tree species in Asia, was shown to harbor a rich diversity of novel viruses from different families. Four *P. castaneae* isolates collected from *Chamaecyparis hodginsii* in a semi-natural montane forest site in Vietnam were investigated for viral presence by traditional and next-generation sequencing (NGS) techniques, i.e., double-stranded RNA (dsRNA) extraction and high-throughput sequencing (HTS) of small RNAs (sRNAs) and total RNA. Genome organization, sequence similarity, and phylogenetic analyses indicated that the viruses were related to members of the order *Bunyavirales* and families *Endornaviridae*, *Megabirnaviridae*, *Narnaviridae*, *Totiviridae,* and the proposed family "Fusagraviridae." The study describes six novel viruses: Phytophthora castaneae RNA virus 1–5 (PcaRV1-5) and Phytophthora castaneae negative-stranded RNA virus 1 (PcaNSRV1). All six viruses were detected by sRNA sequencing, which demonstrates an active RNA interference (RNAi) system targeting viruses in *P. castaneae*. To our knowledge, this is the first report of viruses in *P. castaneae* and the whole *Phytophthora* major Clade 5, as well as of the activity of an RNAi mechanism targeting viral genomes among Clade 5 species. PcaRV1 is the first megabirnavirus described in oomycetes and the genus *Phytophthora*.

Keywords: mycovirus, dsRNA, RNA interference, multiple viral infections, forest pathogen, RdRp, ssRNA, oomycetes

INTRODUCTION

Phytophthora spp. are oomycetes taxonomically grouped under the kingdom Straminipila (Heterokonta; Beakes et al., 2014; Thines and Choi, 2016). Although similar in habitat and morphology to filamentous fungi, they are phylogenetically more closely related to brown algae and diatoms. Ubiquitous in marine, freshwater, and terrestrial environments (Beakes et al., 2014), many *Phytophthora* spp. are major plant and forest pathogens causing substantial economic

losses in agriculture, horticulture, and silviculture, along with environmental damage in natural ecosystems, thereby impacting biodiversity worldwide (Erwin and Ribeiro, 1996; Jung et al., 2018). *Phytophthora castaneae* Katsura & K. Uchida (≡ *P. katsurae*, nom. Illegit.) is a plant pathogen mostly known for causing trunk rot on Japanese chestnut (*Castanea crenata* Sieb. et Zucc.). It is a homothallic species native to Southeast and East Asia (Jung et al., 2017, 2020) with papillate sporangia, easily distinguished from related species by its ornamented oogonia. It resides in *Phytophthora* phylogenetic Clade 5, one of the smallest *Phytophthora* clades, which currently comprises four species, i.e., *P. agathidicida* B.S. Weir, Beever, Pennycook & Bellgard, *P. castaneae*, *P. cocois* B.S. Weir, Beever, Pennycook, Bellgard & J.Y. Uchida, *P. heveae* A.W. Thomps., and the as yet undescribed taxon *P.sp.* "*novaeguineae*" (Martin et al., 2014; Weir et al., 2015). *Phytophthora* species from Clade 5 are distributed in East and Southeast Asia, Papua New Guinea, Eastern Australia, New Zealand, Hawaii, and Central and South America (Erwin and Ribeiro, 1996; Bruce, 1999; Martin et al., 2014; Weir et al., 2015; Jung et al., 2017, 2020; Legeay et al., 2020).

Fungal viruses (mycoviruses) have been discovered in many major taxa of phytopathogenic fungi (Ghabrial et al., 2015). In oomycetes, an increasing number of viruses with double-stranded (ds) RNA, positive-sense (+) single-stranded (ss) RNA, and negative-sense (−) single-stranded (ss) RNA genomes have been reported over the last few years (Shiba et al., 2018, 2019; Cai et al., 2019a,b; Botella et al., 2020; Botella and Jung, 2021; Fukunishi et al., 2021; Poimala et al., 2021). However, many species have yet not been screened for the presence of viruses. Eight novel linear (−)ssRNA viruses putatively belonging to the order *Bunyavirales* were recently described in *Halophytophthora* (Botella et al., 2020), a primarily marine sister genus of *Phytophthora* (Maia et al., 2022). A metagenomics study revealed a rich (−)ssRNA, (+)ssRNA, and dsRNA virome in *Plasmopara viticola* (Berk. & M.A. Curtis) Berl. & De Toni-associated lesions (Chiapello et al., 2020); a (+)ssRNA virus has been described in *Pl. halstedii* (Farl.) Berl. & De Toni and *Sclerophthora macrospora* (Sacc.) Thirum., C.G. Shaw & Naras (Yokoi et al., 1999, 2003; Heller-Dohmen et al., 2011; Grasse and Spring, 2017). In the oomycete genus *Pythium*, dsRNA and virus-like particles were observed in Australian isolates of *Py. irregulare* Buisman (Gillings et al., 1993). In an isolate of the mycoparasitic *Globisporangium nunn* (Lifsh., Stangh. & R.E.D. Baker) Uzuhashi, Tojo & Kakish., formerly classified as *Py. nunn* Lifsh., Stangh. & R.E.D. Baker, Pythium nunn virus 1 (PnV1), a distinct member of the viral genus *Gammapartitivirus* in the family *Partitiviridae*, was described (Shiba et al., 2018). Viruses related to the fusarivirus [Pythium ultimum RNA virus 1 (PuRV1)] and the totivirus [Pythium ultimum RNA virus 2 (PuRV2)] were discovered from a Japanese isolate of the plant-parasitic oomycete *G. ultimum* (Trow) Uzuhashi, Tojo & Kakish. (Fukunishi et al., 2021). A toti-like virus, Pythium polare RNA virus 1 (PpRV1), was reported from the heterothallic oomycete species *Py. polare* Tojo, van West & Hoshino isolated from a moss in the Arctic together with Pythium polare RNA virus 2 (PpRV2) and Pythium polare bunya-like RNA virus 1 (PpBRV1; Sasai et al., 2018). Another toti-like virus closely related to PpRV1 was reported in Japanese isolates of *G. splendens* (Hans Braun) Uzuhashi, Tojo & Kakish., formerly *Py. splendens* Hans Braun, Pythium splendens RNA virus 1 (PsRV1; Shiba et al., 2019).

Until very recently, there were only a few viruses described in the genus *Phytophthora*, mainly from the notoriously devastating potato blight pathogen *P. infestans* (Mont.) de Bary (Cai et al., 2009, 2012, 2013, 2019b; Cai and Hillman, 2013). Phytophthora infestans RNA virus 1 (PiRV-1) and Phytophthora infestans RNA virus 2 (PiRV-2) seem to represent entirely novel viral families (Cai et al., 2009, 2019b; Cai and Hillman, 2013). PiRV-2 is found to stimulate sporangia production in its host, *P. infestans* (Cai et al., 2019a). Phytophthora infestans RNA virus 3 (PiRV-3; Cai et al., 2013) clusters with the newly proposed family "Fusagraviridae" (Wang et al., 2016). Phytophthora infestans RNA virus 4 (PiRV-4) is an unclassified narnavirus characterized in the same isolate as PiRV-3 (Cai et al., 2012). Members of alphaendornaviruses were found in isolates of *Phytophthora* taxon "douglasfir," *P. ramorum* Werres, De Cock & Man in 't Veld (Kozlakidis et al., 2010) and in isolates of a *Phytophthora* pathogen of asparagus collected in Japan (Uchida et al., 2021). Recently, a number of novel viruses were identified in several other *Phytophthora* species (Botella and Jung, 2021; Poimala et al., 2021; Xu et al., 2022). Thirteen bunya-like and two toti-like viruses have been detected by HTS of total and small RNA (sRNA) in *P. condilina* T.I. Burgess from *Phytophthora* phylogenetic Clade 6a (Botella and Jung, 2021). In addition to a totivirus Phytophthora cactorum RNA virus 1 (PcRV1; Poimala and Vainio, 2020), seven novel viruses were recently reported from *P. cactorum*. This included three viruses having genome affinities to members of *Bunyavirales*, three alphaendornaviruses, and an usti-like virus (Poimala et al., 2021). In New Zealand, a novel unclassified dsRNA virus designated Phytophthora pluvialis RNA virus 1 (PplRV1) has been described from *P. pluvialis* Reeser, W. Sutton & E.M. Hansen (Xu et al., 2022).

Besides conventional methods, such as dsRNA extraction based on *CF*-11 cellulose affinity chromatography (Morris and Dodds, 1979) novel technologies, such as high-throughput sequencing (HTS), have enabled rapid expansion of knowledge in this field of virology. The screening of isolates by dsRNA extraction followed by purification and *CF*-11 chromatography is a traditional and relatively inexpensive first approach for viral detection, but it has certain limitations. Such a method was developed primarily to detect viruses with dsRNA genomes and, to a certain extent, is able to detect ssRNA viruses during their replication phase. As such, the method is less likely to detect DNA viruses. In addition, ssRNA viruses could potentially go undetected as the replication of intermediate dsRNA may not accumulate sufficiently during cell infection to allow detection. This is likely to be the case in *Heterobasidion parviporum* Niemelä & Korhonen infected by viruses from two (+)ssRNA virus families *Narnaviridae* and *Botourmiaviridae*, detected by HTS of total RNA depleted of rRNA, but not discovered earlier by extensive dsRNA screening (Sutela et al., 2021). Likewise, in three different species of *Armillaria* (Fr.) Staude, the presence of several ssRNA viruses has been revealed by HTS of total

RNA, while no dsRNA elements were detected (Linnakoski et al., 2021). In addition, the dsRNA extraction technique does not allow characterization of the virome of samples containing a complex range of viruses, and viral full-length characterization is time consuming and requires additional costs (Nerva et al., 2016). In contrast, novel technologies such as HTS of sRNA and total RNA are widely used in detection of viruses, with HTS of total RNA being the dominant method nowadays for both detection and full-length genome characterization, despite having higher per sample cost. In order to direct the translational machinery of the host cell to produce viral proteins and ensure their survival, all viruses independent of their genome organization (DNA or RNA) need to express their genetic material as functional messenger RNAs (mRNAs) during the early infection stage. From genome to mRNA, viruses follow different pathways, which can be attributed to their type of genome (Rampersad and Tennant, 2018). Thus, sequencing of total RNA and sRNAs derived from it, potentially allows detection of all types of viruses, including those with DNA genomes (Nerva et al., 2016).

RNA interference (RNAi) is one of several pathways grouped under RNA silencing phenomena, defined as sequence-specific RNA degradation activated by the presence of long dsRNAs (Ketting, 2011; Svoboda, 2020). RNAi is a conserved mechanism used by many eukaryotic organisms, including animals, plants, fungi (Meister and Tuschl, 2004; Chang et al., 2012), and oomycetes (Mascia et al., 2019), for regulating the activity of gene expression. An important role of this biological pathway is to respond to non-self dsRNAs and mediate protective mechanisms against RNA viral infection. One of the first discoveries of RNA silencing pathways and the first experimental evidence of such activity in fungi was reported from the heterothallic filamentous fungus *Neurospora crassa* Shear & B.O. Dodge (Romano et al., 1992), which today serves as a model organism for studying RNA silencing. As RNA silencing is triggered by the occurrence of viral dsRNAs in the cell, virus-infected organisms in which this system is active are enriched for fragments of RNA representing small segments of viral genomes. HTS of sRNAs has been successfully used to detect viruses in plants and fungal pathogens (Kreuze et al., 2009; Vainio et al., 2015a; Donaire et al., 2016; Adalia et al., 2018; Massart et al., 2019; Kocanová et al., 2020; Botella and Jung, 2021).

The objectives of this study were primarily to identify and characterize viruses in *P. castaneae* combining traditional and next-generation sequencing (NGS) techniques, and, in addition, to examine whether *P. castaneae* has an active RNAi mechanism and processes viral RNAs into siRNAs as an act of defense. To the best of our knowledge, this is the first virome investigation of the tree pathogen *P. castaneae* or any species of *Phytophthora* phylogenetic Clade 5.

MATERIALS AND METHODS

Phytophthora Isolates

Four isolates of *P. castaneae*, i.e., VN999, VN1004, VN1008, and VN1012, from the culture collection at Phytophthora Research Centre (PRC), Mendel University in Brno, originally isolated in 2017 from rhizosphere soil of four *Chamaecyparis hodginsii* (Dunn) Rushforth trees growing in a montane *Chamaecyparis-Quercus* forest stand in northern Vietnam close to Sapa (Jung et al., 2020), were selected for this study. The selected isolates were maintained on V8-juice agar [V8A; 100 ml/l V8 juice (Hermann Pfanner Getränke, Lauterach, Austria), 16 g/l of agar (Sigma-Aldrich, St. Louis, MO, United States), 2 g/l of $CaCO_3$, and 900 ml/l of distilled water] at 20°C in the dark. For total RNA extraction, isolates were grown on carrot juice agar [CA; 100 ml/l carrot juice (Hermann Pfanner Getränke), 16 g/l of agar (Sigma-Aldrich), 2 g/l of $CaCO_3$, and 900 ml/l of distilled water] covered with cellophane membrane (EJA08-100; Gel Company Inc., San Francisco, CA, United States). For dsRNA extraction, the isolates were grown in 100 ml broth containing 0.2 g of $CaCO_3$, 10 ml of V8 juice, and 90 ml of distilled water. The Erlenmeyer flasks with the liquid cultures were incubated at 20°C and 120 RPMs for 7–10 days in a Benchtop Incubated Shaker (IST-3075R; Lab Companion; Seoul, Republic of Korea).

Isolation of dsRNA

The four *P. castaneae* isolates were first examined for putative virus presence using a modified dsRNA extraction protocol (Morris and Dodds, 1979) as described by Tuomivirta et al. (2002). Approximately 2 g of mycelium (fresh weight) was collected in a 50 ml Falcon tube together with two 10 mm diameter stainless steel beads, submerged in liquid nitrogen, and homogenized by vortexing using a Vortex-Genie 2 (SI-0256; Scientific Industries, Inc., Bohemia, NY, United States) and Vertical High-Speed 50 ml Tube Holder (SI-V203; Scientific Industries, Inc., Bohemia, NY, United States). After homogenization, dsRNA extraction was performed as described by Tuomivirta et al. (2002). The obtained products were visualized in a 1.2% agarose gel under ultraviolet light.

Isolation of Total RNA

After 7–15 days, approximately 50 mg of fresh mycelium was collected from the cellophane surface in a sterile 2 ml RNAse/DNAse free tube supplied with lysing matrix C (740813.50; MN Bead Tubes Type C; Macherey-Nagel; Düren, Germany). The tubes were then submerged in liquid nitrogen and vortexed using the Vortex-Genie 2 and high-speed bead tube holder until the mycelium was ground to a fine powder. Total RNA was isolated using the SPLIT RNA extraction kit (Lexogen, Vienna, Austria) following the protocol provided by the manufacturer, eluted in 50 µl of SB buffer (Lexogen), and stored at −80°C.

High-Throughput Sequencing of sRNAs

Prior to library construction, the quantity of extracted RNA was measured by a Modulus™ Single Tube Turner BioSystems device (9200–000; Modulus™; Turner BioSystems Inc.; Sunnyvale, CA, United States). The integrity of total RNA was visually assessed by running it on a 1.2% agarose gel stained with SYBR™ Green I Nucleic Acid Gel Stain (S7563; Invitrogen™; Thermo Fisher Scientific, Waltham, MA, United States). The

libraries were generated directly from total RNA using the TruSeq small RNA library preparation kit (Illumina, San Diego, CA, United States) according to the manufacturer's instructions. After reverse transcription and library amplification, libraries were loaded onto an Agilent High Sensitivity DNA chip (Agilent High Sensitivity DNA Kit, Agilent Technologies, Inc., Santa Clara, CA, United States) according to the manufacturer's instructions and run on an Agilent – 2100 Bioanalyzer, Automated Electrophoresis Instrument (Agilent Technologies). cDNA constructs were purified and loaded into a Novex 6% TBE polyacrylamide 15-well gel (Life Technologies, Thermo Fisher Scientific; Waltham, MA, United States) in 5X Novex® TBE Running Buffer (Life Technologies, Thermo Fisher Scientific; diluted to 1X before the run) and ran on the XCell SureLock® Mini-Cell electrophoresis unit (Life Technologies, Thermo Fisher Scientific). After the run, the gel was stained with GelRed® Nucleic Acid Gel Stain (Biotium, Fremont, CA, United States) and viewed under UV light. The bands that correspond to the adapter-ligated constructs derived from the 22-nucleotides (nts) and 30-nts sRNA fragments [the region between 145 base pairs (bp) and 160 bp amplicons] were cut out using a sterile razor blade. Gel slices were placed into 0.5 ml gel breaker tubes. The target products were purified from PAGE using glycogen (Sigma-Aldrich, Steinheim, Germany), 3 M NaOAc (Sigma-Aldrich), and ethanol. The final product was checked for size, purity, and concentration and the pellet resuspended (to reach a concentration of 2 nm) in 15 µl of Tris–HCl 10 mm, pH 8.5.

The libraries were sequenced on an Illumina MiniSeq sequencer using a MiniSeq High Output Reagent kit (75-cycles, Illumina). Constructed and normalized libraries were diluted to 1 nm and denatured. The 500 µl of prepared 1.8 pM library was then loaded to the reagent cartridge's MiniSeq High Output Reagent (75 cycles). Six-base indexes were used as they significantly reduce ligation bias and ensure accurate measurement of miRNA expression. The libraries were sequenced in a single-read mode, 1 × 36 bp for VN999 and VN1008, and 1 × 50 bp for VN1004 and VN1012.

Stranded Total RNA Sequencing

Total RNA of *P. castaneae* isolate VN999 and *P. condilina* BD661 (Botella and Jung, 2021) was pooled together and treated with TURBO DNA-free™ Kit (AM1907; Invitrogen™, Thermo Fisher Scientific). RNA quantity and quality were measured by the Qubit® 2.0 Fluorometer (Life Technologies, Thermo Fisher Scientific) and Tape Station 4200 (Agilent Technologies). Total RNA was sent to Macrogen Inc., the Republic of Korea, for RNA library construction and deep sequencing. Prior to the library preparation, ribosomal RNA (rRNA) was depleted using the Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat). The library was prepared using an Illumina TruSeq® Stranded Total RNA Sample Preparation Kit (Illumina) and sequenced in paired-end (2 × 101 nts) mode on a NovaSeq6000 (Illumina).

Bioinformatics

High-Throughput Sequencing of sRNAs

The raw data quality was checked using the FastQC-0.10.1 program (Andrews, 2010). Adapters, low-quality reads, and too short reads were removed by the FASTX-Toolkit Clipper,[1] stating the Q33 parameter. Individual reads were assembled using a set of algorithms known as Velvet (version 7.0.4; Zerbino and Birney, 2008). Reads below 17 nts were removed. The reads were assembled with Velvet k-mer lengths from 13 to 29. All the assemblies produced by Velvet were merged using the AssemblyAssembler1.3 script.[2] The obtained contigs were aligned to non-redundant (nr) viral protein and nucleotide databases using BlastX and BlastN (NCBI BLAST+ 2.12.0). The BlastX search was limited to records that included "Viruses" (taxid:10239). All the contigs were imported to Geneious Prime® 2019.0.4 bioinformatics software for molecular biology and sequence analysis.

Stranded Total RNA Sequencing

For the data processing of stranded total RNA-Seq reads, two approaches were used. For both, quality control of raw sequencing data was performed using the FastQC-0.10.1 program (Andrews, 2010). In the first approach, the data were cleaned of the host reads by assembling it to the host genome. Because of the unavailability of the *P. castaneae* genome, the reads were assembled to the genome of the most similar species from the same clade, *P. agathidicida* GCA_001314435.1_NZFS3772v2_genomic.fna.gz (Studholme et al., 2016), using the BWA software package and MEM algorithm (Li, 2013). In the second approach, the reads were used in *de novo* assembly directly after quality control. *De novo* assembly for both methods was performed by Trinity v2.6.5 (Grabherr et al., 2011), and all subsequent steps were the same for both data sets. Contigs obtained by Trinity were searched for similarities to a custom virus protein database in BlastX (BLASTX 2.10.0+) algorithm (Altschul et al., 1997) with an E value set to 10^{-5}. All sequences showing significant similarity to known viruses were also aligned in the nucleotide database BlastN (NCBI[3] BLAST+ 2.12.0) to exclude contigs potentially originating from the host. Potential protein encoding segments were detected with coding open reading frame (ORF) finder at NCBI website.[4] In order to find putative viral contigs having low sequence similarity with the viral proteins deposited to GenBank, Trinity contigs originating from cleaned total RNA-Seq reads were examined and ORFan contigs were retrieved as follows. The *de novo* assembly of total RNA library was mapped to the sRNA library of isolate BD661 (Botella and Jung, 2021) as well as the putative virus sequences hosted by VN999 using Geneious 10.2.6 assembler with medium-low sensitivity. The contigs not mapped were selected and assembled next to the sRNA library of VN999 using the same settings. Of mapped contigs those over 800 nts in length were selected and run with BlastX against the NCBI nr database (E value $10e^{-6}$) and, subsequently, the contigs having no hits were examined further with blast searches and Geneious software.

[1] http://hannonlab.cshl.edu/fastx_toolkit/
[2] https://github.com/dzerbino/velvet/tree/master/contrib/AssemblyAssembler1.3
[3] https://blast.ncbi.nlm.nih.gov
[4] https://www.ncbi.nlm.nih.gov/orffinder/

Depth of Coverage

In order to calculate coverage depth for each virus, and to get information on the number of reads representing individual viruses, raw reads were mapped to the final contigs of each virus. The number of total RNA reads for each virus was calculated by Bowtie 2 v2.3.0 (Langmead and Salzberg, 2012). The number of sRNA reads for each virus was calculated by the alignment program Bowtie 1 v1.1.2 suitable for aligning the relatively short sequencing reads (Langmead et al., 2009). Both types of alignments were done in Geneious Prime® 2020.2.3. For calculation of the coverage depth, the following formula was used as:

$$\frac{\text{total no. of reads}\left(\text{sRNA / total RNA}\right)^{*} \times \text{average read length}}{\text{virus genome length}}$$

*Total number of reads mapped to the final virus sequence.

Detection and Validation of Viruses by RT-PCR Amplification and Sanger Sequencing

After 7–15 days of growth, approximately 100 mg of mycelium was collected and frozen in liquid nitrogen. Total RNA was isolated by RNAzol® RT (Sigma-Aldrich, Steinheim, Germany) following the manufacturer's recommendations with some modifications. The samples were homogenized in 1 ml of RNAzol® RT (Sigma-Aldrich) using a Mixer Mill MM 400 (Retsch GmbH, Haan, Germany) for 2 min at 30 Hz. After the isopropanol precipitation step, the samples were loaded on RNA extraction columns (Lexogen, Vienna, Austria) and centrifuged at high speed for 30 s (RT). This step was repeated until the whole sample was loaded. The columns were then washed twice with 600 µl 75% ethanol and spin-dried after the last wash for at least 1 min at high speed. The columns were then transferred to sterile nuclease-free 1.5 ml tubes. The RNA was eluted by adding 50 µl Elution buffer SPLIT RNA extraction kit (Lexogen) preheated to 70°C, incubated for 1 min, and then centrifuged for 1 min. The quality of total RNA was assessed by visualization under UV light in a 1.2% agarose gel (visible presence of 18 S and 28 S rRNA bands). A Qubit® 2.0 Fluorometer (Life Technologies, Thermo Fisher Scientific) was used to estimate RNA quantity. The final product was stored at −80°C. Prior to cDNA synthesis, total RNA was incubated for 10 min at 65°C. cDNA was synthesized using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosciences, Park Ave, NY, United States). To confirm the successful synthesis of cDNA, the actin housekeeping gene was amplified with primers MIDFWACT and MIDREVACT (Weiland and Sundsbak, 2000). Every cDNA was checked, and if amplification was successful, the cDNA was used in a PCR reaction with virus-specific primers. The virus-specific primers were designed to amplify a fragment of the RNA-dependent RNA polymerase (RdRp) or capsid protein (CP) gene of putative viruses. All virus-specific primers were designed by Primer 3 2.3.7 in Geneious Prime® 2020.2.3 (detailed information is presented in **Supplementary Table 1**). PCR amplification was performed with 20 µl Platinum™ II Hot-Start Green PCR Master Mix (2X; 14001014; Invitrogen™; Thermo Scientific) designed for universal primer annealing, 1 µl of each 10 mm primer, 3–5 µl of cDNA depending on the virus and its abundance and PCR grade water in a total volume of 50 µl. Cycling conditions followed the manufacturer's recommendations, and the annealing temperature of 60°C was universal for all primers. Amplification was checked in 1.2% agarose gels stained with DNA Stain G (39803; SERVA, Heidelberg, Germany) after separation by electrophoresis (120 V, 50 min). All amplicons of the appropriate size were sent to Eurofins Genomics (Ebersberg, Germany) to be sequenced in both directions.

Validation of Nearly Complete Viral Genomes and the Reference Sequences by Sanger Sequencing and Detection of Viral Strains

For verifying the accuracy of the final viral contigs obtained after the *de novo* assembly of total RNA reads and for determination of the exact nucleotide sequence of virus variants among studied isolates, virus-specific primers (**Supplementary Table 2**) covering almost entire genome length were designed and obtained amplicons were Sanger sequenced. Total RNA was extracted by RNAzol® RT (Sigma-Aldrich) and transcribed to cDNA as described above. RT-PCR was conducted as described previously in 25 µl reaction volumes using 1.5–2.5 µl of cDNA. Obtained amplicons were sent to Eurofins Genomic to be Sanger sequenced in both directions. Pairwise sequence alignments were done by MAFFT v7.450 (Katoh and Standley, 2013) in Geneious Prime® 2020.2.3. The accuracy of the Trinity contigs was validated in reference isolate VN999 by mapping the Sanger sequences obtained from VN999 viral amplicons to the final viral contigs by Geneious mapper using Medium-Low sensitivity settings.

RNA Structures and Conserved Domains

NCBI CD-search tool[5] (last accessed on 08.03.2022) was used for search of putative conserved domains. For the prediction of H-type pseudoknots in viral RNA sequences, DotKnot[6] was used with maximum free energy (MFE). In order to look for putative conserved motifs of RdRp regions of putative viruses, predicted amino acid (aa) sequences were aligned using MUSCLE 3.8.425 (Edgar, 2004) in Geneious Prime® 2020.2.3 to aa sequences of related viruses retrieved from the GeneBank. The % of similarities were calculated based on the Blosum62 score matrix with a threshold of 1.

Phylogenetic Analyses

Amino acid (aa) sequences of RdRp regions of each virus were included in phylogenetic analyses. Sequences were aligned in Geneious 2020.2.3 by MUSCLE 3.8.425 (Edgar, 2004) together with known aa sequences of viruses considered to be related. Phylogenetic trees were built using the maximum likelihood

[5]https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[6]https://dotknot.csse.uwa.edu.au/

method (Stamatakis et al., 2008) in RAxML-HPC v.8 on XSEDE running in the CIPRES Science Gateway web portal (Miller et al., 2010). Bootstrapping was performed by using the recommended parameters provided by the CIPRES Science Gateway portal. In order to avoid thorough optimization of the best-scoring ML tree at the end of the run, the GAMMA model was used. As the substitution model for proteins, the Jones–Taylor–Thornton (JTT) model was used. The trees were visualized in FIGTREE (V1.4.4).

RESULTS

DsRNA Screening

Screening of possible viral dsRNA molecules of each of four Vietnamese isolates of *P. castaneae* (VN999, VN1004, VN1008, and VN1012) revealed different dsRNA banding patterns (**Supplementary Figure 1**). Fragments of *ca.* 7 and 0.9 kb were visible in all four isolates. In addition, bands of *ca.* 1.9 and *ca.* 0.5 and a band of *ca.* 10–20 kb could also be observed. In isolate VN999, an additional band of *ca.* 4 kb occurred. There was a lack in the consistency of the banding patterns, i.e., not all the bands appeared in every dsRNA extraction performed. As dsRNA occurrence indicates putative viral presence, all four isolates underwent HTS of sRNAs. Isolate VN999 was selected for stranded RNA sequencing.

High-Throughput Sequencing, *de novo* Assembly, Size Profiling of sRNAs, and Identification of sRNA Contigs

In the first run, sequencing of three separately indexed libraries, VN1004, VN1012, and a third library that is not part of this study, gave 11,205,897 reads. 10,836,989 reads passed through the quality filtering stage; 33.34 and 33.59% of reads belonged to VN1004 and VN1012, respectively. The second run including libraries of isolates VN1008 and VN999 resulted in 19,467,986 reads, of which 18,094,637 passed quality filtering; 27.86 and 37.22% of reads belonged to VN999 and VN1008, respectively. The remaining reads from both runs belonged to libraries that are not part of this study.

A graphical representation of the total number of raw sequences and sRNA size profiles per isolate are presented in **Supplementary Figure 2**. The number of individual contigs produced by *de novo* assembly, usage of assembly assembler script, and details of contigs length are presented in **Supplementary Table 3**. Virus derived sRNA profiles were obtained by mapping raw sRNA reads to the final viral contigs (**Figure 1**). A similar read length distribution was observed between the data sets and viruses. The highest proportion of reads were 21 nts in length while 20 and 22 nts peaks were also prominent (**Figure 1**).

The BlastX search revealed a high number of potential novel viruses from different families in each of the four isolates (**Supplementary Tables 4–7**). After excluding contigs that could represent the host genome with the aid of the BlastN search, 30 contigs of isolate VN999 (**Supplementary Table 4**), 74 contigs of VN1004 (**Supplementary Table 5**), 31 contig of VN1008 (**Supplementary Table 6**), and 61 contig of VN1012 (**Supplementary Table 7**) were selected based on their similarity to known viruses. Based on Blast homology searches, all of the four studied isolates were found to be infected by viruses sharing genomic affinities with viruses of the order *Bunyavirales* and families *Megabirnaviridae*, *Narnaviridae,* and *Totiviridae*. In addition, the BlastX search revealed contigs of the proposed virus family "Fusagraviridae" (Wang et al., 2016) in VN1004 and *Endornaviridae* in VN1008.

Identification of Final Viral Sequences in Total RNA

Isolate VN999 was selected for carrying out stranded total RNA sequencing, which generated 127,332,406 reads. Raw data stats are displayed in **Supplementary Table 8**. *De novo* assembly of reads cleaned from the host produced 100,903 contigs with a minimum length of 201 nts and a maximum length of 21,824 nts. Blast comparison of the contigs revealed that isolate VN999 hosts several novel putative viruses with sequence similarities to known members of the order *Bunyavirales*, family *Megabirnaviridae*, family *Narnaviridae*, family *Totiviridae,* and the proposed family "Fusagraviridae" (Wang et al., 2016). Five viruses (one representative per viral family) from this data set were chosen for further characterization (**Table 1**). The accuracy of the final virus contigs was confirmed by Sanger sequencing of near full-length coding regions (**Supplementary Table 9**).

The Six Novel Viruses Hosted by *Phytophthora castaneae* Isolates

The RT-PCR with virus-specific primers was utilized to confirm that the four *P. castaneae* isolates included in the sRNA analyses were infected with five putative virus species in addition to a sixth virus hosted by one of the *P. castaneae* isolates. The Trinity assembly of the total RNA library contained the complete coding virus sequences of the five putative viruses showing similarity with members of *Bunyavirales*, *Megabirnaviridae*, *Narnaviridae*, *Totiviridae,* and "Fusagraviridae" (**Table 1**). Furthermore, virus sequences having similarity with *Endornaviridae* were detected in the sRNA of VN1008, and a partial sequence of a novel endorna-like virus was obtained with Sanger sequencing (**Table 1**).

Megabirna-Like Virus

Following the pipeline in which host reads were removed prior to *de novo* assembly, one single contig of 6,006 nts was detected. Following the other pipeline where the host reads were not removed, a slightly bigger contig of 6,337 nts was retrieved from the analyzed data. The two putative virus sequences are identical in ORFs regions. The two contigs differ in four nts in the 3' UTR region. The sequence of 6,337 nts was chosen as a representative sequence for Phytophthora castaneae RNA virus 1 (PcaRV1). The putative virus shows similarities to the large (L) segment of megabirnaviruses and has a 60.2% GC content, two overlapping ORFs likely coding for the CP (ORF 1–5′ proximal) and the RdRp (ORF2-3′ proximal; **Figure 2**).

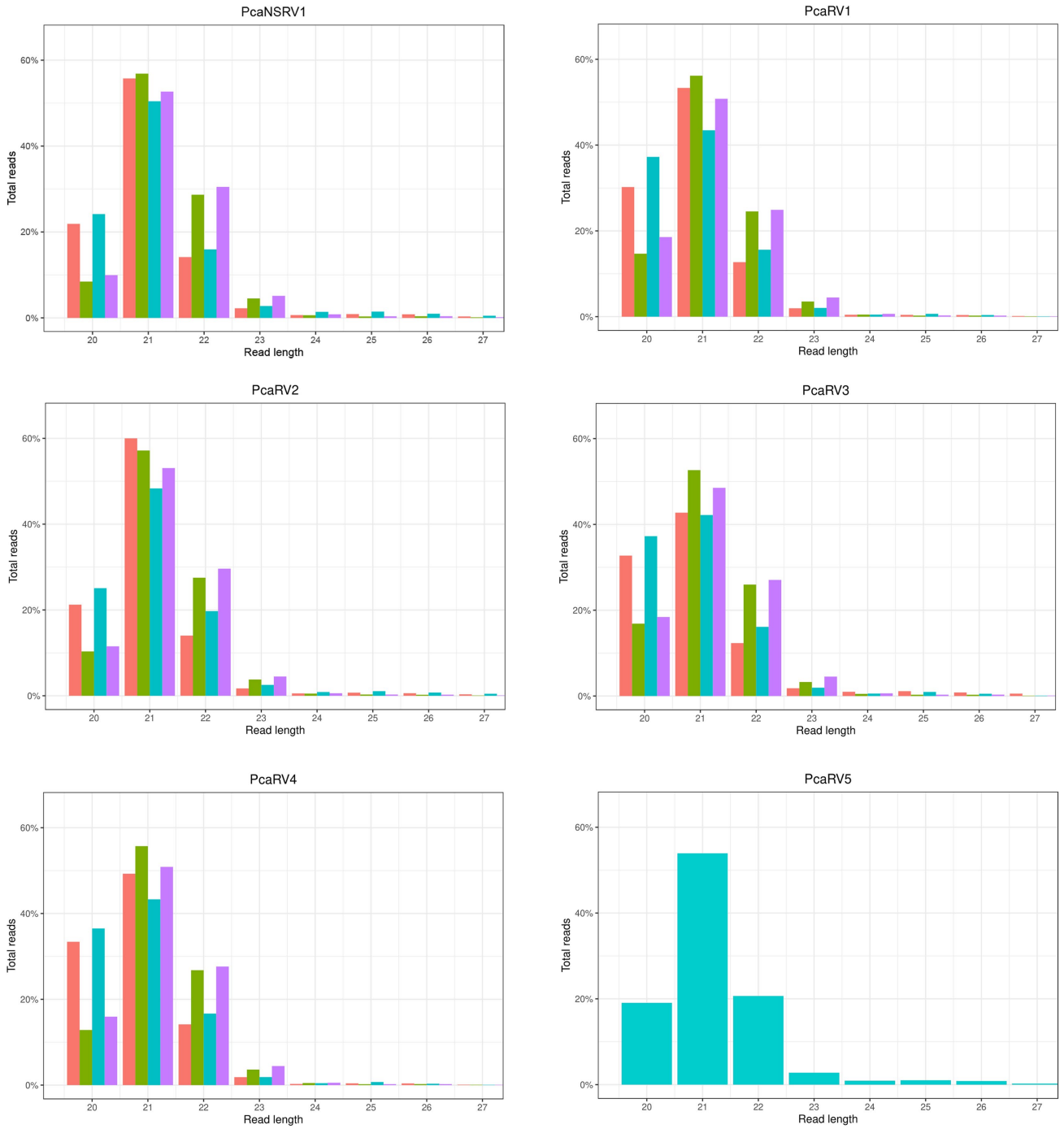

FIGURE 1 | sRNA size distribution of raw VN999 ,VN1004 , VN1008 , and VN1012 reads, respectively, after mapping to the final genomes of viruses PcaNSRV1 and PcaRV1-5, respectively. Mapping was done in Geneious Prime® 2020.2.3 using Bowtie mapper. The final graphs were constructed in R Core Team (2020). Sizes of sRNA profiles starting from 28 nt that were represented with less than 1.5% of total reads per virus were omitted during graph preparation for a better visualization.

The overlapping region consists of 4 nts (AUGA) including the stop codon of the ORF 1 and the start codon of the ORF2. Interestingly, two continuous methionine codons occur at the beginning of the ORF2. No typical heptanucleotide slippery site or shifty heptamer motif facilitating the −1 programmed ribosomal frameshifting was observed at the end of the ORF1 with the general sequence X XXY YYZ (spaced triplets represent pre-frameshift codons; Bekaert et al., 2005)

TABLE 1 | Identification of the viruses found by total RNA seq in *P. castaneae* isolate VN999 and a virus found in VN1008 by sRNA seq and the most similar sequences in GenBank based on a BlastX search.

Virus name	Acronym[1]	Length[2]	GenBank[3]	Putative family/ Order[4]	BlastX best hit[5]	E value	QC(%)[6]	I (%)[7]	Identified by BlastX search of Velvet contigs (yes/no)
Phytophthora castaneae RNA virus 1	PcaRV1	6,337	MZ269516	*Megabirnaviridae*	RdRp Charybdis toti-like virus (DAZ87259.1)	$2e^{-59}$	34	29.08	yes
Phytophthora castaneae RNA virus 2	PcaRV2	2,891	MZ269517	*Narnaviridae*	RdRp Sanya narnavirus 11 (UHM27569.1)	$2e^{-102}$	81	32.20	yes
Phytophthora castaneae RNA virus 3	PcaRV3	5,470	MZ269518	*Totiviridae*	putative RdRp polymerase Totiviridae sp. (UHS72506.1)	0.0	39	50.75	yes
Phytophthora castaneae RNA virus 4	PcaRV4	6,884	MZ269519	*"Fusagraviridae"*	ORF2 Bremia lactucae associated fusagravirus1 (QIP68010.1)	$7e^{-47}$	31	28.15	no*
Phytophthora castaneae RNA virus 5	PcaRV5	2,182[x]	ON131873	*Endornaviridae*	hypothetical polyprotein Diatom colony associated dsRNA virus 15 (YP_009552081.1)	$2e^{-123}$	99	34.82	yes
Phytophthora castaneae negative-stranded RNA virus 1	PcaNSRV1	8,345	MZ269515	*Bunyavirales*	polyprotein Phytophthora cactorum bunyavirus 1 (QUA12643.1)	0.0	97	57.16	yes

[1]*Acronym of Phytophthora castaneae RNA virus 1–5 and Phytophthora castaneae negative-stranded RNA virus 1.*
[2]*Length, virus sequence length obtained through Trinity assembly.*
[3]*GenBank accession number of each virus.*
[4]*Putative Family/Order, placement of the viruses according to their genome organization and phylogenetic analysis.*
[5]*The most similar virus according to BlastX search when the whole virus sequence is blasted against the (nr)protein sequence database in the online version of NCBI BlastX (date of the last search: 01.04.2022).*
[6]*QC, Query cover.*
[7]*I, Identity.*
**PcaRV4 was identified by BlastX search of Velvet contigs only in one isolate (VN1012).*
[x]*Sequence of PcaRV5 is surely partial. The date of the last BlastX search 01.04.2022.*

where X represents A/G/C/U, Y defines A/U, and Z represents A/C/U. Likewise, neither a possible slippery site similar to those found in Sclerotinia sclerotiorum megabirnavirus 1 (SsMBV1; AAAAAAC; Wang et al., 2015) or Fusarium pseudograminearum megabirnavirus 1 (FpgMB1; GAAAAAC; Zhang et al., 2018) was detected. Also, no clear H-type pseudoknot was found upstream of the ORF2. According to the BlastX search, PcaRV1 shares the highest identity with Charybdis toti-like virus (**Table 1**), a divergent virus with uncertain placement inside *Ghabrivirales* and undetermined genome organization due to lack of a full sequence being available (Charon et al., 2021). PcaRV1 shows sequence homology also to the FpgMB1 (query cover: 28%; E value:$1e^{-51}$; identity; and 31.30% accession: AYJ09269.1). The putative dsRNA1 of PcaRV1 shows a 29.15% identity (query cover: 25%; E value: 9 e^{-50}; and accession: QTF98696.1) to the Rosellinia necatrix megabirnavirus 2 (RnMBV1). With Rhizoctonia solani RNA virus HN008 (RsRV-HN008), PcaRV1 shares 26.60% identity in its RdRp region (query cover: 22%; E value: 6 e^{-27}; and accession: YP_009158860.1). The second segment of PcaRV1 was not detected based on the initial BlastX search against the virus protein database. Thus, the dsRNA2 was searched among Trinity contigs encoding for proteins having no detectable homology with proteins of the NCBI nr database (e.g., ORFan contigs; **Supplementary Table 10**), but no reliable dsRNA2 candidate could be retrieved. The CDD search detected two conserved domains in both ORFs (**Figure 2**). In the ORF2, a RdRp domain RT_like family (E value 6.08 e^{-20}; interval 4,215–5,243 nts) was detected. Besides, in the interval from 1,547 to 3,028 nts of the ORF1 with unknown function, the search revealed a significant record (E value 1.09 e^{-04}) with the provisional large tegument protein UL36 (PHA03247; PHA03247), which is the largest protein of herpesviruses. Multiple alignment of the RdRp aa sequence of the PcaRV1 with those of other related viruses enabled detection of conserved sequence motifs (I-VIII; **Supplementary Figure 3**). To determine the phylogenetic position of PcaRV1, the RdRp protein sequence of this virus was compared with RdRp sequences of other related unclassified viruses and viruses of families *Totiviridae*, *Megabirnaviridae,* and proposed family "Fusagraviridae." Phylogenetically, PcaRV1 is related to unclassified megabirnaviruses and, together with them, putatively falls into the *Megabirnaviridae* virus family (**Figure 3**). Megabirnaviruses are found to infect filamentous fungi (Sato et al., 2019). PcaRV1 would be the first megabirnavirus discovered in oomycetes and the genus *Phytophthora*, respectively.

Narna-Like Virus

The final sequence of Phytophthora castaneae RNA virus 2 (PcaRV2) is 2,891 nts long, coding a single protein, putatively RdRp (**Figure 2**). The sequence has a GC content of 49.2% and one single ORF putatively corresponding to the RdRp. The ORF size is 2,613 nts or 870 aa. No conserved domains were detected within the PcaRV2 genome sequence (**Figure 2**). However, when the aa sequence

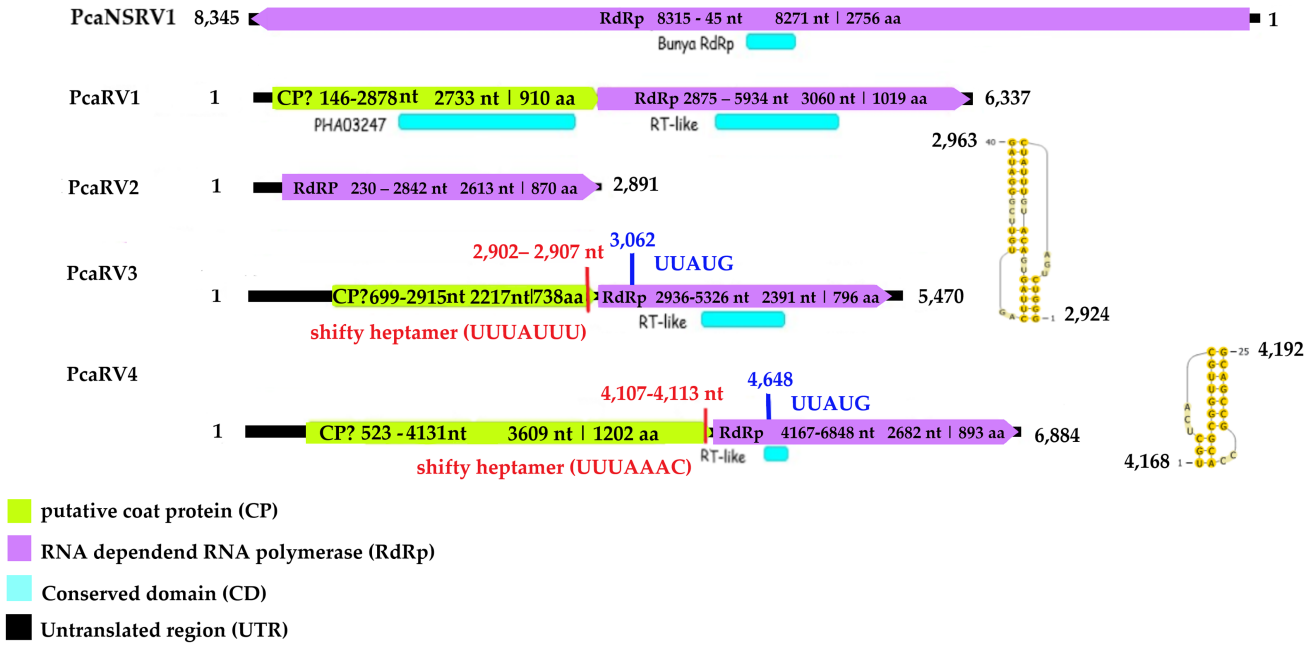

FIGURE 2 | Genome organization of viruses PcaRV1-4 and PcaNSRV1. Genome sequence of PcaRV5 is not graphically represented as it is partial and requires further characterization.

of the ORF1 was aligned to RdRp of other narnaviruses, conserved motifs from III to VII were detected (**Supplementary Figure 4**). A phylogenetic tree was constructed using (aa) sequences of the RdRp of PcaRV2 together with 24 other (+)ssRNA viruses grouped under the families *Narnaviridae*, *Mitoviridae*, *Botourmiaviridae*, and *Leviviridae* (**Figure 4**). The ML tree shows that PcaRV2 clusters with narnaviruses and is phylogenetically close to Phytophthora infestans RNA virus 4 (**Figure 4**). According to the BlastX results, the most similar virus to PcaRV2 is the Sanya narnavirus 11 (UHM27569.1; **Table 1**).

Toti-Like Virus

The final contig corresponding to the genome of Phytophthora castaneae RNA virus 3 (PcaRV3) has a length of 5,470 nts and 59.3% GC content. It has two overlapping ORFs, ORF1 (3'end-proximal) putatively encoding the CP and ORF2 (5′ end-proximal) corresponding to the RdRp. Conserved motifs of the RdRp (I-VIII) have been detected by the CDD search (**Figure 2**; **Supplementary Figure 3**). The ORFs overlap by 70 nts (nucleotide (nt) positions 2,846–2,915), which include the stop codon for ORF 1 (UAG) and the start codon for ORF2 (GUG, Valine). GUG is an alternative non-canonical codon that, under certain circumstances, can also be recognized as an initiator codon in RNA viruses. A heptanucleotide slippery site or shifty heptamer motif was observed at the end of the ORF1 (UUUAUUU, nt positions 2,902– 2,907), which may facilitate −1 programmed ribosomal frameshifting in PcaRV3 transcripts. A short spacer region of 11 nts appears to precede an H-type pseudoknot between nt 2,924 and 2,963 (**Figure 2**) from the slippery site. Upstream at nt position 3,062 UUAUG is found, a key motif for translation of the downstream ORF. PcaRV3 genome organization and its phylogenetic grouping (**Figure 3**) indicate that PcaRV3 belongs to *Totiviridae*, a family of linear uncapped dsRNA viruses approximately 4.6 to 7.0 kb in size (Wickner et al., 2012). According to Blast analyses, PcaRV3 is most similar in its RdRp region to Totiviridae sp. (UHS72506.1; **Table 1**) but it also shares 49.34% of identity with RdRp of Pythium polare RNA virus 1 (YP_009552275.1; Sasai et al., 2018), 33.62% with Plasmopara viticola lesion associated totivirus-like 1 (QGY72634.1; Chiapello et al., 2020), 35.01% with Phytophthora condilina RNA virus 1 (QTT60989.1; Botella and Jung, 2021), and 32.86% with Phytophthora cactorum RNA virus 1 (QUE45741.1; Poimala et al., 2021) and to other toti-like viruses (data not shown). The phylogenetic tree shows that PcaRV3 clusters with Phytophthora condilina RNA virus 1 and Pythium polare RNA virus 1 (**Figure 3**).

Fusagra-Like Virus

The final sequence of a virus designated Phytophthora castaneae RNA virus 4 (PcaRV4) is 6,884 nts long. PcaRV4 has two large discontinuous ORFs (**Figure 2**), ORF1 (putative CP, nt 517–4,131) and ORF2 (RdRp, nt 4,167–6,848). The elements needed to accomplish −1 ribosomal frameshifting in some RNA viruses were found. Shifty heptamer (UUUAAAC) at nts 4,107 to 4,113; a spacer region (54 nts from the slippery site) the recording stimulatory element consisting of an H-type pseudoknot structure between nts 4,168 and 4,192, and finally, 534 nts downstream at nt 4,648, the UUAUG motif. A putative conserved domain containing RNA-dependent RNA polymerase superfamily 4 (pfam02123: RdRP_4) was detected by CDD search in translated ORF2 at the position from 4611 to 5822 nts (E value $8.37e^{-30}$; **Figure 2**; **Supplementary Figure 3**).

FIGURE 3 | Phylogenetic RAxML tree including the three dsRNA viruses described in this study, PcaRV1 (megabirna-like virus), PcaRV3 (toti-like virus), and PcaRV4 (fusagra-like virus; all three indicated by a yellow star), and related members belonging to the families *Megabirnaviridae*, *Toriviridae* proposed family "Fusagraviridae." The viruses described from oomycetes are indicated by a green hexagone . Nodes are labeled with bootstrap percentages ≥50% only. Family classification and the corresponding pBLAST accession numbers are shown next to the virus names. The tree is rooted in the midpoint. Branch lengths are scaled to the expected underlying number of amino acid substitutions per site. The scale bar is indicatig 2.0 aa substitutions per site, per branch.

Pfam02123 family includes RdRp proteins of members of the genera *Luteovirus*, *Totivirus,* and *Rotavirus* and it is a member of the superfamily cl02808. According to the BlastX search, PcaRV4 is associated with proteins of viruses putatively belonging to the newly proposed family "Fusagraviridae." The ORF1 of PcaRV4 is most similar to Bremia lactucae associated fusagravirus 1 ORF2 (QIP68010.1) with 28.15% identity and 59% coverage (E value $1e^{-47}$), hypothetical protein 1 of Wuhan insect virus 28 (YP_009342429.1) with which it shares 22.48% similarity (query coverage 41%; E value $5e^{-18}$) and to hypothetical protein EXH54_gp1 of Phytophthora infestans RNA virus 3 (YP_009551327.1), with 29.96% but less coverage (23%). ORF1 is also similar to Phytophthora infestans RNA virus 3 (YP_009551328.1), RdRp of Bremia lactucae associated fusagravirus 2 (QIP68015.1), hypothetical protein of Macrophomina phaseolina fusagravirus 1 (QDM35289.1), and other viruses. In its RdRp region, PcaRV4 is most similar to RdRp of Bremia lactucae associated fusagravirus 1 (QIP68009.1), with which it shares 29.96% (query coverage 59%; E value $1e^{-45}$) but also with RdRp of Phytophthora infestans RNA virus 3 (YP_009551328.1), RdRp of Spissistilus festinus virus 1 (YP_003800001.1), and other viruses. The whole sequence is most similar to ORF2 of Bremia lactucae associated fusagravirus 1 (QIP68010.1; **Table 1**). The phylogenetic tree including viruses of the families *Megabirnaviridae*, *Totiviridae,* and putative members of proposed "Fusagraviridae" (**Figure 3**) revealed that PcaRV4 clusters with the fusagraviruses but groups separately with PiRV3, indicating a close phylogenetic relationship with this virus, also described from *Phytophthora*. When short sRNA reads of the four libraries were mapped individually to the PcaRV4 genome, high coverage was achieved. However, due to its low similarity with known viruses, the BlastX search failed in identification of the short contigs representing PcaRV4 produced by Velvet in isolates VN999, VN1008, and VN1012, respectively. By BlastX, such contigs were identified only in isolate VN1004, although the genomic sequence of PcaRV4 was obtained from Trinity assemblies of isolate VN999 and

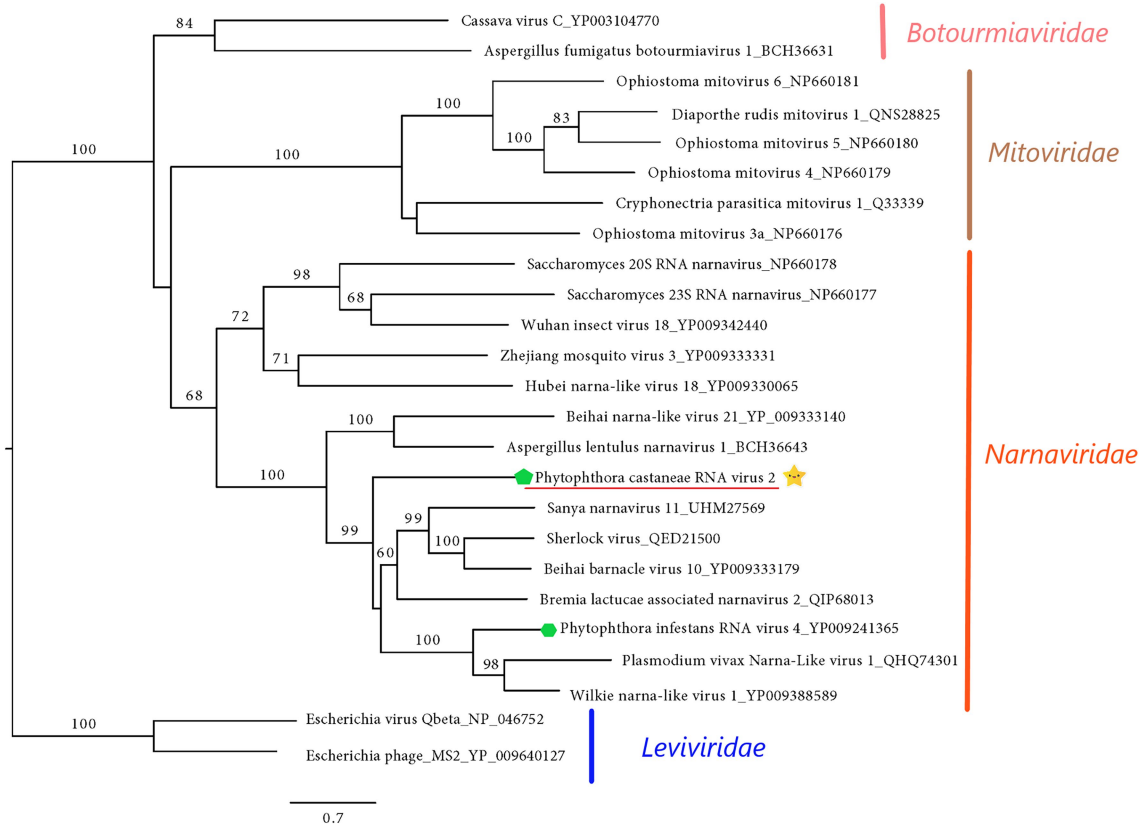

FIGURE 4 | Maximum likelihood tree (RAxML) showing the phylogenetic relationships of PcaRV2 (indicated by the yellow star) with other (+)ssRNA viruses belonging to families *Narnaviridae*, *Mitoviridae*, *Leiviviridae*, and *Botourmiaviridae*. The viruses described from oomycetes are indicated by a green hexagone. Nodes are labeled with bootstrap percentages ≥50%. Branch lengths are scaled to the expected underlying number of amino acid substitutions per site. Family classification and the corresponding pBLAST accession numbers are shown next to the virus names. The tree is rooted in the family *Leviviridae*.

almost its entire length was confirmed by RT-PCR and Sanger sequencing in all four Vietnamese isolates.

Bunya-Like Virus

The nearly complete (−)ssRNA genome of a virus designated Phytophthora castaneae negative-stranded RNA virus 1 (PcaNSRV1) was 8,345 nts long, with one single 8,271 nts long ORF (nt 8,315 to 45), putatively coding for a polyprotein/RdRp (**Figure 2**) and L segment of viruses in the order *Bunyavirales*. The phylogenetic analysis based on the partial RdRp (aa) sequence of PcaNSRV1 and those of related viruses revealed that PcaNSRV1 clusters with unclassified bunyaviruses of other *Phytophthora* species, i.e., *P. cactorum* (Poimala et al., 2021), *P. condilina* (Botella and Jung, 2021), and those described from the genus *Halophytophthora* (Botella et al., 2020; **Figure 5**). According to the BlastX search, PcaNSRV1 is most similar to Phytophthora cactorum bunyavirus 1, with whom it shares 57.16% identity (**Table 1**). PcaNSRV1 is also similar to Phytophthora condilina negative-stranded RNA virus 5 (PcoNSRV5; query coverage 99%; E value 0.0; and percent identity 52.14%), Halophytophthora RNA virus 6 (HRV6; query coverage 68%; E value 0.0; and percent identity 62.47%), Phytophthora cactorum bunyavirus 2 (PcBV2; query coverage 96%; E value 0.0; and percent identity 46.83%), and others. Conserved motifs of bunya RdRp (pfam04196), the only member of superfamily cl20265, were detected in the region from 945 to 1415 aa (E value $6.16e^{-07}$; **Figure 2**; **Supplementary Figure 5**). In addition, pfam15518: L_protein_N endonuclease domain was detected at aa position 245–298 (E value $4.18e^{-03}$). Many bunyavirus L proteins have this domain at the N-terminus (Reguera et al., 2010).

Endorna-Like Virus

As the RNA of isolate VN1008 was sequenced only by HTS of sRNAs, and contigs of viruses similar to *Endornaviridae* were not detected in VN999 *de novo* assembled total RNA reads, contigs 139, 274, and 295 (**Supplementary Table 6**) were used for primer design and Sanger sequencing of 2,182 nts long sequence resembling members of *Endornaviridae*. Virus-specific primers (**Supplementary Table 2**) were designed to fill sequence gaps between the short contigs, and the amplicons obtained through RT-PCR were Sanger sequenced in both directions allowing sequence identification of a longer sequence of a virus here Phytophthora castaneae RNA virus 5 (PcaRV5). PcaRV5 shares the most similarity with Diatom colony associated dsRNA virus 15 (DcaRV15; **Table 1**). It

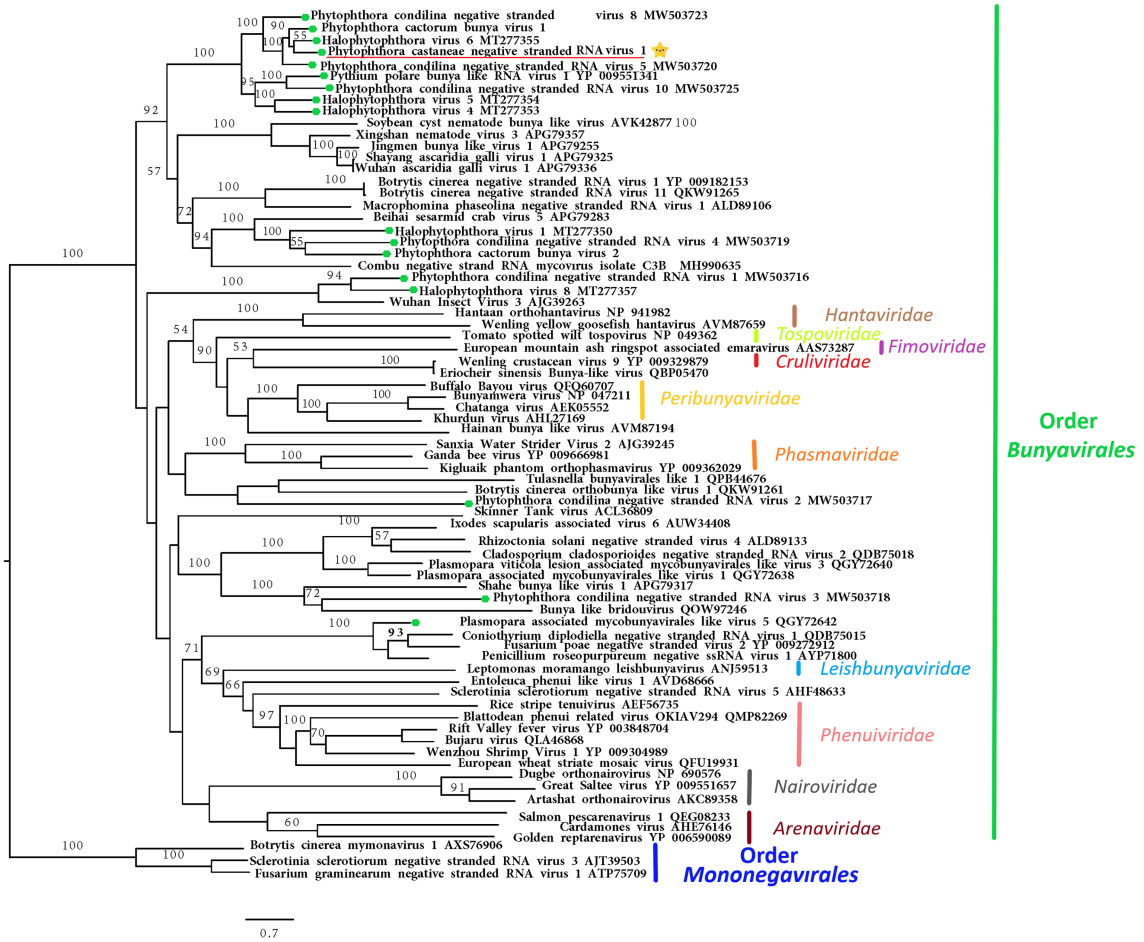

FIGURE 5 | Maximum likelihood tree (RAxML) depicting the phylogenetic relationship of the predicted RdRp of PcaNSRV1 (indicated by the yellow star) with other complete RdRp belonging to related (–)ssRNA viruses from the orders *Bunyavirales* and *Mononegavirales*. The viruses described from oomycetes are indicated by a green hexagone. Nodes are labeled with bootstrap support values. Branch lengths present calculated evolutionary distannce and are scaled to the expected underlying number of amino acid substitutions per site. Nodes are marked with bootstrap percentages ≥50% only. The tree is rooted in *Mononegaviales* viruses, which are classified within the family *Mymonaviridae*. Family classification and the corresponding pBLAST accession numbers are shown next to the virus names. The scale bar is indicatig 0.7 aa substitutions per site, per branch.

is 33.42% identical to the polyprotein of Erysiphe cichoracearum alphaendornavirus (YP_009225663.1; query coverage 99%; E value $6e^{-119}$), 48.92% to the polyprotein of Chalara endornavirus CeEV1 (ADN43901.1; query coverage 57%; E value $1e\text{-}^{118}$), and 46.14% to the polyprotein of Phytophthora cactorum alphaendornavirus 3 (QUA12642.1; query coverage 67%; E value $5e^{-117}$). Conserved domain of pfam00978 RdRP_2 super family (E value $2.88e^{-08}$; accession cl03049) was detected at interval from 1,544 to 2,014 bp of nucleotide sequence. A longer sequence of PcaRV5 was not determined and the putative virus was not further characterized.

Intraspecific Genetic Variability of *Phytophthora castaneae* Viruses

The intraspecific genetic variability of PcaRV1-4 and PcaNSRV1 hosted by the four *P. castaneae* isolates was examined using Sanger sequencing. The intraspecific genetic variability among virus strains present in the four isolates of *P. castaneae* appeared to be very low and homogeneous. Thus, the overall pairwise (pw) sequence comparison (PASC) percentage between PcaRV1 nucleotide sequences (strains) was 99.8% for an alignment of 5,985 bp; in PcaRV2 strains alignment (length: 1,625 bp), it was 98.8%; for PcaRV3 strains, it was 98.9% (alignment length: 4,306 bp); for PcaRV4, it was 98.5% (alignment length: 1,895 pb); and for PcaNSRV1 strains, PASC had the lowest value, 95.6% (alignment length: 7,858 bp). Within PcaNSRV1, the highest pw variability was shown between the strain VN1008 and the rest of the strains (PASC 91.21–91.23%), while the pw similarity between VN999, VN1004, and VN1012 ranged 99.89–99.97%.

Depth of Coverage

The high number of reads and depth of coverage show that the five virus genomic sequences obtained from Trinity assembly

are properly supported. PcaRV1 had the highest number of reads mapped to its almost complete genome in all four sRNA datasets, as well as the deepest coverage in all sRNA libraries, but one (**Table 2**). Overall, PcaNSRV1 is the virus with the lowest coverage depth, although it did not have the lowest number of reads in all the four data sets. In the library of isolate VN999, PcaRV3 showed the smallest number of reads mapped to its genome in comparison to other viruses (**Table 2**). According to total RNA seq, the virus with the most reads covering its genome was PcaRV1 that also had the deepest coverage, while PcaRV4 was covered the least (**Table 2**).

DISCUSSION

A three-method approach including two HTS methods, small and stranded total RNA sequencing, and the conventional dsRNA extraction method was employed in order to characterize and have a better understanding of the viral community hosted by the oomycete pathogen *P. castaneae*.

In the current study, the dsRNA detection technique provided first insights into the potential occurrence of viruses and, based on the presence of dsRNA banding patterns, isolates were chosen for further sequencing analyses. Linking sizes of dsRNA molecules with the described viruses was attempted, but the patterns were not consistent through all performed extractions (**Supplementary Figure 1**). Furthermore, due to the similar virus genome sizes, i.e., PcaRV1 and PcaRV4 larger segments of 6,337 bp and 6,884 bp, respectively, it would have been difficult to confirm the corresponding dsRNA bands of each virus.

Concerning the two tested HTS approaches, RNA-Seq was found superior to sRNA sequencing for virus detection and identification in *P. castaneae*. While sRNA sequencing was found not to be the best method for that purpose, this study demonstrates that *P. castaneae* hosts a diverse virome processed into sRNAs, suggesting that *P. castaneae* has an active RNAi mechanism. To our knowledge, *P. castaneae* is the first *Phytophthora* species reported with such a diverse virome and high number of viruses from different viral families co-occurring in a single isolate, although an abundant occurrence of viral co-infections and extreme virus prevalence among host isolates was also found in *P. cactorum* (Poimala et al., 2021). Three of the analyzed *P. castaneae* isolates (VN999, VN1004, and VN012) appear to host viruses belonging to the families *Megabirnaviridae*, *Narnaviridae*, *Totiviridae*, "Fusagraviridae," and the order *Bunyavirales*. However, in isolate VN1008, in addition to the above-mentioned viruses, the occurrence of a putative endornavirus was also detected by sRNA sequencing. It is common for mycoviruses to co-infect and interact in a variety of ways, forming interactions from synergistic to neutral and antagonistic (Thapa and Roossinck, 2019). Multiple virus infections from more than one virus family have been previously reported in *P. infestans* (hosting a narnavirus and a virus related to the proposed family "Fusagraviridae"; Cai et al., 2012, 2013), *P. condilina* (toti-like and bunya-like viruses; Botella and Jung, 2021), and *P. cactorum* (bunya-like, toti-like, alpahendronaviruses, and an usti-like virus; Poimala et al., 2021). Likewise, multiple

TABLE 2 | Information about the depth of coverage and the total number of sRNA and total RNA reads mapped to each virus.

Virus[I]	Read length total RNA seq VN999	Reads in total RNA seq VN999[II]	Depth of Coverage total RNA seq (VN999)[x]	Read length sRNA seq (VN999 VN1008)[I,II]	Read length sRNA seq (VN1004 VN1012)[I,II]	Reads sRNA seq (VN999)[II]	Depth of coverage in VN999*	Reads sRNA seq (VN1004)[II]	Depth of coverage in VN1004*	Reads sRNA seq (VN1008)[II]	Depth of coverage in VN1008*	Reads sRNA seq (VN1012)[II]	Depth of coverage in VN1012*
PcaRV1	101	2,427,087	38,683	36	50	217,861	1,238	226,485	1,787	168,840	959	295,192	2,329
PcaRV2	101	250,248	8,743	36	50	58,916	734	68,560	1,186	121,361	1,511	102,384	1,771
PcaRV3	101	777,368	14,354	36	50	18,935	125	59,923	548	13,679	90	49,552	453
PcaRV4	101	60,512	889	36	50	72,976	382	136,977	995	73,740	386	183,367	1,332
PcaNSRV1	101	285,239	3,453	36	50	24,637	106	21,332	128	6,624	29	25,261	151

[I]*Virus, Acronym of a virus.*
[II]*Reads sRNA seq*, the number of reads that mapped to each virus.*
[III]*Read length (insert size) of sRNA and total RNA libraries.*
[x]*The depth of coverage was calculated using a formula specified in the Materials and Methods section.*

virus infections seem to be common in other oomycete species including *Py. polare* hosting dsRNA (toti) and (−)ssRNA (bunya) viruses (Sasai et al., 2018), and *Pl. viticola*-associated lessions in which 283 novel viruses resembling genomes of (+)ssRNA, (−) ssRNA, and dsRNA from different virus families and ORFan RdRp viral segments were identified using a metagenomics approach. Interestingly, members of the families *Narnaviridae* [(+)ssRNA], *Totiviridae* (dsRNA), and the order *Bunyavirales* [(−)ssRNA] have been described in the majority of oomycete species screened for viruses so far, suggesting a strong coevolution history of these virus clades with oomycetes. Multiple viral infections are also reported from various fungal plant pathogens (Doherty et al., 2007; Deakin et al., 2017; Botella and Hantula, 2018; Hillman et al., 2018; Zhu et al., 2018; Hantula et al., 2020) but, as indicated above, they seem to be particularly prevalent in oomycetes. As oomycete hyphae lack septa, it has been speculated that viruses are able to passively move with the flowing cytoplasm through the mycelium (Poimala et al., 2021).

This is the first study to report a megabirna-like virus on the genus *Phytophthora*, or any other oomycete genus. PcaRV1 appears to be the first megabirna-like virus detected outside of the kingdom of Fungi. *Megabirnaviridae* is a family of viruses with non-enveloped linear dsRNA genomes organized in two segments, dsRNA 1 (8.9 kb) and dsRNA 2 (7.2 kb; in total 16.1 kb), each with two ORFs. The only known member of the family *Megabirnaviridae* and its single genus *Megabirnavirus* is Rosellinia necatrix megabirnavirus 1 (RnMBV1; Chiba et al., 2009; Sato et al., 2019). Since the genus is monospecific, species demarcation criteria are not yet defined, and no other viruses have been assigned to the family or the genus (Sato et al., 2019). Despite the description of many viruses in other filamentous fungi that share similar characteristics and phylogenetic positions to RnMBV1 such as SsMBV1 (Wang et al., 2015), Rosellinia necatrix megabirnavirus 2 (Sasaki et al., 2016), Pleosporales megabirnavirus 1 (Nerva et al., 2016), Entoleuca megabirnavirus 1 (Velasco et al., 2018), Rosellinia necatrix megabirnavirus 3 (Arjona-Lopez et al., 2018), Rhizoctonia solani megabirnavirus 1 (Bartholomäus et al., 2016), RsRV-HN008 (Zhong et al., 2015), and FpgMB1, these viruses remain unclassified (Sato et al., 2019). The genomic properties of PcaRV1 are similar to other reported megabirnaviruses except for the seemingly smaller size (*ca.* 6.4 kb) of its larger segment (dsRNA1). However, as the 3′- and 5′- termini were not confirmed, we cannot rule out the possibility of the larger segment being longer. The putative smaller segment (dsRNA 2) of PcaRV1 was not detected in the total RNA seq data. Similarly, the second segment has not been reported from some other megabirnaviruses. For example, RsRV-HN008, characterized from *Rhizoctonia solani* J.G. Kühn, has a genome of 7,596 nts organized on one segment containing two non-overlapping coding regions (Zhong et al., 2015). Detection of PcaRV1 by sRNA seq suggests it is targeted by the oomycete RNAi system.

Phylogenetic analyses of this study, based on the partial RdRp sequences, placed virus PcaRV3 from *P. castaneae*, together with putative members of the virus family *Totiviridae*. PcaRV3 also has a genome organization and size typical for a toti-like virus. According to the International Committee on Taxonomy of Viruses (ICTV) 2021 report,[7] the family is organized into five genera, i.e., *Giardiavirus*, *Leishmaniavirus,* and *Trichomonasvirus,* all associated with infections of protozoa, and *Totivirus* and *Victorivirus* infecting fungi. However, many viruses with similarities to totiviruses are unclassified and the host range is much wider than previously assumed (De Lima et al., 2019). Totiviruses are correlated with latent infections in fungal and protozoan hosts (Wickner et al., 2012), but they have also been reported from fish, plants, insects, and oomycetes (Koyama et al., 2015; Chen et al., 2016; Martinez et al., 2016; Mor and Phelps, 2016; Sasai et al., 2018; De Lima et al., 2019; Poimala and Vainio, 2020; Botella and Jung, 2021; Poimala et al., 2021). *Totiviridae* members have distant phylogenetic relationships with the megabirnaviruses, and both belong to the order *Ghabrivirales* (Sato et al., 2019). Similarly, to the dsRNA1 of megabirnaviruses, the RdRP gene of totiviruses of the genera *Totivirus* and *Victorivirus* is encoded downstream of the CP gene. The *Totiviridae* whose dsRNA genomes are not divided are phylogenetically distinct from megabirnaviruses. Coding strands of these totiviruses have shorter UTR at 5′ (<0.5 kb) and smaller virus particles of *ca.* 40 nm diameter (Sato et al., 2019).

The putative new fusagravirus, PcaRV4, has a relatively long 5′ UTR of 522 bp and a small 36 nts long 3′ end UTR. The 5' UTR of fusagraviruses ranges in size from 865 to 1,310 bp, while 3′ UTR can be from 7 to 131 bp long. PcaRV4 is slightly smaller (6,884 bp) than other viruses putatively grouped in the proposed "Fusagraviridae" family, but the full length of its genome is not confirmed. Genome organization and phylogenetic placement of PcaRV4, as well as the occurrence of programmed −1 ribosomal frameshifting and presence of putative shifty heptamer motif located immediately upstream of the stop codon of ORF1, are in accordance with other "Fusagraviridae" members (Wang et al., 2016). PcaRV4 was also detected by sRNA sequencing, suggesting that its presence in the cell activates the RNAi mechanism of *P. castaneae*.

Bunyavirales is an order of (−)ssRNA viruses accommodating 12 families infecting a broad range of hosts including plants, vertebrates, and invertebrates, and includes the genus *Coguvirus* unassigned to any viral family with a single species Citrus coguvirus associated with citrus disease (Abudurexiti et al., 2019). The number of genomic segments of bunyaviruses can vary from two to six. However, most commonly, their genomes are organized in three ss segments, designated according to their size: the small (S; *ca.* 1–2 kb), medium (M; *ca.* 3.7–5 kb), and the large (L) segment (*ca.* 6.8–12 kb; Garrison et al., 2020; Hughes et al., 2020; Leventhal et al., 2021). The L segment codes for a single protein, the RdRp. The M segment codes for Gn and Gc glycoproteins and sometimes a non-structural protein (NSm). The S segment codes for nucleocapsid protein (N) and, typically, a non-structural protein (NSs). The RNA genomes of these viruses are coated with N protein which,

[7]https://talk.ictvonline.org/

together with L protein, form ribonucleoprotein complexes (RNPs; Reguera et al., 2010). PcaNSRV1 codes for a single protein, the RdRp, representing the L segment. No other segments of this virus, except the L segment, were retrieved from either total RNA or sRNA sequencing data after BlastX and BlastN searches. The absence of the putative NP and other non-structural (Ns) associated proteins may indicate their low copy numbers or low level of conservation, making it difficult to detect them through a homology search (Botella and Jung, 2021). The size and genome organization of PcaNSRV1 are in accordance with other *Bunyavirales* members and bunya-like viruses found in *P. cactorum* (Poimala et al., 2021), *P. condilina* (Botella and Jung, 2021), and *Halophytophthora* sp. (Botella et al., 2020), in which no segments other than L were identified.

According to the ICTV (see footnote 7), the family *Narnaviridae* has only one genus *Narnavirus*, containing only two species, *Saccharomyces 20S narnavirus* and *Saccharomyces 23S narnavirus* described from the yeast *Saccharomyces cerevisiae* Meyen ex E.C. Hansen. Besides these two officially recognized viral species, a number of narna-like viruses have been described in a range of other organisms including arthropods (Shi et al., 2016; Harvey et al., 2019), algae (Waldron et al., 2018), fungi (Osaki et al., 2016), oomycetes (Cai et al., 2012), mosquitoes (Cook et al., 2013; Chandler et al., 2015; Shi et al., 2017; Göertz et al., 2019), protozoa (Akopyants et al., 2016; Lye et al., 2016; Grybchuk et al., 2018), and in association with *Plasmodium* parasites causing human malaria (Charon et al., 2019). Interestingly, apart from typical narna- and narna-like viruses with genetic information organized on one segment, several other narna-like viruses with bi- and polysegmented genomes have been described in recent years. The first observed bisegmentation of the narna-like virus genome was reported in Leptomonas seymouri Narna-like virus 1 (Grybchuk et al., 2018) and in Matryoshka RNA virus 1 (Charon et al., 2019). In addition, narna-like viruses with a split polymerase palm domain "splipalmiviruses" have been discovered in the ascomycetus fungus *Oidiodendron maius* Barron (Sutela et al., 2020) and similar multisegmented narna-like viruses with a divided RdRp gene were reported from other ascomycetes, i.e., *Aspergillus fumigatus* Fresen., *Magnaporthe oryzae* B.C. Couch (Chiba et al., 2020), *Sclerotinia sclerotiorum* (Lib.) de Bary (Jia et al., 2021), and *Botrytis cinerea* Pers. (Ruiz-Padilla et al., 2021). "Orfanplasmoviruses" have been discovered in the virome associated with the downy mildew *Pl. viticola* (Chiapello et al., 2020) and in the conifer fungal pathogen *Heterobasidion parviporum* (Sutela et al., 2021). In some narnaviruses, negative-sense coding ORFs were observed (rORFs; DeRisi et al., 2019; Dinan et al., 2020). Due to recent discoveries of a number of remarkable features of narna-like viruses, several studies suggest the establishment of novel families, genera, and clades in order to accommodate the diverse groups of narnaviruses (Chiapello et al., 2020; Dinan et al., 2020; Ruiz-Padilla et al., 2021). The virus PcaRV2 has a size, genome organization, and phylogenic position in accordance with traditional, monosegmented members of *Narnaviridae*, indicating that it indeed represents a putative novel member of this family.

The family *Endornaviridae* contains linear (+)ssRNA viruses with genome sizes ranging from 9.7 to 17.6 kb. There are two genera within this family, to which endornaviruses are assigned based on the size of their genome, presence of distinct domains, and their host association (Valverde et al., 2019). *Alphaendornavirus* genera members are found infecting plants, oomycetes (Fukuhara and Gibbs, 2012), and fungi. *Betaendornavirus* includes viruses infecting ascomycete fungi (Valverde et al., 2019). Based on the low similarity of the *ca*. 2 kb fragment of PcaRV5 to other previously described endornaviruses, we conclude that this virus may represent a putative novel species belonging to the genus *Alphaendornavirus*. However, the full-length genome of this virus needs to be characterized to allow correct determination of its taxonomical position. As aforementioned alphaendornaviruses appear to be common in *Phytophthora* species. Interestingly, two alphaendornaviruses reported from Japanese isolates pathogenic to asparagus, Phytophthora endornavirus 2 (PEV2) and Phytophthora endornavirus 3 (PEV3) were found to inhibit hyphal growth and stimulate production of zoosporangia. In addition, PEV2 and PEV3 reportedly had a possible effect on both reduced as well as increased fungicide sensitivity of their host isolates (Uchida et al., 2021). Effects of viruses on their oomycete host have been also reported from the potato-late blight pathogen *P. infestans* infected by PiRV-2. PiRV-2 has been demonstrated to stimulate sporangia production in its host isolates, having a possible impact on ecological fitness of *P. infestans*, probably *via* downregulation of ammonium and amino acid intake (Cai et al., 2019b).

P. castaneae isolates VN999, VN1004, VN1008, and VN1012 are part of an ongoing broader study on viral diversity of *Phytophthora* Clade 5 and exploration of the possible effect of these viruses on their hosts. Repetitive attempts using thermo- and chemotherapy have been made in order to cure the isolates of viruses and study the isogenic strains for phenotypic effects (M. Raco unpublished data). However, due to the high viral load, no virus-free isolates have been obtained so far. Such rich mixed viral infections as observed in the studied isolates of these *P. castaneae* might have an effect on the overall energy consumption and vitality of the oomycete, which could be related to our observation that compared to other isolates these *P. castaneae* isolates have reduced survival time in storage and need more frequent transfers when compared to other isolates of PRC collection. Similar behavior has been previously observed by Cai et al. (2009) in *P. infestans* infected by PiRV-1. The latter authors also reported difficulties in curing PiRV-1 from host isolates, as well as the inability of transferring this virus *via* somatic fusion to a virus-free isolate.

All six viruses detected in *P. castaneae* are targeted by the *P. castaneae* RNAi machinery, and the most abundant size of sRNA reads in all four isolates was 21 nts, consistent with other *Phytophthora* species (Fahlgren et al., 2013; Jia et al., 2017; Botella and Jung, 2021) and the oomycete *Hyaloperonospora arabidopsidis* (Gäum.) Göker, Riethm., Voglmayr, Weiß & Oberw., a natural, obligate biotrophic downy mildew pathogen of *Arabidopsis thaliana* (L.) Heynh. (Dunker et al., 2020). In

P. condilina, *P. infestans*, *P. ramorum*, *P. sojae* Kaufm. & Gerd., and *H. arabidopsidis*, besides the 21-nts sRNAs, the next most abundant sRNA population had size of 25-nts (Fahlgren et al., 2013; Dunker et al., 2020; Botella and Jung, 2021). In *P. parasitica*, two distinct types of sRNAs were 25–26 nts and 21 nts, but the 25–26 nts profiles were predominant (Jia et al., 2017). 21-nts sRNAs seem to derive primarily from inverted repeats, including a novel conserved microRNA family, several gene families, and Crinkler effectors (Fahlgren et al., 2013). Although it was demonstrated in this study that sRNA sequencing could be used to detect novel and relatively divergent viral species, difficulties in detecting the presence of PcaRV4 in *P. castaneae* isolates VN999, VN1008, and VN1004 by BlastX analyses suggest that the detection of novel virus species sharing a very low percentage of similarity with previously described viruses could be challenging. Therefore, it is necessary to have a more specific database. In cases where known viruses are expected, as shown, for example, in the wood rot fungus *Heterobasidion annosum sensu lato* (Vainio et al., 2015b), and in the pine pitch canker pathogen, *Fusarium circinatum* Nirenberg & O'Donnell infected by mitoviruses (Adalia et al., 2018), HTS of sRNAs can be used as a successful tool for virus diagnosis. If the genomes are not available and novel viruses show low levels of similarity to known viral species, such sequencing methods cannot be used alone for assembling long contigs or nearly complete genomes as shown in *P. condilina* (Botella and Jung, 2021) and the present study. Such findings indicate that sRNA seq is not the best technique for describing novel viral species divergent from previously described viruses due to short contig lengths. However, sRNA reads efficiently provided good coverage and high read numbers (**Table 2**) when mapped to the contigs obtained through the *de novo* assembly of total RNA reads, as demonstrated before (Botella and Jung, 2021).

The isolates used in this study originated from a *Chamaecyparis-Quercus* forest stand on Sau-Chua mountain in northern Vietnam, where *P. castaneae* was the only *Phytophthora* species recovered (Jung et al., 2020). Based on the sequences obtained through RT-PCR and Sanger sequencing of almost the entire length of the final viral contigs, the viruses PcaRV1-PcaRV4 and PcaNSRV1 show very low intraspecific variability in each of the four isolates. This might be explained by the homothallic mainly inbreeding sexual system of *P. castaneae* and its clonal mode of spread *via* asexual zoospores (Erwin and Ribeiro, 1996). Similarly, the intraspecific genetic variability of mitoviruses and partitiviruses infecting *Gremmeniella abietina* in one Spanish pine stand was very low (Botella et al., 2012, 2015). In *Heterobasidion parviporum,* a study showed that during persistent viral infections, single point mutations accumulate, resulting in virus diversification and occurrence of nearly identical viral sequence variants within single host clones (Vainio et al., 2015b).

In conclusion, this study represents the first evidence of multiple viral infections in the tree pathogen *P. castaneae* collected from its natural montane habitat in Vietnam. Sequencing of total stranded RNA was proven to be a powerful tool for detecting and *de novo* assembling of novel viral species and their genomes. The RNAi mechanism appears to actively target all six reported mycoviruses here, and their implications in the virulence of *P. castaneae* should be further investigated.

DATA AVAILABILITY STATEMENT

The RNA-Seq raw datasets presented in this study can be found in the online repository Sequence Read Archive (SRA), and they are available at this link https://www.ncbi.nlm.nih.gov/sra/PRJNA818851 (BioProject ID: PRJNA818851). The sequences of virus genomes are available in the NCBI GenBank database under the accession numbers listed in **Table 1** of this manuscript.

AUTHOR CONTRIBUTIONS

MR, LB, and TJ: original idea and funding. AE, EH, EV, MR, LB, and SS: formal analysis and investigation. AE, EV, and MR: bioinformatics sRNA. MR and SS: bioinformatics of total RNA. TJ provided *P. castaneae* isolates. MR prepared the original draft and wrote the manuscript. LB contributed to the writing. All authors have read, critically reviewed, and edited the manuscript and agreed to the published version of the manuscript.

FUNDING

This research was supported by the European Regional Development Fund, project "Phytophthora Research Centre," Reg. No. CZ.02.1.01/0.0/0.0/15_003/0000453 and Specific University Research Fund of the Faculty of Forestry and Wood Technology, Mendel University in Brno LDF_VP_2019016.

ACKNOWLEDGMENTS

MR would like to thank Martin S. Mullett (Mendel University in Brno) for English language revision and his help and support in graph preparation in R studio. The authors thank Marilia Horta Jung for the molecular identification of the *Phytophthora* isolates and Minh Chi Nguyen (Vietnamese Academy of Forest Sciences, Hanoi), Bruno Scanu (University of Sassari, Italy), Joan Webber and Clive Brasier (Forest Research, Farnham, UK), and Tamara Corcobado (Mendel University in Brno) for assistance with sampling and isolation of the *Phytophthora* isolates. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.911474/full#supplementary-material

REFERENCES

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S. V., Avšič-Županc, T., Ballinger, M. J., et al. (2019). Taxonomy of the order *Bunyavirales*: update 2019. *Arch. Virol.* 164, 1949–1965. doi: 10.1007/s00705-019-04253-6

Adalia, E. J. M., Diez, J. J., Fernández, M. M., and Vainio, J. H. E. J. (2018). Characterization of small RNAs originating from mitoviruses infecting the conifer pathogen *Fusarium circinatum*. *Arch. Virol.* 163, 1009–1018. doi: 10.1007/s00705-018-3712-2

Akopyants, N. S., Lye, L. F., Dobson, D. E., Lukeš, J., and Beverley, S. M. (2016). A narnavirus in the trypanosomatid protist plant pathogen *Phytomonas serpens*. *Genome Announc.* 4:e00711-16. doi: 10.1128/GENOMEA.00711-16

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Oxford University Press Available at: https://academic.oup.com/nar/article/25/17/3389/1061651 (Accessed January 6, 2021).

Andrews, S. (2010). Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data. Soil Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed January 5, 2021).

Arjona-Lopez, J. M., Telengech, P., Jamal, A., Hisano, S., Kondo, H., Yelin, M. D., et al. (2018). Novel, diverse RNA viruses from Mediterranean isolates of the phytopathogenic fungus, *Rosellinia necatrix*: insights into evolutionary biology of fungal viruses. *Environ. Microbiol.* 20, 1464–1483. doi: 10.1111/1462-2920.14065

Bartholomäus, A., Wibberg, D., Winkler, A., Pühler, A., Schlüter, A., and Varrelmann, M. (2016). Deep sequencing analysis reveals the mycoviral diversity of the virome of an avirulent isolate of *Rhizoctonia solani* AG-2-2 IV. *PLoS One* 11:e0165965. doi: 10.1371/journal.pone.0165965

Beakes, G. W., Honda, D., and Thines, M. (2014). "Systematics of the Straminipila: Labyrinthulomycota, Hyphochytriomycota, and Oomycota," in *Systematics and Evolution Part A. The Mycota Book Series. Vol. 7A*. eds. D. J. McLaughli and J. W. Spatafora (Luxemburg: Springer Science+Business Media), 39–97.

Bekaert, M., Richard, H., Prum, B., and Rousset, J. P. (2005). Identification of programmed translational −1 frameshifting sites in the genome of *Saccharomyces cerevisiae*. *Genome Res.* 15, 1411–1420. doi: 10.1101/gr.4258005

Botella, L., and Hantula, J. (2018). Description, distribution, and relevance of viruses of the forest pathogen *Gremmeniella abietina*. *Viruses* 10, 1–14. doi: 10.3390/v10110654

Botella, L., Janoušek, J., Maia, C., Jung, M. H., Raco, M., and Jung, T. (2020). Marine oomycetes of the genus *Halophytophthora* harbor viruses related to *Bunyaviruses*. *Front. Microbiol.* 11:1467. doi: 10.3389/fmicb.2020.01467

Botella, L., and Jung, T. (2021). Multiple *viral infections detected in Phytophthora condilina* by Total and small RNA sequencing. *Viruses* 13:620. doi: 10.3390/v13040620

Botella, L., Tuomivirta, T. T., Hantula, J., Diez, J. J., and Jankovsky, L. (2015). The European race of *Gremmeniella abietina* hosts a single species of Gammapartitivirus showing a global distribution and possible recombinant events in its history. *Fungal Biol.* 119, 125–135. doi: 10.1016/J.FUNBIO.2014.12.001

Botella, L., Tuomivirta, T. T., Vervuurt, S., Diez, J. J., and Hantula, J. (2012). Occurrence of two different species of mitoviruses in the European race of *Gremmeniella abietina* var. *abietina*, both hosted by the genetically unique Spanish population. *Fungal Biol.* 116, 872–882. doi: 10.1016/J.FUNBIO.2012.05.004

Bruce, B. (1999). "Occurrence and impact of *Phytophthora cinnamomi* and other *Phytophthora* species in rainforests of the wet tropics world heritage area, and of the mackay region, qld" in *Patch Deaths in Tropical Queensland Rainforests: Association and Impact of Phytophthora*. ed. P. A. Gadek (Cairns, Australia: Cooperative Research Centre for Tropical Rainforest Ecology and Management)

Cai, G., Fry, W. E., and Hillman, B. I. (2019a). PiRV-2 stimulates sporulation in *Phytophthora infestans*. *Virus Res.* 271:197674. doi: 10.1016/j.virusres.2019.197674

Cai, G., and Hillman, B. I. (2013). Phytophthora viruses. *Adv. Virus Res.* 86, 327–350. doi: 10.1016/B978-0-12-394315-6.00012-X

Cai, G., Krychiw, J. F., Myers, K., Fry, W. E., and Hillman, B. I. (2013). A new virus from the plant pathogenic oomycete *Phytophthora infestans* with an 8 kb dsRNA genome: The sixth member of a proposed new virus genus. *Virology* 435, 341–349. doi: 10.1016/J.VIROL.2012.10.012

Cai, G., Myers, K., Fry, W. E., and Hillman, B. I. (2012). A member of the virus family *Narnaviridae* from the plant pathogenic oomycete *Phytophthora infestans*. *Arch. Virol.* 157, 165–169. doi: 10.1007/s00705-011-1126-5

Cai, G., Myers, K., Fry, W. E., and Hillman, B. I. (2019b). Phytophthora infestans RNA virus 2, a novel RNA virus from *Phytophthora infestans*, does not belong to any known virus group. *Arch. Virol.* 164, 567–572. doi: 10.1007/s00705-018-4050-0

Cai, G., Myers, K., Hillman, B. I., and Fry, W. E. (2009). A novel virus of the late blight pathogen, *Phytophthora infestans*, with two RNA segments and a supergroup 1 RNA-dependent RNA polymerase. *Virology* 392, 52–61. doi: 10.1016/j.virol.2009.06.040

Chandler, J. A., Liu, R. M., and Bennett, S. N. (2015). RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Front. Microbiol.* 6:185. doi: 10.3389/FMICB.2015.00185/BIBTEX

Chang, S.-S., Zhang, Z., and Liu, Y. (2012). RNA interference pathways in fungi: mechanisms and functions. *Annu. Rev. Microbiol.* 66, 305–323. doi: 10.1146/annurev-micro-092611-150138

Charon, J., Grigg, M. J., Eden, J. S., Piera, K. A., Rana, H., William, T., et al. (2019). Novel RNA viruses associated with *plasmodium vivax* in human malaria and Leucocytozoon parasites in avian disease. *PLoS Pathog.* 15:e1008216. doi: 10.1371/journal.ppat.1008216

Charon, J., Murray, S., Holmes, E., and C., (2021). Revealing RNA virus diversity and evolution in unicellular algae transcriptomes. *Virus Evol.* 7, 1–18. doi: 10.1093/ve/veab070

Chen, S., Cao, L., Huang, Q., Qian, Y., and Zhou, X. (2016). The complete genome sequence of a novel maize-associated totivirus. *Arch. Virol.* 161, 487–490. doi: 10.1007/s00705-015-2657-y

Chiapello, M., Rodríguez-Romero, J., Ayllón, M. A., and Turina, M. (2020). Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus Evol.* 6, 1–18. doi: 10.1093/ve/veaa058

Chiba, Y., Oiki, S., Yaguchi, T., Urayama, S.-I., and Hagiwara, D. (2020). Discovery of divided RdRp sequences and a hitherto unknown genomic complexity in fungal viruses. *Virus Evol.* 7:1. doi: 10.1093/ve/veaa101

Chiba, S., Salaipeth, L., Lin, Y.-H., Sasaki, A., Kanematsu, S., and Suzuki, N. (2009). A novel bipartite double-stranded RNA Mycovirus from the white root rot fungus *Rosellinia necatrix*: molecular and biological characterization, taxonomic considerations, and potential for biological control. *J. Virol.* 83, 12801–12812. doi: 10.1128/JVI.01830-09

Cook, S., Chung, B. Y. W., Bass, D., Moureau, G., Tang, S., McAlister, E., et al. (2013). Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* 8:e80720. doi: 10.1371/JOURNAL.PONE.0080720

De Lima, J. G. S., Teixeira, D. G., Freitas, T. T., Lima, J. P. M. S., and Lanza, D. C. F. (2019). Evolutionary origin of 2A-like sequences in *Totiviridae* genomes. *Virus Res.* 259, 1–9. doi: 10.1016/j.virusres.2018.10.011

Deakin, G., Dobbs, E., Bennett, J. M., Jones, I. M., Grogan, H. M., and Burton, K. S. (2017). Multiple viral infections in *Agaricus bisporus* – characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Sci. Rep.* 7, 2469–2413. doi: 10.1038/s41598-017-01592-9

DeRisi, J. L., Huber, G., Kistler, A., Retallack, H., Wilkinson, M., and Yllanes, D. (2019). An exploration of ambigrammatic sequences in narnaviruses. *Sci. Rep.* 9:17982. doi: 10.1038/s41598-019-54181-3

Dinan, A. M., Lukhovitskaya, N. I., Olendraite, I., and Firth, A. E. (2020). A case for a negative-strand coding sequence in a group of positive-sense RNA viruses. *Virus Evol.* 6:veaa007. doi: 10.1093/ve/veaa007

Doherty, M., Sanganee, K., Kozlakidis, Z., Coutts, R. H. A., Brasier, C. M., and Buck, K. W. (2007). Molecular characterization of a totivirus and a partitivirus from the genus *Ophiostoma*. *J. Phytopathol.* 155, 188–192. doi: 10.1111/j.1439-0434.2007.01207.x

Donaire, L., Pagán, I., and Ayllón, M. A. (2016). Characterization of Botrytis cinerea negative-stranded RNA virus 1, a new mycovirus related to plant viruses, and a reconstruction of host pattern evolution in negative-sense ssRNA viruses. *Virology* 499, 212–218. doi: 10.1016/J.VIROL.2016.09.017

Dunker, F., Trutzenberg, A., Rothenpieler, J. S., Kuhn, S., Schreiber, T., Tissier, A., et al. (2020). Oomycete small RNAs bind to the plant RNA-induced silencing complex for virulence. *eLife* 9:e56096. doi: 10.7554/eLife.56096

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Erwin, C. D., and Ribeiro, K. O. (1996). *Phytophthora Diseases Worldwide*. Minesota: The American Phytopathological Society.

Fahlgren, N., Bollmann, S. R., Kasschau, K. D., Cuperus, J. T., and Press, C. M. (2013). *Phytophthora* have distinct endogenous small RNA populations that include short interfering and micrornas. *PLoS One* 8:77181. doi: 10.1371/journal.pone.0077181

Fukuhara, T., and Gibbs, M. J. (2012). "Family *Endornaviridae*," in *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of The International Committee on Taxonomy of Viruses*. eds. A. M. Q. King, M. J. Adams, E. B. Carstens and E. J. Lefkowitz (Elsevier Academic Press), 1055–1057.

Fukunishi, M., Sasai, S., Tojo, M., and Mochizuki, T. (2021). Novel fusari- and toti-like viruses, with probable different origins, in the plant pathogenic oomycete *Globisporangium ultimum*. *Viruses* 13:1931. doi: 10.3390/v13101931

Garrison, A. R., Alkhovsky, S. V., Avšič-Županc, T., Bente, D. A., Bergeron, É., Burt, F., et al. (2020). ICTV virus taxonomy profile: *Nairoviridae* ICTV virus taxonomy profile ICTV. *J. Gen. Virol.* 101, 798–799. doi: 10.1099/jgv.0.001485

Ghabrial, S. A., Castón, J. R., Jiang, D., Nibert, M. L., and Suzuki, N. (2015). 50-plus years of fungal viruses. *Virology* 480, 356–368. doi: 10.1016/j.virol.2015.02.034

Gillings, M. R., Ttesoriero, L. A., and Gunn, L. V. (1993). Detection of double-stranded RNA and virus-like particles in Australian isolates of *Pythium irregulare*. *Plant Pathol.* 42, 6–15. doi: 10.1111/j.1365-3059.1993.tb01466.x

Göertz, G. P., Miesen, P., Overheul, G. J., Van Rij, R. P., Van Oers, M. M., and Pijlman, G. P. (2019). Mosquito small RNA responses to West Nile and insect-specific virus infections in Aedes and Culex mosquito cells. *Viruses* 11:271. doi: 10.3390/V11030271

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grasse, W., and Spring, O. (2017). ssRNA viruses from biotrophic Oomycetes form a new phylogenetic group between *Nodaviridae* and *Tombusviridae*. *Arch. Virol.* 162, 1319–1324. doi: 10.1007/s00705-017-3243-2

Grybchuk, D., Akopyants, N. S., Kostygov, A. Y., Konovalovas, A., Lye, L. F., Dobson, D. E., et al. (2018). Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite Leishmania. *Proc. Natl. Acad. Sci. U. S. A.* 115, E506–E515. doi: 10.1073/PNAS.1717806115/-/DCSUPPLEMENTAL

Hantula, J., Mäkelä, S., Xu, P., Brusila, V., Nuorteva, H., Kashif, M., et al. (2020). Multiple virus infections on *Heterobasidion* sp. *Fungal Biol.* 124, 102–109. doi: 10.1016/j.funbio.2019.12.004

Harvey, E., Rose, K., Eden, J. S., Lawrence, A., Doggett, S. L., and Holmes, E. C. (2019). Identification of diverse arthropod associated viruses in native Australian fleas. *Virology* 535, 189–199. doi: 10.1016/j.virol.2019.07.010

Heller-Dohmen, M., Göpfert, J. C., Pfannstiel, J., and Spring, O. (2011). The nucleotide sequence and genome organization of Plasmopara halstedii virus. *Virol. J.* 8:123. doi: 10.1186/1743-422X-8-123

Hillman, B. I., Annisa, A., and Suzuki, N. (2018). Viruses of plant-interacting fungi. *Adv. Virus Res.* 100, 99–116. doi: 10.1016/bs.aivir.2017.10.003

Hughes, H. R., Adkins, S., Alkhovskiy, S., Beer, M., Blair, C., Calisher, C. H., et al. (2020). ICTV virus taxonomy profile: *Peribunyaviridae*. *J. Gen. Virol.* 101, 1–2. doi: 10.1099/jgv.0.001365

Jia, J., Fu, Y., Jiang, D., Mu, F., Cheng, J., Lin, Y., et al. (2021). Interannual dynamics, diversity and evolution of the virome in *Sclerotinia sclerotiorum* from a single crop field. *Virus Evol.* 7:veab032. doi: 10.1093/ve/veab032

Jia, J., Lu, W., Zhong, C., Zhou, R., Xu, J., Liu, W., et al. (2017). The 25-26 nt small RNAs in *Phytophthora parasitica* are associated with efficient silencing of homologous endogenous genes. *Front. Microbiol.* 8, 1–15. doi: 10.3389/fmicb.2017.00773

Jung, T., Chang, T. T., Bakonyi, J., Seress, D., Perez-Sierra, A. P., Yang, X., et al. (2017). Diversity of *Phytophthora* species in natural ecosystems of Taiwan and association with disease symptoms. *Plant Pathol.* 66, 194–211. doi: 10.1111/ppa.12564

Jung, T., Pérez-Sierra, A., Durán, A., Jung, M. H., Balci, Y., and Scanu, B. (2018). Canker and decline diseases caused by soil- and airborne *Phytophthora* species in forests and woodlands. *Persoonia Mol. Phylogeny Evol. Fungi* 40, 182–220. doi: 10.3767/persoonia.2018.40.08

Jung, T., Scanu, B., Brasier, C. M., Webber, J., Milenković, I., Corcobado, T., et al. (2020). A survey in natural forest ecosystems of Vietnam reveals high diversity of both new and described phytophthora taxa including *P. ramorum*. *Forests* 11:93. doi: 10.3390/f11010093

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Ketting, R. F. (2011). The many faces of RNAi. *Dev. Cell* 20, 148–161. doi: 10.1016/j.devcel.2011.01.012

Kocanová, M., Eichmeier, A., and Botella, L. (2020). A novel mitovirus detected in *Diaporthe rudis*, a fungus associated with Phomopsis dieback on grapevines. *Arch. Virol.* 165, 2405–2408. doi: 10.1007/s00705-020-04755-8

Koyama, S., Urayama, S. I., Ohmatsu, T., Sassa, Y., Sakai, C., Takata, M., et al. (2015). Identification, characterization and full-length sequence analysis of a novel dsRNA virus isolated from the arboreal ant *Camponotus yamaokai*. *J. Gen. Virol.* 96, 1930–1937. doi: 10.1099/vir.0.000126

Kozlakidis, Z., Brown, N. A., Jamal, A., Phoon, X., and Coutts, R. H. A. (2010). Incidence of endornaviruses in *Phytophthora* taxon douglasfir and *Phytophthora ramorum*. *Virus Genes* 40, 130–134. doi: 10.1007/s11262-009-0421-7

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7. doi: 10.1016/j.virol.2009.03.024

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25–R10. doi: 10.1186/GB-2009-10-3-R25

Legeay, J., Husson, C., Boudier, B., Louisanna, E., Baraloto, C., Schimann, H., et al. (2020). Surprising low diversity of the plant pathogen *Phytophthora* in Amazonian forests. *Environ. Microbiol.* 22, 5019–5032. doi: 10.1111/1462-2920.15099

Leventhal, S. S., Wilson, D., Feldmann, H., Hawman, D. W., and Brennan, B. (2021). A look into bunyavirales genomes: functions of non-structural (ns) proteins. *Viruses* 13:314. doi: 10.3390/v13020314

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available at: http://github.com/lh3/bwa. (Accessed January 11, 2021).

Linnakoski, R., Sutela, S., Coetzee, M. P. A., Duong, T. A., Pavlov, I. N., Litovka, Y. A., et al. (2021). Armillaria root rot fungi host single-stranded RNA viruses. *Sci. Rep.* 11:7336. doi: 10.1038/s41598-021-86343-7

Lye, L. F., Akopyants, N. S., Dobson, D. E., and Beverley, S. M. (2016). A narnavirus-like element from the trypanosomatid protozoan parasite *Leptomonas seymouri*. *Genome Announc.* 4:e00713-16. doi: 10.1128/GENOMEA.00713-16

Maia, C., Horta Jung, M., Carella, G., Milenković, I., Janoušek, J., Tomšovský, M., et al. (2022). Eight new *Halophytophthora* species from marine and brackish-water ecosystems in portugal and an updated phylogeny for the genus. *Persoonia Mol. Phylogeny Evol. Fungi.* 48, 54–90. doi: 10.3767/PERSOONIA.2022.48.02

Martin, F. N., Blair, J. E., and Coffey, M. D. (2014). A combined mitochondrial and nuclear multilocus phylogeny of the genus *Phytophthora*. *Fungal Genet. Biol.* 66, 19–32. doi: 10.1016/j.fgb.2014.02.006

Martinez, J., Lepetit, D., Ravallec, M., Fleury, F., and Varaldi, J. (2016). Additional heritable virus in the parasitic wasp *Leptopilina boulardi*: prevalence, transmission and phenotypic effects. *J. Gen. Virol.* 97, 523–535. doi: 10.1099/jgv.0.000360

Mascia, T., Labarile, R., Doohan, F., and Gallitelli, D. (2019). Tobacco mosaic virus infection triggers an RNAi-based response in *Phytophthora infestans*. *Sci. Rep.* 9, 2657–2613. doi: 10.1038/s41598-019-39162-w

Massart, S., Chiumenti, M., Jonghe, K.De, Glover, R., Haegeman, A., Koloniuk, I., et al. (2019). Virus detection by high-throughput sequencing of small RNAs: large-scale performance testing of sequence analysis strategies. doi:10.1094/PHYTO-02-18-0067-R, 109, 488–497.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 343–349. doi: 10.1038/nature02873

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Science Gateway for inference of large phylogenetic trees." in *2010 gateway computing environments workshop (GCE)*; November 14, 2010.

Mor, S. K., and Phelps, N. B. D. (2016). Molecular detection of a novel totivirus from golden shiner (*Notemigonus crysoleucas*) baitfish in the USA. *Arch. Virol.* 161, 2227–2234. doi: 10.1007/s00705-016-2906-8

Morris, T. J., and Dodds, J. A. (1979). Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology* 69:854. doi: 10.1094/phyto-69-854

Nerva, L., Ciuffo, M., Vallino, M., Margaria, P., Varese, G. C., Gnavi, G., et al. (2016). Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Res.* 219, 22–38. doi: 10.1016/j.virusres.2015.10.028

Osaki, H., Sasaki, A., Nomiyama, K., and Tomioka, K. (2016). Multiple virus infection in a single strain of *Fusarium poae* shown by deep sequencing. *Virus Genes* 52, 835–847. doi: 10.1007/S11262-016-1379-X/FIGURES/6

Poimala, A., Parikka, P., Hantula, J., and Vainio, E. J. (2021). Viral diversity in *Phytophthora cactorum* population infecting strawberry. *Environ. Microbiol.* 23, 5200–5221. doi: 10.1111/1462-2920.15519

Poimala, A., and Vainio, E. J. (2020). Complete genome sequence of a novel toti-like virus from the plant-pathogenic oomycete *Phytophthora cactorum*. *Arch. Virol.* 165, 1679–1682. doi: 10.1007/s00705-020-04642-2

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rampersad, S., and Tennant, P. (2018). "Replication and expression strategies of viruses," in *Viruses: Molecular Biology, Host Interactions, and Applications to Biotechnology*. eds. T. Paula, G. Fermin and J. Foster (Academic Press), 55–82.

Reguera, J., Weber, F., and Cusack, S. (2010). Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathog.* 6:e1001101. doi: 10.1371/journal.ppat.1001101

Romano, N., Macino, G., Umana, B., Biologia, S., Umberto, C. P., and La, R. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Mol. Microbiol.* 6, 3343–3353. doi: 10.1111/j.1365-2958.1992.tb02202.x

Ruiz-Padilla, A., Rodríguez-Romero, J., Gómez-Cid, I., Pacifico, D., and Ayllón, M. A. (2021). Novel mycoviruses discovered in the mycovirome of a necrotrophic fungus. *MBio* 12:e03705-20. doi: 10.1128/mBio.03705-20

Sasai, S., Tamura, K., Tojo, M., Herrero, M. L., Hoshino, T., Ohki, S. T., et al. (2018). A novel non-segmented double-stranded RNA virus from an Arctic isolate of *Pythium polare*. *Virology* 522, 234–243. doi: 10.1016/j.virol.2018.07.012

Sasaki, A., Nakamura, H., Suzuki, N., and Kanematsu, S. (2016). Characterization of a new megabirnavirus that confers hypovirulence with the aid of a co-infecting partitivirus to the host fungus, *Rosellinia necatrix*. *Virus Res.* 219, 73–82. doi: 10.1016/j.virusres.2015.12.009

Sato, Y., Miyazaki, N., Kanematsu, S., Xie, J., Ghabrial, S. A., Hillman, B. I., et al. (2019). ICTV virus taxonomy profile: Megabirnaviridae. *J. Gen. Virol.* 100, 1269–1270. doi: 10.1099/jgv.0.001297

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543. doi: 10.1038/nature20167

Shi, M., Neville, P., Nicholson, J., Eden, J.-S., Imrie, A., and Holmes, E. C. (2017). High-resolution Metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J. Virol.* 91, 680–697. doi: 10.1128/JVI.00680-17

Shiba, K., Hatta, C., Sasai, S., Tojo, M., Ohki, S. T., and Mochizuki, T. (2019). A novel toti-like virus from a plant pathogenic oomycete *Globisporangium splendens*. *Virology* 537, 165–171. doi: 10.1016/j.virol.2019.08.025

Shiba, K., Hatta, C., Sasai, S., Tojo, M., Ohki, S., Mochizuki, T., et al. (2018). Genome sequence of a novel partitivirus identified from the oomycete *Pythium Nunn*. *Arch. Virol.* 163, 2561–2563. doi: 10.1007/s00705-018-3880-0

Stamatakis, A., Hoover, P., Rougemont, J., Diego, S., and Jolla, L. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642

Studholme, D. J., McDougal, R. L., Sambles, C., Hansen, E., Hardy, G., Grant, M., et al. (2016). Genome sequences of six *Phytophthora* species associated with forests in New Zealand. *Genomics Data* 7, 54–56. doi: 10.1016/j.gdata.2015.11.015

Sutela, S., Forgia, M., Vainio, E. J., Chiapello, M., Daghino, S., Vallino, M., et al. (2020). The virome from a collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome organization. *Virus Evol.* 6:veaa076. doi: 10.1093/VE/VEAA076

Sutela, S., Piri, T., and Vainio, E. J. (2021). Discovery and community dynamics of novel ssRNA mycoviruses in the conifer pathogen *Heterobasidion parviporum*. *Front. Microbiol.* 12:3337. doi: 10.3389/fmicb.2021.770787

Svoboda, P. (2020). Key mechanistic principles and considerations concerning RNA interference. *Front. Plant Sci.* 11, 1–13. doi: 10.3389/fpls.2020.01237

Thapa, V., and Roossinck, M. J. (2019). Determinants of coinfection in the mycoviruses. *Front. Cell. Infect. Microbiol.* 9:169. doi: 10.3389/fcimb.2019.00169

Thines, M., and Choi, Y. J. (2016). Evolution, diversity, and taxonomy of the Peronosporaceae, with focus on the genus *Peronospora*. *Phytopathology* 106, 6–18. doi: 10.1094/PHYTO-05-15-0127-RVW

Tuomivirta, T. T., Uotila, A., and Hantula, J. (2002). Two independent double-stranded RNA patterns occur in the Finnish *Gremmeniella abietina* var. *abietina* type A. *For. Pathol.* 32, 197–205. doi: 10.1046/j.1439-0329.2002.00285.x

Uchida, K., Sakuta, K., Ito, A., Takahashi, Y., Katayama, Y., Fukuhara, T., et al. (2021). Two novel endornaviruses co-infecting a *Phytophthora* pathogen of *Asparagus officinalis* modulate the developmental stages and fungicide sensitivities of the host oomycete. *Front. Microbiol.* 12:633502. doi: 10.3389/fmicb.2021.633502

Vainio, E. J., Jurvansuu, J., Streng, J., Rajamäki, M. L., Hantula, J., and Valkonen, J. P. T. (2015a). Diagnosis and discovery of fungal viruses using deep sequencing of small RNAs. *J. Gen. Virol.* 96, 714–725. doi: 10.1099/jgv.0.000003

Vainio, E. J., Müller, M. M., Korhonen, K., Piri, T., and Hantula, J. (2015b). Viruses accumulate in aging infection centers of a fungal forest pathogen. *ISME J.* 9, 497–507. doi: 10.1038/ismej.2014.145

Valverde, R. A., Khalifa, M. E., Okada, R., Fukuhara, T., and Sabanadzovic, S. (2019). ICTV virus taxonomy profile: Endornaviridae. *J. Gen. Virol.* 100, 1204–1205. doi: 10.1099/jgv.0.001277

Velasco, L., Arjona-Girona, I., Ariza-Fernández, M. T., Cretazzo, E., and López-Herrera, C. (2018). A novel hypovirus species from Xylariaceae fungi infecting avocado. *Front. Microbiol.* 9:778. doi: 10.3389/fmicb.2018.00778

Waldron, F. M., Stone, G. N., and Obbard, D. J. (2018). Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLoS Genet.* 14:e1007533. doi: 10.1371/JOURNAL.PGEN.1007533

Wang, M., Wang, Y., Sun, X., Cheng, J., Fu, Y., Liu, H., et al. (2015). Characterization of a novel megabirnavirus from *Sclerotinia sclerotiorum* reveals horizontal gene transfer from single-stranded RNA virus to double-stranded RNA virus. *J. Virol.* 89, 8567–8579. doi: 10.1128/jvi.00243-15

Wang, L., Zhang, J., Zhang, H., Qiu, D., and Guo, L. (2016). Two novel relative double-stranded RNA mycoviruses infecting *Fusarium poae* strain SX63. *Int. J. Mol. Sci.* 17:641. doi: 10.3390/ijms17050641

Weiland, J. J., and Sundsbak, J. L. (2000). Differentiation and detection of sugar beet fungal pathogens using PCR amplification of actin coding sequences and the ITS region of the rRNA gene. *Plant Dis.* 84, 475–482. doi: 10.1094/PDIS.2000.84.4.475

Weir, B. S., Paderes, E. P., Anand, N., Uchida, J. Y., Pennycook, S. R., Bellgard, S. E., et al. (2015). A taxonomic revision of *Phytophthora* clade 5 including two new species, *Phytophthora agathidicida* and *P. cocois*. *Phytotaxa* 205, 21–38. doi: 10.11646/phytotaxa.205.1.2

Wickner, R. B., Ghabrial, S. A., Nibert, M. L., Patterson, J. L., and Wang, C. C. (2012). "Family *Totiviridae*," in *Virus Taxonomy: Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses*. eds. A. M. Q. King, M. J. Adams, E. B. Carstens and E. J. Lefkowitz (Elsevier Academic Press), 639–650.

Xu, Z., Khalifa, M. E., Frampton, R. A., Smith, G. R., McDougal, R. L., Macdiarmid, R. M., et al. (2022). Characterization of a novel double-stranded RNA virus from *Phytophthora pluvialis* in New Zealand. *Viruses* 14:247. doi: 10.3390/v14020247

Yokoi, T., Takemoto, Y., Suzuki, M., Yamashita, S., and Hibi, T. (1999). The nucleotide sequence and genome organization of Sclerophthora macrospora virus B. *Virology* 264, 344–349. doi: 10.1006/viro.1999.0018

Yokoi, T., Yamashita, S., and Hibi, T. (2003). The nucleotide sequence and genome organization of Sclerophthora macrospora virus A. *Virology* 311, 394–399. doi: 10.1016/S0042-6822(03)00183-1

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Zhang, X., Gao, F., Zhang, F., Xie, Y., Zhou, L., Yuan, H., et al. (2018). The complete genomic sequence of a novel megabirnavirus from *Fusarium pseudograminearum*, the causal agent of wheat crown rot. *Arch. Virol.* 163, 3173–3175. doi: 10.1007/s00705-018-3970-z

Zhong, J., Chen, C. Y., and Da Gao, B. (2015). Genome sequence of a novel mycovirus of *Rhizoctonia solani*, a plant pathogenic fungus. *Virus Genes* 51, 167–170. doi: 10.1007/S11262-015-1219-4/FIGURES/1

Zhu, J. Z., Zhu, H. J., Gao, B.Da, Zhou, Q., and Zhong, J. (2018). Diverse, novel mycoviruses from the virome of a hypovirulent *Sclerotium rolfsii* strain. *Front. Plant Sci.* 9, 1–14. doi:10.3389/fpls.2018.01738.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fmicb.2022.914516

Risk Factors Associated With Human Papillomavirus Infection, Cervical Cancer, and Precancerous Lesions in Large-Scale Population Screening

Di Yang[1], Jing Zhang[1], Xiaoli Cui[1], Jian Ma[2], Chunyan Wang[1] and Haozhe Piao[3]**

[1] Department of Gynecology, Liaoning Cancer Hospital and Institute, Cancer Hospital of China Medical University, Shenyang, China, [2] Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China, [3] Department of Neurosurgery, Liaoning Cancer Hospital and Institute, Cancer Hospital of China Medical University, Shenyang, China

Cervical cancer is the most common gynecological malignancy and screening for risk factors with early detection has been shown to reduce the mortality. In this study, we aimed to analyze the characteristics and risk factors of human papillomavirus (HPV) infection and precancerous lesions in women and provide clinical evidence for developing strategies to prevent cervical precancerous lesions and cancer in women. Furthermore, we evaluated the influencing factors for high-risk HPV infection. From April 2018 to December 2021, 10,628 women were recruited for cervical cancer screening at Liaoning Cancer Hospital, Shenyang Sujiatun District Women's and Infants Hospital, Benxi Manchu Autonomous County People's Hospital, and Shandong Affiliated Hospital of Qingdao University. The study participants were tested to determine if they were HPV-positive (HPV +) or underwent thinprep cytology test (TCT) for atypical squamous cells of undetermined significance (ASCUS) and above. Furthermore, colposcopies and biopsies were performed for the histopathological examination. Finally, 9991 cases were included in the statistical analysis, and the factors influencing HPV infection and those related to cervical cancer and precancerous lesions were analyzed. HPV + infection, high-grade squamous intraepithelial lesion-positive (CINII +) in cervical high-grade intraepithelial neoplasia, and early cervical cancer diagnosis rates were 12.45, 1.09, and 95.41%, respectively. The potential risk factors for HPV were education ≤ high school [odds ratio (OR) = 1.279 (1.129–1.449), $P < 0.001$], age at initial sexual activity ≤ 19 years [OR = 1.517 (1.080–2.129), $P = 0.016$], sexual partners > 1 [OR = 1.310 (1.044–1.644), $P = 0.020$], ASCUS and above [OR = 11.891 (10.105–13.993), $P < 0.001$], non-condom contraception [OR = 1.255 (1.059–1.487), $P = 0.009$], and HSIL and above [OR = 1.541 (1.430–1.662), $P < 0.001$]. Compared with women aged 56–65 and 35–45 years [OR = 0.810 (0.690–0.950), $P = 0.010$] the HPV infection rate was significantly lower in those aged 46–55 years [OR = 0.79 (0.683–0.915), $P = 0.002$]. Furthermore, ≤ high school age [OR = 1.577 (1.042–2.387), $P = 0.031$], not breastfeeding [OR = 1.763 (1.109–2.804), $P = 0.017$], ASCUS and above [OR = 42.396 (28.042–64.098), $P < 0.001$] were potential risk factors for cervical cancer and precancerous lesions. In women with HPV infection, ≤ high school

OPEN ACCESS

Edited by:
Antoinette Van Der Kuyl, University of Amsterdam, Netherlands

Reviewed by:
Suvankar Ghorai, Raiganj University, India
Ana Afonso, University of São Paulo, Brazil
Komsun Suwannarurk, Thammasat University, Thailand

***Correspondence:**
Chunyan Wang
lnzlgyn@163.com
Haozhe Piao
pzpy@163.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *07 April 2022*
Accepted: *10 June 2022*
Published: *30 June 2022*

Citation:
Yang D, Zhang J, Cui X, Ma J, Wang C and Piao H (2022) Risk Factors Associated With Human Papillomavirus Infection, Cervical Cancer, and Precancerous Lesions in Large-Scale Population Screening. Front. Microbiol. 13:914516. doi: 10.3389/fmicb.2022.914516

education level, initial sexual activity at ≤ 19 years of age, number of sexual partners > 1, ASCUS and above, non-condom contraception, HSIL and above were risk factors for HPV infection. Compared with women aged 56–65 years, those aged 35–45 and 46–55 years had significantly lower HPV infection rates, and high school age and below, non-breastfeeding, and ASCUS and above were all potential risk factors for cervical cancer and precancerous lesions.

Keywords: cervical cancer, screening, high-risk factor, human papillomavirus infection, precancerous lesions

INTRODUCTION

Cervical cancer is the most common gynecological malignancy and in 2020, with 604,000 new cases worldwide, with approximately 342,000 related deaths (Sung et al., 2021). Cervical cancer has become the most common cancer in 23 countries and the leading cause of cancer-related deaths in 36 countries. The incidence rate of the disease in developing countries is significantly different from that in developed countries (Bray et al., 2018), with the highest occurring in sub Saharan Africa, South Polynesia, South America, and Southeast Asia (Sung et al., 2021). Inadequate living conditions and resources in developing regions, such as a lack of access to physical contraceptives and poor living and personal hygiene conditions, are considered factors that increase the burden of cervical cancer (Abulizi et al., 2017).

In contrast, developed countries have high coverage and availability of cancer screening and compliance by the target population (Thanapprapasr et al., 2012). The incidence of cervical cancer in China is not encouraging and the International Agency for Research on Cancer (IARC) reported 109,741 new cases in 2020. Furthermore, 59,060 patients died (Cao et al., 2021), accounting for 18.2 and 17.3% (Bray et al., 2018) of the cervical cancer incidence and death rate worldwide, respectively. According to the 2020 annual work report of the National Cancer Center, cervical cancer still has the sixth highest incidence of female cancers in China, and the mortality rate is still the eighth highest among female malignant tumors (Bray et al., 2018).

The incidence rate of cervical cancer at a younger age is showing an increasing trend (Zheng et al., 2019). Furthermore, cervical lesions are the most common diseases in women of childbearing age, and mainly include inflammation, injury, deformity, precancerous lesions, and tumors. Cervical cancer is usually caused by human papillomavirus (HPV) infection. HPV is a spherical DNA virus, which can cause proliferation of human skin mucosal squamous epithelium, and further reproduction may lead to various cervical diseases (McLaughlin-Drubin and Munger, 2008).

The German scholar Hausen first identified HPV infection as the main pathogenic factor of cervical cancer and precancerous lesions, which could be considered a milestone discovery in the prevention and treatment of this disease (Peter et al., 2010). A considerable amount of epidemiological and biological data have also proved that HPV infection is the main cause of cervical cancer and cervical intraepithelial neoplasia (Cordeiro et al., 2018). However, the HPV infection rate varies between regions because it is affected by numerous factors such as region, race, living habits, and HPV vaccination rate. Worldwide, economically developed regions such as South Korea, reported 18,170 women with HPV infection in 2014–2016, including 2,268 (12.5%) who were high-risk HPV positive (HPV +) (Ouh et al., 2018).

HPV E6/E7 mRNA detection is a new cervical cancer screening technology emerging in recent years. It is a screening method with E6/E7 as the detection target. Whether it has advantages in cervical cancer screening remains to be studied. The sensitivity of hpve6/E7 mRNA detection was lower than that of HR—HPV DNA detection, but the missed detection rate of hpve6/E7 mRNA detection was significantly lower than that of TCT and HR—HPV DNA detection ($P < 0.05$) (Giorgi Rossi et al., 2022).

The study of HPV prevalence and its subtype distribution may provide relevant information for routine vaccination and the types of HPV strains used in vaccination. HPV infection rates among women worldwide range from 11.70 to 7.20% (Lewandowska et al., 2021). The highest prevalence rates have been reported in Sub Saharan Africa (24.00%), Latin America and the Caribbean (16.10%), Eastern Europe (14.20%), and South East Asia (14.00%) (Bekmukhambetov et al., 2016). A study in the economically underdeveloped regions of West Africa showed that the high-risk HPV infection rate of 28.6% (Maria et al., 2018), whereas the overall HPV prevalence rate in Kazakhstan was reported to be as high as 43.8–55.8% (Aimagambetova and Azizan, 2018).

In China, significant differences were also reported in HPV infection rates in different regions, ranging from 13 to 31.9% (Zhong et al., 2017; Jiang et al., 2019). HPV also plays a vital role in the development of cervical lesions and cervical cancer, and the World Health Organization has confirmed that the mortality rate can effectively be reduced by screening. This study analyzed the characteristics and high-risk factors of HPV infection, cervical cancer, and high-grade precancerous lesions in a large-scale population screened for cervical cancer, to provide a reference for the prevention and treatment of these conditions.

PATIENTS AND METHODS

Ethics Statement

This study was approved by the Ethics Committee of Liaoning Cancer Hospital and informed consent was obtain from all individual participants included in the study (No.: 20180106).

Research Participants

From April 2018 to December 2021, we recruited 10,628 women for cervical cancer screening in Liaoning Province at Liaoning Cancer Hospital, Shenyang Sujiatun District Women's and Infant Hospital, Benxi Manchu Autonomous County People's Hospital, and Shandong Affiliated Hospital of Qingdao University. The average and median ages of the selected research participants were 49.61 ± 7.195 years, and 49 years, respectively.

Methods

All women in the study underwent HPV testing and the thinprep cytology test (TCT). Community population and hospital outpatient opportunistic screenings were used to evaluate whether participants met the following inclusion criteria: (1) lived in Liaoning for > 3 years; (2) a history of sexual activity; (3) no sexual activity, vaginal medication, or drug flushing within 1 week before examination; (4) no serious organ dysfunction or mental illness; (5) voluntary participation and signed consent; and 5) willingness to complete the questionnaire survey. The exclusion criteria were: (1) women who were pregnant, lactating, or menstruating; (2) history of cervical surgery or hysterectomy; and (3) diagnosed with a tumor and being treated for other serious internal and external diseases. This study was approved by the ethics committee of Liaoning Cancer Hospital.

Sample Collection

Liquid-Based Cytology

TCT (Thinprep cytologic test purchased from Nanjing Xinbaishi Technology Co., Ltd.). The examination application forms of each patient was filled and their name and age and collection date was placed on the specimen bottle (liquid-based cytology preservation solution vial). The completely filled application form was checked to ensure the information was consistent with that on the preservation solution vial. The cervical brush was then placed in the preservation solution and rinsed by plunging it up and down 20 times, while spreading and rotating the bristles back and forth to make the cells fall into the solution.

Finally, the cervical brush was rapidly rotated to release the collected cells into the preservation solution, the sampler was discarded ensuring the brush head was not left in the bottle, and then it was sent to the pathology department for liquid-based analysis. The operating procedures of the system for the TCT, which is a liquid-based cytology production process, includes the following three steps, which were strictly adhered to: cell mixing, collection, and transfer. Briefly, the sample was placed into 95% alcohol for wet fixation, the production was completed, and then the next staining and diagnosis step was performed. The negative slides should be stored for 1 year, whereas the positive slides should be stored for a long time. Finally, the diagnosis was made by a full-time cytological diagnosis doctor in the pathology department.

Human Papillomavirus Detection Technology

The detection kit was purchased from Beijing Haolejie Healthcare Medical Equipment Company (Beijing). The E6/E7 mRNA detection kit was used to detect the following 14 high-risk HPV mRNAs known to cause cervical cancer (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68), and displayed the detection results of 16 and 18 to provide more clinical guidance information. The kit detects E6 and E7 mRNA of high-risk HPV to avoid missing the detection of high-level lesions and cancer, which could result from only detecting the L1 region. Furthermore, no cross reaction occurred with low-risk HPV and there were fewer false positives, which reduced unnecessary colposcopies and over diagnosis.

Questionnaire and Survey

The questionnaire used in this study adopts the principle of voluntariness and was designed professionally by the members of our research group. The questionnaire includes questions on the participants' personal information; relevant medical, reproductive, sexual activity, smoking, contraceptive, and sports history; educational level; and economic income.

Standard of Referral Colposcopy

Cervical exfoliative cells were examined using cervical liquid-based thinprep cytology test (TCT). The diagnostic criteria are based on the Bethesda system (TBS) classification: atypical squamous cells of undetermined significance (ASCUS) without clear diagnostic significance, excluding ACS-cannot exclude high-grade squamous intraepithelial lesion (ASC-H), low-grade squamous intraepithelial lesion (LSIL), high-grade squamous epithelial lesion (HSIL), squamous cell

TABLE 1 | Basic information of cervical cancer screening population.

Characteristics		N	*n* (%)
Age (years)	35–40	1,271	11.95
	41–45	1,891	17.79
	46–50	2,788	26.23
	51–55	2,325	21.88
	56–60	1,546	14.55
	61–65	807	7.60
Ethnicity	Han	8,932	84.04
	Others	1,696	15.9
Marital status	Unmarried	230	2.16
	Married	9,983	93.93
	Divorced	285	2.68
	Widowed	130	1.22
Profession	Head of party and enterprise unit	801	7.54
	Professional skilled worker	2,312	21.75
	Office and related personnel	1,137	10.70
	Social production and life service personnel	1,411	13.28
	Agriculture, forestry, animal husbandry, and fishery production and auxiliary personnel	368	3.4610
	Production and related personnel	1,107	10.42
	Soldier	8	0.08
	Others who are difficult to classify	2,647	24.91
	Others	837	7.88
Educational level	Junior high school and below	3,080	28.98
	Senior high school	3,341	31.44
	College degree or above	4,205	39.57
Total		10,628	100.0

carcinoma (SCC), and atypical glandular epithelial cells (AGCs) (Nayar and Wilbur, 2015).

For HPV+, TBS classified ASCUS or above, or both, or clinically suspicious abnormalities, a colposcopy is further recommended. The examination results suggested the need for a multipoint tissue biopsy of cervical lesions and bite the tissue for pathological examination. The pathological results were the gold standard, and the diagnostic criteria included normal or inflammatory reaction and cervical intraepithelial neoplasia-intraepithelial neoplasia (CIN) and cervical cancer. CIN is divided into CIN I, CIN II, and CIN III (Kclurman et al., 2014) according to the following three levels: light, medium, and heavy, respectively.

Technical Quality Control

The quality control evaluation of the cervical exfoliative cell examination was conducted using 20 and 5–10% of the randomly selected positive and negative smears, respectively. All smears were reviewed by experts in the field and the acceptable quality rate of the smear results was 80%. The quality control of the colposcopy examination involved a spot check of 10 and 20% of the normal and abnormal reports, respectively. The results were rechecked by experts in the field and the standard quality rate of the reported results was expected to reach 90%. The quality control of the histopathological examination was performed by spot checking 10% of the pathological sections and the results were recheck by experts. The coincidence rate of diagnostic results was expected to reach 90%.

Statistical Analysis

The statistical analysis of the relevant data was conducted using the statistical package for the social science (SPSS) version 19.0 software and the count data rate (%) was analyzed using the chi-square (χ^2) test. The influencing factors were analyzed using univariate and multivariate logistic regression to evaluate the correlation between the relevant factors mentioned in the questionnaire and HPV infection, cervical cancer, and precancerous lesions.

RESULTS

Characteristics of the Participants

Basic Information of Population

The patient age range was 35–65 years, and 1,271, 1,891, 2,788, 2,325, 1,536, and 807 were 35–40, 41–45, 46–50, 51–55, 56–60, and 61–65 year-old, respectively, corresponding to 11.95, 17.79,

TABLE 2 | Personal history and family history of cervical cancer screening population.

Characteristics		N	*n* (%)
Age at menarche (years)	<12	84	0.79
	12–18	10,464	98.46
	>18	80	0.75
Menopausal	Yes	4,573	43.03
	No	6,055	56.97
Age at menopause	<50	1,404	30.70
	≥50	3,169	69.30
Breastfeeding history	Yes	9,188	86.45
	No	1,440	13.55
Breastfeeding time	≤6 months	972	10.58
	>6 months	8,216	89.42
Sexual partners	0	15	0.14
	1	9,099	85.61
	≥2	629	5.92
Age at first sexual activity	Never	6	0.06
	≤19	242	2.27
	20–30	9,402	88.46
	≥31	91	0.86
Pregnancy history	Yes	9,981	98.61
	No	141	1.39
History of miscarriage	Yes	5,337	53.34
	No	4,668	46.66
Sexual partner's foreskin is too long	Yes	372	3.67
	No	9,765	96.33
Bleeding during intercourse	Yes	412	4.1
	No	9,721	95.9
Cervical cancer vaccine	Yes	32	0.32
	No	10,099	99.68
Abnormal vaginal discharge	Yes	814	8.10
	No	9,237	91.90
Past history of gynecological disease	Yes	1,362	12.82
Family history of cancer	Yes	1,074	10.11

TABLE 3 | Living conditions of cervical cancer screening population.

Characteristics		N	*n* (%)
Smoking	No	9,935	93.49
	Currently smoking	602	5.66
	Smoking before	90	0.85
Secondhand smoke	Yes	3,575	33.64
	No	7,052	66.36
Cooking fumes	Almost everyday	8,550	80.46
	Sometimes	1,844	17.35
	Almost not	233	2.19
Alcohol consumption	No	8,910	83.84
	Currently drinking alcohol	173	1.63
	Previously drank alcohol	1,544	14.53
Exercise	Yes	2,107	19.83
	No	8,520	81.17
Tea drinking	Yes	904	8.51
	No	9,723	91.49
Vegetable consumption	Never eat	329	3.11
	<5 pounds/week	6,756	63.77
	≥5 pounds/week	3,509	33.12
Fruit consumption	Never eat	268	2.53
	<2.5 pounds/week	6,434	60.68
	≥2.5 pounds/week	3,902	36.78
Livestock meat consumption	Never eat	399	3.76
	≤350 g/week	7,841	73.95
	>350 g/week	2,363	22.29
Coarse grain consumption	Never eat	870	8.21
	<1 pounds/week	8,014	75.58
	≥1 pounds/week	1,719	16.21

21.88, 26.23, 21.88, 14.55, and 7.6% of the total population, respectively (**Table 1**).

Personal and Family History

The age of menarche and menopause for most participants was 12–18 years and > 50 years, respectively, accounting for 98.46 and 43.03%, respectively. Furthermore, 86.45% of participants had a history of breastfeeding and 89.42% had a cumulative breastfeeding time > 6 months. In addition, 5.92% of participant had multiple sexual partners and 2.27% had sex for the first time under the age of 19. The results also showed that 53.34% of participants had a history of miscarriage, whereas 3.67 and 4.10% reported prepuce and bleeding during intercourse, respectively, and a history of gynecological diseases and family history of tumor accounted for 12.82 and 10.11%, respectively (**Table 2**).

TABLE 4 | Health-related emotional factors of cervical cancer screening population.

Characteristics		N	*n* (%)
Self-assessed health status	Very good or good	5,288	49.76
	Generally	4,791	45.08
	Not good	548	5.16
Hypertension	Yes	918	8.64
	No	9,709	91.36
Diabetes	Yes	306	2.88
	No	10,321	97.12
Hyperlipidemia	Yes	1,042	9.81
	No	9,585	90.19
Diagnosed with a mental illness	Yes	27	0.25
	No	10,600	99.75
Experienced a negative life event	No	8,210	77.26
	1–2 piece	2,347	22.09
	3 pieces and above	70	0.66
Mental depression	No	5,997	56.43
	Occasionally	4,077	38.36
	>1 month	281	2.64
	>6 months	272	2.56
Anxiety	No	6,108	57.48
	Occasionally	4,066	38.26
	>1 month	286	2.69
	>6 months	167	1.57
Sleep quality	Good	6,788	63.88
	Hard to fall asleep	793	7.46
	Wake up early	1,266	11.91
	Sleep well	1,649	15.52
	Wake up at night	131	1.23
When you encounter difficulties, can get support from these people	Husband	9,294	87.46
	Parents	5,918	55.69
	Children	6,713	63.17
	Brothers and sisters	6,117	57.56
	Friends	5,302	49.89
	Colleagues	2,281	21.46
	No	80	0.75

Participant Living Habits

Among the cervical cancer screening population, 56% were current smokers, 0.85% had quit smoking, 33.64% were passive smokers, and 80.46% were exposed to cooking fumes almost daily. In addition, 14.53 and 1.63% of the participants had a history of drinking and were still drinking, respectively, whereas 81.17% did not often participate in outdoor physical exercise and 91.49% did not drink tea. More than 60% of the participants had an insufficient intake of fresh vegetables and fruits, consuming > 5 catties/week and > 2.5 catties/week, respectively. Approximately 75% of human and animal meat intake does not meet the 50–100 g daily requirements of the Dietary Guidelines for Chinese Residents, as shown in **Table 3**.

Health and Emotional Status of Study Population

Among the cervical cancer screening population, 49.76, 8.64, 2.88, and 9.81% had very good or good health status, a history of hypertension, a history of diabetes, hyperlipidemia, mental illness, respectively. Furthermore, 22.75, 43.57, and 36.12% of the study population experienced recent negative life events, mental depression or anxiety symptoms, and poor sleep quality, respectively (**Table 4**).

Population Screening Willingness

Among the cervical cancer screening population, 16.54% thought they could easily develop cancer, whereas 27.15% had received cancer screening, and the cancer screening cost was completely covered by the government for up to 70.88% of the individual participants. In addition, > 78.28% of the study population expressed the willingness to fully accept cancer screening and the main reasons for not accepting cancer screening were time and no obvious symptoms.

Most participants were willing to undergo subsequent screening and could bear the expenses. The acceptable proportion of the out-of-pocket expenses was < 200 yuan in most instances. The data showed that 95.29% of the study population reported a willingness to return for a visit in case of abnormal results, and the main reason for those who were unwilling to was concerns that the examination might be painful. Furthermore, 87.1% of participants expressed a willingness to try a more effective new screening technology, with an acceptable out-of-pocket cost of < 200 yuan. The unwillingness to undergo new technology-based screening was mainly attributable to concerns that the interpretation and utilization of the screening results were unclear, as shown in **Table 5**.

Human Papillomavirus and Thinprep Cytology Test Distribution

We recruited 10,628 women, aged 35–65 years, to participate in this study, including 626 who did not qualify and were subsequently excluded through the questionnaire survey results. Furthermore, 10,002 women were screened for cervical cancer, and 11 were subsequently excluded because they lacked a specimen and, thus, had no HPV examination results. Finally, 9991 participants were included in the statistical analysis, with an average and median age of 49.51 ± 7.188 and 49 years, respectively (**Figure 1**). The cytological examination result

TABLE 5 | Screening willingness of cervical cancer screening population.

	Characteristics	N	*n* (%)
Do you think you are prone to cancer	Yes	1,738	16.54
	No	8,769	83.46
Have you ever been screened for cancer	Yes	2,885	27.15
	No	7,741	72.85
Who bears the cost of cancer screening	It is all borne by the government and not paid by individuals	1,979	70.88
	Some expenses shall be borne by individuals	492	17.62
	All expenses shall be borne by individuals	194	6.95
	No idea	127	4.55
To what extent do you accept cancer screening	Totally acceptable	2,166	78.28
	Acceptable	557	20.13
	Difficulty in accepting	31	1.12
	Unacceptable	13	0.47
Reasons for not participating in cancer screening	Economic reasons	2,362	22.23
	Time reason	4,425	41.64
	The procedure is cumbersome and laborious	3,244	37.4
	Examination can cause pain	3,409	30.53
	I do not think there are any symptoms in my body. It's unnecessary	3,728	35.09
	Physical condition does not allow	81	0.76
	Unaccompanied	90	0.85
If the examination result is abnormal would you be willing to be checked again	Yes	9,298	88.06
	No	1,262	11.95
Would you like to be checked again	Yes	8,706	94.72
	No	485	5.28
What is the acceptable examination fee to you	<100 yuan	1,118	12.85
	100–199 yuan	4,037	46.39
	200–299 yuan	1,734	19.92
	≥300 yuan	1,814	20.84
Are you willing to make a return visit	Yes	9,333	95.29
	No	1,154	11.00
Reasons for not willing to make a return visit/recheck	Economic reasons	832	20.2
	Time reason	1,268	30.7
	The inspection is cumbersome and laborious	1,532	37.1
	Examination can cause pain	1,918	46.4
	I do not think there are any symptoms in my body. It is unnecessary	1,429	34.6
	Physical condition does not allow	58	1.4
	Unaccompanied	28	0.6
Are you willing to accept new technology	Yes	9,170	87.1
	No	1,360	12.9
How much are you willing to pay for the new technology at your own expense	<100 yuan	1,353	14.8
	100–199 yuan	4,410	48.3
	200,299 yuan	1,622	17.8
	≥300 yuan	1,736	19.0
Reasons for reluctance to accept new technology screening	Question the scientific validity and safety of the new method	1,748	38.9
	Unclear interpretation and utilization of screening results	2,548	56.7
	High cost	1,866	41.6
	The old method is reliable, there is no need to use the new method	731	16.3
	Concerned about the pain of new screening methods	432	8.8

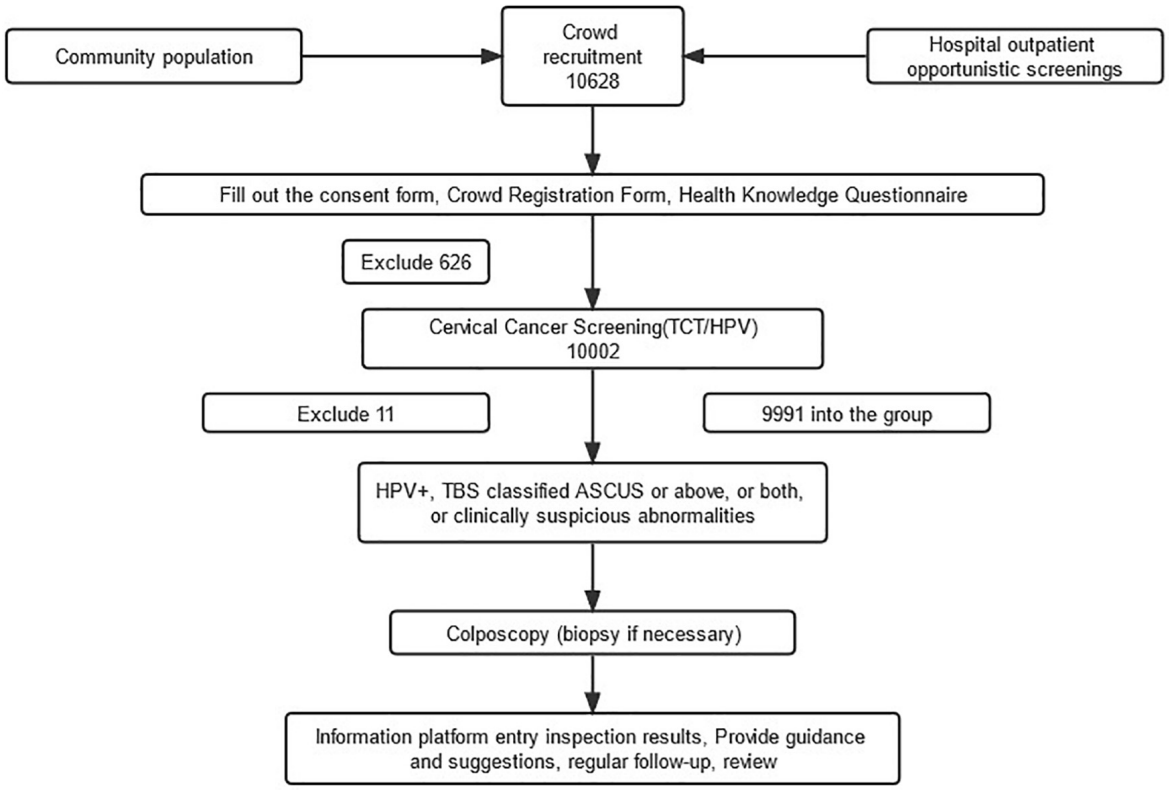

FIGURE 1 | The CONSORT flow of study.

showed 721 TCT + cases, with 7.22% ASCUS + , which included 4.91% ASCUS, 0.49% ASC-H, 1.26% LSIL, 0.33% HSIL, 0.15% AGC-not otherwise specified (AGC-NOS), and 0.06% AGC-prone to cancer (AGC-N). In contrast, SCC, cervical carcinoma *in situ*, and adenocarcinoma were not detected (**Table 6**).

There were 1244 cases of HPV infection, and the positive detection rate was 12.45%. Among the HPV + women, 1213 were further tested for HPV typing, which showed that 184 (15.2%) and 65 (5.4%) were HPV 16 + and 18/45 + with detection rates of 1.84 and 0.65%, respectively. Furthermore, 980 (80.8%) patients were positive for other high-risk types of HPV, with a detection rate of 9.81%. Among the tested patients, 16 women had mixed HPV infections of the above three groups. The positive rate of HPV16 and HPV18/45 in the total screened population was 2.5% (247/9,991) (**Table 7**).

Age Specificity of High-Risk Human Papillomavirus Infection

In this study, 9,991 participants were recruited for high-risk HPV screening in 2018–2021, which included 1,244 (12.45%) that were positive for high-risk HPV. The average age of the women was 49.99 (35–65) years and they were all divided into the following five age groups: 35–40, 41–45, 46–50, 51–55, and 56–65 years old. The prevalence of high-risk HPV infection had two peaks at 35–40, 55–60/61–65 years with infection rates of 13.09, 14.29, and 15.80%, respectively (**Table 8**).

Colposcopy

Further colposcopy was performed in 1,571 cases. There were 1,015 cases with complete colposcopy report, and the referral rate was 64.61%. There were 812 abnormal cases suspected to be diagnosed by colposcopy, 673 cases of low-grade lesions, and 66 cases of high-grade lesions, with a detection rate of 0.66% (**Table 9**).

Single Factor Analysis of Risk Factors of Human Papillomavirus Infection

The results showed that 28 potential risk factors may be related to high-risk HPV infection. Univariate logistic analysis showed that HPV infection was correlated with different age groups, educational level, age of first sexual activity, number of sexual partners, contraceptive methods, and TCT positivity ($P < 0.05$) (**Table 10**).

Logistic Regression Analysis of Risk Factors of Human Papillomavirus Infection

The results of the multivariate unconditional logistic regression analysis showed that high school and below [odds ratio

(OR) = 1.279 (1.129–1.449), $P < 0.001$], initial sexual activity age ≤ 19 years old [OR = 1.517 (1.080–2.129), $P = 0.016$], number of sexual partners > 1 [OR = 1.310 (1.044–1.644), $P = 0.020$], ASCUS and above [OR = 11.891 (10.105–13.993), $P < 0.001$], non-condom contraception [OR = 1.255 (1.059–1.487), $P = 0.009$], and HSIL and above [OR = 1.541 (1.430–1.662), $P < 0.001$] were risk factors for HPV infection. Compared with women aged 56–65 years, the HPV infection rate of those 35–45 [OR = 0.810 (0.690–0.950), $P = 0.010$] and 46–55 [OR = 0.79 (0.683–0.915)] years old ($P = 0.002$) decreased significantly (**Table 11**).

Univariate Logistic Regression Analysis of Cervical Cancer and CIN II/III Univariate Analysis

Among the 9,991 patients included in the analysis, 1,004 had complete pathological results, including follow-up update, whereas 109 and 895 cases were CINII + and LSIL/inflammation, respectively. Furthermore, 136 and 1 patients were referred for a colposcopy because of TCT/HPV and clinically suspected abnormalities, respectively. The result of the colposcopy evaluation was negative and, therefore, no biopsy was performed. In addition, 137 patients were classified as LSIL and inflammation, and in the absence of pathological results, 8,210 women with double negative screening results (TCT-/HPV-) were regarded as having no cervical lesions. Therefore, 109 CINII + , 5 invasive carcinoma and 9, 242 LSIL/inflammation cases, respectively were included in the statistical analysis.

The detection rate of CINII + in cervical high-grade intraepithelial neoplasia was 1.09% (109/9,991), and the early diagnosis rate of cervical cancer was 95.41%(104/109). The results of the 28 logistic regression analyses using CINII + and diet as dependent variables showed a significant correlation between educational level, breastfeeding, and TCT positivity ($P < 0.05$). The detection rate of CINII + increased significantly with increasing age, early sexual activity, multiple sexual partners, total family income ≤ 50,000 yuan, bleeding during sexual intercourse and extramarital sex, but there was no significant difference (**Table 12**).

TABLE 6 | Thinprep cytology test (TCT) cervical cancer screening results.

TBS classification diagnostic criteria	N	*n* (%)
NILM	9,271	9279
ASCUC	491	4.91
ASC-H	49	0.49
LSIL	126	1.26
HSIL	33	0.33
SCC	0	0.00
AGC-NOS	15	0.15
AGC-N	6	0.06
AIS	0	0.00
Adenocarcinoma	0	0.00
Total	9,991	100.00

TBS, the Bethesda system; NILM, negative for intraepithelial lesion or malignancy; ASCUC, atypical squamous cells of undetermined significance; ASC-H, atypical squamous cells cannot exclude high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; SCC, squamous cell carcinoma; AGC-NOS, atypical glandular epithelial cells-not otherwise specified; AGC-N, atypical glandular epithelial cells-not prone to cancer; AIS, adenocarcinoma in situ.

TABLE 7 | Human papillomavirus (HPV) screening results in cervical cancer.

HPV		N	*n* (%)
Detection condition	(+)	1,244	12.45
	(–)	8,747	87.55
Typing	16 type	184	1.84
	18/45 type	65	0.65
	Other	980	9.81
Total		9,991	100.00

TABLE 8 | Age specificity of human papillomavirus (HPV) infection.

Age (years)	Total	HPV+	HPV–	Positive rate (%)	χ^2	*P*
35–40	1,230	161	1,069	13.09	17.44	***0.0037***
41–45	1,792	201	1,591	11.22		
46–50	2,614	308	2,306	11.78		
51–55	2,192	254	1,938	11.59		
56–60	1,435	205	1,230	14.29		
61–65	728	115	613	15.80		

Statistically significant (P < 0.05) values are indicated in bold.

TABLE 9 | Colposcopy screening.

Initial diagnostic impression		N	*n* (%)	Detection rate (%)
Normal		203	20.00	
Abnormal	LSIL	673	66.31	6.74
	HSIL	61	6.01	0.61
	Suspected cancer	5	0.49	0.05
	Other	73	7.19	0.73
Total		1,015	–	–

Multivariate Logistic Regression Analysis of Cervical Cancer and CIN2/3 Risk Factors

The factors that were statistically different in the univariate unconditional logistic regression analysis were further analyzed using multivariate unconditional logistic regression. The results showed that high school and below [OR = 1.577 (1.042–2.387), $P = 0.031$], not breastfeeding [OR = 1.763 (1.109–2.804), $P = 0.017$], ASCUS and above [OR = 42.396 (28.042–64.098), $P < 0.001$] were potential risk factors for cervical cancer and precancerous lesions (**Table 13**).

DISCUSSION

In this multicenter, cross-sectional population study in women, we found that the overall prevalence of HR-HPV was 12.45%, which is lower than the national average. HPV infection can occur at any age and is related to age. In this study, the age-specific distribution showed a bimodal curve, and the first peak appeared in the 35–40-year-old age group, with the infection rate of middle-aged women showing a low trend. The second

TABLE 10 | Comparison of different information and living habits with human papillomavirus (HPV) positive detection rate.

Characteristics		N	(+)	(–)	*n* (%)	χ^2	*P*-value
Age (years)	35–45	3,022	362	2,660	11.98	15.04	***0.0009***
	46–55	4,806	562	4,244	11.69		
	56–65	2,163	320	1,843	14.79		
Family history of cancer	Yes	999	139	860	13.91	2.167	0.1410
	No	8,989	1,105	7,884	12.29		
Level of education	High school and below	6,040	815	5,225	13.49	15.08	***0.0001***
	College degree or above	3,947	429	3,518	10.87		
Age at initial sexual experience	≤19 year	236	42	194	17.8	5.817	***0.0154***
	>19 year	9,262	1,157	8,105	12.49		
Number of sexual partners	Multiple (≥ 2)	613	96	517	15.66	5.48	***0.0192***
	1	8,893	1,104	7,789	12.41		
Number of abortions	>1	2,209	279	1,930	12.63	1.119	0.2902
	1	2,970	405	2,565	13.64		
Number of deliveries	>1	1,043	138	905	13.23	0.6582	0.4172
	≤1	7,557	933	6,624	12.35		
Number of marriages	>1	406	55	351	13.55	0.3209	0.5710
	≤1	9,093	1,145	7,948	12.59		
Contraceptive methods used	Others	6,722	853	5,869	12.69	7.071	***0.0078***
	Condom	1,375	139	1,236	10.11		
Sexual partner has long foreskin	Yes	363	44	319	12.12	0.0456	0.8309
	No	9,537	1,192	8,345	12.50		
Bleeding during intercourse	Yes	405	56	349	13.83	0.7019	0.4022
	No	9,491	1,179	8,312	12.42		
Leucorrhea abnormality	Yes	795	88	707	11.07	1.627	0.2021
	No	9,018	1,139	7,879	12.63		
Vaccination	No	9,863	1,232	8,631	12.49	–	0.7909
	Yes	32	3	29	9.38		
Extramarital sex	Yes	66	11	55	16.67	1.055	0.3044
	No	9,830	1,226	8,604	12.47		
Ethnicity	Han	8,410	1,044	7,366	12.41	2.998	0.2234
	Man	1,431	175	1,256	12.23		
	Others	146	25	121	17.12		
Total household income	≤50,000 yuan	4,786	623	4,163	13.02	2.178	0.1400
	>50,000 yuan	4,181	501	3,680	11.98		
Menopausal	No	4,279	561	3,718	13.11	2.973	0.0846
	Yes	5,711	683	5,028	11.96		
Breastfed	No	1,312	163	1,149	12.42	0.001139	0.9731
	Yes	8,678	1,081	7,597	12.46		
Smoking history	Yes	619	68	551	10.99	1.303	0.2537
	No	9,371	1,176	8,195	12.55		
Alcohol consumption	Yes	1,599	198	1,401	12.38	0.008487	0.9266
	No	8,391	1,046	7,345	12.47		
Physical exercise	No	8,043	993	7,050	12.35	0.3278	0.5130
	Yes	1,947	251	1,696	12.89		
Tea drinking	No	9,145	1,144	8,001	12.51	0.3235	0.5695
	Yes	845	100	745	11.83		
Fresh vegetable consumption	No	6,594	811	5,783	12.30	0.4068	0.5236
	Yes	3,366	429	2,937	12.75		
Fresh fruit consumption	No	6,270	755	5,515	12.04	2.57	0.1089
	Yes	3,699	486	3,213	13.14		
Meat consumption	No	7,765	971	6,794	12.49	0.01798	0.8933
	Yes	2,202	273	1,929	12.40		
Coarse grain consumption	No	8,364	1,035	7,329	12.37	0.5294	0.4669
	Yes	1,604	209	1,395	13.03		
Are you in good health	No	504	71	433	14.09	1.301	0.2540
	Yes	9,486	1,173	8,313	12.36		
TCT	ASCUS and above	720	393	327	54.58	1,264	***<0.0001***
	NILM	9,271	851	8,420	9.18		

TCT, thinprep cytology test; ASCUC, atypical squamous cells of undetermined significance; NILM, negative for intraepithelial lesion or malignancy. Statistically significant (P < 0.05) values are indicated in bold.

TABLE 11 | Multivariate analysis of risk factors affecting human papillomavirus (HPV) infection.

Characteristics	SE	*P*-value	OR (95%)
35–45 year	0.082	0.010	0.810 (0.690∼0.950)
46–55 year	0.075	0.002	0.79 (0.683∼0.915)
High school and below	0.064	<0.001	1.279 (1.129∼1.449)
Initial age of sexual life ≤ 19 year	0.173	0.016	1.517 (1.080∼2.129)
Number of sexual partners > 1	0.116	0.020	1.310 (1.044∼1.644)
ASCUS and above	0.083	<0.001	11.891 (10.105∼13.993)
Contraceptive methods other than condoms	0.087	0.009	1.255 (1.059∼1.487)
HSIL and above	0.038	<0.001	1.541 (1.430∼1.662)

OR, odds ratio; ASCUC, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion.

peak of high-risk HPV infection was observed in women aged 55–60 and 61–65 years, who were born in the 1960s–1970s and the economy of China has improved considerably since then. In addition, the education level of these age groups is significantly lower than that of the younger women, which contributes to their lack of knowledge about HPV infection. In addition, the low level of autoimmunity and hormones further impairs the resistance of the cervix to HPV infection, which may explain the significant increase in their HPV + rate.

Presently, there is no specific or effective drug treatment for HPV infection, and although a vaccine has been developed and listed in China, it is expensive and does not prevent all subtype infections (You et al., 2020). Therefore, the prevention of HPV infection is very important. Number studies have reported the following as some factors to be related to the incidence of cervical cancer: sexual activity, early first sexual activity, premature delivery, prolificacy, high-risk HPV infection, and smoking. The following are some of the factors related to cervical intraepithelial neoplasia: sexual activity, HPV infection, smoking, premature sexual activity, sexually transmitted diseases, low economic status, use of oral contraceptives, and immunosuppression (Shannon et al., 2017).

In this study, multicenter cervical screening was used to analyze the incidence of HPV infection, cervical cancer, and high-grade precancerous lesions in women. The results led us to conclude that high school and below, initial sexual life age ≤ 19 years old, number of sexual partners > 1, ASCUS and above, non-condom contraception, HSIL and above were all risk factors of HPV infection. The reason may be that women with a low educational level lack the awareness of cervical screening and prevention strategies (Williams et al., 2019). Women ≤ 19 years old who are sexually active are prone to HPV infection because of their immature cervical development and incompletely developed autoimmune function (Morris et al., 2019). Furthermore, a higher number of sexual partners increases the potential exposure to HPV infection, rendering an individual prone to HPV infection.

Numerous studies have shown that an active sex life is closely related to the occurrence and development of cervical cancer. Literature reports from countries other than China state that factors such as sexual partners and frequency are closely related to cervical cancer (Torres-Poveda et al., 2019). Condoms block pathogens from damaging the cervical mucosa and inhibiting the immune function, whereas cleaning the vulva reduces the probability of infection. Both processes reduce the stimulatory effects of semen on the cervical mucosa, which is consistent with the research results of Hariri and Warner (2013) on male condoms that indicates that they contribute to reducing the transmission of HPV. The incidence of CINII and above was positively correlated with HPV infection.

Cervical lesions mostly occur in married women and it is the most serious cervical disease, and shows higher occurrence in younger women (Arbyn et al., 2020). Cervical cancer, which is mainly caused by long-term cervical lesions, has a high mortality, which is gradually increasing (Arbyn et al., 2020). Therefore, preventing cervical lesions is extremely significant for women (Cree et al., 2021). HPV infection has always been considered an important factor in the development of cervical lesions, but most women can be protected by the autoimmune system. In addition, a few women will continue to be infected and develop cervical cancer over time. Therefore, the need to actively understand cervical lesions, HPV infection, and the influencing factors has been increasingly attracting clinical attention. Studies suggest that high-risk HPV infection was detected in 99.7% of cervical cancer patients (Crosbie et al., 2013).

HPV has been reported to be transmitted through direct or indirect skin or sexual contact (Wierzbicka et al., 2022). The infection occurs worldwide with approximately 4–20% of healthy individuals harboring the infection, and the cumulative lifetime infection rate is 60–70%. Sustained expression of the E6 and E7 HPV virus subtypes has been found to be closely related to the growth of HPV-infected cancer cells (Pinatti et al., 2017). HPV infection is the most important and definite cause of cervical cancer (Araldi et al., 2018). In this study, we found that an increasing histopathological grade of HR-HPV infection was associated with the incidence of cervical cancer. The results showed that a high school education and below, non-lactation, and TCT positivity were statistically significant ($P < 0.05$) risk factors of cervical cancer and CIN II/III.

The CINII + detection rate increased with increasing age, early sexual activity, multiple sexual partners, total family income ≤ 50,000 yuan, bleeding during sexual intercourse, and extramarital sex, but not significantly. This study has been compared with other similar related research data in the near future (see **Supplementary Table 1** for details) (Kitamura et al., 2021; Tagne Simo et al., 2021; Niu et al., 2022). As a preventable

TABLE 12 | Comparison of detection rates of high-grade squamous intraepithelial lesion positive (CINII +) in populations with different basic characteristics.

Characteristics		N	CINII +	LSIL/inflammation	*n* (%)	χ^2	*P*-value
Age (years)	35–45	2,804	30	2,774	1.07	1.76	0.4147
	46–55	4,542	50	4,492	1.10		
	56–65	2,005	29	1,976	1.45		
Family history of cancer	Yes	952	10	942	1.05	0.06594	0.7973
	No	8,396	96	8,300	1.14		
Level of education	High school and below	5,657	77	5,580	1.36	4.727	***0.0297***
	College degree or above	3,690	32	3,658	0.87		
Age at initial sexual activity	≤19 year	211	3	208	1.42	–	0.7410
	>19 year	8,688	103	8,585	1.19		
Number of sexual partners	>1	567	9	558	1.59	1.177	0.2780
	≤1	8,339	91	8,248	1.09		
Number of abortions	>1	2,088	32	2,056	1.53	0.1755	0.6753
	≤1	2,810	39	2,771	1.39		
Number of deliveries	>1	966	13	953	1.35	0.1977	0.6566
	≤1	7,120	84	7,036	1.18		
Number of marriages	>1	383	7	376	1.83	1.378	0.2405
	≤1	8,516	99	8,417	1.16		
Contraceptive methods	Others	6,332	75	6,257	1.18	0.08034	0.7768
	Condom	1,283	14	1,269	1.09		
Sexual partner has long foreskin	Yes	341	3	338	0.88	–	>0.9999
	No	8,926	104	8,822	1.17		
Bleeding during sexual intercourse	Yes	373	7	366	1.88	1.843	0.1746
	No	8,891	99	8,792	1.11		
Leucorrhea abnormality	Yes	738	10	728	1.36	0.2504	0.6168
	No	8,443	97	8,346	1.15		
Vaccination	No	1,125	106	1,019	9.42	–	>0.9999
	Yes	3	0	3			
Extramarital sex	Yes	63	2	61	3.17	–	0.1646
	No	9,198	105	9,093	1.14		
Ethnicity	Han	7,860	93	7,767	1.18	0.3176	0.8532
	Man	1,350	14	1,336	1.04		
	Others	137	2	135	1.46		
Total household income	≤50,000 yuan	5,247	65	5,182	1.24	3.526	0.0604
	>50,000 yuan	3,856	32	3,824	0.83		
Menopausal	No	5,334	62	5,272	1.16	0.001261	0.9717
	Yes	4,016	47	3,969	1.17		
Breastfed	No	1,240	23	1,217	1.85	5.891	***0.0152***
	Yes	8,110	86	8,024	1.06		
Smoking history	Yes	583	9	574	1.54	0.7709	0.3799
	No	8,767	100	8,667	1.14		
Alcohol consumption	Yes	1,476	18	1,458	1.22	0.04393	0.8340
	No	7,874	91	7,783	1.16		
Physical exercise	Yes	1,836	18	1,818	0.98	0.6814	0.4091
	No	7,514	91	7,423	1.21		
Tea consumption	No	8,567	101	8,466	1.18	0.1539	0.6948
	Yes	783	8	775	1.02		
Fresh vegetable consumption	No	6,124	66	6,058	1.08	1.302	0.2538
	Yes	3,196	43	3,153	1.35		
Fresh fruit consumption	No	5,851	67	5,784	1.15	0.07296	0.7871
	Yes	3,479	42	3,437	1.21		
Meat consumption	No	7,265	85	7,180	1.17	0.0007221	0.9786
	Yes	2,064	24	2,040	1.16		
Coarse grain consumption	No	7,827	91	7,736	1.16	0.01458	0.9039

(Continued)

TABLE 12 | (Continued)

Characteristics		N	CINII +	LSIL/inflammation	*n* (%)	χ^2	*P*-value
	Yes	1,501	18	1,483	1.20		
Are you in good health	No	464	4	460	0.86	0.3908	0.5319
	Yes	8,886	105	8,781	1.18		
TCT	ASCUS and above	513	74	439	14.42	828.3	***<0.0001***
	NILM	8,838	35	8,803	0.40		

TCT, thinprep cytology test; ASCUC, atypical squamous cells of undetermined significance; NILM, negative for intraepithelial lesion or malignancy. Statistically significant (P < 0.05) values are indicated in bold.

TABLE 13 | Multivariate analysis of risk factors for cervical cancer and precancerous lesions.

Characteristics	SE	*P*-value	OR (95%)
High school education and below	0.211	0.031	1.577 (1.042∼2.387)
Not breastfeeding	0.237	0.017	1.763 (1.109∼2.804)
ASCUS and above	0.211	<0.001	42.396 (28.042∼64.098)

SE, standard error; OR, odds ratio; ASCUC, atypical squamous cells of undetermined significance.

malignant tumor, the occurrence and development of cervical cancer is a gradual process that takes several years to progress from intraepithelial lesions to invasive disease. Effective screening methods for cervical cancer could ensure early detection, diagnosis, and treatment of cervical precancerous lesions and, to a certain extent, reduce its incidence rate and mortality.

CONCLUSION

Cervical cancer is a prominent public health problem and proms focusing on its prevention and control should be strengthened. In this study, multicenter cervical screening of women to determine the HPV infection rate and incidence of cervical high-grade precancerous lesions, demonstrated a high school education and below, initial age of sexual activity $\leq$ 19 years old, more than one sexual partner, ASCUS and above, non-condom contraception, and HSIL and above as risk factors of HPV infection. Furthermore, a high school education and below, non-lactation, and a positive TCT result were identified as risk factors of cervical cancer and high-grade precancerous lesion infection. In conclusion, the investigation of risk factors for HPV infection and cervical high-grade precancerous lesions are of great significance for reducing the incidence rate of cervical cancer. Consequently, we recommend the establishment of strategies to enhance the awareness of the risk factors of high-risk HPV infection; to promote healthy sexual behavior, life, and health habits; and to improve immunity and strengthen physical exercise, which could effectively reduce the incidence rate of cervical cancer.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Liaoning Cancer Hospital (No. 20180106). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HP and CW designed and supervised the project. DY, JZ, and XC collected clinical data samples. DY collected and processed data and data analysis, and drafted the manuscript. JM translated and polished the manuscript. All authors reviewed, discussed, and edited versions of the final report manuscript.

FUNDING

This study was supported by the National Key Research and Development Program (2016YFC1303001).

ACKNOWLEDGMENTS

We thank the staff from Cancer Hospital of China Medical University, who took part in the study. We would also like to thank JM for the great help offered in manuscript revision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.914516/full#supplementary-material

REFERENCES

Abulizi, G., Li, H., Mijiti, P., Abulimiti, T., Cai, J., Gao, J., et al. (2017). Risk factors for human papillomavirus infection prevalent among Uyghur women from Xinjiang. *Chin. Oncotarget* 8, 97955–97964.

Aimagambetova G, and Azizan A. (2018). Epidemiology of HPV Infection and HPV-related cancers in Kazakhstan:a review. *Asian Pac. J. Cancer Prev*. 19, 1175–1180.

Araldi, R. P., Sant'Ana, T. A., Módolo, D. G., Melo, T. C., Spadacci-Morena, D. D., de Cassia Stocco, R., et al. (2018). The human papillomavirus (HPV)-related cancer biology: an overview. *Biomed. Pharmacother.* 106, 1537–1556.

Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J., et al. (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob. Health* 8:e191–e203.

Bekmukhambetov, Y. Z., Balmagambetova, S. K., Jarkenov, T. A., Nurtayeva, S. M., Mukashev, T. Z., and Koyshybaev, A. K. (2016). Distribution of high risk human papillomavirus types in Western Kazakhstan — retrospective analysis of PCR data. *Asian Pac. J. Cancer Prev.* 17, 2667–2672.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: gLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424.

Cao, W., Chen, H. D., Yu, Y. W., Li, N., and Chen, W. Q. (2021). Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin. Med. J.* 134, 783–791. doi: 10.1097/CM9.0000000000001474

Cordeiro, M. N., De Lima, R. C. P., Paolini, F., Melo, A. R. D. S., Campos, A. P. F., Venuti, A., et al. (2018). Current research into novel therapeutic vaccines against cervical cancer. *Expert Rev. Anticancer Ther. Apr.* 18, 365–376.

Cree, I. A., Indave Ruiz, B. I., Zavadil, J., McKay, J., Olivier, M., Kozlakidis, Z., et al. (2021). IC3R participants. The International Collaboration for Cancer Classification and Research. *Int. J. Cancer.* 148, 560-571. doi: 10.1002/ijc.33260

Crosbie, E. J., Einstein, M. H., Franceschi, S., and Kitchener, H. C. (2013). Human papillomavirus and cervical cancer. *Lancet* 382, 889–899.

Giorgi Rossi, P., Ronco, G., Mancuso, P., Carozzi, F., Allia, E., Bisanzi, S., et al. (2022). NTCC2 Working Group. Performance of HPV E6/E7 mRNA Assay as Primary Screening Test. Results from the NTCC2 Trial. *Int. J. Cancer* [Epub ahead of print]. doi: 10.1002/ijc.34120

Hariri, S., and Warner, L. (2013). Condom use and human papillomavirus in men. *J. Infect. Dis.* 208, 367–369.

Jiang, L., Tian, X., Peng, D., Zhang, L., Xie, F., Bi, C., et al. (2019). HPV prevalence and genotype distribution among women in Shandong Province, China:analysis of 94,489 HPV genotyping results from Shandong's largest independent pathology laboratory. *PLoS One* 14, e0210311. doi: 10.1371/journal.pone.0210311

Kclurman, R. J., Carcangiu, M. L., Herrington, C. S., et al. (2014). *WHO Classigication of Tumors of Female Reproductive organs [M] 4thed*. Lyon: International Agency for Research on Cancer.

Kitamura, T., Suzuki, M., Shigehara, K., and Fukuda, K. (2021). Prevalence and Risk Factors of Human Papillomavirus Infection among Japanese Female People: a Nationwide Epidemiological Survey by Self-Sampling. *Asian Pac. J. Cancer Prev.* 22, 1843–1849. doi: 10.31557/APJCP.2021.22.6.1843

Lewandowska, A. M., Lewandowski, T., Rudzki, M., Rudzki, S., and Laskowska, B. (2021). Cancer prevention - review paper. *Ann. Agric. Environ. Med.* 2020:12.

Maria, H., Dana, H., Francoise, M., Michael, P., and Jurgen, W. (2018). Human papillomaviruses in Western Africa: prevalences and risk factors in Burkina Faso. *Arch. Gynecol. Obstet.* 298, 789–796.

McLaughlin-Drubin, M. E., and Munger, K. (2008). Viruses associated with human cancer. *Biochim. Biophys. Acta Mol. Basis Dis.* 1782, 127–150.

Morris, B. J., Hankins, C. A., Banerjee, J., Lumbers, E. R., Mindel, A., Klausner, J. D., et al. (2019). Does Male Circumcision Reduce Women's Risk of Sexually Transmitted Infections, Cervical Cancer, and Associated Conditions? *Front. Public Health* 7, 4. doi: 10.3389/fpubh.2019.00004

Nayar, R., and Wilbur, D. (2015). The Pap Test and Bethesda 2014. *Acta Cytol.* 59, 121–132.

Niu, J., Pan, S., Wei, Y., Hong, Z., Gu, L., Di, W., et al. (2022). Epidemiology and analysis of potential risk factors of high-risk human papillomavirus (HPV) in Shanghai China: a cross-sectional one-year study in non-vaccinated women. *J. Med. Virol.* 94, 761–770. doi: 10.1002/jmv.27453

Ouh, Y. T., Min, K. J., Cho, H. W., Ki, M, Oh, J. K, Shin, S. Y et al. (2018). Prevalence of human papillomavirus genotypes and precancerous cervical lesions in a screening population in the Republic of Korea, 2014–2016. *J. Gynecol. Oncol.* 29:e14. doi: 10.3802/jgo.2018.29.e14

Peter, M., Stransky, N., Couturier, J., Hupé, P., Barillot, E., de Cremoux, P., et al. (2010). Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. *J. Pathol.* 221, 320–330. doi: 10.1002/path.2713

Pinatti, L. M., Walline, H. M., and Carey, T. E. (2017). Human Papillomavirus Genome Integration and Head and Neck Cancer. *J. Dent. Res.* 97, 691–700.

Shannon, B., Yi, T. J., Perusini, S., Gajer, P., Ma, B., Humphrys, M. S., et al. (2017). Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. *Mucosal. Immunol.* 10, 1310–1319.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: gLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 71, 209–249.

Tagne Simo, R., Djoko Nono, A. G., Fogang Dongmo, H. P., Seke Etet, P. F., Fonyuy, B. K., Kamdje, A. H. N., et al. (2021). Prevalence of precancerous cervical lesions and high-risk human papillomavirus types in Yaounde. *Cameroon. J. Infect. Dev. Ctries.* 15, 1339–1345.

Thanapprapasr, D., Deesamer, S., Sujintawong, S., Udomsubpayakul, U., and Wilailak, S. (2012). Cervical cancer screening behaviours among Thai women: results from a cross-sectional survey of 2112 healthcare providers at Ramathibodi Hospital. *Thailand. Eur. J. Cancer Care* 21, 542–547. doi: 10.1111/j.1365-2354.2012.01333.x

Torres-Poveda, K., Ruiz-Fraga, I., Madrid-Marina, V., Chavez, M., and Richardson, V. (2019). High risk HPV infection prevalence and associated cofactors: a population-based study in female ISSSTE beneficiaries attending the HPV screening and early detection of cervical cancer program. *BMC Cancer.*19:1205. doi: 10.1186/s12885-019-6388-4

Wierzbicka, M., San Giorgi, M. R. M., and Dikkers, F. G. (2022). Transmission and clearance of human papillomavirus infection in the oral cavity and its role in oropharyngeal carcinoma - A review. *Rev. Med. Virol.* 22:e2337. doi: 10.1002/rmv.2337

Williams, M. S., Kenu, E., Adanu, A., Yalley, R. A., Lawoe, N. K., Dotse, A. S., et al. (2019).).Awareness and beliefs about cervical cancer, the HPV vaccine, and cervical cancer screening among Ghanaian women with diverse education levels. *J. Cancer Educ.* 34, 897–903. doi: 10.1007/s13187-018-1392-y

You, D., Han, L., Li, L., Hu, J., Zimet, G. D., and Alias, H. (2020). Human Papillomavirus (HPV) Vaccine Uptake and the Willingness to Receive the HPV Vaccination among Female College Students in China: a Multicenter Study. *Vaccines* 8:31. doi: 10.3390/vaccines8010031

Zheng, R. S., Sun, K. X., Zhang, S. W., Zeng, H. M., Zou, X. N., Chen, R., et al. (2019). Report of cancer epidemiology in China, 2015. *Zhonghua Zhong Liu Za Zhi* 41, 19–28.

Zhong, T. Y., Zhou, J. C., Hu, R., Fan, X. N., Xie, X. Y., Liu, Z. X., et al. (2017). Prevalence of human papillomavirus infection among 71,435 women in Jiangxi Province, China. *J. Infect. Public Health* 10, 783–788. doi: 10.1016/j.jiph.2017.01.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 03 August 2022
DOI 10.3389/fmicb.2022.960676

OPEN ACCESS

EDITED BY
Christine A. King,
Upstate Medical University,
United States

REVIEWED BY
Miguel Angel Martinez,
IrsiCaixa, Spain
Flor Pujol,
Instituto Venezolano de
Investigaciones Científicas
(IVIC), Venezuela

*CORRESPONDENCE
Esteban Domingo
edomingo@cbm.csic.es
Celia Perales
celia.perales@cnb.csic.es

SPECIALTY SECTION
This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 03 June 2022
ACCEPTED 11 July 2022
PUBLISHED 03 August 2022

CITATION
García-Crespo C, Vázquez-Sirvent L, Somovilla P, Soria ME, Gallego I, de Ávila AI, Martínez-González B, Durán-Pastor A, Domingo E and Perales C (2022) Efficacy decrease of antiviral agents when administered to ongoing hepatitis C virus infections in cell culture.
Front. Microbiol. 13:960676.
doi: 10.3389/fmicb.2022.960676

Efficacy decrease of antiviral agents when administered to ongoing hepatitis C virus infections in cell culture

Carlos García-Crespo[1,2], Lucía Vázquez-Sirvent[1,3], Pilar Somovilla[1,4], María Eugenia Soria[1,2,3], Isabel Gallego[1,2], Ana Isabel de Ávila[1,2], Brenda Martínez-González[3,5], Antoni Durán-Pastor[1], Esteban Domingo[1,2]* and Celia Perales[1,2,3,5]*

[1]Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, [2]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd) del Instituto de Salud Carlos III, Madrid, Spain, [3]Department of Clinical Microbiology, IIS-Fundación Jiménez Díaz, UAM. Av. Reyes Católicos, Madrid, Spain, [4]Departamento de Biología Molecular, Universidad Autónoma de Madrid, Madrid, Spain, [5]Department of Molecular and Cell Biology, Centro Nacional de Biotecnología (CNB-CSIC), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

We report a quantification of the decrease of effectiveness of antiviral agents directed to hepatitis C virus, when the agents are added during an ongoing infection in cell culture vs. when they are added at the beginning of the infection. Major determinants of the decrease of inhibitory activity are the time post-infection of inhibitor administration and viral replicative fitness. The efficacy decrease has been documented with antiviral assays involving the combination of the direct-acting antiviral agents, daclatasvir and sofosbuvir, and with the combination of the lethal mutagens, favipiravir and ribavirin. The results suggest that strict antiviral effectiveness assays in preclinical trials may involve the use of high fitness viral populations and the delayed administration of the agents, relative to infection onset.

KEYWORDS

daclatasvir, sofosbuvir, favipiravir, ribavirin, direct acting antivirals, viral fitness, delayed drug administration, lethal mutagenesis

Introduction

The intra-host population dynamics of RNA viruses can influence the efficacy of antiviral treatments by providing dynamic mutant spectra in which some genomes may encode amino acids that confer resistance to antiviral agents. Resistance-associated substitutions (RASs) may preexist in viral populations prior to treatment administration, or they may be generated and selected during replication in the presence of the corresponding antiviral agent. RASs are generally specific for an antiviral drug, and they have been described for most viruses for which antiviral treatments have been investigated or implemented. The presence of RAS at a sufficient frequency to permit

virus escape depends on interconnected sets of factors such as the genetic barrier (number and types of mutations required to produce a RAS), the phenotypic barrier (the fitness cost inflicted by RAS), the error rate of the virus during replication, the distance of the population from a clonal origin, and the viral population size [examples and reviews of the effect of such multiple factors can be found in (Richman, 1996; Ribeiro et al., 1998; Ribeiro and Bonhoeffer, 2000; Domingo and Perales, 2012; Perales et al., 2017; Perales, 2018; Li and Chung, 2019; Domingo, 2020)].

The administration of direct-acting antiviral (DAA) agents has been highly successful for the control of chronic hepatitis C virus (HCV) infections, with sustained virological responses of around 95% (Janjua et al., 2021). Selection of RAS is responsible for a large proportion of HCV treatment failures (Di Maio et al., 2017, 2021; Kai et al., 2017; Ceccherini-Silberstein et al., 2018; Dietz et al., 2018; Perpinan et al., 2018; Sorbo et al., 2018; Lombardi et al., 2019; Chen et al., 2020; Malandris et al., 2021; Sarrazin, 2021). RASs have been identified in treatment-naive patients (Costantino et al., 2015; Kai et al., 2017; Li et al., 2017a,b; Esposito et al., 2018; Perales et al., 2018; Yang et al., 2018; Morishita et al., 2020), reflecting the impact of the basal HCV mutation rate and viral population dynamics on the response to antiviral agents.

An alternative mechanism of antiviral resistance was identified in HCV from chronically infected patients. It was revealed by the presence of a class of highly represented amino acid substitutions (HRSs) in the basal viral samples (before treatment onset) of patients who then failed therapy; HRSs were associated with resistance to several DAA treatments, comprising double and triple combinations that included sofosbuvir (isopropyl (2S)-2-[[[(2R,3R,4R,5R)-5-(2,4-dioxopyrimidin-1-yl)-4-fluoro-3-hydroxy-4-methyl-tetrahydrofuran-2-yl]methoxy-phenoxy-phosphoryl]amino]propanoate) and ribavirin (1-β-D-ribofuranosyl-1-*H*-1,2,4-triazole-3-carboxamide) (Soria et al., 2020). This RAS-independent mechanism may account for a sizeable proportion of HCV treatment failures that have been reported in several patient cohorts (Kim et al., 2014; Nakamoto et al., 2014; Mawatari et al., 2018; Uchida et al., 2018; Bellocchi et al., 2019). The molecular mechanism of HRS-mediated antiviral resistance is not known but it may relate to the effect of HCV replicative fitness on antiviral resistance. Antiviral resistance conferred by high viral fitness was revealed in studies with HCV in cell culture, and HRS may reflect a similar fitness effect *in vivo*, although this proposal needs to be proven. The high HCV fitness-mediated resistance to anti-HCV inhibitors in cell culture was documented with disparate classes of antiviral agents. The studies involved measurements with DAAs (sofosbuvir, telaprevir, or daclatasvir), interferon-α (IFN-α), cyclosporine A—which targets the cellular cyclophilin A—lethal mutagens such as the nucleoside analogs favipiravir (T-705; 6-fluoro-3-hydroxy-2-pirazinecarboxamide) and ribavirin [(Sheldon et al., 2014; Gallego et al., 2016, 2018), or the metabolite inhibitor guanosine (Sabariegos et al., 2022); reviewed in (Domingo et al., 2019)].

As a consequence of these observations, we became interested in antiviral designs that can be effective in inhibiting high fitness HCV. In this line, we described that favipiravir and ribavirin exert a synergistic antiviral effect on HCV during replication on human hepatoma Huh-7.5 cells (Gallego et al., 2019). Importantly, combinations of these two analogs extinguished high fitness HCV that was resistant to equivalent total doses of one of the analogs. Synergy may be due to differences in the mechanism of antiviral activity between favipiravir and ribavirin (Furuta et al., 2009; Beaucourt and Vignuzzi, 2014; De Clercq and Li, 2016) in addition to their lethal mutagenesis (error rate enhancing) activity during RNA synthesis. A difference in the preference for the genomic sites mutated by the two analogs—as revealed by deep sequencing—may have contributed to synergy (Gallego et al., 2019).

A critical question that has not been addressed in preclinical experiments with HCV is the effect on antiviral efficacy of the time of addition of the antiviral agents, relative to the time of initiation of the infection. Classic findings on the treatment of microbial infections in general, and in the use of antiretroviral agents for AIDS, indicated benefits of the implementation of treatments early after infection (Ehrlich, 1913; Ho, 1995). These observations fit predictions of models of viral dynamics (Nowak and May, 2000; Hadjichrysanthou et al., 2016). One of the models that integrated virological and immunological data in a system consisting of cynomolgus macaques infected with Ebola virus indicated that, to be effective, antiviral treatments with favipiravir and remdesivir (GS-5734) had to be administered at least 2 days before the peak of viremia and cytokine storm (Madelain et al., 2018). Similar predictions of the advantage of early antiviral administration were made using models of monotherapy to block SARS-CoV-2 using infection parameters from patients (Goncalves et al., 2020). The potential benefit of favipiravir to treat arenavirus infections (Mendenhall et al., 2011a) was reinforced by the observation that favipiravir was effective in guinea pigs infected with Pichinde virus, even when the drug was administered when the animals were ill (Mendenhall et al., 2011b).

The HCV-Huh-7.5 cell culture system—with the availability of HCV populations displaying different fitness (Sheldon et al., 2014; Moreno et al., 2017; Gallego et al., 2020; Delgado et al., 2021)—offered a unique opportunity to quantify the consequences for antiviral efficacy of adding DAAs and lethal mutagens once the infection has been already initiated. With this aim, we used two related HCV populations that belong to the same evolutionary lineage in Huh-7.5 cells in cell culture, but that differed in fitness. The low fitness HCV was the clonal population HCV p0, obtained by the

transcription of plasmid Jc1FLAG2(p7-nsGluc2A) (Marukian et al., 2008), cell transfection with the RNA transcripts, and virus amplification in Huh-7.5 cells (Perales et al., 2013). HCV p0 was arbitrarily given the fitness value of 1. It was subjected to 200 serial passages in Huh-7.5 cells at a multiplicity of infection (MOI) of 0.03 $TCID_{50}$/cell to yield population HCV p200 which displayed a fitness 2.3-fold higher than that of HCV p0 (Gallego et al., 2020). We report the consequences of adding DAAs [combinations of daclatasvir (dimethyl N,N'-([1,1'-biphenyl]-4,4'-diylbis{1H-imidazole-5,2-diyl-[(2S)-pyrrolidine-2,1-diyl][(2S)-3-methyl-1-oxobutane-1,2-diyl]}) dicarbamate) and sofosbuvir] or combinations of lethal mutagens, favipiravir and ribavirin, once the infection of Huh-7.5 cells by HCV p0 or its high fitness derivative HCV p200 is ongoing. Delaying inhibitor addition relative to initiation of the infection reduced significantly the antiviral efficacy, and the reduction was accentuated with the DAAs when cells were infected with the high fitness HCV population. We discuss the results in terms of replicative parameters and fitness effects on antiviral sensitivity. We suggest that preclinical protocols for the evaluation of antiviral agents would benefit from including the use of high fitness viral populations and tests of addition of the agents once the infection is well underway.

Materials and methods

Cells and viruses

Huh-7.5 reporter cells were used for all HCV infections, and Huh-7.5 cells were used for virus titration. Cells were grown in Dulbecco's modification of Eagle's medium (DMEM) at 37°C in a 5% CO_2 atmosphere, following previously described procedures (Blight et al., 2002; Jones et al., 2010; Gallego et al., 2018). Cells were passaged a maximum of 30 times, using a split ratio of 1:4 before their use in the experiments.

The initial HCV population was obtained following virus rescue by the expression of plasmid Jc1FLAG2(p7-nsGluc2A) (a chimera of J6 and JFH-1 from HCV serotype 2a) (Marukian et al., 2008) and subsequent expansions in Lunet and Huh-7.5 human hepatoma cells to obtain HCV p0, as previously described (Perales et al., 2013). Likewise, a replication-defective HCV GNN [that includes a mutation in NS5B that abolishes polymerase activity (Marukian et al., 2008)] was rescued and used as negative infection control. HCV was passaged 200 times in Huh-7.5 cells to produce HCV p200, as previously described (Sheldon et al., 2014; Moreno et al., 2017). To control the absence of contamination, mock-infected and HCV GNN-infected cells were handled in parallel, and their supernatants were titrated; no evidence of contamination was obtained in any of the experiments.

Antiviral agents and antiviral protocols

Stock solutions of daclatasvir [10 mM in dimethyl sulfoxide (DMSO); Selleck Chemicals], sofosbuvir (10 mM in DMSO; Selleck Chemicals), favipiravir (20 mM in water; Atomax Chemicals Co. Ltd.), and ribavirin (100 mM in PBS; Sigma) were prepared, sterilized by filtration, and stored at−70°C, as detailed previously (Sheldon et al., 2014; Gallego et al., 2016, 2018). Drugs were diluted in DMEM before their use to reach the desired concentrations for the experiments. In all experiments, 4 x 10^5 Huh-7.5 reporter cells were infected with either HCV p0 or HCV p200 at a multiplicity of infection (MOI) of 0.03 $TCID_{50}$/cell. After 5 h of virus adsorption to cells, the inoculum was removed, cells were washed, and fresh medium without or with the antiviral compounds was added. In different protocols, the time of addition of antiviral compounds relative to the initiation of infection and the duration of the infection varied, as indicated for each experiment. For serial viral passages, 0.5 ml of the cell culture supernatant from the previous infection was used to infect 4 x 10^5 Huh-7.5 reporter cells. The infection continued in the absence or presence of the drugs for 72 to 96 h, as indicated in the relevant experiment.

HCV titration and test of HCV extinction

For titration of infectious HCV, serial dilutions of the cell culture supernatants were applied to Huh-7.5 cells that had been seeded 16 h earlier in 96-well plates at 6,400 cells/well. Three days post-infection, the monolayers were washed with PBS, fixed with ice-cold methanol, and stained with NS5A-specific monoclonal antibody 9E10, as previously described (Lindenbach et al., 2005; Perales et al., 2013). Viral titers are expressed as 50% tissue culture infective dose ($TCID_{50}$/ml) (Reed and Muench, 1938).

When no infectivity was detected, the cell culture supernatant was analyzed by the HCV extinction test (de Avila et al., 2016). It consists in subjecting the undiluted cell culture supernatant to three blind passages in Huh-7.5 reporter cells, in the absence of any drug; when no infectivity

TABLE 1 CC_{50} for Huh-7.5 cells and IC_{50} values for inhibition of HCV progeny production by antiviral agents.

Antiviral agent	CC_{50}	IC_{50}
Daclatasvir[a]	14,900 ± 600 nM	10 ± 0.3 pM
Sofosbuvir[b]	>50 μM	20 ± 3 nM
Ribavirin[a]	108 ± 4.2 μM	6.9 ± 0.9 μM
Favipiravir[c]	865 ± 59 μM	7.4 ± 6 μM

[a]Data from Sheldon et al., 2014.
[b]Data from Gallego et al., 2016.
[c]Data from de Avila et al., 2016.

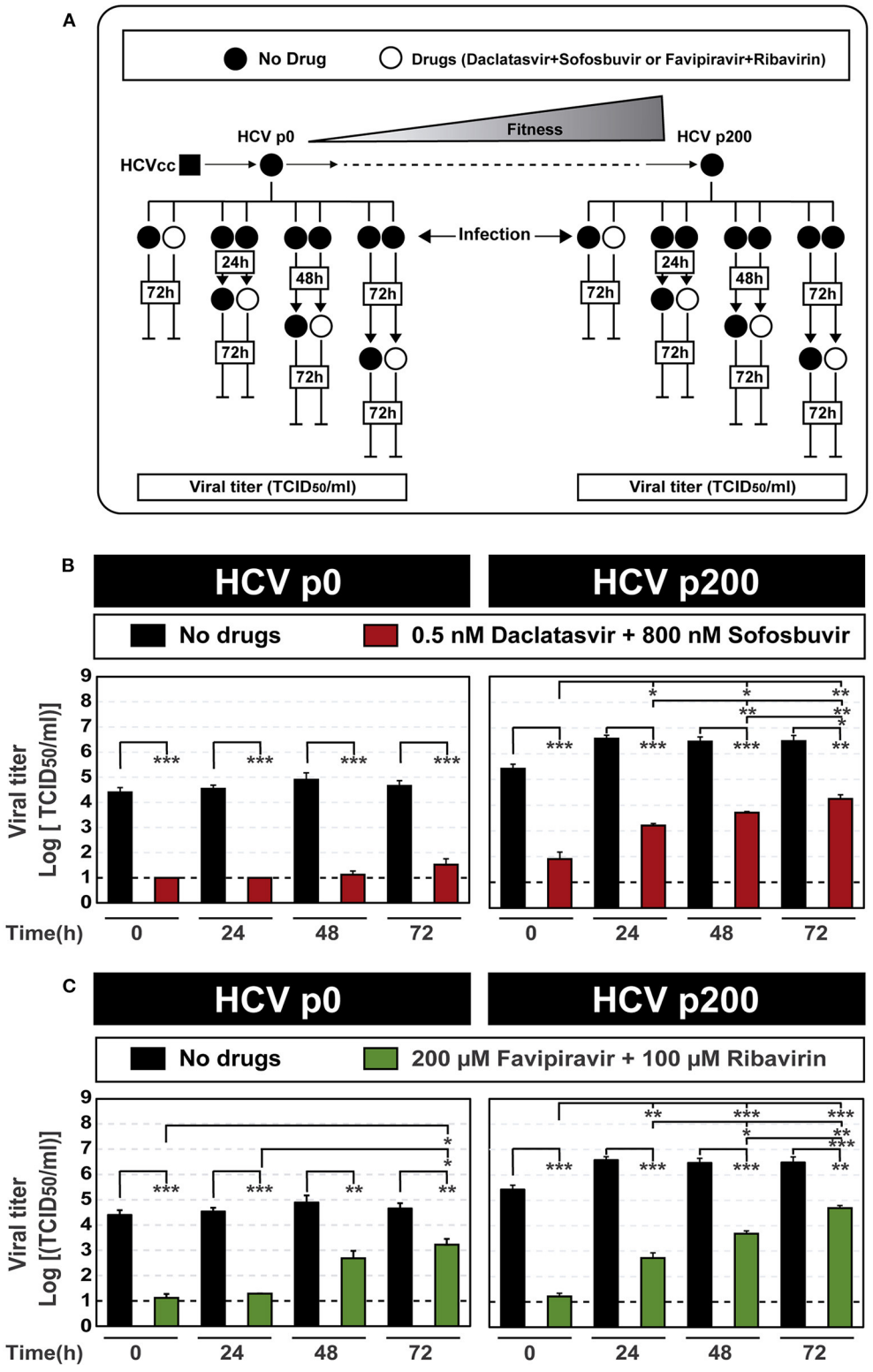

FIGURE 1
Effect of the time of addition of inhibitor combinations after initiation of HCV infection. **(A)** Scheme of the experiment, displaying fitness gain upon subjecting HCV p0 [derived from HCVcc obtained by transfection of Lunet cells with RNA transcribed from plasmid Jc1FLAG2(p7-nsGluc2A)] to 200 serial passages in Huh-7.5 reporter cells to yield HCV p200. The time of infection and of addition of the inhibitors are indicated (empty boxes inserted in vertical lines); in all cases, cell culture supernatants were titrated for HCV infectivity at 72 h

(Continued)

FIGURE 1
following the last addition of inhibitors. **(B)** Virus titer upon the addition of combinations of the DAAs daclatasvir and sofosbuvir. The virus used for infection is indicated in the upper filled boxes, and the inhibitor concentrations in the cell culture medium are given in the empty box. The abscissa shows the time post-infection of addition of the drug combination. **(C)** Same as **(B)** except that the inhibitors used were combinations of the mutagenic nucleoside analogs favipiravir and ribavirin. For **(B,C)**, the statistical significance of the differences between values given in the bars are given as follows: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; t-test). Viral titer values can be found in Supplementary Table S2. Extracellular RNA values can be found in Supplementary Table S3 and Supplementary Figure S2. Experiments were performed in triplicate (replicas A, B, and C in Supplemental Tables S2, S3). Procedures followed for drug preparation, cell infections, and titration of infectivity are detailed in Materials and Methods.

was detected in the cell culture medium of the third blind passage and no RT-PCR amplifiable material was detected in the intracellular material, the virus was considered extinct. The oligonucleotide primers used for the RT-PCR NS5A-F1 and NS5A-R1 are included in Supplementary Table S1.

Quantification of HCV RNA

Total cellular RNA was extracted from infected cells or cell culture supernatants using QIAamp Viral RNA kit (Qiagen), according to the manufacturer's instructions; for intracellular RNA, the Qiagen RNeasy kit (Qiagen) was used. HCV RNA was quantified by real-time quantitative RT-PCR (qRT-PCR) using the LightCycler RNA Master SYBR Green kit (Roche) (Perales et al., 2013). A fragment of the 5' untranslated region (UTR) was used for the quantification; the oligonucleotide primers used for the amplification, termed HCV-5UTR-F2 and HCV-5UTR-R2, are included in Supplementary Table S1. Quantification was relative to a standard curve obtained with known amounts of HCV RNA synthesized by *in vitro* transcription of plasmid GNNFLAG2(p7-nsGluc2A). To ascertain the absence of contamination with undesired templates, negative controls (consisting of samples without added template RNA or including RNA from mock-infected cells) were run in parallel. This procedure and primers for HCV RNA quantification have been previously used (Perales et al., 2013; de Avila et al., 2016).

PCR amplification and Sanger sequencing

RT-PCR amplification was carried out using AccuScript (Agilent) following the manufacturer's instructions. Amplification products were analyzed by agarose gel electrophoresis, using GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific) as molar mass standard. Amplification controls in the absence of RNA were run in parallel to ascertain the absence of contamination by undesired templates. Amplified DNA was sequenced using the 23 ABI 3730 XLS sequencer (Macrogen, Inc.). All the oligonucleotide primers used in this study are listed in Supplementary Table S1.

Statistics

The statistical significance of differences in infectivity and RNA levels was determined using the t-test and software GraphPad Prism 8.00. The differences between viral titers and RNA levels along the different passages were determined using ANCOVA test and software GraphPad Prism 8.00.

Results

Kinetics of hepatitis C progeny production and time-dependent decrease of antiviral efficacy

We previously determined CC_{50} for Huh-7.5 cells and IC_{50} values for inhibition of HCV progeny production for the DAAs, daclatasvir and sofosbuvir, and for the mutagenic nucleoside analogs, favipiravir and ribavirin (Table 1); these are the four inhibitors used in this study. Upon infection of Huh-7.5 cells, the progeny production of both HCV p0 and HCV p200 increased exponentially up to 72 h post-infection (Moreno et al., 2017); this result was confirmed in this study, with the additional note that the extracellular infectivity was maintained up to 144 h post-infection (Supplementary Figure S1).

Based on the kinetics of progeny production, and the CC_{50} and IC_{50} values for the inhibitors used in this study, we evaluated the influence of the time elapsed between infection initiation and inhibitor addition on the efficacy of inhibitor combinations. With this aim, DAAs and lethal mutagen combinations were added at 0 h, 24 h, 48 h, and 72 h after infection, and viral titer was determined at 72 h after the addition of the inhibitors. The results (Figure 1) documented a decrease of antiviral efficacy that was more intense the longer was the infection time prior to inhibitor addition. The effect was more accentuated in the infections with HCV p200 than with HCV p0, particularly for the DAA combination (compare Figures 1B,C; the numerical values are given in Supplementary Table S2). The inhibitions were confirmed with quantifications of extracellular viral RNA (Supplementary Table S3; Supplementary Figure S2).

The variation of efficacy with time of inhibitor addition following the initiation of infection approximated an exponential function of the form $y = y_0 - Ae^{(x/t)}$, implying that the decrease of efficacy is more accentuated the longer

10.3389/fmicb.2022.960676

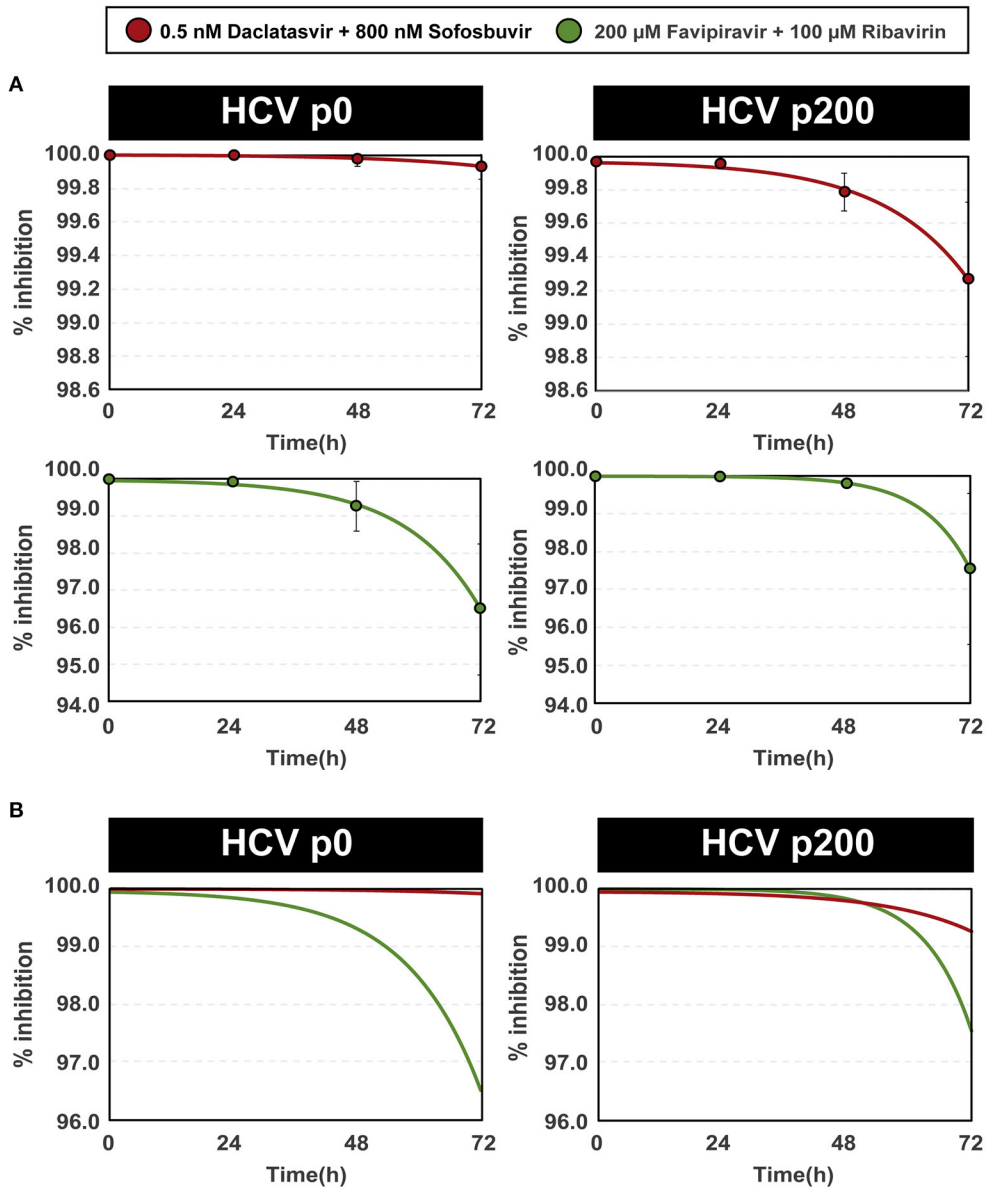

FIGURE 2
Percentage of inhibition of HCV progeny production as a function of the time of inhibitor addition after initiation of infection. **(A)** The percentages of inhibition have been calculated with the data given in Figure 1 and Supplemental Table S2. The inhibitors (color coded) and HCV used are indicated in the upper boxes. The abscissae give the time elapsed between the initiation of the infection and the addition of inhibitors. The experimental points are given in the four panels, and the lines indicate the function that best fits the data. Note that the scale in ordinate differs among panels of the same column. The functions are the following: DAAs with HCV p0 (upper left panel): $y=100.00-(5.68\cdot10^{-4}e^{(x/0.84)})$; DAAs with HCV p200 (upper right panel): $y=99.97-(2.21\cdot10^{-3}e^{(x/0.69)})$; mutagens with HCV p0 (bottom left panel): $y=99.97-(4.69\cdot10^{-3}e^{(x/0.60)})$; mutagens with HCV p200 (bottom right panel): $y=100.00-(6.33\cdot10^{-5}e^{(x/0.38)})$. **(B)** For comparative purposes, the four adjusted functions [color coded as in (A)] are depicted with the same scale in ordinate. Procedures are detailed in Materials and Methods.

the infection is allowed to progress before adding the antiviral combinations. The exponential function exhibited a good fit with the experimental points ($R^2 \geq 0.98$) (Figure 2A). The A and t terms of the equations reflected a higher inhibition, despite the delayed addition of inhibitors, for the DAAs than for the mutagenic analog combinations (Figure 2B). The delayed administration of DAAs decreased their efficacy more for the high fitness HCV p200 than for HCV p0. No such difference was noted with the mutagenic analog combinations (Figure 2B).

Thus, the addition of DAAs or mutagenic nucleoside analogs to actively replicating HCV infections decreases significantly their antiviral efficacy.

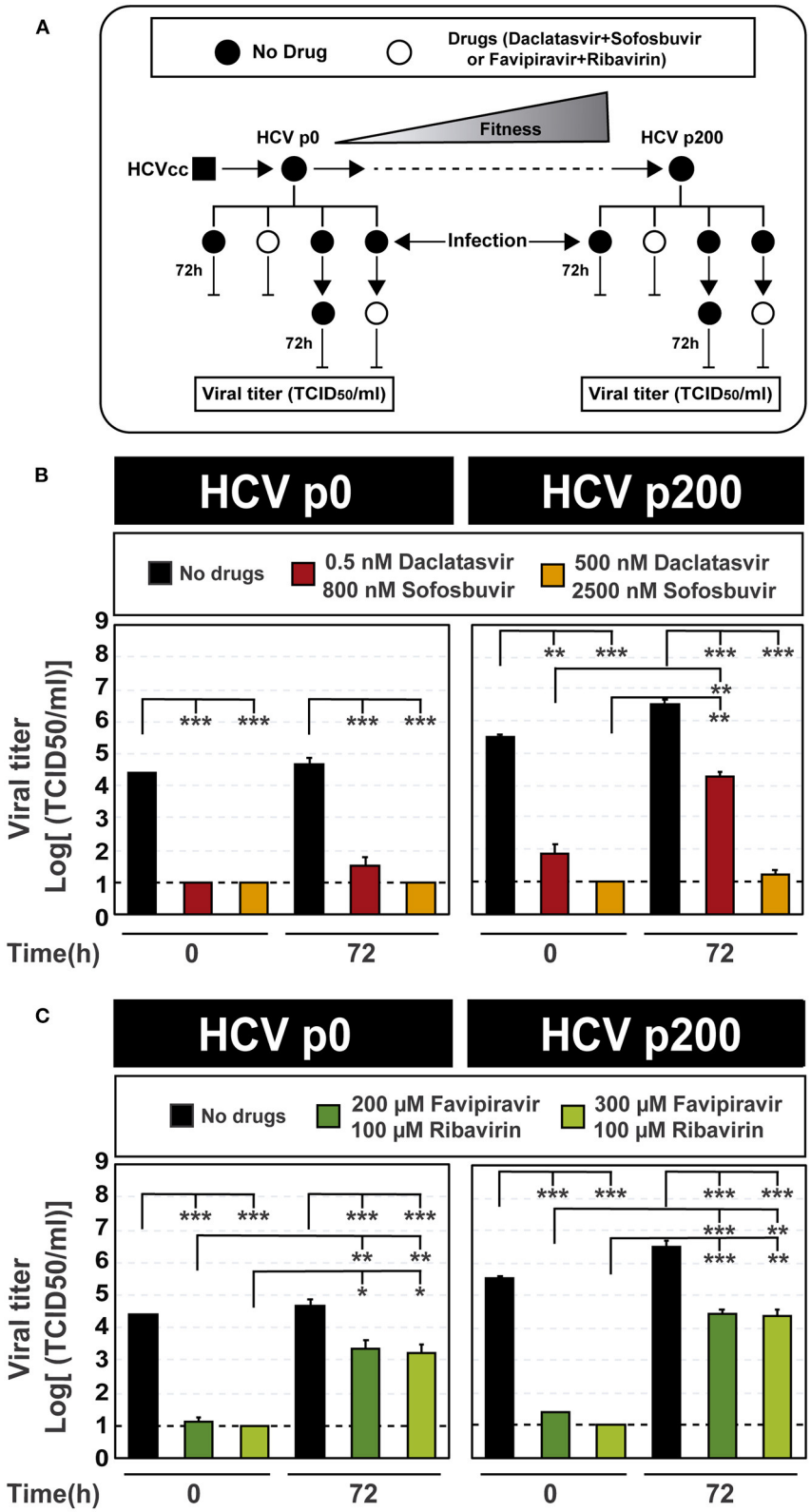

FIGURE 3
Inhibition of HCV infectious progeny production by combinations of antiviral agents added at the time of infection or 72 h post-infection. **(A)** Scheme of the experiment displaying at the top the evolution of HCV p0 toward its high fitness derivative HCV p200. The time of infection and of addition of the antiviral inhibitors are indicated next to the vertical lines; in all cases, virus titer in the cell culture supernatant was determined

(Continued)

FIGURE 3
72 h after addition of the inhibitors. **(B)** Virus titer upon the addition of combinations of the DAAs daclatasvir and sofosbuvir (concentrations in the cell culture medium given in the boxes) either at the time of infection or 72 h post-infection (indicated in abscissa); infections were carried out in parallel with HCV p0 and HCV p200. **(C)** Same as **(B)** except that the inhibitors used were combinations of the mutagenic nucleoside analogs favipiravir and ribavirin. For B and C, the statistical significance of the differences between values given in the bars are given as follows: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; *t*-test. Experiments were performed in triplicate (replicas A, B, and C in Supplemental Tables S4, S5). Procedures followed for drug preparation, cell infections, and titration of infectivity are detailed in Materials and Methods.

Extinction of hepatitis C progeny production at 72 h post-initiation of the infection

To evaluate conditions that could extinguish high fitness HCV p200, we performed a new experiment by fixing at 72 h the time elapsed between infection initiation and inhibitor addition, and using higher inhibitor concentrations. The maximum concentrations of daclatasvir, sofosbuvir, and favipiravir tested were based on our former results with these inhibitors and HCV (Sheldon et al., 2014; Gallego et al., 2016, 2018), and they were three- to 30-fold lower than their CC_{50} values. Ribavirin was kept at 100 μM because in previous studies this concentration was discriminatory of fitness effects: it produced extinction of HCV p0 after four serial passages, while it failed to extinguish HCV p200 even after ten serial passages (Gallego et al., 2018). Under these treatment conditions (Figure 3A), again, there was a significantly lower inhibition of HCV progeny production when either DAA or lethal mutagen combinations were added at 72 h post-infection, as compared with their addition at the time of infection onset (Figures 3B,C; the numerical values are given in Supplementary Table S4). The decrease of efficacy of the DAAs was more accentuated with HCV p200 than HCV p0, in agreement with the HCV fitness-associated resistance to the DAA inhibitors when used individually (Sheldon et al., 2014; Gallego et al., 2016, 2018). The fitness effect was not noted with the combination of mutagens.

Since no virus extinction was attained under any of the infection conditions tested in our experiments, we subjected the initial viral populations HCV p0 and HCV p200, as well as the populations obtained at 72h post-infection in the absence and presence of drugs (HCV populations described in Figure 3), to serial infections in the absence and presence of the same drug concentrations. The results indicate that, in all cases, titers below the limit of detection were achieved by passage 3 or earlier, except when the input virus was HCV p200 subjected to 0.5 nM daclatasvir and 800 nM sofosbuvir (Figure 4; the numerical values are given in Supplementary Table S4). For this population, the virus titer remained at around 10^3-10^5 $TCID_{50}$/ml for at least five passages (Figure 4A). The inhibitions were confirmed with quantifications of extracellular viral RNA (Supplementary Table S5; Supplementary Figure S3).

To investigate the possible presence of daclatasvir- or sofosbuvir-escape substitutions, the initial HCV p200 population and the populations at passage 5 in the absence and presence of drugs were subjected to Sanger sequencing. The results show several amino acid substitutions whose frequency increased with passage number in the presence of the daclatasvir plus sofosbuvir combination, independently of adding the drugs at 0 or 72 h.p.i. (Supplementary Table S6). Of these substitutions, T24A, F28I, and L31M may account for the lack of effectiveness of the lower DAA concentrations because they have been described as resistance-associated substitutions (RASs) for NS5A inhibitors (Lontok et al., 2015; Sarrazin, 2016; European Association for the Study of the Liver, 2020). The escape effect of these substitutions may be reinforced by the high fitness of HCV p200 (Sheldon et al., 2014; Gallego et al., 2016, 2018; Sabariegos et al., 2022). Some additional substitutions found of HCV p200 passaged in the absence of drugs have been previously described. For example, T245A resulted in a perinuclear-restricted localization of NS5A (Goonawardane et al., 2018), and C465S resulted in an increase of viral titer (Han et al., 2013). No substitutions that coincide with HRSs defined in patients (Soria et al., 2020) have been found in the HCV p200 populations. The same HCV p200 population was effectively extinguished by the higher DAA doses, and all populations were extinguished by the mutagenic analog concentration; no evidence of infectivity or viral RNA was obtained when the populations that yielded a titer below detection level were subjected to the extinction test (Figure 4). Thus, the time post-infection of inhibitor addition is critical to determine inhibitor and lethal mutagenic effectiveness, but continuing treatment may achieve extinction in a modest number of passages.

Discussion

Models of viral population dynamics predict loss of antiviral efficacy when the inhibitors are added when a viral infection is actively progressing (Nowak and May, 2000; Hadjichrysanthou et al., 2016; Madelain et al., 2018; Goncalves et al., 2020). In this study with HCV in cell culture, we have documented significant reductions of antiviral efficacy correlated with the time of inhibitor addition relative to infection onset. The study has been performed with the well-characterized daclatasvir and sofosbuvir and the lethal mutagens, favipiravir and ribavirin (Perales et al., 2019; Bhattacharjee et al., 2021). The activity of

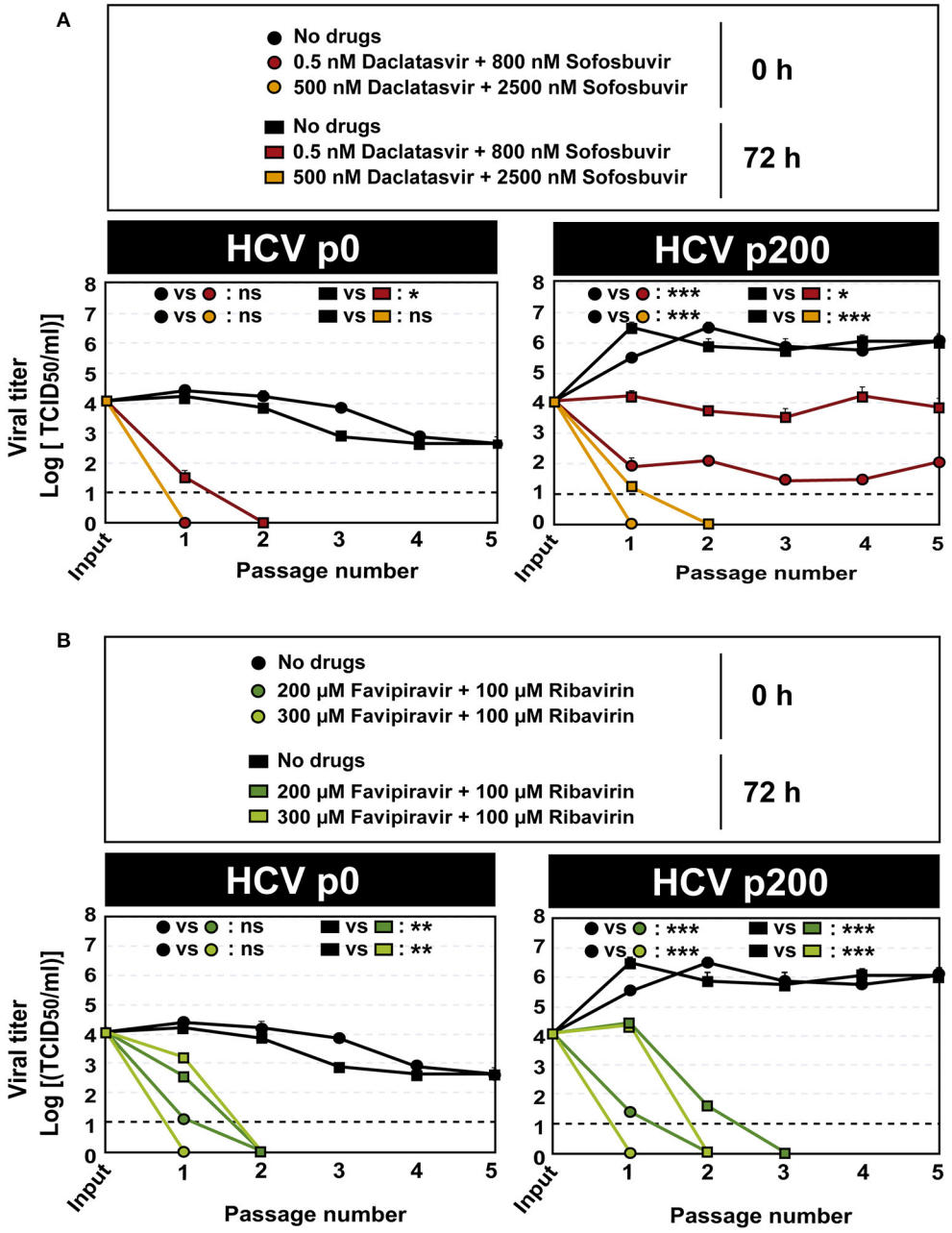

FIGURE 4
Extinction of HCV populations by serial passages in the presence of antiviral agent combinations. **(A)** The responses of HCV p0 and HCV p200 subjected to five serial infections in the absence or presence of the daclatasvir and sofosbuvir are indicated in the box. **(B)** The responses of HCV p0 and HCV p200 subjected to five serial infections in the absence or presence of the favipiravir and ribavirin are indicated in the box. The initial and the viruses at passage 1 are those described in Figure 3. The time of infection and of addition of the antiviral inhibitors are indicated; in all cases, virus titer in the cell culture supernatant was determined 72 h after addition of the inhibitors. The differences between the viral titers along the passages as given at the top of each graph are as follows: ns: no significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; ANCOVA test. Viral titer values can be found in Supplemental Table S4. Extracellular RNA values can be found in Supplemental Table S5 and Supplementary Figure S3. Experiments were performed in triplicate (replicas A, B, and C in Supplemental Tables S4, S5). Procedures followed for drug preparation, cell infections, and titration of infectivity are detailed in Materials and Methods.

these agents was previously investigated in our laboratory with the clonal population HCV p0 and its high fitness derivative HCV p200 (Ortega-Prieto et al., 2013; Sheldon et al., 2014; de Avila et al., 2016; Gallego et al., 2016, 2018, 2019). This study has confirmed an HCV fitness-mediated antiviral resistance in protocols in which inhibitors were added to an ongoing

HCV infection. Due to experimental limitations derived from the number of variables covered, only a reduced range of inhibitor concentrations has been explored, based on CC_{50} and IC_{50} values, as well as on previous analyses with the same inhibitors used individually on inhibition and extinction of HCV p0 (Ortega-Prieto et al., 2013; Sheldon et al., 2014; de Avila et al., 2016; Gallego et al., 2016, 2018). With the inhibitors used, and under the conditions tested, higher viral suppression was achieved with the combination of the DAAs, daclatasvir and sofosbuvir, than with the combination of the mutagens, favipiravir and ribavirin. The relevance of the higher suppression of viral replication attained by these DAAs relative to the nucleoside analog combinations is reinforced by the fact that the concentrations used were on average 12-fold lower than their IC_{50} values for DAAs than for the mutagen combinations (compare Table 1; Figures 1, 2). The higher HCV suppressive effect of the DAAs observed cannot be given the rank of general conclusion on relative efficacy of standard (non-mutagenic) vs. mutagenic inhibitors for HCV. Clarification of this point for HCV will require a comparison of additional DAAs and mutagenic base and nucleoside analogs over a broader range of viral population size and viral fitness.

The exponential functions that relate the percentage of inhibition with the time elapsed between initiation of HCV infection and the addition of inhibitors to the infected cells (Figure 2) suggest that there is a margin of early times after infection in which inhibitor efficacy is minimally affected. Following this initial period, delay in administration has a progressively increasing negative effect on avoiding HCV progeny production. The progressing time-dependent trend of antiviral efficacy loss is in agreement with the competition model proposed to explain the fitness effect of antiviral efficacy (Domingo et al., 2019). According to the competition model, the amount of active forms of the inhibitors per replication complex equivalent may be lower for high fitness HCV. A similar inhibitor limitation effect may occur with an advanced infection as compared with the initial phases of the same infection.

Despite environmental heterogeneity in an *in vivo* setting, the results of our simplified and controlled experimental system are in agreement with observations reported with inhibitor efficacy with several viruses tested in animal models (Mendenhall et al., 2011b; Madelain et al., 2018; Abdelnabi et al., 2021) and with general theoretical predictions (Nowak and May, 2000; Hadjichrysanthou et al., 2016).

The consequences of delayed treatment implementation bear indirectly on the issue of at which point in the development of HCV-related hepatic disease (i.e. liver fibrosis stage) should an antiviral treatment be initiated (Chahal et al., 2016; El Sayed et al., 2017; Ruggeri et al., 2018). Long-term HCV infection does not necessarily mean a steady increase of viral load due to the many host and viral factors involved, as well as viral turnover and clearance (Dahari et al., 2005). However, it is expected that prolonged replication in the same liver may favor fitness increase of the resident HCV. This emphasizes the success of current DAA-based treatments of chronic HCV infection, which achieve sustained virological responses of around 95% (Janjua et al., 2021). The process required for the approval of new antiviral agents (even if they are a modified form of a previously licensed agent or one investigated from drug repositioning) is generally lengthy and costly. Therefore, we suggest that the preclinical screening of antiviral agents in cell culture would benefit from using high fitness viral populations and administration of the agents once the infection is underway. High fitness viral populations can presently be derived in a straightforward manner provided the host cell or organism—in which the antiviral tests will be conducted—is available to support serial virus passages (Domingo and Holland, 1997). A multiple passaged virus acquires significant fitness increases relative to the parental population when fitness is measured in the same environment (Domingo et al., 2022). The influence of high HCV fitness on the decrease of inhibitory activity was noted with both combinations, as expected from previous results on the higher effectiveness of extinction of low fitness HCV than high fitness HCV by the individual antiviral agents (Sheldon et al., 2014; Gallego et al., 2016, 2018). With the combinations used in this study, the efficacy decrease associated with high HCV fitness was clearly noted with the DAAs and far less with the lethal mutagens (Figures 1, 3). In the plots that represented the decrease of inhibitory activity as a function of the time post-infection of inhibitors administration, an attenuating effect of HCV fitness on the loss of inhibition was noted with the DAA combination and not with the mutagen combination (Figure 2). Additional data with other inhibitors and lethal mutagens are needed to suggest if this can be an advantage of lethal mutagens vs. non-mutagenic inhibitors for HCV and other RNA viruses. Protocols using high fitness virus and delayed administration of inhibitors (either individually or in combination) provide a better filter to avoid expensive biological assays prior to clinical tests with humans.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

This study used human cell lines obtained from Dr. Charles Rice's laboratory. The institutional Bioethics Committee from Consejo Superior de Investigaciones Científicas (CSIC), in accord with Spanish regulations did not require the study to be reviewed or approved, because only established cell lines and no human samples were involved.

Author contributions

ED and CP conceived, designed, and supervised the project. CG-C, LV-S, PS, MS, IG, AÁ, and AD-P performed the experiments and analyzed the results. BM-G performed the statistical analyses. ED, CP, and CG-C wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Instituto de Salud Carlos III, Spanish Ministry of Science and Innovation (COVID-19 Research Call COV20/00181) and cofinanced by European Development Regional Fund A way to achieve Europe and grants CSIC-COV19-014 from Consejo Superior de Investigaciones Científicas (CSIC), project 525/C/2021 from Fundació La Marató de TV3, PID2020-113888RB-I00 from Ministerio de Ciencia e Innovación, BFU2017-91384-EXP from Ministerio de Ciencia, Innovación y Universidades (MCIU), PI18/00210 and PI21/00139 from Instituto de Salud Carlos III, and S2018/BAA-4370 (PLATESA2 from Comunidad de Madrid/FEDER). This research work was also funded by the European Commission NextGenerationEU (Regulation EU 2020/2094), through the CSIC's Global Health Platform (PTI Salud Global). CP was supported by the Miguel Servet Program of the Instituto de Salud Carlos III (CP14/00121 and CPII19/00001), cofinanced by the European Regional Development Fund (ERDF). Centro de Investigación en Red de Enfermedades Hepáticas y Digestivas (CIBERehd) is funded by Instituto de Salud Carlos III. Institutional grants from the Fundación Ramón Areces and Banco Santander to the CBMSO are also acknowledged. The team at CBMSO belongs to the Global Virus Network (GVN). CG-C was supported by predoctoral contract PRE2018-083422 from MCIU. PS was supported by postdoctoral contract Margarita Salas CA1/RSUE/2021 from MCIU. BM-G was supported by predoctoral contract PFIS FI19/00119 from ISCIII, cofinanced by Fondo Social Europeo (FSE).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.960676/full#supplementary-material

References

Abdelnabi, R., Foo, C. S., Kaptein, S. J. F., Zhang, X., Do, T. N. D., Langendries, L., et al. (2021). The combined treatment of Molnupiravir and Favipiravir results in a potentiation of antiviral efficacy in a SARS-CoV-2 hamster infection model. *EBioMedicine* 72, 103595. doi: 10.1016/j.ebiom.2021.103595

Beaucourt, S., and Vignuzzi, M. (2014). Ribavirin: a drug active against many viruses with multiple effects on virus replication and propagation. Molecular basis of ribavirin resistance. *Curr. Opin. Virol.* 8, 10–15. doi: 10.1016/j.coviro.2014.04.011

Bellocchi, M. C., Aragri, M., Carioti, L., Fabeni, L., Pipitone, R. M., Brancaccio, G., et al. (2019). NS5A gene analysis by next generation sequencing in HCV nosocomial transmission clusters of HCV genotype 1b infected patients. *Cells* 8:666. doi: 10.3390/cells8070666

Bhattacharjee, C., Singh, M., Das, D., Chaudhuri, S., and Mukhopadhyay, A. (2021). Current therapeutics against HCV. *VirusDisease*, 32, 228–243. doi: 10.1007/s13337-021-00697-0

Blight, K. J., McKeating, J. A., and Rice, C. M. (2002). Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J. Virol.* 76, 13001–13014. doi: 10.1128/JVI.76.24.13001-13014.2002

Ceccherini-Silberstein, F., Cento, V., Di Maio, V. C., Perno, C. F., and Craxi, A. (2018). Viral resistance in HCV infection. *Curr Opin Virol* 32, 115–127. doi: 10.1016/j.coviro.2018.10.005

Chahal, H. S., Marseille, E. A., Tice, J. A., Pearson, S. D., Ollendorf, D. A., Fox, R. K., et al. (2016). Cost-effectiveness of early treatment of Hepatitis C virus genotype 1 by stage of liver fibrosis in a US treatment-naive population. *JAMA Intern. Med.* 176, 65–73. doi: 10.1001/jamainternmed.2015.6011

Chen, Q., Perales, C., Soria, M. E., Garcia-Cehic, D., Gregori, J., Rodriguez-Frias, F., et al. (2020). Deep-sequencing reveals broad subtype-specific HCV resistance mutations associated with treatment failure. *Antiviral Res.* 174, 104694. doi: 10.1016/j.antiviral.2019.104694

Costantino, A., Spada, E., Equestre, M., Bruni, R., Tritarelli, E., Coppola, N., et al. (2015). Naturally occurring mutations associated with resistance to HCV NS5B polymerase and NS3 protease inhibitors in treatment-naive patients with chronic hepatitis C. *Virol. J.* 12, 186. doi: 10.1186/s12985-015-0414-1

Dahari, H., Major, M., Zhang, X., Mihalik, K., Rice, C. M., Perelson, A. S., et al. (2005). Mathematical modeling of primary hepatitis C infection: noncytolytic clearance and early blockage of virion production. *Gastroenterology* 128, 1056–1066. doi: 10.1053/j.gastro.2005.01.049

de Avila, A. I., Gallego, I., Soria, M. E., Gregori, J., Quer, J., Esteban, J. I., et al. (2016). Lethal mutagenesis of Hepatitis C virus induced by favipiravir. *PLoS ONE* 11, e0164691. doi: 10.1371/journal.pone.0164691

De Clercq, E., and Li, G. (2016). Approved antiviral drugs over the past 50 years. *Clin Microbiol Rev* 29, 695–747. doi: 10.1128/CMR.00102-15

Delgado, S., Perales, C., Garcia-Crespo, C., Soria, M. E., Gallego, I., de Avila, A. I., et al. (2021). A Two-Level, Intramutant spectrum haplotype profile of Hepatitis C virus revealed by self-organized maps. *Microbiol. Spectr.* 9, e0145921. doi: 10.1128/Spectrum.01459-21

Di Maio, V. C., Barbaliscia, S., Teti, E., Fiorentino, G., Milana, M., Paolucci, S., et al. (2021). Resistance analysis and treatment outcomes in hepatitis C virus genotype 3-infected patients within the Italian network VIRONET-C. *Liver Int.* 41, 1802–1814. doi: 10.1111/liv.14797

Di Maio, V. C., Cento, V., Lenci, I., Aragri, M., Rossi, P., Barbaliscia, S., et al. (2017). Multiclass HCV resistance to direct-acting antiviral failure in real-life patients advocates for tailored second-line therapies. *Liver Int.* 37, 514–528. doi: 10.1111/liv.13327

Dietz, J., Susser, S., Vermehren, J., Peiffer, K. H., Grammatikos, G., Berger, A., et al. (2018). Patterns of resistance-associated substitutions in patients with chronic HCV infection following treatment with direct-acting antivirals. *Gastroenterology 154(4)* 976-988 e974. doi: 10.1053/j.gastro.2017.11.007

Domingo, E. (2020). *Virus as Populations.* 2nd ed. Amsterdam: Academic Press; Elsevier.

Domingo, E., de Avila, A. I., Gallego, I., Sheldon, J., and Perales, C. (2019). Viral fitness: history and relevance for viral pathogenesis and antiviral interventions. *Pathog. Dis.* 77, ftz021. doi: 10.1093/femspd/ftz021

Domingo, E., and Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51, 151–178. doi: 10.1146/annurev.micro.51.1.151

Domingo, E., and Perales, C. (2012). From quasispecies theory to viral quasispecies: how complexity has permeated virology. *Math. Model. Nat. Phenom.* 7, 32–49. doi: 10.1051/mmnp/20127508

Domingo, E., García-Crespo, C., Soria, M. E., and Perales, C. (2022). Viral fitness, population complexity, host interactions, and resistance to antiviral agents. *Curr. Top Microbiol. Immunol*. (in press).

Ehrlich, P. (1913). *Address in Pathology on Chemiotherapy*. Report from the International Medical Congress, London. p. 353–359.

El Sayed, A., Barbati, Z. R., Turner, S. S., Foster, A. L., Morey, T., Dieterich, D. T., et al. (2017). Sofosbuvir in the treatment of early HCV infection in HIV-infected men. *HIV Clin. Trials* 18, 60–66. doi: 10.1080/15284336.2017.1280594

Esposito, I., Marciano, S., Haddad, L., Galdame, O., Franco, A., Gadano, A., et al. (2018). Prevalence and factors related to natural resistance-associated substitutions to direct-acting antivirals in patients with Genotype 1 Hepatitis C virus infection. *Viruses* 11, 3. doi: 10.3390/v11010003

European Association for the Study of the Liver. (2020). EASL recommendations on treatment of hepatitis C: final update of the series. *J. Hepatol.* 73, 1170–1218. doi: 10.1016/j.jhep.2020.08.018

Furuta, Y., Takahashi, K., Shiraki, K., Sakamoto, K., Smee, D. F., Barnard, D. L., et al. (2009). T-705 (favipiravir) and related compounds: novel broad-spectrum inhibitors of RNA viral infections. *Antiviral Res.* 82, 95–102. doi: 10.1016/j.antiviral.2009.02.198

Gallego, I., Gregori, J., Soria, M. E., Garcia-Crespo, C., Garcia-Alvarez, M., Gomez-Gonzalez, A., et al. (2018). Resistance of high fitness hepatitis C virus to lethal mutagenesis. *Virology* 523, 100–109. doi: 10.1016/j.virol.2018.07.030

Gallego, I., Sheldon, J., Moreno, E., Gregori, J., Quer, J., Esteban, J. I., et al. (2016). Barrier-independent, fitness-associated differences in sofosbuvir efficacy against Hepatitis C Virus. *Antimicrob Agents Chemother* 60, 3786–3793. doi: 10.1128/AAC.00581-16

Gallego, I., Soria, M. E., Garcia-Crespo, C., Chen, Q., Martinez-Barragan, P., Khalfaoui, S., et al. (2020). Broad and dynamic diversification of infectious Hepatitis C virus in a cell culture environment. *J. Virol.* 94, e01856–e0119. doi: 10.1128/JVI.01856-19

Gallego, I., Soria, M. E., Gregori, J., de Avila, A. I., Garcia-Crespo, C., Moreno, E., et al. (2019). Synergistic lethal mutagenesis of hepatitis C virus. *Antimicrob Agents Chemother.* 63, e01653–e01619. doi: 10.1128/AAC.01653-19

Goncalves, A., Bertrand, J., Ke, R., Comets, E., de Lamballerie, X., Malvy, D., et al. (2020). Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load. *CPT Pharmacometrics Syst. Pharmacol.* 9, 509–514. doi: 10.1002/psp4.12543

Goonawardane, N., Ross-Thriepland, D., and Harris, M. (2018). Regulation of hepatitis C virus replication via threonine phosphorylation of the NS5A protein. *J. Gen. Virol.* 99, 62–72. doi: 10.1099/jgv.0.000975

Hadjichrysanthou, C., Cauet, E., Lawrence, E., Vegvari, C., de Wolf, F., and Anderson, R. M. (2016). Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. *J R Soc. Interface* 13(119). doi: 10.1098/rsif.2016.0289

Han, Q., Manna, D., Belton, K., Cole, R., and Konan, K. V. (2013). Modulation of hepatitis C virus genome encapsidation by nonstructural protein 4B. *J. Virol.* 87, 7409–7422. doi: 10.1128/JVI.03523-12

Ho, D. D. (1995). Time to hit HIV, early and hard. *N. Engl. J. Med.* 333, 450–451. doi: 10.1056/NEJM199508173330710

Janjua, N. Z., Wong, S., Abdia, Y., Jeong, D., Buller-Taylor, T., Adu, P. A., et al. (2021). Impact of direct-acting antivirals for HCV on mortality in a large population-based cohort study. *J. Hepatol.* 75, 1049–1057. doi: 10.1016/j.jhep.2021.05.028

Jones, C. T., Catanese, M. T., Law, L. M., Khetani, S. R., Syder, A. J., Ploss, A., et al. (2010). Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system. *Nat. Biotechnol.* 28, 167–171. doi: 10.1038/nbt.1604

Kai, Y., Hikita, H., Morishita, N., Murai, K., Nakabori, T., Iio, S., et al. (2017). Baseline quasispecies selection and novel mutations contribute to emerging resistance-associated substitutions in hepatitis C virus after direct-acting antiviral treatment. *Sci. Rep.* 7, 41660. doi: 10.1038/srep41660

Kim, D. W., Lee, S. A., Kim, H., Won, Y. S., and Kim, B. J. (2014). Naturally occurring mutations in the nonstructural region 5B of hepatitis C virus (HCV) from treatment-naive Korean patients chronically infected with HCV genotype 1b. *PLoS One* 9, e87773. doi: 10.1371/journal.pone.0087773

Li, D. K., and Chung, R. T. (2019). Overview of direct-acting antiviral drugs and drug resistance of Hepatitis C virus. *Methods Mol. Biol.* 1911, 3–32. doi: 10.1007/978-1-4939-8976-8_1

Li, Z., Chen, Z. W., Li, H., Ren, H., and Hu, P. (2017a). Prevalence of hepatitis C virus-resistant association substitutions to direct-acting antiviral agents in treatment-naive hepatitis C genotype 1b-infected patients in western China. *Infect. Drug Resist.* 10, 377–392. doi: 10.2147/IDR.S146595

Li, Z., Liu, Y., Zhang, Y., Shao, X., Luo, Q., Guo, X., et al. (2017b). Naturally occurring resistance-associated variants to Hepatitis C virus direct-acting antiviral agents in treatment-naive HCV Genotype 6a-infected patients. *Biomed. Res. Int.* 2017, 9849823. doi: 10.1155/2017/9849823

Lindenbach, B. D., Evans, M. J., Syder, A. J., Wolk, B., Tellinghuisen, T. L., Liu, C. C., et al. (2005). Complete replication of hepatitis C virus in cell culture. *Science* 309, 623–626. doi: 10.1126/science.1114016

Lombardi, A., Mondelli, M. U., and Hepatitis, E. S. G.,f.V. (2019). Hepatitis C: is eradication possible? *Liver Int.* 39, 416–426. doi: 10.1111/liv.14011

Lontok, E., Harrington, P., Howe, A., Kieffer, T., Lennerstrand, J., Lenz, O., et al. (2015). Hepatitis C virus drug resistance-associated substitutions: state of the art summary. *Hepatology* 62, 1623–1632. doi: 10.1002/hep.27934

Madelain, V., Baize, S., Jacquot, F., Reynard, S., Fizet, A., Barron, S., et al. (2018). Ebola viral dynamics in nonhuman primates provides insights into virus immuno-pathogenesis and antiviral strategies. *Nat. Commun.* 9, 4013. doi: 10.1038/s41467-018-06215-z

Malandris, K., Kalopitas, G., Theocharidou, E., and Germanidis, G. (2021). The role of RASs /RVs in the current management of HCV. *Viruses* 13(10). doi: 10.3390/v13102096

Marukian, S., Jones, C. T., Andrus, L., Evans, M. J., Ritola, K. D., Charles, E. D., et al. (2008). Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* 48, 1843–1850. doi: 10.1002/hep.22550

Mawatari, S., Oda, K., Tabu, K., Ijuin, S., Kumagai, K., Fujisaki, K., et al. (2018). The co-existence of NS5A and NS5B resistance-associated substitutions is associated with virologic failure in Hepatitis C Virus genotype 1 patients treated with sofosbuvir and ledipasvir. *PLoS One* 13, e0198642. doi: 10.1371/journal.pone.0198642

Mendenhall, M., Russell, A., Juelich, T., Messina, E. L., Smee, D. F., Freiberg, A. N., et al. (2011a). T-705 (favipiravir) inhibition of arenavirus replication in cell culture. *Antimicrob. Agents Chemother.* 55, 782–787. doi: 10.1128/AAC.01219-10

Mendenhall, M., Russell, A., Smee, D. F., Hall, J. O., Skirpstunas, R., Furuta, Y., et al. (2011b). Effective oral favipiravir (T-705) therapy initiated after the onset of clinical disease in a model of arenavirus hemorrhagic fever. *PLoS Negl. Trop. Dis.* 5, e1342. doi: 10.1371/journal.pntd.0001342

Moreno, E., Gallego, I., Gregori, J., Lucia-Sanz, A., Soria, M. E., Castro, V., et al. (2017). Internal disequilibria and phenotypic diversification during replication of Hepatitis C virus in a noncoevolving cellular environment. *J. Virol.* 91, e02505-16. doi: 10.1128/JVI.02505-16

Morishita, N., Sakamori, R., Yamada, T., Kai, Y., Tahata, Y., Urabe, A., et al. (2020). Pre-existing minor variants with NS5A L31M/V-Y93H double substitution are closely linked to virologic failure with asunaprevir plus daclatasvir treatment for genotype 1b hepatitis C virus infection. *PLoS One* 15, e0234811. doi: 10.1371/journal.pone.0234811

Nakamoto, S., Kanda, T., Wu, S., Shirasawa, H., and Yokosuka, O. (2014). Hepatitis C virus NS5A inhibitors and drug resistance mutations. *World J. Gastroenterol.* 20, 2902–2912. doi: 10.3748/wjg.v20.i11.2902

Nowak, M. A., and May, R. M. (2000). *Virus Dynamics. Mathematical Principles of Immunology and Virology*. New York, NY: Oxford University Press Inc.

Ortega-Prieto, A. M., Sheldon, J., Grande-Perez, A., Tejero, H., Gregori, J., Quer, J., et al. (2013). Extinction of hepatitis C virus by ribavirin in hepatoma cells involves lethal mutagenesis. *PLoS One* 8, e71039. doi: 10.1371/journal.pone.0071039

Perales, C. (2018). Quasispecies dynamics and clinical significance of hepatitis C virus (HCV) antiviral resistance. *Int. J. Antimicrob Agents* 56, 105562. doi: 10.1016/j.ijantimicag.2018.10.005

Perales, C., Beach, N. M., Gallego, I., Soria, M. E., Quer, J., Esteban, J. I., et al. (2013). Response of hepatitis C virus to long-term passage in the presence of alpha interferon: multiple mutations and a common phenotype. *J. Virol.* 87, 7593–7607. doi: 10.1128/JVI.02824-12

Perales, C., Chen, Q., Soria, M. E., Gregori, J., Garcia-Cehic, D., Nieto-Aponte, L., et al. (2018). Baseline hepatitis C virus resistance-associated substitutions present at frequencies lower than 15% may be clinically significant. *Infect. Drug Resist.* 11, 2207–2210. doi: 10.2147/IDR.S172226

Perales, C., Gallego, I., de Avila, A. I., Soria, M. E., Gregori, J., Quer, J., et al. (2019). The increasing impact of lethal mutagenesis of viruses. *Future Med. Chem.* 11, 1645–1657. doi: 10.4155/fmc-2018-0457

Perales, C., Ortega-Prieto, A. M., Beach, N. M., Sheldon, J., Menendez-Arias, L., and Domingo, E. (2017). Quasispecies and drug resistance. In: *Handbook of Antimicrobial Resistance*. New York, NY: Springer Science+Business Media.

Perpinan, E., Caro-Perez, N., Garcia-Gonzalez, N., Gregori, J., Gonzalez, P., Bartres, C., et al. (2018). Hepatitis C virus early kinetics and resistance-associated substitution dynamics during antiviral therapy with direct-acting antivirals. *J. Viral. Hepat.* 25, 1515–1525. doi: 10.1111/jvh.12986

Reed, L. J., and Muench, H. (1938). A simple method for estimating fifty per cent endpoint. *Am. J. Hyg.* 27, 493–497. doi: 10.1093/oxfordjournals.aje.a118408

Ribeiro, R. M., and Bonhoeffer, S. (2000). Production of resistant HIV mutants during antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* 97, 7681–7686. doi: 10.1073/pnas.97.14.7681

Ribeiro, R. M., Bonhoeffer, S., and Nowak, M. A. (1998). The frequency of resistant mutant virus before antiviral therapy. *Aids* 26, 461–465. doi: 10.1097/00002030-199805000-00006

Richman, D. D. (1996). *Antiviral Drug Resistance*. New York, NY: John Wiley and Sons Inc.

Ruggeri, M., Romano, F., Basile, M., Coretti, S., Rolli, F. R., Drago, C., et al. (2018). Cost-effectiveness analysis of early treatment of chronic HCV with Sofosbuvir/Velpatasvir in Italy. *Appl. Health Econ. Health Policy* 16, 711–722. doi: 10.1007/s40258-018-0410-x

Sabariegos, R., Ortega-Prieto, A. M., Diaz-Martinez, L., Grande-Perez, A., Garcia Crespo, C., Gallego, I., et al. (2022). Guanosine inhibits hepatitis C virus replication and increases indel frequencies, associated with altered intracellular nucleotide pools. *PLoS Pathog.* 18, e1010210. doi: 10.1371/journal.ppat.1010210

Sarrazin, C. (2016). The importance of resistance to direct antiviral drugs in HCV infection in clinical practice. *J. Hepatol.* 64, 486–504. doi: 10.1016/j.jhep.2015.09.011

Sarrazin, C. (2021). Treatment failure with DAA therapy: Importance of resistance. *J Hepatol* 74, 1472–1482. doi: 10.1016/j.jhep.2021.03.004

Sheldon, J., Beach, N. M., Moreno, E., Gallego, I., Pineiro, D., Martinez-Salas, E., et al. (2014). Increased replicative fitness can lead to decreased drug sensitivity of hepatitis C virus. *J. Virol.* 88, 12098–12111. doi: 10.1128/JVI.01860-14

Sorbo, M. C., Cento, V., Di Maio, V. C., Howe, A. Y. M., Garcia, F., Perno, C. F., et al. (2018). Hepatitis C virus drug resistance associated substitutions and their clinical relevance: Update 2018. *Drug Resist. Updat.* 37, 17–39. doi: 10.1016/j.drup.2018.01.004

Soria, M. E., Garcia-Crespo, C., Martinez-Gonzalez, B., Vazquez-Sirvent, L., Lobo-Vega, R., de Avila, A. I., et al. (2020). Amino acid substitutions associated with treatment failure for Hepatitis C virus infection. *J. Clin. Microbiol.* 58(12). doi: 10.1128/JCM.01985-20

Uchida, Y., Nakamura, S., Kouyama, J. I., Naiki, K., Motoya, D., Sugawara, K., et al. (2018). Significance of NS5B substitutions in Genotype 1b Hepatitis C virus evaluated by bioinformatics analysis. *Sci. Rep.* 8, 8818. doi: 10.1038/s41598-018-27291-7

Yang, S., Xing, H., Feng, S., Ju, W., Liu, S., Wang, X., et al. (2018). Prevalence of NS5B resistance-associated variants in treatment-naive Asian patients with chronic hepatitis C. *Arch. Virol.* 163, 467–473. doi: 10.1007/s00705-017-3640-6

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 September 2022
DOI 10.3389/fmicb.2022.979087

OPEN ACCESS

EDITED BY
Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands

REVIEWED BY
Assunta Venuti,
International Agency for Research on Cancer (IARC), France
Adriana Aguilar Lemarroy,
Centro de Investigación Biomédica de Occidente (CIBO), Mexico

*CORRESPONDENCE
Qingling Ren
yfy0047@njucm.edu.cn

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 27 June 2022
ACCEPTED 16 August 2022
PUBLISHED 14 September 2022

CITATION
Dai R, Tao R, Li X, Shang T, Zhao S and Ren Q (2022) Expression profiling of mRNA and functional network analyses of genes regulated by human papilloma virus E6 and E7 proteins in HaCaT cells.
Front. Microbiol. 13:979087.
doi: 10.3389/fmicb.2022.979087

Expression profiling of mRNA and functional network analyses of genes regulated by human papilloma virus E6 and E7 proteins in HaCaT cells

Renjinming Dai[1], Ran Tao[2], Xiu Li[1], Tingting Shang[1], Shixian Zhao[1] and Qingling Ren[1]*

[1]The Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China, [2]Laboratory of Clinical Applied Anatomy, Department of Human Anatomy, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China

Human papillomavirus (HPV) oncogenes E6 and E7 are essential for HPV-related cancer development. Here, we developed a cell line model using lentiviruses for transfection of the HPV16 oncogenes E6 and E7 and investigated the differences in mRNA expression during cell adhesion and chemokine secretion. Subsequently, RNA sequencing (RNA-seq) analysis was performed to explore the differences in mRNA expression. Compared to levels in the control group, 2,905 differentially expressed mRNAs (1,261 downregulated and 1,644 upregulated) were identified in the HaCaT-HPV16E6E7 cell line. To predict the functions of these differentially expressed genes (DEGs) the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes databases were used. Protein–protein interactions were established, and the hub gene was identified based on this network. Real-time quantitative-PCR (RT-qPCR) was conducted to confirm the levels of 14 hub genes, which were consistent with the RNA-seq data. According to this, we found that these DEGs participate in the extracellular matrix (ECM), cell adhesion, immune control, and cancer-related signaling pathways. Currently, an increasing number of clinicians depend on E6/E7mRNA results to make a comprehensive judgment of cervical precancerous lesions. In this study, 14 hub genes closely related to the expression of cell adhesion ability and chemokines were analyzed in HPV16E6E7-stably expressing cell lines, which will open up new research ideas for targeting E6E7 in the treatment of HPV-related cancers.

KEYWORDS

cell adhesion, cervical epithelioma, differentially expressed genes, HPV, oncogenes, RNA sequencing

Introduction

The human papillomavirus (HPV) is a DNA virus that encodes approximately eight genes. It can cause cancers of the antrum, genital tract, upper respiratory tract, and digestive tract (Dunne and Park, 2013). High-risk HPV E6 and E7 proteins reduce immune recognition by disrupting cell adhesion, polarity, and transcription via several pathways, such as the cell adhesion pathway (Oldak et al., 2006; Whiteside et al., 2008; Howie et al., 2009; Läubli and Borsig, 2019; Kombe Kombe et al., 2020). Studies have shown that the cervical squamous epithelium, which is the target of HPV, consists of keratinocytes, immature dendritic cells (DCs), and Langerhans cells (LCs), all of which perform immunosurveillance functions on the epithelium (Schiffman et al., 2007). LCs are the only specialized antigen-presenting cells found in the epidermis, and their abnormal expression affects antiviral immune responses. HPV-associated cervical squamous epithelial neoplasia (SIL) is strongly associated with immature LCs (Habbous et al., 2012). Cell adhesion between cells and the extracellular matrix (ECM) is required for the integrity of the epidermis and the formation of an immune barrier (Honig and Shapiro, 2020). Cell surface adhesion molecules, the major components of cell-to-cell and cell-to-extracellular interactions (Login et al., 2019), are frequently expressed in epithelial (Sumigray and Lechler, 2015), endothelial (Schimmel and Gordon, 2018), and immune (Dustin, 2009) cells. They consist of intracellular and extracellular components (CC) and can be divided into four types, cadherins, integrins, selectins, and immunoglobulins. Many studies have focused on malignant metastasis and cell adhesion (Janiszewska et al., 2020); however, the mechanisms related to interactions among HPV proteins, cell adhesion, and the epithelial immune response remains unclear. Activated leukocyte cell adhesion molecule (ALCAM) (Ferragut et al., 2021) is a cell-surface glycoprotein that has been recognized as a vital prognostic marker for various tumors, including HPV-associated head and neck squamous cell carcinoma (Zhu et al., 2020). Based on current research, ALCAM can regulate adhesion to immune cells (Te Riet et al., 2014) and control disease progression by affecting cell proliferation, migration, and invasion (von Lersner et al., 2019).

Chemokines are the largest subfamily of cytokines (Bule et al., 2021). It is well known that HPV is capable of preventing a robust immune response to the infected cells based on many mechanisms, such as by reducing the expression of chemokines (Attademo et al., 2020; Acevedo-Sánchez et al., 2021). When chemokines are unable to guide immune cells to migrate to the infected site in an orderly fashion, immune tolerance collapses and immune monitoring becomes ineffective (Kolaczkowska and Kubes, 2013).

With the increasing prevalence of Illumina technology in recent years, RNA-sequencing (RNA-Seq) technology has been effectively used in biomedical research (Tao et al., 2021; Wang et al., 2021). The carcinogenetic effect of E6 and E7 is apparent, but molecules that are induced by these proteins in HaCaT cells still need to be researched, and high-throughput sequencing would contribute vastly to this question. In this research, the differential mRNA expression profiles of HPV16E6E7-stably expressing cell lines were determined via RNA-seq. Our results showed that transcriptome levels were significantly different between the two groups of non-transfected control and HPV16E6E7-stably expressing cell lines. Transcriptome data are essential in providing a research target for mechanistic studies aimed at reversing cervical epithelioma and blocking the malignant progression of epithelial cells.

Materials and methods

Cell adhesion assays

E6-IREs-E7 was constructed and inserted into the plenti-Gm-cMV vector to obtain the vector plasmidpLenti-III-HPV-16E6-IRES-E7Vector (Puro). It was co-transfected with Lenti-ComboPackingMix and lentifectin into HEK293T. Cell supernatants were collected after the transfection, and the lentivirus was designated as Lenti-III-HPV-16E6-IRES-E7Virus (Puro). HaCaT cells were seeded onto 6-well plates at a density of approximately 20–30%, and then were infected with the lentivirus and incubated at 5% CO2, 37°C for 48 h. The infected HaCaT cells were plated in 96-well plates forstable transfection strain resistance screening. After HaCaT in logarithmic growth phase was infected by Lenti-III-HPV-16 E6-IRES-E7 Virus (Puro), four monoclonal cell lines,clone#4, clone#6, clone#10, clone#11, were obtained. Clone#11 showed the highest expression of target gene, validated by Real-time quantitative-PCR (RT-qPCR). Our team collaborated with ABM (Applied Biological Materials, Jiangsu, China) to establish this HaCaT-HPV16E6/E7. According to the introduction of the commercial cell adhesion kit (BestBio, Shanghai, China), the adhesion ability of parental HaCaT and HaCaT-HPV16E6E7 cells was defined. The optical density was recorded at 450 nm using a microplate reader (TECAN, SCHWEIZ). Finally, the level of cell adhesion was analyzed based on the OD value.

Western blot analysis

Protein samples were extracted from parental HaCaT and HaCaT-HPV16E6/E7 Clone#11 cell which were harvested in three different generations using RIPA buffer (Beyotime Biotechnology, Shanghai, China). Determination of total protein was performed using a commercial kit (Beyotime Biotechnology, Shanghai, China), and then, samples were incubated at 100°C for 10 min. Protein samples were separated

with 2–8% precast gels (ACE Biotechnology, Nangjing, China) and transferred onto PVDF (Millipore, Massachusetts, United States). The membranes were then soaked with 5% skim milk at 20–22°C for 1 h. After this, the solution was removed, and the membranes were incubated with primary antibodies, including those against ALCAM (1:1,000, Abmart, T57159), HPV16 E7 (1:500, bs-4623R), and HPV16 E6 (1:500, Abmart, V028329). The following day, the membrane was incubated with the secondary antibody (1:5,000, Cell Signaling Technology, 7076/7074) at 20–22°C for 1 h. The protein bands were developed using an ultrasensitive luminescent liquid reagent (Tanon, Shanghai, China) and analyzed using Image Lab version 5.1. The band density of the target protein to that of β-actin was used as the relative target protein content. All western blot experiments consisted of at least 3 independent replicates.

Real-time quantitative-PCR

We examined expression levels of the chemokine genes *CCL2*, *CCL3*, *CCL5*, *CCL7*, and *CXCL10* using RT-qPCR. The mRNA levels were determined by real time PCR. Total RNA was extracted from the parental HaCaT and HaCaT-HPV16E6/E7 Clone#11 cell which were harvested in three different generations using an RNA extraction kit (Vazyme, Nanjing, China). The cDNA product was obtained via reverse transcription using a commercial regent at 37°C for 15 min and then 85°C for 5 s (Vazyme, Nanjing, China). GAPDH and β-actin were used as internal controls. Primers were pre-designed and synthesized from a company (Sangon Biotech, Shanghai, China), and primer sequences are described in **Table 1**. Products were amplified using QUANTSTUDIO 7 FLEX QUANTSTUDIO (ABAppliedBiosystems, United States) according to the instructions for the Taq Pro Universal SYBR qPCR Master Mix kit (Vazyme, Nanjing, China). RT-qPCR was performed in technical triplicate for each biological replicate. Finally, Real-time PCR results were calculated according to the ΔΔCT method by using GAPDH and β-ACTIN as housekeeping genes.

RNA-Seq and gene expression analysis

RNA samples harvested from the parental HaCaT and HaCaT-HPV16E6/E7 Clone#11 cell, which were harvested in four different generations and were used for sequencing. Libraries for RNA-Seq were obtained from total RNA, and sequencing was performed on the NovaSeq 6000 platform by Gene *Denovo* Biotechnology Co., Ltd (Guangzhou, China). To quantify gene expression, HTSeq was used to obtain the original read count for each gene; alternatively, reads per kilobase of transcript per million mapped reads (RPKM) were used depending on the total map read count and the gene length for each sample. DEGs were determined based on the gene expression level measured using $\log_2$-transformed RPKM ($|\log_2 FC|$ (fold change) > 2, FDR < 0.01). The read counts data were entered and analyzed using EdgeR and DESeq2 software, including normalization of read counts and calculation of *P*-values and FDR values. Transcriptome data (SRA accession: PRJNA850539) were stored in the NCBI SRA database.[1]

Enrichment analysis

GO enrichment analysis includes biological process (BP), molecular function (MF), and CC, three aspects that are based on significantly enriched DEGs. The differential genes are mapped to each term of the GO database and the number of genes per term is counted to obtain a list of genes with a certain GO function and gene number statistics. Hypergeometric tests were applied to identify GO terms that were significantly enriched in differential genes compared to the whole genomic background. Briefly, all DEGs were mapped to GO terms in the GO database.[2] Then, gene numbers were calculated for every term, and GO terms associated with significantly enriched DEGs compared to the genome background were defined using a hypergeometric test. The calculated *p*-value was subjected to false discovery rate (FDR) correction, with FDR ≤ 0.05 as the threshold. Pathways meeting this condition were defined as significantly enriched pathways among the DEGs. The KEGG analysis was carried out according to KEGG.[3] GSEA was also performed with the GSEA software.[4]

Generation of the protein–protein interaction network

The protein–protein interaction (PPI) network was constructed using the String database,[5] for which genes and interactions were determined and designated as nodes and lines, respectively, in the network. To show the biological interactions among DEGs, their relationship network was visualized using Cytoscape (v3.9.2) software. Moreover, the cytoHubba plug-in was used to mine the hub genes using the MNC, Stress, Degree, Closeness, and Radiality calculation method in the Cytoscape MCODE plug-in Cytoscape, which was used to further extract the core sub-networks.

1 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA850539
2 http://www.geneontology.org/
3 http://www.genome.jp/kegg/
4 http://www.broadinstitute.org/gsea
5 https://cn.string-db.org/

TABLE 1 Primer lists.

Gene	Forward primer (5′–3′)	Reverse primer (5′–3′)
GAPDH	CATCATCCCTGCCTCTACTG	CTGCTTCACCACCTTCTTG
β-actin	GGCACCCAGCACAATGAAG	CCGATCCACACGGAGTACTTG
CCL2	CAGCCAGATGCAATCAATGCC	TGGAATCCTGAACCCACTTCT
CCL5	CCAGCAGTCGTCTTTGTCAC	CTCTGGGTTGGCACACACTT
CCL7	GCCTCTGCAGCACTTCTGTG	CACTTCTTGTGTGGGGTCAGC
CXCL10	GTGGCATTCAAGGAGTACCTC	TGATGGCCTTCGATTCTGGATT
CCL3	CAGCCAGATGCAATCAATGCC	TGGAATCCTGAACCCACTTCT
CDH1	CGAGAGCTACACGTTCACGG	GGGTGTCGAGGGAAAAATAGG
CDH2	TCAGGCGTCTGTAGAGGCTT	ATGCACATCCTTCGATAAGACTG
EGF	TGGATGTGCTTGATAAGCGG	ACCATGTCCTTTCCAGTGTGT
FGF2	AGAAGAGCGACCCTCACATCA	CGGTTAGCACACACTCCTTTG
BDNF	GGCTTGACATCATTGGCTGAC	CATTGGGCCGAACTTTCTGGT
SOX2	GCCGAGTGGAAACTTTTGTCG	GGCAGCGTGTACTTATCCTTCT
TLR4	AGACCTGTCCCTGAACCCTAT	CGATGGACTTCTAAACCAGCCA
IGF1	GCTCTTCAGTTCGTGTGTGGA	GCCTCCTTAGATCACAGCTCC
DLG4	TCGGTGACGACCCATCCAT	GCACGTCCACTTCATTTACAAAC
COL1A1	GAGGGCCAAGACGAAGACATC	CAGATCACGTCATCGCACAAC
FYN	ATGGGCTGTGTGCAATGTAAG	GAAGCTGGGGTAGTGCTGAG
CXCR4	ACTACACCGAGGAAATGGGCT	CCCACAATGCCAGTTAAGAAGA
WNT5A	ATTCTTGGTGGTCGCTAGGTA	CGCCTTCTCCGATGTACTGC
IL-1β	TCGCCAGTGAAATGATGGCT	TGGAAGGAGCACTTCATCTGTT

Statistical analysis

Data are presented from three independent experiments based on the means ± SEMs. For statistical analyses, GraphPad Prism 9 software was adopted (GraphPad, United States). When comparing two groups, unpaired *t*-tests were used, and differences among groups were analyzed using one-way ANOVA. A p-value < 0.05 was considered statistically significant.

Results

Human papillomavirus 16 oncogenes E6 and E7 affect cell adhesion in HaCaT cells

HaCaT cells stably expressed the E6 and E7 proteins after lentiviral transduction (**Supplementary Figure 1**). A cell adhesion reagent kit was used to determine whether this inhibited cell adhesion by measuring the optical density value in each group of cells. The overexpression of E6 and E7 significantly reduced the cell absorbance, indicating that they negatively affected cell adhesion (**Figure 1B**). Simultaneously, the protein expression level of the cell adhesion molecule (CAM) ALCAM was significantly decreased when examined via western blotting analysis (**Figure 1A** and **Supplementary Figures 2**, **3**).

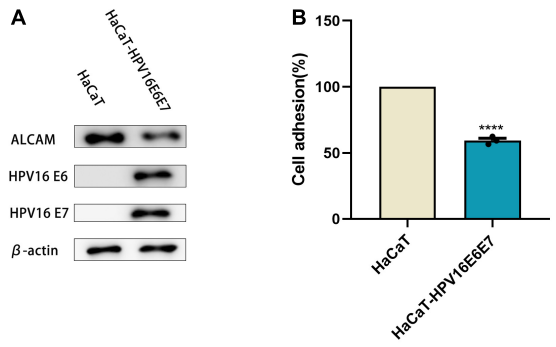

FIGURE 1
E6 and E7 injured the ability of cell adhesion in HaCaT **(A)** WB results showed that the expression of β-actin as an internal reference HPV16 E6E7, ALCAM was reduced in protein level. **(B)** The OD value of HaCaT-HPV16E6E7 was found to be significantly reduced in comparison with HaCaTHPV16 oncogenes E6 and E7 affect the secretion of chemokines in HaCaT. ****$p < 0.0001$.

Human papillomavirus16 oncogenes E6 and E7 affect HaCaT cell adhesion

Chemokines are effective chemical attractants for natural killer cells, macrophages, and monocytes, and they engage in the host immune response against HPV-infected cervical epithelia. CCL2, CCL3, CCL5, and CCL7 attract monocytes/macrophages

to exert anti-infective, anti-tumor, and immunomodulatory effects at sites of inflammation. In response to pro-inflammatory stimuli, such as IL-1, TNF-α, LPS, or viruses, the expression of CCL2, CCL3, CCL5, and CXCL10 recruit immune cells to sites of inflammation. However, after overexpression of the HPVE6E7 protein, the expression levels of CCL2, CCL3, CCL5, CCL7, and CXCL10 decreased with different magnitudes. The selective loss of chemokine expression, including that of CCL2, was observed after lentiviral transduction. Moreover, their mRNA levels were significantly lower than those in keratinocyte HaCaT cells (**Figure 2** and **Supplementary Figure 4**).

Overview of differential mRNA expression

The original RNA-seq data (Read Count) and differential expression analysis results (DESeq2_analysis_results) used in this study were uploaded, and these tables can be found in the **Supplementary Material 1**. To identify DEGs between the cell groups, the Limma package was used to analyze data. The criteria were as follows: $|\log_2 FC| > 2$ and $p < 0.01$ (**Figure 3A**). After this step, the mRNA expression of 1,644 genes were found to be upregulated, whereas that of 1,261 genes was downregulated; according to our data, DEGs based on mRNA expression were presented as a volcano map using the ggplot2 package (**Figure 3B**). Heatmaps of the DEGs were created with the heatmap package (**Figure 3C**).

Enrichment analyses of differentially expressed mRNAs

The top 20 KEGG pathways (**Figure 4A**) indicated that DEGs were enriched in the Hippo and Wnt pathways and signaling pathways that regulate the ECM receptor interaction, among others. These signaling pathways exhibit crosstalk with cell adhesion by regulating the transcription of chemokines, CAMs, and growth factors, including cadherin 1 (CDH1), CDH2, epidermal growth factor (EGF), brain-derived neurotrophic factor (BDNF), interleukin (IL)-1β, fibroblast growth factor (FGF2), disks large homolog 4 (DLG4), Toll-like receptor 4 (TLR4), insulin-like growth factor 1 (IGF1), collagen type I alpha 1 (COL1A1), CXCR4, and WNT5A. In the GO analysis, DEGs closely related to several BP, MF, and CC included genes involved in adhesion, immune system, cell aggregation, transcription regulator activity, and cell junction (**Figure 4B**). In addition, GESA and KEGG analyses showed similar results, such as genes involved in tight junctions (**Supplementary Figure 5**).

Analysis of protein–protein interaction network

To provide a better understanding of interactions among the DEGs, their corresponding PPI network was generated using Cytoscape software and the STRING database. The "cytoHubba" plug-in identified the 14 core genes based on differences in the level of linkage (**Figures 5A–E**). These 14hub genes are obtained by taking the intersection of the DEGs obtained from five different algorithms, and are presented with the form of Venn diagram (**Figure 5F**). The interactive density region in the PPI network was also determined using the "MCODE" plug-in (**Figures 6A–D**). The figure below shows the densest regions (**Figure 5E**). For example, BDNF binds to CDH2 and CXCR4, TLR4 to CXCR4 and IL-1β, and CDH1 to FYN and IGF1. In this study, it was considered that these proteins play useful roles in regulating chemokine binding and macrophage-mediated cytokine production.

Verification of differentially expressed mRNAs

RT-qPCR was performed to verify the transcriptome sequencing results before and after lentiviral transduction and define the possible signaling network associated with the E6 and E7 oncogenes upon HPV16 infection. The RT-qPCR data showed that the expression of *CXCR4* was decreased in cells overexpressing the HPV16 oncogenes E6 and E7, as with the findings of RNA-seq. However, *SOX2* and *FYN* expression levels were overexpressed, in contrast to RNA-seq results. The mRNA expression levels of *CDH1*, *CDH2*, *EGF*, *IL-1β*, *FGF2*, *BDNF*, *TLR4*, *IGF1*, *DLG4*, *COL1A1*, *CXCR4*, and *WNT5A* were analyzed by RT-qPCR, and their expression was in agreement with the outcomes of RNA-Seq. (**Figure 7** and **Supplementary Figure 6**).

Discussion

It is well known that E6 and E7 can attach to the ECM of the cervical squamous epithelium. Further, HPV E6 and E7 can be detected in a substantial proportion of women with cervical intraepithelial neoplasia (CIN) (Guan et al., 2012); they also serve as attractive targets for HPV-related diseases (Hoppe-Seyler et al., 2018). Owing to its specific sites and functions in antigen presentation, LCs are considered essential for the immune response against HPV infection (Cunningham et al., 2010). In the epidermis of mice (Kaplan et al., 2005) and female genital tract (Britto et al., 2020), LCs produce effector cytokines to stimulate T lymphocytes and B lymphocytes through interactions with C-type lectin (CLR receptors) and TLRs. Hubert et al. (Hubert et al., 2005) confirmed decreased levels

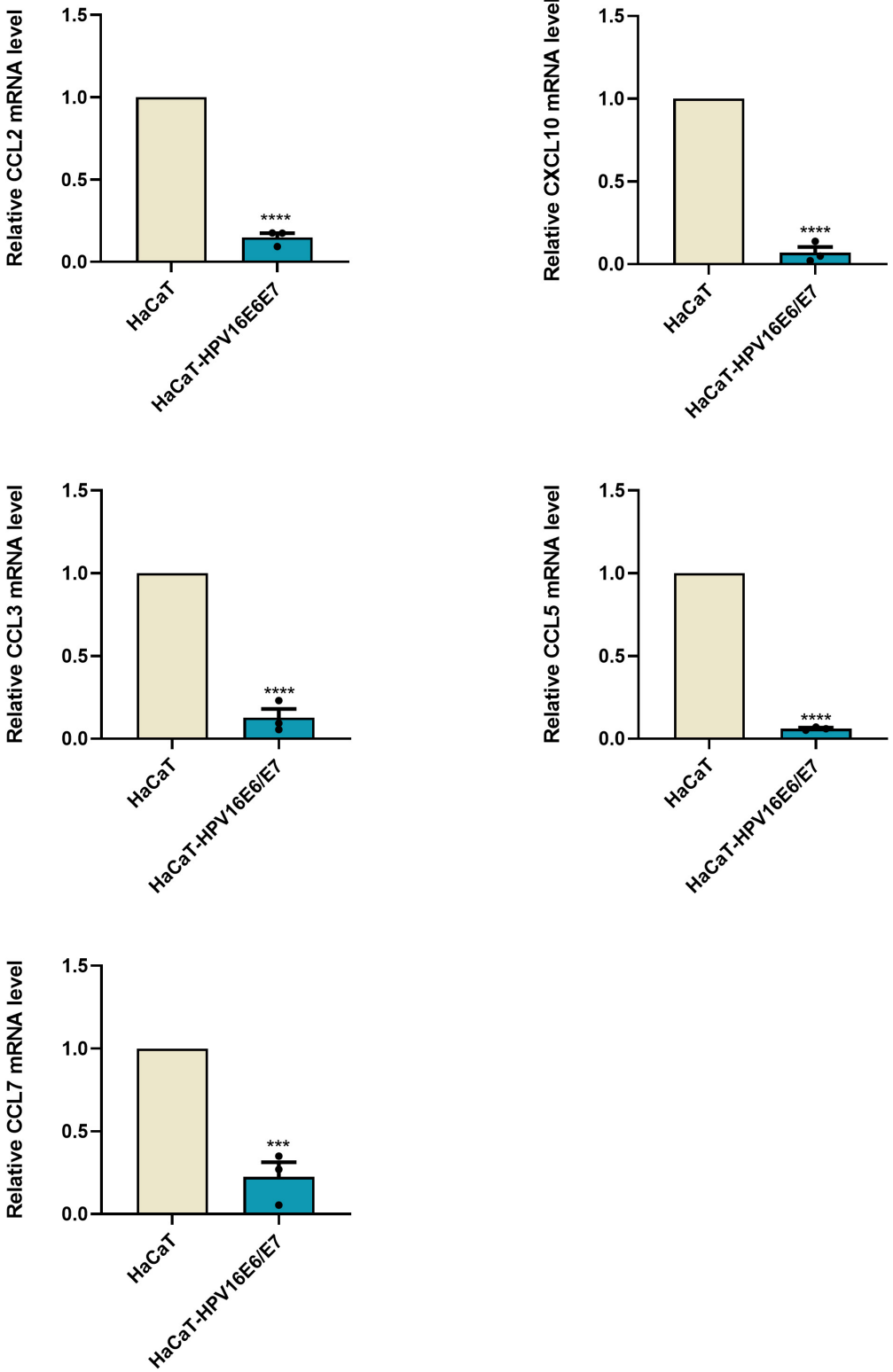

FIGURE 2
E6 and E7 damaged the secretion of chemokines in HaCaT RT-qPCR results were quantified by real-time PCR using GAPDH and β-ACTIN as internal control genes and presented in bar graph. It can be seen that overexpression of E6E7 does lead to a decrease in the relative expression of CCL2, CCL3, CCL5, CCL7, and CXCL10. ***$p < 0.001$, ****$p < 0.0001$.

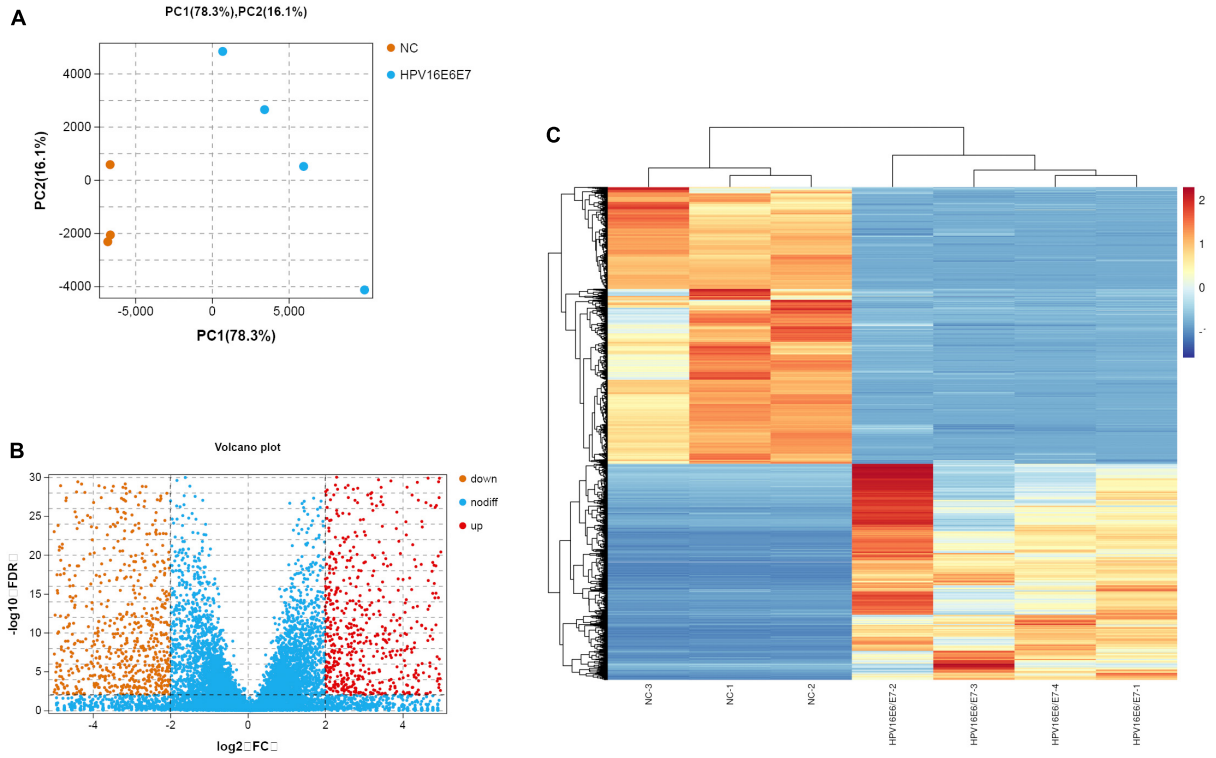

FIGURE 3
Overview of the differential mRNA expression between two groups **(A)** Use R software to carry out PCA, the horizontal and vertical coordinate represent PC1 and PC2 representatively; The distance in the graph reflects the degree of the difference between groups. **(B)** In the volcano plot, orange and red represent down- and up-regulated genes, respectively, while blue expresses genes that are not significantly different. **(C)** In the Heat Map, the closer the value of the vertical coordinate to red indicates a greater relative gene expression, and vice versa.

of *CDH1* mRNA and numbers of CD1α^{+} LCs in squamous cell carcinoma and SILs compared with those in normal cervical epithelia. According to the high throughput screening of clinical specimens (Britto et al., 2020), the decrease observed in the mRNA levels of *CDH1*, *Claudin 1*, *Claudin 4*, and *ZO-1* mRNA and IFN signaling upon HPV infection indicate that it could result in weakness of the epithelial barrier (Aggarwal et al., 2019; Giorgi Rossi et al., 2021). Collectively, these findings suggest that the E6 and E7 oncoproteins are strongly associated with cell adhesion in HPV-induced immunosuppression. The loss of cell adhesion caused by the imbalance in adhesion molecules and chemokines might be another immune-escape strategy that has evolved in HPV.

CXC motif chemokine ligand 10 (CXCL10), CC motif chemokine ligand 2 (CCL2), CCL3, and CCL5 are pro-inflammatory chemokines that actively participate in the inflammatory response after viral exposure (Kolaczkowska and Kubes, 2013). CCL2 (Nakamura et al., 1995; Kunstfeld et al., 1998; Hacke et al., 2010) is a key chemokine that attracts cells, including T lymphocytes, eosinophil granulocytes, macrophages, mast cells, and monocytes, to inflammatory sites. In general, CCL2 expression is activated after viral infection and recruits Langerhans and DCs, respectively. However, co-expression of the E6 and E7 proteins inhibits CCL2 expression in primary epithelial cells of the female genital tract (Hacke et al., 2010). An *in vitro* study (Nakamura et al., 1995) has suggested that HPV E7 interacts with interferon regulatory factor 1 (IRF-1), a significant regulator of cellular immune responses, to suppress CCL2 expression. This partly explains the almost undetectable activation of potent anti-tumor immune responses in primary HPV-infected epithelial tissues. It has been reported that humans lack a robust inflammatory response against primary HPV infection and progressively lose immune cells in the cervical stroma during CIN (Kleine-Lowinski et al., 2003; Hacke et al., 2010; Garrido et al., 2021). CIN is strongly associated with CCL2 expression in the tumor microenvironment. Habbous et al. (2012) suggested that as CIN progresses, levels of many hosts immune response markers, including the chemokine CCL2, are reduced. However, the specific mechanisms involved remain unclear. As epithelial dysplasia progresses to cervical cancer, macrophages, T cells, and LCs of the cervical mucosa are depleted. Restoring CCL2 expression can specifically improve the anti-tumor immunity of the host and block disease progression. Therefore, exploring

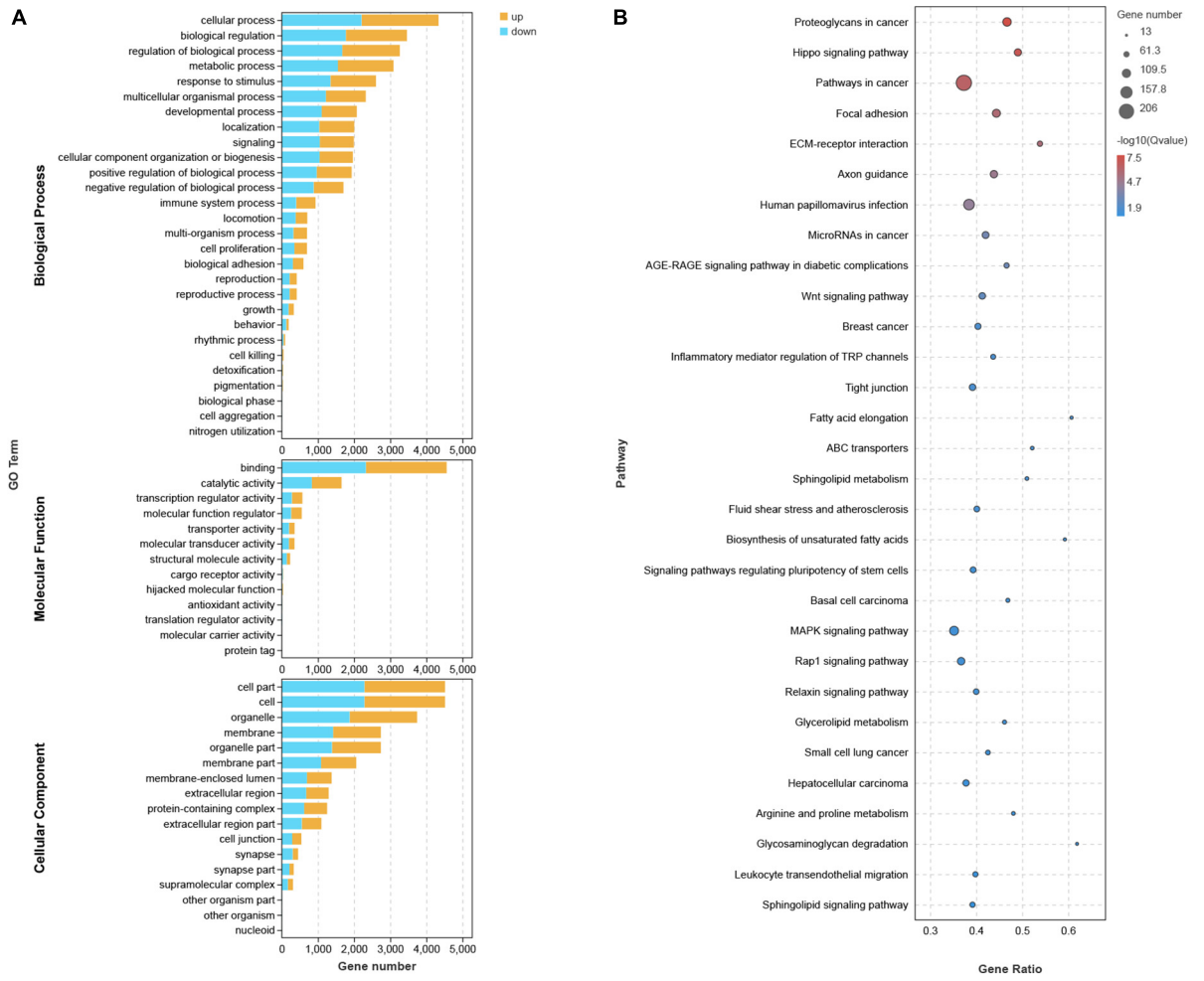

FIGURE 4
GO, KEGG Analysis. **(A)** Each bar in GO analysis represents a pathway, where the up-regulated and down-regulated genes involved in this pathway are indicated in yellow and blue, respectively. **(B)** In KEGG Pathway analysis, the size of the circle indicates the number of enriched genes.

the selective loss of CCL2 expression after E6/E7 transduction is crucial to blocking disease progression. However, in cell adhesion, the role of chemokines still needs to be explored in many ways. CXCL10 (Halle et al., 2017; Takeuchi and Saito, 2017) is associated with increased quantities of tumor-infiltrating $CD8^+$ T cells, inducing the release of cytotoxic molecules (such as granzyme B and perforin) and apoptosis, which are associated with lower levels of cancer metastasis and improved patient survival in ovarian and colon cancer. CCL5 (Bruand et al., 2021) expression was found to be positively correlated with $CD8^+$ T cell infiltration. T cell infiltration reportedly requires the secretion of CCL5 by cells in both epithelial cancer cell models and animal models. Zhang et al. (2020) suggested that CCL7 promotes anti-PD-1 therapy (CDC1) in a mouse model by recruiting conventional DCs to the tumor microenvironment to promote T cell proliferation. In addition, chemokines contribute to anti-tumor immunity by recruiting tumor-associated antigen-presenting cells, promoting their binding to T cells, and influencing tumorigenesis by interacting with tumor stem cell-like cells and stromal cells (Kombe Kombe et al., 2020). In summary, different lymphocyte subsets are recruited into the tumor microenvironment through different chemokine–chemokine receptor signaling pathways. The restricted expression of chemokines directly or indirectly leads to the failure of anti-tumor immunity.

Our study revealed 1,644 upregulated mRNAs and 1,261 downregulated mRNAs after E6/E7 overexpression. According to RNA-seq analysis, there is a tight connection between DEGs and cell adhesion, focal adhesion, leukocyte–endothelial migration, tight junctions, and Hippo and Wnt signaling pathways. KEGG analysis revealed that these DEGs were highly focused in cancer proliferation-related pathways, including the MAPK, Hippo, Wnt, and cell adhesion signaling pathways. It has been suggested that E6/E7 might exhibit crosstalk

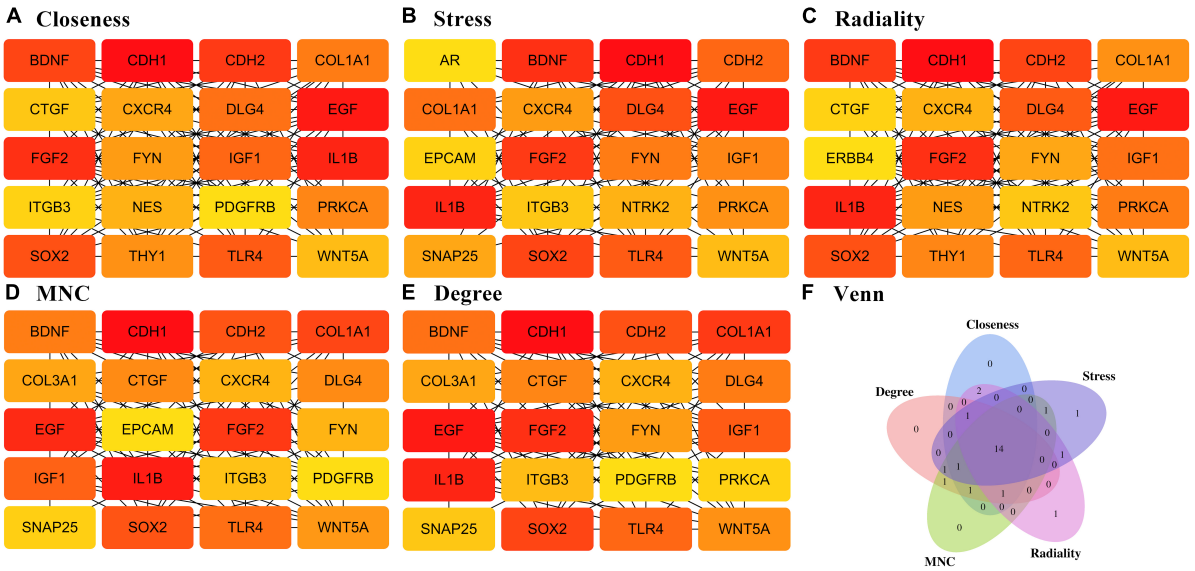

FIGURE 5
PPI network based on DEGs. **(A–E)** Different algorithms for interaction analysis of the top 20 pivotal genes. **(F)** Wayne diagram showing the 14-hub gene intersected by different algorithms.

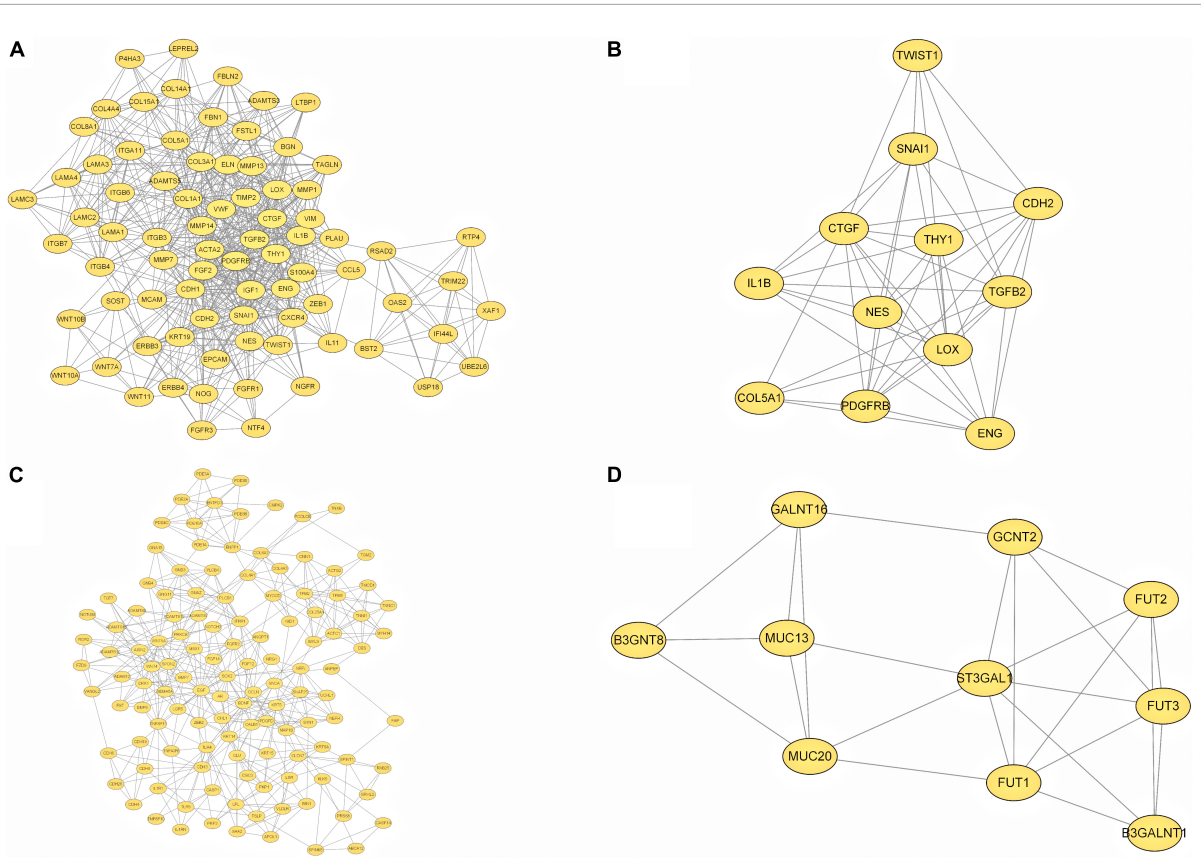

FIGURE 6
PPI network based on DEGs. **(A–D)** The top four interactive density region in the PPI network by "MCODE" plug-in was also discovered.

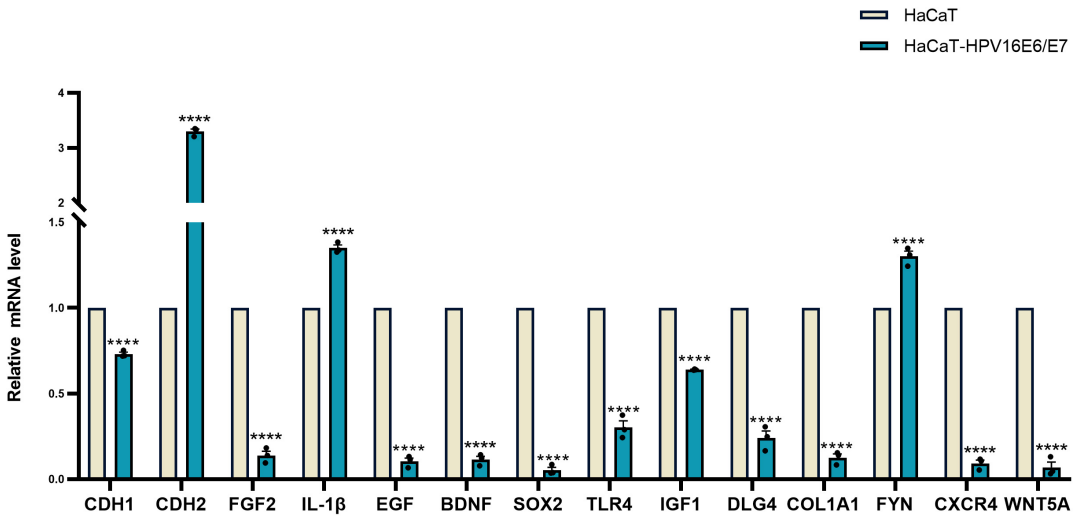

FIGURE 7
Verification of 14 hub gene Relative expression results of 14 hub gene were quantified by real-time PCR using GAPDH and β-ACTIN as internal control genes and shown by the bar graph whose horizontal and vertical coordinates represent the gene names and relative expressions, respectively. The verification is consistent with High-throughput sequencing results. ****$p < 0.0001$.

with these signaling pathways, affecting cell adhesion and limiting the immune response. CXCR4 (Zou et al., 1998; Goïta and Guenot, 2022) is a chemokine G protein-coupled transmembrane receptor released by stromal cells. Stromal cell-derived factor-1 (SDF-1; CXCL12), a ligand of CXCR4, induces B-cell proliferation and T-cell recruitment by activating CXCR4-expressing cells. Studies have demonstrated that the immune infiltration of B cells in the tumor microenvironment is associated with a survival advantage in patients with cervical (Abdul Rahman et al., 2020), breast (Hassel et al., 2021), and high-grade serous ovarian (Montfort et al., 2017) cancer; B cells are recruited through the binding of CXCL12 and CXCR4 (Greenbaum et al., 2013). Furthermore, CXCL12/CXCR4 enhances Src phosphorylation by activating MAPK or mitogen-activated protein kinase (MEK)/extracellular signal-regulated kinase (ERK) kinases (Kasina and Macoska, 2012). CXCR4-induced signaling cascades are commonly referred to as the "CXCR4–SDF-1 axis" and are frequently associated with lymphocyte trafficking and homing. To participate in the immune response (Peled et al., 1999), CXCL12 binds to CXCR4 to stimulate the firm adhesion of leukocytes by increasing integrin stability, which binds to intracellular adhesion molecules (ICAM-1) and vascular CAMs (VCAM-1). Researchers (Zheng et al., 2017) have used the CRISPR-Cas9 system in a cellular model to knock out CXCR4 and observed that colorectal cancer cell lines showed reduced adhesion to the ECM and endothelium, further demonstrating that CXCR4 is closely related to the Akt and type 1 insulin-like growth factor receptor (IGF1R) signaling pathways. In mouse models (Aggarwal et al., 2019), the CXCL12/CXCR4 axis enhances the function of M1-type macrophages, stimulates the production of inflammatory cytokines, and increases the phagocytosis of pathogens. Furthermore, it has been shown that CXCL12 primes IL-1 production in M2 macrophages by inducing its receptor CXCR4 and directly mediating the activation of $CD4^{+}$T cells (Yu et al., 2020). Macrophages, an essential component of the innate immune system, are highly heterogeneous and malleable, and the reduced expression of CXCR4 might affect the dysregulation of M1/M2 phenotypes (Beneduce et al., 2019). This could play a vital role in the immune escape mediated by E6 and E7. These results suggest that CXCL12/CXCR4 is a potential therapeutic target for treating autoimmune diseases, including human immunodeficiency virus infection, cancer, warts, and immunodeficiency (Comba et al., 2020).

The expression of E6 and E7 resulted in the detection of many DEGs using RNA-seq and RT-qPCR. Fourteen of these hub genes were linked to cell adhesion (Chute, 2006; Teicher and Fricker, 2010; Paolillo and Schinelli, 2019; Marotta et al., 2021), leukocyte–endothelial migration (Yakovlev et al., 2019; Morein et al., 2020; Yan et al., 2021), stem cell signaling (Davidson et al., 2007; Bijlmakers, 2009), and epithelial–mesenchymal transition (Pastushenko et al., 2021; Wang et al., 2021). Based on these experimental and bioinformatics analyses, we hypothesized that HPV16 E6 and E7 signaling causes differential mRNA expression through direct or indirect mRNA pathways that regulate cell adhesion and chemokine secretion.

In this study, we used RNA-seq, RT-qPCR, and bioinformatics to validate the HPV-associated signaling pathways and possible downstream signaling molecules of the E6 and E7 proteins in human keratinocytes. This study

provides a new direction and basis for further elucidation of the mechanism of E6/E7-mediated cell adhesion suppression. However, on the one hand, our study was restricted to *in vitro* conditions and the mode of interaction between CCL2 and hub genes is unclear in keratinocytes immortalized with both HPV E6 and E7 genes. The detailed mechanism will be investigated profoundly in our future studies. On the other hand, although some studies have indicated that FGF stimulation produces phosphorylation of E-cadherin and β-catenin on tyrosine residues, as well as increased E-cadherin localization to the cytoplasmic membrane and association with FGFR1 demonstrable by co-immunoprecipitation in the human pancreatic adenocarcinoma cell lines (BxPc3, T3M4 and HPAF), more direct interactions between hub genes need to be verified by co-immunoprecipitation or immunoblastic hybridization in our cellular models. The levels of mRNA only respond to the transcriptional aspect, while the process of post-translational modifications is highly variable, with different modifications resulting in different protein expression. Such differences may be influenced by post-transcriptional mechanisms such as methylation and acetylation. A previous study revealed that the level of m6A-modified mRNA is increased in cancer cells during the epithelial-mesenchymal transition of hepatocellular carcinoma cell metastasis (Lin et al., 2019). Although several studies have confirmed that the mRNA levels of certain hub genes such as CHD1, CDH2 (Castillo et al., 2016; Wang et al., 2020), IL-1β, EGF (Fu et al., 2020), BDNF (McDole et al., 2015), SOX2 (Scapin et al., 2019), and TLR4 (Jin et al., 2017) were consistent with the protein levels, it has not been validated in HaCaT-HPV16E6E7 cells. Considering this special feature, we will focus on discovering and exploring the differences in hub gene at the mRNA and protein levels in our future studies. It is undeniable that lentiviral vectors have an effect on gene expression, and we chose parental HaCaT as a negative control in our study, which makes it difficult to exclude the differential gene changes that would result from null loading at this time.

Data availability statement

The data presented in this study are deposited in the SRA repository, accession number: PRJNA850539 can be found in the article/**Supplementary material**.

Author contributions

RD and QR conceived and designed the experiments and revised the manuscript. RD and XL performed the experiments. RT carried out RNA-seq. RD, RT, SZ, TS, and QR analyzed the data. RD wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (General Program, 8207150903), the Chinese Clinical Medicine Innovation Center of Obstetrics, Gynecology, and Reproduction in Jiangsu Province (ZX202102), and the Postgraduate Research and Practice Innovation Program of Jiangsu Province, China (KYCX21_1680).

Acknowledgments

We thank Micaella Aliza from Editage for editing the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.979087/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Verification of HPV 16 E6/E7 Stable Expression HaCaT Cell Line Clone#4, #6, #10, #11 The results suggested that HPV 16 E6/E7 Stable Expression HaCaT Cell Line Clone#4, #6, #10, #11 all displayed significant overexpression of HPV16 E6 and E7. While, Clone#11 showed the highest expression of target gene, validated by RT-qPCR method.

SUPPLEMENTARY FIGURE 2
The original western blots of HPV16E6 with protein markers WB results showed that the expression of β-actin as an internal reference HPV16 E6 was reduced in protein level.

SUPPLEMENTARY FIGURE 3
The original western blots of HPV16E7 with protein markers WB results showed that the expression of β-actin as an internal reference HPV16 E7 was reduced in protein level.

SUPPLEMENTARY FIGURE 4
(A) RT-qPCR results that used β-ACTIN as internal control genes. **(B)** RT-qPCR results that used GAPDH as internal control genes. Verification of 14 hub gene RT-qPCR results were quantified by real-time PCR using GAPDH and β-ACTIN as internal control genes, respectively. It can be seen in bar graph that the verification is consistent with the results of **Figure 1**. ***$p < 0.001$, ****$p < 0.0001$.

SUPPLEMENTARY FIGURE 5
GESA analyses The vertical coordinate is the ES value and the horizontal coordinate is the sequencing of the gene. According to the result of gene sequencing, the genes in that pathway/term are scored plus and the genes not in that pathway/term are scored minus. The final maximum value obtained is the maximum value for that pathway/term. The points that are added will be marked with "vertical lines" below the graph.

SUPPLEMENTARY FIGURE 6
Verification of 14 hub gene Relative expression results of 14 hub gene were quantified by real-time PCR using GAPDH and β-ACTIN as internal control genes, respectively and shown by the bar graph whose horizontal and vertical coordinates represent the gene, names and relative expressions, respectively. The verification is consistent with the results of **Figure 7**. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

SUPPLEMENTARY MATERIAL
The differential expression analysis results (DESeq2_analysis_results) used in this study were uploaded, and these tables were placed in the Supplementary Material.

References

Abdul Rahman, S. F., Xiang Lian, B. S., and Mohana-Kumaran, N. (2020). Targeting the B-cell lymphoma 2 anti-apoptotic proteins for cervical cancer treatment. *Future Oncol.* 16, 2235–2249. doi: 10.2217/fon-2020-0389

Acevedo-Sánchez, V., Rodríguez-Hernández, R. M., Aguilar-Ruíz, S. R., Torres-Aguilar, H., and Romero-Tlalolini, M. L. A. (2021). Extracellular vesicles in cervical cancer and HPV infection. *Membranes (Basel)* 11:453. doi: 10.3390/membranes11060453

Aggarwal, C., Cohen, R. B., Morrow, M. P., Kraynyak, K. A., Sylvester, A. J., Knoblock, D. M., et al. (2019). Immunotherapy targeting HPV16/18 generates potent immune responses in HPV-associated head and neck cancer. *Clin. Cancer Res.* 25, 110–124. doi: 10.1158/1078-0432.Ccr-18-1763

Attademo, L., Tuninetti, V., Pisano, C., Cecere, S. C., Di Napoli, M., Tambaro, R., et al. (2020). Immunotherapy in cervix cancer. *Cancer Treat Rev.* 90:102088. doi: 10.1016/j.ctrv.2020.102088

Beneduce, E., Matte, A., De Falco, L., Mbiandjeu, S., Chiabrando, D., Tolosano, E., et al. (2019). Fyn kinase is a novel modulator of erythropoietin signaling and stress erythropoiesis. *Am. J. Hematol.* 94, 10–20. doi: 10.1002/ajh.25295

Bijlmakers, M. J. (2009). Protein acylation and localization in T cell signaling (Review). *Mol. Membr. Biol.* 26, 93–103. doi: 10.1080/09687680802650481

Britto, A. M. A., Goes, L. R., Sivro, A., Policarpo, C., Meirelles, ÂR., Furtado, Y., et al. (2020). HPV Induces changes in innate immune and adhesion molecule markers in cervical mucosa with potential impact on HIV infection. *Front. Immunol.* 11:2078. doi: 10.3389/fimmu.2020.02078

Bruand, M., Barras, D., Mina, M., Ghisoni, E., Morotti, M., Lanitis, E., et al. (2021). Cell-autonomous inflammation of BRCA1-deficient ovarian cancers drives both tumor-intrinsic immunoreactivity and immune resistance via STING. *Cell Rep.* 36:109412. doi: 10.1016/j.celrep.2021.109412

Bule, P., Aguiar, S. I., Aires-Da-Silva, F., and Dias, J. N. R. (2021). Chemokine-directed tumor microenvironment modulation in cancer immunotherapy. *Int. J. Mol. Sci.* 22:9804. doi: 10.3390/ijms22189804

Castillo, L. F., Tascón, R., Lago Huvelle, M. R., Novack, G., Llorens, M. C., Dos Santos, A. F., et al. (2016). Glypican-3 induces a mesenchymal to epithelial transition in human breast cancer cells. *Oncotarget* 7, 60133–60154. doi: 10.18632/oncotarget.11107

Chute, J. P. (2006). Stem cell homing. *Curr. Opin. Hematol.* 13, 399–406. doi: 10.1097/01.moh.0000245698.62511.3d

Comba, A., Dunn, P. J., Argento, A. E., Kadiyala, P., Ventosa, M., Patel, P., et al. (2020). Fyn tyrosine kinase, a downstream target of receptor tyrosine kinases, modulates antiglioma immune responses. *Neuro Oncol.* 22, 806–818. doi: 10.1093/neuonc/noaa006

Cunningham, A. L., Abendroth, A., Jones, C., Nasr, N., and Turville, S. (2010). Viruses and langerhans cells. *Immunol. Cell Biol.* 88, 416–423. doi: 10.1038/icb.2010.42

Davidson, D., Schraven, B., and Veillette, A. (2007). PAG-associated FynT regulates calcium signaling and promotes anergy in T lymphocytes. *Mol. Cell Biol.* 27, 1960–1973. doi: 10.1128/mcb.01983-06

Dunne, E. F., and Park, I. U. (2013). HPV and HPV-associated diseases. *Infect. Dis. Clin. North Am.* 27, 765–778. doi: 10.1016/j.idc.2013.09.001

Dustin, M. L. (2009). Supported bilayers at the vanguard of immune cell activation studies. *J. Struct. Biol.* 168, 152–160. doi: 10.1016/j.jsb.2009.05.007

Ferragut, F., Vachetta, V. S., Troncoso, M. F., Rabinovich, G. A., and Elola, M. T. (2021). ALCAM/CD166: A pleiotropic mediator of cell adhesion, stemness and cancer progression. *Cytokine Growth Factor Rev.* 61, 27–37. doi: 10.1016/j.cytogfr.2021.07.001

Fu, B., Yin, S., Lin, X., Shi, L., Wang, Y., Zhang, S., et al. (2020). PTPN14 aggravates inflammation through promoting proteasomal degradation of SOCS7 in acute liver failure. *Cell Death Dis.* 11:803. doi: 10.1038/s41419-020-03014-7

Garrido, F., Wild, C. M., Mittelberger, J., Dobler, F., Schneider, M., Ansorge, N., et al. (2021). The role of chemokines in cervical cancers. *Medicina (Kaunas)* 57:1141. doi: 10.3390/medicina57111141

Giorgi Rossi, P., Carozzi, F., Ronco, G., Allia, E., Bisanzi, S., Gillio-Tos, A., et al. (2021). p16/ki67 and E6/E7 mRNA accuracy and prognostic value in triaging HPV DNA-positive women. *J. Natl. Cancer Inst.* 113, 292–300. doi: 10.1093/jnci/djaa105

Goïta, A. A., and Guenot, D. (2022). Colorectal cancer: The contribution of CXCL12 and its receptors CXCR4 and CXCR7. *Cancers (Basel)* 14:1810. doi: 10.3390/cancers14071810

Greenbaum, A., Hsu, Y. M., Day, R. B., Schuettpelz, L. G., Christopher, M. J., Borgerding, J. N., et al. (2013). CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* 495, 227–230. doi: 10.1038/nature11926

Guan, P., Howell-Jones, R., Li, N., Bruni, L., de Sanjosé, S., Franceschi, S., et al. (2012). Human papillomavirus types in 115,789 HPV-positive women: A meta-analysis from cervical infection to cancer. *Int. J. Cancer* 131, 2349–2359. doi: 10.1002/ijc.27485

Habbous, S., Pang, V., Eng, L., Xu, W., Kurtz, G., Liu, F. F., et al. (2012). p53 Arg72Pro polymorphism. HPV status and initiation, progression, and development of cervical cancer: A systematic review and meta-analysis. *Clin. Cancer Res.* 18, 6407–6415. doi: 10.1158/1078-0432.Ccr-12-1983

Hacke, K., Rincon-Orozco, B., Buchwalter, G., Siehler, S. Y., Wasylyk, B., Wiesmüller, L., et al. (2010). Regulation of MCP-1 chemokine transcription by p53. *Mol. Cancer* 9:82. doi: 10.1186/1476-4598-9-82

Halle, S., Halle, O., and Förster, R. (2017). Mechanisms and dynamics of T cell-mediated cytotoxicity in vivo. *Trends Immunol.* 38, 432–443. doi: 10.1016/j.it.2017.04.002

Hassel, C., Gausserès, B., Guzylack-Piriou, L., and Foucras, G. (2021). Ductal macrophages predominate in the immune landscape of the lactating mammary gland. *Front. Immunol.* 12:754661. doi: 10.3389/fimmu.2021.754661

Honig, B., and Shapiro, L. (2020). Adhesion protein structure. Molecular affinities, and principles of cell-cell recognition. *Cell* 181, 520–535. doi: 10.1016/j.cell.2020.04.010

Hoppe-Seyler, K., Bossler, F., Braun, J. A., Herrmann, A. L., and Hoppe-Seyler, F. (2018). The HPV E6/E7 oncogenes: Key Factors for viral carcinogenesis and therapeutic targets. *Trends Microbiol.* 26, 158–168. doi: 10.1016/j.tim.2017.07.007

Howie, H. L., Katzenellenbogen, R. A., and Galloway, D. A. (2009). Papillomavirus E6 proteins. *Virology* 384, 324–334. doi: 10.1016/j.virol.2008.11.017

Hubert, P., Caberg, J. H., Gilles, C., Bousarghin, L., Franzen-Detrooz, E., Boniver, J., et al. (2005). E-cadherin-dependent adhesion of dendritic and Langerhans cells to keratinocytes is defective in cervical human papillomavirus-associated (pre)neoplastic lesions. *J. Pathol.* 206, 346–355. doi: 10.1002/path.1771

Janiszewska, M., Primi, M. C., and Izard, T. (2020). Cell adhesion in cancer: Beyond the migration of single cells. *J. Biol. Chem.* 295, 2495–2505. doi: 10.1074/jbc.REV119.007759

Jin, X., Chen, C., Li, D., Su, Q., Hang, Y., Zhang, P., et al. (2017). PRDX2 in myocyte hypertrophy and survival is mediated by TLR4 in acute infarcted myocardium. *Sci. Rep.* 7:6970. doi: 10.1038/s41598-017-06718-7

Kaplan, D. H., Jenison, M. C., Saeland, S., Shlomchik, W. D., and Shlomchik, M. J. (2005). Epidermal langerhans cell-deficient mice develop enhanced contact hypersensitivity. *Immunity* 23, 611–620. doi: 10.1016/j.immuni.2005.10.008

Kasina, S., and Macoska, J. A. (2012). The CXCL12/CXCR4 axis promotes ligand-independent activation of the androgen receptor. *Mol. Cell. Endocrinol.* 351, 249–263. doi: 10.1016/j.mce.2011.12.015

Kleine-Lowinski, K., Rheinwald, J. G., Fichorova, R. N., Anderson, D. J., Basile, J., Münger, K., et al. (2003). Selective suppression of monocyte chemoattractant protein-1 expression by human papillomavirus E6 and E7 oncoproteins in human cervical epithelial and epidermal cells. *Int. J. Cancer* 107, 407–415. doi: 10.1002/ijc.11411

Kolaczkowska, E., and Kubes, P. (2013). Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* 13, 159–175. doi: 10.1038/nri3399

Kombe Kombe, A. J., Li, B., Zahid, A., Mengist, H. M., Bounda, G. A., Zhou, Y., et al. (2020). Epidemiology and Burden of human papillomavirus and related diseases, molecular pathogenesis, and vaccine evaluation. *Front. Public Health* 8:552028. doi: 10.3389/fpubh.2020.552028

Kunstfeld, R., Lechleitner, S., Wolff, K., and Petzelbauer, P. (1998). MCP-1 and MIP-1alpha are most efficient in recruiting T cells into the skin in vivo. *J. Invest. Dermatol.* 111, 1040–1044. doi: 10.1046/j.1523-1747.1998.00410.x

Läubli, H., and Borsig, L. (2019). Altered cell adhesion and glycosylation promote cancer immune suppression and metastasis. *Front. Immunol.* 10:2120. doi: 10.3389/fimmu.2019.02120

Lin, X., Chai, G., Wu, Y., Li, J., Chen, F., Liu, J., et al. (2019). RNA m(6)A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. *Nat. Commun.* 10:2065. doi: 10.1038/s41467-019-09865-9

Login, F. H., Jensen, H. H., Pedersen, G. A., Koffman, J. S., Kwon, T. H., Parsons, M., et al. (2019). Aquaporins differentially regulate cell-cell adhesion in MDCK cells. *FASEB J.* 33, 6980–6994. doi: 10.1096/fj.201802068RR

Marotta, G., Basagni, F., Rosini, M., and Minarini, A. (2021). Role of fyn kinase inhibitors in switching neuroinflammatory pathways. *Curr. Med. Chem.* 29, 4738–4755. doi: 10.2174/0929867329666211221153719

McDole, B., Isgor, C., Pare, C., and Guthrie, K. (2015). BDNF over-expression increases olfactory bulb granule cell dendritic spine density in vivo. *Neuroscience* 304, 146–160. doi: 10.1016/j.neuroscience.2015.07.056

Montfort, A., Pearce, O., Maniati, E., Vincent, B. G., Bixby, L., Böhm, S., et al. (2017). A strong B-cell response is part of the immune landscape in human high-grade serous ovarian metastases. *Clin. Cancer Res.* 23, 250–262. doi: 10.1158/1078-0432.Ccr-16-0081

Morein, D., Erlichman, N., and Ben-Baruch, A. (2020). Beyond cell motility: The expanding roles of chemokines and their receptors in malignancy. *Front. Immunol.* 11:952. doi: 10.3389/fimmu.2020.00952

Nakamura, K., Williams, I. R., and Kupper, T. S. (1995). Keratinocyte-derived monocyte chemoattractant protein 1 (MCP-1): Analysis in a transgenic model demonstrates MCP-1 can recruit dendritic and Langerhans cells to skin. *J. Invest. Dermatol.* 105, 635–643. doi: 10.1111/1523-1747.ep12324061

Oldak, M., Smola-Hess, S., and Maksym, R. (2006). Integrin beta4, keratinocytes and papillomavirus infection. *Int. J. Mol. Med.* 17, 195–202.

Paolillo, M., and Schinelli, S. (2019). Extracellular matrix alterations in metastatic processes. *Int. J. Mol. Sci.* 20:4947. doi: 10.3390/ijms20194947

Pastushenko, I., Mauri, F., Song, Y., de Cock, F., Meeusen, B., Swedlund, B., et al. (2021). Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature* 589, 448–455. doi: 10.1038/s41586-020-03046-1

Peled, A., Grabovsky, V., Habler, L., Sandbank, J., Arenzana-Seisdedos, F., Petit, I., et al. (1999). The chemokine SDF-1 stimulates integrin-mediated arrest of CD34(+) cells on vascular endothelium under shear flow. *J. Clin. Invest.* 104, 1199–1211. doi: 10.1172/jci7615

Scapin, C., Ferri, C., Pettinato, E., Zambroni, D., Bianchi, F., Del Carro, U., et al. (2019). Enhanced axonal neuregulin-1 type-III signaling ameliorates neurophysiology and hypomyelination in a Charcot-Marie-Tooth type 1B mouse model. *Hum. Mol. Genet.* 28, 992–1006. doi: 10.1093/hmg/ddy411

Schiffman, M., Castle, P. E., Jeronimo, J., Rodriguez, A. C., and Wacholder, S. (2007). Human papillomavirus and cervical cancer. *Lancet* 370, 890–907. doi: 10.1016/s0140-6736(07)61416-0

Schimmel, L., and Gordon, E. (2018). The precise molecular signals that control endothelial cell-cell adhesion within the vessel wall. *Biochem. Soc. Trans.* 46, 1673–1680. doi: 10.1042/bst20180377

Sumigray, K. D., and Lechler, T. (2015). Cell adhesion in epidermal development and barrier formation. *Curr. Top. Dev. Biol.* 112, 383–414. doi: 10.1016/bs.ctdb.2014.11.027

Takeuchi, A., and Saito, T. (2017). CD4 CTL, a Cytotoxic Subset of CD4(+) T Cells, Their differentiation and function. *Front. Immunol.* 8:194. doi: 10.3389/fimmu.2017.00194

Tao, R., Huang, F., Lin, K., Lin, S. W., Wei, D. E., and Luo, D. S. (2021). Using RNA-Seq to explore the hub genes in the trigeminal root entry zone of rats by compression injury. *Pain Physician* 24, E573–E581.

Te Riet, J., Helenius, J., Strohmeyer, N., Cambi, A., Figdor, C. G., and Müller, D. J. (2014). Dynamic coupling of ALCAM to the actin cortex strengthens cell adhesion to CD6. *J. Cell Sci.* 127(Pt 7), 1595–1606. doi: 10.1242/jcs.141077

Teicher, B. A., and Fricker, S. P. (2010). CXCL12 (SDF-1)/CXCR4 pathway in cancer. *Clin. Cancer Res.* 16, 2927–2931. doi: 10.1158/1078-0432.Ccr-09-2329

von Lersner, A., Droesen, L., and Zijlstra, A. (2019). Modulation of cell adhesion and migration through regulation of the immunoglobulin superfamily member ALCAM/CD166. *Clin. Exp. Metastasis* 36, 87–95. doi: 10.1007/s10585-019-09957-2

Wang, F., Tao, R., Zhao, L., Hao, X. H., Zou, Y., Lin, Q., et al. (2021). Differential lncRNA/mRNA Expression profiling and functional network analyses in Bmp2 deletion of mouse dental papilla cells. *Front. Genet.* 12:702540. doi: 10.3389/fgene.2021.702540

Wang, H., Wang, Z., Li, Y., Lu, T., and Hu, G. (2020). Silencing snail reverses epithelial-mesenchymal transition and increases radiosensitivity in hypopharyngeal carcinoma. *Onco Targets Ther.* 13, 497–511. doi: 10.2147/ott.S237410

Whiteside, M. A., Siegel, E. M., and Unger, E. R. (2008). Human papillomavirus and molecular considerations for cancer risk. *Cancer* 113(10 Suppl), 2981–2994. doi: 10.1002/cncr.23750

Yakovlev, S., Cao, C., Galisteo, R., Zhang, L., Strickland, D. K., and Medved, L. (2019). Fibrin-VLDL Receptor-dependent pathway promotes leukocyte transmigration by inhibiting Src kinase fyn and is a target for fibrin β15-42 peptide. *Thromb. Haemost.* 119, 1816–1826. doi: 10.1055/s-0039-1695008

Yan, S. L., Hwang, I. Y., Kamenyeva, O., Kabat, J., Kim, J. S., Park, C., et al. (2021). Unrestrained Gα(i2) Signaling disrupts neutrophil trafficking, aging, and clearance. *Front. Immunol.* 12:679856. doi: 10.3389/fimmu.2021.679856

Yu, J., Zhou, Z., Wei, Z., Wu, J., OuYang, J., Huang, W., et al. (2020). FYN promotes gastric cancer metastasis by activating STAT3-mediated epithelial-mesenchymal transition. *Transl. Oncol.* 13:100841. doi: 10.1016/j.tranon.2020.100841

Zhang, M., Yang, W., Wang, P., Deng, Y., Dong, Y. T., Liu, F. F., et al. (2020). CCL7 recruits cDC1 to promote antitumor immunity and facilitate checkpoint immunotherapy to non-small cell lung cancer. *Nat. Commun.* 11:6119. doi: 10.1038/s41467-020-19973-6

Zheng, F., Zhang, Z., Flamini, V., Jiang, W. G., and Cui, Y. (2017). The axis of CXCR4/SDF-1 Plays a role in colon cancer cell adhesion through regulation of the AKT and IGF1R signalling pathways. *Anticancer Res.* 37, 4361–4369. doi: 10.21873/anticanres.11830

Zhu, S., Li, Y., Bennett, S., Chen, J., Weng, I. Z., Huang, L., et al. (2020). The role of glial cell line-derived neurotrophic factor family member artemin in neurological disorders and cancers. *Cell Prolif.* 53:e12860. doi: 10.1111/cpr.12860

Zou, Y. R., Kottmann, A. H., Kuroda, M., Taniuchi, I., and Littman, D. R. (1998). Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature* 393, 595–599. doi: 10.1038/31269

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fmicb.2022.985561

OPEN ACCESS

EDITED BY
Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands

REVIEWED BY
Shashanka Prasad,
JSS Academy of Higher Education and Research, India
Abhik Saha,
Presidency University, India

*CORRESPONDENCE
Chunyan Wang
lnzlgyn@163.com
Haozhe Piao
pzpy@163.com

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 04 July 2022
ACCEPTED 22 August 2022
PUBLISHED 15 September 2022

CITATION
Yang D, Zhang J, Cui X, Ma J, Wang C and Piao H (2022) Status and epidemiological characteristics of high-risk human papillomavirus infection in multiple centers in Shenyang.
Front. Microbiol. 13:985561.
doi: 10.3389/fmicb.2022.985561

Status and epidemiological characteristics of high-risk human papillomavirus infection in multiple centers in Shenyang

Di Yang[1], Jing Zhang[1], Xiaoli Cui[1], Jian Ma[2], Chunyan Wang[1]* and Haozhe Piao[3]*

[1]Department of Gynecology, Liaoning Cancer Hospital and Institute, Cancer Hospital of China Medical University, Shenyang, China, [2]Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China, [3]Department of Neurosurgery, Liaoning Cancer Hospital and Institute, Cancer Hospital of China Medical University, Shenyang, China

The different human papillomavirus (HPV) strains cause warts in various regions of the body. However, considering that the status and genotype distribution of HPV infection in women in Shenyang remain unknown, herein, we investigated the epidemiological characteristics of high-risk HPV (HR-HPV) infection in women in Shenyang, as well as the current state of HPV infection in Shenyang, to provide a theoretical basis for the prevention and treatment of cervical cancer. From December 2018 to December 2021, 6,432 urban and rural women from the Liaoning Cancer Hospital and the Sujiatun Women and Infants' Hospital were assessed via the Thinprep cytology test (TCT) and HR-HPV detection. Of the 5,961 women enrolled, 739 were HPV positive (12.40%) and 562 were TCT positive (9.43%). Statistical analyses identified the following HPV risk factors: high school education or lower [OR = 1.426 (1.199–1.696), $p < 0.001$], age at first sexual encounter $\leq$ 19 years [OR = 1.496 (1.008–2.220), $p = 0.046$], and number of sexual partners $>$ 1 [OR = 1.382 (1.081–1.768), $p = 0.010$], atypical squamous cells of undetermined significance (ASCUS) and above [OR = 10.788 (8.912–13.060), $p < 0.001$], non-condom-based contraception [OR = 1.437 (1.103–1.871), $p = 0.007$], nationalities other than Han [OR = 1.690 (1.187–2.406), $p = 0.004$], rural residence [OR = 1.210 (1.031–1.419), $p = 0.020$]. Compared to the HPV infection rate of women aged 56–65, that in women aged 35–45 [OR = 0.687 (0.549–0.860), $p = 0.001$] and 46–55 [OR = 0.740 (0.622–0.879), $p = 0.001$] decreased significantly. To conclude, risk factors of HPV infection among female patients include high school age and below, initial sexual encounter at age $\leq$ 19 years, number of sexual partners $>$ 1, ASCUS and above, non-condom contraception, nationalities other than Han nationality and rural population. Collectively, this study provides insights for the improved prevention and treatment of cervical cancer.

KEYWORDS

screening, human papillomavirus, high risk type, Shenyang area, epidemiological characteristics

Introduction

In women, cervical cancer ranks fourth among all cancers in incidence (13.1%) and mortality (6.9%), only after breast cancer, colorectal cancer, and lung cancer (Arbyn et al., 2020). In 2020, approximately 604,127 new cases were diagnosed worldwide, while 341,831 people died of cervical cancer (Sung et al., 2021). A total of 85% of cases occur in developing countries, one third of which occur in China and India. In developing countries, cervical cancer ranks second in incidence rate among female tumors, and first in mortality (Bray et al., 2018). Moreover, an increasing trend in incidence among younger individuals has been noted, thus posing a serious risk to women's health (Bruni et al., 2019). Indeed, within 2020, 109,741 new cervical cancer cases and 59,060 cervical cancer-related deaths were reported in China (Zheng et al., 2019; Goodman et al., 2020).

Human papillomavirus (HPV), a double-stranded circular DNA virus, is a papillomavirus belonging to the family Papillomaviridae. To date, more than 100 HPV genotypes have been characterized, forty of which are associated with human reproductive tract infections via uncontrolled induction of squamous epithelia proliferation within the mucosa (Ardekani et al., 2022). HPV represents a common sexually transmitted infection worldwide. Approximately 75% of sexually active people will experience HPV infection in their lifetime (Trottier and Franco, 2006). Meanwhile, the main risk factor for cervical cancer and its precancerous lesions is persistent infection with high-risk HPV (HR-HPV) (de Sanjosé et al., 2018). Approximately 90% of cervical cancers contain DNA sequences of specific HPV genotypes with most cases caused by HR-HPV infection transmitted through sexual intercourse. Therefore, primary prevention (HPV vaccination) and secondary prevention (screening and treatment of cervical precancerous lesions) can effectively prevent cervical cancer (Arbyn et al., 2020). Therefore, the prevention and treatment of HR-HPV infection is an important way of preventing female cervical cancer.

Significant differences have been reported in HPV infection rates and types among people in different regions (Duan et al., 2022). For example, the positivity rate, as well as the distribution and characteristics of the HPV types, differ in women of different ages and regions. In particular, the HPV infection rate in women in China is 15.5%, which is lower than that in African countries (20.9–23.4%) and higher than that in other Asian countries (5.5–7.5%), Europe (7.8–8.4%), and North America (12.4–13.5%). Moreover, within China, differences in the HPV infection rate and subtype distribution have been reported among different regions and ethnic groups (Choi et al., 2018). For example, the infection rate is 33.05% in Hebei, 28.43% in Guangzhou, 27.09% in Fujian, 24.31% in Guizhou, and 16.64% in Shanghai (Solomon et al., 2002).

From a clinical point of view, epidemiological studies investigating HPV infection among women in different regions have been conducted to help reduce the prevalence of infection among women. However, to date, no study has reported on the status and genotype distribution of HPV infection in women in Shenyang. Therefore, the current cross-sectional study examines the prevalence and high-risk factors of HPV infection within women in Shenyang. Moreover, the epidemiological characteristics of HR-HPV in women in this region are assessed, and a novel strategy for the prevention and treatment of cervical cancer is proposed.

Materials and methods

Ethics statement

This study was approved by the Ethics Committee of the Liaoning Cancer Hospital (ethics batch number: 20180106), and informed consent was obtained from all participants included in this study.

Research participants

From April 2018 through December 2021, 6,432 women in Liaoning were screened for cervical cancer at the Liaoning Cancer Hospital and the Shenyang Sujiatun District Women's and Infant Hospital. The women selected for this study had an average age of 51.56 ± 6.621 years and a median age of 51 years. Community population and hospital outpatient opportunistic screenings were employed to evaluate whether participants met the following inclusion criteria: (1) the screening participants were selected by the community (village committee) and did not represent a single occupational group; (2) women from urban and rural populations with high incidence rates and mortalities and good population compliance were selected as the target population; (3) only participants with a registered residence in their regions (living in the local area for more than 3 years) were screened; (4) women with a history of sexual life, aged between 35 and 65 years old (subject to the date of birth on the ID card); and (5) women with no serious organ dysfunction or mental disease, who voluntarily participate in and can accept the questionnaire survey. Exclusion criteria for screening objects included the following: (1) Women who have been diagnosed with a tumor; (2) women who are currently being treated for other serious internal and external diseases; (3) women who have had a hysterectomy; and (4) women who are pregnant and lactating.

Methods

The Thinprep cytology test (TCT) was combined with HPV detection technology to screen cervical lesions. Each medical

examiner took the stone cutting position. The vulva was fully disinfected and the cervix exposed with a vaginal dilator. Next, the cervix was wiped with a sterile cotton swab to remove excess cervical secretions. Finally, a disposable cervical cell sampling brush was rotated ten times within the scaly columnar junction of the cervix, removed, and stored in a cell preservation solution.

Liquid-based cytology

A TCT (Thinprep cytologic test purchased from Hologic.lnc 20160621) test was used to obtain materials for production. The accompanying report was prepared based on The Bethesda System (TBS) for Reporting Cervical Cytology (2014). Abnormal results included negative for intraepithelial lesion or malignant tumor (NILM), atypical squamous cells of undetermined significance (ASCUS), atypical squamous cells cannot exclude high-grade squamous intraepithelial lesion (ASCH), low-grade squamous intraepithelial lesion (LSIL), high grade squamous intraepithelial lesion (HSIL), squamous cell carcinoma (SCC), and atypical glandular cells (AGC). A positive TCT diagnosis included patients with ASCUS, ASC-H, LSIL, HSIL, SCC, or AGC.

Human papillomavirus detection technology

The E6/E7 mRNA detection kit (Hologic Gen-Probe, San Diego, CA 20183401863)—an FDA-approved assay for detecting HPV—was used to detect 14 HR-HPV-mRNA causing cervical cancers (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68). The results for HPV type 16 and 18 detection can be made available to provide clinical guidance. The kit detects E6 and E7, rather than L1, mRNAs of the HR-HPV virus to avoid the missed diagnosis of high-level lesions and cancer. Moreover, the test avoids cross-reactivity with low-risk HPV types, has fewer false positives, and avoids unnecessary colposcopy and overdiagnosis.

Standard referral colposcopy

Cervical exfoliative cells were examined using cervical liquid-based TCT. The diagnostic criteria are based on the TBS classification as outlined above (Kclurman et al., 2014). Colposcopies were also performed in patients that were HPV positive and/or assigned a TBS classification of ASCUS or above, or found to have clinically suspicious abnormalities. If the results of a multi-point tissue biopsy for cervical lesions were suspicious, the bite tissues were sent for pathological examinations. The pathological results were of the gold standard. Diagnostic criteria included normal or inflammatory reactions, cervical intraepithelial neoplasia (CIN), and cervical cancer. Recently, CIN has gained secondary classifications based on the severity of the disease and can be classified on a scale from one to three (CINI-CINIII) (Liang et al., 2016).

Technical quality control

For quality control purposes, 20% of the positive cervical exfoliative cell smears and 5–10% of the negative smears were randomly selected. All selected smears were rechecked by experts. The qualified rate of smear results reached 80%. Quality control of colposcopy samples entailed a spot check for 10% of the normal reports and 20% of abnormal reports, as well as subsequent analysis by experts. The standard rate of the report results reached 90%. Quality control of histopathological examination entailed a spot check of 10% of the pathological sections, subsequently rechecked by experts. The coincidence rate of the diagnostic results reached 90%.

Investigation of risk factors and awareness of high-risk-human papillomavirus

Analysis was performed on the basic population information, as well as the living habits, physiological indicators, psychological and emotional conditions, screening willingness, and other high-risk factors. The questionnaire posed question regarding demographic characteristics and risk factors that may be associated with HR-HPV infections. Additionally, participants answered questions meant to reflect their level of knowledge about, and overall attitudes toward, cervical cancer, HPV, and the correlation between the two. All questions were answered directly by each participant or by dictating answers to a study nurse.

Statistical analysis

Statistical analysis of the relevant data was conducted using the statistical package for the social science (SPSS) version 19.0 software. The count data rate (%) was analyzed using the chi-square (χ^2) test. The influencing factors were analyzed using univariate and multivariate logistic regression to evaluate the correlation between the relevant factors mentioned in the questionnaire and HPV infection.

Results

Demographic characteristics

Basic population information

Of the 6,432 participants, 262 were 35–40 years old (4.40%), 941 were 41–45 years old (15.79%), 1,581 were 46–50 years old

TABLE 1 Basic information of the cervical cancer screening population.

Characteristics		N	n (%)
Age (years)	35–40	262	4.40
	41–45	941	15.79
	46–50	1,581	26.52
	51–55	1,443	24.21
	56–60	1,115	18.70
	61–65	619	10.38
Ethnicity	Han	5,058	84.85
	Others	179	3.00
Marital status	Unmarried	146	2.45
	Married	5,578	93.57
	Divorced	168	2.82
	Widowed	68	1.14
Profession	Head of party and enterprise unit	423	7.10
	Professional skilled worker	1,064	17.85
	Office and related personnel	466	7.82
	Social production and life service personnel	974	16.34
	Agriculture, forestry, animal husbandry, fishery production, and auxiliary personnel	158	2.65
	Production and related personnel	408	6.84
	Soldiers	4	0.07
	Others who are difficult to classify	1,984	33.28
	Others	479	8.04
Educational level	Junior high school and below	1,990	33.38
	Senior high school	2,009	33.70
	College degree or above	1,961	32.90
Total		5,961	100.00

(26.52%), 1,443 were 51–55 years old (24.21%), 1,115 were 56–60 years old (18.70%), and 619 were 61–65 years old (10.38%; Table 1).

Personal and family history

Menarche age occurred between 12 and 18 years old in 97.94% of the participants. Menopause had occurred in 52.86% of the participants, most of whom were over 50 years of age. Moreover, 83.14% of the participants had a history of breast-feeding, and 71.26% breast-fed for more than 6 months. The population with multiple sexual partners accounted for 8.92 and 3.00% of the participants were aged < 19 years old at the time of their first sexual encounter. A total of 71.25% had a history of abortion; those with lover's redundant prepuce and sexual bleeding accounted for 5.54 and 6.46%, respectively. A total of 22.18 and 17.55% women had a history of gynecological diseases and family histories of tumors, respectively (Table 2).

Living habits

Current smokers accounted for 8.37% of all participants, while those who had quit smoking accounted for 1.21%. Passive smokers accounted for 62.44 and 82.17% of them were exposed to cooking oil smoke almost every day. Additionally, 11.06% of the participants had a history of alcohol consumption, and 1.02% reported current alcohol consumption. Moreover, 85.82% did not often participate in outdoor physical exercise, and 92.70% did not drink tea. More than 50% of the women had an insufficient intake of fresh vegetables and fruits with vegetable intake < 5 kg per week, and fruit intake < 2.5 kg per week. Approximately 70% of women did not meet the dietary guideline requirements set forth for Chinese residents (Table 3).

TABLE 2 Personal and family history of the cervical cancer screening population.

Characteristics		N	n (%)
Age at menarche (years)	<12	68	1.14
	12–18	5,838	97.94
	>18	55	0.92
Menopausal	Yes	3,151	52.86
	No	2,810	47.14
Age at menopause	<50	966	16.21
	≥50	2,185	36.65
Breastfeeding history	Yes	4,956	83.14
	No	1,005	16.86
Breastfeeding time	≤6 Months	708	11.88
	>6 Months	4,248	71.26
Sexual partners	0	11	0.18
	1	5,364	89.98
	≥2	532	8.92
Age at first sexual activity	Never	2	0.03
	≤19	179	3.00
	20–30	5,656	94.88
	≥31	68	1.14
Pregnancy history	Yes	5,810	97.47
	No	75	1.26
History of miscarriage	Yes	4,247	71.25
	No	1,571	26.35
Sexual partner's foreskin is too long	Yes	323	5.42
	No	5,567	93.39
Bleeding during intercourse	Yes	385	6.46
	No	5,499	92.25
Cervical cancer vaccine	Yes	31	0.52
	No	5,858	98.27
Abnormal vaginal discharge	Yes	679	11.39
	No	5,146	86.33
History of gynecological disease	Yes	1,322	22.18
Family history of cancer	Yes	1,046	17.55

TABLE 3 Living conditions of the cervical cancer screening population.

Characteristics		N	n (%)
Smoking	No	5,390	90.42
	Currently smoking	499	8.37
	Smoked before	72	1.21
Second hand smoke	Yes	3,722	62.44
	No	2,239	37.56
Cooking fumes	Almost everyday	4,898	82.17
	Sometimes	951	15.95
	Almost not	112	1.88
Alcohol consumption	No	6,241	104.70
	Currently drinking alcohol	61	1.02
	Previously drank alcohol	659	11.06
Exercise	Yes	845	14.18
	No	5,116	85.82
Tea drinking	Yes	435	7.30
	No	5,526	92.70
Vegetable consumption	Never	212	3.56
	<5 Pounds/week	3,407	57.15
	≥5 Pounds/week	2,317	38.87
Fruit consumption	Never	209	3.51
	<2.5 Pounds/week	3,462	58.08
	≥2.5 Pounds/week	2,274	38.15
Livestock meat consumption	Never	258	4.33
	≤350 g/Week	4,578	76.80
	>350 g/Week	1,109	18.60
Coarse grain consumption	Never	617	10.35
	<1 Pound/week	4,731	79.37
	≥1 Pound/week	591	9.91

Health history

Of the participants, 60.69% had a good health status or self-rated their health status as good. An additional 9.21% had a history of hypertension, 3.62% of diabetes, 11.84% of hyperlipidemia, and 0.22% of mental health illness. Moreover, 20.33% of the participants experienced negative life events, 32.52% had depression or anxiety symptoms, and 35.31% had poor sleep quality (Table 4).

Population screening willingness

A total of 22.38% of the participants believed they could easily develop cancer, while 31.74% had been previously screened for cancer. Cancer screening for 27.38% of the participants was paid for by the government. Moreover, 28.07% indicated that they would fully accept cancer screening. Those who did not accept cancer screening listed time concerns and no conscious symptoms.

In the case of abnormal inspection results, most women were willing to be re-inspected and felt they were able to cover the associated expenses. However, the acceptance rate for self-paid expenses was typically below 200 Yuan. Moreover, 92.35% of the participants indicated that they were willing to accept a follow-up visit in the case of abnormal results. The primary reason preventing participants from accepting subsequent appointments was the time requirement.

Additionally, 91.01% of the participants indicated that they were willing to try new, more effective screening technologies, with acceptable out-of-pocket expenses for these technologies concentrated below 200 Yuan. The participants who were unwilling to accept new, more effective screening technologies,

TABLE 4 Health-related emotional factors of the cervical cancer screening population.

Characteristics		N	n (%)
Self-assessed health status	Very good or good	3,618	60.69
	General	1,947	32.66
	Not good	396	6.64
Hypertension	Yes	549	9.21
	No	5,412	90.79
Diabetes	Yes	216	3.62
	No	5,745	96.38
Hyperlipidemia	Yes	706	11.84
	No	5,255	88.16
Diagnosed with a mental illness	Yes	13	0.22
	No	5,948	99.78
Experienced a negative life event	No	4,749	79.67
	1–2 Events	1,160	19.46
	≥3 Events	52	0.87
Mental depression	No	4,111	68.96
	Occasionally	1,461	24.51
	>1 Month	186	3.12
	>6 Months	203	3.41
Anxiety	No	4,022	67.47
	Occasionally	1,633	27.39
	>1 Month	191	3.20
	>6 Months	115	1.93
Sleep quality	Good	3,856	64.69
	Difficult to fall asleep	524	8.79
	Wake up early	535	8.98
	Sleep well	960	16.10
	Wake up at night	86	1.44
People offering support in difficult times	Husband	5,371	90.10
	Parents	3,255	54.60
	Children	3,954	66.33
	Siblings	3,300	55.36
	Friends	3,100	52.00
	Colleagues	1,203	20.18
	No support	68	1.14

TABLE 5 Screening willingness of the cervical cancer screening population.

Characteristics		N	n (%)
Do you think you are prone to cancer?	Yes	1,334	22.38
	No	4,532	76.03
Have you ever been screened for cancer?	Yes	1,892	31.74
	No	3,987	66.88
Who bears the cost of cancer screening?	It is all borne by the government and not paid by individuals	1,632	27.38
	Some expenses shall be borne by individuals	199	3.34
	All expenses shall be borne by individuals	40	0.67
	No idea	42	0.70
To what extent do you accept cancer screening?	Totally acceptable	1,673	28.07
	Acceptable	218	3.66
	Difficulty in accepting	9	0.15
	Unacceptable	0	0.00
Reasons for not participating in cancer screening	Economic reasons	1,556	26.10
	Time reason	2,776	46.57
	The procedure is cumbersome and laborious	1,350	22.65
	Examination can cause pain	1,719	28.84
	I do not think there are any symptoms in my body. It's unnecessary	1,787	29.98
	Physical condition does not allow	31	0.52
	Unaccompanied	52	0.87
If the examination result is abnormal, would you be willing to be checked again?	Yes	5,530	92.77
	No	380	6.37
Would you like to be checked again?	Yes	5,166	86.66
	No	276	4.63
What is an acceptable examination fee for you?	<100 Yuan	537	9.01
	100–199 Yuan	2,470	41.44
	200–299 Yuan	985	16.52
	≥300 Yuan	1,165	19.54
Are you willing to make a return visit?	Yes	5,505	92.35
	No	373	6.26
Reasons for not willing to make a return visit/recheck.	Economic reasons	619	10.38
	Time reason	887	14.88
	The inspection is cumbersome and laborious	582	9.76
	Examination can cause pain	208	3.49
	I do not think there are any symptoms in my body. It is unnecessary	149	2.50
	Physical condition does not allow	17	0.29
	Unaccompanied	7	0.12

(Continued)

TABLE 5 (Continued)

Characteristics		N	n (%)
Are you willing to accept new technology?	Yes	5,425	91.01
	No	478	8.02
How much are you willing to pay for the new technology?	<100 Yuan	671	11.26
	100–199 Yuan	2,628	44.09
	200–299 Yuan	950	15.94
	≥300 Yuan	1,121	18.81
Reasons for reluctance to accept new technology screening.	Question the scientific validity and safety of the new method	1,145	19.21
	Unclear interpretation and utilization of screening results	697	11.69
	High cost	512	8.59
	The old method is reliable, there is no need to use the new method	112	1.88
	Concerned about the pain of new screening methods	4	0.07

primarily showed concern for the scientific validity and safety of the new methods (Table 5).

Gynecological examination

All participants (5,961) completed gynecological examinations for cervical cancer screenings (Table 6). The detection rates of leukoplakia, ulcers, condyloma, and tumors were 0.34, 0.02, 0.03, and 0.05%, respectively. The detection rates of congestion, condyloma, and tumors in vaginal examinations were 1.16, 0.02, and 0.02%, respectively. The rates of tofu-like, purulent, foam, blood, and peculiar smell were 0.12, 1.69, 0.82, 0.10, and 0.55%, respectively, whereas the rates of cervical hypertrophy, ulcers, erosion, atrophy, polyps, and cysts were 1.93, 0.12, 5.54, 10.97, 2.89, and 9.56%, respectively.

Human papillomavirus and thinprep cytology test distribution

Of the 6,432 participants, 471 were excluded from the survey results. The remaining 5,961 participants were screened for cervical cancer and included in the statistical analysis (Table 7). The average age of the remaining participants was 51.51 ± 6.61 years, with a median age of 51 years (Figure 1). Cytological examinations revealed 562 cases that were TCT positive; 9.43% (562/5961) were ASCUS +, of which 6.71% had ASCUS (not statistically significant), an additional 0.50% had ASC-H, 1.63% had LSIL, 0.27% had HSIL, and 0.22% had AGC-NOS. Additionally, AGC-N was reported in 0.10% of the participants, while SCC, cervical carcinoma *in situ*, and adenocarcinoma were not detected. Moreover, TCT-/HPV + accounted for 7.68% of the cases, TCT+/HPV- for 4.71%, and TCT + /HPV + for 4.71%.

Of the 5,961 participants, 739 (12.40%) tested positive for HPV (Table 8), of which 106 were HPV-16 positive (14.34, 1.78% positive rate), 46 were HPV-18/45 positive (6.22, 0.77% positive rate), and 589 were positive for other high-risk types of HPV genotypes (79.70; 9.88% positive rate). The positive rate of HPV-16 and HPV-18/45 in the total screening population was 2.52%. Two of the participants tested positive for both type 16 and type 18/45.

Age specificity of thinprep cytology test positive and high-risk-human papillomavirus infection

The 5,961 participants that were enrolled for TCT and HR-HPV screening from December 2018 to December 2021 were divided into six age groups, and the positive detection rates were determined for each. It was found that 35–40, 41–45, 46–50, 51–55, 56–60, and 61–65 year old participants had TCT positive detection rates of 9.92, 10.51, 10.94, 9.36, 7.71, and 6.95%, respectively (Table 9). The prevalence of HR-HPV infection reached two peaks at the ages of 56–60 and 61–65, with infection rates of 14.71 and 15.51%, respectively (Table 10).

Single-factor analysis of risk factors for human papillomavirus infection

A total of 28 potential risk factors may be related to HR- HPV infection. Univariate logistic analysis showed that HPV infection was correlated with different age groups, educational level, age at first sexual encounter, number of sexual

TABLE 6 Gynecological examination of the cervical cancer screening population.

Characteristics		N	n (%)
Vulva	Normal	5,810	97.47
	Vitiligo	20	0.34
	Ulcer	1	0.02
	Condyloma	2	0.03
	Tumor	3	0.05
	Other	32	0.54
Vaginal	Normal	5,773	96.85
	Congestion	69	1.16
	Condyloma	1	0.02
	Tumor	1	0.02
	Other	24	0.40
Secretion	Normal	5,475	91.85
	Tofu-like	7	0.12
	Purulent	101	1.69
	Foam	49	0.82
	Bloody	6	0.10
	Peculiar smell	33	0.55
	Excessive	157	2.63
	Other	40	0.67
Cervix	Normal	3,646	61.16
	Hypertrophy	115	1.93
	Congestion	48	0.81
	Touch blood	24	0.40
	Ulcer	7	0.12
	Erosion	390	6.54
	Shrink	654	10.97
	Polyp	172	2.89
	Cervical cyst	570	9.56
	IUD tail wire	24	0.40
	Other	217	3.64
Total		5,961	100.00

partners, contraceptive methods, TCT positivity, nationality, and population ($p < 0.05$; **Table 11**).

Logistic regression analysis of risk factors of human papillomavirus infection

The results of the multivariate unconditional logistic regression analysis showed that high school education or lower [odds ratio (OR) = 1.426 (1.199–1.696), $p < 0.001$], age at first sexual encounter $\leq$ 19 years [OR = 1.496 (1.008–2.220), $p = 0.046$], number of sexual partners > 1 [OR = 1.382 (1.081–1.768), $p = 0.010$], assigned TBS classification of ASCUS or above [OR = 10.788 (8.912–13.060), $p < 0.001$], non-condom-based contraception [OR = 1.437 (1.103–1.871), $p = 0.007$], other nationalities except Han [OR = 1.690 (1.187–2.406), $p = 0.004$], and rural residence [OR = 1.210 (1.031–1.419), $p = 0.020$] were risk factors for HPV infection (**Table 12**). Compared with women aged 56–65, the infection rates in women aged 35–45 [OR = 0.687 (0.549–0.860), $p = 0.001$] and 46–55 [OR = 0.740 (0.622–0.8709), $p = 0.001$] were significantly reduced.

TABLE 7 Thinprep cytology test (TCT) and HPV cervical cancer screening results.

TBS classification diagnostic criteria	N	HPV (+)	HPV (–)	n (%)
NILM	5,399	458	4,941	7.68
ASC-UC	400	154	246	2.58
ASC-H	30	28	2	0.47
LSIL	97	79	18	1.33
HSIL	16	15	1	0.25
SCC	0	0	0	0.00
AGC-NOS	13	3	10	0.05
AGC-N	6	2	4	0.03
AIS	0	0	0	–
Adenocarcinoma	0	0	0	–
Total	5,961	739	5,222	–

TBS, the Bethesda system; NILM, negative for intraepithelial lesion or malignancy; ASCUC, atypical squamous cells of undetermined significance; ASC-H, atypical squamous cells cannot exclude high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; SCC, squamous cell carcinoma; AGC-NOS, atypical glandular epithelial cells-not otherwise specified; AGC-N, atypical glandular epithelial cells-not prone to cancer; AIS, adenocarcinoma in situ.

Discussion

Currently, cervical cancer is the second most common female malignant tumor after breast cancer. The incidence rate of cervical cancer has increased significantly and has become increasingly prevalent in younger women. As such, establishing effective preventative and therapeutic strategies for cervical cancer has become increasing important. Various studies (Sui et al., 2018) have reported that HPV infection represents the primary biological cause of cervical precancerous lesions and cervical cancer. The pathological process from HPV infection, to cervical precancerous lesion development, to carcinoma, to invasive carcinoma *in vivo*, typically occurs over a span of approximately 10 years. Hence, interrupting the early stages of HPV infection, or early stages of cervical lesion formation may effectively prevent, or delay, the occurrence and development of cervical lesions and cervical cancer.

A recent study (Schettino et al., 2008) found that the positive rate of HPV infection was closely related to geographical region, population, age, and living habits, among other factors. In particular, HPV infection was found to have a strong regional characteristic, as the positivity rate and distribution of HPV type varies in different countries or regions around the world.

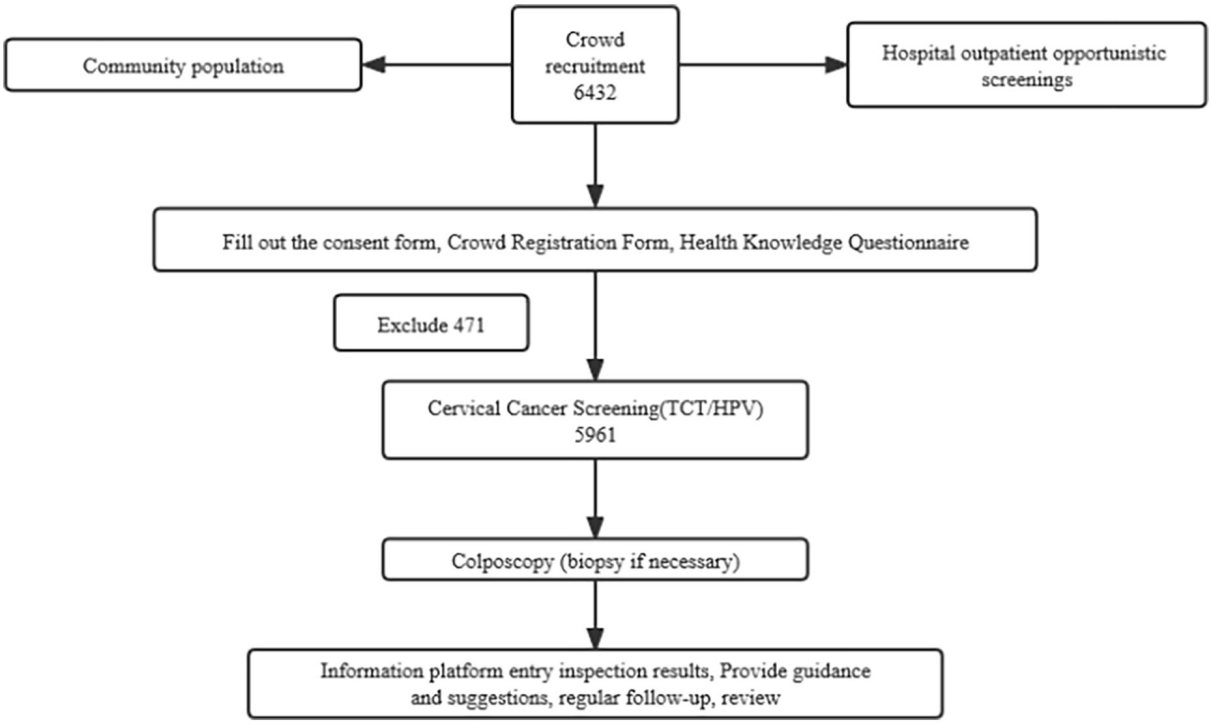

FIGURE 1
The CONSORT flow of study.

That is, regions with relatively developed economies exhibit lower positivity rates. For instance, a study in South Korea, comprising 18,170 women had a HR-HPV positive rate of 12.5% (Ouh et al., 2018). Similarly, a study in Turkey showed that the incidence rate of HR-HPV infection was ∼9.17% (Taskin et al., 2022). Meanwhile, in countries with relatively underdeveloped economies, such as sub-Sahara Africa and Bangladesh, the HR-HPV infection rate was 34% (Seyoum et al., 2022) and ∼41.86% (Sharmin et al., 2021), respectively. Moreover, although the total positivity rate of HPV infection in China is ∼21.69%, it varies between economically developed and underdeveloped areas. For instance, a study in Shanghai included 6,619 women, of whom the HR-HPV infection rate was 9.5% (Wang et al., 2021), whereas another study in Jiangsu included 36,500 women and reported a 28.95% HR-HPV infection rate (Niu et al., 2022). Still further, a study in Guangxi from 2016 to 2021 included 41,140 women, for whom the overall HR-HPV infection rate was 18.10% (Wei et al., 2022). In the current study, 5,961 cases of HPV were detected in urban and rural women in Shenyang, of which 739 were HPV positive, accounting for 12.40% of the women (739/5961). Significant regional differences were observed in HPV infection rates. These epidemiological studies on HR-HPV distribution and prevalence in different geographical regions can aid the development of strategies to prevent HPV-related cancer, especially with regard to vaccines (Smith et al., 2008; Wang et al., 2019). Early diagnosis, the use of preventive HPV vaccines, and screening tests are critical for the prevention of precancerous lesions and cervical cancer (Alotaibi et al., 2020).

TABLE 8 Human papillomavirus (HPV) screening results in cervical cancer.

HPV		N	n (%)
Detection	(+)	739	12.40
Condition	(–)	5,222	87.60
Typing	16 type	106	1.78
	18/45 type	46	0.77
	Other	589	9.88
Total		5,961	100.00

Age differences were also observed with regard to HPV infection rate, in the current study. Previous studies have reported that HPV infection is more common in elderly women (Hariri et al., 2011). Here, we examined the prevalence of HPV in six age groups and found the highest prevalence in women aged 61–65, at a rate of 15.51%. Similarly, women aged 55–60 showed a positive infection rate of 14.7%. The TCT positive rate was lower in the 51–55 and 56–65 age groups; however, their HR-HPV positivity rates were higher than those of the other age groups. These results may have been missed TCT diagnoses,

TABLE 9 Age specificity of TCT positivity.

TBS classification diagnostic criteria	35–40 (year)	41–45 (year)	46–50 (year)	51–55 (year)	56–60 (year)	61–65 (year)
NILM	236	842	1,408	1,308	1,029	576
ASC-UC	16	72	121	99	65	27
ASC-H	3	3	7	7	6	4
LSIL	6	17	31	23	13	7
HSIL	0	5	3	3	1	4
SCC	0	0	0	0	0	0
AGC-NOS	1	2	8	1	1	0
AGC-N	0	0	3	2	0	1
AIS	0	0	0	0	0	0
Adenocarcinoma	0	0	0	0	0	0
Total	262	942	1,581	1,443	1,115	619
ASCUS +	26	99	173	135	86	43
Positive n (%)	9.92%	10.51%	10.94%	9.36%	7.71%	6.95%

which can result from the inward movement of the cervical squamous column junction in perimenopausal women. The close relationship between HPV infection rate and age may be due to the development of clearance over time, with changes in sexual activity and immunity acquired from previous infections (Arbyn et al., 2011). The increased prevalence of HPV in the elderly can further be explained by the reactivation of potential infection (Giuliano et al., 2019). Therefore, a cervical HPV screening program is particularly important for perimenopausal women (Shi et al., 2017; Li et al., 2020). In women of all ages, the positive rate of HPV infection showed a "U" shaped change. The first peak of infection typically occurred during the early stages of their sexual life, which was likely related to factors that affect the cervical self-clearance ability. These factors include the age at first sexual encounter, immature cervical development, failure to practice safe sex, frequent sex, multiple sexual partners, and personal health problems. Meanwhile, the second peak occurred primarily during perimenopause, which may be related to a change in sex hormone levels in aging women, a decline in immune function, a decrease in cervical resistance, an increase in virus susceptibility, or the reactivation of latent viral infection. Therefore, it is important to continue promoting the benefits of safe sex, while also implementing public health education resources related to health issues, individual immunity, healthy habits, and the importance of regular health checks. Moreover, the use of endocrine therapy by perimenopausal women may reduce the incidence of HPV infection. Hence, more attention should be dedicated to screening for HPV infection in elderly women.

TABLE 10 Age specificity of human papillomavirus (HPV) infection.

Year	N	HPV(+)	HPV(−)	n (%)	χ^2	*P*
35–40	262	29	233	11.07	15.9	0.0071
41–45	941	101	840	10.73		
46–50	1,581	184	1,397	11.64		
51–55	1,443	165	1,278	11.43		
56–60	1,115	164	951	14.71		
61–65	619	96	523	15.51		

Previous studies have reported that first sexual encounters occurring before the age of 19 years represents an independent risk factor for HR-HPV infection. The cervical epithelial repair function and autoimmune function of women under 19 years of age are not yet fully mature, rendering them susceptible to HR-HPV (Liu et al., 2015). Moreover, puberty hormones can promote HPV infection, with the hormone concentration in women younger than 19 years being relatively high. Sexual intercourse serves as the primary route of HPV transmission, while having multiple sexual partners increases the risk of HPV infection in women. Other studies have shown an association between the number of sexual partners and the risk of cervical cancer (Kitamura et al., 2021), a finding that was also demonstrated by the international cooperation, comprising 21 epidemiological studies (International Collaboration of Epidemiological Studies of Cervical Cancer, 2009). Here, we also found an association between the number of sexual partners and the risk of HVP infection. Although studies have shown that an active sexual life is closely related to the occurrence and development of cervical cancer, condoms can block the destruction of cervical mucosa by pathogens, inhibit immune function, reduce the probability of infection, reduce the cervical mucosa stimulation by semen, and help reduce HPV transmission (Hariri and Warner, 2013).

Consistent with other studies (Niu et al., 2022), the current study shows that women with a high school education or lower, have a higher risk of HPV infection than women with a higher education level. Additionally, socioeconomic status is closely related to the incidence of cervical cancer, particularly in developing countries. A higher education level may improve

TABLE 11 Comparison of information and living habits of women with a positive HPV detection rate.

Characteristics		N	(+)	(−)	n (%)	χ^2	*P*
Age (years)	35–45	1,203	130	1,073	10.81	15.61	0.0004
	46–55	3,024	349	2,675	11.54		
	56–65	1,734	260	1,474	14.99		
Family history of cancer	Yes	987	139	848	14.08	3.095	0.0785
	No	4,974	600	4,374	12.06		
Level of education	High school and below	3,999	544	3,455	13.60	16.22	0.0001
	College degree or above	1,961	195	1,766	9.94		
Age at initial sexual experience	≤19 Years	179	31	148	17.32	4.04	0.0444
	>19 Years	5,723	703	5,020	12.28		
Number of sexual partners	Multiple (≥2)	532	85	447	15.98	6.704	0.0096
	1	5,375	650	4,725	12.09		
Number of abortions	>1	1,949	249	1,700	12.78	0.9141	0.3390
	1	2,279	314	1,965	13.78		
Number of deliveries	>1	544	78	466	14.34	1.518	0.2179
	≤1	5,124	640	4,484	12.49		
Number of marriages	>1	320	45	275	14.06	0.8146	0.3668
	≤1	5,587	690	4,897	12.35		
Contraceptive methods used	Others	4,508	582	3,926	12.91	7.283	0.0070
	Condom	727	68	659	9.35		
Sexual partner has long foreskin	Yes	323	40	283	12.38	0.001146	0.9728
	No	5,567	693	4,874	12.45		
Bleeding during intercourse	Yes	385	53	332	13.77	0.6647	0.4149
	No	5,499	679	4,820	12.35		
Leucorrhea abnormality	Yes	679	73	606	10.75	1.988	0.1585
	No	5,146	651	4,495	12.65		
Vaccination	No	759	535	224	70.49	–	> 0.999
	Yes	3	2	1	66.67		
Extramarital sex	Yes	63	10	53	15.87	0.6793	0.4098
	No	5,819	723	5,096	12.42		
Ethnicity	Han	5,749	699	5,050	12.16	8.661	0.033
	Others	211	40	171	18.96		
Total household income	≤50,000 Yuan	3,188	429	2,759	13.46	4.147	0.0417
	>50,000 Yuan	2,120	245	1,875	11.56		
Breastfed	No	1,005	120	885	11.94	0.2324	0.6297
	Yes	4,956	619	4,337	12.49		
Smoking history	Yes	571	63	508	11.03	1.082	0.2983
	No	5,390	676	4,714	12.54		
Alcohol consumption	Yes	720	82	638	11.39	0.7667	0.3812
	No	5,241	657	4,584	12.54		
Physical exercise	Yes	845	118	727	13.96	2.227	0.1356
	No	5,116	621	4,495	12.14		
Tea drinking	No	5,526	692	4,384	12.52	2.76	0.0966
	Yes	435	47	388	10.80		
Fresh vegetable consumption	No	3,619	451	3,168	12.46	0.03397	0.8538
	Yes	2,317	285	2,032	12.30		
Fresh fruit consumption	No	3,671	449	3,222	12.23	0.1968	0.6573
	Yes	2,274	287	1,987	12.62		
Meat consumption	No	4,836	595	4,241	12.30	0.3845	0.5352
	Yes	1,109	144	965	12.98		

(Continued)

TABLE 11 (Continued)

Characteristics		N	(+)	(–)	n (%)	χ^2	*P*
Coarse grain consumption	No	5,348	659	4,689	12.32	0.6942	0.4047
	Yes	592	80	512	13.51		
Are you in good health?	No	396	60	336	15.15	2.963	0.0852
	Yes	5,565	679	4,886	12.20		
TCT	ASCUS and above	562	281	281	50.00	807.9	< 0.0001
	NILM	5,399	458	4,941	8.48		
Population area	Urban population	3,961	463	3,498	11.69	5.453	0.0195
	Rural population	2,000	276	1,724	13.80		

TCT, Thinprep cytology test; ASCUS, atypical squamous cells of undetermined significance; NILM, negative for intraepithelial lesion or malignancy.

TABLE 12 Multivariate analysis of risk factors affecting human papillomavirus (HPV) infection.

Characteristics	SE	*P*	OR (95%)
35–45 year	0.115	0.001	0.687 (0.549∼0.860)
46–55 year	0.088	0.001	0.740 (0.622∼0.879)
High school and below	0.088	<0.001	1.426 (1.199∼1.696)
Initial age of sexual life ≤ 19 years	0.202	0.046	1.496 (1.008∼2.220)
Number of sexual partners > 1	0.126	0.010	1.382 (1.081∼1.768)
ASCUS and above	0.097	<0.001	10.788 (8.912∼13.060)
Contraceptive methods other than condoms	0.135	0.007	1.437 (1.103∼1.871)
Nationalities other than Han	0.180	0.004	1.690 (1.187∼2.406)
Rural population	0.082	0.020	1.210 (1.031∼1.419)

OR, odds ratio; ASCUS, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion.

an individual's awareness of beneficial health practices, thus reducing their risk of disease (Sharmin et al., 2021). Hence, education pertaining to HPV and the HPV vaccine should be strengthened to improve self-protection awareness.

This study also found that living in rural areas represents a high-risk factor for HPV infection, which may be due to a lack of cervical cancer screening services, poverty in rural areas, and poor education. Indeed, Vhuromu et al. (2018) reported that the utilization rate of cervical cancer screening in rural areas is low. Hence, new approaches to improve health systems and cervical cancer screening services in rural areas may improve these disparities, thereby reducing cervical cancer cases in rural areas.

An international study conducted by IARC in 2011 showed that compared to those who never smoked, the risk of cervical cancer for current smokers was OR = 1.94 (95% confidence interval: 1.26–2.98) (Louie et al., 2011); however, no significant difference was observed between women that smoked and those that did not.

Currently, HPV vaccine research (preventive and therapeutic) has significantly increased. A study in the United States found that HPV preventive vaccines (bivalent, tetravalent, and non-valent vaccines) are safe and have preventive and protective effects (Blake and Middleman, 2017). However, given that different countries and regions have different HPV infection conditions, the vaccine benefits may vary based on geographical location. As such, China should focus on developing an HPV-specific preventive vaccine suitable for Chinese nationals, which takes into consideration the dual factors of regional typing and age stratification during vaccination.

Conclusion

The occurrence and development of cervical cancer is a multi-factor, complex, and progressive biological process. Currently, the relationship between HPV infection, cervical cancer, and precancerous lesions is clear. Through cervical screening of female HPV infection rate, this study identified that a high school education or lower, age at first sexual encounter ≤ 19 years, number of sexual partners > 1, TBS classification of ASCUS or above, non-condom-based contraception, other nationalities except Han, and rural population serve as risk factors for HPV infection. It is, therefore, necessary to establish and improve the cervical cancer screening system according to the relevant factors affecting the incidence of cervical cancer, screen out and follow-up with high-risk groups, strengthen public education on prevention and treatment of cervical cancer, and improve awareness of cancer prevention among the general population. Moreover, targeted improvement measures for women in different regions

are fundamental to reducing the positivity rate of HPV infection. Improving the living environment, improving education level, expanding health education, developing beneficial health habits, and actively preventing and treating HPV infection are the basic and effective ways to reduce the positive rate of HPV infection. Only by providing a large amount of epidemiological data can we carry out effective vaccine research and development while also optimizing the cost-effectiveness of vaccinating the Chinese population. Collectively, the results of this study provide epidemiological data for the development of an HPV vaccine, to effectively reduce the incidence of cervical cancer in Shenyang.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was approved by the Ethics Committee of the Liaoning Cancer Hospital (Ethics batch number: 20180106), and informed consent was obtained from all participants included in this study. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

HP and CW designed and supervised the project. DY, JZ, and XC collected the clinical data samples. DY collected and processed the data, performed the data analysis, and drafted the manuscript. JM translated and edited the manuscript. All authors reviewed, discussed and edited version of the manuscript, and agreed with the final manuscript.

Funding

This study was supported by the National Key Research and Development Program (Grant no. 2016YFC1303001).

Acknowledgments

We thank the staff of the Cancer Hospital of China Medical University who partook in the study and JM for his great help in revising this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alotaibi, H. J., Almajhdi, F. N., Alsaleh, A. N., Obeid, D. A., Khayat, H. H., Al-Muammer, T. A., et al. (2020). Association of sexually transmitted infections and human papillomavirus co-infection with abnormal cervical cytology among women in Saudi Arabia. *Saudi J. Biol. Sci.* 27, 1587–1595. doi: 10.1016/j.sjbs.2020.03.021

Arbyn, M., Castellsague, X., de Sanjose, S., Bruni, L., Saraiya, M., Bray, F., et al. (2011). Worldwide burden of cervical cancer in 2008. *Ann. Oncol.* 22, 2675–2686. doi: 10.1093/annonc/mdr015

Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J., et al. (2020). Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis. *Lancet Glob. Health* 8, e191–e203. doi: 10.1016/S2214-109X(19)30482-6

Ardekani, A., Sepidarkish, M., Mollalo, A., Afradiasbagharani, P., Rouholamin, S., Rezaeinejad, M., et al. (2022). Worldwide prevalence of human papillomavirus among pregnant women: A systematic review and meta-analysis. *Rev. Med. Virol.* 9:e2374. doi: 10.1002/rmv.2374

Blake, D. R., and Middleman, A. B. (2017). Human papillomavirus vaccine update. *Pediatr. Clin. North Am.* 64, 321–329. doi: 10.1016/j.pcl.2016.11.003

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A., et al. (2018). Global cancer statistics 2018:Globocan estimates of incidence and mortalityworldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Bruni, L., Albero, G., Serrano, B., Mena, M., Collado, J. J., Gómez, D., et al. (2019). *Human Papilloma-Virus And Relatated Diseases In The World*. Barcelona: ICO/IARC Information Centre on HPV and Cancer.

Choi, H. C. W., Jit, M., Leung, G. M., Tsui, K. L., and Wu, J. T. (2018). Simultaneously characterizing the comparative economics of routine female adolescent nonavalent human papillomavirus (HPV) vaccination and assortativity of sexual mixing in Hong Kong Chinese: A modeling analysis. *BMC Med.* 16:127. doi: 10.1186/s12916-018-1118-3

de Sanjosé, S., Brotons, M., and Pavo'n, M. A. (2018). The natural history of human papillomavirus infection. *Best Pract. Res. Clin. Obstet. Gynaecol.* 47, 2–13. doi: 10.1016/j.bpobgyn.2017.08.015

Duan, R., Xu, K., Huang, L., Yuan, M., Wang, H., Qiao, Y., et al. (2022). Temporal trends and projection of cancer attributable to human papillomavirus infection in china, 2007-2030. *Cancer Epidemiol. Biomark. Prev.* 31, 1130–1136. doi: 10.1158/1055-9965.EPI-21-1124

Giuliano, A. R., Joura, E. A., Garland, S. M., Huh, W. K., Iversen, O. E., and Kjaer, S. K. (2019). Nine-valent HPV vaccine efficacy against related diseases and definitive therapy: Comparison with historic placebo population. *Gynecol. Oncol.* 154, 110–117. doi: 10.1016/j.ygyno.2019.03.253

Goodman, J. E., Mayfield, D. B., Becker, R. A., Hartigan, S. B., and Erraguntla, N. K. (2020). Recommendations for further revisions to improve the International Agency for Research on Cancer (IARC) Monograph program. *Regul. Toxicol. Pharmacol.* 113:104639. doi: 10.1016/j.yrtph.2020.104639

Hariri, S., and Warner, L. (2013). Condom use and human papillomavirus in men. *J. Infect. Dis.* 208:367-369. doi: 10.1093/infdis/jit193

Hariri, S., Unger, E. R., Sternberg, M., Dunne, E. F., Swan, D., Patel, S., et al. (2011). Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006. *J. Infect. Dis.* 204, 566–573. doi: 10.1093/infdis/jir341

International Collaboration of Epidemiological Studies of Cervical Cancer (2009). Cervical carcinoma and sexual behavior: Collaborative reanalysis of individual data on 15,461 women with cervical carcinoma and 29,164 women without cervical carcinoma from 21 epidemiological studies. *Cancer Epidemiol. Biomarkers Prev.* 18, 1060–1069.

Kclurman, R. J., Carcangiu, M. L., Herrington, C. S., and Young, R. H. (2014). *WHO Classigication Of Tumors Of Female Reproductive Organs [M]*, 4th Edn. Lyon: International Agency for Research on Cancer.

Kitamura, T., Suzuki, M., Shigehara, K., and Fukuda, K. (2021). Prevalence and risk factors of human papillomavirus infection among japanese female people: A nationwide epidemiological survey by self-sampling. *Asian Pac. J. Cancer Prev.* 22, 1843–1849. doi: 10.31557/APJCP.2021.22.6.1843

Li, L., Chen, Y., Chen, J., Su, Q., Tang, J., Yang, P., et al. (2020). Prevalence and genotype distribution of high-risk human papillomavirus among chinese women in sichuan province. *Jpn. J. Infect. Dis.* 73, 96–101. doi: 10.7883/yoken.JJID.2019.181

Liang, H., Fu, M., Zhou, J., and Song, L. (2016). Evaluation of 3D-CPA, HR-HPV, and TCT joint detection on cervical disease screening. *Oncol. Lett.* 12, 887–892. doi: 10.3892/ol.2016.4677

Liu, Z. C., Liu, W. D., Liu, Y. H., Ye, X. H., and Chen, S. D. (2015). Multiple sexual partners as a potential independent risk factor for cervical cancer: A meta-analysis of epidemiological studies. *Asian Pac. J. Cancer Prev.* 16, 3893–3900. doi: 10.7314/APJCP.2015.16.9.3893

Louie, K. S., Castellsague, X., de Sanjose, S., Herrero, R., Meijer, C. J., Shah, K., et al. (2011). International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Smoking and passive smoking in cervical cancer risk: Pooled analysis of couples from the IARC multicentric case-control studies. *Cancer Epidemiol. Biomark. Prev.* 20, 1379–1390. doi: 10.1158/1055-9965.EPI-11-0284

Niu, J., Pan, S., Wei, Y., Hong, Z., Gu, L., Di, W., et al. (2022). Epidemiology and analysis of potential risk factors of high-risk human papillomavirus (HPV) in Shanghai China: A cross-sectional one-year study in non-vaccinated women. *J. Med. Virol.* 94, 761–770. doi: 10.1002/jmv.27453

Ouh, Y. T., Min, K. J., Cho, H. W., Ki, M., Oh, J. K., and Shin, S. Y. (2018). Prevalence of human papillomavirus genotypes and precancerous cervical lesions in a screening population in the Republic of Korea, 2014-2016. *J. Gynecol. Oncol.* 29:e14. doi: 10.3802/jgo.2018.29.e14

Schettino, M. T., De Franciscis, P., Schiattarella, A., La Manna, V., Della Gala, A., Caprio, F., et al. (2008). Age-specific prevalence of infection with human papillomavirus in females: A global review. *J. Adolesc. Health* 43, S5–S25, S25 e21–41.

Seyoum, A., Assefa, N., Gure, T., Seyoum, B., Mulu, A., and Mihret, A. (2022). Prevalence and genotype distribution of high-risk human papillomavirus infection among sub-saharan african women: A systematic review and meta-analysis. *Front. Public Health* 8:890880. doi: 10.3389/fpubh.2022.890880

Sharmin, S., Sabikunnahar, B., Aditya, A., Khan, M. A., Nessa, A., Ahsan, C. R., et al. (2021). Genotypic distribution and prevalence of human papillomavirus infection in an apparently healthy female population in Bangladesh. *IJID Reg.* 1, 130–134. doi: 10.1016/j.ijregi.2021.10.005

Shi, N., Lu, Q., Zhang, J., Li, L., Zhang, J., and Zhang, F. (2017). Analysis of risk factors for persistent infection of asymptomatic women with high-risk human papilloma virus. *Hum. Vaccin. Immunother.* 13, 1–7. doi: 10.1080/21645515.2016.1239669

Smith, J. S., Melendy, A., Rana, R. K., and Pimenta, J. M. (2008). Age-specific prevalence of infection with human papillomavirus in females: A global review. *J. Adolesc. Health* 43, S5–25, S25.e1–41.

Solomon, D., Davey, D., Kurman, R., Moriarty, A., O'Connor, D., Prey, M., et al. (2002). The 2001 Bethesda System: Terminology for reporting results of cervical cytology. *JAMA* 287, 2114–2119. doi: 10.1001/jama.287.16.2114

Sui, S., Zhu, M., Jiao, Z., Han, L., Wang, L., Niyazi, M., et al. (2018). Prognosis and related factors of HPV infections in postmenopausal Uyghur women. *J. Obstet. Gynaecol.* 38, 1010–1014. doi: 10.1080/01443615.2018.1440285

Sung, H., Ferlay, J., Siegel, R. L., Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of incidence and mortality worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Taskin, M. H., Nursal, A. F., Oruc, M. A., and Kariptas, E. (2022). Genotype distribution and prevalence of high-risk human papillomavirus infection among women in samsun province of turkey. *Asian Pac. J. Cancer Prev.* 23, 2477–2482. doi: 10.31557/APJCP.2022.23.7.2477

Trottier, H., and Franco, E. L. (2006). Human papillomavirus and cervical cancer:burden of illness and basis for prevention. *Am. J. Manag. Care* 12, S462–S472.

Vhuromu, E. N., Goon, D. T., Maputle, M. S., Lebese, R. T., and Okafor, B. U. (2018). Utilization of cervical cancer screening services among women in vhembe district, south africa: A cross-sectional study. *Open Public Health J.* 11, 451–463. doi: 10.2174/1874944501811010451

Wang, J., Tang, D., Wang, J., Zhang, Z., Chen, Y., Wang, K., et al. (2019). Genotype distribution and prevalence of human papillomavirus among women with cervical cytological abnormalities in Xinjiang, China. *Hum. Vaccin. Immunother* 15, 1889–1896. doi: 10.1080/21645515.2019.1578598

Wang, L., Chen, G., and Jiang, J. (2021). Genotype distribution and prevalence of human papillomavirus infection in women in northern jiangsu province of China. *Cancer Manage. Res.* 13, 7365–7372.

Wei, L., Ma, L., Qin, L., and Huang, Z. (2022). The prevalence and genotype distribution of human papillomavirus among women in Guangxi, southern China. *Infect. Agent Cancer* 17:19. doi: 10.1186/s13027-022-00431-5

Zheng, R. S., Sun, K. X., Zhang, S. W., Zeng, H. M., Zou, X. N., Chen, R., et al. (2019). [Report of cancer epidemiology in China, 2015]. *Zhonghua Zhong Liu Za Zhi* 41, 19–28.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 16 September 2022
DOI 10.3389/fmicb.2022.951040

OPEN ACCESS

EDITED BY
Ana Grande-Pérez,
Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Spain

REVIEWED BY
MahmoudReza Pourkarim,
KU Leuven, Belgium
Mauro Pistello,
University of Pisa, Italy

*CORRESPONDENCE
Lia van der Hoek
c.m.vanderhoek@amsterdamumc.nl

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 23 May 2022
ACCEPTED 30 August 2022
PUBLISHED 16 September 2022

CITATION
Kaczorowska J, Cicilionytė A, Wahdaty AF, Deijs M, Jebbink MF, Bakker M and van der Hoek L (2022) Transmission of anelloviruses to HIV-1 infected children.
Front. Microbiol. 13:951040.
doi: 10.3389/fmicb.2022.951040

Transmission of anelloviruses to HIV-1 infected children

Joanna Kaczorowska[1,2], Aurelija Cicilionytė[1,2], Annet Firouzi Wahdaty[1,2], Martin Deijs[1,2], Maarten F. Jebbink[1,2], Margreet Bakker[1,2] and Lia van der Hoek[1,2]*

[1]Laboratory of Experimental Virology, Department of Medical Microbiology and Infection Prevention, University of Amsterdam, Amsterdam, Netherlands, [2]Amsterdam Institute for Infection and Immunity, Amsterdam, Netherlands

Anelloviruses (AVs) are widespread in the population and infect humans at the early stage of life. The mode of transmission of AVs is still unknown, however, mother-to-child transmission, e.g., via breastfeeding, is one of the likely infection routes. To determine whether the mother-to-child transmission of AVs may still occur despite the absence of natural birth and breastfeeding, 29 serum samples from five HIV-1-positive mother and child pairs were Illumina-sequenced. The Illumina reads were mapped to an AV lineage database "Anellometrix" containing 502 distinct ORF1 sequences. Although the majority of lineages from the mother were not shared with the child, the mother and child anellomes did display a significant similarity. These findings suggest that AVs may be transmitted from mothers to their children via different routes than delivery or breastfeeding.

KEYWORDS

anellome, anelloviridae, anellovirus, early-life infections, mother-to-child transmission, virome

Introduction

Anelloviruses (AVs) are small, circular single-stranded DNA viruses. So far, AV infection has not been associated with any disease, however, the levels of AVs seem to be connected with the levels of host immunosuppression (Focosi et al., 2016). AVs are detected in a majority of body fluids in people of all ages (Spandole et al., 2015), but they dominate in blood, being one of the most abundant viruses in this environment (Cebriá-Mendoza et al., 2021). Not much is known about the ways of transmission of AVs. There is evidence of blood transfusion transmission (Arze et al., 2021). However, considering the enormous abundance of AVs in the human population, including young children, a mother-to-child transmission is highly likely.

Children are born AV-negative (Lim et al., 2015; Liang and Bushman, 2021), and there was so far no convincing evidence of vertical transmission of AVs

(Tyschik et al., 2017). AV DNA was detected in children's blood (Tyschik et al., 2017, 2018; Kaczorowska et al., 2022a) and feces (Lim et al., 2015) in the first months of life, thus it is likely that AVs are transmitted from the mother to child during the delivery or the post-partum period. Children born naturally tend to have higher loads of AVs compared to the ones born via the caesarian section (McCann et al., 2018). Moreover, AV DNA has been detected in breastmilk (Ohto et al., 2002; Maqsood et al., 2021), and we described recently that beta- and gammatorqueviruses are dominating both in breastmilk and in the blood of children younger than 6 months, which suggests a role of breastfeeding in AV transmission (Kaczorowska et al., 2022a). However, other means of mother-to-child transmission, such as fecal-oral or respiratory, are yet to be explored.

Here, we Illumina-sequenced a total of 29 serum samples derived from five mother and child pairs. We compared the viromes of the mother and child pairs and assessed the presence of shared anellovirus genomes. All the subjects were HIV-1 positive, thus the children were born via cesarean section and they were not breastfed.

Materials and methods

Clinical samples

Twenty-nine serum samples were collected from five mothers and their children. All individuals were HIV-1 positive. There was no information available regarding the presence of antiretroviral therapy in the subjects at the moment of samples' collection, except the mother 3, who was on the therapy from the second time point onward. Four mothers (M1, M3, M4, and M5) and two children (C1 and C5) had more than one serum sample available, and the samples were collected longitudinally at irregular intervals. All samples are listed in **Supplementary Table 1**.

Nucleic acid isolations

After thawing at room temperature, the serum samples were centrifuged for 10 min at 5,000 *g*. A 100 μL of supernatant was transferred to a new tube and treated with TURBO DNase (Invitrogen). The total nucleic acids were extracted using the Boom isolation method (Boom et al., 1990), and stored at −80°C until further use.

Genus-specific qPCR

Three quantitative PCRs (qPCRs) were performed on all the selected samples to assess the prevalence and concentration of AV genera in the tested samples. The first qPCR detected the genus alphatorquevirus, the second one detected betatorquevirus, and the third—beta- and gammatorquevirus. The qPCRs were performed as described previously (Kaczorowska et al., 2022a). Briefly, the qPCR reaction mixture consisted of 2.5 μL isolated nucleic acids (non-rolling circle amplified), 6.25 μL 2× Qiagen RotorGene Probe Master-mix (Qiagen, catalog number 204574), 0.25 μL probe, 0.5 μL forward, and 0.5 μL reverse primer (all 10 μM) and 2.5 μL H_2O. The reaction was performed on a Rotor-Gene machine (Qiagen GmbH, Hilden, Germany) as follows: 95°C for 3 min, followed by 40 cycles of 95°C for 3 s, 60°C for 10 s. The final elongation step was held at 72°C for 3 min.

Illumina libraries preparation

The Illumina libraries were prepared as described previously (Kaczorowska et al., 2022b). Briefly, the nucleic acid samples were first amplified using rolling circle amplification (4 h at 30°C, followed by 10 min at 65°C) and then fragmented using fragmentase enzyme, for 25 min at 37°C. The ends of the fragmented nucleic acids were repaired using polymerase I, large (Klenow) fragment (New England Biolabs, NEB) in combination with NEB2 10× buffer (NEB) and dNTPs (final concentration 500 μM each). A-taling was performed with 3′-5′ exo (-) polymerase I, large (Klenow) fragment (NEB), NEB2 buffer, and dATPs (final concentration 200 μM, NEB). Both reactions were performed at 37°C for 30 min and between each of the mentioned enzymatic reactions, an AMPure XP Beads clean-up was performed. The NEB Next adaptors were mixed with T4 ligase (Invitrogen) and T4 buffer and the reaction was incubated overnight at 16°C. Afterward, an AMPure XP Beads size selection was performed and the eluate was used in the adaptor-enrichment PCR. The PCR master mix consisted of Q5 Hot Start master mix (NEB), NEB Next universal primer (final concentration of 0.5 μM; NEB), NEB Next index primer (unique for each sample; final concentration of 0.5 μM; NEB), and USER enzyme (NEB). Cycling was performed as follows: 37°C for 15 min, 98°C for 30 s, followed by 12 cycles of 98°C for 10 s and 65°C for 75 s, followed by a final extension at 65°C for 5 min. Two rounds of AMPure XP Beads size selection were performed, and the concentration of each sample was measured using Qubit High Sensitivity assay (Invitrogen). The samples were pooled at equal concentrations and run on the Illumina miSeq sequencing machine. The pool containing samples from mothers (n = 22) was processed and ran separately from the children's sample pool (n = 7), to avoid cross-contamination.

Anellometrix sequence database

All complete or nearly complete human-derived AV sequences were downloaded from NCBI database (state for

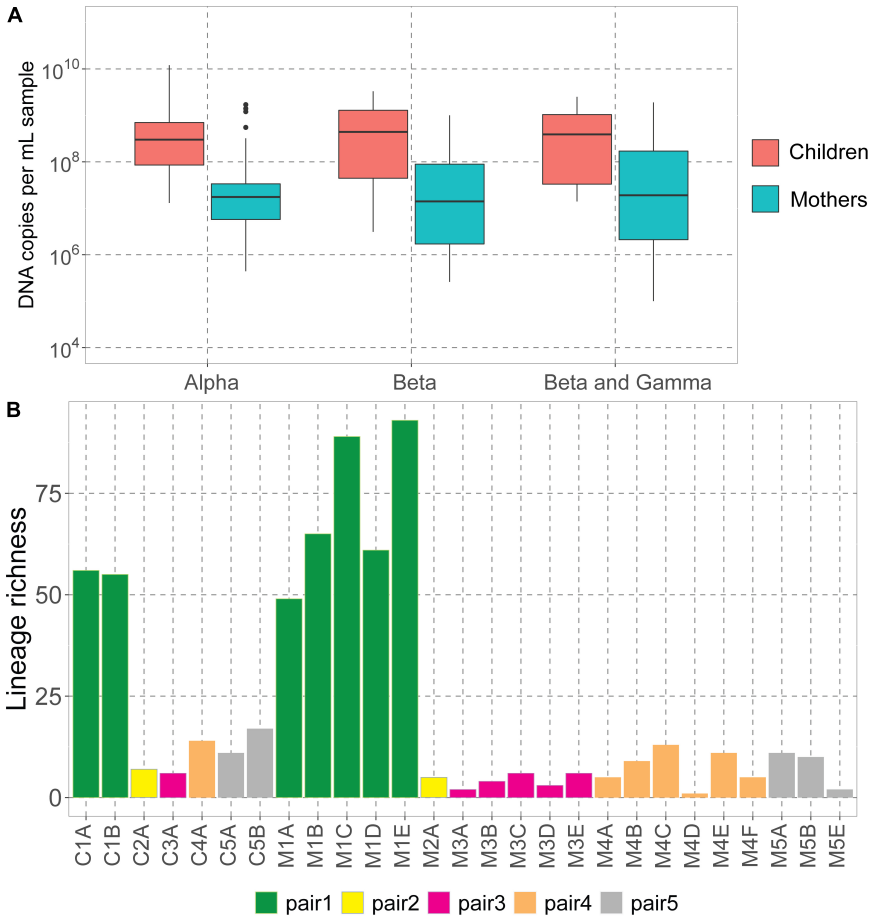

FIGURE 1
Anelloviruses in mothers and their children. **(A)** Alpha-, beta-, and beta- and gammatorquevirus DNA copies per mL serum in mothers and children measured by genus-specific qPCR; only samples positive in the assay are included in the graph. Wilcoxon sum-rank test was used to assess the significance; all differences were insignificant (p-value >0.05). **(B)** Number of lineages (richness) in each sample.

October 2021). The sequences were merged into one fasta file, and the *ORF1* gene nucleotide sequences were extracted using EMBOSS getorf.[1] The sequences were clustered using 95% threshold, and any duplicates were removed using dedupe script from BBTools.[2] By running a BLASTn search against a recently updated database of AV sequences (Varsani et al., 2021), we categorized the majority of the lineages into species. The final Anellometrix genome database consisted of 502 *ORF1* sequences (**Supplementary Table 2**).

Sequencing data processing

The adaptors and low quality reads were trimmed using Trimmomatic version 0.39 (Bolger et al., 2014). Paired reads were aligned to the Anellometrix database using bowtie2 (Langmead et al., 2009), using –very-sensitive setting. The obtained sam files were indexed, converted to bam, and sorted using SAMtools (Li et al., 2009). A table with mapped read numbers was generated using SAMtools idxstats function, and the percentage coverage was obtained using bbmap mpileup command.[3] The mapping was considered valid when the genome was covered at least once for ≥75% of the genome length. The read counts were normalized to reads per million (RPM), or, for the heatmap graphs, to relative abundance.

To compare the anellomes between the samples, we used unweighted UniFrac on QIIME2 (Lozupone et al., 2006; Bolyen et al., 2019). Unweighted UniFrac calculates the distance matrix based on presence-absence tables and phylogenetic relationships between the tested samples. The

1 http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf
2 https://jgi.doe.gov/data-and-tools/software-tools/bbtools/
3 https://sourceforge.net/projects/bbmap/

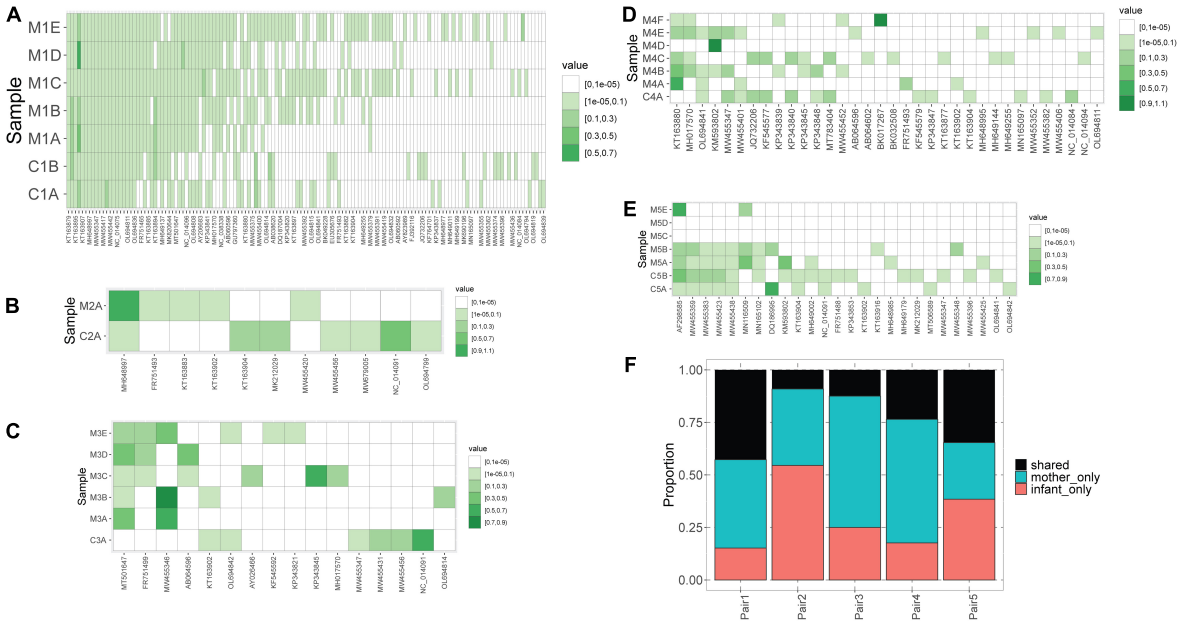

FIGURE 2
Blood anellomes of mothers and children. Heatmaps of relative abundance of different anellovirus lineages (*x*-axis) in samples (*y*-axis) within pair #1 **(A)**, pair #2 **(B)**, pair #3 **(C)**, pair #4 **(D)** and pair #5 **(E)**. The stronger the color, the higher the relative abundance (indicated by "value"), and the white color indicates the absence of a lineage. **(F)** Proportion of shared and not shared lineages within each pair. The unshared lineages were either detected solely in the mother ("mother-only") or in the child ("infant-only").

ORF1 sequences of Anellometrix database were aligned using MAFFT (E-INS algorithm) (Katoh et al., 2018) and the maximum-likelihood phylogenetic tree was constructed using RAXML (Stamatakis, 2014). All the abovementioned programs were run on a high-performance cluster computer Lisa (Surfsara).[4]

Statistical analysis

All graphs were constructed in R version 4.1.3 using the tidyverse, vegan, reshape2, and ggpubr packages, and the statistics were calculated using rstatix and vegan packages.

Results

We assessed the DNA concentrations of the 3 human-infecting anellovirus (AV) genera using the genus-specific qPCRs. All samples were positive in alphatroquevirus qPCR, while five samples were negative in betatorquevirus and four in beta- and gammatorquevirus qPCR (**Supplementary Table 1**). Interestingly, we observed negative values only in mother samples after having given birth, but there were no significant differences between AV DNA concentrations between mother and child samples (**Figure 1A**).

The highest AV DNA concentrations among mothers were noted for mother #1 (**Supplementary Figure 1**). Alphatorquevirus DNA loads were stable for all mothers (**Supplementary Figure 1A**), while beta- and gammatorquevirus loads were fluctuating over time (**Supplementary Figures 1B,C**).

Illumina sequencing resulted in a total of 5.4×10^7 paired reads, ranging from 3.0×10^5 (sample M1A) and 6.5×10^6 (C3A) reads per sample (**Supplementary Table 1**). Sample M5D was excluded from further analysis due to a low number of reads. We mapped the reads to the Anellometrix database, which consists of 502 ORF1 sequences derived from a broad spectrum of AV lineages. We obtained between 0 reads per million (rpm; sample M5C) and almost 5×10^5 rpm (C1B) reads mapping to Anellometrix reference sequences. The AV richness differed among the subjects, with the highest richness observed in mother #1 samples (**Figure 1B**), which also showed the highest AV DNA concentrations in genus-specific qPCRs (**Supplementary Figure 1**). One of the samples from this subject, M1E, had reads mapping to as many as 98 distinguishable AV lineages (**Figure 1B**). Interestingly, of all children, child #1 also possessed the highest number of lineages (56 lineages). The lowest richness values were observed in child #2 (7 lineages)/mother #2 (5 lineages), and child #3 (6

4 https://www.surf.nl/

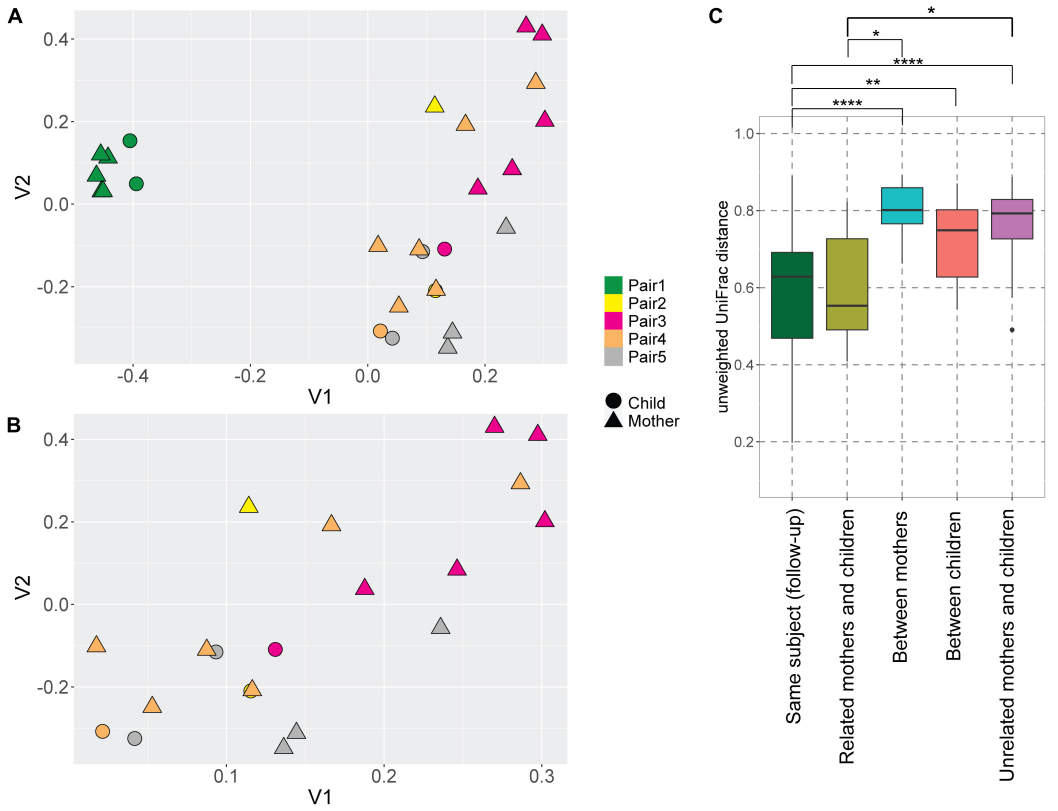

FIGURE 3
UniFrac distances of the anellomes. Principal Component Analysis (PCoA) plot using unweighted UniFrac distances between all samples **(A)** and samples excluding pair #1 **(B)**. The colors represent different pairs, while the shapes represent the category of a sample (child or mother). **(C)** Unweighted UniFrac pairwise comparison between the longitudinal samples from the same subject, and related and unrelated samples (thus within or outside the pairs). To compare related and unrelated samples, only selected samples (**Supplementary Table 1**—samples marked with an asterisk) were used. Statistical significance was assessed using Wilcoxon sum-rank test. Only the significant values are marked on the graph; explanations of the symbols: *$P \leq 0.05$, **$P \leq 0.01$, ****$P \leq 0.0001$.

lineages)/mother #3 (6 lineages). One of the samples derived from mother #5 (M5C) had a proper number of sequence reads, a reasonably high virus load (10^7 DNA copies/mL), yet no detectable AV Illumina reads. Apparently, the library preparation did not work in this sample, possibly due to a lack of circular genomes needed for rolling circle amplification. There was no significant difference between the number of lineages detected in mothers and in children (Wilcoxon sum-rank test, p-value >0.05).

We observed that mothers and children within the same pair possessed a similar number of lineages (**Figure 1B**), however, the lineages were not always the same (**Figure 2**). In pair #1, which had the broadest anellome, 59 out of the total of 138 (43%) lineages were shared between mother and children samples (**Figure 2A**). There was just 1 lineage shared between the mother and the child in pair #2 (out of 11 lineages; 9%; **Figure 2B**), 2 out of 16 in pair #3 (12.5%; **Figure 2C**), 8 out of 34 in pair #4 (24%; **Figure 2D**), and 9 out of 26 in pair #5 (35%; **Figure 2E**). For pairs #1, #3, and #4 the majority of not shared lineages were present in mothers only, while in pairs #2 and #5 most of the unshared lineages were detected in infants (**Figure 2F**). In pair #1 and pair #5, the shared lineages were significantly more abundant in mother samples than the unshared lineages (Wilcoxon sum-rank test; p-value $= 7.3 \times 10^{-11}$ for pair #1 and 0.01 for pair #5; **Supplementary Table 3**). Moreover, the most abundant AV lineage of mother #2 was also the only one that was shared with the child (**Figure 2B**). No correlation between the relative abundance and sharing of the lineage was observed in pairs #3, #4, and #5 (p-value >0.05).

Of note, we observed that anellomes of the mothers that had more than 1 sample available (#1, #3, #4, and #5) were relatively stable in time in terms of the anellome breadth. The broad anellome of mother #1 remained so for the whole follow-up period (**Figure 2A**), while the smaller anellomes of mothers #3 (**Figure 2C**) and #4 (**Figure 2D**) were narrow throughout the follow-up. However, only a small number of lineages persisted throughout the whole follow-up period. In mother #1, 37 out of 117 lineages were detected in all the time points, in mother #3 it was 1 out of 16, and there were no such lineages in mother #4 and mother #5. Moreover, in mother #5, a large drop in the

number of lineages was observed over time—with samples M5C and M5D having no detectable lineages, and the last sample possessing only two lineages (**Figure 2E**).

Even though the majority of AV lineages were not shared between the mothers and children, we still hypothesized that the AV lineages of mother and child within a pair are phylogenetically related, and unrelated samples possess more phylogenetically divergent anellomes. We compared the anellome diversity and the phylogenetic relationships of all samples using the unweighted UniFrac. Principle component analysis (PCoA) of pairwise unweighted UniFrac showed strong clustering of samples from pair #1 (**Figure 3A**), but no such clustering was observed in the case of samples from the remaining pairs (**Figure 3B**).

Next, we compared the unweighted UniFrac distances between the samples (**Figure 3C**). Unsurprisingly, the mean pairwise distances between the longitudinal samples derived from the same subject were significantly smaller than the ones measured between different subjects: different mothers (Wilcoxon sum-rank test; p-value = 9.34×10^{-8}), and different children (p-value = 0.01; **Figure 3C**, **Supplementary Table 4**). Thus, to avoid the bias caused by the high similarity of follow-up samples from the same subject, we selected just one sample per subject when comparing the related and unrelated mothers and children. We chose the mother and child samples within each pair that were closest to each other in terms of UniFrac distances (samples marked with asterisks in **Supplementary Table 1**). We selected these samples because they were most likely collected at the moment of the most prominent mother-to-child transmission. There was indeed a significant difference in pairwise distances between mothers and related children versus unrelated mothers and children (p-value = 0.04; **Figure 3C**; **Supplementary Table 4**). This result shows that even though children do not share exactly the same AV lineages with their mother, their anellome composition is more similar to their own mother compared to unrelated mothers or children. Next to that, we also compared the UniFrac distances using mother samples collected closest to the delivery moment (samples marked by "$" in the **Supplementary Table 1**). In this case, we lost significance between related samples on the one hand and unrelated samples on the other hand (p-value >0.05; **Supplementary Figure 2**, **Supplementary Table 5**). This finding suggests that the moment of delivery is not the main moment of AV transmission.

Discussion

In this study we compared the anellovirus (AV) viromes (anellomes) of 5 HIV-1 infected mother and child pairs. We observed that related mothers and children possess more similar anellomes than unrelated people—both in the terms of AV lineage richness and phylogenetic relationships. We found an inconsistent number of shared lineages within pairs, which suggests that there may not be a general pattern of transmission when children are born from HIV-1 infected mothers. One mother and child pair shared almost half of the lineages (pair #1), while another shared as little as 1 out of 11 lineages (pair #2), and another only 2 out of 16 (pair #3). We hypothesize that the AV lineages that were present only in children may have been acquired from siblings, other family members tending to the child, or from the environment.

The healthy human virome is regarded as highly dynamic in the first months of life, both in the gut and in blood (Lim et al., 2015; Tyschik et al., 2017; Bushman and Liang, 2021; Kaczorowska et al., 2022a). The AVs colonize infants within the first 6 months of life (Lim et al., 2015; Kaczorowska et al., 2022a), yet the main source or route of the first infection is still unknown. All children in our study were born via caesarian section, were not breastfed, and still were infected by AVs in the very first months of life—thus the transmission must have occurred via a different route. Fecal-oral transmission is likely—feces are positive for AVs both in children and in adults (Lim et al., 2015; Shkoporov et al., 2019; Liang and Bushman, 2021; Beller et al., 2022). Airway transmission is another possible route since AVs are frequently detected in nasal secretions (Maggi et al., 2003), saliva (Inami et al., 2000; Naganuma et al., 2008; Liang and Bushman, 2021), and bronchoalveolar lavage liquid (Young et al., 2015; Segura-Wang et al., 2018). In one longitudinal study by Maggi and colleagues, two children were initially PCR-positive for alphatorquevirus DNA in the nasal secretions and negative in plasma. After a month since the initial detection in the respiratory tract, both children became positive for alphatorquevirus in plasma (Maggi et al., 2003), which suggests that the airway may be an important transmission route of AV infection in children. A metagenomic study involving paired nasal, fecal, and blood samples from mothers and their children would shed more light on the importance of the airway and fecal-oral transmission routes in early-life AV acquisition.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA838480.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee of the Amsterdam University Medical Center (location AMC) of the University of Amsterdam, Netherlands (MEC 07/182). Written informed

consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

JK, AC, AW, and LH designed the study. MB selected and provided the samples and extracted the available metadata. JK, AC, AW, MD, MFJ, and MB performed the wet lab experiments. AC, AW, and JK performed the analysis of next-generation sequencing results, designed the graphs, and performed the statistical analysis. JK wrote the initial draft. LH reviewed and edited the manuscript and supervised the experiments. All authors contributed to the manuscript revision.

Funding

This study was supported by a grant from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie agreement No. 721367 (HONOURs) and Amsterdam University Medical Center funding connected to HONOURs.

Acknowledgments

The authors gratefully acknowledge the Amsterdam Cohort Studies on HIV infection and AIDS (ACS). The authors thank all study participants for their contribution, as well as the nurses, data managers, and lab technicians. The authors would also like to thank Anne L. Timmerman, Lisanne Commandeur, and Cormac M. Kinsella from Amsterdam UMC for useful discussions and help in establishing the Anellometrix lineage database.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.951040/full#supplementary-material

References

Arze, C. A., Springer, S., Dudas, G., Patel, S., Bhattacharyya, A., Swaminathan, H., et al. (2021). Global genome analysis reveals a vast and dynamic anellovirus landscape within the human virome. *Cell Host Microbe* 29, 1305.e6–1315.e6. doi: 10.1016/j.chom.2021.07.001

Beller, L., Deboutte, W., Vieira-Silva, S., Falony, G., Tito, R. Y., Rymenans, L., et al. (2022). The virota and its transkingdom interactions in the healthy infant gut. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2114619119. doi: 10.1073/pnas.2114619119

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., and van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* 28, 495–503. doi: 10.1128/JCM.28.3.495-503.1990

Bushman, F., and Liang, G. (2021). Assembly of the virome in newborn human infants. *Curr. Opin. Virol.* 48, 17–22. doi: 10.1016/j.coviro.2021.03.004

Cebriá-Mendoza, M., Arbona, C., Larrea, L., Díaz, W., Arnau, V., Peña, C., et al. (2021). Deep viral blood metagenomics reveals extensive anellovirus diversity in healthy humans. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-86427-4

Focosi, D., Antonelli, G., Pistello, M., and Maggi, F. (2016). Torquetenovirus: the human virome from bench to bedside. *Clin. Microbiol. Infect.* 22, 589–593. doi: 10.1016/j.cmi.2016.04.007

Inami, T., Konomi, N., Arakawa, Y., and Abe, K. (2000). High prevalence of TT virus DNA in human saliva and semen. *J. Clin. Microbiol.* 38, 2407–2408.

Kaczorowska, J., Cicilionyt, A., Timmerman, A. L., Deijs, M., and van der Hoek, L. (2022a). Early-life colonization by anelloviruses in infants. *Viruses* 14:865. doi: 10.3390/v14050865

Kaczorowska, J., Deijs, M., Klein, M., Bakker, M., Jebbink, M. F., Sparreboom, M., et al. (2022b). Diversity and long-term dynamics of human blood anelloviruses. *J. Virol.* [Epub ahead of print]. doi: 10.1128/jvi.00109-22

Katoh, K., Rozewicki, J., and Yamada, K. D. (2018). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liang, G., and Bushman, F. D. (2021). The human virome: assembly, composition and host interactions. *Nat. Rev. Microbiol.* 19, 514–527. doi: 10.1038/s41579-021-00536-5

Lim, E. S., Zhou, Y., Zhao, G., Bauer, I. K., Droit, L., Ndao, I. M., et al. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* 21, 1228–1234. doi: 10.1038/nm.3950

Lozupone, C., Hamady, M., and Knight, R. (2006). UniFrac - An online tool for comparing microbial community diversity in a

phylogenetic context. *BMC Bioinform.* 7:371. doi: 10.1186/1471-2105-7-371

Maggi, F., Pifferi, M., Fornai, C., Andreoli, E., Tempestini, E., Vatteroni, M., et al. (2003). TT virus in the nasal secretions of children with acute respiratory diseases: relations to viremia and disease severity. *J. Virol.* 77, 2418–2425. doi: 10.1128/jvi.77.4.2418-2425.2003

Maqsood, R., Reus, J. B., Wu, L. I., Holland, L. A., Nduati, R., Mbori-Ngacha, D., et al. (2021). Breast milk virome and bacterial microbiome resilience in kenyan women living with HIV. *mSystems* 6:e01079-20. doi: 10.1128/mSystems.01079-20

McCann, A., Ryan, F. J., Stockdale, S. R., Dalmasso, M., Blake, T., Anthony Ryan, C., et al. (2018). Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* 2018, 1–13. doi: 10.7717/peerj.4694

Naganuma, M., Tominaga, N., Miyamura, T., Soda, A., Moriuchi, M., and Moriuchi, H. (2008). TT virus prevalence, viral loads and genotypic variability in saliva from healthy Japanese children. *Acta Paediatr. Int. J. Paediatr.* 97, 1686–1690. doi: 10.1111/j.1651-2227.2008.00962.x

Ohto, H., Ujiie, N., Takeuchi, C., Sato, A., Hayashi, A., Ishiko, H., et al. (2002). TT virus infection during childhood. *Transfusion* 42, 892–898. doi: 10.1046/j.1537-2995.2002.00150.x

Segura-Wang, M., Görzer, I., Jaksch, P., and Puchhammer-Stöckl, E. (2018). Temporal dynamics of the lung and plasma viromes in lung transplant recipients. *PLoS One* 13:e0200428. doi: 10.1371/journal.pone.0200428

Shkoporov, A. N., Clooney, A. G., Sutton, T. D. S., Ryan, F. J., Daly, K. M., Nolan, J. A., et al. (2019). The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* 26, 527.e5–541.e5. doi: 10.1016/j.chom.2019.09.009

Spandole, S., Cimponeriu, D., Berca, L. M., and Mihăescu, G. (2015). Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch. Virol.* 160, 893–908. doi: 10.1007/s00705-015-2363-9

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tyschik, E. A., Rasskazova, A. S., Degtyareva, A. V., Rebrikov, D. V., and Sukhikh, G. T. (2018). Torque teno virus dynamics during the first year of life. *Virol. J.* 15, 1–4. doi: 10.1186/s12985-018-1007-6

Tyschik, E. A., Shcherbakova, S. M., Ibragimov, R. R., and Rebrikov, D. V. (2017). Transplacental transmission of torque teno virus. *Virol. J.* 14, 1–3. doi: 10.1186/s12985-017-0762-0

Varsani, A., Opriessnig, T., Celer, V., Maggi, F., Okamoto, H., Blomström, A. L., et al. (2021). Taxonomic update for mammalian anelloviruses (family Anelloviridae). *Arch. Virol.* 166, 2943–2953. doi: 10.1007/s00705-021-05192-x

Young, J. C., Chehoud, C., Bittinger, K., Bailey, A., Diamond, J. M., Cantu, E., et al. (2015). Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am. J. Transplant.* 15, 200–209. doi: 10.1111/ajt.13031

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fmicb.2022.1020147

OPEN ACCESS

EDITED BY
Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands

REVIEWED BY
Jose A. Usme-Ciro,
Universidad Cooperativa de Colombia, Colombia
Limin Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Zeineb Belaiba
zeineb.belaiba@pasteur.utm.tn
Anissa Chouikha
chouikhaanissa@gmail.com

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 15 August 2022
ACCEPTED 15 September 2022
PUBLISHED 17 October 2022

CITATION
Belaiba Z, Ayouni K, Gdoura M, Kammoun Rebai W, Touzi H, Sadraoui A, Hammemi W, Yacoubi L, Abdelati S, Hamzaoui L, Msaddak Azzouz M, Chouikha A and Triki H (2022) Whole genome analysis of hepatitis B virus before and during long-term therapy in chronic infected patients: Molecular characterization, impact on treatment and liver disease progression.
Front. Microbiol. 13:1020147.
doi: 10.3389/fmicb.2022.1020147

Whole genome analysis of hepatitis B virus before and during long-term therapy in chronic infected patients: Molecular characterization, impact on treatment and liver disease progression

Zeineb Belaiba[1,2]*, Kaouther Ayouni[1,2], Mariem Gdoura[1,2], Wafa Kammoun Rebai[3], Henda Touzi[1], Amel Sadraoui[1], Walid Hammemi[1], Lamia Yacoubi[1], Salwa Abdelati[4], Lamine Hamzaoui[5], Mohamed Msaddak Azzouz[5], Anissa Chouikha[1,2]* and Henda Triki[1,2]

[1]Laboratory of Clinical Virology, WHO Reference Laboratory for Poliomyelitis and Measles in the Eastern Mediterranean Region, Pasteur Institute of Tunis, University Tunis El Manar (UTM), Tunis, Tunisia, [2]Research Laboratory "Virus, Vectors and Hosts: One Health Approach and Technological Innovation for a Better Health," LR20IPT02, Pasteur Institute of Tunis, University Tunis El Manar (UTM), Tunis, Tunisia, [3]Laboratory of Biomedical Genomics and Oncogenetics (LR16IPT05), Pasteur Institute of Tunis, University Tunis El Manar (UTM), Tunis, Tunisia, [4]Department of Gastroenterology, Polyclinic of CNSS, Sousse, Tunisia, [5]Department of Gastroenterology, Hospital of Tahar Maamouri, Nabeul, Tunisia

Hepatitis B virus (HBV) infection remains a serious public health concern worldwide despite the availability of an efficient vaccine and the major improvements in antiviral treatments. The aim of the present study is to analyze the mutational profile of the HBV whole genome in ETV non-responder chronic HBV patients, in order to investigate antiviral drug resistance, immune escape, and liver disease progression to Liver Cirrhosis (LC) or Hepatocellular Carcinoma (HCC). Blood samples were collected from five chronic hepatitis B patients. For each patient, two plasma samples were collected, before and during the treatment. Whole genome sequencing was performed using Sanger technology. Phylogenetic analysis comparing the studied sequences with reference ones was used for genotyping. The mutational profile was analyzed by comparison with the reference sequence M32138. Genotyping showed that the studied strains belong to subgenotypes D1, D7, and D8. The mutational analysis showed high genetic variability. In the RT region of the polymerase gene, 28 amino acid (aa) mutations were detected. The most significant mutations were the pattern rtL180M+rtS202G+rtM204V, which confer treatment resistance. In the S gene, 35 mutations were detected namely sP120T, sT126S, sG130R, sY134F, sS193L, sI195M, and sL216stop were previously described to lead to vaccine, immunotherapy, and/or diagnosis escape. In the C gene, 34 mutations were found. In particular, cG1764A, cC1766G/T, cT1768A, and cC1773T in the BCP; cG1896A and cG1899A in

the precore region and cT12S, cE64D, cA80T, and cP130Q in the core region were associated with disease progression to LC and/or HCC. Other mutations were associated with viral replication increase including cT1753V, cG1764A/T, cC1766G/T, cT1768A, and cC1788G in the BCP as well as cG1896A and cG1899A in the precore region. In the X gene, 30 aa substitutions were detected, of which substitutions xT36D, xP46S, xA47T, xI88F, xA102V, xI127T, xK130M, xV131I, and xF132Y were previously described to lead to LC and/or HCC disease progression. In conclusion, our results show high genetic variability in the long-term treatment of chronic HBV patients causing several effects. This could contribute to guiding national efforts to optimize relevant HBV treatment management in order to achieve the global hepatitis elimination goal by 2030.

KEYWORDS

HBV, antiviral resistance, liver cirrhosis, PCR, whole genome, Sanger sequencing, hepatocellular carcinoma

Introduction

Hepatitis B virus (HBV) infection remains a serious public health concern worldwide despite the availability of an efficient vaccine and the major improvements in antiviral treatments. The World Health Organization (WHO) estimates that, in 2021, approximately 296 million persons are chronic HBV carriers. Among them, 820,000 represent a high risk of mortality caused by developing progressive liver diseases including hepatocellular carcinoma (HCC) and liver cirrhosis (LC) (WHO, 2021).

The genome of HBV is a circular DNA partially double-stranded of 3.2 kb and classified into 10 genotypes from A to J (Sunbul, 2014). It is organized into four main open overlapped reading frames (ORFs; pre-S1/pre-S2/S, pre-C/C, P, and X), encoding several proteins including the surface proteins S, M, and L holding the HBs antigen (HBsAg), the precore/core proteins holding HBeAg and HBcAg antigens, the polymerase (P), and the X protein holding the antigen HBxAg. Thus, mutations that occur in one gene can result in significant changes in the other overlapping genes.

Long-term treatment of HBV chronic patients with the available antiviral molecules can lead to the emergence of mutations throughout the whole genome. Mutations that occur within the reverse transcriptase (RT) domain of the P gene, target of antiviral treatment, may lead to treatment failure (Locarnini and Mason, 2006). Potential resistance-related mutations are grouped into 4 categories, primary mutations (category 1) could reduce antiviral susceptibility and HBV replication fitness. Secondary/compensatory mutations (category 2) developed subsequently and could restore functional defects in the RT activity of HBV caused by primary mutations. Putative antiviral resistance mutations (category 3) were reported as possible drug-resistant mutations but not verified experimentally and may be related to prolonged treatment or replication compensation. Pre-treatment mutations (category 4) could be found among treatment-naive patients but their role in antiviral treatment resistance has not been elucidated (Liu et al., 2010; Ciftci et al., 2014).

Moreover, mutations that emerge throughout a prolonged therapy could affect not only the RT region (Locarnini and Mason, 2006) but also the different overlapping genes. Therefore, such variations might result in hepatitis B immunoglobulin (HBIG) therapy escape, vaccine escape, misdiagnosis, and immune escape. They also could enhance viral replication capacity and viral persistence leading to the progression of severe liver diseases such as HCC or LC (Sheldon et al., 2006; Sheldon, 2008; Rajoriya et al., 2017).

On the other hand, it has been found that the presence of pre-existing naturally occurring mutations in treatment-naive patients may influence the efficacy of antiviral treatments. Therefore, knowledge of the mutational profile by whole genome sequencing of the HBV genome, for chronically infected patients, is of great interest for a complete diagnosis toward an efficient therapy scheme.

For HBV chronic patients in Tunisia, the national therapeutic schema is based on Entecavir (ETV) as a first-line of HBV treatment, and it is fully covered by the National Health Insurance Fund (NHIF), in case of resistance, Tenofovir disoproxil fumarate

Abbreviations: ADV, Adefovir; aa, Amino acid; DNA, Deoxyribonucleic acid; ETV, Entecavir; HBcAg, hepatitis B core antigen; HBeAg, hepatitis B e antigen; HBIG, Hepatitis B immunoglobulin; HBsAg, Hepatitis B surface antigen; HBV, Hepatitis B virus; HCC, Hepatocellular carcinoma; LdT, Telbivudine; LMV, Lamivudine; MHR, Major hydrophilic region; NCBI, National Center for Biotechnology Information; NGS, Next Generation Sequencing; ORF, open reading frame; PCR, Polymerase chain reaction; RT, reverse transcriptase; TDF, tenofovir disoproxil fumarate.

TABLE 1 Virological, treatment molecules, and treatment duration data for the five Tunisian chronic HBV infected patients before and during therapy.

Patients	Gender	Age	Viral load (UI/ml)/ year		Date of treatment beginning	Treatment duration (months)	Clinical status	Treatment molecules	Subgenotype	HBsAg	HBeAg
			Before treatment	During treatment							
Patient 1	Male	53	>1,1.10^8/2012	8,85.10^5/2018	10/2012	72	CHB$^+$	ETV*	D7	+	+
Patient 2	Male	N.A	1,85.10^7/2011	6,79.10^4/2017	04/2012	58	HCC^{++} (Deceased in 2020)	ETV* and TDF**	D1	+	+
Patient 3	Male	N.A	2,38.10^6/2006	4,74.10^6/2016	12/2010	68	CHB$^+$	ETV* and TDF**	D1	+	+
Patient 4	Female	33	>1,1.10^8/2016	5,89.10^3/2018	02/2017	14	CHB$^+$	ETV*	D7	+	+
Patient 5	Female	23	3,57.10^6/2012	9,05.10^2/2014	11/2012	16	CHB$^+$	ETV*	D8	+	+

N.A, not available.$^+$CHB, Chronic HBV Infection.
$^{++}$HCC, Hepatocellular Carcinoma.
*ETV, Entecavir.
**TDF, tenofovir disoproxil fumarate.

(TDF), is recommended alone or combined to ETV. However, the TDF is not covered by the NHIF.

The aim of the present study is to analyze the mutational profile through the HBV whole genome in ETV non-responder chronic HBV patients, in order to investigate antiviral drug resistance, immune escape, and liver disease progression to LC or HCC.

Materials and methods

Patients and samples

HBV chronic patients with quantifiable viral load and suspected to be ETV non-responders after viral breakthrough were included in the study. Blood samples were collected from five chronic hepatitis B patients investigated during the routine diagnostic activity of the Laboratory of Clinical Virology in Pasteur Institute of Tunis. For each included patient two plasma samples were collected: one before treatment as part of the pre-treatment diagnostic, and one during the treatment upon request of the treating physician. The period separating the second sample from the date of treatment beginning ranging between 14 and 72 months depending on the time of the viral breakthrough for each patient. Virological and clinical data are shown in Table 1.

Methods

DNA extraction, amplification, and sequencing

DNA was extracted from 200 µl of plasma using the Qiagen QIAamp® DNA extraction kit (QIAGEN® Inc., Hilden, Germany) according to the manufacturer's instructions. Three pairs of primers previously described by Chekaraou et al. (2010) were used to amplify 3 overlapping amplicons of 1,228-bp (nt 2,817–863), 1,253 bp (nt 448–1,701), and 1,653 bp (nt 1,609–80) covering the whole HBV genome as shown in Figure 1.

PCR reactions were performed in 50 µl of reaction mixture containing 1X polymerase buffer, 1.5 mM MgCl2, 0.2 mM dNTPs, 0.2 µM of each primer, 1.25 U of Taq Core MP® (Applied Biosystems) and nuclease-free water. The amount of DNA extract added varied between 10 to 35 µl depending on the viral load. PCR cycling was as follows: 94°C for 5 min, 40 cycles (94°C for 1 min, 56°C/57°C/62.5°C for regions 1, 2, and 3, respectively, 72°C for 1 min) with a final extension step at 72°C for 10 min. PCR products were analyzed by electrophoresis on 1% agarose gels stained with 1,25X of Red gel™ dye Nucleic Acid (Biotium®) and visualized by UV transilluminator.

The purified template DNA was sequenced using a BigDye Terminator Ready Reaction Cycle Sequencing Kit (Applied Biosystems) using the same primers pairs on an ABI Prism 3130 Genetic Analyzer (Applied Biosystems).

HBV genome assembly, genotyping, and subtyping

The obtained overlapping sequences were then assembled using BLAST multiple sequences software by comparison with a reference sequence (M32138).[1] The generated final sequences were submitted to Genbank under accession numbers: MT591274-MT591281 and OP121186.

Sequence alignment was performed with MAFFT online server using default parameters[2] by comparing the obtained

1 https://blast.ncbi.nlm.nih.gov/Blast.cgi.
2 https://mafft.cbrc.jp/alignment/server/

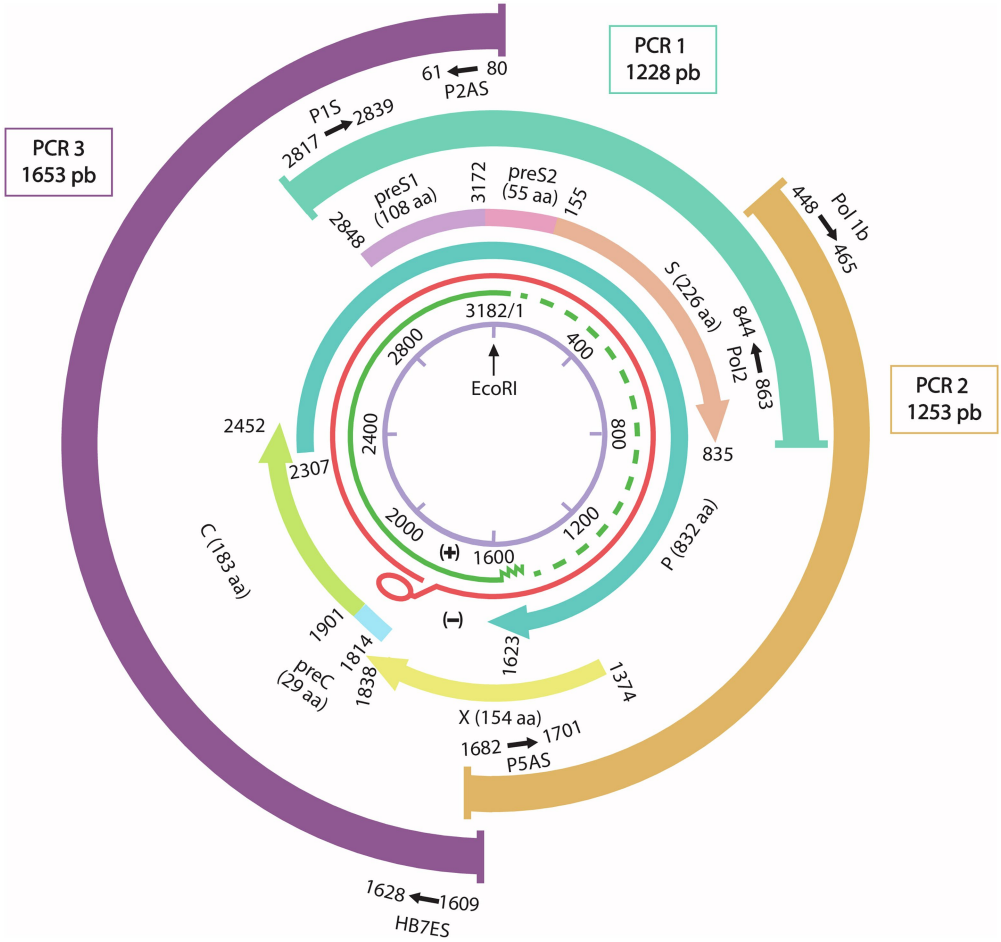

FIGURE 1
Representation of the HBV whole genome with the three overlapping PCR products and their primers sets positions. From inside to outside: schematic representation of the HBV genome with the EcoR1 restriction enzyme site representing nucleotide number 1. In green the incomplete positive strand, in Red the complete negative strand of the HBV genome. The overlapping genes P: Polymerase gene, S: Surface gene, C: capsid gene and X gene. Three overlapping fragments covering the whole genome are represented with the used primers P1S-Pol2 for PCR1 (1,228nt), Pol1b-P5as for PCR2 (1,253bp), and P2AS-HB7ES for PCR3 (1,653bp).

genomic sequences with 58 reference sequences representing the 10 HBV genotypes (A–J) and their corresponding subgenotypes. The resulting alignment was used to build a maximum likelihood phylogenetic tree using the IQ-TREE web server, supported by 1,000 bootstrap replicates.[3] The phylogenetic tree was then visualized using Figtree software.[4] The tree was rooted using the midpoint rooting method. Genotypes were also confirmed by the National Center for Biotechnology Information's (NCBI) E-genotype online software.[5] HBV subtypes were inferred from sequences of the S gene by identifying amino acids (aa) at positions 122, 160, 127, 140, and 159 according to an algorithm previously described (Purdy et al., 2006).

3 http://iqtree.cibiv.univie.ac.at
4 http://tree.bio.ed.ac.uk/software/figtree/
5 https://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi

Mutation analysis

Mutational profiles of the nucleotide or amino acid sequences were determined by comparing each gene (P, S, C, and X) before and during treatment with the corresponding reference sequence using Mega 7.026 (Kumar et al., 2016). Mutations' impacts on treatment, immune response, and liver disease progression were analyzed based on the literature.

Results

HBV whole genome assembly

Whole genome sequences were obtained before and during treatment for 3 patients (1, 2, and 3). For the two remaining patients (4 and 5) we succeeded to obtain the whole genome before treatment. During treatment, the obtained sequence of patient 4 was lacking 408 bp (from nucleotide 45 to nt 453) and for

patient 5 we could not be able to amplify the HBV genome which could be due to the low viral load.

HBV genotyping and subtyping

Phylogenetic analysis (Figure 2) showed that all the sequences belong to genotype D. Subgenotyping showed that patients 1 and 4 were infected with subgenotype D7; patients 2 and 3 with D1 and patient 5 with D8, supported by high bootstrap values: 100, 100, and 85, respectively.

Subtyping showed that the studied HBV strains belonged to the ayw2 subtype based on Arg122, Lys160, Pro127, Tyr140, and Gly159 positions.

Genetic variability in the P, S, C, and X genes

Mutational profile of the RT region in the polymerase gene

The mutational analysis of the RT region revealed a total of 28 aa substitutions ranging between 7 and 12 per patient, among them, several potential resistance-related mutations were detected. Primary mutations (category 1), rtS202G and rtM204V, occurred in patients 2 and 3 during treatment. Secondary/compensatory mutations (category 2), were found in 8 aa replacements; 5 were detected before treatment (rtL91I and rtT128N in patient 2; rtQ149K and rtP237T in patients 1, 4, and 5; rtQ267H in patients 4 and 5) and 3 changes emerged during treatment (rtL180M in patients 2 and 3; rtQ215S and rtF221Y in patient 2).

Three putative antiviral resistance mutations (category 3) were detected: rtR153W in treatment-naïve patients 1, 4, and 5 as well as rtD134E and rtC256S during treatment in patients 3 and 1, respectively. Three pre-treatment mutations (category 4) were also found: rtR110G and rtI266R in patient 1 and rtD263E in patient 5.

Mutations that did not fit categories 1 to 4 were classified into "novel amino acid substitutions" and were observed in 12 aa positions. Six variations namely rtE11D, rtH54Y, rtW257Y, rtD263E, rtQ267Y, and rtE271D were found in treatment-naïve patients and six variations namely rtL145M, rtL260F, rtQ267R, rtK270R, rtM309K, and rtN337T occurred during treatment. The aa changes detected in the RT region of the P gene are mentioned in Table 2; Figure 3.

Mutational analysis of the pre-S/S coding regions

A total of 35 aa substitutions were observed in the whole S gene ranging between 3 and 18 mutations per patient, most of them ($n = 16$) were located in the S region. In the pre-S1 and pre-S2 regions, $n = 9$ and $n = 10$ substitutions were observed, respectively. Mutations detected in the S gene are summarized in Table 3; Figure 3. Out of the 16 aa changes in the S region, 10 were clustered in HBsAg epitopes including B-cells epitopes (aa100-160) as follows: 6 (sN3T, sL42R, sL49R, sT57I, sC76Y, and sQ101H) in HBs1 (upstream of aa120); 1 (sP120T) in HBs2 (aa120–123) and 3 (sT126S, sG130R and sY134F) in HBs3 (aa124–137). Furthermore, the major hydrophilic region (MHR, aa99–169) of the S region had accumulated 5 aa variations of which three were within the HBsAg "a" determinant region (aa124–147).

In addition, nine mutations out of the 16 substitutions in the S gene occurred in different CD4 and CD8 recognition epitopes with the following distribution: 6 aa changes (sL42R, sL49R, sT57I, sS193L, sI195M, and sL216*) in T-helper CD4 epitopes (aa21–65/aa186-197/aa215–223) and 3 (sS207R, sP211R, sL213I) within cytotoxic T lymphocyte CD8 epitopes (aa206-215).

Immune escape mutations (sQ101H, sG130R, sY134F, sS207R, and sL213I) were detected in patients 2 and 5, HBsAg vaccine escape mutations (sP120T, sT126S, sS193L, and sI195M) were found in patients 2 and/or 3, HBIG immunotherapy escape mutations (sP120T and sT126S) were detected in patients 2 and 3, respectively, and misdiagnosis mutations (sP120T, sT126S, sI195M, and sL216stop) were observed in patients 2 and/or 3. Other mutations such as sN3T, sL42R, sC76Y, and sP211R are either not reported or with unknown impacts are also detected in our study.

Mutational analysis of the basal core promotor BCP, precore, and core coding regions

The analysis of the 9 HBV genomic sequences bearing the BCP, pre-C, and Core regions is summarized in Table 4. In the BCP region, 14 different aa mutations were identified over 10 sites ranging between 2 and 8 per patient. Nucleotide mutation G1757A was detected in patients 2, 3, 4, and 5; A1762T, G1764A, and T1753V were detected in patients 1, 3, and 5; C1773T in patients 2 and 3, G1764T/C1766G in patient 2 and C1766T/T1768A in patient 3. The double mutation G1764A/A1762T was found in patients 1, 3, and 5.

In the precore region, 2 mutations were identified: G1896A (patient 2) and G1899A (patients 1, 2, and 3). Whereas, 22 mutations were observed in the core region ranging between 2 and 8 per patient. Among them, several amino acid substitutions were found within different antigen immunogenic epitopes. In particular, 4 aa changes were detected in the T-helper CD4 epitopes (aa 35–45 and 48–69; cE40D/Q, cE64D, cT67N, cA69G), 5 in the B-cell epitopes (aa76–89, 105–116, 130–135; cP79Q, cA80T/S, cV85I, cI116V and cP130Q) and 7 in the CTL CD8 epitopes (aa 18–27, 50–69, 74–83, 141–151; cS21T, cE64D, cT67N, cA69G, cV74S/G, cA80T/S, and cR151Q). The detected nucleotide/aa mutations found in the C gene are presented in Table 4; Figure 3.

Mutational analysis in the X coding region

In total, 30 aa substitutions were found in the X gene region ranging between 4 and 14 aa substitutions per patient of which 29 were before treatment beginning and only 1 was during it. Among

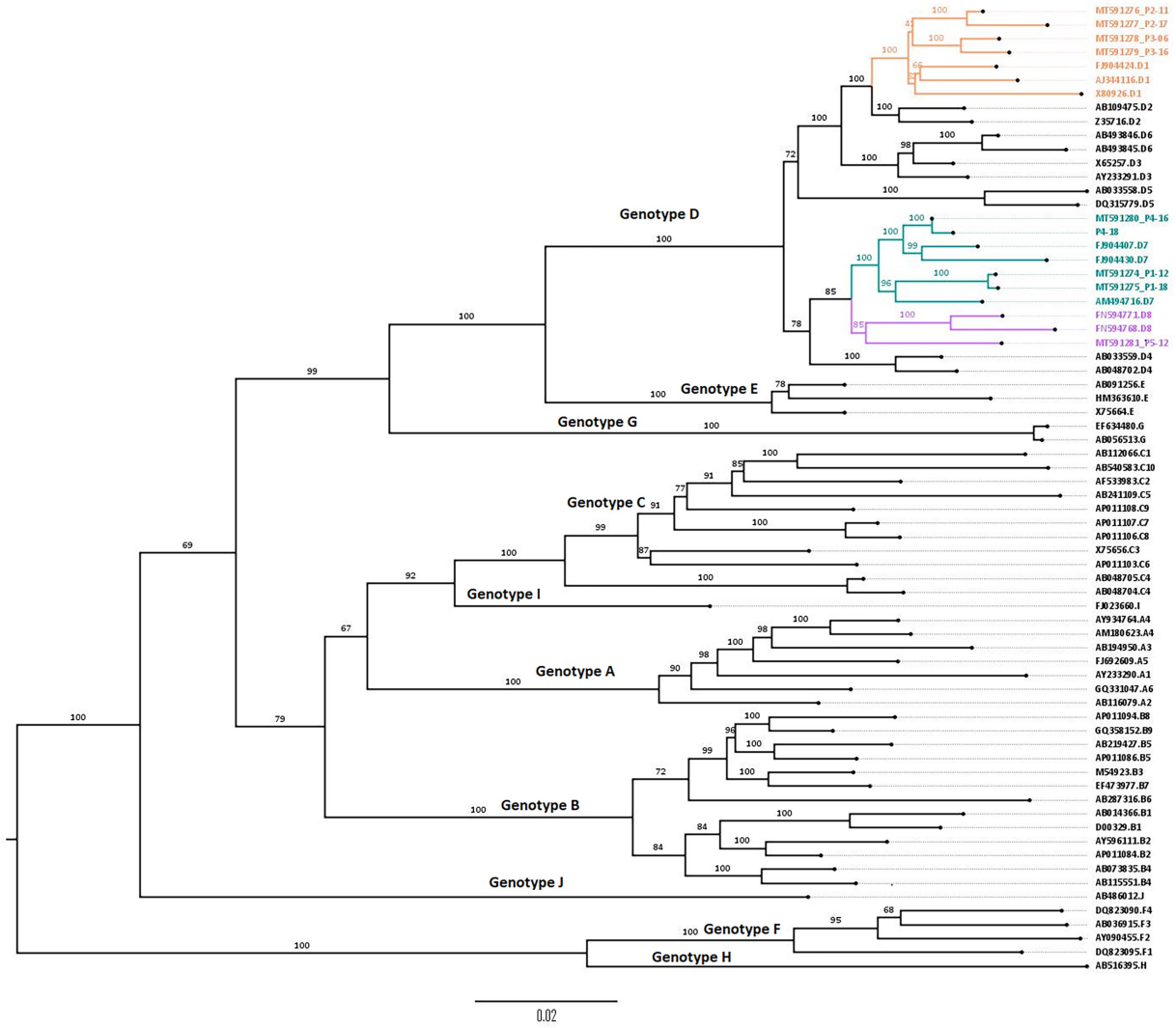

FIGURE 2
Phylogenetic tree constructed using the maximum likelihood method, showing D1, D7, and D8 subgenotypes from 5 Tunisian HBV patients. Phylogenetic tree of obtained genomic sequences with 58 reference sequences representing the 10 HBV genotypes (A–J) and their corresponding subgenotypes. The tree was constructed using the maximum likelihood method using the IQ tree web server and visualized by FigTree. Topology was supported by 1,000 bootstrap replicates. The tree was rooted using the midpoint rooting method.

them 10 variations were detected in the B-cell epitope (aa 29–48) namely: xL34I, xT36D/G, xS38P, xS39P, xP40S, xS41P, xL42P, xS43P, xP46S and xA47T. Four mutations; xK95N, xL98I, xA102, and xT105M; were detected within the T-helper CD4 epitope (aa 91–105) and 2 substitutions (xD119N and xL123W) were detected in CTL CD8 epitope (aa115–123).

As BCP overlaps partially with the HBx coding sequence, mutations at nucleotide positions T1753C/A/G, T1762T, G1764A/T, and C1788G; induce amino acid changes xI127T/D/G, xK130M, xV131I/L and xH139D near the C-terminus of the HBx protein, respectively.

The amino acid substitutions detected in the HBV X gene and their impact are summarized in Table 5; Figure 3.

Discussion

In the present study, we have succeeded to generate the whole HBV genome by amplifying 3 overlapping PCR products covering the entire genome (3.2 kb) using Sanger technology. This technology remains of great importance despite the transition of most laboratories to Next generation sequencing (NGS) technologies. In fact, for small genomes, such as HBV, the Sanger technology is cost effective and more efficient for low viral loads $< 10^3$ IU/ml.

The whole genome was assembled for the 5 patients before the treatment and for 4 patients during the treatment. HBV genome was used for genotyping as well as to study the mutational profile

TABLE 2 Amino acid substitutions detected within the RT region sequences of the five HBV Chronic infected patients with their reported antiviral resistance.

Amino acid substitution	Mutation category	Patients		Drug resistance	Change in overlapping genes	References
		Treatment naïve	During treatment			
E11D	Novel mutation	P4	–	Unknown	N.C	Horikita et al. (1994)
H54Y	Novel mutation	P4/P5	P3	Unknown	N.C	Yang et al. (2002)
N76D	Novel mutation	–	P2	Clinical failure of famciclovir	N.C	Günther et al. (1999); Delaney et al. (2001); Schildgen (2007)
L91I	Secondary/ compensatory	P2	P2	LMV ETV	N.C	Ciftci et al. (2014); Mahabadi et al. (2013); Karatayli et al. (2012); Yamani et al. (2017)
R110G	Pre-treatment	P1	P1	Potential resistance	N.C	Ciftci et al. (2014); Biswas et al. (2013); Azarkar et al. (2018)
T128N	Secondary/ compensatory	P2	P2	LMV	sP120T	Torresi et al. (2002a); Locarnini et al. (2003)
D134E	Putative	–	P3	TDF	sT126S	Liu et al. (2010); Park et al. (2019); Zheng et al. (2012); Choi et al. (2018)
L145M	Novel mutation	–	P4	Unknown	N.C	Katsoulidou et al. (2009)
Q149K	Secondary/ compensatory	P1/P4/P5	P1/P4	Unknown	N.C	Germer et al. (2003)
R153W	Putative	P1/P4/P5	P1/P4	TDF	N.C	Mokaya et al. (2020); Ismail et al. (2011); Li et al. (2012); Mokaya et al. (2019); Olusola et al. (2021)
L180M	Secondary/ compensatory	–	P2/P3	LMV, ETV, LdT, TDF	N.C	He et al. (2015); Choi et al. (2018); Yang et al. (2005)
S202G	Primary	–	P2/P3	LMV, ETV	sS193L	Villet et al. (2007); Mukaide et al. (2010)
M204V	Primary	–	P2/P3	LMV, LdT, ETV, TDF	sI195M	He et al. (2015); Li et al. (2005)
Q215S	Secondary/ compensatory	–	P2	LMV, ADV	sS207R	Shaw et al. (2006); Moriconi et al. (2007); Amini-Bavil-Olyaee et al. (2009); Liu et al. (2009); Wang et al. (2017)
F221Y	Secondary/ compensatory	–	P2	ADV	sL213I	Pollicino et al. (2009); Li et al. (2017); Choi et al. (2018)
P237T	Secondary/ compensatory	P1/P4/P5	P1/P4	ADV	N.A	Pollicino et al. (2009)
C256S	Putative	–	P1	LMV, TDF	N.A	Ciftci et al. (2014); Mokaya et al., (2020); Ciancio et al. (2004)
W257Y	Novel mutation	P2/P3	P2/P3	Unknown	N.A	Ismail et al. (2011)
L260F	Novel mutation	–	P4	Unknown	N.A	Not reported
D263E	Pre-treatment	P5	–	Potential partial resistance to TDF	N.A	Bakhshizadeh et al. (2015)
I266R	Pre-treatment	P1	P1	Unknown	N.A	Westland (2003)
Q267H/R/Y	H:Secondary/ compensatory	P4/P5	P4	H: LMV, LdT	N.A	Qin et al. (2013b)
	R:Novel mutation	–	P2	Unknown	N.A N.A	Qin et al. (2013a)
	Y:Novel mutation	P1	P1	Unknown		Not reported
K270R	Not reported	-	P2	Unknown	N.A	Quiros-Roldan et al. (2008)
E271D	Novel mutation	P4	P4	Unknown	N.A	Quiros-Roldan et al. (2008)
M309K	Novel mutation	–	P3	Unknown	N.A	Wu Y et al. (2014)
N337T	Not reported	–	P2	unknown	N.A	Boyd et al. (2019)

P1–P5 = patients 1–5. NC, no change; mutation is silent in the surface antigen reading frame. NA, not applicable; polymerase substitution is downstream of the surface antigen reading frame. ADV, Adefovir dipivoxil; ETV, Entecavir; LdT, Telbivudine; LMV, Lamivudine; TDF, Tenofovir disoproxil fumarate.

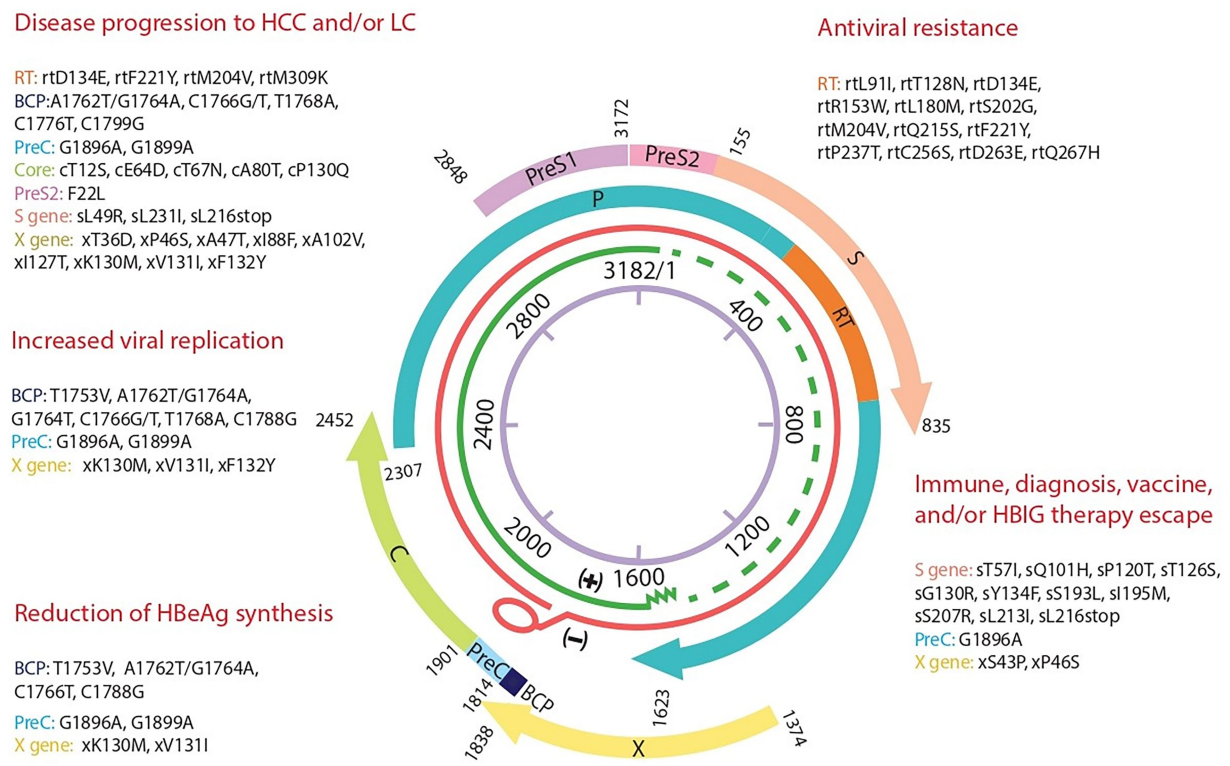

FIGURE 3
Hepatitis B genome map with summary of the most significant mutations and their clinical impact detected in 5 Tunisian HBV patients. A representation of the four overlapping genes encoding the polymerase P, the surface gene S, the capsid gene C, and X gene. The detected significant mutations are grouped depending on their clinical impact based on the literature. BCP, Basal core Promotor; C, core; PreC, precore; S, surface; RT, reverse transcriptase; HBIG, Immunoglobulin.

in all the genes (S, P, X, and C) in order to give scientific proof of antiviral treatment resistance.

Genotyping showed that genotype D was detected in the 5 studied patients. This genotype was previously described as a predominant HBV genotype in Tunisia and the Maghreb region as well as in the Middle East with a low co-circulation rate of genotype E (Ayed et al., 2007; Ezzikouri et al., 2008; Ouneissa et al., 2013).

Subgenotypes D1 and D7, found in the present study, were previously described as the most prevalent subgenotypes circulating in Tunisia (Meldal et al., 2009). However, subgenotype D8 is to our knowledge detected for the first time in Tunisia. This subgenotype has been firstly detected in Niger and has been described as a recombinant strain between genotypes D and E (Chekaraou et al., 2010). The recombination analysis of the detected D8 strain, using the NCBI viral genotyping tool, was in line with the previous findings. Further studies are needed on larger population size to estimate the prevalence of this subgenotype in Tunisia.

In the second part of the present study, we have analyzed the mutational profile of all HBV genes P, S, C, and X.

The mutational profile of the RT region in the P gene showed high genetic variability with 28 different mutations. Before the treatment, 14 aa mutations were detected of which patient 2 had already 2 secondary/compensatory substitutions: rtL91I and T128N, described to be a resistance mutation to ETV and/or to LMV, respectively (Torresi et al., 2002a; Mahabadi et al., 2013; Ziaee et al., 2016). For the remaining patients, four mutations were detected and reported to be resistant to at least one of the following antivirals: rtQ267H in patients 4 and 5 to LMV and LdT; rtP237T in patients 1, 4, and 5 to ADV; rtR153W in patients 1, 4, and 5 in addition to rtD263E in patient 5 potentially to TDF (Pollicino et al., 2009; Qin et al., 2013b; Bakhshizadeh et al., 2015; Mokaya et al., 2020). The eight remaining substitutions were not previously described to have an impact on antiviral treatment.

During the treatment, 14 additional aa substitutions occurred. The most significant ones were rtM204V, rtL180M, and rtS202G detected in patients 2 and 3. Indeed, it has been described that the rtM204V substitution is usually associated with the compensatory mutation rtL180M, which restores the replication capacity of rtM204V mutants (Tenney et al., 2004). Thus, the pattern rtL180M, rtS202G, and rtM204V act synergistically not only to increase viral load but also to reduce treatment susceptibility and confer cross-resistance to ETV, TDF, LMV, and LdT (Kamiya, 2003; Li et al., 2005; He et al., 2015; Mokaya et al., 2020). Other emerged aa variations have been detected in our patients and previously described as resistance mutations that reduce the affinity and susceptibility to antiviral drugs namely rtQ215S and

TABLE 3 Amino acid substitutions within the HBV surface gene sequences from the studied patients with their impact.

Region	Cell subsets	Amino acid substitution	Patients		Effects	References
			Treatment Naive	During treatment		
Pre S1 region		A28T	P1/P5	P1	Unknown	Taghiabadi et al. (2019)
		A28N	P3	P3	Unknown	Feeney et al. (2013)
		T40P	P5	–	Unknown	Pourkarim et al. (2014)
		H60D	P3	–	Unknown	Not reported
		P78T				
		S85C				
		I74L	P2/P3	P2/P3	Unknown	Mondal et al. (2015)
		S90L	P1	P1	Unknown	Not reported
		N103D	P5	P2	Unknown	Mondal et al. (2015)
Pre S2 region		T11N	P3	P3	Unknown	Pourkarim et al. (2014)
		R16K	P2	–	Unknown	Pourkarim et al. (2014)
		R18K	P1	P1/P2	Unknown	Lago et al. (2014)
		F22L	P1	P1/P2	Association with HCC progression	Gopalakrishnan (2013); Chaudhuri et al. (2004)
		N33D	–	P2	Unknown	Kim et al. (2013)
		A39V	P2/P3/P4	P2/P3	Unknown	Pollicino et al. (2007)
		P41H	P2/P3/P5	P2/P3	Unknown	Pollicino et al. (2007)
		I42T	P3	P2/P3	Unknown	Kim et al. (2010)
		F46S	P2	–	Unknown	Olinger et al. (2007)
		P52L	P3	–	Unknown	Pollicino et al. (2007)
S region	Other	N3T	P4	–	Unknown	Not reported
	T-helper (CD4) epitope (aa21-65)	L42R	-	P2	Unknown	Chaouch et al. (2016)
		L49R	P3	–	- Association with LC progression	Chaouch et al. (2016)
		T57I	P5	–	- Reduced HBsAg antigenicity	Duda (2020)
	Other B-cell epitope (aa 100–160)	C76Y	P5	–	Unknown	Wei et al. (2011)
		Q101H	P2	P2	- Immune escape	Tokgöz et al. (2018)
		P120T	P2	P2	- HBIG therapy escape - Misdiagnosis - Vaccine escape - Reduced HbsAg secretion	Amini-Bavil-Olyaee et al. (2010); Bahramali et al. (2008)
		T126S	–	P3	- HBIG therapy escape - Vaccine escape - Misdiagnosis	Moerman et al. (2004); Sitnik et al. (2004)
		G130R	P5		- Immune escape	Kwei et al. (2013); Tokgöz et al. (2018)
		Y134F	P2	–	- Immune escape	Chaouch et al. (2016); Coppola, (2015)
	T-helper (CD4) epitope (aa 186–197)	S193L	–	P2	- Vaccine escape	Aydın et al. (2019); Suntur et al. (2019)
		I195M	–	P2/P3	- Vaccine escape - Misdiagnosis - Reduced *in vitro* affinity to anti-HBs antibodies.	Colagrossi et al. (2018); Torresi et al. (2002b); Araujo et al. (2008)
	CTL (CD8) epitope (aa 206–215)	S207R	–	P2	- Immune escape	Hosseini et al. (2019)
		P211R	–	P4	Unknown	Choga et al. (2020)
		L213I	–	P2	- Immune escape	Hosseini et al. (2019); Datta et al. (2014)
	T-helper (CD4) epitope (aa 215–223)	L216stop	P3	–	- Truncated HbsAg protein - Misdiagnosis - Reduced HbsAg secretion - Association with HCC progression	Araujo et al. (2008); Hosseini et al. (2019)

P1–P5 = Patients 1–5. HCC, hepatocellular carcinoma; LC, liver cirrhosis.

TABLE 4 Amino acid/nucleotide substitutions detected within the BCP, recure, and core sequences of the five chronic HBV infected patients with their impact.

Region	Cell subsets	Substitution		Patients		Effects	References
		Amino acid	Nucleotide	Treatment naïve	During treatment		
Basal core promotor		N.A	A1752G	P3	P3	- Low viral replication capacity	Quarleri (2014); Ng et al. (2005)
		N.A	T1753V (C/A/G)	P1/P5 (C) P3 (A)	P1 (C) P3 (G)	- Increase viral replication - Reduction in HbeAg synthesis	Caligiuri et al. (2016); Parekh et al. (2003)
		N.A	G1757A	P2/P3/P4/P5	P2/P3/P4	- Protection from liver disease progression to LC and/or HCC	Poustchi et al. (2008); Ducancelle et al. (2013); Mohamadkhani et al. (2011)
		N.A	A1762T	P1/P3/P5	P1/P3	- Reduction in HbeAg synthesis	Quarleri (2014); Chen et al. (2005); Leng et al. (2015); Yan et al. (2015); Fang et al. (2008)
		N.A	G1764A	P1/P3/P5	P1	May increase viral transcription and replication - HbeAg seroconversion - Association with liver disease progression to HCC or LC	
		N.A	G1764T	P2	P2	- Increase viral replication - Increase core promoter activity	Sendi et al. (2005); Poustchi et al. (2008)
		N.A	C1766G	P2	P2	- increase core promoter activity - Increase viral replication - Association with liver disease progression to LC	Sendi et al. (2005); Poustchi et al. (2008); Salarneia et al. (2016)
		N.A	C1766T	P3	P3	- Increase viral replication - Reduction in HbeAg synthesis - Association with liver disease progression to HCC and LC	Tong et al. (2013); Kitab et al. (2012); Nishizawa et al. (2016)
		N.A	T1768A	P3	P3	- Increase viral replication - Association with liver disease progression to HCC and LC	Yin et al. (2011); Jammeh et al. (2008); Huang et al. (2011)
		N.A	C1773T	P2/P3	P2/P3	- Association with liver disease progression to HCC and LC	Ghosh et al. (2012); Yin et al. (2011); Gil-García et al. (2019)
		N.A	C1788G	–	P4	- Reduction in HbeAg synthesis -Increase viral replication	Tong et al. (2013)
		N.A	C1799G	P2	P2	- Inversely associated with HCC and significantly associated with LC	Chen et al. (2005); Yin et al. (2011)
Precore		W28stop	G1896A	–	P2	- Inhibition of HbeAg synthesis - Immune escape to anti-Hbe - Increase viral replication - Association with liver progression to LC and HCC	Kargar Kheirabad et al. (2017); Al-Qahtani et al. (2018); Tong et al. (2007)
		G29D	G1899A	P1/P2/P3	P1/P3	- Inhibition of the recognition and cleavage of HbeAg precursor - May increase viral replication - Association with disease progression to LC and HCC	Thompson et al. (2010); Liao et al. (2012); Ouneissa et al. (2012); Al-Qahtani et al. (2018)

(Continued)

TABLE 4 (Continued)

Region	Cell subsets	Substitution		Patients		Effects	References
		Amino acid	Nucleotide	Treatment naïve	During treatment		
Core	Other	T12S	A1934T	P1	P1	- Association with disease severity	Datta et al. (2014); Saha et al. (2014)
	CTL (CD8) epitope (aa 18–27)	S21T	T1961A	P1	P1	Unknown	Sominskaya et al. (2011)
	Other	D29H	G1985C	P5	-	Unknown	Not reported
	T-helper (CD4) epitope (aa 35–45)	E40D	A2020T	P1/P4/P5	P1/P4	Unknown	Pollicino et al. (2007)
		E40Q	G2018C A2020T	–	P3	Unknown	Homs et al. (2012)
	CTL (CD8) epitope (aa 50–69) + T-helper (CD4) epitope (aa 48–69)	E64D	A2092C	P3/P4	P2/P4	- Association with disease progression to LC and HCC - Reduction in T-cell proliferation in association with T67N	Pollicino et al. (2007); Al-Qahtani et al. (2018); Homs et al. (2011)
		T67N	C2100A	P4	P4	- Same effects as E64D	Datta et al. (2014); Saha et al. (2014); Pollicino et al. (2007); Homs et al. (2011)
		A69G	C2106G	P4	P4	Unknown	Sominskaya et al. (2011)
	CTL (CD8) epitope (aa 74–83)	V74G	T2121G	P2/P3	P3	- Reduction in HBe and HBc antigenicity	Pollicino et al. (2007); Homs et al. (2012)
		V74S	G2120A T2121G	–	P2	Unknown	Not reported
	B-cell epitope (aa 76–89)	P79Q	C2136A	P1	P1	- Reduction in HBe and HBc antigenicity	Pollicino et al. (2007); Huang et al. (2014)
		A80T	G2138A	P2/P3	P2/P3	- Truncated HBcAg protein → Negativity for anti-HBc. - Reduction in HBe and HBc antigenicity. Association with disease progression to HCC or LC	Pollicino et al. (2007); Bajpai et al. (2017); Al-Qahtani et al. (2018)
		A80S	G2138T	P1	P1	Unknown	Not reported
		V85I	G2153A	–	P2	Unknown	Pollicino et al. (2007)
	Other	M93V	A2177G	P3	-	Unknown	Al-Qahtani et al. (2018)
	B-cell epitope (aa 105–116)	I116V	A2246G	P1	P1	Unknown	Pollicino et al. (2007)
	B-cell epitope (aa 130–135)	P130Q	C2289A	–	P2	- Association with disease progression to HCC or LC	Datta et al. (2014); Pollicino et al. (2007)
	CTL (CD8) epitope (aa 141–151)	R151Q	G2352A	P1	P1	Unknown	Not reported
	Other	G153C	G2357T	–	P2	Unknown	Wu J et al. (2014)
		S155T	T2363A	–	P2	Unknown	Pollicino et al. (2007)
		P156S	C2366T	P1	P1	Unknown	Not reported
		R166P	C2366T	–	P4	Unknown	Not reported

P1–P5 = Patients 1–5. HCC, Hepatocellular carcinoma; LC, liver cirrhosis.

TABLE 5 Amino acid/nucleotide substitutions detected within the X gene sequences of the patients with their reported effects.

Cell subsets	Aa substitution	Nucleotide mutation	Patients		Effects	References
			Treatment naïve	During treatment		
Other	C26S	T1449A	P2	P2	Unknown	Not reported
	C26R	T1449C	P3	P3	Unknown	Pollicino et al. (2007)
B-cell epitope (aa 29–48)	L34I	C1473A	P5	–	Unknown	Not reported
	T36D	A1479G C1480A	P1/P4	P1/P4	- Association with HCC progression	Pollicino et al. (2007); Sominskaya et al. (2011); Javanmard et al. (2020)
	T36G	A1479G C1480G	P5	–	Unknown	Not reported
	S38P	T1485C	P1	P1	Unknown	Mani et al. (2019)
	S39P	T1488C	P5	–	Unknown	Pollicino et al. (2007)
	P40S	C1491T	P1/P4	P1/P4	Unknown	León et al. (2005)
	S41P	T1494C	P5	–	Unknown	Xu et al. (2007)
	L42P	T1498C	P2/P3	P2/P3	Unknown	Not reported
	S43P	T1500C	P1/P4/P5	P1/P4	- Immune escape (B-cell epitope affected)	Putri et al. (2019); Wang et al. (2012); Li et al. (2018)
	P46S	C1509T	P2/P3	P2/P3	- Immune escape (B-cell epitope affected) - Association with HCC progression	Pollicino et al. (2007); Li et al. (2018)
	A47T	G1512A	P2/P3/P5	P2/P3	- Association with HCC progression	Al-Qahtani et al. (2017); Artarini et al. (2016)
Other	T82S	A1617T	P2	P2	Unknown	Not reported
	H86R	A1630G	P5	–	Unknown	Huang et al. (2014)
	I88F	A1635T	P2	P3	- Association with HCC progression	Javanmard et al. (2020); Pollicino et al. (2007)
	I88C	A1635T T1636G	P3	P2	Unknown	Abdel Hamid and Salama (2018)
T-helper (CD4) epitope (aa 91–105)	K95N	G1658C	P1	P1	Unknown	Not reported
	L98I	C1665A	P1	P1	Unknown	Pollicino et al. (2007)
	A102V	C1678T	P2/P3	P2/P3	- Association with HCC progression	Ghosh et al. (2012); Mani et al. (2019); Pollicino et al. (2007)
	T105M	C1687T	P3	P3	Unknown	Mani et al. (2019)
CTL (CD8) epitope (aa 115–123)	D119N	G1728A	P3	–	Unknown	Zhu et al. (2008)
	L123W	T1741G	P3	–	Unknown	Not reported
Other	I127T	T1753C	P1/P5	P1	- Association with HCC progression - Promote transactivation and increase anti-proliferative activity	Al-Qahtani et al. (2017); Artarini et al. (2016); Lin et al. (2005); Elkady et al. (2008)
	I127D/G	T1753A/G	P3 (D)	P3 (G)	Unknown	Not reported
	K130M	A1762T	P1/P3/P5	P1/P3	- Increase viral replication and cell invasion - Decrease the expression of HBeAg - Association with HCC progression	Mani et al. (2019); Lin et al. (2005); Yuan et al. (2009)

(Continued)

TABLE 5 (Continued)

Cell subsets	Aa substitution	Nucleotide mutation	Patients		Effects	References
			Treatment naïve	During treatment		
	V131I	G1764A	P1/P3/P5	P1	- Increase viral replication and cell invasion - Decrease the expression of HBeAg - Association with disease progression	Al-Qahtani et al. (2017); Mani et al. (2019); Lin et al. (2005); Kim et al. (2016)
	V131L	G1764T	P2	P2	Unknown	Pollicino et al. (2007)
	F132Y	T1768A	P3	P3	- Increase viral replication and cell invasion - Association with HCC progression	Pollicino et al. (2007); Al-Qahtani et al. (2017); Mani et al. (2019)
	H139D	C1788G	–	P4	Unknown	Not reported

P1–P5 = Patients 1–5. HCC, Hepatocellular carcinoma.

rtC256S to LMV; rtQ215S and rtF221Y to ADV; rtD134E, rtQ215S, and rtC256S to TDF (Moriconi et al., 2007; Amini-Bavil-Olyaee et al., 2009; Liu et al., 2009; Pollicino et al., 2009; Ciftci et al., 2014; Park et al., 2019; Mokaya et al., 2020).

Thus, our results support the need to introduce HBV genome sequencing as a pre-treatment diagnosis to predict potential resistance to available antiviral molecules, as well as to monitor the evolution of treatment response.

In addition, we have studied the mutational profile in the preS1, preS2, and S genes. As the coding sequence of the HBsAg is completely overlapped with the RT domain of the HBV polymerase, some mutations occurring in the RT region may lead to the emergence of escape mutants in the S region and vice versa. Thus, rtT128N, rtD134E, rtS202G, rtM204V, rtQ215S and rtF221Y substitutions observed in the RT region result in sP120T, sT126S, sS193L sI195M, sS207R and sL213I in the HBsAg gene, respectively. These mutations in addition to sT57I, sQ101H, sG130R, sY134F, and sL216stop could alter the antigenicity of HBsAg and reduce its expression and/or recognition by antibodies. Therefore, they could induce immune, vaccine, HBIG therapy, and/or diagnosis escape as well as influence HBsAg expression and treatment efficacy (Moerman et al., 2004; Sitnik et al., 2004; Bahramali et al., 2008; Amini-Bavil-Olyaee et al., 2010; Coppola, 2015; Ziaee et al., 2016; Rendon et al., 2017; Tokgöz et al., 2018; Aydın et al., 2019; Hosseini et al., 2019; Duda, 2020).

Regarding the mutational profile of BCP (nt 1,742–1,849), precore (nt 1,814–1,900), and core regions that code HBeAg and HBcAg proteins, the double mutants A1762T/G1764A, G1764T/C1766G and C1766T/T1768A, as well as the single mutations A1752G, T1753V (C/A/G), C1766T and C1788G, detected in BCP region, have been reported to enhance viral replication and/or reduce HBeAg synthesis by suppressing the transcription of the pre-C region (Parekh et al., 2003; Sendi et al., 2005; Poustchi et al., 2008; Tong et al., 2013; Caligiuri et al., 2016; Lazarevic et al., 2019). The single nucleotide mutations G1896A and G1899A in the precore region have been suggested to be mutational hotspots occurring most frequently in genotype D and were previously reported in Tunisian studies with an occurrence alone or in association (Triki et al., 2000; Bahri et al., 2006; Ayed et al., 2007; Poustchi et al., 2008; Ouneissa et al., 2012). These mutants result in a stop codon at position W28* and a substitution at position G29D, respectively, leading to the production of a truncated precore protein and then the abolition of HBeAg expression (Kobayashi et al., 2003; Thompson et al., 2010; Ducancelle et al., 2016). These variations are the major immune escape mutants of HBV as HBeAg is the main target for both cellular and humoral immune responses leading to a higher risk of liver HCC and LC progression (Tong et al., 2005; Liao et al., 2012; Suppiah et al., 2015; Pahal et al., 2016). In addition, precore mutants impose serious consequences on the treatment and enhance viral replication (Ouneissa et al., 2012; Kargar Kheirabad et al., 2017; Boyd et al., 2018).

Concerning the core mutations, cT67N within the T-helper CD4 epitope might be able to escape the host immune response (Datta et al., 2014; Saha et al., 2014). Moreover, cV74G, cP79Q, and cA80T mutations are known to reduce both HBe and HBc antigenicity (Pollicino et al., 2007; Huang et al., 2014). In addition, cA80T has resulted in the production of altered and truncated HBcAg protein leading potentially to abnormal immune reaction and negativity of anti-HBc (Bajpai et al., 2017).

In the last part of this study, we studied the mutational profile in the X gene. Substitutions xS43P and xP46S located in the B-cell epitope were detected in our study and have been suggested to be related with immune escape (Putri et al., 2019). Mutations xP46S, xA47T, xI88F, xA102V, xI127T, xK130M, xV131I, and xF132Y, were previously reported as significant HCC-related HBx mutants alone or combined such as (I127T + K130M + V131I) in patients 1, 3 and 5 and (xK130M + xV131I + xF132Y) in patient 3 (Pollicino et al., 2007; Ghosh et al., 2012; Ali et al., 2014; Al-Qahtani et al., 2017). Moreover, the double mutant xK130M + xV131I has been suggested to exacerbate the host's immune response, increase viral replication, and lead to a truncated HBx protein (Wungu et al., 2019). In addition, it is associated with the activation of proto-oncogenes and inactivation of the tumor suppressor gene

leading to a rapid progression of liver cirrhosis and/or HCC cell invasion and metastasis (Wang et al., 2016).

Several mutations previously reported to be significantly associated with an increased risk of severe liver disease progression to HCC and/or LC progression were also detected in other genes namely (rtD134E/rtF221Y/rtM204V/rtM309k) in the RT region; (sF22L) in the preS2 region; (sL49R, sL213I and sL216*) in the S region; (C1766T/T1768A double mutant, C1773T, C1799G, and C1766G) in the BCP region; and (cT12S/cE64D/cT67N/cA80T/cP130Q) in the core region of the C gene. These HCC-related mutations could be used as markers of HCC evolution in particular rtF221Y mutant which has been indicated as an independent risk factor for poor overall survival (Jammeh et al., 2008; Yin et al., 2011; Kitab et al., 2012; Zheng et al., 2012; Gopalakrishnan, 2013; Tong et al., 2013; Datta et al., 2014; Chaouch et al., 2016; Nishizawa et al., 2016; Kim et al., 2017; Li et al., 2017; Al-Qahtani et al., 2018; Choi et al., 2018; Hosseini et al., 2019). In contrast, the early development of G1757A in the BCP reduces the oncogenic potential of HBV suggesting that it might be a protective biomarker in chronic hepatitis B (Poustchi et al., 2008; Mohamadkhani et al., 2011; Ducancelle et al., 2013).

In addition to the commonly mentioned substitutions in all genes (P, S, C, and X), several nucleotide/amino acid substitutions have been detected in our patients (see Tables 2–5) but have never been reported previously or have been reported with unknown impact. Therefore, further studies are necessary to better understand and elucidate the effect of these mutations on HBV treatment, antigenicity, and disease evolution.

Conclusion

In conclusion, we would propose the whole genome sequencing as a pre-treatment diagnosis to predict potential resistance to available antiviral molecules, as well as to monitor the evolution of treatment response and prevent progression to cirrhosis or hepatocellular carcinoma. Thus, this could contribute to guiding national efforts to optimize relevant HBV treatment management in order to achieve the global hepatitis elimination goal by 2030.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

ZB, AC, HTr, MG, SA, LH, and MM: conceptualization. ZB, HTo, AS, WH, WK, and LY: methodology. ZB, AC, and HTr: validation. ZB and AC: formal analysis. ZB, AC, HTo, AS, WH, WK, and LY: investigation. ZB, AC, MG, SA, LH, and MM: data curation. ZB and KA: writing—original draft preparation. AC and HTr: editing and reviewing. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the Tunisian Ministry of Higher Education [Programme d'Encouragement des Jeunes Chercheurs PEJC, 1ère Edition (2018; project code: 18PJEC07-09)], the Research Laboratory LR20IPT02: "Virus, Vectors and Hosts: One Health approach and technological innovation for a better health," and the Clinical Investigation Center (CIC).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdel Hamid, M., and Salama, A. (2018). X Gene variability and genotyping of hepatitis B virus (HBV) in sudanese patients with liver diseases, Khartoum, Sudan. Available at: https://www.researchgate.net/publication/325477946_X_GENE_VARIABILITY_AND_GENOTYPING_OF_HEPATITIS_B_VIRUS_HBV_IN_SUDANESE_PATIENTS_WITH_LIVER_DISEASES_KHARTOUM_SUDAN (Accessed October 22, 2020).

Al-Qahtani, A. A., Al-Anazi, M. R., Nazir, N., Ghai, R., Abdo, A. A., Sanai, F. M., et al. (2017). Hepatitis B virus (HBV) X gene mutations and their association with liver disease progression in HBV-infected patients. *Oncotarget* 8, 105115–105125.

Ali, A., Abdel-Hafiz, H., Suhail, M., Al-Mars, A., Zakaria, M. K., Fatima, K., et al. (2014). Hepatitis B virus, HBx mutants and their role in hepatocellular carcinoma. *World J. Gastroenterol.* 20, 10238–10248. doi: 10.3748/wjg.v20.i30.10238

Al-Qahtani, A. A., Al-Anazi, M. R., Nazir, N., Abdo, A. A., Sanai, F. M., Al-Hamoudi, W. K., et al. (2018). The correlation between Hepatitis B virus Precore/Core mutations and the progression of severe liver disease. *Front. Cell. Infect. Microbiol.* 8. doi: 10.3389/fcimb.2018.00355

Amini-Bavil-Olyaee, S., Herbers, U., Mohebbi, S. R., Sabahi, F., Zali, M. R., Luedde, T., et al. (2009). Prevalence, viral replication efficiency and antiviral drug susceptibility of rtQ215 polymerase mutations within the hepatitis B virus genome. *J. Hepatol.* 51, 647–654. doi: 10.1016/j.jhep.2009.04.022

Amini-Bavil-Olyaee, S., Vucur, M., Luedde, T., Trautwein, C., and Tacke, F. (2010). Differential Impact of Immune Escape Mutations G145R and P120T on the Replication of Lamivudine-Resistant Hepatitis B Virus e Antigen-Positive and -Negative Strains. *J. Virol.* 84, 1026–1033. doi: 10.1128/JVI.01796-09

Araujo, N. M., Branco-Vieira, M., Silva, A. C. M., Pilotto, J. H., Grinsztejn, B., de Almeida, A. J., et al. (2008). Occult hepatitis B virus infection in HIV-infected patients: evaluation of biochemical, virological and molecular parameters. *Hepatol. Res.* 38, 1194–1203. doi: 10.1111/j.1872-034X.2008.00392.x

Artarini, A., Geby Jessica, H., Rini Kartikasari, R., Riani, C., and Soefie Retnoningrum, D. (2016). Detection of Hepatitis B virus X gene mutation from local clinical samples. *Microbiol Indones.* 10, 9–14. doi: 10.5454/mi.10.1.2

Aydın, M., Tekin, S., Sayan, M., and Akhan, S. (2019). Molecular characterization of Hepatitis B virus strains isolated from chronic Hepatitis B patients in southeastern region of Turkey. *Viral Hepat. J.* 25, 40–44. doi: 10.4274/vhd.galenos.2019.2019.0015

Ayed, K., Gorgi, Y., Ayed-Jendoubi, S., Aouadi, H., Sfar, I., Najjar, T., et al. (2007). Hepatitis B virus genotypes and precore/core-promoter mutations in Tunisian patients with chronic hepatitis B virus infection. *J. Infect.* 54, 291–297. doi: 10.1016/j.jinf.2006.05.013

Azarkar, Z., Ziaee, M., Ebrahimzadeh, A., Sharifzadeh, G., and Javanmard, D. (2018). Epidemiology, risk factors, and molecular characterization of occult hepatitis B infection among anti-hepatitis B core antigen alone subjects. *J. Med. Virol.* 91.

Bahramali, G., Sadeghizadeh, M., Amini-Bavil-Olyaee, S., Alvaian, S. M., Behzad-Behbahani, A., Adeli, A., et al. (2008). Clinical, virologic and phylogenetic features of hepatitis B infection in Iranian patients. *World J. Gastroenterol.* 14, 5448–5453. doi: 10.3748/wjg.14.5448

Bahri, O., Cheikh, I., Hajji, N., Djebbi, A., Maamouri, N., Sadraoui, A., et al. (2006). Hepatitis B genotypes, precore and core promoter mutants circulating in Tunisia. *J. Med. Virol.* 78, 353–357. doi: 10.1002/jmv.20554

Bajpai, V., Gupta, E., Kundu, N., Sharma, S., and Shashtry, S. M. (2017). Hepatitis B core antibody negativity in a chronic hepatitis b infected patient: report of an unusual serological pattern. *J. Clin. Diagn. Res.* 11, DD04–DD06. doi: 10.7860/JCDR/2017/26821.10498

Bakhshizadeh, F., Hekmat, S., Keshvari, M., Alavian, S. M., Mostafavi, E., Keivani, H., et al. (2015). Efficacy of tenofovir disoproxil fumarate therapy in nucleoside-analogue naive Iranian patients treated for chronic hepatitis B. *Hepat. Mon.* 15. doi: 10.5812/hepatmon.15(5)2015.25749

Biswas, A., Panigrahi, R., Chandra, P. K., Banerjee, A., Datta, S., Pal, M., et al. (2013). Characterization of the occult hepatitis B virus variants circulating among the blood donors from eastern India. *Sci. World J.* doi: 10.1155/2013/212704

Boyd, A., Kouamé, M. G., Houghtaling, L., Moh, R., Gabillard, D., Maylin, S., et al. (2019). Hepatitis B virus activity in untreated hepatitis B e antigen-negative human immunodeficiency virus-hepatitis B virus co-infected patients from sub-Saharan Africa. *Trans. R. Soc. Trop. Med. Hyg.* 113, 437–445.

Boyd, A., Moh, R., Maylin, S., Abdou Chekaraou, M., Mahjoub, N., Gabillard, D., et al. (2018). Precore G1896A mutation is associated with reduced rates of HBsAg seroclearance in treated HIV hepatitis B virus co-infected patients from Western Africa. *J. Viral Hepat.* 25, 1121–1131. doi: 10.1111/jvh.12914

Caligiuri, P., Cerruti, R., Icardi, G., and Bruzzone, B. (2016). Overview of hepatitis B virus mutations and their implications in the management of infection. *World J. Gastroenterol.* 22, 145–154. doi: 10.3748/wjg.v22.i1.145

Chaouch, H., Taffon, S., Villano, U., Equestre, M., Bruni, R., Belhadj, M., et al. (2016). Naturally occurring surface antigen variants of Hepatitis B virus in Tunisian patients. *Intervirology* 59, 36–47. doi: 10.1159/000445894

Chaudhuri, V., Tayal, R., Nayak, B., Acharya, S. K., and Panda, S. K. (2004). Occult hepatitis B virus infection in chronic liver disease: full-length genome and analysis of mutant surface promoter. *Gastroenterology* 127, 1356–1371. doi: 10.1053/j.gastro.2004.08.003

Chekaraou, M. A., Brichler, S., Mansour, W., Gal, F., Garba, A., Dény, P., et al. (2010). A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *J. Gen. Virol.* 91, 1609–1620. doi: 10.1099/vir.0.018127-0

Chen, C. H., Lee, C. M., Lu, S. N., Changchien, C. S., Eng, H. L., Huang, C. M., et al. (2005). Clinical significance of hepatitis B virus (HBV) genotypes and precore and core promoter mutations affecting HBV e antigen expression in Taiwan. *J. Clin. Microbiol.* 43, 6000–6006. doi: 10.1128/JCM.43.12.6000-6006.2005

Choga, W. T., Anderson, M., Zumbika, E., Phinius, B. B., Mbangiwa, T., Bhebhe, L. N., et al. (2020). In Silico prediction of human leukocytes antigen (HLA) class II binding Hepatitis B virus (HBV) peptides in Botswana. *Viruses* 12:731. doi: 10.3390/v12070731

Choi, Y. M., Lee, S. Y., and Kim, B. J. (2018). Naturally occurring hepatitis B virus reverse transcriptase mutations related to potential antiviral drug resistance and liver disease progression. *World J. Gastroenterol.* 24, 1708–1724. doi: 10.3748/wjg.v24.i16.1708

Ciancio, A., Smedile, A., Rizzetto, M., Lagget, M., Gerin, J., and Korba, B. (2004). Identification of HBV DNA Sequences That Are Predictive of Response to Lamivudine Therapy. *Hepatology* 39, 64–73. doi: 10.1002/hep.20019

Ciftci, S., Keskin, F., Cakiris, A., Akyuz, F., Pinarbasi, B., Abaci, N., et al. (2014). Analysis of potential antiviral resistance mutation profiles within the HBV reverse transcriptase in untreated chronic hepatitis B patients using an ultra-deep pyrosequencing method. *Diagn. Microbiol. Infect. Dis.* 79, 25–30. doi: 10.1016/j.diagmicrobio.2014.01.005

Colagrossi, L., Hermans, L. E., Salpini, R., Di Carlo, D., Pas, S. D., Alvarez, M., et al. (2018). Immune-escape mutations and stop-codons in HBsAg develop in a large proportion of patients with chronic HBV infection exposed to anti-HBV drugs in Europe. *BMC Infect. Dis.* 18:251. doi: 10.1186/s12879-018-3161-2

Coppola, N. (2015). Clinical significance of hepatitis B surface antigen mutants. *World J. Hepatol.* 7:2729. doi: 10.4254/wjh.v7.i27.2729

Datta, S., Ghosh, A., Dasgupta, D., Ghosh, A., Roychoudhury, S., Roy, G., et al. (2014). Novel point and combo-mutations in the genome of hepatitis B virus-genotype D: characterization and impact on liver disease progression to hepatocellular carcinoma. *PLoS One* 9. doi: 10.1371/journal.pone.0110012

Delaney, W. E., Locarnini, S., and Shaw, T. (2001). Resistance of hepatitis B virus to antiviral drugs: Current aspects and directions for future investigation. *Antivir. Chem. Chemother.* 12, 1–35. doi: 10.1177/095632020101200101

Ducancelle, A., Abgueguen, P., Birguel, J., Mansour, W., Pivert, A., le Guillou-Guillemette, H., et al. (2013). High Endemicity and Low Molecular Diversity of Hepatitis B Virus Infections in Pregnant Women in a Rural District of North Cameroon. *PLoS One* 8:80346

Ducancelle, A., Pivert, A., Bertrais, S., Boursier, J., Balan, V., Veillon, P., et al. (2016). Different precore/core mutations of hepatitis B interact with, limit, or favor liver fibrosis severity. *J. Gastroenterol. Hepatol.* 31, 1750–1756. doi: 10.1111/jgh.13338

Duda, A. (2020). Influence de la variabilité des glycoprotéines d'enveloppe du virus de l'hépatite B sur la clairance de l'Ag HBs chez des patients co-infectés par le VIH suivis au CHU de Nancy. Available at: http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm (Accessed September 29, 2020).

Elkady, A., Tanaka, Y., Kurbanov, F., Oynsuren, T., and Mizokami, M. (2008). Virological and clinical implication of core promoter C1752/V1753 and T1764/G1766 mutations in hepatitis B virus genotype D infection in Mongolia. *J. Gastroenterol. Hepatol.* 23, 474–481. doi: 10.1111/j.1440-1746.2008.05321.x

Ezzikouri, S., Chemin, I., Chafik, A., Wakrim, L., Nourlil, J., Malki, A., et al. (2008). Genotype determination in Moroccan hepatitis B chronic carriers. *Infect. Genet. Evol.* 8, 306–312. doi: 10.1016/j.meegid.2008.01.010

Fang, Z. L., Sabin, C. A., Dong, B. Q., Ge, L. Y., Wei, S. C., Chen, Q. Y., et al. (2008). HBV A1762T, G1764A mutations are a valuable biomarker for identifying a subset of male HBsAG carriers at extremely high risk of hepatocellular carcinoma: a prospective study. *Am. J. Gastroenterol.* 103, 2254–2262. doi: 10.1111/j.1572-0241.2008.01974.x

Feeney, S. A., Mccaughey, C., Watt, A. P., Agnaf, M. R. E., Mcdougall, N., Wend, U. C., et al. (2013). Reactivation of occult hepatitis B virus infection following cytotoxic lymphoma therapy in an anti-HBc negative patient. *J. Med. Virol.* 85, 597–601. doi: 10.1002/jmv.23513

Germer, J. J., Charlton, M. R., Ishitani, M., Forehand, C. D., and Patel, R. (2003). Characterization of Hepatitis B Virus Surface Antigen and Polymerase Mutations in Liver Transplant Recipients Pre- and Post-Transplant. *Am. J. Transplant.* 3, 743–753. doi: 10.1034/j.1600-6143.2003.00149.x

Ghosh, S., Mondal, R. K., Banerjee, P., Nandi, M., Sarkar, S., Das, K., et al. (2012). Tracking the naturally occurring mutations across the full-length genome of hepatitis B virus of genotype D in different phases of chronic e-antigen-negative infection. *Clin. Microbiol. Infect.* 18, E412–E418. doi: 10.1111/j.1469-0691.2012.03975.x

Gil-García, A. I., Madejón, A., Francisco-Recuero, I., López-López, A., Villafranca, E., Romero, M., et al. (2019). Prevalence of hepatocarcinoma-related hepatitis B virus mutants in patients in grey zone of treatment. *World J. Gastroenterol.* 25, 5883–5896. doi: 10.3748/wjg.v25.i38.5883

Gopalakrishnan, D. (2013). Hepatitis B virus subgenotype A1 predominates in liver disease patients from Kerala, India. *World J. Gastroenterol.* 19:9294. doi: 10.3748/wjg.v19.i48.9294

Günther, S., von Breunig, F., Santantonio, T., Jung, M. C., Gaeta, G. B., Fischer, L., et al. (1999). Absence of mutations in the YMDD motif/B region of the hepatitis B

virus polymerase in famciclovir therapy failure. *J. Hepatol.* 30, 749–754. doi: 10.1016/S0168-8278(99)80124-X

He, X., Wang, F., Huang, B., Chen, P., and Zhong, L. (2015). Detection and analysis of resistance mutations of hepatitis B virus. *Int. J. Clin. Exp. Med.* 8, 9630–9639. PMID: 26309637

Homs, M., Buti, M., Tabernero, D., Quer, J., Sanchez, A., Corral, N., et al. (2012). Quasispecies dynamics in main core epitopes of hepatitis B virus by ultra-deep-pyrosequencing. *World J. Gastroenterol.* 18, 6096–6105. doi: 10.3748/wjg.v18.i42.6096

Homs, M., Jardi, R., Buti, M., Schaper, M., Tabernero, D., Fernandez-Fernandez, P., et al. (2011). HBV core region variability: effect of antiviral treatments on main epitopic regions. *Antivir. Ther.* 16, 37–49. doi: 10.3851/IMP1701

Horikita, M., Itoh, S., Yamamoto, K., Shibayama, T., Tsuda, F., and Okamoto, H. (1994). Differences in the entire nucleotide sequence between hepatitis B virus genomes from carriers positive for antibody to hepatitis B e antigen with and without active disease. *J. Med. Virol.* 44, 96–103. doi: 10.1002/jmv.1890440118

Hosseini, S. Y., Sanaei, N., Fattahi, M. R., Malek-Hosseini, S. A., and Sarvari, J. (2019). Association of HBsAg mutation patterns with hepatitis B infection outcome: Asymptomatic carriers versus HCC/cirrhotic patients. *Ann. Hepatol.* 18, 640–645. doi: 10.1016/j.aohep.2018.12.006

Huang, Y., Tong, S., Tai, A. W., Hussain, M., and Lok, A. S. F. (2011). Hepatitis B virus core promoter mutations contribute to hepatocarcinogenesis by deregulating SKP2 and its target, p21. *Gastroenterology* 141, 1412–1421.e5. doi: 10.1053/j.gastro.2011.06.048

Huang, F. Y., Wong, D. K. H., Seto, W. K., Zhang, A. Y., Lee, C. K., Lin, C. K., et al. (2014). Sequence variations of full-length hepatitis B virus genomes in Chinese patients with HBsAg-negative hepatitis B infection. *PLoS One* 9. doi: 10.1371/journal.pone.0115743

Ismail, A. M., Samuel, P., Eapen, C. E., Kannangai, R., and Abraham, P. (2011). Antiviral resistance mutations and genotype-associated amino acid substitutions in treatment-Naïve Hepatitis B virus-infected individuals from the Indian subcontinent. *Intervirology* 55, 36–44.

Jammeh, S., Tavner, F., Watson, R., Thomas, H. C., and Karayiannis, P. (2008). Effect of basal core promoter and pre-core mutations on hepatitis B virus replication. *J. Gen. Virol.* 89, 901–909. doi: 10.1099/vir.0.83468-0

Javanmard, D., Niya, M. H. K., Kalafkhany, D., Najafi, M., Ziaee, M., Babaei, M. R., et al. (2020). Downregulation of gsk3β and upregulation of urg7 in hepatitis b-related hepatocellular carcinoma. *Hepat. Mon.* 20, 1–11.

Kamiya, N. (2003). The mechanisms of action of antivirals against hepatitis B virus infection. *J. Antimicrob. Chemother.* 51, 1085–1089. doi: 10.1093/jac/dkg236

Karatayli, E., Karatayli, S. C., Cinar, K., Gokahmetoglu, S., Güven, K., Idilman, R., et al. (2012). Molecular characterization of a novel entecavir mutation pattern isolated from a multi-drug refractory patient with chronic hepatitis B infection. *J. Clin. Virol.* 53, 130–134. doi: 10.1016/j.jcv.2011.10.011

Kargar Kheirabad, A., Farshidfar, G., Nasrollaheian, S., and Gouklani, H. (2017). Prevalence and Characteristics of Precore Mutation in Iran and Its Correlation with Genotypes of Hepatitis B. *Electron. Physician* 9, 4114–4123. doi: 10.19082/4114

Katsoulidou, A., Paraskevis, D., Magiorkinis, E., Moschidis, Z., Haida, C., Hatzitheodorou, E., et al. (2009). Molecular characterization of occult hepatitis B cases in Greek blood donors. *J. Med. Virol.* 81, 815–825. doi: 10.1002/jmv.21499

Kim, J. H., Jung, Y. K., Joo, M. K., Kim, J. H., Yim, H. J., Park, J. J., et al. (2010). Hepatitis B viral surface mutations in patients with adefovir resistant chronic hepatitis B with A181T/V polymerase mutations. *J. Korean Med. Sci.* 25, 257–264. doi: 10.3346/jkms.2010.25.2.257

Kim, H., Lee, S. A., and Kim, B. J. (2016). X region mutations of hepatitis B virus related to clinical severity. *World J. Gastroenterol.* 22, 5467–5478. doi: 10.3748/wjg.v22.i24.5467

Kim, J. E., Lee, S. Y., Kim, H., Kim, K. J., Choe, W. H., and Kim, B. J. (2017). Naturally occurring mutations in the reverse transcriptase region of hepatitis B virus polymerase from treatment-naïve Korean patients infected with genotype C2. *World J. Gastroenterol.* 23, 4222–4232. doi: 10.3748/wjg.v23.i23.4222

Kim, H., Lee, S. A., Kim, D. W., Lee, S. H., and Kim, B. J. (2013). Naturally occurring mutations in large surface genes related to occult Infection of Hepatitis B virus Genotype C. Blackard J, editor. *PLoS One* 8:e54486. doi: 10.1371/journal.pone.0084194

Kitab, B., Essaid El Feydi, A., Afifi, R., Trepo, C., Benazzouz, M., Essamri, W., et al. Variability in the Precore and Core Promoter Regions of HBV Strains in Morocco: Characterization and Impact on Liver Disease Progression. R. Ray editor. *PLoS One.* (2012). 7:e42891, doi: 10.1371/journal.pone.0042891.

Kobayashi, M., Arase, Y., Ikeda, K., Tsubota, A., Suzuki, Y., Saitoh, S., et al. (2003). Precore wild-type hepatitis B virus with G1896 in the resolution of persistent hepatitis B virus infection. *Intervirology* 46, 157–163. doi: 10.1159/000071456

Kumar, S., Stecher, G., Tamura, K., and Dudley, J. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets Downloaded from. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kwei, K., Tang, X., Lok, A. S., Sureau, C., Garcia, T., Li, J., et al. (2013). Impaired Virion secretion by Hepatitis B virus immune escape mutants and its rescue by wild-type envelope proteins or a second-site mutation. *J. Virol.* 87, 2352–2357. doi: 10.1128/JVI.02701-12

Lago, B. V., Mello, F. C., Ribas, F. S., Valente, F., Soares, C. C., Niel, C., et al. Analysis of Complete Nucleotide Sequences of Angolan Hepatitis B Virus Isolates Reveals the Existence of a Separate Lineage within Genotype E. I. A. Chemin, editor. *PLoS One.* (2014). 9:e92223, doi: 10.1371/journal.pone.0092223.

Lazarevic, I., Banko, A., Miljanovic, D., and Cupic, M. (2019). Immune-escape hepatitis B virus mutations associated with viral reactivation upon immunosuppression. *Viruses* 11:778. doi: 10.3390/v11090778

Leng, X. H., Chen, E. Q., Du, L. Y., Bai, L., Gong, D. Y., Cheng, X., et al. (2015). Biological characteristics of the A1762T/G1764A mutant strain of hepatitis B virus in vivo. *Mol. Med. Rep.* 12, 5141–5148. doi: 10.3892/mmr.2015.4072

León, B., Taylor, L., Vargas, M., Luftig, R. B., Albertazzi, F., Herrero, L., et al. (2005). HBx M130K and V131I (T-A) mutations in HBV genotype F during a follow-up study in chronic carriers. *Virol. J.* 4:60

Li, M. W., Hou, W., Wo, J. E., and Liu, K. Z. (2005). Character of HBV (hepatitis B virus) polymerase gene rtM204V/I and rtL180M mutation in patients with lamivudine resistance. *J. Zhejiang Univ. Sci.* 6, 664–667. doi: 10.1631/jzus.2005.B0664

Li, H., Jia, J., Wang, M., Wang, H., Gu, X., Fang, M., et al. (2017). F221Y mutation in hepatitis B virus reverse transcriptase is associated with hepatocellular carcinoma prognosis following liver resection. *Mol. Med. Rep.* 15, 3292–3300. doi: 10.3892/mmr.2017.6362

Li, X. Y., Liang, C. H., Parkman, V., and Lv, Z. T. (2018). The association between TNF-a 238A/G and 308A/G polymorphisms and juvenile idiopathic arthritis: An updated PRISMA-compliant meta-analysis. *Medicine* 97. doi: 10.1097/MD.0000000000013964

Li, X. G., Liu, B. M., Xu, J., Liu, X. E., Ding, H., and Li, T. (2012). Discrepancy of potential antiviral resistance mutation profiles within the HBV reverse transcriptase between nucleos(t)ide analogue-untreated and -treated patients with chronic hepatitis B in a hospital in China. *J. Med. Virol.* 84, 207–216. doi: 10.1002/jmv.23182

Liao, Y., Hu, X., Chen, J., Cai, B., Tang, J., Ying, B., et al. (2012). Precore mutation of hepatitis B virus may contribute to hepatocellular carcinoma risk: evidence from an updated meta-analysis. *PLoS One* 7. doi: 10.1371/journal.pone.0038394

Lin, X., Xu, X., Huang, Q. L., Liu, Y. Q., Zheng, D. L., Chen, W. N., et al. (2005). Biological impacts of "hot-spot" mutations of hepatitis B virus X proteins are genotype B and C differentiated. *World J. Gastroenterol.* 11, 4703–4708. doi: 10.3748/wjg.v11.i30.4703

Liu, B. M., Li, T., Xu, J., Li, X. G., Dong, J. P., Yan, P., et al. (2010). Characterization of potential antiviral resistance mutations in hepatitis B virus reverse transcriptase sequences in treatment-naïve Chinese patients. *Antiviral Res.* 85, 512–519. doi: 10.1016/j.antiviral.2009.12.006

Liu, S., Zhang, H., Gu, C., Yin, J., He, Y., Xie, J., et al. (2009). Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. *J. Natl. Cancer Inst.* 101, 1066–1082.

Locarnini, S., and Mason, W. S. (2006). Cellular and virological mechanisms of HBV drug resistance. *J. Hepatol.* 44, 422–431. doi: 10.1016/j.jhep.2005.11.036

Locarnini, S., McMillan, J., and Bartholomeusz, A. (2003). The hepatitis B virus and common mutants. *Semin. Liver Dis.* 23, 5–20.

Mahabadi, M., Norouzi, M., Alavian, S. M., Samimirad, K., Azad, T. M., Saberfar, E., et al. (2013). Drug-related mutational patterns in hepatitis B virus (HBV) reverse transcriptase proteins from Iranian treatment-naïve chronic HBV patients. *Hepat. Mon.* 13. doi: 10.5812/hepatmon.6712

Mani, M., Vijayaraghavan, S., Sarangan, G., Barani, R., Abraham, P., and Srikanth, P. (2019). Hepatitis B virus X protein: the X factor in chronic hepatitis B virus disease progression. *Indian J. Med. Microbiol.* doi: 10.4103/ijmm.IJMM_19_421

Meldal, B. H. M., Moula, N. M., Barnes, I. H. A., Boukef, K., and Allain, J. P. (2009). A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors. *J. Gen. Virol.* 90, 1622–1628. doi: 10.1099/vir.0.009738-0

Moerman, B., Moons, V., Sommer, H., Schmitt, Y., and Stetter, M. (2004). Evaluation of sensitivity for wild type and mutant forms of hepatitis B surface antigen by four commercial HBsAg assays. *Clin. Lab.* 50, 159–162. PMID: 15074469

Mohamadkhani, A., Montazeri, G., and Poustchi, H. (2011). The importance of hepatitis B virus genome diversity in basal core promoter region. *Middle East J. Dig. Dis.* 3, 13–19. PMID: 25197527

Mokaya, J., Maponga, T. G., McNaughton, A. L., van Schalkwyk, M., Hugo, S., Singer, J. B., et al. (2020). Evidence of tenofovir resistance in chronic hepatitis B virus

(HBV) infection: An observational case series of South African adults. *J. Clin. Virol.* 129. doi: 10.1016/j.jcv.2020.104548

Mokaya, J., McNaughton, A., Bester, P., Goedhals, D., Barnes, E., Marsden, B., et al. (2019). Hepatitis B virus resistance to tenofovir: fact or fiction? A synthesis of the evidence to date. medRxiv [Preprint].

Mokaya, J., Mcnaughton, A. L., Bester, P. A., Goedhals, D., Barnes, E., Marsden, B. D., et al. (2020). *Hepatitis B virus resistance to tenofovir: fact or fiction? A systematic literature review and structural analysis of drug resistance mechanisms [version 1; peer review: awaiting peer review]*.

Mondal, R. K., Khatun, M., Ghosh, S., Banerjee, P., Datta, S., Sarkar, S., et al. (2015). Immune-driven adaptation of hepatitis B virus genotype D involves preferential alteration in B-cellepitopes and replicative attenuation-an insight from human immunodeficiency virus/hepatitis B virus coinfection. *Clin. Microbiol. Infect.* 21, 710.e11–710.e20. doi: 10.1016/j.cmi.2015.03.004

Moriconi, F., Colombatto, P., Coco, B., Ciccorossi, P., Oliveri, F., Flichman, D., et al. (2007). Emergence of hepatitis B virus quasispecies with lower susceptibility to nucleos(t)ide analogues during lamivudine treatment. *J. Antimicrob. Chemother.* 60, 341–349. doi: 10.1093/jac/dkm187

Mukaide, M., Tanaka, Y., Shin-I, T., Yuen, M. F., Kurbanov, F., Yokosuka, O., et al. (2010). Mechanism of entecavir resistance of hepatitis B virus with viral breakthrough as determined by long-term clinical assessment and molecular docking simulation. *Antimicrob. Agents Chemother.* 54, 882–889. doi: 10.1128/AAC.01061-09

Ng, L. F. P., Chan, M., Chan, S. H., Cheng, P. C. P., Leung, E. H. C., Chen, W. N., et al. (2005). Host heterogeneous ribonucleoprotein K (hnRNP K) as a potential target to suppress hepatitis B virus replication. *PLoS Med.* 2, 0673–0683.

Nishizawa, T., Hoshino, T., Naganuma, A., Kobayashi, T., Nagashima, S., Takahashi, M., et al. (2016). Enhanced pregenomic RNA levels and lowered precore mRNA transcription efficiency in a genotype A hepatitis B virus genome with C1766T and T1768A mutations obtained from a fulminant hepatitis patient. *J. Gen. Virol.* 97, 2643–2656. doi: 10.1099/jgv.0.000566

Olinger, C. M., Weber, B., Otegbayo, J. A., Ammerlaan, W., Van Der Taelem-Brulé, N., and Muller, C. P. (2007). Hepatitis B virus genotype E surface antigen detection with different immunoassays and diagnostic impact of mutations in the preS/S gene. *Med. Microbiol. Immunol.* 196, 247–252. doi: 10.1007/s00430-007-0050-5

Olusola, B. A., Faneye, A. O., Oluwasemowo, O. O., Motayo, B. O., Adebayo, S., Oludiran-Ayoade, A. E., et al. (2021). Profiles of mutations in hepatitis B virus surface and polymerase genes isolated from treatment-naïve Nigerians infected with genotype E. *J. Med. Microbiol.* 70. doi: 10.1099/jmm.0.001338

Ouneissa, R., Bahri, O., Alaya-Bouafif, N., Ben Chouaieb, S., Yahia, A., Ben Sadraoui, A., et al. (2012). Frequency and clinical significance of core promoter and precore region mutations in Tunisian patients infected chronically with hepatitis B. *J. Med. Virol.* 84, 1719–1726. doi: 10.1002/jmv.23394

Ouneissa, R., Bahri, O., Ben Yahia, A., Touzi, H., Msaddak Azouz, M., Ben Mami, N., et al. (2013). Evaluation of PCR-RFLP in the Pre-S Region as Molecular Method for Hepatitis B Virus Genotyping. *Hepat. Mon.* 13. doi: 10.5812/hepatmon.11781

Pahal, V., Singh, J., and Dadhich, K. S. (2016). Hepatitis B virus core promoter and precore mutations and their relatedness to genotypes and disease pathogenesis. *Pelagia Res. Lib. Adv. Appl. Sci. Res.* 7, 70–80.

Parekh, S., Zoulim, F., Ahn, S. H., Tsai, A., Li, J., Kawai, S., et al. (2003). Genome replication, Virion secretion, and e antigen expression of naturally occurring Hepatitis B virus Core promoter mutants. *J. Virol.* 77, 6601–6612. doi: 10.1128/JVI.77.12.6601-6612.2003

Park, E. S., Lee, A. R., Kim, D. H., Lee, J. H., Yoo, J. J., Ahn, S. H., et al. (2019). Identification of a quadruple mutation that confers tenofovir resistance in chronic hepatitis B patients. *J. Hepatol.* 70, 1093–1102. doi: 10.1016/j.jhep.2019.02.006

Pollicino, T., Isgrò, G., di Stefano, R., Ferraro, D., Maimone, S., Brancatelli, S., et al. (2009). Variability of reverse transcriptase and overlapping S gene in hepatitis B virus isolates from untreated and lamivudine-resistant chronic hepatitis B patients. *Antivir. Ther.* 14, 649–654. doi: 10.1177/135965350901400504

Pollicino, T., Raffa, G., Costantino, L., Lisa, A., Campello, C., Squadrito, G., et al. (2007). Molecular and functional analysis of occult hepatitis B virus isolates from patients with hepatocellular carcinoma. *Hepatology* 45, 277–285. doi: 10.1002/hep.21529

Pourkarim, M. R., Vergote, V., Amini-Bavil-Olyaee, S., Sharifi, Z., Sijmons, S., Lemey, P., et al. (2014). Molecular characterization of hepatitis B virus (HBV) strains circulating in the northern coast of the Persian Gulf and its comparison with worldwide distribution of HBV subgenotype D1. *J. Med. Virol.* 86, 745–757. doi: 10.1002/jmv.23864

Poustchi, H., Mohamadkhani, A., Bowden, S., Montazeri, G., Ayres, A., Revill, P., et al. (2008). Clinical significance of precore and core promoter mutations in genotype D hepatitis B-related chronic liver disease. *J. Viral Hepat.* 15, 753–760. doi: 10.1111/j.1365-2893.2008.00998.x

Purdy, M. A., Talekar, G., Swenson, P., Araujo, A., and Fields, H. (2006). A new algorithm for deduction of hepatitis B surface antigen subtype determinants from the amino acid sequence. *Intervirology* 50, 45–51.

Putri, W. A., Yano, Y., Yamani, L. N., Lusida, M. I., Soetjipto, L. Y., et al. (2019). Association between HBx variations and development of severe liver disease among indonesian patients. *Kobe J. Med. Sci.* 65, 28–35.

Qin, B., Pei, R. J., He, T. T., Huang, Z. H., Pan, G. S., Tu, C. Y., et al. (2013a). Polymerase mutations rtN238R, rtT240Y and rtN248H of hepatitis B virus decrease susceptibility to adefovir. *Chin. Sci. Bull.* 58, 1760–1766. doi: 10.1007/s11434-013-5770-x

Qin, B., Zhang, B., Zhang, X., He, T., Xu, W., Fu, L., et al. (2013b). Substitution Rtq267h of hepatitis B virus increases the weight of replication and lamivudine resistance. *Hepat. Mon.* 13:12160. doi: 10.5812/hepatmon.12160

Quarleri, J. (2014). Core promoter: a critical region where the hepatitis B virus makes decisions. *World J. Gastroenterol.* 20, 425–435. doi: 10.3748/wjg.v20.i2.425

Quiros-Roldan, E., Calabresi, A., Lapadula, G., Tirelli, V., Costarelli, S., Cologni, G., et al. (2008). Evidence of long-term suppression of hepatitis B virus DNA by tenofovir as rescue treatment in patients coinfected by HIV. *Antivir. Ther.* 13, 341–348. doi: 10.1177/135965350801300315

Rajoriya, N., Combet, C., Zoulim, F., and Janssen, H. L. A. (2017). How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *J. Hepatol.* 67, 1281–1297. doi: 10.1016/j.jhep.2017.07.011

Rendon, J. C., Cortes-Mancera, F., Restrepo-Gutierrez, J. C., and Hoyos, S. (2017). Navas MC. Molecular characterization of occult hepatitis B virus infection in patients with end-stage liver disease in Colombia. *PLoS One* 12:e0180447. doi: 10.1371/journal.pone.0180447

Saha, D., Pal, A., Biswas, A., Panigrahi, R., Sarkar, N., Das, D., et al. (2014). Molecular characterization of HBV strains circulating among the treatment-naive HIV/HBV co-infected patients of eastern India. *PLoS One* 9. doi: 10.1371/journal.pone.0090432

Salarneia, F., Zhand, S., Khodabakhshi, B., Tabarraei, A., Vakili, M. A., Javid, N., et al. (2016). Mutations at nucleotide 1762, 1764 and 1766 of Hepatitis B virus X gene in patients with chronic Hepatitis B and Hepatitis B-related cirrhosis. *Med. Lab. J* 10, 31–35.

Schildgen, O. (2007). Novel lamivudine resistance [3]. *Antimicrob. Agents Chemother.* 51:4533. doi: 10.1128/AAC.00840-07

Sendi, H., Mehrab-Mohseni, M., Zali, M. R., Norder, H., and Magnius, L. O. (2005). T1764G1766 core promoter double mutants are restricted to Hepatitis B virus strains with an A1757 and are common in genotype D. *J. Gen. Virol.* 86, 2451–2458. doi: 10.1099/vir.0.81023-0

Shaw, T., Bartholomeusz, A., and Locarnini, S. (2006). HBV drug resistance: mechanisms, detection and interpretation. *J. Hepatol.* 44, 593–606. doi: 10.1016/j.jhep.2006.01.001

Sheldon, J. (2008). Soriano V. Hepatitis B virus escape mutants induced by antiviral therapy. *J. Antimicrob. Chemother.* 61, 766–768. doi: 10.1093/jac/dkn014

Sheldon, J., Rodès, B., Zoulim, F., Bartholomeusz, A., and Soriano, V. (2006). Mutations affecting the replication capacity of the hepatitis B virus [Internet]. *J. Viral Hepat.* 13, 427–434. doi: 10.1111/j.1365-2893.2005.00713.x

Sitnik, R., Pinho, J. R. R., Bertolini, D. A., Bernardini, A. P., da Silva, L. C., and Carrilho, F. J. (2004). Hepatitis B virus genotypes and precore and core mutants in Brazilian patients. *J. Clin. Microbiol.* 42, 2455–2460. doi: 10.1128/JCM.42.6.2455-2460.2004

Sominskaya, I., Mihailova, M., Jansons, J., Legzdina, D., Sudmale, G., Pumpens, P., et al. (2011). Hepatitis B virus genotypes in Latvia. *Open Hepatol. J.* 3. doi: 10.2174/1876517301103010007

Sunbul, M. (2014). Hepatitis B virus genotypes: global distribution and clinical importance. *World J. Gastroenterol.* 20:5427. doi: 10.3748/wjg.v20.i18.5427

Suntur, B. M., Sayan, M., Kaya, H., and Ünal, N. (2019). Overlapping Pol/S gene analysis in chronic hepatitis B patients with coexisting Hbsag and Anti-HBs. *Acta Medica Mediterranea.* 35:3113

Suppiah, J., Zain, R. M., Bahari, N., Nawi, S. H., and Saat, Z. (2015). G1896A precore mutation and association with HBeAG status, genotype and clinical status in patients with chronic hepatitis B. *Hepat. Mon.* 15. doi: 10.5812/hepatmon.31490

Taghiabadi, M., Hosseini, S. Y., Gorzin, A. A., Taghavi, S. A., Monavari, S. H. R., and Sarvari, J. (2019). Comparison of pre-S1/S2 variations of hepatitis B virus between asymptomatic carriers and cirrhotic/hepatocellular carcinoma-affected individuals. *Clin Exp Hepatol.* 5, 161–168. doi: 10.5114/ceh.2019.84781

Tenney, D. J., Levine, S. M., Rose, R. E., Walsh, A. W., Weinheimer, S. P., Discotto, L., et al. (2004). Clinical emergence of entecavir-resistant hepatitis B virus requires additional substitutions in virus already resistant to lamivudine. *Antimicrob. Agents Chemother.* 48, 3498–3507. doi: 10.1128/AAC.48.9.3498-3507.2004

Thompson, A. J. V., Nguyen, T., Iser, D., Ayres, A., Jackson, K., Littlejohn, M., et al. (2010). Serum hepatitis B surface antigen and hepatitis B e antigen titers: Disease phase influences correlation with viral load and intrahepatic hepatitis B virus markers. *Hepatology* 51, 1933–1944. doi: 10.1002/hep.23571

Tokgöz, Y., Terlemez, S., Sayan, M., and Kırdar, S. (2018). Investigation of antiviral resistance and escape mutations in children with naive chronic hepatitis B patients and their parents. *Turk. J. Pediatr.* 60, 514–519. doi: 10.24953/turkjped.2018.05.007

Tong, M. J., Blatt, L. M., Kao, J. H., Cheng, J. T., and Corey, W. G. (2007). Basal core promoter T1762/A1764 and precore A1896 gene mutations in hepatitis B surface antigen-positive hepatocellular carcinoma: a comparison with chronic carriers. *Liver Int.* 27, 1356–1363. doi: 10.1111/j.1478-3231.2007.01585.x

Tong, S., Kim, K. H., Chante, C., Wands, J., and Li, J. (2005). Hepatitis B virus e antigen variants. *Int. J. Med. Sci.* 2, 2–7. doi: 10.7150/ijms.2.2

Tong, S., Li, J., Wands, J. R., and Wen, Y. (2013). Hepatitis B virus genetic variants: biological properties and clinical implications. *Emerg. Microbes Infect.* 2, 1–11.

Torresi, J., Earnest-Silveira, L., Civitico, G., Walters, T. E., Lewin, S. R., Fyfe, J., et al. (2002a). Restoration of replication phenotype of lamivudine-resistant hepatitis B virus mutants by compensatory changes in the "fingers" subdomain of the viral polymerase selected as a consequence of mutations in the overlapping S gene. *Virology* 299, 88–99. doi: 10.1006/viro.2002.1448

Torresi, J., Earnest-Silveira, L., Deliyannis, G., Edgtton, K., Zhuang, H., Locarnini, S. A., et al. (2002b). Reduced antigenicity of the Hepatitis B virus HBsAg protein arising as a consequence of sequence changes in the overlapping polymerase gene that are selected by lamivudine therapy. *Virology* 293, 305–313. doi: 10.1006/viro.2001.1246

Triki, H., Ben Slimane, S., Ben Mami, N., Sakka, T., Ben Ammar, A., and Dellagi, K. (2000). High circulation of hepatitis B virus (HBV) precore mutants in Tunisia, North Africa. *Epidemiol. Infect.* 125, 169–174. doi: 10.1017/S0950268899003921

Villet, S., Ollivet, A., Pichoud, C., Barraud, L., Villeneuve, J. P., Trépo, C., et al. (2007). Stepwise process for the development of entecavir resistance in a chronic hepatitis B virus infected patient. *J. Hepatol.* 46, 531–538. doi: 10.1016/j.jhep.2006.11.016

Wang, L., Han, F., Duan, H., Ji, F., Yan, X., Fan, Y., et al. (2017). Hepatitis B virus pre-existing drug resistant mutation is related to the genotype and disease progression. *J. Infect. Dev. Ctries.* 11, 727–732. doi: 10.3855/jidc.9021

Wang, Y., Zeng, L., and Chen, W. (2016). HBV X gene point mutations are associated with the risk of hepatocellular carcinoma: A systematic review and meta-analysis. *Mol. Clin. Oncol.* 4, 1045–1051. doi: 10.3892/mco.2016.847

Wang, Q., Zhang, T., Ye, L., Wang, W., and Zhang, X. (2012). Analysis of hepatitis B virus X gene (HBx) mutants in tissues of patients suffered from hepatocellular carcinoma in China. *Cancer Epidemiol.* 36, 369–374. doi: 10.1016/j.canep.2011.11.006

Wei, C., Yu-tian, C., Ji-zhi, W., Yong-wei, L., and Gang, L. (2011). Characterization of hepatitis virus B isolated from a multi-drug refractory patient. *Virus Res.* 155, 254–258. doi: 10.1016/j.virusres.2010.10.018

Westland, C. (2003). Week 48 resistance surveillance in two phase 3 clinical studies of adefovir dipivoxil for chronic hepatitis B. *Hepatology* 38, 96–103. doi: 10.1053/jhep.2003.50288

WHO (2021). Hepatitis B WHO guidlines. Available at: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b (Accessed April 25, 2022).

Wu, Y., Gan, Y., Gao, F., Zhao, Z., Jin, Y., Zhu, Y., et al. (2014). Novel natural mutations in the Hepatitis B virus reverse transcriptase domain associated with hepatocellular carcinoma. *PLoS One* 9. doi: 10.1371/journal.pone.0115141

Wu, J. F., Ni, Y. H., Chen, H. L., Hsu, H. Y., and Chang, M. H. (2014). The impact of hepatitis B virus precore/core gene carboxyl terminal mutations on viral biosynthesis and the host immune response. *J Infect Dis* 209, 1374–1381. doi: 10.1093/infdis/jit638

Wungu, C. D. K., Amin, M., Ruslan, S. E. N., Purwono, P. B., Kholili, U., Maimunah, U., et al. (2019). Association between host TNF-α, TGF-β1, p53 polymorphisms, hbv x gene mutation, hbv viral load and the progression of hbv-associated chronic liver disease in indonesian patients. *Biomed. Rep.* 11, 145–153. doi: 10.3892/br.2019.1239

Xu, R., Zhang, X., Zhang, W., Fang, Y., Zheng, S., and Yu, X. F. (2007). Association of human APOBEC3 cytidine deaminases with the generation of hepatitis virus B x antigen mutants and hepatocellular carcinoma. *Hepatology* 46, 1810–1820. doi: 10.1002/hep.21893

Yamani, L. N., Yano, Y., Utsumi, T., Wasityastuti, W., Rinonce, H. T., Widasari, D. I., et al. (2017). Profile of mutations in the reverse transcriptase and overlapping surface genes of hepatitis B virus (HBV) in treatment-naïve indonesian HBV carriers. *Jpn. J. Infect. Dis.* 70, 647–655. doi: 10.7883/yoken.JJID.2017.078

Yan, L., Zhang, H., Ma, H., Liu, D., Li, W., Kang, Y., et al. (2015). Deep sequencing of hepatitis B virus basal core promoter and precore mutants in HBeAg-positive chronic hepatitis B patients. *Sci. Rep.* 9:5

Yang, H., Qi, X., Sabogal, A., Miller, M., Xiong, S., and Delaney, W. E. (2005). Cross-resistance testing of next-generation nucleoside and nucleotide analogues against lamivudine-resistant HBV. *Antivir. Ther.* 1359

Yang, H., Westland, C. E., Delaney, W. E., Heathcote, E. J., Ho, V., Fry, J., et al. (2002). Resistance surveillance in chronic hepatitis B patients treated with adefovir dipivoxil for up to 60 weeks. *Hepatology* 36, 464–473. doi: 10.1053/jhep.2002.34740

Yin, J., Xie, J., Liu, S., Zhang, H., Han, L., Lu, W., et al. (2011). Association between the various mutations in viral core promoter region to different stages of hepatitis B, ranging of asymptomatic carrier state to hepatocellular carcinoma. *Am. J. Gastroenterol.* 106, 81–92. doi: 10.1038/ajg.2010.399

Yuan, J. M., Ambinder, A., Fan, Y., Gao, Y. T., Yu, M. C., and Groopman, J. D. (2009). Prospective evaluation of hepatitis B 1762 T/1764 a mutations on hepatocellular carcinoma development in Shanghai, China. *Cancer Epidemiol. Biomarkers Prev.* 18, 590–594. doi: 10.1158/1055-9965.EPI-08-0966

Zheng, J., Zeng, Z., Zhang, D., Yu, Y., Wang, F., and Pan, C. Q. (2012). Prevalence and significance of Hepatitis B reverse transcriptase mutants in different disease stages of untreated patients. *Liver Int.* 32, 1535–1542. doi: 10.1111/j.1478-3231.2012.02859.x

Zhu, R., Zhang, H. P., Yu, H., Li, H., Ling, Y. Q., Hu, X. Q., et al. (2008). Hepatitis B virus mutations associated with in situ expression of hepatitis B core antigen, viral load and prognosis in chronic hepatitis B patients. *Pathol. Res. Pract.* 204, 731–742. doi: 10.1016/j.prp.2008.05.001

Ziaee, M., Javanmard, D., Sharifzadeh, G., Namaei, M. H., and Azarkar, G. (2016). Genotyping and mutation pattern in the overlapping MHR region of HBV isolates in southern khorasan, eastern Iran. *Hepat. Mon.* 16:37806. doi: 10.5812/hepatmon.37806

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 26 October 2022
DOI 10.3389/fmicb.2022.1032918

OPEN ACCESS

EDITED BY
Ana Grande-Pérez,
Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Spain

REVIEWED BY
Manasvi Lingam,
Florida Institute of Technology, United States
Richard Allen White,
University of North Carolina at Charlotte, United States

*CORRESPONDENCE
Ester Lázaro
lazarole@cab.inta-csic.es

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 31 August 2022
ACCEPTED 10 October 2022
PUBLISHED 26 October 2022

CITATION
de la Higuera I and Lázaro E (2022) Viruses in astrobiology. *Front. Microbiol.* 13:1032918. doi: 10.3389/fmicb.2022.1032918

Viruses in astrobiology

Ignacio de la Higuera[1] and Ester Lázaro[2]*

[1]Department of Biology, Center for Life in Extreme Environments, Portland State University, Portland, OR, United States, [2]Centro de Astrobiología (CAB), CSIC-INTA, Torrejón de Ardoz, Spain

Viruses are the most abundant biological entities on Earth, and yet, they have not received enough consideration in astrobiology. Viruses are also extraordinarily diverse, which is evident in the types of relationships they establish with their host, their strategies to store and replicate their genetic information and the enormous diversity of genes they contain. A viral population, especially if it corresponds to a virus with an RNA genome, can contain an array of sequence variants that greatly exceeds what is present in most cell populations. The fact that viruses always need cellular resources to multiply means that they establish very close interactions with cells. Although in the short term these relationships may appear to be negative for life, it is evident that they can be beneficial in the long term. Viruses are one of the most powerful selective pressures that exist, accelerating the evolution of defense mechanisms in the cellular world. They can also exchange genetic material with the host during the infection process, providing organisms with capacities that favor the colonization of new ecological niches or confer an advantage over competitors, just to cite a few examples. In addition, viruses have a relevant participation in the biogeochemical cycles of our planet, contributing to the recycling of the matter necessary for the maintenance of life. Therefore, although viruses have traditionally been excluded from the tree of life, the structure of this tree is largely the result of the interactions that have been established throughout the intertwined history of the cellular and the viral worlds. We do not know how other possible biospheres outside our planet could be, but it is clear that viruses play an essential role in the terrestrial one. Therefore, they must be taken into account both to improve our understanding of life that we know, and to understand other possible lives that might exist in the cosmos.

KEYWORDS

astrobiology, virosphere, origin of life, quasispecies, horizontal gene transfer, virus biosignatures, experimental evolution

Introduction

Astrobiology is a scientific discipline concerned with the origin, evolution, distribution and future of life in the universe (Des Marais et al., 2008; Hays et al., 2015; Cockell, 2020; O'rourke et al., 2020; Lingam and Loeb, 2021; Bennett et al., 2022). The use of the term life in reference to the object of study of astrobiology may raise doubts

about whether viruses should be included in its field of research. Some of the difficulties in deciding this question derive from the absence of a universal definition of life (Dix, 2003; Benner, 2010; Cleland, 2011, 2019), which is partly due to the fact that the only example of life we know is that on Earth. That all living beings on this planet share the same ancestor (Woese et al., 1990) makes it difficult to identify the fundamental properties of life. What we think is essential because of its presence in all forms of known life, may simply represent a characteristic that has been inherited from the common progenitor.

NASA defines life as a self-sustaining chemical system capable of darwinian evolution. Although viruses are chemical systems composed of the same molecules as life, they cannot use matter and energy from the external environment to build internal order in the same way that life does. Thus, if we regard the concept of self-sustaining system, viruses should not be considered living entities (Moreira and López-García, 2009). Nevertheless, the fact that viruses contain a genome that encodes proteins following the same rules as cells –together with the reality that they can reproduce and evolve– are sufficient arguments for other scientists to consider them as life (Hegde et al., 2009; Navas-Castillo, 2009; Forterre, 2016; Harris and Hill, 2020). To further complicate the question, some recently discovered viruses have genomes and dimensions comparable to some cellular microorganisms (Van Etten and Meints, 1999; La Scola et al., 2003; Abergel and Claverie, 2020) and can even be infected by other viral entities called virophages (Fischer, 2011; Mougari et al., 2019).

If we go back to the times when life was taking its first steps, the boundaries between life and non-life were more blurred than they are today (Szostak, 2012). Although most life definitions contain both a thermodynamic and an inheritance aspect, it is easy to imagine that at primitive times there were entities in which these two properties were not present in the same way they are in modern life. An example of these gray zones would be represented by autocatalytic networks, in which all molecules can be synthesized through the reactions catalyzed within the set (Vitas and Dobovišek, 2019), giving rise to a self-maintained system. In these networks, information is not stored in homopolymers such as DNA or RNA and, therefore, evolution would reside in variations in inter-molecular interactions or in the mechanisms of obtaining energy from the environment. Viruses would be another example of gray zone that can exist as an inert form, the virion, and an active form, the virus multiplying inside the cell. In their active form, viruses are self-organized systems that store and transmit information, and are able to maintain their organization despite changes in the environment. Although viruses always need to infect a cell to reproduce, all the information necessary to manipulate the cell metabolism and produce a viral progeny is contained in the virus genome itself, which brings them closer to the definition of life. Forterre introduced the concept of virocell (Forterre, 2011) as a living system corresponding to any cell infected by a virus, whose main function is to produce new virions that would act as seeds or spores.

Whether viruses are considered living things or not, there is no doubt that finding viruses on a planet other than Earth would immediately make us think about the possibility of the existence of life on it. That life could be well established, taking its first steps, or even be extinct, having left viruses as the last vestiges of its existence. Viruses are inseparable companions of life as we know it. As we will discuss extensively in this review, viruses are probably very ancient and were present in the pool of genetic elements that gave rise to life (Koonin et al., 2006; Forterre and Prangishvili, 2009; Villarreal and Witzany, 2010; Durzyńska and Goździcka-Józefiak, 2015; Krupovic et al., 2019). Since that time, they have influenced, and continue influencing, evolution and distribution of life (Filée et al., 2003; Forterre, 2006; Koonin et al., 2006; Koonin, 2016). And there is no doubt of their relevance to conform the properties of the life that will populate our planet in the future.

Traditionally, viruses have been seen as disease-causing agents, a view that has changed radically in recent years. Pathogenic viruses are only a small fraction of the wide variety of viruses that exist. Many of them coexist peacefully with their hosts, sometimes even causing benefits. Often, advantages provided by viruses derive from their high capacity to modify and exchange pieces of their genomes, which convert them in great inventors of genes that, thanks to horizontal gene transfer, can be subsequently transferred to the cellular world (Muniesa et al., 2011; Penadés et al., 2015; Gilbert and Cordaux, 2017; Weiss, 2017; Irwin et al., 2022), giving rise to many of the intricate connections among the different branches of the phylogenetic tree of life.

Viruses exist practically everywhere in our planet, including extreme environments (Säwström et al., 2008; Yoshida-Takashima et al., 2012; Munson-McGee et al., 2020), which in some cases have physicochemical conditions that resemble some of those present in extraterrestrial environments. They are the most abundant biological entities on Earth and it is believed that all cellular organisms can be infected by some kind of virus (Suttle, 2005; Mushegian, 2020), which raises the question of whether they are an inevitable consequence of the emergence of living systems (Iranzo et al., 2016; Koonin et al., 2017) and must be present in any potential biosphere. Viruses, in the virion form, can withstand more extreme conditions than most cellular life. For this reason, sometimes they have even been considered as possible containers for the transport of genetic material between planets (Griffin, 2013; Berliner et al., 2018).

Finally, the rapidity of viral evolution is being exploited in multiple laboratories to carry out studies devoted to the research of biological evolution in real time (Elena et al., 2008; Kawecki et al., 2012; McDonald, 2019), which allows to establish some of the general principles governing this process. The importance

in evolution of the error rate value, the intensity of selective pressures, the population size or the standing genetic diversity are just some examples of the topics that can be studied using viruses as experimental system.

The conclusion that arises from all the considerations described above is that, although viruses are not considered *sensu stricto* life, both current and past life in our planet has been shaped thanks to their action. As knowledge has advanced, viruses have gone from being considered enemies of life to being recognized highly relevant in fields as diverse as molecular ecology, evolutionary biology, structural biology or geomicrobiology. The same is happening with astrobiology, which is increasingly realizing that to understand life in a wide sense it is absolutely necessary to include viruses in its scenario (Griffin, 2013; Berliner et al., 2018).

What can viruses teach us about the origin of life?

How old are viruses? Does their simplicity mean that they are intermediate forms between inert and living matter? Or do they represent an alternative path followed by the same precursors that gave rise to life? Some proposals also state that viruses are modern entities that have derived from cells. So far, we do not have a definite answer for these questions, although most of the evidence indicates that viruses, particularly those with an RNA genome, are very old. Therefore, their study can provide relevant information about the first steps toward life.

Origin of viruses

There are three main hypotheses to explain the origin of viruses (Forterre, 2006; Koonin et al., 2006; Forterre and Prangishvili, 2009; Krupovic et al., 2019) that are briefly described below. Each of them has its implications in the relationships that the viral world establishes with the cellular world and in their mutual influence throughout life history. The three hypotheses are not exclusive in the sense that different groups of viruses could have been originated through different routes and at different times. For example, viruses with a DNA genome could not have existed until the emergence of this molecule, which places their origin at a later time than that of RNA viruses.

Escape hypothesis

Viruses are genetic cell elements that escaped from cells, becoming infectious entities unable to replicate on their own. According to this idea, prior to the emergence of DNA cells, there was a stage in the evolution of life that was dominated by RNA cells containing fragmented genomes (Woese, 1987, 2002). In contrast to the current world –in which the processes of genomic replication and cell division are perfectly integrated– in the primitive world, there were probably no mechanisms that regulated this integration in a precise way. Thus, it would be easier at that time that a genomic fragment could be released from a cell and become an infectious unit capable of replicating at the expense of the resources produced by others. For this hypothesis to be true, there should be more similarities between the genes of viruses that infect a particular domain of life and the cellular genes of that same domain. However, when virus protein structural domains have been analyzed, in many cases they are found to be similar in bacteriophages, archaeal viruses and eukaryotic viruses (Bamford, 2003; Benson et al., 2004; Rice et al., 2004; Trus et al., 2004; Krupovic and Bamford, 2008; Abrescia et al., 2012; Brum et al., 2013).

Reductive hypothesis

According to this scenario, viruses are the product of the degeneration of ancestral cells that lost their machinery for protein synthesis and energy production. These deteriorated cells became parasites of others that kept all their capacities. No evolutionary intermediary between viruses and cells has been found to date. Moreover, when viruses are compared with other parasites, the latter always retain some characteristics of their free-living time. The discovery of giant viruses that infect protists, and that sometimes possess genes of the protein translation machinery, triggered a resurgence of the reductive hypothesis (Schulz et al., 2017; Abrahão et al., 2018; Rodrigues et al., 2020). However, it is now widely accepted that these genes have been acquired from the host (Koonin and Yutin, 2018; Brahim Belhaouari et al., 2022).

Virus-first hypothesis

This hypothesis states that viruses, particularly those with RNA genomes, descend from the first molecules with replicative capacity that likely existed on Earth before the appearance of cellular life. For a time, this thinking received little consideration, because actual viruses always need a cell to replicate and, therefore, cells should have preceded viruses. However, primitive viruses were probably very different from today's viruses. They could have been mere genetic parasites that emerged in replicator networks endowed with catalytic capacities. Parasites would multiply at the expense of the products of the network without contributing to their generation, as it is shown in some theoretical studies (Iranzo et al., 2016; Koonin et al., 2017). Additional, indirect support for this hypothesis comes from the fact that viruses use a great diversity of molecules [single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), single-stranded DNA (ssDNA), and double-stranded DNA (dsDNA)] and strategies to store and replicate their genetic information, all of which brings to mind a time when different ways of preserving and processing genetic information were being tested (Koonin et al., 2006; Holmes, 2011b; Krupovic et al., 2019).

The RNA world

How was the pre-cellular state of biological evolution where viruses could first emerge? In modern life, functional proteins can be synthesized because DNA stores information about the order in which amino acids should be linked together. But going from the DNA sequence to the sequence of a protein is an intricate process, which in turn requires the intervention of other proteins and complex structures such as ribosomes. Therefore, separation of information and function into two different molecules poses a paradox that can only be resolved if, in primitive life, information and function resided in the same molecule. Many scientists accept that at some stage of the evolution of life this molecule was RNA, although this does not preclude the existence of pre-RNA worlds in which catalysis and information could reside in other molecules that were more stable or easier to assemble than RNA. In this context, PNAs (peptide nucleic acids), which are synthetic polymers with a simple, achiral chemical structure composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds (Nielsen, 2007) have received much consideration.

That RNA can be used to store hereditary information is demonstrated by the existence of viroids and RNA viruses. Due to internal base pairing, RNA molecules fold into secondary and tertiary structures that maximize their stability. As it happens with proteins in the current world, these structures can lead to the formation of active centers that facilitate the catalysis of certain chemical reactions. RNA molecules with catalytic capacity are called ribozymes. Their existence was first demonstrated in the eighties (Kruger et al., 1982; Bass and Cech, 1984; Altman, 1989) and confirmed in many subsequent studies (Chapman and Szostak, 1994; Hager et al., 1996; Lee and Lee, 2017; Janzen et al., 2020).

Everything described above led to postulate the existence of a hypothetical RNA world, made up of ensembles of molecules, the so-called primitive replicators, capable of storing and transmitting information (Gilbert, 1986; Oro et al., 1990; Sankaran, 2016). The catalytic properties of these molecules could have facilitated the emergence of a simple metabolism, which, once individualized in a compartment, would have been the basis for the appearance of the first cells (Koonin et al., 2006; Koonin, 2014; Szostak, 2017; Joyce and Szostak, 2018) and probably of the first parasites.

The first experiment demonstrating the capacity of RNA molecules to evolve in response to the environment was carried out by Sol Spiegelman in 1967 (Mills et al., 1967), using the bacteriophage Qβ, a virus with an RNA genome (**Figure 1**). The experiment carried out by Spiegelman consisted in mixing in a test tube a small amount of viral RNA with all the components necessary for its replication. This mixture was incubated under optimal conditions for the time necessary for the RNA to be copied. Then, a fraction of the molecules produced were transferred to a new tube containing fresh substrates for replication. This process of serial transfers was repeated over time, with the aim of finding out what happened to the RNA molecules. The result was that the virus genome lost a large part of its sequence, which allowed it to increase the copying rate (**Figure 1**). In new experiments it was found that, if the environment changed, the RNA molecules were capable of specific adaptations, increasing their replication speed under conditions such as imbalanced nucleotide concentrations or presence of ribonucleases (Levisohn and Spiegelman, 1969; Kacian et al., 1972; Biebricher and Gardiner, 1997). These experiments were highly relevant to demonstrate that molecules are also susceptible of optimization through darwinian evolution, a concept of crucial importance for the idea that before cellular life there must have been a phase of molecular evolution.

Theoretical studies inspired in Spiegelman's experiments showed that error-prone replication combined with natural selection, and acting for a sufficiently long time in an infinite population of replicators (RNA-like molecules) differing in their fitness values, would result in the generation of a steady-state distribution of mutants in which each of them has a constant frequency (Eigen, 1971; Eigen and Schuster, 1979). This population structure was referred to as quasispecies, and the framework describing its dynamics quasispecies theory (Biebricher and Eigen, 2006). The major achievement of quasispecies theory was to describe mathematically a system involving multiplication of molecules with a regular production of error copies. The sequence displaying the highest replication rate within the quasispecies was denoted master sequence and corresponded to that most represented in the whole set of mutants, also known as mutant spectrum. The ensemble of all possible variations that can be generated from a genetic sequence is known as the sequence space. For a replicator of length l the sequence space is 4^l, a number that can be enormous, even for short replicators. Quasispecies population structure was an important component of the catalytic hypercycle, a primitive model of organization of self-replicating molecules connected in a cyclic and autocatalytic fashion (Eigen and Schuster, 1979).

In addition to defining in quantitative terms a system of replicators that are copied with high error rate, quasispecies theory established the value of the error rate (the so-called error threshold) that was compatible with the conservation of genetic information (Biebricher and Eigen, 2005). When the error threshold is exceeded, the superiority of the master sequence disappears, all possible sequences become equally probable and the genetic information is lost.

Almost in parallel, experimental studies carried out with bacteriophage Qβ (Batschelet et al., 1976; Domingo et al., 1978) showed that this virus replicated with very high error rate and showed a great heterogeneity in its populations. The typical error rates of RNA viruses are on the order of 10^{-6} to 10^{-4} errors per nucleotide copied (Holland et al., 1982; Duffy et al., 2008; Sanjuán et al., 2010; Domingo et al., 2021).

For comparison, error rate values (expressed with the same units as above) are around 10^{-6} for ssDNA viruses and in the ranges between 10^{-8} and 10^{-7} for dsDNA viruses, 10^{-10} and 10^{-9} for bacteria, and 10^{-11} and 10^{-10} for eukaryotes [(Gago et al., 2009) and references therein]. These high error rate values, together with the usually large sizes of viral populations, lead to the generation of highly diverse populations whose behavior could mimic that of the ensembles of replicators present in the RNA world and described in quasispecies theory. Nevertheless, factors such as the mechanism of genome replication, the variability and intensity of the selective pressures or the frequency of population bottlenecks are also highly relevant to determine the structure of a virus population and its evolutionary dynamics. Consequently, the substitution rate per nucleotide and year in some eukaryotic ssDNA viruses is similar to that of ssRNA viruses with the same genome length (Duffy et al., 2008).

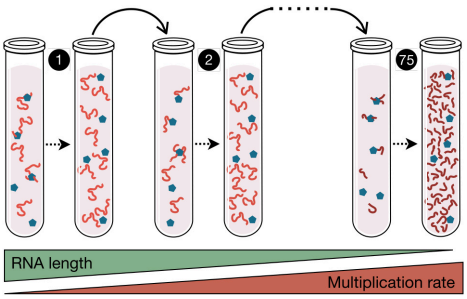

FIGURE 1
Spigelman's Monster. In this experiment published in 1967, the RNA from phage Qβ (red lines) and the enzyme responsible for its copy (blue pentagons) were incubated in the presence of nucleotides, and subjected to *in vitro* replication. The products of each polymerization reaction were serially transferred to new tubes. After 75 of these transfers, 83% of the RNA genome was eliminated, which in turn increased the multiplication rate of the RNA population. This seminal paper by Mills et al. (1967) demonstrated that darwinian evolution can occur in a test tube. Similar serial transfers experiments are used in the study of virus evolution.

Viruses in the RNA world

If viruses are considered as simple genetic parasites, it is easy to think that their predecessors could have been present in the ensemble of primitive replicators that made up the RNA or the pre-RNA world. Their emergence would require the generation of mutants capable of being copied without contributing to the generation of the resources necessary for the copying process. The occurrence of these mutants would be relatively frequent, given the low fidelity of primitive replication (Krakauer and Sasaki, 2002). However, since parasites are not capable of a self-copying process, their permanence would require that they remain stable in the environment for long periods of time. This means that resistance to damage caused by external physicochemical variables probably constituted a relevant trait subjected to the action of natural selection in the RNA world, which later could have led to the emergence of viral capsids. The current existence of defective virus mutants—which lack the ability to encode all the proteins needed to complete their infectious cycle, but can give rise to a progeny when these proteins are provided by other viruses—corroborates the ease of the emergence of this kind of parasites (Huang, 1973; Cole, 1975; Grande-Pérez et al., 2005).

A valid approach to investigate whether viruses could originate in the RNA world is to examine the genes that encode specific viral functions, such as genome replication. In this case, there are two activities that are only found in the viral world and, more specifically, in RNA viruses. These are the RNA-dependent RNA polymerase (RdRp), and the reverse transcriptase (RT). In the cellular world, the RdRp and the RT perform very specialized functions, but are not involved in genome replication. It is striking that both enzymes share with DNA-dependent DNA polymerases a structural folding located in the main catalytic domain (Iyer et al., 2005; Kazlauskas et al., 2016), which seems to correspond to an ancestral RNA recognition motif (RRM) that was probably crucial for replication of both RNA and DNA (Cléry et al., 2008), serving as a cofactor for ribozymes in the RNA world. Other virus genes that also contain the RRM are the rolling-circle replication endonucleases, a superfamily of helicases, and some protein-primed DNA polymerases.

A relevant question is whether the primitive genetic parasites also encoded the proteins necessary for the formation of capsids. The current view is that virus capsid proteins evolved from cellular proteins at different stages of evolution (Krupovic et al., 2019). The high prevalence of some structural domains, such as the single jelly roll (a β-barrel fold consisting of eight antiparallel β-strands organized in two sheets that form the opposite sides of the barrel) or the double jelly roll in the capsids of viruses infecting organisms belonging to the three domains of life is probably due to their antiquity (Abrescia et al., 2012; Yutin et al., 2018; Santos-Pérez et al., 2019). In contrast to this, other capsid proteins may have been acquired more recently.

In conclusion, there are multiple lines of evidence indicating that the primitive RNA world could gave rise to two parallel and interconnected paths that led to the emergence of the first cells and the first viruses. Since then, both worlds have evolved in close link, influencing each other and contributing to increase their diversity, as we will see in the following sections.

Role of viruses on the evolution of life

In the last decades, it has been recognized that viruses are major players in the history of life. They influence cellular evolution in different ways: acting as selective pressures,

behaving as vehicles for horizontal gene transfer, and creating new genes that provide evolutionary novelty.

Viruses as selective pressure

Virus infections favor organisms that possess some type of defense mechanisms. In the prokaryotic world, these mechanisms include abortive infections, modification of the molecules used as receptors, or the selection of systems, such as restriction endonucleases or CRISPRs, that recognize the virus genetic material and degrade it (Gao et al., 2020; Hampton et al., 2020; Wang et al., 2020b). In turn, the appearance of microorganisms that better resist the infection exerts a selective pressure on the viral population that will favor viruses able to counteract the cellular defenses (Geoghegan and Holmes, 2018; Stanley and Maxwell, 2018; Koonin and Krupovic, 2020). The repetition of these cycles involves concerted changes in the pathogen and its host, an *arms race* that accelerates the evolution of both the virus and the host (Buckling and Brockhurst, 2012; Koskella and Brockhurst, 2014; Watson et al., 2021).

A similar scheme works for viruses infecting eukaryotic cells. In the particular case of vertebrates (Kaján et al., 2020), it has led to the selection of the adaptive immune system that, with a limited number of genes, is able to cope with a wide variety of antigens. It is believed that one of the possible causes of the emergence of this system has been the need to deal with highly changing agents such as viruses (Holmes, 2004). Sometimes viruses reduce their virulence in certain host species. In this way, they do not induce a strong immune response and can remain multiplying longer in the same host. When, on the one hand, a protective immune response to the virus is generated and, on the other hand, virulence is reduced, a pacific coexistence between the virus and its host can be reached (Barnard, 1984; Holmes and Duchêne, 2019). When one of these viruses that are well adapted to a particular animal species is able to infect the human population, a zoonosis occurs. The absence of immunity to the pathogen, together with the fact that the virulence has not been attenuated in the new host, can lead to the emergence of a new disease (French and Holmes, 2020). COVID-19, AIDS, or the pandemics caused by new variants of influenza virus are just some examples of diseases caused by zoonotic viruses. In the short term, the introduction of these or other viruses in the human population can be very negative, but in the long term they may contribute to make us stronger against new threats.

Horizontal gene transfer in prokaryotes

Horizontal or lateral gene transfer allows innovations that arise in one group of organisms to be shared by a much larger set. Through transduction, both specialized and generalized, viruses are relevant agents of horizontal gene transfer in bacteria and archaea.

Specialized transduction takes places during the lysogenic cycle and implies that the virus, instead of destroying the cell, integrates its genome into the cell chromosome, being transferred to daughter cells by vertical transmission (Feiner et al., 2015). Most sequenced bacterial genomes contain at least one prophage (a phage genome integrated) and, in some cases, these elements constitute 10–20% of their total genetic material. Sometimes, after the lysogenic integration of a prophage, some signals occur that provoke its release, initiating a lytic cycle that will end with the production of new phages and the lysis of the bacterium. In the process, the phage frequently carries some cellular genes with it, which can be transferred to other bacteria in the course of new infections. Among the most frequently phage-transferred genes there are some that increase the virulence of the bacterium (Plunkett et al., 1999; Nishida et al., 2021), confer resistance to antibiotics (Gabashvili et al., 2020, 2022), help it to survive under difficult circumstances (Lindell et al., 2004, 2005; Sharon et al., 2007; Fridman et al., 2017), or allow the exploration of new ecological niches (Sullivan et al., 2005). In all cases, the advantages obtained from the integration of virus genetic material is what contributes to their persistence. Taking as an example the most abundant photosynthesizers on Earth –*Prochlorococcus* and *Synechococcus* (Flombaum et al., 2013)–, genes for the components of these cyanobacteria's photosystems are also present in the genomes of their phages (Mann et al., 2003), which seems to have amplified the evolution and expansion of these systems. Furthermore, the expression of the viral photosynthetic genes during infection enhances the photosynthetic activity of the host (Fridman et al., 2017), which increases viral replication, especially under low-nutrient conditions (Hurwitz et al., 2013).

Generalized transduction occurs during the lytic cycle of some viruses that, at the beginning of the infection, cause cellular DNA breakage (Thierauf et al., 2009; Waddell et al., 2009). When the progeny viruses are assembled, it is frequent that they do not distinguish between their own genome and the fragments of the cellular genome. The foreign DNA can be introduced into the virus capsids and, thanks to homologous recombination mechanisms, become integrated in the genome of the new infected microorganism.

Virus endogenization

The insertion of virus genomes into the cell chromosome is not exclusive to the prokaryotic world. When it happens in the chromosomes of the host reproductive cells of vertebrates and the inserted provirus is maintained over time, the process is called viral endogenization (Holmes, 2011a; Feschotte and Gilbert, 2012; Greenwood et al., 2018). After the initial insertion event, the provirus can spread in the cellular genome through a mechanism similar to that used by retrotransposons. During their dissemination, the viral sequences accumulate

mutations or undergo epigenetic modifications that inactivate the expression of most of their genes (Wolf and Goff, 2008; Rowe and Trono, 2011; Yang et al., 2022), so that after a time they can no longer lead to the production of infectious viral particles.

The integrated viral elements can disrupt the sequence of genes or their regulatory elements, causing loss of function. Consequently, individuals carrying this kind of exogenous DNA are normally eliminated by natural selection. However, sometimes viral genomes persist through generations, expanding in the population until they become fixed. In these cases, the DNA of viral origin becomes host DNA. The great number of viral insertions that have been fixed in vertebrate lineages (Hayward et al., 2015) represents a substantial source of genetic material that adds variability to genomes and that sometimes can provide benefits. For instance, the insertion of a provirus in a region near the pancreatic amylase gene has allowed this enzyme to be also expressed in saliva in humans, so that starch digestion can begin in the mouth (Ting et al., 1992). This is just an example of how virus regulatory sequences can modify gene expression in the host. In addition to the effects on individual genes, the process of provirus amplification can lead to the dispersion of viral regulatory elements throughout the cellular genome, producing large reorganizations of gene expression that can produce major evolutionary innovations (Feschotte, 2008; Koonin, 2016). The amplification process also facilitates the exchange of genetic material between different sequences of the genome, resulting in the loss of some regions, the duplication of others, or changes in their location (Hughes and Coffin, 2001; Bai et al., 2021).

Despite the serious consequences that the expression of virus genes usually has for the host, there are some situations in which their products have been recruited to perform cellular functions. A classic example is the expression of viral proteins that provide immunity to infection by related viruses (Arnaud et al., 2008). But perhaps the best example of genes from endogenous viruses that have allowed the emergence of novel functions are those encoding for syncytins. Syncytins are proteins that promote the fusion of a type of cells called trophoblasts to give rise to the formation of the syncytiotrophoblast, a cell layer that forms part of the placenta. Syncytin genes have their origin in a gene from retroviruses, specifically in the gene env, which has fusogenic activity (Dupressoir et al., 2011). It appears that the event of domestication of env genes of viral origin has occurred at least six independent times throughout evolution, in different mammalian species, including humans (Rawn and Cross, 2008).

Role of viruses in major evolutionary transitions

Darwinian theories explain evolution as the result of the progressive accumulation of heritable changes of small effect. However, these small changes cannot explain the increases in biological complexity associated with some evolutionary innovations that require drastic changes in the amount of genetic information and/or in the way it is processed. When these innovations involve the integration of entities from a lower level of organization into a higher one that comes to constitute a new level of selection, we speak of major evolutionary transitions (Szathmáry and Smith, 1995; Szathmáry, 2015). The continuous arms-race between viruses and their hosts, together with the frequent exchange of genetic material driven by viruses, has frequently contributed to the increase of biological complexity (Forterre, 2006; Koonin, 2016; Mizuuchi et al., 2022), as illustrated in some of the examples given in the previous section.

It is believed that the need to defend against primitive viruses promoted that the molecules carrying genetic information became grouped in compartments that, in addition to create barriers for the spread of parasites, facilitated the cooperation between non-parasitic replicators, contributing to stabilize the whole system, which emerged as a new selection unit (Szathmáry and Demeter, 1987; Higgs and Lehman, 2015). Within compartments, selection for increased genome size would also have been generally favored, as a mechanism to fix the most favorable genetic combinations. Primitive compartments could be protective micro-environments in mineral surfaces or simple lipid vesicles.

Mathematical modeling of replicator systems also suggests that resistance to parasites increases when information and function are separated in different molecules. It is easy to understand the advantages of this fact, since parasites could only take advantage of the functional molecules and not of the informative ones, which would facilitate their survival. However, we still do not know much about how DNA replaced RNA as an information storage molecule and about the appearance of the first protein catalysts, which probably coexisted for some time with RNA or RNA-like catalysts. The presence in retroviruses of reverse transcriptase activity, which converts viral RNA to DNA, has led some authors (Forterre, 2006) to think that this enzyme could be involved in the replacement of RNA for DNA as informative molecule, but this is something that with the current evidence cannot be affirmed with certainty.

Viruses may also have played a role in the origin of eukaryotic cells, which possess mitochondria and an internal membrane system that delimits a nucleus in which the genomic DNA is located. Mitochondria have their origin in a process of endosymbiosis between a bacterium and another cell, probably an archaeon. It is intriguing that in all them, the bacterial RNA polymerase has been replaced by a phage RNA polymerase (Filée and Forterre, 2005), although it is difficult to elucidate the advantages provided by this fact. The eukaryotic nucleus allows the uncoupling between transcription and translation. The discovery of viral factories (Fridmann-Sirkis et al., 2016; Chaikeeratisak et al., 2017), showing that some bacterial and eukaryotic viruses can establish a similar separation, led to

propose that the eukaryotic nucleus has a viral origin (Koonin, 2015; Bell, 2020; Takemura, 2020). However, this idea is not exempt of controversy and there are other hypotheses that state that the nucleus originated from invaginations of the plasmatic membrane in an ancestral prokaryote or has a symbiotic origin without virus participation.

Viruses at the planetary scale: The virosphere

The ongoing advances in high-throughput sequencing technologies and metagenomic analyses have been crucial to reveal the immense diversity and pervasive distribution of viruses throughout our planet (Zhang et al., 2018). As previously indicated, viruses seem to have colonized –presumably associated with cellular hosts– every place where life can be sustained, while exploring vast areas of the evolutionary sequence space. The viral composition of the Earth's biosphere is known as the virosphere.

The immensity of the virosphere and its diversity

The abundance of viruses in aquatic and terrestrial environments has been estimated to be 10–100-fold higher than that of unicellular organisms, although this ratio is dynamic and varies across ecological niches (Wigington et al., 2016). A milliliter of seawater can contain up to hundreds of millions of viruses (Bergh et al., 1989), and up to a billion of them can be found in a single gram of soil (Swanson et al., 2009). Thus, it has been estimated that the total number of viral particles in the planet is in the astronomical order of 10^{31} (Hendrix et al., 1999; Mushegian, 2020), outnumbering all cellular organisms.

Metagenomic studies aimed at unveiling the viral makeup of different environments around the globe have barely scratched the surface of the virosphere's diversity. A remarkable effort is the deep sequencing of samples collected during the Tara Oceans expeditions, in which nearly 35,000 samples from the ocean were collected between 2009 and 2013 in a ship transect around the globe (Brum et al., 2015; Roux et al., 2016; Gregory et al., 2019; Sunagawa et al., 2020). This data set consisted mostly of viruses with dsDNA genomes (excluding large portions of the virosphere), and yet, they are taxonomically more abundant than bacteria, archaea and micro-eukaryotes together (Abreu et al., 2022).

The rate of discovery of RNA viruses has lagged behind that of DNA viruses, mostly due to the technical requirement of converting RNA to DNA prior to sequencing. However, thanks to advances in metatranscriptomics, there has been a recent explosion in the amount of RNA virus data. The collection of RNA viruses in the databases doubled in 2016 after the metatranscriptomic analysis of numerous invertebrates (Shi et al., 2016), and doubled again in 2020 after sequencing just 10 L of sea water (Wolf et al., 2020). In 2022, by mining data from databases, the amount of known RNA virus sequences increased fivefold with the description of ∼330,000 novel RNA viruses (Neri et al., 2022), in addition to the discovery of ∼130,000 highly divergent RdRp sequences from novel viruses (Edgar et al., 2022). In each of these RNA virus data expansions, novel virus clades have been revealed. For instance, the exploration of the RNA virus oceanic composition has reached enough depth to reveal a novel group of capsid-less viruses, the *taraviricots*, which provides a missing link between retroelements and RNA viruses. The lineage of the taraviricots seems to predate the split between RdRPs and retrotranscriptases, suggesting that virus-like replicators encoding an RdRp are ancestral to RNA viruses and retroviruses, and probably derived from the primordial pool of replicators of the RNA world (Zayed et al., 2022).

Organization of the virosphere and its challenges

Considering the colossal diversity of viruses, classifying the virosphere is not a trivial task. For decades, the main virus classification scheme was based on the nature of the packaged nucleic acid and traits such as virion morphology, replication strategy, host organism and type of disease. Given the explosion of virus sequence information fueled by metagenomics and for which no phenotypic information is available, these criteria seem no longer practical. The 7 Baltimore classes provided an organization framework based on the nature of the packaged nucleic acid and the strategies by which mRNAs get produced from these DNA or RNA genomes (**Figure 2**; Baltimore, 1971). However, the Baltimore classification does not accurately represent the evolutionary relationships among viral lineages.

A phylogenetic approach to virus classification is difficult, due to the absence of a common marker for virus classification. Moreover, viral genomes are modular and virus evolution is greatly shaped by lateral gene transfer. Despite these difficulties, by using phylogenetic analyses of virus hallmark genes and gene sharing networks, a new "megataxonomy" of viruses has recently been established (Koonin et al., 2020). In this new classification of viruses, a new taxonomic rank –the realm– had to be defined to distinguish between virus taxa with no common ancestry. The six virus realms are: (i) *Riboviria*, which contains all RNA viruses and retroviruses, and that may have derived from the primordial pool of replicons, (ii) *Monodnaviria*, with viruses whose DNA genomes (most single stranded) encode an endonuclease for replication, (iii) *Varidnaviria*, dsDNA viruses with vertical jelly roll capsid proteins, (iv) *Duplodnaviria*, dsDNA viruses with HK97-fold capsid proteins, (v) *Adnaviria*, which helical virions containing A-form DNA genomes, and (vi) *Ribozyviria*, whose members have circular RNA genomes

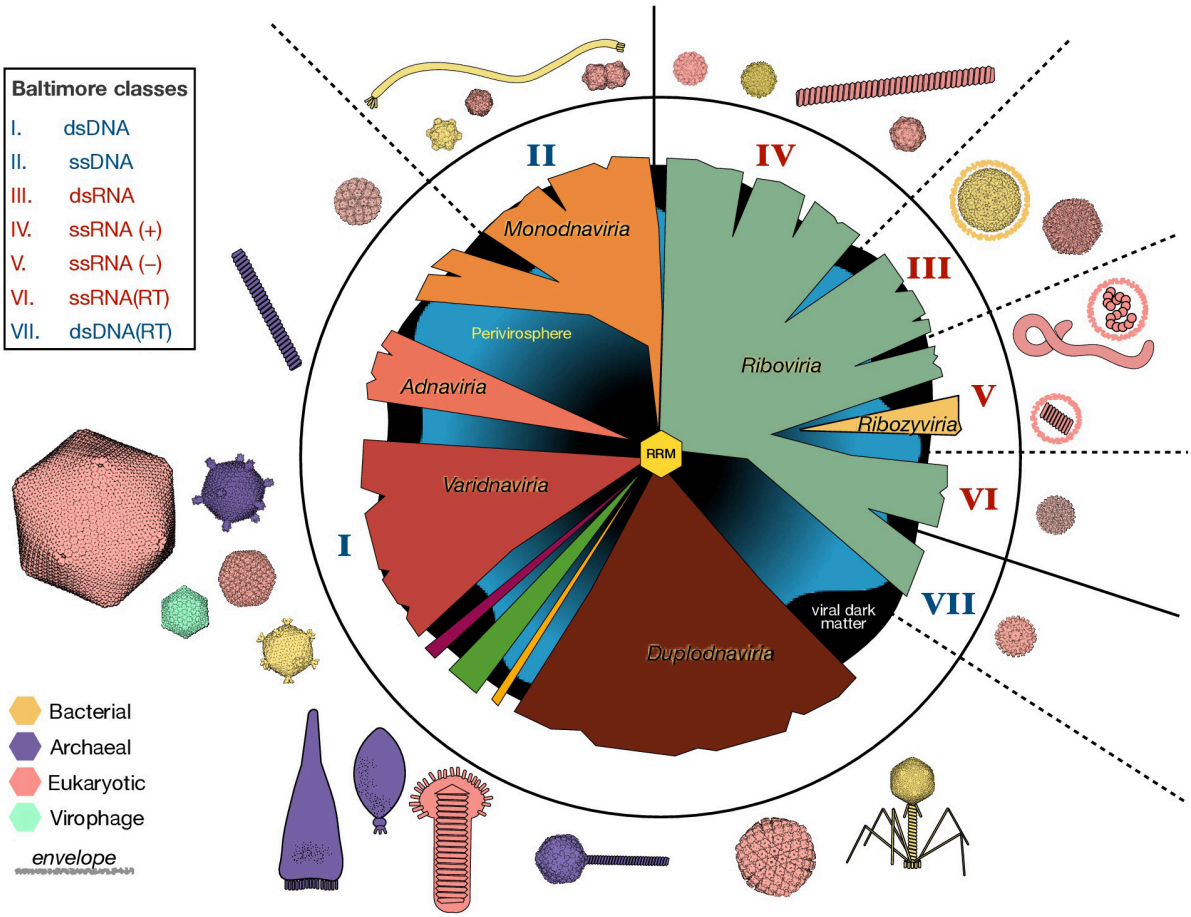

FIGURE 2
Organization and Diversity of the Virosphere. The vast genetic diversification of viruses is reflected in a variety of viral genome lengths (1.8 Kb–2.5 Mb) and organizations (e.g., circular vs. linear, segmented or not), virion sizes (17 nm–1 μm) and morphologies, host ranges, and in the types of interactions between viruses and their hosts. Here, we represent the six virus realms in the middle circle, as they emerge from an ancestral RNA-recognition motif (RRM) in the center. The roman numbers indicate the nature of the packaged nucleic acid according to the Baltimore classification scheme, where blue numbers represent DNA and red numbers RNA. Non-viral mobile genetic elements (perivirosphere) and the unknown viral sequences (viral dark matter) are included in the background. Some representative virus morphologies from viruses infecting all three domains are depicted with approximate relative sizes and colored according to their host. The available molecular structures were downloaded from RCSB Protein Data Bank and visualized with UCSF ChimeraX. ss, single-stranded; ds, double-stranded; RT, retro-transcriptase.

similar to viroids (**Figure 2**). Although different viruses probably emerged at different times and through different routes, all viruses in these six realms use replication proteins derived from the ancestral RNA-recognition motif (RRM), that probably has its origin in the RNA world (see section "Viruses in the RNA world") and subsequently predated cellular life (Koonin et al., 2020). This new classification scheme has been recently ratified by the International Committee on Taxonomy of Viruses.

The six virus realms currently contain more than 200 viral families. However, the detection of viral sequences with no recognizable homologs is far from uncommon, and a vast portion of sequence data remain unclassified. Current advances in sequence analysis are key to disentangle this "viral dark matter" (Krishnamurthy and Wang, 2017). Another challenge to virus classification is the existence of viral chimeras. While recombination between viral genomes within viral groups is frequent (Pérez-Losada et al., 2015), gene sharing extends beyond the borders between the virus realms. The cruciviruses, for example, is a group of chimeric viruses whose circular ssDNA genomes encode a capsid protein also found in RNA viruses (de la Higuera et al., 2020).

The virosphere also include mobile genetic elements that are not bona fide viruses, such as viroids, plasmids and retroelements, which have been relegated to the "perivirosphere" (Koonin et al., 2021). The delimitation of the perivirosphere, however, is blurry, as mobile genetic elements can acquire capsid genes to become viruses (Krupovic et al., 2019; Koonin et al., 2022). For instance, ssDNA viruses belonging to the

realm Monodnaviria appear to have emerged multiple times from plasmids by capturing capsid genes from RNA viruses (Kazlauskas et al., 2019).

Impact of viruses on Earth's biogeochemical cycles and organismal biodiversity

Microbes are the main drivers of the biogeochemical cycles that define ecosystems, but they also are susceptible to virus infection, e.g., it is estimated that ∼20% of the bacteria in the oceans are lysed by viruses per day (Suttle, 1994). These ongoing viral infections shape microbial community structures and are necessary for nutrient cycling (Dominguez-Huerta et al., 2022).

Viruses can impact food webs and biogeochemical cycles in different ways. Cell death caused by lytic viruses leads to the release of dissolved organic matter, which is then available for its consumption by low trophic levels. This process, called "viral shunt," allows nutrients to be reused by other microorganisms rather than getting passed on to higher trophic levels (Wilhelm and Suttle, 1999; Zimmerman et al., 2020). The released organic matter from viral lysis can be particularly rich in nucleotides and amino acids (Ankrah et al., 2014). Thus, the viral shunt is important for the cycling of not just carbon, but other nutrients such as nitrogen (Shelford et al., 2012), phosphorus, and iron (Gobler et al., 1997; Poorvin et al., 2004). A complementary process, the "viral shuttle," describes how viruses help the export of carbon by sinking organic matter to the bottom of the ocean (Guidi et al., 2016; Sullivan et al., 2017). The virus shuttle has been observed in algal blooms –often terminated through virus infection (Brussaard, 2004)–, in which the production of virally induced polymeric matrices enhances the sink of cellular aggregates to the sea floor (Laber et al., 2018). Virus composition have proven to be a great predictor of carbon flux in aquatic (Guidi et al., 2016; Kaneko et al., 2021) and terrestrial ecosystems (Emerson et al., 2018; Trubl et al., 2018; Starr et al., 2019), which further demonstrates the importance of virus infections in biogeochemical cycles.

The lysis of cells upon infection can have a significant effect on top-to-bottom control on biogeochemical cycles. However, viruses do not need to kill their host to impact ecosystems. During infection, viruses reprogram the cellular metabolism to divert resources to progeny production, which at times involves the expression of virally encoded metabolic genes that were likely acquired from cells (see section "Horizontal gene transfer in prokaryotes") (Howard-Varona et al., 2020). Since a significant portion of the Earth's microbiome is infected at any given time, these viral "auxiliary metabolic genes" may carry functions that can have an impact on planetary processes (Zimmerman et al., 2020). The analysis of auxiliary metabolic genes from virus genomic data can be used as a proxy to understand the impact that viruses exert on ecosystems, including the control of sulfur and nitrogen cycling in the ocean (Roux et al., 2016).

Viruses are also community structure regulators. They can propagate more effectively when their host is in abundance, thus, according to the "kill-the-winner" hypothesis, lytic viruses help maintain a balance and promote diversity in microbial communities (Thingstad and Lignell, 1997). This seems the case in the collapse of algal blooms by viral infection (Suttle, 2007). However, there are other examples where a particular species seems to be pervious to viral infection and become dominant (Yooseph et al., 2010). This scenario, called "king-of-the-mountain" (Giovannoni et al., 2013), suggests that high levels of recombination in these organisms allow the rapid adaptation of defense systems against their viruses (Zhao et al., 2013). This positive feedback loop in the co-evolution of virus and host is another example of the arms race described in section "Viruses as selective pressure," by which viruses fuel organismal evolution and diversification (Giovannoni et al., 2013).

The realization of the important roles that viruses play in Earth's habitats underscore how pertinent it is to consider genetic parasites in the study of biological systems. For that matter, it is relevant to also look at those ecological niches where life defies its own physicochemical boundaries.

Viruses in extreme environments

The study of life in harsh conditions is essential to astrobiology. On one hand, extremophilic organisms shed light on the limits of habitability, i.e., the range of environments capable of supporting life in terms of temperature, pH, salinity, pressure, humidity and radiation (An Astrobiology Strategy for the Search for Life in the Universe, 2019). On the other hand, the biology of extreme environments can help to understand the conditions for life emergence (Baaske et al., 2007; Morowitz and Smith, 2007; Weiss et al., 2016).

Extreme environments are no exception to the wide distribution of viruses. They can thrive at temperature ranges between −12 and 96°C (Munson-Mcgee et al., 2018), pH values from 1 to 11 (at the least), saturated salinity (Pietilä et al., 2013), and extreme dryness (Gil et al., 2021). According to this, viruses have been found in hydrothermal vents (Russell and Hall, 1997), including those located at great depths in the oceans (Thomas et al., 2021), in oil fields (Zheng et al., 2020), in the soils of hyper-arid deserts (Zablocki et al., 2015), at hundreds of meters in the subsurface (Daly et al., 2019), or in places with salt concentrations far above the optimal for most living beings (Santos et al., 2012). Many of these locations share environmental characteristics that also occur in extraterrestrial environments. For example, submarine hydrothermal vents could be considered a terrestrial analog of those postulated to exist at the bottom of the inner oceans of the icy moons of some gas giants, such as Enceladus, Europa, or Ganymede. Oil field

biology could help us to understand whether life could develop in the hydrocarbon lakes that exist on Titan, while the study of the subsurface could help us to figure out the possible life that existed (or still exists) on Mars. As we have already indicated, wherever there is cellular life there are also viruses, which reinforces the idea that the search for these entities could be used as a proxy for the search for life beyond Earth, something that is not receiving all the consideration it deserves.

Like in other ecosystems, extreme environment viruses can encode auxiliary metabolic genes that may be transferred between organisms. In hydrothermal vents, these genes are involved in metabolic pathways that are critical for survival in this environment including –but not limited to– sulfur-oxidizing enzymes (Anantharaman et al., 2014; Castelán-Sánchez et al., 2019). Viruses can also increase the adaptability of multicellular organisms to extreme environments. The triple-symbiosis in the thermotolerant terrestrial plant *Curvularia protuberata* is a classic example of how a virus infection –*via* fungal interaction in this case– can provide tolerance to heat stress (Márquez et al., 2007).

Archaea are typically found in the whole spectrum of extreme environments, and their viruses (archaeoviruses) –given the striking diversity of their unique morphologies (**Figure 2**)– are one of the most captivating types of extremophilic viruses. The virions of archaeal viruses have shapes suggestive of lemons (*Fuselloviridae*), bottles (*Ampullaviridae*), droplets (*Guttaviridae*), or spirals (*Spiraviridae*) (**Figure 2**; Baquero et al., 2020). Most archaeoviral families encode proteins with no detectable homologs in other viral groups, which hinders their assignment into any of the established realms of the virosphere.

The molecular adaptions of viruses to withstand the extreme extracellular conditions in these habitats remain underexplored, but structural studies have revealed special features in virions isolated from these locations. Some archaeoviruses package their genomes in the A-form of DNA (DiMaio et al., 2015). This DNA conformation is more compact than the more usual B-form, suggesting that this DNA stabilization mechanism is an adaptation to extreme conditions (Wang et al., 2020a). Another archaeovirus, *Aeropyrum coil-shaped virus* (not only the largest known ssDNA virus, but also stable at 95°C), forms a nucleoprotein with its circular ssDNA genome, which coils into an intertwining fiber that compacts adopting the shape of a spiral (Mochizuki et al., 2012). Archaeoviruses can also incorporate ether lipids from their extremophilic host in their virions, which seem to protect against chemical stress (Boyd et al., 2013). Archaeal lipids are also present in the structure of hyperthermophilic spindle-shaped viruses, but its location remains under debate (Quemin et al., 2015; Han and Yuan, 2022).

These structural adaptations to attain virion stability in extreme conditions highlight the ability of viruses to disperse within, but also beyond, the natural habitat of their hosts. Spindle-shaped archaeoviruses are found in a wide variety of extreme environments around the globe, including deep-sea hydrothermal vents, hypersaline environments, anoxic freshwaters, cold Antartic lakes, terrestrial hot springs and acidic mines (Wiedenheft et al., 2004; Krupovic et al., 2014), even though their hosts are highly adapted to isolated geographies (Whitaker et al., 2003). However, other studies show that viruses present in hydrothermal vents have restricted host range and are not widely distributed among vent sites (Thomas et al., 2021). All this raises questions about how viruses disperse at a planetary scale (Stedman et al., 2006), and possibly beyond (Griffin, 2013).

Viral (inter?) planetary distribution

Viruses can disperse passively throughout different geographic scales in their stable virion form. According to the seed-bank hypothesis, viruses can be transported by oceanic or wind currents, and be recruited in ecosystems where they have an opportunity to thrive (Breitbart and Rohwer, 2005). Thus, viruses constitute a global genetic reservoir from which local virus communities reshape within biomes across ecological zones (Brum et al., 2015; Gregory et al., 2019).

Viruses disseminate in the atmosphere associated with dust particles or aerosolized liquid droplets. Therefore, vast amounts of viral particles can be globally spread from events such as desert storms or the release of sea spray (Baylor et al., 1977; Griffin et al., 2001). Some of these virions can be transported past the planetary boundary layer into the troposphere, further favoring long-range viral dispersal (Reche et al., 2018). The deposition of viruses from the troposphere has been observed for air masses coming from both the ocean and the desert, which reflects the magnitude of virus movement across the globe (Reche et al., 2018).

Given the potential of viruses to disperse at a planetary scale, it is likely that some of the viral particles escape the Earth. The same possibility could be applied to extraterrestrial virospheres. In addition to viral adaptations to extreme environments, viruses have been shown –in experimental setups– to survive high doses of UV radiation, ionizing radiation, X rays, high vacuum, desiccation, or microgravity conditions (Koike et al., 1992; Hegedüs et al., 2006; Horneck et al., 2010; Vaishampayan and Grohmann, 2019; Sharma and Curtis, 2022), a fact that would support their possible involvement in the transport of genetic material between planets. Particular consideration deserves the finding that under simulated hot spring conditions, viruses can fossilize upon the aggregation of silica deposits (Laidler and Stedman, 2010; Orange et al., 2011). Since virus silicification renders virions resistant to desiccation, it has been suggested as a potential long-range dispersal mechanism (Laidler et al., 2013). Although the ability of viruses to endure outer space conditions and disperse cosmically remains poorly understood, the field of astrobiology would benefit

from considering viruses as potential carriers of life signatures (Griffin, 2013; Berliner et al., 2018).

Viruses as biosignatures

The search for extraterrestrial life relies on the study of indicators of extant or past life. Biosignatures are defined as detectable substances, objects or patterns that are likely originated from life processes, and not abiotically (Chan et al., 2019). We could detect extraterrestrial life: (1) remotely, by the observation of astronomical bodies; (2) *in situ*, through space exploration, or (3) in transported samples. A good biosignature should be abundant, unlikely to originate in the absence of life, persistent in time (especially for the detection of past life or in transported samples), and easily detectable (An Astrobiology Strategy for the Search for Life in the Universe, 2019). The expectations of what life looks like in other planets is largely based on our understanding of our own biosphere. However, special attention needs to be put into universal and agnostic biosignatures that could apply to unknown forms of life.

The abundance and importance of viruses in Earth's history and processes suggest that virus-like agents should be abundant members of other biospheres. Would these extraterrestrial molecular parasites form detectable structures like viruses in our planet do?

Icosahedral capsids as agnostic biosignatures

To increase their stability and protect their genetic material, viruses coat their genomes with a protein capsid encoded in its own genome. However, at least in the life we know, it is impossible to translate a protein large enough to package its own code (i.e., the molecular weight of a codon is ∼1,000 g/mol and amino acids are 75–203 g/mol, making genetic information bulkier than its product). This size dilemma is overcome in viruses by using capsid proteins with the ability to self-assemble. Thus, a multimeric container can be formed from identical copies of a single gene product. In this case, these protein subunits have the exact chemical and spatial properties and must interact with each other in an equivalent or quasi-equivalent manner, which is achieved by the formation of symmetrical arrangements (**Figure 3A**; Caspar, 1956; Watson and Crick, 1956). Structural symmetry is a staple of the genetic economy characteristic of viruses, as it allows the creation of containers using the minimal amount of information.

Icosahedral symmetry is a prevalent virus architecture, as the spherical containers it provides are the most efficient in terms of volume-to-surface ratio. Capsids with icosahedral symmetry have emerged multiple times throughout the evolution of the virosphere (Krupovic and Koonin, 2017), and are present in viruses infecting all domains of life, and packaging any form of nucleic acid (**Figure 2**). The principle of quasiequivalence (Caspar and Klug, 1962) explains, by a simple rule of triangulation, how any multiple of 60 subunits can assemble into a wide spectrum of structures with icosahedral symmetry. Peptides as short as 24 residues have been shown self-assembly into icosahedral structures (Matsuura et al., 2010), suggesting that these apparently complex architectures could be possible even at early stages of life. Moreover, the icosahedral arrangement of capsid proteins is a direct consequence of a free energy minimization of the chemical interactions between subunits (Zandi et al., 2004), which allows the spontaneous assembly of highly stable structures (Garmann et al., 2019). The geometrical and thermodynamic constraints that pressure the selection of icosahedral symmetries for the formation of molecular containers are probably universal, and as such, in principle should be expected in extraterrestrial virospheres. However, constraints different from those operating in terrestrial life could be relevant in the architecture of the capsids of possible extraterrestrial viruses. When talking about life outside the Earth, it is necessary to take special care that the properties of what we know do not bias our expectations about what we can find. Even in terrestrial viruses, there are other types of capsids apart from the icosahedral ones (**Figure 2**). An example already cited is the wide variety of morphologies exhibited by archaeal viruses. Another problem that can arise is the detection of structures similar to viral capsids, but whose origin is due to chemical or geological processes. Distinguishing whether an apparently biological structure originates from biotic or abiotic processes is not easy, as has been shown on several occasions, including the controversial detection of microfossils in the ALH84001 meteorite.

Viral structures are good candidates for biosignatures as they are probably abundant components of other biospheres and –given their major role in protecting and dispersing information– can be stable in a wide range of environmental conditions. Despite the small size of viral particles, a plethora of techniques can be used to detect and analyze them. Among them, the most relevant to obtain meaningful structural information are electron microscopy, atomic force microscopy (AFM) and X-ray crystallography. Electron microscopy (cryo-EM in particular) is a powerful method broadly used for structure determination (Jiang and Tang, 2017). However, it requires complex sample manipulation and its throughput is still limited. AFM analyzes the contour of a sample by scanning it mechanically with an ultrafine probe, obtaining virus structure information at high resolution (**Figure 3B**). AFM technology is advancing at a fast pace, allowing the analysis of several mm^2 per second at nanometer resolution (Marchesi et al., 2021). AFM devices can be very compact, which is a clear advantage for biosignature detection *in situ*, as they can be –and have been (Smith et al., 2008)– incorporated into spacecrafts. The usage of

X-ray diffraction for the study of virus morphology has seemed to succumb to more convenient techniques such as cryo-EM; however, new advances are revolutionizing the field (Shoemaker and Ando, 2018).

The cutting edge of X-ray diffraction technology is the X-ray free-electron laser (XFEL). The XFEL produces coherent and ultra-bright X-ray femtoseconds-long pulses that can be applied to an aerosolized sample (Chapman et al., 2011). This not only allows for the observation of ultra-fast protein motion at near-atomic resolution (Pandey et al., 2020), it could –most importantly to our matter– scan a sample for molecular structures of interest at a great speed and depth. The XFEL offers a great flexibility in terms of temperature and sample conditions, and it can work under non-equilibrium while penetrating through matter without interfering with external fields (Meents and Wiedorn, 2019). Because icosahedral capsids are highly symmetrical, XFEL can produce clear diffraction patterns from single viral particles with no previous crystallization (**Figure 3B**; Seibert et al., 2011), which –coupled with machine learning (Suzuki et al., 2020)–, would offer great sensitivity for the detection of icosahedral symmetry in a complex sample. To our knowledge, however, there has not been any astrobiological research efforts in this direction. XFEL technology requires big structures to operate, which would only allow the analysis of samples transported back to Earth; however, technological improvements aimed at reducing the size of XFEL instrumentation are underway (Rosenzweig et al., 2020). Although abiotic molecules with icosahedral symmetry exist –e.g., fullerene and closo-carboranes–, these are much smaller than what is expected from biotic icosahedral structures and should be easily discriminated by the methodologies described above.

Other viral signatures

Biosignatures based on the detection of chiral excesses, isotopic fractionation or the products of degradation of the biomolecules composing viruses can also be considered as possible proofs of their existence in extraterrestrial biospheres that may even be extinct. Carrying out experiments devoted to determining the chemical and morphological footprint left by viruses immobilized on mineral substrates, before and after being subjected to simulated space conditions in laboratory chambers, can be very useful to know what we can expect to find in another place of space where viruses have been present at some point in its history.

Additionally, there are biosignatures that are not necessarily virus-specific, but that would imply their detection in one way or another. For instance, the usage of chemical polymers to store and transmit information is a likely possibility in other life forms. Given the role of viruses as reservoirs and vehicles of genetic information on Earth and their abundance and dispersibility, there are better chances to detect virus-like information packages than more complex life-forms. There are procedures that allow the concentration and purification of viral particles (flocculation with polyethylene glycol and salt, or differential ultracentrifugation) that could be used to increase the probability of life detection in a sample. Sequences technologies based on electrochemistry (e.g., Nanopore) could then be used to detect information patterns from possible biopolymers in a sample, even if they are distinct from DNA or RNA (Rezzonico, 2014). Nanopore-based technology is very compact and, in synergy with other techniques, has successfully been used in samples from a Mars-analog environment, where DNA –including viral sequences– was detected at very low concentrations (Maggiori et al., 2020).

Other life detection methods that are being investigated include immunoassays using antibody microarrays for the recognition of a broad spectrum of common prokaryotic antigens (García-Descalzo et al., 2019). Virus hallmark motifs should be added to this collection of antigens. Antibodies for the detection of the jelly roll fold characteristic of many viral capsids or the ancestral reconstruction of the primitive RRM that antecedes most of viral and cellular replicases (Koonin et al., 2022) would be exceptional candidates for future versions of the LDChip [Life Detector Chip; (García-Descalzo et al., 2019)].

Additional roles of viruses in the detection of life have been discussed (Berliner et al., 2018), including the remote identification of virally provoked phenomena such as the formation of huge calcium carbonate deposits upon lytic termination of algal blooms described in section " Impact of viruses on Earth's biogeochemical cycles and organismal biodiversity." Viruses may have also played a role in the transformation of microbial communities, including the generation of microbial mats and stromatolites. If we take into account that the most ancient evidence of life on Earth correspond to stromatolites, the study of whether viruses can manipulate the microbial metabolism to influence carbonate precipitation and other processes related to the lithification of microbial communities can be highly relevant to interpret mineral biosignatures through geological time (White et al., 2021). Research efforts toward the study of viral biomarkers have been scarce, but increasing understanding of the role of viruses in origin and evolution of life will likely put viruses in the spot-light of astrobiological investigations in the near future.

Viruses as a tool for determining general evolutionary principles

A fundamental topic in Astrobiology is the study of how life has diversified since its origin to give rise to all the forms in which it currently manifests. However, the study of the evolutionary process is not an easy task, because its primary cause -the generation of mutations- occurs by chance and is

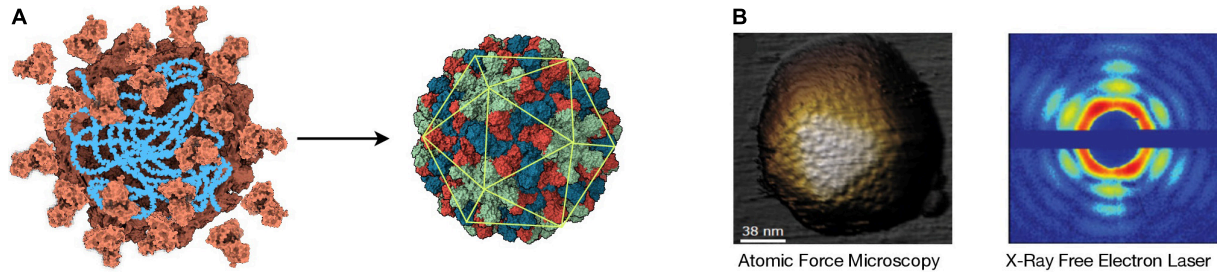

FIGURE 3
Icosahedral Symmetry Formation and Detection. **(A) Left:** Cartoon representing the assembly of identical capsid protein subunits (orange) around a nucleic acid (blue). **Right:** The resulting virion has icosahedral symmetry, as indicated by the overlapping icosahedron (green). The different colors represent the three different positions that the subunits can occupy in this $T = 3$ structure that corresponds to the plant-infecting RNA virus TBSV (pdb: 2TBV). In this case, a capsid is assembled from 180 copies of the capsid protein. The structural data were downloaded from RCSB Protein Data Bank and visualized with UCSF ChimeraX. **(B) Left:** Image of an adenovirus particle detected by atomic force microscopy (AFM) [adapted from de Pablo and San Martín (2022)]. **Right:** X-ray diffraction pattern caused by a mimivirus particle using X-ray free-electron laser (XFEL) [adapted from Ekeberg et al. (2015)].

subjected to multiple contingencies that condition the action of selective processes. Moreover, its results usually require long periods of time, which makes this process difficult to observe in real time. A valid approach consists of analyzing how biological diversity has changed throughout life history and as a function of the physicochemical parameters of the environment. However, natural environments are complex and are affected by multiple interacting variables whose values cannot be controlled and whose history is frequently unknown. Therefore, relationships between environment, phenotype and genotype are difficult to establish through this strategy, which causes that many basic questions remain unanswered. These include the effect of mutations on fitness, the relative contribution of natural selection and genetic drift, how diverse and reproducible are adaptive pathways, whether error rate is a character subjected to the action of natural selection, or how interactions between mutations and mutants influence adaptation. Experimental evolution arises in this context of uncertainty to provide a framework in which variables can be controlled by the experimenter, allowing a more precise relationship between environmental changes and the response of organisms to them (Bell, 2008; Elena et al., 2008; Kawecki et al., 2012; Elena, 2016; Roux et al., 2016; Geoghegan and Holmes, 2018; Van den Bergh et al., 2018). Populations used in these studies must evolve fast and be easy to handle, two conditions met by many microorganisms, including viruses.

In addition to all exposed above, viruses, and particularly those with an RNA genome, can be a good model for the study of the molecular evolution processes that took place before the emergence of cellular life, when natural selection acted on primitive RNA or RNA-like replicators. RNA viruses, together with viroids and some subgenomic elements, are the only current biological entities that use RNA to store genetic information. This RNA, although copied by protein enzymes encoded in the viral genome, is actively involved in its own replication, not only by providing binding sites for enzymes and ligands, but also because its copying capacity depends on the three-dimensional structure it adopts according to its primary sequence. Finally, it is widely documented that RNA viruses form population structures in quasispecies, similar to those described theoretically for the ensembles of primitive replicators. All this, together with their lower complexity with respect to cellular systems, means that RNA viruses can be considered a suitable model for the study of evolution in the hypothetical RNA world.

It is impossible to include in this review all the evolutionary questions that viruses have helped to understand. Here we do not intend to make an exhaustive list, and we will simply describe some illustrative examples.

Effect of mutations on fitness

This question has been the focus of many experimental and theoretical studies (Cuevas et al., 2012; Vale et al., 2012; Acevedo et al., 2014; Bataillon and Bailey, 2014; Minicka et al., 2017). Domingo-Calap and Sanjuán (2011) carried out a comparative study of three phages with a ssRNA genome and three others with a ssDNA genome. Their findings showed that mutations usually have negative effects and confirmed that RNA viruses accumulated mutations at a faster rate than DNA viruses (Sanjuán et al., 2004; Carrasco et al., 2007; Domingo-Calap et al., 2009).

Small genomes, as such of RNA viruses, contain highly compacted information in which mutations are more prone to interact than in larger genomes. This kind of interactions, called epistasis, can make it very difficult to determine the relative contribution of particular mutations to the observed

changes in fitness. Epistatic effects can change depending on the environment, as it was demonstrated in a study carried out with Tobacco Etch Virus that compared the effect of pairs of mutations in different hosts (Cervera et al., 2016). Epistasis may also determine evolutionary trajectories, depending on the first mutations that arise during adaptation (Zhao et al., 2019). A particular case of epistasis are compensatory mutations that are neutral or deleterious mutations that turn beneficial in the context of other deleterious mutations. Their existence is widely documented in drug-resistant mutants where the fitness cost that usually has the resistance mutation is compensated by the acquisition of new mutations (Buckheit, 2004; Boucher et al., 2009).

Standing genetic diversity vs. diversity generated *de novo*

An evolutionary relevant question is the relative importance for adaptation of the mutations generated *de novo* vs. those pre-existing in populations. Quantitative genetics studies predict that, initially, natural selection acts on the pre-existing genetic variation, while mutations generated *de novo* are more relevant in the long term. The high genetic diversity contained in RNA virus populations provides an excellent system to study this question.

It has been shown that minority genomes, which had selective advantages against previous selective pressures, can speed up adaptation when the population faces a similar condition (Ruíz-Jarabo et al., 2002; Briones and Domingo, 2008). Recent studies carried out with hepatitis C virus showed that the diversification of the mutant spectrum that takes place when it is propagated under constant conditions included the presence of some variants that were resistant to particular antiviral drugs to which the virus had not been previously exposed (Gallego et al., 2020). Other study showed that populations of bacteriophage Qβ propagated at 37°C contain a considerable fraction of low-frequency mutations that may facilitate adaptation when the virus is exposed to 43°C (Somovilla et al., 2022). The conclusion is that the spread of populations on the space of sequences is unavoidable and play an essential role in adaptation (Koelle et al., 2006; Quakkelaar et al., 2007).

Influence of error rate on adaptation

High error rates can constitute a great advantage for RNA virus adaptation. However, since most mutations are deleterious, a possible consequence is that small increases in the error rate could lead to the generation of increasingly unfit populations, which could ultimately become extinct. There are few studies on the effect of environmental conditions on the virus error rate, which would be highly relevant to determine its importance for virus permanence in a particular environment. In contrast, there are many experiments in which the error rate has been artificially modified through the use of mutagens (Domingo et al., 2021). This kind of assays has led to considerable progress in the knowledge of the relationships between error rate, adaptive capacity and the risk of population extinction. Many of the experiments performed in this context were inspired in the concept of error catastrophe of quasispecies theory (Biebricher and Eigen, 2005). It was found that, although in most cases viral populations eventually lost their infectivity when treated with mutagens, no evidence was found of the predicted loss of genetic information. Rather, what seemed to happen was a progressive decrease in fitness due to the increased accumulation of deleterious mutations (Bull et al., 2007).

In some cases, viral populations subjected to mutagenic action were able to select mutants with higher fidelity than the original virus (Pfeiffer and Kirkegaard, 2003; Agudo et al., 2010). Sometimes, these mutations only manifested their effect in the presence of the mutagen (Cabanillas et al., 2014), but in others, the increase in fidelity was maintained regardless of whether the mutagen was present or not (Agudo et al., 2016). These results are highly relevant, since they indicate that the error rate is a character subjected to the action of natural selection, and can be modified. Subsequent experiments carried out with fidelity mutants showed that they had a disadvantage in some infections (Vignuzzi et al., 2006; Coffey et al., 2011), suggesting that the presence of a complex mutant spectrum is necessary for the expression of certain phenotypic traits that require cooperation between different components of the population.

Finally, studies carried out with mutagens have made it possible to identify a new pathway to extinction based on the increase of defective interactions in the mutant spectrum (Grande-Pérez et al., 2005).

Adaptation to adverse conditions in the extracellular medium

For a virus, as for primitive replicators, the capacity to resist adverse environmental conditions in the inter-replication periods is as important as the capacity to replicate successfully. The ability of viruses to increase their stability in the environment has been explored in several studies (McGee et al., 2014; Costello et al., 2015; Hanson et al., 2015; Lázaro et al., 2018; Whittington and Rokyta, 2019), as well as the possible trade-offs between increased stability and the virus multiplicative capacity (De Paepe and Taddei, 2006; Dessau et al., 2012). A study carried out with vesicular stomatitis virus evolved through a regime that involved an increase in the time that the virus spent out of the host showed the

selection for virus variants with increased extracellular survival and lower fecundity (Elena, 2001; Wasik et al., 2015), which seemed to confirm the trade-off. In contrast to this, other studies carried out with phages — propagated through successive cycles of exposure to adverse extracellular conditions followed by replication at optimal conditions — showed that it was possible to increase the resistance to the harsh extracellular environment without apparent trade-offs on virus replication (McGee et al., 2014; Lázaro et al., 2018).

Discussion

For the first time in human history, in recent decades we have reached a technological level that allows us to investigate the possible existence of life on other worlds. This is a huge challenge that, if achieved, would help us understand the meaning of life and which of its properties are essential and must be present in any of its manifestations, be they terrestrial or extraterrestrial. The search for life outside of Earth is inevitably biased by the properties of the life we know. Although this fact introduces limitations, it also provides a starting point that can guide us, provided that we are able to stop thinking about the concrete manifestations of life on Earth and start thinking in terms of the processes that life carries out. In this sense, life can be understood as a process that transforms energy using a genetic substrate to store the instructions on how to do it. Viruses do not transform energy by themselves, but they contain in their genome all the necessary instructions to manipulate the metabolism of the cells they infect, so that they manage to multiply and transmit between cells and organisms. Ultimately, the idea of metabolic autonomy is nothing more than an illusion, since, to a greater or lesser degree, most of the living beings on Earth depend on others to exist.

The transition between living and inert matter was probably a continuous process in which it is not clear where to draw the dividing line. In addition to viruses, on Earth there is a set of biological molecules –viroids, transposons, plasmids, prions– that are capable of moving between cells and organisms, producing copies of themselves with a certain independence, and that interact with life in multiple ways, accelerating its evolution and contributing to shaping it. If we want to understand life in its broadest sense, all these entities must be included in the field of study of astrobiology. We do not know if the existence of a biosphere without viruses or similar pathogens is possible, but if this were so, it would almost certainly be less rich and diverse than the one we know. As we have widely described in this review, viruses can help understand the origin, evolution, distribution, and future of life. Therefore, we encourage the astrobiology community and the funding agencies to give more support to the research on these fascinating entities, as well as to progress in the design of techniques that allow us to include viruses as signatures of the existence of life.

Author contributions

Both authors designed the structure of this review, wrote the text, carried out the bibliographic search, reviewed the manuscript, and approved the submitted version.

Funding

This work has been funded by grant PID2020-113284GB-C22, given by the Spanish Ministry of Science and Innovation/State Agency of Research (MCIN/AEI/10.13039/501100011033), and by "ERDF A way of making Europe". IH was supported by NSF grant MCB-2025305.

Acknowledgments

We thank our colleagues at the Centro de Astrobiología and the Center for Life in Extreme Environments for the numerous discussions on the topic covered here. We would also like to acknowledge the valuable comments of two reviewers who have undoubtedly contributed to improve the quality of this review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abergel, C., and Claverie, J. M. (2020). Giant viruses. *Curr. Biol.* 30, R1108–R1110. doi: 10.1016/j.cub.2020.08.055

Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., et al. (2018). Tailed giant *Tupanvirus* possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* 9:749. doi: 10.1038/S41467-018-03168-1

Abrescia, N. G. A., Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2012). Structure unifies the viral universe. *Annu. Rev. Biochem.* 81, 795–822. doi: 10.1146/annurev-biochem-060910-095130

Abreu, A., Bourgois, E., Gristwood, A., Troublé, R., Acinas, S. G., Bork, P., et al. (2022). Priorities for ocean microbiome research. *Nat. Microbiol.* 7, 937–947. doi: 10.1038/s41564-022-01145-5

Acevedo, A., Brodsky, L., and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505, 686–690. doi: 10.1038/NATURE12861

Agudo, R., De la Higuera, I., Arias, A., Grande-Pérez, A., and Domingo, E. (2016). Involvement of a joker mutation in a polymerase-independent lethal mutagenesis escape mechanism. *Virology* 494, 257–266. doi: 10.1016/j.virol.2016.04.023

Agudo, R., Ferrer-Orta, C., Arias, A., de la Higuera, I., Perales, C., Pérez-Luque, R., et al. (2010). A multi-step process of viral adaptation to a mutagenic nucleoside analogue by modulation of transition types leads to extinction-escape. *PLoS Pathog.* 6:e1001072. doi: 10.1371/journal.ppat.1001072

Altman, S. (1989). Ribonuclease P: An enzyme with a catalytic RNA subunit. *Adv. Enzymol. Relat. Areas Mol. Biol.* 62, 1–36. doi: 10.1002/9780470123089.CH1

An Astrobiology Strategy for the Search for Life in the Universe. (2019). *An astrobiology strategy for the search for life in the universe*. Washington, DC: National Academies Press. doi: 10.17226/25252

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M., and Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344, 757–760. doi: 10.1126/science.1252229

Ankrah, N. Y. D., May, A. L., Middleton, J. L., Jones, D. R., Hadden, M. K., Gooding, J. R., et al. (2014). Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J.* 8, 1089–1100. doi: 10.1038/ismej.2013.216

Arnaud, F., Varela, M., Spencer, T. E., and Palmarini, M. (2008). Coevolution of endogenous betaretroviruses of sheep and their host. *Cell. Mol. Life Sci.* 65, 3422–3432. doi: 10.1007/S00018-008-8500-9

Baaske, P., Weinert, F. M., Duhr, S., Lemke, K. H., Russell, M. J., and Braun, D. (2007). Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9346–9351. doi: 10.1073/pnas.0609592104

Bai, J., Yang, Z., Li, H., Hong, Y., Fan, D., Lin, A., et al. (2021). Genome-wide characterization of zebrafish endogenous retroviruses reveals unexpected diversity in genetic organizations and functional potentials. *Microbiol. Spectr.* 9:e0225421. doi: 10.1128/spectrum.02254-21

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241. doi: 10.1128/MMBR.35.3.235-241.1971

Bamford, D. H. (2003). Do viruses form lineages across different domains of life? *Res. Microbiol.* 154, 231–236. doi: 10.1016/S0923-2508(03)00065-2

Baquero, D. P., Liu, Y., Wang, F., Egelman, E. H., Prangishvili, D., and Krupovic, M. (2020). *Structure and assembly of archaeal viruses*, 1st Edn. Amsterdam: Elsevier Inc. doi: 10.1016/bs.aivir.2020.09.004

Barnard, C. J. (1984). Stasis: A coevolutionary model. *J. Theor. Biol.* 110, 27–34. doi: 10.1016/S0022-5193(84)80013-2

Bass, B. L., and Cech, T. R. (1984). Specific interaction between the self-splicing RNA of *Tetrahymena* and its guanosine substrate: Implications for biological catalysis by RNA. *Nature* 308, 820–826. doi: 10.1038/308820a0

Bataillon, T., and Bailey, S. F. (2014). Effects of new mutations on fitness: Insights from models and data. *Ann. N.Y. Acad. Sci.* 1320, 76–92. doi: 10.1111/nyas.12460

Batschelet, E., Domingo, E., and Weissmann, C. (1976). The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* 1, 27–32. doi: 10.1016/0378-1119(76)90004-4

Baylor, E. R., Baylor, M. B., Blanchard, D. C., Syzdek, L. D., and Appel, C. (1977). Virus transfer from surf to wind. *Science* 198, 575–580. doi: 10.1126/science.918656

Bell, G. (2008). Experimental evolution. *Heredity (Edinb).* 100, 441–442. doi: 10.1038/hdy.2008.19

Bell, P. J. L. (2020). Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res.* 289:198168. doi: 10.1016/J.VIRUSRES.2020.198168

Benner, S. A. (2010). Defining life. *Astrobiology* 10, 1021–1030. doi: 10.1089/AST.2010.0524

Bennett, J. O., Shostak, G. S., Schneider, N., and MacGregor, M. (2022). *Life in the universe*, 5th Edn. Princeton, NJ: Princeton University Press.

Benson, S. D., Bamford, J. K. H., Bamford, D. H., and Burnett, R. M. (2004). Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* 16, 673–685. doi: 10.1016/J.MOLCEL.2004.11.016

Bergh, O., Børsheim, K. Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature* 340, 467–468. doi: 10.1038/340467a0

Berliner, A. J., Mochizuki, T., and Stedman, K. M. (2018). Astrovirology: Viruses at large in the universe. *Astrobiology* 18, 207–223. doi: 10.1089/ast.2017.1649

Biebricher, C. K., and Eigen, M. (2005). The error threshold. *Virus Res.* 107, 117–127. doi: 10.1016/J.VIRUSRES.2004.11.002

Biebricher, C. K., and Eigen, M. (2006). "What is a quasispecies?," in *Quasispecies: Concept and implications for virology. Current topics in microbiology and immunology*, ed. E. Domingo (Berlin: Springer-Verlag).

Biebricher, C. K., and Gardiner, W. C. (1997). Molecular evolution of RNA in vitro. *Biophys. Chem.* 66, 179–192. doi: 10.1016/S0301-4622(97)00059-8

Boucher, C. A. B., Nijhuis, M., and Van Maarseveen, N. M. (2009). Antiviral resistance and impact on viral replication capacity: Evolution of viruses under antiviral pressure occurs in three phases. *Handb. Exp. Pharmacol.* 189, 299–320. doi: 10.1007/978-3-540-79086-0_11

Boyd, E. S., Hamilton, T. L., Wang, J., He, L., and Zhang, C. L. (2013). The role of tetraether lipid composition in the adaptation of thermophilic archaea to acidity. *Front. Microbiol.* 4:62. doi: 10.3389/fmicb.2013.00062

Brahim Belhaouari, D., Pires De Souza, G. A., Lamb, D. C., Kelly, S. L., Goldstone, J. V., Stegeman, J. J., et al. (2022). Metabolic arsenal of giant viruses: Host hijack or self-use? *Elife* 11:e78674. doi: 10.7554/eLife.78674

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 13, 278–284. doi: 10.1016/j.tim.2005.04.003

Briones, C., and Domingo, E. (2008). Minority report: Hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev.* 10, 93–109.

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., et al. (2015). Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. doi: 10.1126/science.1261498

Brum, J. R., Schenck, R. O., and Sullivan, M. B. (2013). Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J.* 7, 1738–1751. doi: 10.1038/ISMEJ.2013.67

Brussaard, C. P. D. (2004). Viral control of phytoplankton populations-a review. *J. Eukaryot. Microbiol.* 51, 125–138. doi: 10.1111/j.1550-7408.2004.tb00537.x

Buckheit, R. W. (2004). Understanding HIV resistance, fitness, replication capacity and compensation: Targeting viral fitness as a therapeutic strategy. *Expert Opin. Investig. Drugs* 13, 933–958. doi: 10.1517/13543784.13.8.933

Buckling, A., and Brockhurst, M. (2012). Bacteria-virus coevolution. *Adv. Exp. Med. Biol.* 751, 347–370. doi: 10.1007/978-1-4614-3567-9_16

Bull, J. J., Sanjuán, R., and Wilke, C. O. (2007). Theory of lethal mutagenesis for viruses. *J. Virol.* 81, 2930–2939. doi: 10.1128/JVI.01624-06

Cabanillas, L., Sanjuán, R., and Lázaro, E. (2014). Changes in protein domains outside the catalytic site of the bacteriophage Qβ replicase reduce the mutagenic effect of 5-azacytidine. *J. Virol.* 88, 10480–10487. doi: 10.1128/JVI.00979-14

Carrasco, P., de la Iglesia, F., and Elena, S. F. (2007). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco Etch virus. *J. Virol.* 81, 12979–12984. doi: 10.1128/JVI.00524-07

Caspar, D. L. D. (1956). Structure of bushy stunt virus. *Nature* 177, 475–476. doi: 10.1038/177475a0

Caspar, D. L. D., and Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27, 1–24. doi: 10.1101/SQB.1962.027.001.005

Castelán-Sánchez, H. G., Lopéz-Rosas, I., García-Suastegui, W. A., Peralta, R., Dobson, A. D. W., Batista-García, R. A., et al. (2019). Extremophile deep-sea viral

communities from hydrothermal vents: Structural and functional analysis. *Mar. Genomics* 46, 16–28. doi: 10.1016/j.margen.2019.03.001

Cervera, H., Lalić, J., and Elena, S. F. (2016). Effect of host species on topography of the fitness landscape for a plant RNA virus. *J. Virol.* 90, 10160–10169. doi: 10.1128/JVI.01243-16

Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A. F., Erb, M. L., Coker, J. K. C., et al. (2017). Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355, 194–197. doi: 10.1126/SCIENCE.AAL2130

Chan, M. A., Hinman, N. W., Potter-Mcintyre, S. L., Schubert, K. E., Gillams, R. J., Awramik, S. M., et al. (2019). Deciphering biosignatures in planetary contexts. *Astrobiology* 19, 1075–1102. doi: 10.1089/ast.2018.1903

Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., et al. (2011). Femtosecond X-ray protein nanocrystallography. *Nature* 470, 73–78. doi: 10.1038/nature09750

Chapman, K. B., and Szostak, J. W. (1994). In vitro selection of catalytic RNAs. *Curr. Opin. Struct. Biol.* 4, 618–622. doi: 10.1016/S0959-440X(94)90227-5

Cleland, C. E. (2011). Life without definitions. *Synthese* 1851, 125–144. doi: 10.1007/S11229-011-9879-7

Cleland, C. E. (2019). Moving beyond definitions in the search for extraterrestrial life. *Astrobiology* 19, 722–729. doi: 10.1089/ast.2018.1980

Cléry, A., Blatter, M., and Allain, F. H. T. (2008). RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* 18, 290–298. doi: 10.1016/J.SBI.2008.04.002

Cockell, C. (2020). *Astrobiology?: Understanding life in the universe*, 2nd Edn. New York, NY: Wiley-Blackwell.

Coffey, L. L., Beeharry, Y., Bordería, A. V., Blanc, H., and Vignuzzi, M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16038–16043. doi: 10.1073/PNAS.1111650108

Cole, C. N. (1975). Defective interfering (DI) particles of poliovirus. *Progr.Med.Virol. Vol.* 20, 180–207.

Costello, D. A., Whittaker, G. R., and Daniel, S. (2015). Variations in pH sensitivity, acid stability, and fusogenicity of three influenza virus H3 subtypes. *J. Virol.* 89, 350–360. doi: 10.1128/JVI.01927-14

Cuevas, J. M., Domingo-Calap, P., and Sanjuán, R. (2012). The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol. Biol. Evol.* 29, 17–20. doi: 10.1093/molbev/msr179

Daly, R. A., Roux, S., Borton, M. A., Morgan, D. M., Johnston, M. D., Booker, A. E., et al. (2019). Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat. Microbiol.* 4, 352–361. doi: 10.1038/S41564-018-0312-6

de la Higuera, I., Kasun, G. W., Torrance, E. L., Pratt, A. A., Maluenda, A., Colombet, J., et al. (2020). Unveiling crucivirus diversity by mining metagenomic data. *mBio* 11, e1410–e1420. doi: 10.1128/mBio.01410-20

de Pablo, P. J., and San Martín, C. (2022). Seeing and touching adenovirus: Complementary approaches for understanding assembly and disassembly of a complex virion. *Curr. Opin. Virol.* 52, 112–122. doi: 10.1016/j.coviro.2021.11.006

De Paepe, M., and Taddei, F. (2006). Viruses' life history: Towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol.* 4:e193. doi: 10.1371/JOURNAL.PBIO.0040193

Des Marais, D. J., Nuth, J. A., Allamandola, L. J., Boss, A. P., Farmer, J. D., Hoehler, T. M., et al. (2008). The NASA astrobiology roadmap. *Astrobiology* 8, 715–730. doi: 10.1089/ast.2008.0819

Dessau, M., Goldhill, D., McBride, R. L., Turner, P. E., and Modis, Y. (2012). Selective pressure causes an RNA virus to trade reproductive fitness for increased structural and thermal stability of a viral enzyme. *PLoS Genet.* 8:e1003102. doi: 10.1371/JOURNAL.PGEN.1003102

DiMaio, F., Yu, X., Rensen, E., Krupovic, M., Prangishvili, D., and Egelman, E. H. (2015). A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* 348, 914–917. doi: 10.1126/science.aaa4181

Dix, D. E. (2003). What is life? Prerequisites for a definition. *Yale J. Biol. Med.* 75, 313–321.

Domingo, E., García-Crespo, C., Lobo-Vega, R., and Perales, C. (2021). Mutation rates, mutation frequencies, and proofreading-repair activities in rna virus genetics. *Viruses* 13:1882. doi: 10.3390/v13091882

Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13, 735–744. doi: 10.1016/0092-8674(78)90223-4

Domingo-Calap, P., and Sanjuán, R. (2011). Experimental evolution of rna versus dna viruses. *Evolution* 65, 2987–2994. doi: 10.1111/j.1558-5646.2011.01339.x

Domingo-Calap, P., Cuevas, J. M., and Sanjuán, R. (2009). The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* 5:e1000742. doi: 10.1371/journal.pgen.1000742

Dominguez-Huerta, G., Zayed, A. A., Wainaina, J. M., Guo, J., Tian, F., Pratama, A. A., et al. (2022). Diversity and ecological footprint of global ocean RNA viruses. *Science* 376, 1202–1208. doi: 10.1126/science.abn6358

Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008). Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* 9, 267–276. doi: 10.1038/nrg2323

Dupressoir, A., Vernochet, C., Harper, F., Guégan, J., Dessen, P., Pierron, G., et al. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1164–E1173. doi: 10.1073/PNAS.1112304108

Durzyńska, J., and Goździcka-Józefiak, A. (2015). Viruses and cells intertwined since the dawn of evolution. *Virol. J.* 12:169. doi: 10.1186/s12985-015-0400-7

Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., et al. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature* 602, 142–147. doi: 10.1038/s41586-021-04332-2

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523.

Eigen, M., and Schuster, P. (1979). *The hypercycle a principle of natural self-organization*. Berlín: Springer.

Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R. N. C., Seltzer, V., Claverie, J. M., et al. (2015). Three-dimensional reconstruction of the giant mimivirus particle with an X-ray free-electron laser. *Phys. Rev. Lett.* 114:098102. doi: 10.1103/PHYSREVLETT.114.098102

Elena, S. F. (2001). Evolutionary history conditions the timing of transmission in vesicular stomatitis virus. *Infect. Genet. Evol.* 1, 151–159. doi: 10.1016/S1567-1348(01)00022-3

Elena, S. F. (2016). Evolutionary transitions during RNA virus experimental evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 371:20150441. doi: 10.1098/RSTB.2015.0441

Elena, S. F., Agudelo-Romero, P., Carrasco, P., Codoñer, F. M., Martín, S., Torres-Barceló, C., et al. (2008). Experimental evolution of plant RNA viruses. *Heredity (Edinb)* 100, 478–483. doi: 10.1038/sj.hdy.6801088

Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. B., et al. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* 3, 870–880. doi: 10.1038/s41564-018-0190-y

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* 1310, 641–650. doi: 10.1038/nrmicro3527

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi: 10.1038/NRG2337

Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: Insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296. doi: 10.1038/NRG3199

Filée, J., and Forterre, P. (2005). Viral proteins functioning in organelles: A cryptic origin? *Trends Microbiol.* 13, 510–513. doi: 10.1016/J.TIM.2005.08.012

Filée, J., Forterre, P., and Laurent, J. (2003). The role played by viruses in the evolution of their hosts: A view based on informational protein phylogenies. *Res. Microbiol.* 154, 237–243. doi: 10.1016/S0923-2508(03)00066-4

Fischer, M. G. (2011). Sputnik and mavirus: More than just satellite viruses. *Nat. Rev. Microbiol.* 10:88. doi: 10.1038/NRMICRO2676-C1

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus and Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829. doi: 10.1073/pnas.1307701110

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 117, 5–16. doi: 10.1016/J.VIRUSRES.2006.01.010

Forterre, P. (2011). Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Comptes Rendus Chim.* 14, 392–399. doi: 10.1016/J.CRCI.2010.06.007

Forterre, P. (2016). To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. *Stud. Hist. Philos. Biol. Biomed. Sci.* 59, 100–108. doi: 10.1016/j.shpsc.2016.02.013

Forterre, P., and Prangishvili, D. (2009). The origin of viruses. *Res. Microbiol.* 160, 466–472. doi: 10.1016/J.RESMIC.2009.07.008

French, R. K., and Holmes, E. C. (2020). An ecosystems perspective on virus evolution and emergence. *Trends Microbiol.* 28, 165–175. doi: 10.1016/J.TIM.2019.10.010

Fridman, S., Flores-Uribe, J., Larom, S., Alalouf, O., Liran, O., Yacoby, I., et al. (2017). A myovirus encoding both photosystem i and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat. Microbiol.* 2, 1350–1357. doi: 10.1038/s41564-017-0002-9

Fridmann-Sirkis, Y., Milrot, E., Mutsafi, Y., Ben-Dor, S., Levin, Y., Savidor, A., et al. (2016). Efficiency in complexity: Composition and dynamic nature of mimivirus replication factories. *J. Virol.* 90, 10039–10047. doi: 10.1128/JVI.01319-16

Gabashvili, E., Kobakhidze, S., Chkhikvishvili, T., Tabatadze, L., Tsiklauri, R., Dadiani, K., et al. (2022). Bacteriophage-mediated risk pathways underlying the emergence of antimicrobial resistance via intrageneric and intergeneric recombination of antibiotic efflux genes across natural populations of human pathogenic bacteria. *Microb. Ecol.* 84, 213–226. doi: 10.1007/s00248-021-01846-0

Gabashvili, E., Osepashvili, M., Koulouris, S., Ujmajuridze, L., Tskhitishvili, Z., and Kotetishvili, M. (2020). Phage transduction is involved in the intergeneric spread of antibiotic resistance-associated bla CTX-M, mel, and tetM Loci in natural populations of some human and animal bacterial pathogens. *Curr. Microbiol.* 77, 185–193. doi: 10.1007/S00284-019-01817-2

Gago, S., Elena, S. F., Flores, R., and Sanjuán, R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* 323:1308. doi: 10.1126/SCIENCE.1169202

Gallego, I., Soria, M. E., García-Crespo, C., Chen, Q., Martínez-Barragán, P., Khalfaoui, S., et al. (2020). Broad and dynamic diversification of infectious hepatitis C virus in a cell culture environment. *J. Virol.* 94, e01856–19. doi: 10.1128/JVI.01856-19

Gao, L., Altae-Tran, H., Böhning, F., Makarova, K. S., Segel, M., Schmid-Burgk, J. L., et al. (2020). Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* 369, 1077–1084. doi: 10.1126/SCIENCE.ABA0372

García-Descalzo, L., Parro, V., García-Villadangos, M., Cockell, C. S., Moissl-Eichinger, C., Perras, A., et al. (2019). Microbial markers profile in anaerobic mars analogue environments using the ldchip (Life detector chip) antibody microarray core of the solid (signs of life detector) platform. *Microorganisms* 7:365. doi: 10.3390/microorganisms7090365

Garmann, R. F., Goldfain, A. M., and Manoharan, V. N. (2019). Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proc. Natl. Acad. Sci.* 116, 22485–22490. doi: 10.1073/pnas.1909223116

Geoghegan, J. L., and Holmes, E. C. (2018). Evolutionary virology at 40. *Genetics* 210, 1151–1162. doi: 10.1534/GENETICS.118.301556

Gil, J. F., Mesa, V., Estrada-Ortiz, N., Lopez-Obando, M., Gómez, A., and Plácido, J. (2021). Viruses in extreme environments, current overview, and biotechnological potential. *Viruses* 13, 1–23. doi: 10.3390/v13010081

Gilbert, C., and Cordaux, R. (2017). Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Curr. Opin. Virol.* 25, 16–22. doi: 10.1016/J.COVIRO.2017.06.005

Gilbert, W. (1986). Origin of life: The RNA world. *Nature* 319:618. doi: 10.1038/319618a0

Giovannoni, S., Temperton, B., and Zhao, Y. (2013). Giovannoni et al. reply to "SAR11 viruses and defensive host strains.". *Nature* 499, E4–E5. doi: 10.1038/nature12388

Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M., and Sañudo-Wilhelmy, S. A. (1997). Release and bioavailability of C. N, P, Se, and Fe following viral lysis of a marine chrysophyte. *Limnol. Oceanogr.* 42, 1492–1504. doi: 10.4319/lo.1997.42.7.1492

Grande-Pérez, A., Lázaro, E., Lowenstein, P., Domingo, E., and Manrubia, S. C. (2005). Suppression of viral infectivity through lethal defection. *Proc. Natl. Acad. Sci. U.S.A.* 102:4448. doi: 10.1073/PNAS.0408871102

Greenwood, A. D., Ishida, Y., O'Brien, S. P., Roca, A. L., and Eiden, M. V. (2018). Transmission, Evolution, and endogenization: Lessons learned from recent retroviral invasions. *Microbiol. Mol. Biol. Rev.* 82, e00044–17. doi: 10.1128/mmbr.00044-17

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., et al. (2019). Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177, 1109.e–1123.e. doi: 10.1016/j.cell.2019.03.040

Griffin, D. W. (2013). The quest for extraterrestrial life: What about the viruses? *Astrobiology* 13, 774–783. doi: 10.1089/AST.2012.0959

Griffin, D. W., Garrison, V. H., Herman, J. R., and Shinn, E. A. (2001). African desert dust in the Caribbean atmosphere: Microbiology and public health. *Aerobiologia (Bologna)* 17, 203–213. doi: 10.1023/A:1011868218901

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Europe PMC funders group plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532, 465–470. doi: 10.1038/nature16942.Plankton

Hager, A. J., Pollard, J. D., and Szostak, J. W. (1996). Ribozymes: Aiming at RNA replication and protein synthesis. *Chem. Biol.* 3, 717–725. doi: 10.1016/S1074-5521(96)90246-X

Hampton, H. G., Watson, B. N. J., and Fineran, P. C. (2020). The arms race between bacteria and their phage foes. *Nature* 577, 327–336. doi: 10.1038/S41586-019-1894-8

Han, Z., and Yuan, W. (2022). Structural insights into a spindle-shaped archaeal virus with a sevenfold symmetrical tail. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2119439119. doi: 10.1073/pnas.2119439119

Hanson, A., Imai, M., Hatta, M., McBride, R., Imai, H., Taft, A., et al. (2015). Identification of stabilizing mutations in an H5 hemagglutinin influenza virus protein. *J. Virol.* 90, 2981–2992. doi: 10.1128/JVI.02790-15

Harris, H. M. B., and Hill, C. (2020). A place for viruses on the tree of life. *Front. Microbiol.* 11:604048. doi: 10.3389/FMICB.2020.604048

Hays, L., Achenbach, L., Bailey, J., Barnes, R., Baross, J., Bertka, C., et al. (2015). *NASA astrobiology strategy 2015*. Washington, DC: National Aeronautics and Space Administration (NASA).

Hayward, A., Cornwallis, C. K., and Jern, P. (2015). Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl. Acad. Sci. U.S.A.* 112, 464–469. doi: 10.1073/PNAS.1414980112

Hegde, N. R., Maddur, M. S., Kaveri, S. V., and Bayry, J. (2009). Reasons to include viruses in the tree of life. *Nat. Rev. Microbiol.* 7:615. doi: 10.1038/NRMICRO2108-C1

Hegedüs, M., Kovács, G., Módos, K., Rontó, G., Lammer, H., Panitz, C., et al. (2006). Exposure of phage T7 to simulated space environment: The effect of vacuum and UV-C radiation. *J. Photochem. Photobiol. B.* 82, 94–104. doi: 10.1016/J.JPHOTOBIOL.2005.09.002

Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl. Acad. Sci.* 96, 2192–2197. doi: 10.1073/pnas.96.5.2192

Higgs, P. G., and Lehman, N. (2015). The RNA world: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* 16, 7–17. doi: 10.1038/NRG3841

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. *Science* 215, 1577–1585. doi: 10.1126/SCIENCE.7041255

Holmes, E. C. (2004). Adaptation and immunity. *PLoS Biol.* 2:E307. doi: 10.1371/JOURNAL.PBIO.0020307

Holmes, E. C. (2011b). What does virus evolution tell us about virus origins? *J. Virol.* 85, 5247–5251. doi: 10.1128/JVI.02203-10

Holmes, E. C. (2011a). The evolution of endogenous viral elements. *Cell Host Microbe* 10, 368–377. doi: 10.1016/J.CHOM.2011.09.002

Holmes, E. C., and Duchêne, S. (2019). Evolutionary stasis of viruses? *Nat. Rev. Microbiol.* 17:329. doi: 10.1038/S41579-019-0168-7

Horneck, G., Klaus, D. M., and Mancinelli, R. L. (2010). Space microbiology. *Microbiol. Mol. Biol. Rev.* 74, 121–156. doi: 10.1128/MMBR.00016-09

Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H. B., et al. (2020). Phage-specific metabolic reprogramming of virocells. *ISME J.* 14, 881–895. doi: 10.1038/s41396-019-0580-z

Huang, A. S. (1973). Defective interfering viruses. *Annu. Rev. Microbiol.* 27, 101–117. doi: 10.1146/ANNUREV.MI.27.100173.000533

Hughes, J. F., and Coffin, J. M. (2001). Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat. Genet.* 29, 487–489. doi: 10.1038/NG775

Hurwitz, B. L., Hallam, S. J., and Sullivan, M. B. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* 14:R123. doi: 10.1186/gb-2013-14-11-r123

Iranzo, J., Puigbo, P., Lobkovsky, A. E., Wolf, Y. I., and Koonin, E. V. (2016). Inevitability of genetic parasites. *Genome Biol. Evol.* 8, 2856–2869. doi: 10.1093/gbe/evw193

Irwin, N. A. T., Pittis, A. A., Richards, T. A., and Keeling, P. J. (2022). Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat. Microbiol.* 7, 327–336. doi: 10.1038/s41564-021-01026-3

Iyer, L. M., Koonin, E. V., Leipe, D. D., and Aravind, L. (2005). Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: Structural insights and new members. *Nucleic Acids Res.* 33, 3875–3896. doi: 10.1093/NAR/GKI702

Janzen, E., Janzen, E., Blanco, C., Peng, H., Kenchel, J., Kenchel, J., et al. (2020). Promiscuous ribozymes and their proposed role in prebiotic evolution. *Chem. Rev.* 120, 4879–4897. doi: 10.1021/acs.chemrev.9b00620

Jiang, W., and Tang, L. (2017). Atomic Cryo-EM structures of viruses. *Curr. Opin. Struct. Biol.* 46:122. doi: 10.1016/J.SBI.2017.07.002

Joyce, G. F., and Szostak, J. W. (2018). Protocells and RNA Self-Replication. *Cold Spring Harb. Perspect. Biol.* 10:a034801. doi: 10.1101/CSHPERSPECT.A034801

Kacian, D. L., Mills, D. R., Kramer, F. R., and Spiegelman, S. (1972). A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. *Proc. Natl. Acad. Sci. U.S.A.* 69, 3038–3042. doi: 10.1073/pnas.69.10.3038

Kaján, G. L., Doszpoly, A., Tarján, Z. L., Vidovszky, M. Z., and Papp, T. (2020). Virus-host coevolution with a focus on animal and human DNA viruses. *J. Mol. Evol.* 88, 41–56. doi: 10.1007/S00239-019-09913-4

Kaneko, H., Blanc-Mathieu, R., Endo, H., Chaffron, S., Delmont, T. O., Gaia, M., et al. (2021). Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience* 24:102002. doi: 10.1016/j.isci.2020.102002

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M. C. (2012). Experimental evolution. *Trends Ecol. Evol.* 27, 547–560. doi: 10.1016/J.TREE.2012.06.001

Kazlauskas, D., Krupovic, M., and Venclovas, C. (2016). The logic of DNA replication in double-stranded DNA viruses: Insights from global analysis of viral genomes. *Nucleic Acids Res.* 44, 4551–4564. doi: 10.1093/NAR/GKW322

Kazlauskas, D., Varsani, A., Koonin, E. V., and Krupovic, M. (2019). Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-11433-0

Koelle, K., Cobey, S., Grenfell, B., and Pascual, M. (2006). Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* 314, 1898–1903. doi: 10.1126/SCIENCE.1132745

Koike, J., Oshima, T., Koike, K. A., Taguchi, H., Tanaka, R., Nishimura, K., et al. (1992). Survival rates of some terrestrial microorganisms under simulated space conditions. *Adv. Sp. Res.* 12, 271–274. doi: 10.1016/0273-1177(92)90182-W

Koonin, E. V. (2014). The origins of cellular life. *Antonie Van Leeuwenhoek* 106, 27–41. doi: 10.1007/S10482-014-0169-5

Koonin, E. V. (2015). Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: Eukaryogenesis just made easier? *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140333. doi: 10.1098/rstb.2014.0333

Koonin, E. V. (2016). Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 371:20150442. doi: 10.1098/RSTB.2015.0442

Koonin, E. V., and Krupovic, M. (2020). Phages build anti-defence barriers. *Nat. Microbiol.* 5, 8–9. doi: 10.1038/S41564-019-0635-Y

Koonin, E. V., and Yutin, N. (2018). Multiple evolutionary origins of giant viruses. *F1000Res* 7. doi: 10.12688/F1000RESEARCH.16248.1

Koonin, E. V., Dolja, V. V., and Krupovic, M. (2022). The logic of virus evolution. *Cell Host Microbe* 30, 917–929. doi: 10.1016/j.chom.2022.06.008

Koonin, E. V., Dolja, V. V., Krupovic, M., and Kuhn, J. H. (2021). Viruses defined by the position of the virosphere within the replicator space. *Microbiol. Mol. Biol. Rev.* 85:e0019320. doi: 10.1128/MMBR.00193-20

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 84, 1–33. doi: 10.1128/MMBR.00061-19

Koonin, E. V., Senkevich, T. G., and Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biol. Direct* 1:29. doi: 10.1186/1745-6150-1-29

Koonin, E. V., Wolf, Y. I., and Katsnelson, M. I. (2017). Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. *Biol. Direct* 12:31. doi: 10.1186/S13062-017-0202-5

Koskella, B., and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* 38, 916–931. doi: 10.1111/1574-6976.12072

Krakauer, D. C., and Sasaki, A. (2002). Noisy clues to the origin of life. *Proceedings. Biol. Sci.* 269, 2423–2428. doi: 10.1098/RSPB.2002.2127

Krishnamurthy, S. R., and Wang, D. (2017). Origins and challenges of viral dark matter. *Virus Res.* 239, 136–142. doi: 10.1016/j.virusres.2017.02.002

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 31, 147–157. doi: 10.1016/0092-8674(82)90414-7

Krupovic, M., and Bamford, D. H. (2008). Virus evolution: How far does the double β-barrel viral lineage extend? *Nat. Rev. Microbiol.* 6, 941–948. doi: 10.1038/nrmicro2033

Krupovic, M., and Koonin, E. V. (2017). Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2401–E2410. doi: 10.1073/PNAS.1621061114

Krupovic, M., Dolja, V. V., and Koonin, E. V. (2019). Origin of viruses: Primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* 17, 449–458. doi: 10.1038/S41579-019-0205-6

Krupovic, M., Quemin, E. R. J., Bamford, D. H., Forterre, P., and Prangishvili, D. (2014). Unification of the globally distributed spindle-shaped viruses of the archaea. *J. Virol.* 88, 2354–2358. doi: 10.1128/jvi.02941-13

La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/SCIENCE.1081867

Laber, C. P., Hunter, J. E., Carvalho, F., Collins, J. R., Hunter, E. J., Schieler, B. M., et al. (2018). Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* 3, 537–547. doi: 10.1038/s41564-018-0128-4

Laidler, J. R., and Stedman, K. M. (2010). Virus silicification under simulated hot spring conditions. *Astrobiology* 10, 569–576. doi: 10.1089/ast.2010.0463

Laidler, J. R., Shugart, J. A., Cady, S. L., Bahjat, K. S., and Stedman, K. M. (2013). Reversible inactivation and desiccation tolerance of silicified viruses. *J. Virol.* 87, 13927–13929. doi: 10.1128/jvi.02825-13

Lázaro, E., Arribas, M., Cabanillas, L., Román, I., and Acosta, E. (2018). Evolutionary adaptation of an RNA bacteriophage to the simultaneous increase in the within-host and extracellular temperatures. *Sci. Rep.* 8:8080. doi: 10.1038/S41598-018-26443-Z

Lee, K. Y., and Lee, B. J. (2017). Structural and biochemical properties of novel self-cleaving ribozymes. *Molecules* 22:678. doi: 10.3390/molecules22040678

Levisohn, R., and Spiegelman, S. (1969). Further extracellular Darwinian experiments with replicating RNA molecules: Diverse variants isolated under different selective conditions. *Proc. Natl. Acad. Sci. U.S.A.* 63, 805–811. doi: 10.1073/PNAS.63.3.805

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M., and Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438, 86–89. doi: 10.1038/nature04111

Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11013–11018. doi: 10.1073/pnas.0401526101

Lingam, M., and Loeb, A. (2021). *Life in the cosmos?: From biosignatures to technosignatures*. Cambridge, MA: Harvard University Press.

Maggiori, C., Stromberg, J., Blanco, Y., Goordial, J., Cloutis, E., García-Villadangos, M., et al. (2020). The limits, capabilities, and potential for life detection with MinION sequencing in a paleochannel mars analog. *Astrobiology* 20, 375–393. doi: 10.1089/ast.2018.1964

Mann, N. H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003). Bacterial photosynthesis genes in a virus. *Nature* 424, 741–741. doi: 10.1038/424741a

Marchesi, A., Umeda, K., Komekawa, T., Matsubara, T., Flechsig, H., Ando, T., et al. (2021). An ultra-wide scanner for large-area high-speed atomic force microscopy with megapixel resolution. *Sci. Rep.* 11:13003. doi: 10.1038/s41598-021-92365-y

Márquez, L. M., Redman, R. S., Rodriguez, R. J., and Roossinck, M. J. (2007). A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science* 315, 513–515. doi: 10.1126/science.1136237

Matsuura, K., Watanabe, K., Matsuzaki, T., Sakurai, K., and Kimizuka, N. (2010). Self-assembled synthetic viral capsids from a 24-mer viral peptide fragment. *Angew. Chem. Int. Ed. Engl.* 49, 9662–9665. doi: 10.1002/anie.201004606

McDonald, M. J. (2019). Microbial experimental evolution–a proving ground for evolutionary theory and a tool for discovery. *EMBO Rep.* 20:e46992. doi: 10.15252/EMBR.201846992

McGee, L. W., Aitchison, E. W., Caudle, S. B., Morrison, A. J., Zheng, L., Yang, W., et al. (2014). Payoffs, not tradeoffs, in the adaptation of a virus to ostensibly conflicting selective pressures. *PLoS Genet.* 10:e1004611. doi: 10.1371/JOURNAL.PGEN.1004611

Meents, A., and Wiedorn, M. O. (2019). Virus structures by X-Ray free-electron lasers. *Annu. Rev. Virol.* 6, 161–176. doi: 10.1146/ANNUREV-VIROLOGY-092818-015724

Mills, D. R., Peterson, R. L., and Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U.S.A.* 58, 217–224. doi: 10.1073/PNAS.58.1.217

Minicka, J., Elena, S. F., Borodynko-Filas, N., Rubiś, B., and Hasiów-Jaroszewska, B. (2017). Strain-dependent mutational effects for Pepino mosaic virus in a natural host. *BMC Evol. Biol.* 17:67. doi: 10.1186/s12862-017-0920-4

Mizuuchi, R., Furubayashi, T., and Ichihashi, N. (2022). Evolutionary transition from a single RNA replicator to a multiple replicator network. *Nat. Commun.* 13:1460. doi: 10.1038/s41467-022-29113-x

Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., Sako, Y., Forterre, P., and Prangishvili, D. (2012). Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13386–13391. doi: 10.1073/pnas.1203668109

Moreira, D., and López-García, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* 7, 306–311. doi: 10.1038/nrmicro2108

Morowitz, H., and Smith, E. (2007). Energy flow and the organization of life. *Complexity* 13, 51–59. doi: 10.1002/cplx.20191

Mougari, S., Sahmi-Bounsiar, D., Levasseur, A., Colson, P., Scola, B., and La. (2019). Virophages of Giant viruses: An update at eleven. *Viruses* 11:733. doi: 10.3390/V11080733

Muniesa, M., Imamovic, L., and Jofre, J. (2011). Bacteriophages and genetic mobilization in sewage and faecally polluted environments. *Microb. Biotechnol.* 4, 725–734. doi: 10.1111/j.1751-7915.2011.00264.x

Munson-Mcgee, J. H., Peng, S., Dewerff, S., Stepanauskas, R., Whitaker, R. J., Weitz, J. S., et al. (2018). A virus or more in (nearly) every cell: Ubiquitous networks of virus-host interactions in extreme environments. *ISME J.* 12, 1706–1714. doi: 10.1038/s41396-018-0071-7

Munson-McGee, J. H., Rooney, C., and Young, M. J. (2020). An uncultivated virus infecting a nanoarchaeal parasite in the hot springs of Yellowstone National Park. *J. Virol.* 94, e1213–e1219. doi: 10.1128/jvi.01213-19

Mushegian, A. R. (2020). Are there 10 31 virus particles on earth, or more, or fewer? *J. Bacteriol.* 202, e52–e20. doi: 10.1128/JB.00052-20

Navas-Castillo, J. (2009). Six comments on the ten reasons for the demotion of viruses. *Nat. Rev. Microbiol.* 7:615. doi: 10.1038/NRMICRO2108-C2

Neri, U., Wolf, Y. I., Roux, S., Camargo, A. P., Lee, B., Kazlauskas, D., et al. (2022). Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell* 185, 4023–4037.e18. doi: 10.1016/j.cell.2022.08.023

Nielsen, P. E. (2007). Peptide nucleic acids and the origin of life. *Chem. Biodivers.* 4, 1996–2002. doi: 10.1002/CBDV.200790166

Nishida, R., Nakamura, K., Taniguchi, I., Murase, K., Ooka, T., Ogura, Y., et al. (2021). The global population structure and evolutionary history of the acquisition of major virulence factor-encoding genetic elements in Shiga toxin-producing *Escherichia coli* O121:H19. *Microb. Genomics* 7:000716. doi: 10.1099/MGEN.0.000716

O'rourke, A., Zoumplis, A., Wilburn, P., Lee, M. D., Lee, Z., Vecina, M., et al. (2020). Following the astrobiology roadmap: Origins, habitability and future exploration. *Curr. Issues Mol. Biol.* 38, 1–32. doi: 10.21775/cimb.038.001

Orange, F., Chabin, A., Gorlas, A., Lucas-Staat, S., Geslin, C., Le Romancer, M., et al. (2011). Experimental fossilisation of viruses from extremophilic Archaea. *Biogeosciences* 8, 1465–1475. doi: 10.5194/bg-8-1465-2011

Oro, J., Miller, S. L., and Lazcano, A. (1990). The origin and early evolution of life on earth. *Annu. Rev. Earth Planet. Sci.* 18, 317–356. doi: 10.1146/ANNUREV.EA.18.050190.001533

Pandey, S., Bean, R., Sato, T., Poudyal, I., Bielecki, J., Cruz Villarreal, J., et al. (2020). Time-resolved serial femtosecond crystallography at the European XFEL. *Nat. Methods* 17, 73–78. doi: 10.1038/s41592-019-0628-z

Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R. P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* 23, 171–178. doi: 10.1016/j.mib.2014.11.019

Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., and González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infect. Genet. Evol.* 30, 296–307. doi: 10.1016/j.meegid.2014.12.022

Pfeiffer, J. K., and Kirkegaard, K. (2003). A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7289–7294. doi: 10.1073/PNAS.1232294100

Pietilä, M. K., Atanasova, N. S., Oksanen, H. M., and Bamford, D. H. (2013). Modified coat protein forms the flexible spindle-shaped virion of haloarchaeal virus His1. *Environ. Microbiol.* 15, 1674–1686. doi: 10.1111/1462-2920.12030

Plunkett, G., Rose, D. J., Durfee, T. J., and Blattner, F. R. (1999). Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* 181, 1767–1778. doi: 10.1128/JB.181.6.1767-1778.1999

Poorvin, L., Rinta-Kanto, J. M., Hutchins, D. A., and Wilhelm, S. W. (2004). Viral release of iron and its bioavailability to marine plankton. *Limnol. Oceanogr.* 49, 1734–1741. doi: 10.4319/lo.2004.49.5.1734

Quakkelaar, E. D., Beaumont, T., van Nuenen, A. C., van Alphen, F. P. J., Boeser-Nunnink, B. D. M., van 't Wout, A. B., et al. (2007). T cell line passage can select for pre-existing neutralization-sensitive variants from the quasispecies of primary human immunodeficiency virus type-1 isolates. *Virology* 359, 92–104. doi: 10.1016/J.VIROL.2006.09.021

Quemin, E. R. J., Pietilä, M. K., Oksanen, H. M., Forterre, P., Rijpstra, W. I. C., Schouten, S., et al. (2015). Sulfolobus spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J. Virol.* 89, 11681–11691. doi: 10.1128/jvi.02270-15

Rawn, S. M., and Cross, J. C. (2008). The evolution, regulation, and function of placenta-specific genes. *Annu. Rev. Cell Dev. Biol.* 24, 159–181. doi: 10.1146/ANNUREV.CELLBIO.24.110707.175418

Reche, I., D'Orta, G., Mladenov, N., Winget, D. M., and Suttle, C. A. (2018). Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J.* 12, 1154–1162. doi: 10.1038/S41396-017-0042-4

Rezzonico, F. (2014). Nanopore-based instruments as biosensors for future planetary missions. *Astrobiology* 14, 344–351. doi: 10.1089/ast.2013.1120

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., et al. (2004). The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7716–7720. doi: 10.1073/PNAS.0401773101

Rodrigues, R. A. L., da Silva, L. C. F., and Abrahão, J. S. (2020). Translating the language of giants: translation-related genes as a major contribution of giant viruses to the virosphere. *Arch. Virol.* 165, 1267–1278. doi: 10.1007/s00705-020-04626-2

Rosenzweig, J. B., Majernik, N., Robles, R. R., Andonian, G., Camacho, O., Fukasawa, A., et al. (2020). An ultra-compact x-ray free-electron laser. *N. J. Phys.* 22:093067. doi: 10.1088/1367-2630/abb16c

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537, 689–693. doi: 10.1038/NATURE19366

Rowe, H. M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. *Virology* 411, 273–287. doi: 10.1016/J.VIROL.2010.12.007

Ruíz-Jarabo, C. M., Arias, A., Molina-París, C., Briones, C., Baranowski, E., Escarmís, C., et al. (2002). Duration and fitness dependence of quasispecies memory. *J. Mol. Biol.* 315, 285–296. doi: 10.1006/jmbi.2001.5232

Russell, M. J., and Hall, A. J. (1997). The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* 154, 377–402. doi: 10.1144/GSJGS.154.3.0377

Sanjuán, R., Moya, A., and Elena, S. F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8396–8401. doi: 10.1073/PNAS.0400146101

Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., and Belshaw, R. (2010). Viral mutation rates. *J. Virol.* 84, 9733–9748. doi: 10.1128/JVI.00694-10

Sankaran, N. (2016). The RNA world at thirty: A look back with its author. *J. Mol. Evol.* 83, 169–175. doi: 10.1007/s00239-016-9767-3

Santos, F., Yarza, P., Parro, V., Meseguer, I., Rosselló-Móra, R., and Antón, J. (2012). Culture-independent approaches for studying viruses from hypersaline environments. *Appl. Environ. Microbiol.* 78, 1635–1643. doi: 10.1128/AEM.07175-11

Santos-Pérez, I., Charro, D., Gil-Carton, D., Azkargorta, M., Elortza, F., Bamford, D. H., et al. (2019). Structural basis for assembly of vertical single β-barrel viruses. *Nat. Commun.* 10:1184. doi: 10.1038/S41467-019-08927-2

Säwström, C., Lisle, J., Anesio, A. M., Priscu, J. C., and Laybourn-Parry, J. (2008). Bacteriophage in polar inland waters. *Extremophiles* 12, 167–175. doi: 10.1007/s00792-007-0134-6

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356, 82–85. doi: 10.1126/SCIENCE.AAL4657

Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., et al. (2011). Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* 470, 78–82. doi: 10.1038/nature09748

Sharma, G., and Curtis, P. D. (2022). The impacts of microgravity on bacterial metabolism. *Life (Basel)* 12:774. doi: 10.3390/LIFE12060774

Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D. B., et al. (2007). Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J.* 1, 492–501. doi: 10.1038/ISMEJ.2007.67

Shelford, E. J., Middelboe, M., Møller, E. F., and Suttle, C. A. (2012). Virus-driven nitrogen cycling enhances phytoplankton growth. *Aquat. Microb. Ecol.* 66, 41–46. doi: 10.3354/ame01553

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543. doi: 10.1038/nature20167

Shoemaker, S. C., and Ando, N. (2018). X-rays in the Cryo-EM era: Structural biology's dynamic future. *Biochemistry* 57:277. doi: 10.1021/ACS.BIOCHEM.7B01031

Smith, P. H., Tamppari, L., Arvidson, R. E., Bass, D., Blaney, D., Boynton, W., et al. (2008). Introduction to special section on the phoenix mission: Landing site characterization experiments, mission overviews, and expected science. *J. Geophys. Res.* 113:E00A18. doi: 10.1029/2008JE003083

Somovilla, P., Rodríguez-Moreno, A., Arribas, M., Manrubia, S., and Lázaro, E. (2022). Standing genetic diversity and transmission bottleneck size drive adaptation in bacteriophage Qβ. *Int. J. Mol. Sci.* 23:8876. doi: 10.3390/IJMS23168876

Stanley, S. Y., and Maxwell, K. L. (2018). Phage-encoded anti-CRISPR defenses. *Annu. Rev. Genet.* 52, 445–464. doi: 10.1146/annurev-genet-120417-031321

Starr, E. P., Nuccio, E. E., Pett-Ridge, J., Banfield, J. F., and Firestone, M. K. (2019). Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc. Natl. Acad. Sci. U.S.A.* 116, 25900–25908. doi: 10.1073/pnas.1908291116

Stedman, K. M., Clore, A. J., and Combet-Blanc, Y. (2006). "Biogeographical diversity of archaeal viruses," in *GM symposium 66: Prokaryotic diversity: Mechanisms and significance*, eds N. A. Logan, H. M. Lappin-Scott, and P. C. F. Oyston (Cambridge: Cambridge University Press), 131–144.

Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., and Chisholm, S. W. (2005). Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* 3:e144. doi: 10.1371/journal.pbio.0030144

Sullivan, M. B., Weitz, J. S., and Wilhelm, S. (2017). Viral ecology comes of age. *Environ. Microbiol. Rep.* 9, 33–35. doi: 10.1111/1758-2229.12504

Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Acinas, S. G., Babin, M., et al. (2020). Tara oceans: Towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* 18, 428–445. doi: 10.1038/s41579-020-0364-5

Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microb. Ecol.* 28, 237–243. doi: 10.1007/BF00166813

Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160

Suttle, C. A. (2007). Marine viruses–major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750

Suzuki, Y., Hino, H., Hawai, T., Saito, K., Kotsugi, M., and Ono, K. (2020). Symmetry prediction and knowledge discovery from X-ray diffraction patterns using an interpretable machine learning approach. *Sci. Rep.* 101, 1–11. doi: 10.1038/s41598-020-77474-4

Swanson, M. M., Fraser, G., Daniell, T. J., Torrance, L., Gregory, P. J., and Taliansky, M. (2009). Viruses in soils: Morphological diversity and abundance in the rhizosphere. *Ann. Appl. Biol.* 155, 51–60. doi: 10.1111/j.1744-7348.2009.00319.x

Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10104–10111. doi: 10.1073/PNAS.1421398112

Szathmáry, E., and Demeter, L. (1987). Group selection of early replicators and the origin of life. *J. Theor. Biol.* 128, 463–486. doi: 10.1016/S0022-5193(87)80191-1

Szathmáry, E., and Smith, J. M. (1995). The major evolutionary transitions. *Nature* 374, 227–232. doi: 10.1038/374227A0

Szostak, J. W. (2012). Attempts to define life do not help to understand the origin of life. *J. Biomol. Struct. Dyn.* 29, 599–600. doi: 10.1080/073911012010524998

Szostak, J. W. (2017). The narrow road to the deep past: In search of the chemistry of the origin of life. *Angew. Chem. Int. Ed. Engl.* 56, 11037–11043. doi: 10.1002/ANIE.201704048

Takemura, M. (2020). Medusavirus ancestor in a proto-eukaryotic cell: Updating the hypothesis for the viral origin of the nucleus. *Front. Microbiol.* 11:571831. doi: 10.3389/FMICB.2020.571831

Thierauf, A., Perez, G., and Maloy, A. S. (2009). Generalized transduction. *Methods Mol. Biol.* 501, 267–286. doi: 10.1007/978-1-60327-164-6_23

Thingstad, T. F., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* 13, 19–27. doi: 10.3354/ame013019

Thomas, E., Anderson, R. E., Li, V., Rogan, L. J., and Huber, J. A. (2021). Diverse viruses in deep-sea hydrothermal vent fluids have restricted dispersal across ocean basins. *mSystems* 6, e68–e21. doi: 10.1128/MSYSTEMS.00068-21

Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., and Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* 6, 1457–1465. doi: 10.1101/GAD.6.8.1457

Trubl, G., Jang, H., Bin, Roux, S., Emerson, J. B., Solonenko, N., Vik, D. R., et al. (2018). Soil viruses are underexplored players in ecosystem carbon processing. *mSystems* 3, e76–e18. doi: 10.1128/msystems.00076-18

Trus, B. L., Cheng, N., Newcomb, W. W., Homa, F. L., Brown, J. C., and Steven, A. C. (2004). Structure and polymorphism of the UL6 portal protein of herpes simplex virus type 1. *J. Virol.* 78, 12668–12671. doi: 10.1128/JVI.78.22.12668-12671.2004

Vaishampayan, A., and Grohmann, E. (2019). Multi-resistant biofilm-forming pathogens on the international space station. *J. Biosci.* 44:125. doi: 10.1007/s12038-019-9929-8

Vale, P. F., Choisy, M., Froissart, R., Sanjuán, R., and Gandon, S. (2012). The distribution of mutational fitness effects of phage φX174 on different hosts. *Evolution* 66, 3495–3507. doi: 10.1111/J.1558-5646.2012.01691.X

Van den Bergh, B., Swings, T., Fauvart, M., and Michiels, J. (2018). Experimental design, population dynamics, and diversity in microbial experimental evolution. *Microbiol. Mol. Biol. Rev* 82:e00008–e18. doi: 10.1128/mmbr.00008-18

Van Etten, J. L., and Meints, R. H. (1999). Giant viruses infecting algae. *Annu. Rev. Microbiol.* 53, 447–494. doi: 10.1146/ANNUREV.MICRO.53.1.447

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–348. doi: 10.1038/NATURE04388

Villarreal, L. P., and Witzany, G. (2010). Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* 262, 698–710. doi: 10.1016/J.JTBI.2009.10.014

Vitas, M., and Dobovišek, A. (2019). Towards a general definition of life. *Orig. Life Evol. Biosph.* 49, 77–88. doi: 10.1007/s11084-019-09578-5

Waddell, T. E., Franklin, K., Mazzocco, A., Kropinski, A. M., and Johnson, R. P. (2009). Generalized transduction by lytic bacteriophages. *Methods Mol. Biol.* 501, 293–303. doi: 10.1007/978-1-60327-164-6_25

Wang, W., Zhao, H., and Han, G.-Z. (2020b). Host-virus arms races drive elevated adaptive evolution in viral receptors. *J. Virol.* 94, e684–e620. doi: 10.1128/jvi.00684-20

Wang, F., Baquero, D. P., Beltran, L. C., Su, Z., Osinski, T., Zheng, W., et al. (2020a). Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. *Proc. Natl. Acad. Sci. U.S.A.* 117, 19643–19652. doi: 10.1073/PNAS.2011125117

Wasik, B. R., Bhushan, A., Ogbunugafor, C. B., and Turner, P. E. (2015). Delayed transmission selects for increased survival of vesicular stomatitis virus. *Evolution* 69, 117–125. doi: 10.1111/EVO.12544

Watson, B. N. J., Steens, J. A., Staals, R. H. J., Westra, E. R., and van Houte, S. (2021). Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. *Cell Host Microbe* 29, 715–725. doi: 10.1016/j.chom.2021.03.018

Watson, J. D., and Crick, F. H. C. (1956). Structure of small viruses. *Nature* 177, 473–475.

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1, 1–8. doi: 10.1038/nmicrobiol.2016.116

Weiss, R. A. (2017). Exchange of genetic sequences between viruses and hosts. *Curr. Top. Microbiol. Immunol.* 407, 1–29. doi: 10.1007/82_2017_21

Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978. doi: 10.1126/science.1086909

White, R. A., Visscher, P. T., and Burns, B. P. (2021). Between a rock and a soft place: The role of viruses in lithification of modern microbial mats. *Trends Microbiol.* 29, 204–213. doi: 10.1016/J.TIM.2020.06.004

Whittington, A. C., and Rokyta, D. R. (2019). Biophysical spandrels form a hot-spot for kosmotropic mutations in bacteriophage thermal adaptation. *J. Mol. Evol.* 87, 27–36. doi: 10.1007/S00239-018-9882-4

Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A.-K., Zoeller, L., et al. (2004). Comparative genomic analysis of hyperthermophilic archaeal fuselloviridae viruses. *J. Virol.* 78, 1954–1961. doi: 10.1128/jvi.78.4.1954-1961.2004

Wigington, C. H., Sonderegger, D., Brussaard, C. P. D., Buchan, A., Finke, J. F., Fuhrman, J. A., et al. (2016). Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.* 1, 4–11. doi: 10.1038/nmicrobiol.2015.24

Wilhelm, S. W., and Suttle, C. A. (1999). Viruses and nutrient cycles in the sea. *Bioscience* 49, 781–788. doi: 10.2307/1313569

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271. doi: 10.1128/MR.51.2.221-271.1987

Woese, C. R. (2002). On the evolution of cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8742–8747. doi: 10.1073/PNAS.132266999

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea. Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/PNAS.87.12.4576

Wolf, D., and Goff, S. P. (2008). Host restriction factors blocking retroviral replication. *Annu. Rev. Genet.* 42, 143–163. doi: 10.1146/ANNUREV.GENET.42.110807.091704

Wolf, Y. I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., et al. (2020). Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat. Microbiol.* 5, 1262–1270. doi: 10.1038/s41564-020-0755-4

Yang, B., Fang, L., Gao, Q., Xu, C., Xu, J., Chen, Z. X., et al. (2022). Species-specific KRAB-ZFPs function as repressors of retroviruses by targeting PBS regions. *Proc. Natl. Acad. Sci. U.S.A* 119:e2119415119. doi: 10.1073/pnas.2119415119

Yooseph, S., Nealson, K. H., Rusch, D. B., McCrow, J. P., Dupont, C. L., Kim, M., et al. (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* 468, 60–66. doi: 10.1038/nature09530

Yoshida-Takashima, Y., Nunoura, T., Kazama, H., Noguchi, T., Inoue, K., Akashi, H., et al. (2012). Spatial distribution of viruses associated with planktonic and attached microbial communities in hydrothermal environments. *Appl. Environ. Microbiol.* 78, 1311–1320. doi: 10.1128/AEM.06491-11

Yutin, N., Bäckström, D., Ettema, T. J. G., Krupovic, M., and Koonin, E. V. (2018). Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virol. J.* 15:67. doi: 10.1186/S12985-018-0974-Y

Zablocki, O., Adriaenssens, E. M., and Cowan, D. (2015). Diversity and ecology of viruses in hyperarid desert soils. *Appl. Environ. Microbiol.* 82, 770–777. doi: 10.1128/AEM.02651-15

Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M., and Rudnick, J. (2004). Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15556–15560. doi: 10.1073/pnas.0405844101

Zayed, A. A., Wainaina, J. M., Dominguez-Huerta, G., Pelletier, E., Guo, J., Mohssen, M., et al. (2022). Cryptic and abundant marine viruses at the evolutionary origins of earth's RNA virome. *Science* 376, 156–162. doi: 10.1126/science.abm5847

Zhang, Y. Z., Shi, M., and Holmes, E. C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172. doi: 10.1016/j.cell.2018.02.043

Zhao, L., Seth-Pasricha, M., Stemate, D., Crespo-Bellido, A., Gagnon, J., Draghi, J., et al. (2019). Existing host range mutations constrain further emergence of RNA viruses. *J. Virol.* 93, e1385–e1318. doi: 10.1128/JVI.01385-18

Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., et al. (2013). Abundant SAR11 viruses in the ocean. *Nature* 494, 357–360. doi: 10.1038/nature11921

Zheng, L., Liang, X., Shi, R., Li, P., Zhao, J., Li, G., et al. (2020). Viral abundance and diversity of production fluids in oil reservoirs. *Microorganisms* 8, 1–15. doi: 10.3390/MICROORGANISMS8091429

Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., et al. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* 18, 21–34. doi: 10.1038/s41579-019-0270-x

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fmicb.2022.1035359

OPEN ACCESS

EDITED BY
Antoinette Van Der Kuyl,
University of Amsterdam,
Netherlands

REVIEWED BY
Mahmoud Mohamed Fayez,
Veterinary Serum and Vaccine Research Institute, Egypt
Shimon Ginath,
Wolfson Medical Center,
Israel

*CORRESPONDENCE
Xiuhui Zheng
zhengxiuhui1972@126.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 02 September 2022
ACCEPTED 20 October 2022
PUBLISHED 08 November 2022

CITATION
Liu D, Zhang Y, Chen D, Wang X, Huang F, Long L and Zheng X (2022) Evaluation of the presence of SARS-CoV-2 in vaginal and anal swabs of women with omicron variants of SARS-CoV-2 infection. *Front. Microbiol.* 13:1035359. doi: 10.3389/fmicb.2022.1035359

Evaluation of the presence of SARS-CoV-2 in vaginal and anal swabs of women with omicron variants of SARS-CoV-2 infection

Ding Liu[1†], Yunfu Zhang[2†], Dongfeng Chen[3], Xianhua Wang[4], Fuling Huang[4], Ling Long[4] and Xiuhui Zheng[4*]

[1]Department of Disease Prevention and Control, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing, China, [2]Department of Medical Affair and Research, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing, China, [3]Department of Gastroenterology, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing, China, [4]Department of Gynecology and Obstetrics, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing, China

Objectives: The study aimed to determine whether SARS-CoV-2 Omicron variant could be detected in the vaginal fluid and anal swabs of reproductive-aged and postmenopausal women infected with SARS-CoV-2 Omicron variant.

Methods: Included in this study were 63 women who were laboratory confirmed as having SARS-CoV-2 Omicron variant infection and admitted to the responsible ward of Daping Hospital of at the National Exhibition and Convention Center(Shanghai) Makeshift Hospital from May 1–24, 2022.From them, vaginal and anal swabs were obtained with informed consent. The demographic and baseline clinical characteristics and the swab test results were analyzed.

Results: The 63 included patients ranged in age from 18 to 72 years with a median of 47.71±15.21years. Of them, 38 women (60.3%) were in their reproductive years. Most of the participants (77.8%) were healthy without significant underlying diseases. Fourteen patients (22.2%) had asymptomatic infection and the remaining 49 (77.8%) had mild infection. The upper respiratory tract symptoms including cough (40/63.5%) and sore throat (18/28.6%)were the most common clinical manifestations of these mildly infected patients. Only 5 patients (7.8%) had gastrointestinal (GI) symptoms, including simple diarrhea in 4 patients, and diarrhea with vomiting in one patient. Pharyngeal,vaginal and anal swabs were collected simultaneously from all 63 patients 8–16 (mean 11.25±2.23) days after SARS-Cov-2 Omicron variant infection. The vaginal swabs were negative for SARS-CoV-2 in all 63 patients, and the anal swabs were positive in 4 patients (6.5%). The overall median hospitalization duration was 16.73±3.16days.

Conclusion: The results of the present study suggest that there is a low possibility of SARS-Cov-2 Omicron variant transmission *via* the digestive tract and vaginal fluid. The correlation between the GI symptoms and the presence of viral RNA in anal swabs is uncertain.

KEYWORDS

coronavirus, SARS-CoV-2, omicron, vaginal fluid, anal swab

Introduction

The Omicron variant of the novel coronavirus (COVID-19) has become the dominant epidemic strain in the world since it was first reported in South Africa in November 2021 (Vaughan, 2021). There has been an outbreak in Shanghai, China since March 2022, and the virus strain is Omicron variant BA.2. From March 1 to April 18, 2022, a total of 397,933 infected cases were reported, of which asymptomatic infections accounted for 93.06% of all infected people (Xian et al., 2022)，reflecting the characteristics of high infectivity and low pathogenicity of Omicron (Guo et al., 2022). Previous studies reported that they failed to detect severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) in vaginal fluid swabs of patients with moderate to severe COVID-19 (Cui et al., 2020; Qiu et al., 2020; Uslu Yuvacı et al., 2021). However, viral nucleic acid was detected in patients with mild COVID-19, but only 3 positive cases were reported (Barber et al., 2021; Schwartz et al., 2021). Therefore, whether there is existence of infectious virions of SARS-CoV-2 in the vaginal fluid remains unclear. It was also reported that viruses could exist in the gastrointestinal (GI) tract for a long time and therefore anal swab detection could be used to judge the severity of the patient's condition and guide epidemic prevention and control (Xu et al., 2020; Zhang et al., 2020). However, all these studies were carried out prior to the emergence of the current Omicron rampancy. With the virus mutation, the transmission ability, antibody resistance, immune escape and other abilities of the Omicron mutant have been enhanced (Cameroni et al., 2022), which can be transmitted through the mixture of aerosols and droplets. However, whether the virus will infect the female reproductive and GI tract has not been reported. As different strains may have different relative tropisms to tissues (Chertow et al., 2021), it is necessary to study female vaginal secretions and anal swabs to ascertain whether there are other transmission routes of SARS-CoV-2 Omicron variant beyond the respiratory system for the sake of providing a more scientific basis for its prevention and treatment.

The study aimed to determine whether SARS-CoV-2 Omicron variant was present in the vaginal fluid and anal swab of both reproductive-aged and postmenopausal women with SARS-CoV-2 Omicron variant infection.

Materials and methods

General information

Included in this study were 63 nonpregnant women who were laboratory confirmed as having SARS-CoV-2 Omicron variant infection and admitted to the responsible ward of Daping Hospital of at the National Exhibition and Convention Center(Shanghai) Makeshift Hospital from May 1 to May 24, 2022. They were diagnosed as asymptomatic or having mild infections based on the New Coronavirus Pneumonia Prevention and Control Program (9thedition) published by the National Health Commission of China (NHCC) (National Health Commission of China, 2022). Pharyngeal swab nucleic acid test was performed daily during hospitalization. Participants in pregnancy, menstruation and vaginitis symptoms were excluded from the study. Discharge criteria were as follows: the body temperature had returned normal for more than 3 days, the clinical symptoms of the respiratory tract were significantly ameliorated, and the CT values of SARS-CoV-2 nucleic acid N gene and ORF gene were ≥ 35 on pharyngeal swabs for two consecutive tests by qRT-PCR, the threshold value was 40, and the sampling time was at least 24 h apart.

This study was approved by the ethics committee of the said hospital (2022 No. 122–01), and written informed consent was obtained from all participants.

Treatment measures

According to the NHCC New Coronavirus Pneumonia Prevention and Control Program (9thedition), asymptomatic and mild infections will be subjected to centralized isolation management. During isolation, symptomatic treatments and condition monitoring of the patients were performed. If the condition was aggravated, they would be transferred to designated hospitals for treatment.

Reagents and instruments

The reagents and instruments use in this study were Sansure Natch96 semi-automatic extractor (Changsha, China); Hongshi SLAN PCR instrument(Shanghai, China); new coronavirus nucleic acid detection kit, nucleic acid extraction kit(Sansure Biotech, Changsha, China), and new coronavirus nucleic acid detection kit (Da An Gene Co.Ltd., Guangzhou, China). The fragments detected by the nucleic acid detection kits were ORF1ab/N genes, using the qRT-PCR method.

Methods

Sampling method

Pharyngeal, vaginal and anal swabs were obtained simultaneously from all 63 patients between 8 and 16 days after onset of SARS-CoV-2 infection for SARS-CoV-2 nucleic acid. The pharyngeal swab test was performed by using the sterilized cotton swab rinsed with water to wipe the patient's pharyngeal lateral wall and posterior pharyngeal wall several times and then sealing the swab in the sampling tube. The anal swab test was performed by using the sterilized cotton swab dipped with a small amount of 0.9% sterile sodium chloride solution, inserting it into the anal canal at a depth of 3 cm, rotating it for about 5 s, and sealing the swab in the sampling tube. The vaginal swab test

TABLE 1 Demographic and clinical characteristics of the patients.

Item	Total ($n = 63$)	Asymptomatic ($n = 14$)	Mild infection ($n = 49$) (%)	p-value[a]
Age (mean ± SD), years	47.71 ± 15.21	50.86 ± 16.28	46.82 ± 14.94	0.385
Status of menopause, *n* (%) Postmenopausal	25(39.7)	7(50.0)	18(36.7)	0.371
Reproductive-aged	38(60.3)	7(50.0)	31(63.3)	
Vaccination, *n* (%) Vaccination (2–3 shots)	43(68.3)	8(57.1)	35(71.4)	0.492
Unvaccinated or one shot	20(31.7)	6(42.9)	14(28.6)	
Underlying diseases status, *n* (%)				0.240
None	49(77.8)	13 (92.9)	36 (73.5)	
Combined underlying diseases	14(22.2)	1(7.1)	13(26.5)	
Chronic hypertension, *n* (%)	8(12.7)	0(0.0)	8 (16.3)	0.245
DM2, *n* (%)	5(7.9)	0(0.0)	5 (10.2)	0.578[c]
Others, (%)	5(7.9)	1(7.1)	4 (8.2)	0.258[c]
anal swabs, *n* (%)	4(6.5)	0(0.0)	4(8.3)	0.566[c]
Hospitalization (mean ± SD),days	16.73 ± 3.16	15.79 ± 3.47	17.00 ± 3.06	0.208[b]
Interval between infectiononset and sampling (mean ± SD), days	11.25 ± 2.23	11.14 ± 2.51	11.29 ± 2.17	0.834[b]

[a]chi-square test, [b]independent-sample t test, [c]Fisher exact test.

was performed by inserting the sterilized cotton swab into the vagina in a 5–6 cm depth, rotating it for 4–5 rounds in about 7 s, and then sealing the swab in the sampling tube. All specimens were sent to the laboratory immediately after sampling.

SARS-CoV-2 nucleic acid detection: qRT-PCR was used to detect the open reading frame 1ab (ORF1ab) and nucleocapsid protein (N) in the SARS-CoV-2 genome.

Viral nucleic acid amplification Ct value detection method: the nucleic acid was extracted and placed in a reaction tube for amplification. The SARS-CoV-2 nucleocapsid protein N gene and the open reading frame OR-Flab were used as detection targets, and reverse transcription was performed at 50°C for 10 min (1 cycle); pre denaturation at 97°C for 1 min (1 cycle); denaturation at 97°C for 5 s; annealing or extension or fluorescence detection at 58°C for 30 s, 45 cycles as the amplification conditions.

Interpretation of the SARS-CoV-2 nucleic acid test results:(1) Target gene detection and interpretation. ①Negative: Ct value ≥35 or not detected; ②Positive: S-shaped amplification curve and Ct value<35.(2) Interpretation criteria for positive SARS-CoV-2 detection: ① Both targets (ORF lab, N) in SARS-CoV-2 in the same specimen were positive ② Re-sample and re-test for patients with positive single target. Single target positive, judged as positive.

Statistical processing

In this study, SPSS 21.0 software was used to statistically process and describe the population distribution characteristics of the patients. The measurement data were described in the form of the mean ± standard deviation (x ± s), and the counting data were described in frequency (percentage, n (%)). Comparison of measurement data between two groups was performed by using two independent sample t-test, and comparison of counting data between two groups was performed by using chi square test and accurate probability method.

Results

Epidemiological characteristics

A total of 63 SARS-Cov-2 Omicron variant–infected female cases were collected, who ranged in age from 18 to 72 years with a median of 47.71 ± 15.21 years. Of them, 38 patients were in their reproductive period and 25 postmenopausal women. Twenty-six patients had received 3 vaccinations, 17 had received 2 vaccinations, one had received one vaccination, and 19 had not received any vaccination. Fourteen patients had underlying diseases, including hypertension in 8, type II diabetes mellitus (DM2) in 5, coronary heart disease (CHD) in 1, emphysema in 1, Hashimoto's thyroiditis in 1, 1 week after abortion in 1, and knee joint replacement in 1. Four patients had two underlying diseases at the same time: hypertension+DM2 in 2, DM2 + CHD in 1, and hypertension+knee joint replacement in 1. The interval between onset of SARS-CoV-2 infection and sampling was 8–16 days with a mean of 11.25 ± 2.23 days (Table 1).

Clinical symptoms

Among the 63 patients, 14 (22.2%) were asymptomatic and 49 (77.8%) had mild infections. The main clinical symptoms were cough, sore throat, and fever (>38.0°C), and may have myalgia, diarrhea, loss of taste, and fatigue. Other symptoms included chest tightness, nasal congestion, runny nose, hyposmia and chills (Table 2).

TABLE 2 Clinical symptoms.

Clinical symptoms	*n* (%)
Mild disease: Cough	40 (63.5)
Sore throat	18 (28.6)
Fever (>38°C)	10 (15.9)
Muscle soreness	7 (10.9)
Diarrhea	5 (7.9)
Loss of taste	4 (6.3)
Fatigue	3 (4.8)
Chest tightness	3 (4.8)
Stuffy nose	2 (3.2)
Decreased smell	2 (3.2)
Runny nose	1 (1.6)
Chills	1 (1.6)
Vomiting	1 (1.6)
Total	49 (77.8)
Asymptomatic	14 (22.2)

Nucleic acid test results of specimens

Of the 63 patients, pharyngeal swab test was positive in 54 patients (85.7%), negative in 9 patients (14.3%), Among whom, 6 were asymptomatic and 3 had mild symptoms; anal swab test was positive in 4 patients (6.35%), all of whom had mild symptoms, only one had GI symptoms with diarrhea. Two had underlying diseases with well controlled hypertension, and 2 patients were healthy. Vaginal secretion test was negative in all patients.

Discussion

SARS-COV-2 is a ribonucleic acid (RNA) virus with high mutation potential. The virus adapts to the new host through "mutation-selection-adaptation" (Cameroni et al., 2022), which means that the virus is a mutant strain with a strong replication ability but does not cause disease for most hosts; rather it has the opportunity to be selected in the process of popularity and rises to the main epidemic strain. The host cells actively capture the virus through receptors in the process of SARS-COV-2 infection. With the evolution of the virus, the affinity of the virus and host receptor ACE2 gradually improves, resulting in further reduction in the concentration of virus required for infection. Therefore, a stronger spreading ability and weaker pathogenicity should be expected as the evolution direction of SARS-COV-2 (Cao, 2020; Liu et al., 2021; Cameroni et al., 2022).

SARS-COV-2 has been found in nasopharyngeal secretion, feces, urine, semen, and tear since the outbreak of COVID-19 (Wang et al., 2020). Previous studies on the GI tract mostly focused on moderate-to severe-disease patients, in whom SARS-CoV-2 RNA has been detected in 40–85% fecal samples (Brooks and Bhatt, 2021), indicating that SARS-CoV-2 viral RNA can be detected in feces nearly as frequently as that in respiratory secretions. In addition, the positive rate of nucleic acid detection *via* anal swabs is related to the severity of the disease, and therefore it can be used as a supplement to the detection of respiratory specimens (Xu et al., 2020; Zhang et al., 2020). Natarajan et al. (2022) first studied the feces of patients with mild to moderate COVID-19, and found that fecal SARS-CoV-2 RNA was detected in 49.2% of the participants within the first week of diagnosis. With the extension of time, the positive rate gradually decreased, and the viral RNA was still detected in the feces of 3.8% patients 7 months after the diagnosis. They found that viral RNA in feces was the extended presence and fecal viral RNA shedding was correlated with GI symptoms. Their study is valuable for inferring population-level prevalence of COVID-19 from wastewater studies. For the study of SARS-CoV-2 in the female reproductive tract, no viral RNA was detected in vaginal fluid in patients with severe infection (Cui et al., 2020; Qiu et al., 2020; Uslu Yuvacı et al., 2021). But Schwartz et al. (2021) detected the SARS-CoV-2 virus in the vaginal secretions in 2 out of 35 women aged 21 and 86 years respectively, and both patients had mild symptoms. Sampling was conducted from the patients on day 6 and 11 after infection, respectively. Barber et al.[8]described 35 non-pregnant women with mild to moderately severe disease. SARS-CoV-2 was detected in one post-menopausal patient aged 60 years. Vaginal secretion sampling was performed on the same day of admission and diagnosis of COVID-19， In addition to the 21-year-old healthy woman, the other two patients have a variety of chronic diseases. Therefore, further clarifying the situation of the virus in the reproductive tract will have great significance for determining the risk of sex transmission and mother to child transmission during childbirth.

The Omicron variant is substantially mutated compared to any previously described SARS-CoV-2 isolates. It includes 37 substitutions of residues in the spike protein, 15 of which are clustered in the receptor-binding domain (RBD) (Cameroni et al., 2022), the key position of the ACE-2 receptor-cell interaction[19] and the main target of neutralizing antibodies after infection or vaccination (Liu et al., 2021). A study (Tian et al., 2022) reported that the Omicron RBD had 2.4-fold increased binding affinity to human ACE2 as compared with Wuhan-Hu-1 (Cameroni et al., 2022). The infectivity of Omicron might be more than 10-fold high as that of the original strain and approximately twice high as that of Delta. In the sera of volunteers with 3 vaccinations, the neutralizing antibody titer of the Omicron variant and the Delta variant reduced by 16.5-fold and 3.3-fold, respectively, as compared with that of the original strain (Wang et al., 2022). Therefore, while the gene mutation of the Omicron variant enhances the transmission ability, the ability to escape antibodies is also enhanced, which is the new transmission characteristic of the Omicron variant. To effectively prevent the spread of the virus, it is necessary to determine whether there are other different routes of transmission.

To answer this question, we detected nucleic acid simultaneously in pharyngeal, vaginal and anal swabs in 63 non-pregnant infected women at day 8–16 after the onset of SARS-CoV-2 infection, and found that the positive rate of pharyngeal swabs was 85.7%, positive anal swabs were detected in

only 4 (6.35%) of the 63 included women, which is far lower than previous research[9、10、18], indicating that the ability of SARS-Cov-2 Omicron variant to invade the digestive tract is limited and the possibility of fecal oral transmission is low. No SARS-Cov-2 was detected in vaginal fluid. Our findings are supported by previous reports (Cui et al., 2020; Qiu et al., 2020; Uslu Yuvacı et al., 2021). But Schwartz and Barber reported 3 positive cases in the vaginal fluid (Barber et al., 2021; Schwartz et al., 2021), for which there are several possible explanations. First, although the binding affinity between the Omicron variant and Angiotensin-converting enzyme 2(ACE2) receptors is enhanced (Cameroni et al., 2022; Tian et al., 2022), the expression of the SARS-COV-2 ACE2 receptor in the vaginal tissue is low (Dimitrov, 2003; Uhlén et al., 2015; Morelli et al., 2021). it is possible that the SARS-Cov-2 Omicron variant does not seem to enter the vaginal fluid. Second, the incidence of viremia in patients with COVID-19 is low (Fajnzylber et al., 2020; Kim et al., 2020; Wang et al., 2020; Bwire et al., 2021; Hagman et al., 2021), and detection of the virus is feasible only in the viremia stage of the disease, indicating that the sampling time will affect the positive results. Additionally, this may be related to the Omicron inherent biology, the immune status of the host due to underlying disorders, natural immunization, vaccination, the duration of the virus invasion, and the sample size (Morelli et al., 2021).

The clinical characteristics of the 63 patients in our series were preliminarily analyzed, who ranged in age from 18 to 72 years with a median of 47.71 ± 15.21 years. The proportion of patients with mild infection was significantly higher than that of patients with asymptomatic infection(49/77.8% vs. 14/22.2%). The severity of SARS-Cov-2 infection depends on population characteristics, strain characteristics, and vaccination (Chen et al., 2022). Our statistical analysis showed that age, the menopausal status, underlying diseases, length of hospitalization, and other factors had no significant impact on symptoms (asymptomatic vs. mild) after infection. This may be due to the relatively small number of the participants and the relatively long hospitalization duration as compared with the mean hospitalization duration of 7 days. In our study, 43 patients (68.3%) were vaccinated more than twice, and the incidence of mild disease had no significant difference from that of the uncompleted vaccination group. The four patients with positive anal swabs all received more than 2 vaccinations, suggesting that the gene mutation of the Omicron variant may weaken the effectiveness of SARS-Cov-2 vaccine. However, another study (Puhach et al., 2022) showed that vaccination still had a certain protective effect in reducing the incidence of severe diseases, the risk of death and transmission. In the current epidemic in Shanghai, asymptomatic infections accounted for 93.06%, low severe infections and low mortality also supported the above view.

The most common clinical manifestations of mild infected patients were upper respiratory tract symptoms such as cough and sore throat, which may be related to the fast replication rate of the Omicron variant in bronchial epithelial cells, and the weak replication rate in alveolar cells (Hui et al., 2022). Of the five patients with diarrhea, the anal swab test was positive in only one patient, suggesting that the correlation between the GI symptoms and the presence of viral RNA in the GI tract remains uncertain, and the GI symptoms may largely be the manifestation of systemic symptoms after infection.

This study has some limitations. First, the samples were only extracted from patients 8–16 days after the infection, and dynamic research on other periods is absent. It cannot fully explain the characteristics of virus transmission through the lower genital tract during the whole infection process. In addition, although asymptomatic patients and those with mild infection could be complementary to previous studies to some extent, the study did not include patients with other subtypes of Omicron variants. More comprehensive epidemiological studies with larger sample sizes are required to further verify the findings and conclusions in this study.

In summary, our results demonstrated that among asymptomatic and mildly infected non-pregnant women, the Omicron variant has limited dissemination capacity through the GI tract, and does not possibly spread through the lower genital tract. However, this conclusion needs to be confirmed by further in-depth and comprehensive research. At present, the number of patients infected with Omicron mutants continues to increase globally. It seems that the variation and transmission of SARS-CoV-2 has reached normalcy. This study will not only give guidance to choosing the approach of laboring for pregnant women with Omicron infection, but also provide useful references for the prevention and control of COVID-19.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Daping Hospital, Army Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XZ had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis and supervision. XZ and DL: study concept and design. DL and YZ: acquisition of data and drafting of the manuscript. DL, YZ, DC, XW, FH, and LL: analysis and interpretation of data and statistical analysis. DL, YZ, and XZ: critical revision of the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

Funding

The work was supported by the Talent Innovation Ability Training Plan of the Army Medical Center (2019CXLCC014); Chongqing Natural Science Foundation (cstc2021jcyj-msxmX1027); Chongqing Medical Scientific Research Project (Joint Project of Chongqing Health Commission and Science and Technology Bureau) (2019ZDXM009).

Acknowledgments

The authors would like to extend sincere gratitude to XZ for instructive advice and useful suggestions in preparing the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Barber, E., Kovo, M., Leytes, S., Sagiv, R., Weiner, E., Schwartz, O., et al. (2021). Evaluation of SARS-CoV-2 in the vaginal secretions of women with COVID-19: a prospective study. *J. Clin. Med.* 10:2735. doi: 10.3390/jcm10122735

Brooks, E. F., and Bhatt, A. S. (2021). The gut microbiome: a missing link in understanding the gastrointestinal manifestations of COVID-19? Cold spring Harb. *Mol. Case Stud.* 7:a006031. doi: 10.1101/mcs.a006031

Bwire, G. M., Majigo, M. V., Njiro, B. J., and Mawazo, A. (2021). Detection profile of SARS-CoV-2 using RT-PCR in different types of clinical specimens: a systematic review and meta-analysis. *J. Med. Virol.* 93, 719–725. doi: 10.1002/jmv.26349

Cameroni, E., Bowen, J. E., Rosen, L. E., Saliba, C., Zepeda, S. K., Culap, K., et al. (2022). Broadly neutralizing antibodies overcome SARS-CoV-2 omicron antigenic shift. *Nature* 602, 664–670. doi: 10.1038/s41586-021-04386-2

Cao, G. W. (2020). Key issues on the evolution, related epidemiological features, and specific prophylaxis of severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2)[J]. *Shanghai J. Prevent. Med.* 32, 697–703. doi: 10.19428/j.cnki.sjpm.2020.20587

Chen, X. H., Yan, X. M., Sun, K. Y., Zheng, N., Sun, R. J., Zhou, J. X., et al. (2022). Estimation of disease burden and clinical severity of COVID-19 caused by Omicron BA.2 in Shanghai. *Emerging Microbes & Infections*, 1–10. doi: 10.1101/2022.07.11.22277504

Chertow, D., Stein, S., Ramelli, S., Grazioli, A., Chung, J. Y., Singh, M., et al. (2021). SARS-CoV-2 infection and persistence throughout the human body and brain. *Res. Square.* doi: 10.21203/rs.3.rs-1139035/v1

Cui, P., Chen, Z., Wang, T., Dai, J., Zhang, J., Ding, T., et al. (2020). Severe acute respiratory syndrome coronavirus 2 detection in the female lower genital tract. *Am. J. Obstet. Gynecol.* 223, 131–134. doi: 10.1016/j.ajog.2020.04.038, indexed in Pub Med: 32376320

Dimitrov, D. S. (2003). The secret life of ACE2 as a receptor for the SARS virus. *Cells* 115, 652–653. doi: 10.1016/s0092-8674(03)00976-0

Fajnzylber, J., Regan, J., Coxen, K., Corry, H., Wong, C., Rosenthal, A., et al. (2020). SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat. Commun.* 11:5493. doi: 10.1038/s41467-020-19057-5

Guo, Z., Zhao, S., Lee, S. S., Mok, C. K. P., Wong, N. S., Wang, J., et al. (2022). Superspreading potential of COVID-19 outbreak seeded by omicron variants of SARS-CoV-2 in Hong Kong. *J. Travel Med.* 29:taac049. doi: 10.1093/jtm/taac049

Hagman, K., Hedenstierna, M., Gille-Johnson, P., Hammas, B., Grabbe, M., Dillner, J., et al. (2021). SARS-CoV-2 RNA in serum as predictor of severe outcome in COVID-19: a retrospective cohort study. *Clin. Infect. Dis.* 73:e2995-e 3001. doi: 10.1093/cid/ciaa1285

Hui, K. P. Y., Ho, J. C. W., Cheung, M. C., Ng, K. C., Ching, R. H. H., Lai, K. L., et al. (2022). SARS-CoV-2 omicron variant replication in human bronchus and lung *ex vivo*. *Nature* 603, 715–720. doi: 10.1038/s41586-022-04479-6

Kim, J. M., Kim, H. M., Lee, E. J., Jo, H. J., Yoon, Y., Lee, N. J., et al. (2020). Detection and isolation of SARS-CoV-2 in serum, urine, and stool specimens of COVID-19 patients from the Republic of Korea. *Osong. Public Health Res. Perspect.* 11, 112–117. doi: 10.24171/j.phrp.2020.11.3.02

Liu, J., Chen, X., Liu, Y., et al. (2021). Characterization of SARS-CoV-2 worldwide transmission based on evolutionary dynamics and specific viral mutations in the spike protein [J]. *Infect. Dis. Poverty* 10:112. doi: 10.1186/s40249-021-00895-4

Morelli, F., Meirelles, L. E. F., de Souza, M. V. F., Mari, N. L., Mesquita, C. S. S., Dartibale, C. B., et al. (2021). COVID-19 infection in the Human reproductive tract of men and nonpregnant women. *Am. J. Trop. Med. Hyg.* 104, 814–825. doi: 10.4269/ajtmh.20-1098

Natarajan, A., Zlitni, S., Brooks, E. F., Vance, S. E., Dahlen, A., Hedlin, H., et al. (2022). Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection [J]. *Med (NY)* 3, 371–387. doi: 10.1016/j.medj.2022.04.001

National Health Commission of China. (2022). New coronavirus pneumonia prevention and control program. *9th Edn.* (in Chinese). [EB/OL].

Puhach, O., Adea, K., Hulo, N., Sattonnet, P., Genecand, C., Iten, A., et al. (2022). Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or omicron SARS-CoV-2. *Nat. Med.* 28, 1491–1500. doi: 10.1038/s41591-022-01816-0

Qiu, L., Liu, X., Xiao, M., Xie, J., Cao, W., Liu, Z., et al. (2020). SARS-CoV-2 is not detectable in the vaginal fluid of women with severe COVID-19 infection. *Clin. Infect. Dis.* 71, 813–817. doi: 10.1093/cid/ciaa375, indexed in Pub Med: 32241022

Schwartz, A., Yogev, Y., Zilberman, A., Alpern, S., Many, A., Yousovich, R., et al. (2021). Detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in vaginal swabs of women with acute SARS-CoV-2 infection: a prospective study. *BJOG Int. J. Obstet. Gynaecol.* 128, 97–100. doi: 10.1111/1471-0528.16556

Tian, D., Sun, Y., Xu, H., and Ye, Q. (2022). The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 omicron variant [J]. *J. Med. Virol.* 94, 2376–2383. doi: 10.1002/jmv.27643

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419. doi: 10.1126/science.1260419

Uslu Yuvacı, H., Musa Aslan, M., Kose, O., Akdemir, N., Hande, T., Cevrioğlu, A., et al. (2021). Evaluation of the presence of SARS-CoV-2 in the vaginal fluid of reproductive-aged women. *Ginekol. Pol.* 92, 406–409. doi: 10.5603/GPa 2021.0018

Vaughan, A. (2021). Omicron emerges. *New Sci.* 252:7. doi: 10.1016/S0262-4079(21) 02140-0

Wang, K., Jia, Z. J., Bao, L. L., Wang, L., Cao, L., Chi, H., et al. (2022). Memory B cell repertoire from triple vaccinees against diverse SARS-CoV-2 variants [J]. *Nature* 603, 919–925. doi: 10.1038/s41586-022-04466-x

Wang, L., Wang, Y., Ye, D., and Liu, Q. (2020). Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence [J]. *Int. J. Antimicrob. Agents* 55:105948. doi: 10.1016/j.ijantimicag.2020.105948

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., et al. (2020). Detection of SARS-CoV-2 in different types of clinical specimens. *J. Am. Med. Assoc.* 323, 1843–1844. doi: 10.1001/jama.2020.3786

Xian, L. F., Lin, J. S., Yu, S. C., Zhao, Y., Zhao, P., Cao, G. W., et al. (2022). Epidemiological characteristics of SARS-CoV-2 infection outbreak in Shanghai in the spring of 2022. *Shanghai J. Prevent. Med.* 34, 294–299. doi: 10.19428/j.cnki.sjpm.2022.22058

Xu, C. Y., Song, J. F., Liu, S. X., Zheng, H., Kang, X. W., Li, Y., et al. (2020). Analysis of the positive rate of anal swab nucleic acid test and clinical characteristics in patients with different severity of coronavirus disease 2019. *Chin. Crit. Care Med.* 32, 1171–1173. doi: 10.3760/cma.j.cn121430-20200611-00631

Zhang, W., Du, R. H., Li, B., Zheng, X. S., Yang, X. L., Hu, V., et al. (2020). Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg. Microb. Infection.* 9, 386–389. doi: 10.1080/22221751.2020.1729071

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 11 November 2022
DOI 10.3389/fmicb.2022.1040285

OPEN ACCESS

EDITED BY
Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands

REVIEWED BY
Fangbin Song,
Shanghai General Hospital,
China
Xinfeng Qu,
Peking University,
China

*CORRESPONDENCE
You-Lin Qiao
qiaoy@cicams.ac.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 09 September 2022
ACCEPTED 21 October 2022
PUBLISHED 11 November 2022

CITATION
Li Z-F, Jia X-H, Ren X-Y, Wu B-K, Chen W, Feng X-X, Wang L-B and Qiao Y-L (2022) Comparison of the performance of HPV DNA chip test and HPV PCR test in cervical cancer screening in rural China. *Front. Microbiol.* 13:1040285. doi: 10.3389/fmicb.2022.1040285

Comparison of the performance of HPV DNA chip test and HPV PCR test in cervical cancer screening in rural China

Zhi-Fang Li[1,2†], Xin-Hua Jia[1,3†], Xin-Yu Ren[4], Bei-Ke Wu[4], Wen Chen[1], Xiang-Xian Feng[2], Li-Bing Wang[5] and You-Lin Qiao[6*]

[1]Department of Epidemiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of Preventive Medicine, Changzhi Medical College, Changzhi, China, [3]State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, National Institute of Diagnostics and Vaccine Development in Infectious Diseases, Collaborative Innovation Center of Biologic Products, School of Public Health, Xiamen University, Xiamen, Fujian, China, [4]School of Public Health, Shanxi Medical University, Taiyuan, China, [5]Department of Pathology, Affiliated Heping Hospital of Changzhi Meidical College, Changzhi, China, [6]Center for Global Health, School of Population Medicine and Public Health Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Background: This study aimed to evaluate the performance of two different principles of HPV testing in primary cervical cancer screening and ASC-US triage in rural areas.

Methods: 3,328 and 3,913 women were enrolled in Shanxi, China in 2017 and 2018, respectively, and screened using liquid-based cytology and different HPV tests with a 4-year follow-up. Different screening methods commonly used in clinical practice were evaluated.

Results: In the HPV PCR test cohort, the prevalence of HPV infection was 14.90%. A total of 38 cases of CIN2+ were identified at baseline, 2 of which were in the HPV-negative cohort and the rest in the HPV-positive cohort (2=186.85, p<0.001). Fifty-three cases of CIN2+ were accumulated over 4years. The HPV infection rate in the HPV DNA chip test cohort was 21.10%. A total of 26 CIN2+ cases were identified at baseline, all in the HPV-positive population (2=92.96, p<0.001). 54 CIN2+ cases were cumulative over 4years. At 4-year follow-up, HPV-negative results were significantly more protective against cervical intraepithelial neoplasia grade 2 or worse (CIN2+) than normal cytologic results at baseline. HPV screening was more sensitive and specific than cytologic screening (using ASC-US as the threshold) and performed better on the HPV DNA microarray test. In addition, compared with HPV 16/18 testing, sensitivity increases and specificity decreases when using HPV testing for cytologic ASC-US triage, regardless of which HPV test is used.

Conclusion: In the rural areas where we implemented the study, HPV tests performed well for screening than LBC and HPV DNA chip testing performed better than HPV PCR testing in the screening cohort. Optimal screening was achieved technically when used in combination with LBC for ASC-US population triage, without thinking the feasibility for resource availability.

KEYWORDS
human papillomavirus, cervical cancer, HPV DNA chip test, HPV PCR test, screening

Introduction

Cervical cancer is one of the top three most common malignancies among women worldwide. The Global Burden of Cancer 2020 report published by the International Agency for Research on Cancer of the World Health Organization showed that there were about 604,000 new cases of cervical cancer and 342,000 deaths worldwide in 2020; and there were about 110,000 new cases and 59,000 deaths in China, accounting for 18.2 and 17.3% of the global incidence and deaths, respectively. In addition, there was a trend toward younger age (Kang et al., 2014; Sung et al., 2021). Persistent infection with high-risk human papillomavirus (HR-HPV) is considered the leading cause of cervical cancer and precancerous lesions. HR-HPV infection can be detected in 60–70% and 20–40% of cervical lesions CIN2 to CIN3 and CIN1, respectively (Xu et al., 2016). More than 99% of patients with cervical cancer are detected with HPV infection in the cervical specimens (Sias et al., 2019). The 2020 guidelines recommended HPV testing as the primary screening tool for cervical cancers. However, in most developing countries (World Health Organization, 2021), such as China, the more widely used screening tool for cervical cancer is liquid-based cytology testing. HPV testing is gradually replacing cytology because it is fast and has high sensitivity and specificity. However, the feasibility of HPV screening test in resource-constrained areas need to be properly assessed (Banerjee et al., 2022).

The main kit technologies that have been marketed for HPV nucleic acid detection are the fluorescent PCR method, Hybrid Capture 2 (HC2), biochip, and flow-through fluorescent hybridization (Stoler et al., 2020). There is some variability in the results using different HPV DNA detection methods, which is mainly related to the principles of the detection methods and other aspects (Michelli et al., 2011). HPV DNA chip technology, a new HPV detection technology, has been evaluated in fewer studies. In this study, the real-time fluorescent PCR method was compared with the HPV DNA chip method. The aim was to compare the performance of two different HPV test kits using different screening strategies, that is, primary screening for cervical cancer and triage effect in ASC-US population.

Materials and methods

Study population

Two counties of Changzhi City, Shanxi Province (Changzhi county and Wuxiang County) were selected to establish cervical cancer screening cohorts. From May to June 2017, 3,328 rural women in Changzhi county underwent cervical cancer screening; From August to September 2018, 3,913 rural women in Wuxiang County underwent cervical cancer screening.

The inclusion criteria were as follows: (1) age ≥ 21 years with an intact uterus; (2) no history of cervical cancer treatment or cervical surgery; (3) no sexual intercourse within 48 h after screening and no vaginal medications, vaginal contraceptives, or vaginal washings within 48 h; (4) no suspected clinical pregnancy symptoms; pregnant women could participate in the study up to 8 weeks after the end of pregnancy; (5) signed informed consent.

Study design and procedure

All study participants were included in the study after giving informed consent, completing a questionnaire and a gynecological examination. The questionnaires were administered by trained investigators and included general information about the study participants, such as marital status, education level, history of smoking and alcohol consumption, menstrual history and reproductive history.

A trained gynecologist performed gynecological examinations of the vulva, vagina, and cervix for each enrolled woman. Participants in Changzhi County had one cervical exfoliated cell collected by a gynecologist and preserved in cytology preservation solution (PreservCyt solution) for liquid-based cytology testing (LBC) and HPV DNA testing. Participants in Wuxiang County had two cervical samples collected by gynecologists: one cervical exfoliated cell was collected using a single-use cervical sampling swab and preserved in cytology preservation solution (PreservCyt solution) for LBC testing; one cervical sample (cervical exfoliated epithelial cells or secretions, etc.) was collected and preserved in cell preservation solution (Jiangsu Jiangyou Medical Technology Co. Ltd.) in a single-use sampling kit for HPV DNA testing.

Within 12 weeks of collection of cervical exfoliated cell specimens, those with positive results for HPV types 16 and/or 18 and cytology results ≥ASC-US were referred to undertake colposcopy by an experienced gynecologist. If the colposcopy result was abnormal, a cervical biopsy or endocervical curettage (ECC) was performed on the subject. Participants with HPV-negative test results and normal cytology results at baseline were followed up for cervical cytology after 3 years. In contrast, participants with HPV-positive test results or cytology ≥ASC-US at baseline received annual cervical cytology follow-up (3 years). Participants with ≥ASC-US results at follow-up received colposcopy, and participants with abnormal colposcopy received histopathology. During the study, participants with

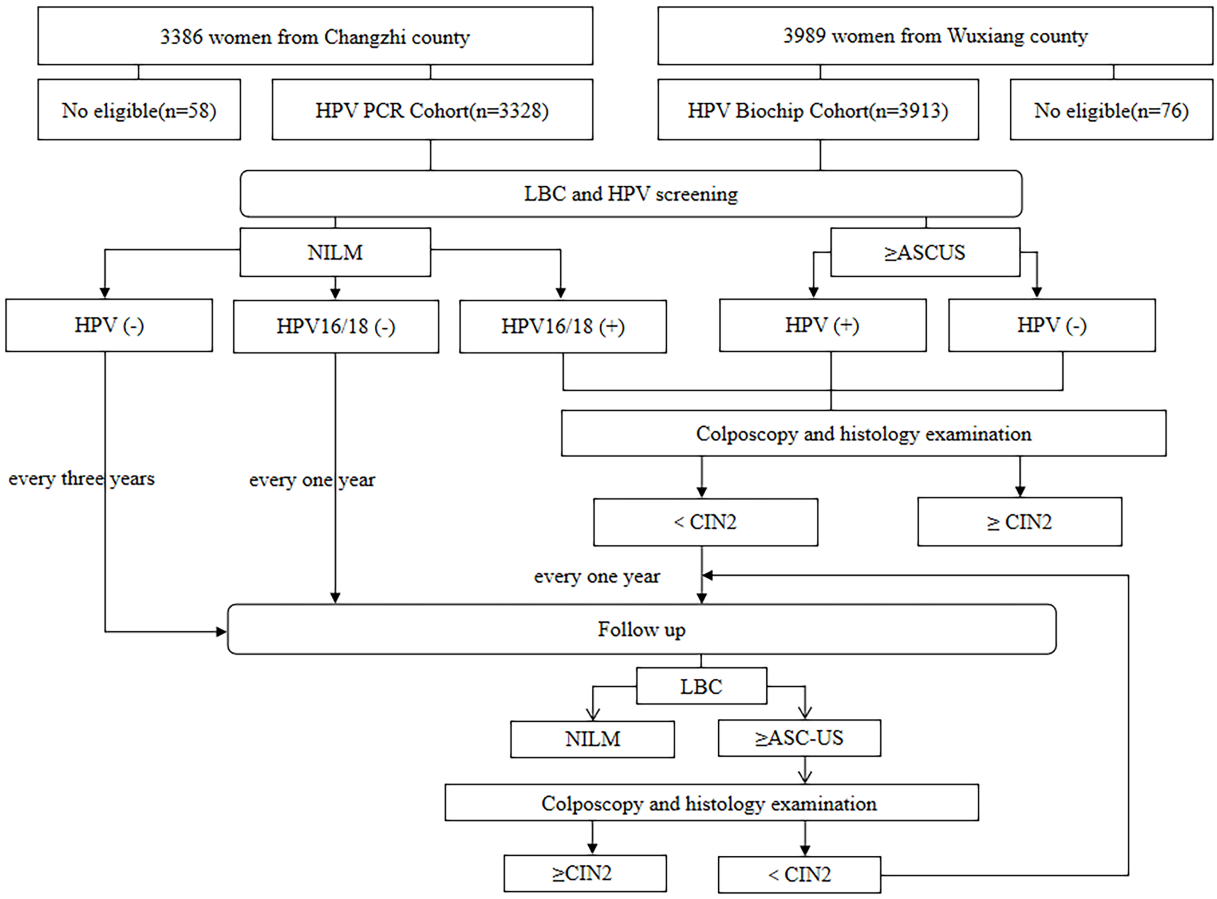

FIGURE 1
Flow chart of HPV PCR cohort and HPV biochip cohort.

histopathological findings ≥CIN2 discontinued follow-up. as detailed in Figure 1.

Before program implementation, standard training was provided for cervical exfoliation cytology sampling (for cytology testing and HPV testing), stain production, reading, laboratory testing, colposcopy, and histopathology. All program staff were required to pass an assessment by the designated medical institution. Senior physicians from the Heping Hospital of Changzhi Medical College were on site for long periods of time to provide quality control guidance during the program.

Laboratory tests

HPV DNA chip test

HPV typing nucleic acid detection kit (biochip method) (Beijing Bohui Innovation Biotechnology Group Co., Ltd.) uses a nucleic acid amplification-reverse spot hybridization technique combining PCR *in vitro* amplification and DNA reverse spot hybridization method to detect 14 (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) high-risk human papillomavirus nucleic acid subtypes.

HPV PCR test

The HPV DNA typing test was performed using the high-risk HPV5 + 9 nucleic acid test kit (Tellgen Corporation). The test kit is based on fluorescent PCR technology and detects HPV 16, 18, 33, 52, 58, and 9 other HPV (31, 35, 39, 45, 51, 56, 59, 66, 68) typing for a total of 14 HPV and human genomic β-Globin DNA in a single reaction tube using a four-channel fluorescent PCR instrument with a minimum detection limit of 100 copies/ test, simultaneous detection of cervical exfoliated cell gene β-Globin.

Pathological diagnosis

All of the following tests and diagnostic procedures were strictly double-blinded. Cytology slides were read by two pathologists results and reported according to the Bethesda 2014 classification. The cytological results were as follows: negative for intraepithelial lesion or malignancy (NILM), atypical cells of undetermined significance (ASC-US),

low-grade squamous intraepithelial lesion (LSIL), atypical squamous cells–cannot exclude high-grade squamous intraepithelial lesion (ASC-H), high-grade squamous intraepithelial lesion (HSIL), atypical glandular cells (AGC), and cervical cancer cells. Diagnoses were reported if the diagnoses by two cytologists were consistent. Otherwise, a third cytologist was consulted.

All participants with positive HPV 18 and/or HPV 16 testing or abnormal cytology (ASC-US or worse) were referred to undertake colposcopy. Two pathologists independently made diagnoses according to the 2014 WHO Classification of Tumors of the Female Genital Tract. If the diagnoses were concordant, they were reported as the pathologic diagnosis. Otherwise, a panel of pathologists was consulted to reach a consensual diagnosis. The histological diagnoses of cervical lesions were classified as normal, LSIL, HSIL/CIN2, HSIL/CIN3. HPV testing, cytology, and pathologic examination were performed with blinding to the results of each test. Participants with both negative HPV and cytology results were not referred for colposcopy and were considered normal.

Statistical analysis

In this study, an ACCESS database was developed, double-entered and checked until complete consistency before analysis. Data were analyzed using R software (version 4.1.2). We reported numbers and percentages for categorical variables. Chi-square test were used to compare the difference between groups. $p < 0.05$(two-sided) was considered statistically significant. Using cervical intraepithelial neoplasia grade 2 or worse as a reference standard, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic (ROC) curve (AUC) were calculated.

Based on a review of the published literature (Cox et al., 2013; Chatzistamatiou et al., 2016), we chose to evaluate six common clinical screening strategies. Strategy 1 used cytology as primary screening, referring participants with LBC ≥ ASC-US for colposcopy and participants with normal cytology for routine screening. Strategy 2 used HPV testing as primary screening and referred participants who were positive for HPV-16/18 for colposcopy. Strategies 3 and 4 used combined screening, with strategy 3 referring participants with ASC-US and HPV-16/18 positivity or LBC ≥ LSIL for colposcopy and strategy 4 referring participants with ASC-US and HPV positivity or LBC ≥ LSIL for colposcopy, and participants who did not undergo colposcopy were followed up at one-year or three-year intervals. Strategies 5 and 6 were to concurrently link cytology with HPV testing, i.e., strategy 5 referred participants for colposcopy if HPV 16/18 were positive or LBC ≥ ASC-US. Strategy 6 referred participants with HPV positivity or LBC ≥ ASC-US for colposcopy.

Results

A total of 3,328 participants were included in the HPV PCR testing cohort, with an HPV prevalence of 14.90%. The majority of the participants were aged 40 years or above. About half of them completed secondary education, and almost all of them were married, non-smokers, and non-drinkers. There were no statistically significant differences in different subgroups of age at menarche, age at first pregnancy, number of pregnancies, and number of births.

A total of 3,919 participants were included in the HPV DNA chip testing cohort, with an HPV prevalence of 21.10%, and the distribution of population characteristics was similar to that of the HPV testing cohort and was overall comparable, as shown in Table 1.

Table 2 demonstrates the distribution of baseline and 4-year cumulative histopathological findings with different cytologic diagnoses at baseline and HPV status. In the population tested with HPV PCR, a total of 38 cases of CIN2+ were identified at baseline, 2 of which were found in the HPV-negative population and the rest in the HPV-positive population ($\chi^2 = 186.85$, $p < 0.001$). In the HPV DNA chip population, a total of 26 cases of CIN2+ were identified at baseline, all of which were found in the HPV-positive population ($\chi^2 = 92.96$, $p < 0.001$). Among the 4-year cumulative cases, a total of 53 cases of CIN2+ were found in the population with HPV PCR testing, 51 of which were in the HPV-positive population ($\chi^2 = 274.36$, $p < 0.001$). Fifty-four cases were found cumulatively in the HPV DNA chip population, all of which were in the HPV-positive population ($\chi^2 = 199.55$, $p < 0.001$), and all of these differences were statistically significant.

In the two cohorts based on different HPV detection methods, the specificity was better, around 90%, if LBC alone was used as a screening strategy, but the sensitivity was poorer and there was a more severe underdiagnosis. And the difference in positive predictive values for baseline CIN2+ and 4-year cumulative CIN2+ was not significant. If strategy 2 is implemented, that is, colposcopy in HPV16/18 positive population, the sensitivity reached around 80% and HPV detection based on HPV DNA chip technology is superior to conventional PCR technology with a sensitivity of 84.62 (95% *CI*: 64.27, 94.95) and 77.36 (95% *CI*: 63.45, 87.27), respectively. There was a relatively good improvement in sensitivity and specificity when LBC was combined with HPV testing for screening. The sensitivity of both strategies 3 and 4 decreased when using CIN2+ as the gold standard. However, precancerous lesions are a progressive state. Therefore, there was a significant improvement in sensitivity and specificity when using the 4-year cumulative CIN2+ as the gold standard. For strategies 3, 4, 5 and 6, HPV DNA chip technology performed better than PCR HPV test. The same trend was shown for CIN2+ over the next 4 years (Table 3).

Figure 2 shows the ROC curves plotted for different screening strategies. When using cytology only as a screening method (strategy 1), the AUC of HPV PCR assay = 0.71, 0.77, respectively, at baseline and 4 year cumulative, which are much lower than

TABLE 1 Population characteristics of different HPV testing cohorts at baseline.

Characteristic	HPV PCR test cohort (*N*, %)		*p*	HPV DNA chip test cohort (*N*, %)		*p*
	+	−		+	−	
Age			0.28			<0.01
21 ~ 29 years	12 (2.42)	109 (3.85)		26 (3.14)	104 (3.37)	
30 ~ 39 years	99 (19.96)	576 (20.34)		123 (14.87)	674 (21.84)	
≥40 years	385 (77.62)	2,147 (75.81)		678 (81.98)	2,308 (74.79)	
Education level			0.14			0.01
Primary School and below	116 (23.39)	577 (20.37)		270 (32.65)	850 (27.54)	
Middle School	227 (45.77)	1,344 (47.46)		357 (43.17)	1,344 (43.55)	
High School	89 (17.94)	455 (16.07)		78 (9.43)	269 (8.72)	
University and above	64 (13.90)	456 (16.10)		122 (14.75)	623 (20.19)	
Marital Status			0.11			0.01
Married	482 (97.18)	2,782 (98.23)		806 (97.46)	3,059 (99.13)	
Unmarried	14 (2.82)	50 (1.77)		21 (2.54)	27 (0.87)	
Smoking status			0.15			0.46
Never smoked	495 (99.80)	2,832 (100.00)		826 (99.88)	3,078 (99.74)	
Smoking	1 (0.20)	0 (0)		1 (0.12)	8 (0.26)	
Drinking status			0.03			0.01
Never drink	412 (83.06)	2,459 (86.83)		807 (97.58)	3,028 (98.12)	
Drink	84 (16.93)	373 (13.17)		20 (2.42)	58 (1.88)	
Age at menarche			0.32			0.01
≤14 years	271 (54.64)	1,616 (57.06)		374 (45.22)	1,602 (51.91)	
>14 years	225 (45.36)	1,216 (42.94)		453 (54.78)	1,484 (48.09)	
Age of first pregnancy			0.11			0.01
≤23 years	333 (67.14)	1795 (63.38)		479 (57.92)	1,510 (48.93)	
>23 years	163 (32.86)	1,037 (36.62)		348 (42.08)	1,576 (51.07)	
Number of pregnancies			0.85			0.13
≤3	394 (79.44)	2,260 (79.80)		527 (63.72)	2054 (66.56)	
>3	102 (20.56)	572 (20.20)		300 (36.28)	1,032 (33.44)	
Number of births			0.27			0.07
≤2	447 (90.12)	2,595 (91.63)		551 (66.63)	2,157 (69.90)	
>2	49 (9.88)	237 (8.37)		276 (33.37)	929 (30.10)	

HPV Chip assay = 0.87, 0.91, respectively. as shown in A and B. C and D indicate ROC curves for strategies 3, 4, 5, and 6 at baseline and 4-year cumulative, respectively. HPV testing as a primary screening method or for ASC-US triage, and it can be concluded that when using CIN2+ as an endpoint, LBC with HPV chip test combined screening can achieve the maximum AUC (0.93) at baseline, for 4 years of cumulative CIN2+ as well.

Discussion

The aim of cervical cancer screening is to detect and treat precancerous lesions early before they occur. In addition, women who participated in organized cervical cancer screening had a 41–92% reduction in cervical cancer mortality (Jansen et al., 2020) and a 50–60% reduction in all-cause mortality (Arbyn et al., 2012) compared to women who did not participate in screening. Therefore, cervical cancer screening remains an effective strategy for the prevention and control of cervical cancer (Simms et al., 2016). In rural areas, the lack of health resources is one of the most challenging factors. Measures to improve screening efficiency, reduce missed diagnoses, and decrease colposcopy referrals can improve health resource utilization. Combining cytology with HPV can improve screening in areas with poor health resources (Mezei et al., 2017; Jansen et al., 2021). Numerous studies have shown that HPV-based cervical screening is more effective than cytology alone (Balasubramanian et al., 2010; Ronco et al., 2014; Schiffman et al., 2015; Zhao et al., 2020). However, increased colposcopy referral rates are a potential problem with HPV-based screening (Zhao et al., 2021), whereas cytology has the inherent advantage of high specificity (Thomsen et al., 2021). In addition, it remains controversial whether combination testing or HPV testing alone should be used as the primary screening tool (Castle et al., 2011; Cox et al., 2013; Blatt et al., 2015; Cuschieri et al., 2018;

TABLE 2 Histopathological distribution in different cytological diagnoses regarding HPV status.

	HPV-negative results (*N*, %)				HPV-positive results (*N*, %)			
	No colposcopy	No CIN	CIN 1	CIN 2+	No colposcopy	No CIN	CIN 1	CIN 2+
Baseline								
HPV PCR test cohort								
NILM	2,478 (99.84)	2 (0.08)	2 (0.08)	0 (0.00)	305 (77.02)	51 (12.88)	23 (5.81)	17 (4.29)
ASC-US	82 (68.91)	28 (23.53)	8 (6.72)	1 (0.84)	16 (55.17)	8 (27.59)	2 (6.90)	3 (10.34)
ASC-H	0 (0.00)	25 (47.17)	27 (50.94)	1 (1.89)	5 (18.52)	7 (25.93)	8 (29.63)	7 (25.93)
AGC	2 (16.67)	9 (75.00)	1 (8.33)	0 (0.00)	1 (20.00)	1 (20.00)	3 (60.00)	0 (0.00)
LSIL	47 (34.31)	55 (40.15)	35 (25.55)	0 (0.00)	5 (20.83)	6 (25.00)	10 (41.67)	3 (12.50)
HSIL+	3 (10.34)	17 (58.62)	9 (31.03)	0 (0.00)	3 (20.00)	2 (13.33)	4 (26.67)	6 (40.00)
Total	2,612 (92.23)	136 (4.80)	82 (2.90)	2 (0.07)	335 (67.54)	75 (15.12)	50 (10.08)	36 (7.26)
HPV DNA chip test cohort								
NILM	2,874 (99.90)	3 (0.10)	0 (0.00)	0 (0.00)	597 (94.31)	24 (3.79)	8 (1.26)	4 (0.63)
ASC-US	66 (55.46)	53 (44.54)	0 (0.00)	0 (0.00)	46 (44.66)	32 (31.07)	19 (18.45)	6 (5.83)
ASC-H	2 (28.57)	4 (57.14)	1 (14.29)	0 (0.00)	4 (23.53)	9 (52.94)	2 (11.76)	2 (11.76)
AGC	0 (0.00)	1 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (100.00)
LSIL	48 (59.26)	30 (37.04)	3 (3.70)	0 (0.00)	13 (24.07)	29 (53.70)	8 (14.81)	4 (7.41)
HSIL+	0 (0.00)	1 (100)	0 (0.00)	0 (0.00)	1 (5.26)	2 (10.53)	7 (36.84)	9 (47.37)
Total	2,990 (96.89)	92 (2.98)	4 (0.13)	0 (0.00)	661 (79.93)	96 (11.61)	44 (5.32)	26 (3.14)
4-year cumulative								
HPV PCR test cohort								
NILM	2,476 (99.84)	2 (0.08)	2 (0.08)	0 (0.00)	300 (77.32)	48 (12.37)	22 (5.67)	18 (4.64)
ASC-US	82 (68.91)	28 (23.53)	8 (6.72)	1 (0.84)	16 (45.71)	8 (22.86)	2 (5.71)	9 (25.71)
ASC-H	0 (0.00)	25 (47.17)	27 (50.94)	1 (1.89)	4 (14.81)	6 (22.22)	8 (29.63)	9 (33.33)
AGC	2 (16.67)	9 (75.00)	1 (8.33)	0 (0.00)	0 (0.00)	1 (25.00)	1 (25.00)	2 (50.00)
LSIL	47 (34.31)	55 (50.15)	35 (25.55)	0 (0.00)	5 (19.23)	6 (23.08)	9 (34.62)	6 (23.08)
HSIL+	3 (10.34)	17 (58.62)	9 (31.03)	0 (0.00)	3 (18.75)	2 (12.50)	4 (25.00)	7 (43.75)
Total	2,610 (92.23)	136 (4.81)	82 (2.90)	2 (0.07)	328 (66.13)	71 (14.31)	46 (9.27)	51 (10.28)
HPV DNA chip test cohort								
NILM	2,874 (99.90)	3 (0.10)	0 (0.00)	0 (0.00)	585 (94.35)	22 (3.55)	8 (1.29)	5 (0.81)
ASC-US	66 (55.46)	53 (44.54)	0 (0.00)	0 (0.00)	46 (43.81)	30 (28.57)	15 (14.29)	14 (13.33)
ASC-H	2 (28.57)	4 (57.14)	1 (14.29)	0 (0.00)	3 (12.50)	9 (37.50)	1 (4.17)	11 (45.83)
AGC	0 (0.00)	1 (100)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (100.00)
LSIL	48 (59.26)	30 (37.04)	3 (3.70)	0 (0.00)	13 (23.64)	26 (47.27)	7 (12.73)	9 (16.36)
HSIL+	0 (0.00)	1 (100.00)	0 (0.00)	0 (0.00)	1 (4.55)	2 (9.09)	5 (22.73)	14 (63.64)
Total	2,990 (96.89)	92 (2.98)	4 (0.13)	0 (0.00)	648 (78.36)	89 (10.76)	36 (4.35)	54 (6.53)

Schiffman et al., 2018; Chan et al., 2020). Therefore, this study evaluated two testing methods by comparing primary screening for HPV 16/18 and triage for the ASC-US population.

In this study, we analyzed two large cohorts of women with baseline screening and 4-year histological follow-up. Both cohorts in this study were from resource-constraint areas and had similar demographic characteristics, so they were comparable. In the PCR cohort, 51 (1.53%) CIN2+ cases were found cumulatively over 4 years, while in the chip cohort, 54 (1.38%) cases were found cumulatively over 4 years, with no statistically significant differences. In the present study, DNA microarray outperformed PCR performance. The DNA microarray performed better when LBC was combined with HPV screening, both for baseline CIN2+ and cumulative CIN2+ over 4 years. Differences in detection principles may be an important reason for this difference. The sensitivity of LBC (ASC-US+) in the two cohorts in this study was 55.26 and 84.62%, respectively. The difference in cytology sensitivity between the PCR and chip cohorts may be related to the performance of cytology physicians and field sampling, and it is important to note that these physicians were from low to moderate resource areas.

In a previous study that pooled 40 studies, it showed that different principles of HPV testing perform differently in screening (Koliopoulos et al., 2017). HC2 is the conventional HPV DNA detection technique. For CIN 2+, the pooled sensitivity of HC2 and LBC (ASC-US+) was estimated to be 89.9 and 72.9%, and the pooled specificity was estimated to be 89.9 and 90.3%. The sensitivity of LBC (ASC-US+) was generally between 52 and 94% (Malloy et al., 2000; Tsiodras et al., 2010). The sensitivity of PCR ranged from 75 to 100%

TABLE 3 The sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV) of various assays for CIN2+ endpoints.

Screening strategy	SE (%, 95CI)	SP (%, 95CI)	PPV (%, 95CI)	NPV (%, 95CI)
HPV PCR test cohort				
CIN2+ at baseline				
Strategy 1	55.26 (38.47, 71.01)	86.96 (85.75, 88.08)	4.67 (2.98, 7.15)	99.40 (99.03, 99.64)
Strategy 2	76.32 (59.39, 87.97)	96.47 (95.77, 97.07)	20.00 (14.01, 27.62)	99.72 (99.44, 99.86)
Strategy 3	52.63 (36.05, 68.69)	91.19 (90.15, 92.12)	6.45(4.09, 9.94)	99.40 (99.04, 99.64)
Strategy 4	52.63 (36.05, 68.69)	90.55 (89.48, 91.51)	6.04 (3.83, 9.33)	99.40 (99.03, 99.63)
Strategy 5	97.37 (84.57, 99.86)	84.07 (82.77, 85.30)	6.60 (4.75, 9.06)	99.96 (99.77, 100.00)
Strategy 6	100.00 (88.57, 100.0)	75.44 (73.93, 76.90)	4.49(3.24, 6.17)	100.00 (99.80, 100.00)
4-year cumulative CIN2+				
Strategy 1	66.04 (51.64, 78.11)	87.08 (85.88, 88.20)	7.64 (5.45, 10.57)	99.37 (98.99, 99.62)
Strategy 2	77.36 (63.45, 87.27)	96.67 (95.98, 97.25)	27.33 (20.54, 35.32)	99.62 (99.32, 99.80)
Strategy 3	80.77 (60.02, 92.69)	94.75 (93.99, 95.42)	9.33 (6.01, 14.10)	99.86 (99.66, 99.95)
Strategy 4	84.62 (64.27, 94.95)	93.26 (92.41, 94.01)	7.75 (5.03, 11.65)	99.89 (99.70, 99.96)
Strategy 5	98.11 (88.62, 99.90)	84.27 (82.97, 85.50)	9.17 (6.98, 11.93)	99.96 (99.77, 100.00)
Strategy 6	100.00 (91.58, 100.0)	75.79 (74.27, 77.24)	6.26 (4.77, 8.17)	100.00 (99.81, 100.00)
HPV DNA chip test cohort				
CIN2+ at baseline				
Strategy 1	84.62 (64.27, 94.95)	90.20 (89.21, 91.11)	5.46 (3.53, 8.27)	99.89 (99.69, 99.96)
Strategy 2	84.62 (64.27, 94.95)	94.88 (94.13, 95.54)	9.95 (6.48, 14.87)	99.89 (99.70, 99.97)
Strategy 3	60.38 (46.02, 73.24)	91.33 (90.30, 92.26)	10.13 (7.13, 14.12)	99.30 (98.92, 99.56)
Strategy 4	64.15 (49.75, 76.51)	90.69 (89.63, 91.65)	10.03 (7.14, 13.85)	99.36 (98.99, 99.61)
Strategy 5	100.0 (83.98, 100.00)	87.03 (85.93, 88.07)	4.91 (3.29, 7.20)	100.00 (99.86, 100.00)
Strategy 6	100.0 (83.98, 100.00)	74.02 (72.60, 75.38)	2.51 (1.68, 3.71)	100.00 (99.83, 100.00)
4-year cumulative CIN2+				
Strategy 1	90.74 (78.94, 96.54)	90.49 (89.51, 91.39)	11.78 (8.92, 15.36)	99.86 (99.65, 99.95)
Strategy 2	72.22 (58.14, 83.14)	95.26 (94.53, 95.90)	17.57 (12.93, 23.36)	99.59 (99.31, 99.76)
Strategy 3	83.33 (70.21, 91.64)	95.08 (94.33, 95.73)	19.15 (14.44, 24.89)	99.76 (99.52, 99.88)
Strategy 4	92.45 (80.93, 97.55)	93.58 (92.74, 94.32)	16.50 (12.56, 21.33)	99.89 (99.70, 99.96)
Strategy 5	98.15 (88.82, 99.90)	87.43 (86.34, 88.45)	9.85(7.53, 12.76)	99.97 (99.81, 100.00)
Strategy 6	100.00 (91.73, 100.0)	74.55 (73.14, 75.92)	5.21 (3.97, 6.79)	100.00 (99.83, 100.00)

and the specificity ranged from 85 to 97%, which is overall consistent with the performance of this study. In most studies (An et al., 2003; Lee et al., 2005; Inoue et al., 2006; Jun et al., 2015), the HPV DNA chip test was used primarily for HPV genotyping.

The advantage of this study is the ability to see the long-term effects of screening through 4-year follow-up. Cytology sampling and diagnosis are from rural areas, which more realistically reflects the performance of cytology in areas with poor and moderate resources, but this is also a disadvantage. In addition, the cytology performance of the two cohorts was inconsistent, which may be related to the inconsistent performance of physicians in the regions or because of differences in population characteristics. Thirdly, as colposcopy was not performed in our programme in the cytologically normal and HPV-negative population, this may have led to partial under-diagnosis. Furthermore, colposcopic images were not reviewed and the subjectivity of the colposcopist can influence sampling. Finally, health resources and demographic characteristics vary widely across different provinces in China. Also, the cohort population was from two counties in Shanxi province, which is not fully representative of rural China. Therefore, the results of this study need to be further validated in other rural areas.

More studies should be conducted in the future to confirm the performance of HPV DNA chip. The development of HPV DNA chip with low price and high performance could provide more options for HPV test-based cervical cancer screening.

Conclusion

In conclusion, we selected two rural areas in China for a four-year follow-up to evaluate cytology and different HPV testing principles and concluded that HPV testing is more effective than LBC for screening and HPV DNA microarray testing is more effective than HPV PCR testing. Optimal screening can be achieved when used in combination with LBC for ASC-US population triage, if consideration is given to the feasibility of resource availability.

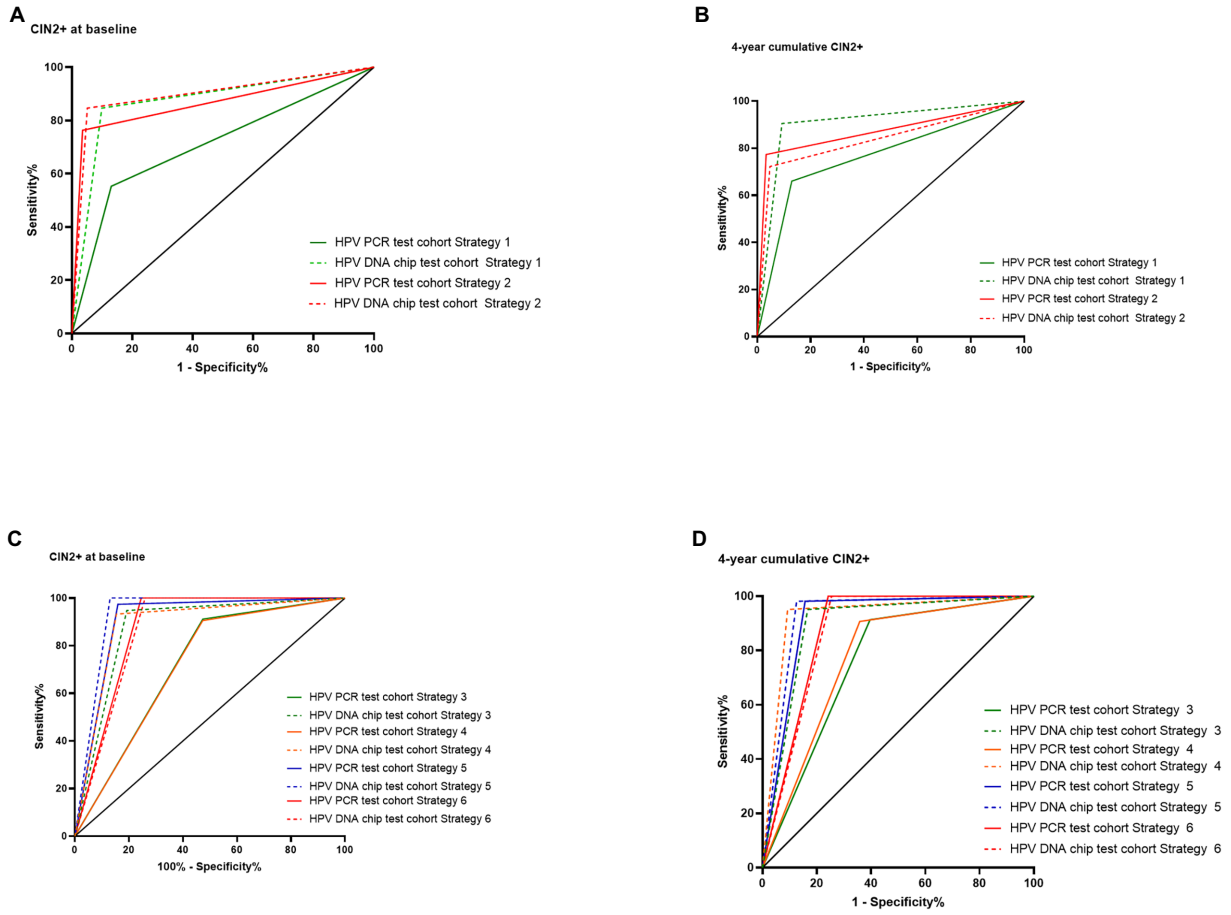

FIGURE 2
ROC curves of HPV DNA Chip and PCR testing techniques in different screening strategies.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Affiliated Heping Hospital of Changzhi Meidical College. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Y-LQ and X-XF contributed to the design. Z-FL and X-HJ wrote the manuscript. B-KW, X-YR, and WC performed statistical analysis and HPV test. L-BW contributed to cytology and histology examination. All authors read and approved the final manuscript.

Funding

This research was supported by CAMS Innovation Fund for Medical Sciences (no. CAMS 2021-I2M-1-004).

Acknowledgments

We thank the support and co-operation of all the study participants, as well as local healthcare workers from the Wuxiang and Zezhou Women and Children Health Service Centers. We are also grateful for the contribution of our coworkers in conducting the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

An, H. J., Cho, N. H., Lee, S. Y., Kim, I. H., Lee, C., Kim, S. J., et al. (2003). Correlation of cervical carcinoma and precancerous lesions with human papillomavirus (HPV) genotypes detected with the HPV DNA chip microarray method. *Cancer* 97, 1672–1680. doi: 10.1002/cncr.11235

Arbyn, M., Weiderpass, E., and Capocaccia, R. (2012). Effect of screening on deaths from cervical cancer in Sweden. *BMJ* 344:e804. doi: 10.1136/bmj.e804

Balasubramanian, A., Kulasingam, S. L., Baer, A., Hughes, J. P., Myers, E. R., Mao, C., et al. (2010). Accuracy and cost-effectiveness of cervical cancer screening by high-risk human papillomavirus DNA testing of self-collected vaginal samples. *J. Low. Genit. Tract Dis.* 14, 185–195. doi: 10.1097/LGT.0b013e3181cd6d36

Banerjee, D., Mittal, S., Mandal, R., and Basu, P. (2022). Screening technologies for cervical cancer: overview. *Cytojournal* 19:23. doi: 10.25259/CMAS_03_04_2021

Blatt, A. J., Kennedy, R., Luff, R. D., Austin, R. M., and Rabin, D. S. (2015). Comparison of cervical cancer screening results among 256, 648 women in multiple clinical practices. *Cancer Cytopathol.* 123, 282–288. doi: 10.1002/cncy.21544

Castle, P. E., Stoler, M. H., Wright, T. C., Sharma, A., Wright, T. L., and Behrens, C. M. (2011). Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: a subanalysis of the ATHENA study. *Lancet Oncol.* 12, 880–890. doi: 10.1016/S1470-2045(11)70188-7

Chan, K. K. L., Liu, S. S., Wei, N., Ngu, S. F., Chu, M. M. Y., Tse, K. Y., et al. (2020). Primary HPV testing with cytology versus cytology alone in cervical screening-a prospective randomized controlled trial with two rounds of screening in a Chinese population. *Int. J. Cancer* 147, 1152–1162. doi: 10.1002/ijc.32861

Chatzistamatiou, K., Moysiadis, T., Moschaki, V., Panteleris, N., and Agorastos, T. (2016). Comparison of cytology, HPV DNA testing and HPV 16/18 genotyping alone or combined targeting to the more balanced methodology for cervical cancer screening. *Gynecol. Oncol.* 142, 120–127. doi: 10.1016/j.ygyno.2016.04.027

Cox, J. T., Castle, P. E., Behrens, C. M., Sharma, A., Wright, T. C., and Cuzick, J. (2013). Comparison of cervical cancer screening strategies incorporating different combinations of cytology, HPV testing, and genotyping for HPV 16/18: results from the ATHENA HPV study. *Am. J. Obstet. Gynecol.* 208, 184.e1–184.e11. doi: 10.1016/j.ajog.2012.11.020

Cuschieri, K., Ronco, G., Lorincz, A., Smith, L., Ogilvie, G., Mirabello, L., et al. (2018). Eurogin roadmap 2017: triage strategies for the management of HPV-positive women in cervical screening programs. *Int. J. Cancer* 143, 735–745. doi: 10.1002/ijc.31261

Inoue, M., Sakaguchi, J., Sasagawa, T., and Tango, M. (2006). The evaluation of human papillomavirus DNA testing in primary screening for cervical lesions in a large Japanese population. *Int. J. Gynecol. Cancer* 16, 1007–1013. doi: 10.1136/ijgc-00009577-200605000-00009

Jansen, E., Naber, S. K., Aitken, C. A., de Koning, H. J., van Ballegooijen, M., and de Kok, I. (2021). Cost-effectiveness of HPV-based cervical screening based on first year results in the Netherlands: a modelling study. *BJOG* 128, 573–582. doi: 10.1111/1471-0528.16400

Jansen, E. E. L., Zielonke, N., Gini, A., Anttila, A., Segnan, N., Vokó, Z., et al. (2020). Effect of organised cervical cancer screening on cervical cancer mortality in Europe: a systematic review. *Eur. J. Cancer* 127, 207–223. doi: 10.1016/j.ejca.2019.12.013

Jun, S. Y., Park, E. S., Kim, J., Kang, J., Lee, J. J., Bae, Y., et al. (2015). Comparison of the Cobas 4800 HPV and HPV 9G DNA Chip tests for detection of high-risk human papillomavirus in cervical specimens of women with consecutive positive HPV tests but negative pap smears. *PLoS One* 10:e0140336. doi: 10.1371/journal.pone.0140336

Kang, L. N., Castle, P. E., Zhao, F. H., Jeronimo, J., Chen, F., Bansil, P., et al. (2014). A prospective study of age trends of high-risk human papillomavirus infection in rural China. *BMC Infect. Dis.* 14:96. doi: 10.1186/1471-2334-14-96

Koliopoulos, G., Nyaga, V. N., Santesso, N., Bryant, A., Martin-Hirsch, P. P., Mustafa, R. A., et al. (2017). Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database Syst. Rev.* 2018:CD008587. doi: 10.1002/14651858.CD008587.pub2

Lee, G. Y., Kim, S. M., Rim, S. Y., Choi, H. S., Park, C. S., and Nam, J. H. (2005). Human papillomavirus (HPV) genotyping by HPV DNA chip in cervical cancer and precancerous lesions. *Int. J. Gynecol. Cancer* 15, 81–87. doi: 10.1136/ijgc-00009577-200501000-00013

Malloy, C., Sherris, J. D., and Herdman, C. A. (2000). HPV/DNA testing: Technical and programmatic issues in the prevention of cervical cancer in low-resource settings. https://screening.iarc.fr/doc/HPV-DNA-Testing-Issues.pdf

Mezei, A. K., Armstrong, H. L., Pedersen, H. N., Campos, N. G., Mitchell, S. M., Sekikubo, M., et al. (2017). Cost-effectiveness of cervical cancer screening methods in low- and middle-income countries: a systematic review. *Int. J. Cancer* 141, 437–446. doi: 10.1002/ijc.30695

Michelli, E., Téllez, L., Mendoza, J.-A., Jürgensen, C., Muñoz, M., Pérez, S., et al. (2011). Comparative analysis of three methods for HPV DNA detection in cervical samples. *Investig. Clin.* 52, 344–357.

Ronco, G., Dillner, J., Elfström, K. M., Tunesi, S., Snijders, P. J., Arbyn, M., et al. (2014). International HPV screening working group: efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet* 383, 524–532. doi: 10.1016/S0140-6736(13)62218-7

Schiffman, M., Boyle, S., Raine-Bennett, T., Katki, H. A., Gage, J. C., Wentzensen, N., et al. (2015). The role of human papillomavirus genotyping in cervical cancer screening: a large-scale evaluation of the cobas HPV test. *Cancer Epidemiol. Biomark. Prev.* 24, 1304–1310. doi: 10.1158/1055-9965.EPI-14-1353

Schiffman, M., Kinney, W. K., Cheung, L. C., Gage, J. C., Fetterman, B., Poitras, N. E., et al. (2018). Relative performance of HPV and cytology components of Cotesting in cervical screening. *J. Natl. Cancer Inst.* 110, 501–508. doi: 10.1093/jnci/djx225

Sias, C., Salichos, L., Lapa, D., Del Nonno, F., Baiocchini, A., and Capobianchi, M. R. (2019). Garbuglia AR: alpha, Beta, gamma human PapillomaViruses (HPV) detection with a different sets of primers in oropharyngeal swabs, anal and cervical samples. *Virol. J.* 16:27. doi: 10.1186/s12985-019-1132-x

Simms, K. T., Smith, M. A., Lew, J. B., Kitchener, H. C., Castle, P. E., and Canfell, K. (2016). Will cervical screening remain cost-effective in women offered the next generation nonavalent HPV vaccine? Results for four developed countries. *Int. J. Cancer* 139, 2771–2780. doi: 10.1002/ijc.30392

Stoler, M. H., Baker, E., Boyle, S., Aslam, S., Ridder, R., Huh, W. K., et al. (2020). Approaches to triage optimization in HPV primary screening: extended genotyping and p 16/Ki-67 dual-stained cytology-retrospective insights from ATHENA. *Int. J. Cancer* 146, 2599–2607. doi: 10.1002/ijc.32669

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Thomsen, L. T., Kjær, S. K., Munk, C., Ørnskov, D., and Waldstrøm, M. (2021). Benefits and potential harms of human papillomavirus (HPV)-based cervical cancer screening: a real-world comparison of HPV testing versus cytology. *Acta Obstet. Gynecol. Scand.* 100, 394–402. doi: 10.1111/aogs.14121

Tsiodras, S., Georgoulakis, J., Chranioti, A., Voulgaris, Z., Psyrri, A., Tsivilika, A., et al. (2010). Hybrid capture vs. PCR screening of cervical human papilloma virus infections. Cytological and histological associations in 1270 women. *BMC Cancer* 10:53. doi: 10.1186/1471-2407-10-53

World Health Organization (2021). *WHO Guideline for Screening and Treatment of Cervical Pre-Cancer Lesions for Cervical Cancer Prevention*. 2nd Edn. Available at: https://www.who.int/publications/i/item/9789240030824 (Accessed October 26, 2022).

Xu, L., Verdoodt, F., Wentzensen, N., Bergeron, C., and Arbyn, M. (2016). Triage of ASC-H: a meta-analysis of the accuracy of high-risk HPV testing and other markers to detect cervical precancer. *Cancer Cytopathol.* 124, 261–272. doi: 10.1002/cncy.21661

Zhao, Y., Bao, H., Ma, L., Song, B., Di, J., Wang, L., et al. (2021). Real-world effectiveness of primary screening with high-risk human papillomavirus testing in the cervical cancer screening programme in China: a nationwide, population-based study. *BMC Med.* 19. doi: 10.1186/s12916-021-02026-0

Zhao, X. L., Xu, X. Q., Duan, X. Z., Rezhake, R., Hu, S. Y., Wang, Y., et al. (2020). Comparative performance evaluation of different HPV tests and triaging strategies using self-samples and feasibility assessment of thermal ablation in 'colposcopy and treat' approach: a population-based study in rural China. *Int. J. Cancer* 147, 1275–1285. doi: 10.1002/ijc.32881

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fmicb.2022.999290

OPEN ACCESS

EDITED BY
Christine A. King,
Upstate Medical University,
United States

REVIEWED BY
Qiyi Tang,
Howard University,
United States
Michael M. Nevels,
University of St. Andrews,
United Kingdom

*CORRESPONDENCE
Emma Poole
elp27@cam.ac.uk

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 20 July 2022
ACCEPTED 02 November 2022
PUBLISHED 24 November 2022

CITATION
Poole E and Sinclair J (2022) Latency-associated upregulation of SERBP1 is important for the recruitment of transcriptional repressors to the viral major immediate early promoter of human cytomegalovirus during latent carriage. *Front. Microbiol.* 13:999290. doi: 10.3389/fmicb.2022.999290

Latency-associated upregulation of SERBP1 is important for the recruitment of transcriptional repressors to the viral major immediate early promoter of human cytomegalovirus during latent carriage

Emma Poole* and John Sinclair

Department of Medicine, University of Cambridge, Cambridge, United Kingdom

Suppression of human cytomegalovirus (HCMV) major immediate early gene (IE) expression from the viral major immediate early promoter (MIEP) is known to be crucial for the establishment and maintenance of HCMV latency in myeloid progenitor cells and their undifferentiated derivatives. This suppression of the MIEP during latent infection is known to result from epigenetic histone modification imparting a repressive chromatin structure around the MIEP in undifferentiated myeloid cells. In contrast, reactivation, resulting from, e.g., myeloid cell differentiation, is associated with activatory chromatin marks around the MIEP. Recently, recruitment of the transcriptional repressor SETDB1, *via* KAP1, to latent HCMV genomes was shown to be involved in latency-associated MIEP suppression in CD34+ progenitor cells. KAP1 is also known to associate with Chromodomain-helicase-DNA-binding protein 3 (CHD3) as part of the NuRD complex which can aid transcriptional silencing. We now show that the cellular protein Plasminogen activator inhibitor 1 RNA-binding protein (SERBP1), a known interactor of CHD3, is significantly upregulated during HCMV latency and that this protein is required for MIEP suppression during latent infection of myeloid cells. We further show that SERBP1 mediates CHD3 association with the MIEP as well as KAP1 association with viral genomic DNA. We suggest that SERBP1 functions as a scaffold protein to recruit transcriptional repressors to the latent viral genome and to mediate transcriptional silencing of the MIEP during latent carriage.

KEYWORDS

human cytomegalovirus, latency, CHD3, SERBP1, KAP1

Introduction

HCMV is the prototypic beta-herpesvirus which is usually asymptomatic upon primary infection but establishes a lifelong infection which is tightly controlled by the host immune system. In contrast, HCMV causes significant disease in the absence of a functioning immune system such as in those individuals undergoing immunosuppression during organ

transplantation or in immunonaive neonates. The reason that HCMV infection is never cleared from the infected individual but maintained for life is, at least partly, due to its ability to establish a "latent" infection whereby cells carry viral genome in the absence of production of infectious virions but in a form that can be reactivated (Poole et al., 2014). For HCMV, it is known that one site of latency *in vivo* is in undifferentiated cells of the myeloid lineage, in particular CD34+ bone marrow progenitor cells and their monocyte derivatives (Sinclair and Poole, 2014). However, once these cells differentiate to macrophages or dendritic cells (DCs), virus sporadically reactivates from latency and virions are produced. In the immunocompetent, such sporadic reactivation is well controlled by the host immune response and is sub-clinical. In the immunocompromised, however, such reactivation events are not controlled and lead to virus dissemination and lytic infection in multiple cell types.

In lytic infection, a temporal cascade of viral gene expression occurs starting with immediate early (IE) gene expression, driven from the viral major IE promoter/enhancer (MIEP), which is key to driving subsequent expression of early (E) and then late (L) viral genes. In contrast, during latent infection, suppression of the viral MIEP occurs which is associated with histone marks of transcriptional repression around the MIEP; in contrast to histone marks of transcriptional activation which are present on the MIEP during reactivation (Sinclair and Poole, 2014). Although a plethora of cellular and viral factors have been identified which negatively or positively regulate the MIEP in a differentiation-dependent manner, how these are orchestrated to control HCMV latency and reactivation is still far from clear. One of the cellular factors known to be involved in MIEP suppression during latency is the Krüppel-associated box (KRAB)-associated protein 1 (KAP1/TRIM28) which has been shown to be recruited to the viral genome in latently infected myeloid cells and phosphorylation of KAP1 appears to act as a switch between latency and reactivation (Rauwel et al., 2015).

Upon chromatin tethering, KAP1 is known to scaffold epigenetic repressive components such as Histone 3 Lysine 9 (H3K9) methyltransferases (e.g., SETDB1; Schultz et al., 2002), Heterochromatin-Protein 1 (HP1) proteins (Nielsen et al., 1999; Ryan et al., 1999) as well as the NuRD histone deacetylase complex (Schultz et al., 2001, 2002), to promote chromatin condensation and transcriptional repression. Members of the NuRD complex, include CHD3 and CHD4 and phosphorylation of KAP1, which prevents KAP1-mediated repression of the HCMV MIEP, is known to cause CHD3/4 to be released from the repressive NuRD complex (Goodarzi et al., 2011) which can result in transcriptional activation (Hoffmeister et al., 2017). Thus, the presence of CHD3 and 4 is often associated with a transcriptional repression.

However, while the NuRD complex is generally associated with transcriptional repression, it is known that CHD3 and CHD4 can also have transcriptional activation properties (Hoffmeister et al., 2017). Indeed, during DNA damage repair, CHD4 and CHD3 act antagonistically to mediate activatory and repressive complexes, respectively. Therefore, CHD4 recruitment to the NuRD complex occurs when Kap1 is phosphorylated and repressive CHD3 is removed (Goodarzi et al., 2011). Interestingly, during lytic infection, it is known that viral pUL38 and pUL29/28 interact with CHD4 NuRD complexes to enhance IE accumulation (Terhune et al., 2010). Thus, it is possible that a balance of CHD3 and CHD4 recruitment to the NuRD complex could act to regulate the latent/lytic switch. While is it known that lytic viral genes are required to recruit CHD4-containing NuRD complexes during lytic infection to activate the MIEP in permissive cells, it is not known if CHD3 or 4 are recruited to viral genomes during latency, and, if so, how this may affect latency.

One known binding partner of CHD3 is SERBP1 (Lemos et al., 2003) which suggested that SERBP1 may play a role in regulation of the NuRD complex. Additionally, SERBP1 has been shown to maintain high levels of methionine in the cell which may also be needed for repressive histone modification (Kosti et al., 2020). Both activities of SERBP1 could lend themselves to functions which might be predicted to be required for the maintenance of a latent infection. Here we show that SERBP1 is significantly upregulated during HCMV latency and that, without SERBP1, the MIEP is active in otherwise latently infected cells. Furthermore, our observations show that, in the absence of SERBP1, the HCMV genome is not associated with either KAP1 or CHD3, suggesting that SERBP1 may act as a scaffold protein to enhance a repressive chromatin structure to aid the maintenance of HCMV latency.

Materials and methods

Cells and viruses

Primary monocytes were isolated from apheresis cones or venous blood as described previously (Aslam et al., 2019). HFFF, THP1, and 293 T cells were obtained from ATCC and maintained as previously (Poole et al., 2021).Two HCMV TB40E-derived viruses were used which have been described previously: TB40E-IE2YFP, which expresses YFP fused to the immediate early gene IE2 (IE86) and TB40E-GATA2mCherry which expresses mCherry from the GATA2 promoter (Elder et al., 2019; Poole et al., 2019). Lentivirus to generate shRNA SERBP1KD cells was obtained from Santa Cruz.

Proteomic screen

The original unbiased proteomic screen identifying the upregulation of SERBP1 during latent HCMV carriage in monocytes has been published in full (Aslam et al., 2019).

Western blot

Cell lysates were analyzed by Western blotting using the following primary antibodies: Anti-actin rabbit polyclonal

(Abcam), anti-SERBP1 (Abcam) followed by the appropriate horseradish peroxidase (HRP)-conjugated secondary antibody, and analyzed by chemiluminescence.

Immunofluorescence

Adherent cells were fixed with 4% paraformaldehyde for 20 min before permeabilization with 0.5% Triton X-100 for 20 min and blocking in 3% bovine serum albumin (BSA)/PBS for 1 h. After this time, primary antibody was added. Anti-SERBP1 primary antibody was added (1 in 200 dilution) for 1 h at RT and, after washing, cells were then incubated with the relevant Alexa Fluor 488 secondary antibody with Hoechst 33342 for 1 h before visualization by fluorescence microscopy.

Chromatin immunoprecipitation

Chromatin immunoprecipitations (ChIPs) were carried out using the Imprint ChIP kit (Sigma) with antibodies anti-histone H3 (Upstate), anti-dimethylated (K9) histone H3 (Upstate), anti-acetylated (K8) histone H4 (Invitrogen) using the manufacturer protocol and as described previously (Poole et al., 2021).

RTqPCR

RT-qPCR analyses were carried out using the QuantiTect (Qiagen) SYBR kit using standard primers and parameters for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and specific IE exon2/3 and UL138 primers previously described (Poole et al., 2021).

Droplet PCR

Genome copy number was assessed using the Biorad QX200 system as previously described (Poole et al., 2021).

shSERBP1 KD cell line

The lentiviral expression vector encoding SERBP1shRNA and puromycin resistance (Santa Cruz Biotechnology).To generate lentiviral particles, 293 T cells were seeded into 6-well plates at 5×105 cells/well were transfected into 293 T cells using transfection TRANSit 293 (Mirus) according to the manufacturer's instructions. 24 h post transduction, media was replaced with 2.5 ml RPMI supplemented with 30% fetal calf serum. 24 h post media change, media was harvested containing lentivirus. This virus was used to inoculate 2.5 × 105 THP-1 cells which were pelleted, and then resuspended in the lentiviral or control supernatant in a 6-well plate. Polybrene was added to the cells at 2 μg/ml and the cells were then centrifuged in the plate at 600 xg for 45 min, and incubated overnight at 37°C/5% CO2. 5 days post transduction the transduced THP-1 cells were pelleted and resuspended in fresh RPMI supplemented with 10% fetal calf serum. 7 days post transduction, puromycin (Sigma) was added at 2 μg/ml and the selective media was refreshed every 2 days.

Analysis of HCMV latency and reactivation

All latent infections in undifferentiated myeloid cells were validated by RTqPCR demonstrating a repression of IE gene expression and the presence of UL138 and the ability to reactivate virus using differentiation factors, such as PML, as described in the text and in previously published studies (Aslam et al., 2019; Poole et al., 2021).

Results

SERBP1 is upregulated during latency

In a previous proteome analysis using unbiased Tandem Mass Tag technology approach (Aslam et al., 2019) we identified that SERBP1 was one of the most highly upregulated proteins in a primary monocyte model of experimental HCMV latency (a reanalysis of this data is shown in Supplementary Figures 1A,B). Consequently, we validated this upregulation of SERBP1 during HCMV latency using a TB40E-SV40GFP virus, which allows the sorting of latently infected CD14+ monocytes expressing GFP, by western blot analysis. Figure 1A shows that latent infection of CD14+ monocytes results in a substantial increase in SERBP1 levels compared to mock infected cells. Two independent experiments are shown which both demonstrate that the SERBP1 upregulation is very high during latency and that it can be difficult to detect the WT SERBP1 due to such high levels of SERBP1. This increase in SERBP1 was also confirmed in infected cells by indirect immunofluorescence (IF) using a TB40E-GATA2mCherry virus which expresses mCherry under the control of the GATA2 promoter, a promoter known to be active in myeloid cells. Figure 1B shows that when latently infected CD14+ monocytes (red) were stained for SERBP1 (green) there was an upregulation of SERBP1 in the latently infected cells compared to uninfected bystander cells (Figure 1B top panel) which was not observed in mock infected cells (Figure 1B bottom panel). To test whether the SERBP1 protein was upregulated as a consequence of mRNA upregulation, levels of SERBP1 RNA were analyzed by RTqPCR. Figure 1C shows that SERBP1 mRNA was upregulated during HCMV latency, which is consistent with the observed protein upregulation. We next addressed whether SERBP1 was also upregulated during lytic infection, to determine whether SERBP1 upregulation was just a consequence of any HCMV infection. Figure 1D shows that SERBP1 protein expression was

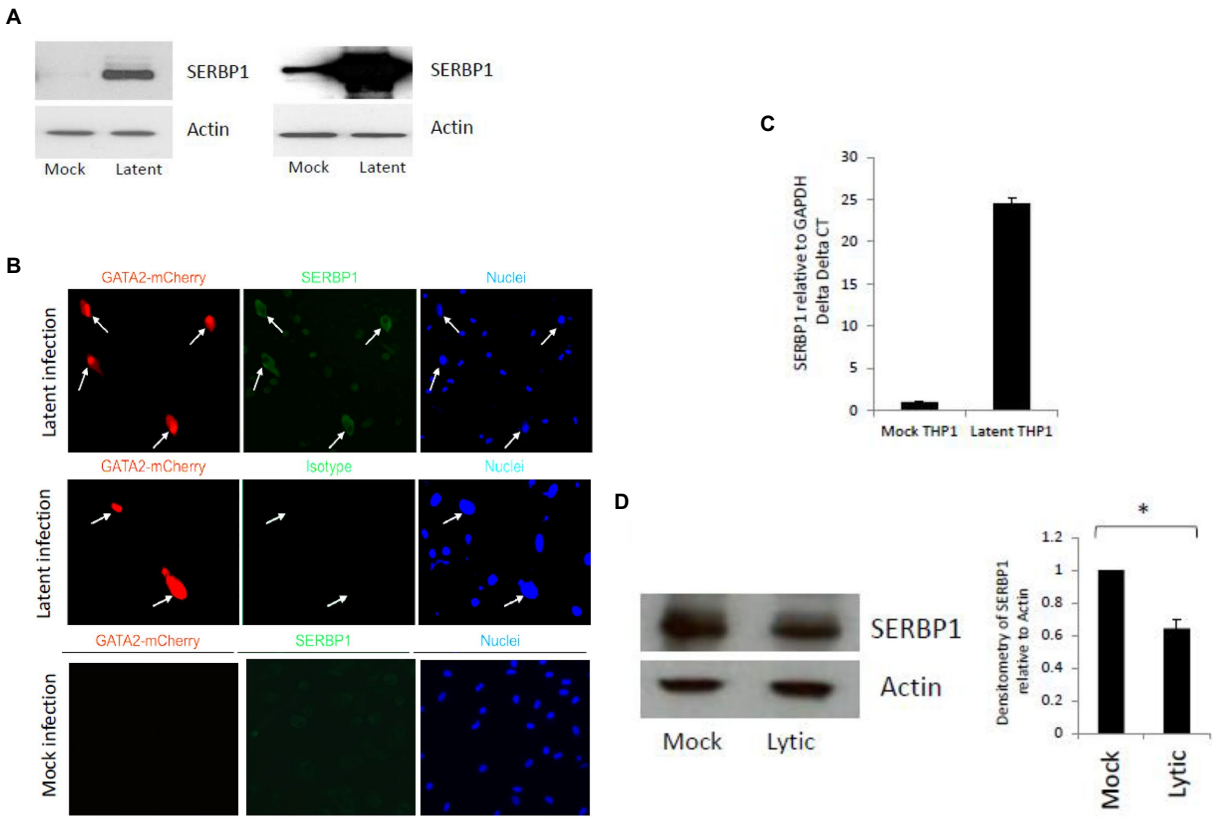

FIGURE 1
Validation of SERBP1 upregulation during HCMV latent and lytic infection. CD14+ cells were infected (latent) or uninfected (mock) with TB40E-GATA2mCherry for 6days and then the latently infected cells were sorted by FACS before western blotting and two independent experiments are shown in the left and right hand panels **(A)**. Alternatively, CD14+ primary cells were either mock or latently infected with TB40E GATA2mCherry before staining for SERBP1 using a rabbit SERBP1 primary antibody followed by FITC anti-rabbit secondary antibody (top and bottom panels, SERBP1) or the equivalent isotype control (middle panel, isotype) and counter stained with Hoechst 33342 (nuclei; **B**). Finally, the cells were harvested for RNA analysis and RTqPCR carried out for GAPDH and SERBP1 **(C)**. These results represent triplicate technical repeats from two different donors. Additionally, HFFF cells were infected for 72h before harvesting for western blot analysis of GAPDH and SERBP1 and densitometry carried out **(D)**. * = *p* value of <0.01.

downregulated during lytic HCMV infection in fibroblasts. These data are consistent with the previously published screen of proteomic changes during HCMV lytic infection (Weekes et al., 2014). Consequently, the upregulation of SERBP1 in latently infected cells was robust and validated using two different analyses and two different recombinant HCMVs and this contrasts the observations for lytic infection.

SERBP1 is required for IE repression during latency

To analyze the significance of this upregulation of SERBP1 during HCMV latency, we generated SERBP1 knock-down (KD) THP1 cells (Supplementary Figure 2) using shRNA technology cells; THP1 cells are a myelomonocytic cell line that has been routinely used as a model of HCMV latent (Poole et al., 2021). These cells were validated for the levels of SERBP1 gene expression which showed an 87% knockdown in these cells as shown by western blotting and densitometry (Supplementary Figures 2A,B). These cells were then infected with recombinant TB40E HCMV carrying an IE2YFP gene cassette which marks lytically infected cells, but not latently infected cells, with YFP (Straschewski et al., 2010). Figure 2A shows that, as expected, undifferentiated wild type (WT) THP1 cells or undifferentiated control THP1 cells in which Beta-2 microglobulin (B2M) was targeted using shRNA technology (removal of B2M has previously been shown to have no impact on the ability of undifferentiated cell types to establish HCMV latency (Poole et al., 2021)) do not express IE2 unless they are differentiated to a permissive macrophage-like phenotype. In contrast, undifferentiated SERBP1 KD THP1 cells failed to repress IE gene expression and express high levels of IE2YFP. All cells tested (WT, B2M or SERBP1 KD cells) express IE after differentiation, as expected. While expression of IE clearly occurred in the SERBP1 KD cells these cells did not progress through full lytic cycle as no virus production was observed upon their co-culture with indicator fibroblasts (Figure 2B, left panel). However, if these cells were differentiated, virus production was observed, as expected (Figure 2B, right panel). Consistent with this, WT THP1 cells infected with HCMV showed low levels of IE

expression concomitant with good levels of UL138 expression (consistent with a latent infection (Poole et al., 2014)) whereas infected SERBP1 KD THP1 cells expressed high levels of IE exon2/3 RNA (Figures 2C,D). Taken together, these data shows that SERBP1 is required for the repression of IE gene expression in latently infected THP1 cells.

SERBP1 mediates recruitment of both Kap1 and CHD3 to the HCMV genome during latency

Since SERBP1 is known to bind to CHD3, which can associate with Kap1 and the NuRD complex, we tested whether Kap1 and CHD3 were able to associate with the viral genome in the absence of SERBP1. Figure 3A shows that, as expected, there is little difference in total histone association with the viral MIEP on infection of WT THP1 or SERBP1 KD THP1 cells when using an antibody for histone H3 ("histone"). Similarly, as expected, in infected WT THP1 cells the viral MIEP is associated with low levels of activatory acetylated histone marks (H4K8ac) and high levels of repressive methylated histone marks (H3K9me)—consistent with suppression of the MIEP and a latent infection. In contrast, in the absence of SERBP1, there is a profound increase in acetylated histone marks (H4K8ac) and a decrease in methylated histone marks (H3K9me) consistent with MIEP transcriptional activity. Importantly, CHD3 can only be found associated with the viral MIEP in the presence of SERBP1—removal of SERBP1 prevents CHD3 association with the promoter. In contrast, PMA induced reactivation of both infected WT and SERBP2 KD cells results in association of the MIEP with activatory histone marks (H4K8ac), and a reduction of repressive histone marks (H3K9me), as expected, with no association with CHD3 (Figure 3B). These data suggest that CHD3 association with the MIEP is required for efficient MIEP repression during latency and that removal of SERBP1 prevents association of CHD3 to the MIEP.

As KAP1 binding to the viral genome had been shown to be involved in MIEP repression during latent infection of primary myeloid cells (Rauwel et al., 2015), and because Kap1-associated protein CHD3 is known to bind SERBP1 (Lemos et al., 2003), we assessed the binding of KAP1 to the viral genome in latently infected THP1 cells in the presence or absence of SERBP1. Figure 3C shows that in the absence of SERBP1, Kap1 no longer associates with the HCMV genomes in infected THP1 cells in these cells in which IE gene expression but not full virus reactivation has occurred. These data suggest that SERBP1 acts as a recruitment factor to facilitate Kap1 association with the viral genome during latent infection.

Discussion

The regulation of the MIEP during latent and lytic infection is complex and controlled by multiple mechanisms involving a plethora of cellular and viral factors. It is clear, though, that during latent infection of undifferentiated myeloid cells the HCMV MIEP is associated with repressive histone marks which are replaced by activatory histone marks as the cell differentiates and reactivation of major IE gene expression is initiated. To date, cellular factors involved in MIEP suppression during latency have identified specific cellular transcription factors that bind to known transcription factor binding sites in the MIEP (Bain et al., 2003; Martinez et al., 2014; Elder et al., 2021; Poole et al., 2021). In this paper, we have identified SERBP1 as an additional factor involved in control of the MIEP during latency.

The data presented do not exclude the possibility that there are other mechanisms by which SERBP1 functions to regulate the MIEP, for example, it may be that these effects are mediated by the known RNA binding functions of SERBP1 which occur in the cytoplasm. Our data show that SERBP1 is located in the cytoplasm and so it is quite likely that it plays a role which perhaps affects CHD3 and KAP1. Additionally, SERBP1 is known to play a role in the regulation of methionine levels which may also be important for generating a repressive chromatin structure (Kosti et al., 2020) and these possibilities need to be investigated in future work. However, given the observed effects on CHD3 and KAP1 shown in this paper, it is does suggest that SERBP1 could act as a "scaffold protein" to recruit repressive accessory factors CHD3 (Lemos et al., 2003) and Kap1 to help to prevent activation of IE gene expression during HCMV latency. A potential model is outlined in Figure 3D where the histone-associating proteins CHD3 (Tencer et al., 2017) and HDAC are shown to be associated with nucleosomes. SERBP1 could mediate CHD3 and, in turn Kap1, association with the viral genome. These proteins being associated with the genome would support a repressive environment because Kap1 mediates methylation of histones *via* SETDB1 and HDACs lead to histone deacetylation.

While we have not addressed the trigger for SERBP1 upregulation during latency in this manuscript, it is interesting to note that the transcription factor YY1 can drive the expression of SMHG8 which, in turn, can lead to an upregulation of SERBP1 (Shan et al., 2022). Since it has been previously published that YY1 is critical for the maintenance of HCMV latency (Poole et al., 2021) it is possible that this contributes to the upregulation of SERBP1 during latency. SERBP1 is known to shuttle between the cytoplasm and the nucleus (Lee et al., 2012, 2014) and certainly in lytic infection has been observed to be located in both locations (Weekes et al., 2014). Our data suggests that SERBP1 is predominatly cytoplasmic and it is possible that SERBP1 is mediating effects *via* a cytoplasmic route and this is, as stated above, is something that should be investigated in future studies. However, given the effects on CHD3 and KAP1, our observations here are consistent with it acting as a scaffold protein to recruit repressive complexes, involving CHD3 and KAP1, to the viral genome in order to help maintain the viral genome in a repressive chromatin structure thereby aiding maintenance of latency. Interestingly, as well as HCMV, KAP1 has also been shown to be required for KSHV latency where the KSHV LANA gene

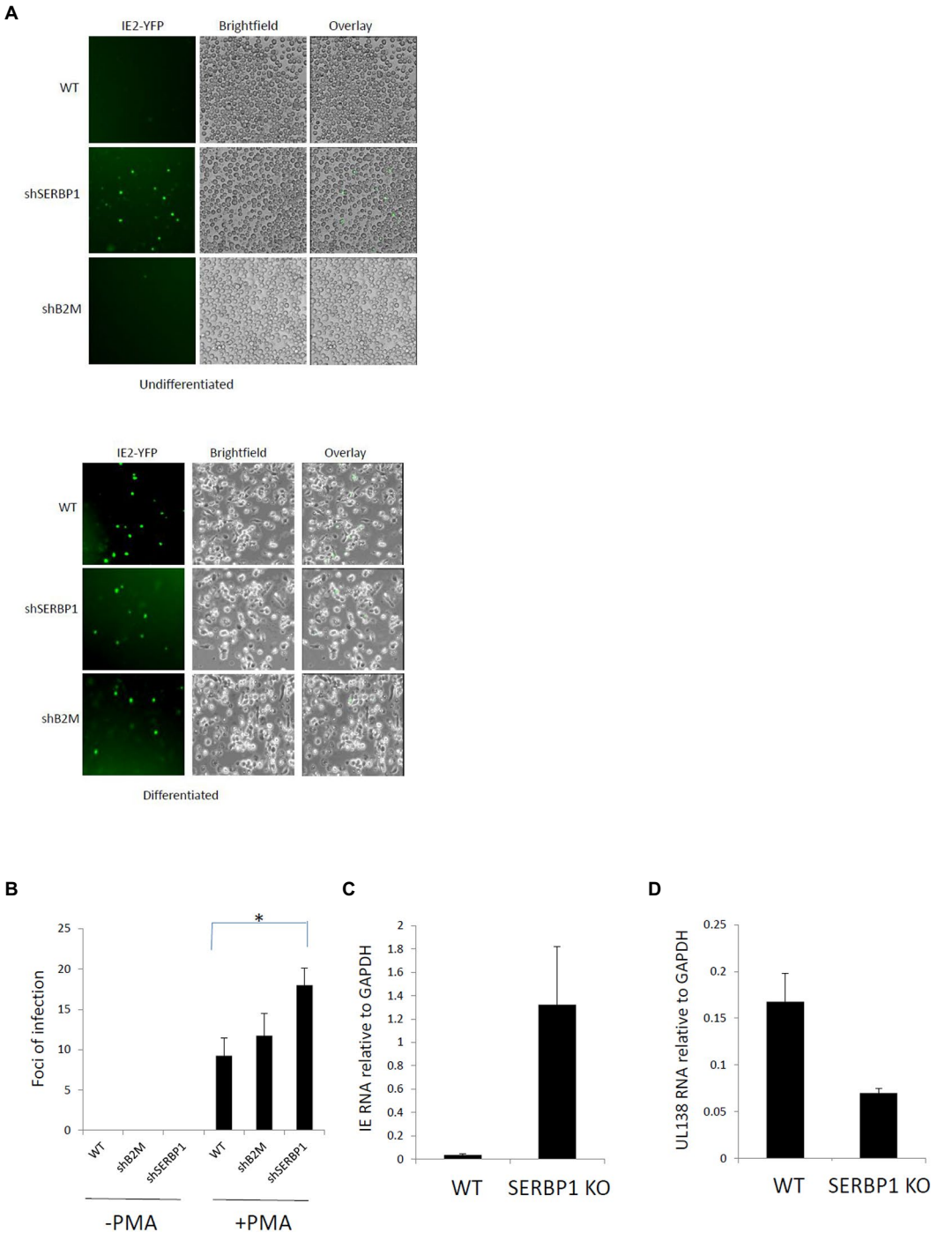

FIGURE 2
SERBP1 is required for repression of the MIEP during latency. WT, shB2M or shSERBP1 THP1 cells were infected with TB40E-IE2YFP for 4days and then either reactivated with PMA or not before either analyzing by microscopy directly **(A)** or supernatants transferred to fibroblasts and foci of infection counted to detect productive virus reactivation **(B)** or harvesting cells for viral RNAs IE exon2/3 **(C)** and UL138 **(D)**. Data represent triplicate technical data from two independent experiments with standard deviation shown and the student's *t*-test used for statistical analysis (* represents a *p*-value of <0.01).

product has been proposed to mediate recruitment of Kap1 to the KSHV genome (Sun et al., 2014). Similarly, depletion of CHD3 in during Herpes Simplex Virus (HSV-1) infection breaks latency in an HSV-1 latency model (Arbuckle and Kristie, 2014). It would be of interest to determine whether these other herpesviruses also use SERBP1 as a scaffold to recruit both KAP1 and CHD3 to their genomes to initiate and maintain a repressive chromatin structure during latent infection.

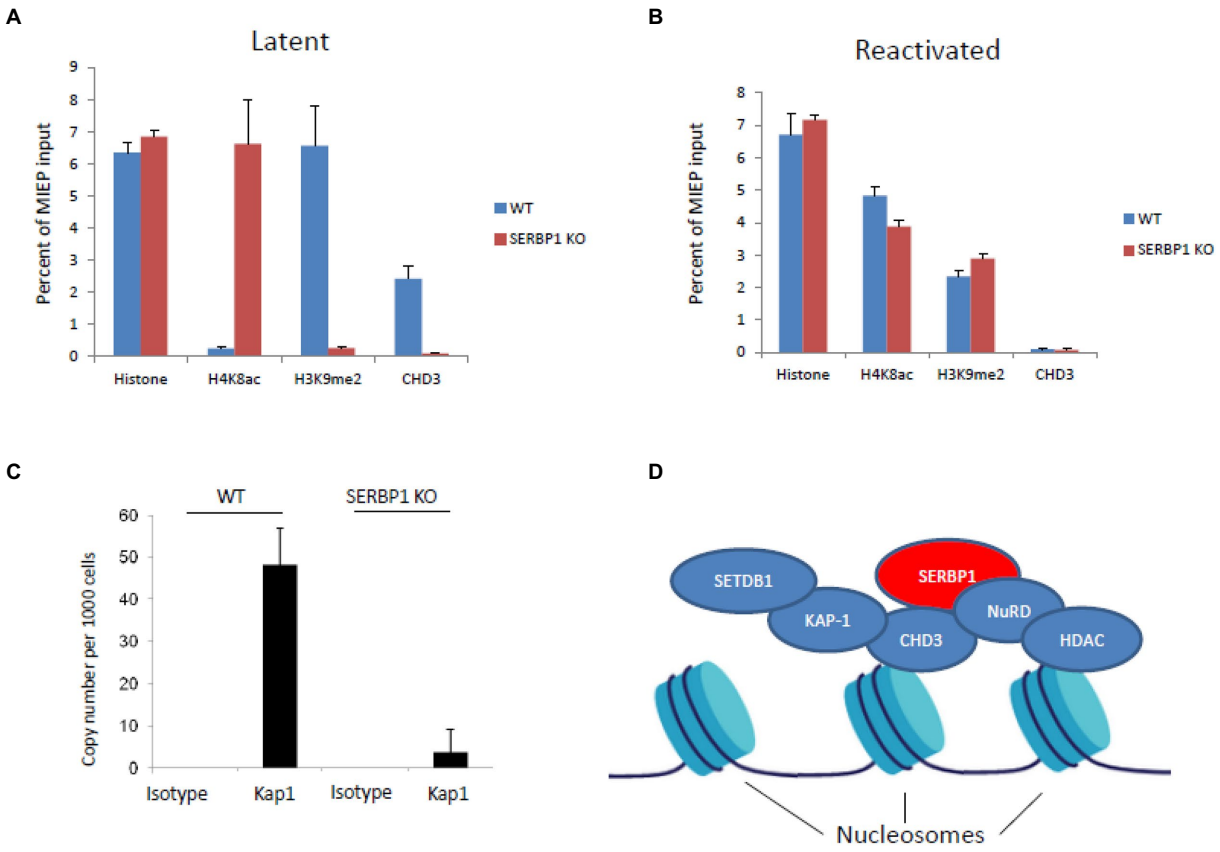

FIGURE 3
SERBP1 can recruit repressive proteins CHD3 and KAP1 to the viral genome. WT or shSERBP1 THP1 cells were either infected or left uninfected with TB40E-IE2YFP for 4days before harvesting for ChIP analysis for histone H3 ("histone") or activated ("H4K8ac")or repressive ("H3K9me2") histone markers or CHD3 during latency (latent, **A**) or reactivation (reactivated, **B**). Graphs represent standard deviation about the mean from two independent experiments with triplicate technical repeats. WT or shSERBP1 THP1 cells were infected with TB40E for 4days before fixing in 1% PFA and then harvesting for ChIP analysis. DNA was then analyzed by ChIP for the presence of KAP1 and the numbers of genomes were quantified by droplet digital PCR using a gB probe **(C)**. Potential schematic shows how SERBP1 may act as a scaffold to form a repressive complex *via* CHD3 and HDAC (which are known to associate with histone proteins) to recruit KAP1 where SETDB1 can mediate methylation and the NuRD complex with HDAC proteins can mediate histone deacetylation **(D)**.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

All research describing studies on primary human material with HCMV were assessed and approved by the Cambridge Local Research Ethics committee. Informed consent was received from blood donors with the Cambridge Local Research Ethics committee and the Cambridge Internal Review Board. Cells were harvested from healthy adult donors, and the decision to use tissue was not affected by gender and age, as this was not important to the studies performed. The patients/participants provided their written informed consent to participate in this study.

Author contributions

EP performed experiments and wrote the manuscript. JS edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the British Medical Research Council, grant number MR/S00081X/1/MRC.

Acknowledgments

We thank Paula Rayner and Linda Teague for technical assistance and the Cambridge University NIHR BRC for cell sorting analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MN declared a past collaboration with one of the authors EP to the handling editor.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.999290/full#supplementary-material

SUPPLEMENTARY FIGURE 1
SERBP1 is upregulated during HCMV latency. Primary CD14 monocytes were infected with TB40E-SV40GFP. The green cells were then sorted by FACS before full proteomic screen analysis as previously described (Aslam et al., 2019). The proteins identified with the highest number of peptides are shown from two independent screens (A). The data for all the hits are plotted on a graph (B) where each + represents a protein and all proteins are represented along the x axis in alphabetical order (proteins).

SUPPLEMENTARY FIGURE 2
Validation of SERBP1 knock down cells. THP1 shSERBP1 cells were western blotted for SERBP1 alongside WT THP1 cells **(A)** and the levels of knock down determined by densitometry with standard deviations shown and the student's T test statistical analysis to determine the statistical significance (*** represents a P value >0.0001) **(B)**.

References

Arbuckle, J. H., and Kristie, T. M. (2014). Epigenetic repression of herpes simplex virus infection by the nucleosome remodeler CHD3. *MBio* 5, e01027–e01013. doi: 10.1128/mBio.01027-13

Aslam, Y., Williamson, J., Romashova, V., Elder, E., Krishna, B., Wills, M., et al. (2019). Human cytomegalovirus Upregulates expression of HCLS1 resulting in increased cell motility and Transendothelial migration during latency. *iScience* 20, 60–72. doi: 10.1016/j.isci.2019.09.016

Bain, M., Mendelson, M., and Sinclair, J. (2003). Ets-2 repressor factor (ERF) mediates repression of the human cytomegalovirus major immediate-early promoter in undifferentiated non-permissive cells. *J. Gen. Virol.* 84, 41–49. doi: 10.1099/vir.0.18633-0

Elder, E. G., Krishna, B. A., Poole, E., Perera, M., and Sinclair, J. (2021). Regulation of host and viral promoters during human cytomegalovirus latency via US28 and CTCF. *J. Gen. Virol.* 102:001609. doi: 10.1099/jgv.0.001609

Elder, E., Krishna, B., Williamson, J., Aslam, Y., Farahi, N., Wood, A., et al. (2019). Monocytes latently infected with human cytomegalovirus evade neutrophil killing. *iScience*. 12, 13–26. doi: 10.1016/j.isci.2019.01.007

Goodarzi, A. A., Kurka, T., and Jeggo, P. A. (2011). KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nat. Struct. Mol. Biol.* 18, 831–839. doi: 10.1038/nsmb.2077

Hoffmeister, H., Fuchs, A., Erdel, F., Pinz, S., Grobner-Ferreira, R., Bruckmann, A., et al. (2017). CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality. *Nucleic Acids Res.* 45, 10534–10554. doi: 10.1093/nar/gkx711

Kosti, A., de Araujo, P. R., Li, W. Q., Guardia, G. D. A., Chiou, J., Yi, C., et al. (2020). The RNA-binding protein SERBP1 functions as a novel oncogenic factor in glioblastoma by bridging cancer metabolism and epigenetic regulation. *Genome Biol.* 21:195. doi: 10.1186/s13059-020-02115-y

Lee, Y. J., Hsieh, W. Y., Chen, L. Y., and Li, C. (2012). Protein arginine methylation of SERBP1 by protein arginine methyltransferase 1 affects cytoplasmic/nuclear distribution. *J. Cell. Biochem.* 113, 2721–2728. doi: 10.1002/jcb.24151

Lee, Y. J., Wei, H. M., Chen, L. Y., and Li, C. (2014). Localization of SERBP1 in stress granules and nucleoli. *FEBS J.* 281, 352–364. doi: 10.1111/febs.12606

Lemos, T. A., Passos, D. O., Nery, F. C., and Kobarg, J. (2003). Characterization of a new family of proteins that interact with the C-terminal region of the chromatin-remodeling factor CHD-3. *FEBS Lett.* 533, 14–20. doi: 10.1016/S0014-5793(02)03737-7

Martinez, F. P., Cruz, R., Lu, F., Plasschaert, R., Deng, Z., Rivera-Molina, Y. A., et al. (2014). CTCF binding to the first intron of the major immediate early (MIE) gene of human cytomegalovirus (HCMV) negatively regulates MIE gene expression and HCMV replication. *J. Virol.* 88, 7389–7401. doi: 10.1128/JVI.00845-14

Nielsen, A. L., Ortiz, J. A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., et al. (1999). Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J.* 18, 6385–6395. doi: 10.1093/emboj/18.22.6385

Poole, E., Carlan da Silva, M. C., Huang, C., Perera, M., Jackson, S., Groves, I. J., et al. (2021). A BMPR2/YY1 signaling Axis is required for human cytomegalovirus latency in undifferentiated myeloid cells. *MBio* 12:e0022721. doi: 10.1128/mBio.00227-21

Poole, E., Groves, I., Jackson, S., Wills, M., and Sinclair, J. (2021). Using primary human cells to analyze human cytomegalovirus biology. *Methods Mol. Biol.* 2244, 51–81. doi: 10.1007/978-1-0716-1111-1_4

Poole, E., Huang, C. J. Z., Forbester, J., Shnayder, M., Nachshon, A., Kweider, B., et al. (2019). An iPSC-derived myeloid lineage model of herpes virus latency and reactivation. *Front. Microbiol.* 10:2233. doi: 10.3389/fmicb.2019.02233

Poole, E., Wills, M., and Sinclair, J. (2014). Human cytomegalovirus latency: targeting differences in the latently infected cell with a view to clearing latent infection. *New J. Sci.* 2014, 1–10. doi: 10.1155/2014/313761

Rauwel, B., Jang, S. M., Cassano, M., Kapopoulou, A., Barde, I., and Trono, D. (2015). Release of human cytomegalovirus from latency by a KAP1/TRIM28 phosphorylation switch. *Elife* 4. doi: 10.7554/eLife.06068

Ryan, R. F., Schultz, D. C., Ayyanathan, K., Singh, P. B., Friedman, J. R., Fredericks, W. J., et al. (1999). KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* 19, 4366–4378. doi: 10.1128/MCB.19.6.4366

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., and Rauscher, F. J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* 16, 919–932. doi: 10.1101/gad.973302

Schultz, D. C., Friedman, J. R., and Rauscher, F. J. 3rd. (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the mi-2alpha subunit of NuRD. *Genes Dev.* 15, 428–443. doi: 10.1101/gad.869501

Shan, B., Qu, S., Lv, S., Fan, D., and Wang, S. (2022). YY1-induced long non-coding RNA small nucleolar RNA host gene 8 promotes the tumorigenesis of melanoma via the microRNA-656-3p/SERPINE1 mRNA binding protein 1 axis. *Bioengineered* 13, 4832–4843. doi: 10.1080/21655979.2022.2034586

Sinclair, J., and Poole, E. (2014). Human cytomegalovirus latency and reactivation in and beyond the myeloid lineage. *Futur. Virol.* 9, 557–563. doi: 10.2217/fvl.14.34

Straschewski, S., Warmer, M., Frascaroli, G., Hohenberg, H., Mertens, T., and Winkler, M. (2010). Human cytomegaloviruses expressing yellow fluorescent fusion proteins--characterization and use in antiviral screening. *PLoS One* 5:e9174. doi: 10.1371/journal.pone.0009174

Sun, R., Liang, D., Gao, Y., and Lan, K. (2014). Kaposi's sarcoma-associated herpesvirus-encoded LANA interacts with host KAP1 to facilitate establishment of viral latency. *J. Virol.* 88, 7331–7344. doi: 10.1128/JVI.00596-14

Tencer, A. H., Cox, K. L., Di, L., Bridgers, J. B., Lyu, J., Wang, X., et al. (2017). Covalent modifications of histone H3K9 promote binding of CHD3. *Cell Rep.* 21, 455–466. doi: 10.1016/j.celrep.2017.09.054

Terhune, S. S., Moorman, N. J., Cristea, I. M., Savaryn, J. P., Cuevas-Bennett, C., Rout, M. P., et al. (2010). Human cytomegalovirus UL29/28 protein interacts with components of the NuRD complex which promote accumulation of immediate-early RNA. *PLoS Pathog.* 6:e1000965. doi: 10.1371/journal.ppat.1000965

Weekes, M. P., Tomasec, P., Huttlin, E. L., Fielding, C. A., Nusinow, D., Stanton, R. J., et al. (2014). Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* 157, 1460–1472. doi: 10.1016/j.cell.2014.04.028

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 02 December 2022
DOI 10.3389/fmicb.2022.1052324

OPEN ACCESS

EDITED BY
Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands

REVIEWED BY
Kemin Li,
Sichuan University,
China
Nikolaos Tsetsos,
424 General Military Hospital, Greece

*CORRESPONDENCE
Zhifang Li
lzfmuzi@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

RECEIVED 23 September 2022
ACCEPTED 16 November 2022
PUBLISHED 02 December 2022

CITATION
Ren X, Hao Y, Wu B, Jia X, Niu M, Wang K and Li Z (2022) Efficacy of prophylactic human papillomavirus vaccines on cervical cancer among the Asian population: A meta-analysis. *Front. Microbiol.* 13:1052324. doi: 10.3389/fmicb.2022.1052324

Efficacy of prophylactic human papillomavirus vaccines on cervical cancer among the Asian population: A meta-analysis

Xinyu Ren[1†], Yubing Hao[1†], Beike Wu[1], Xinhua Jia[2,3], Meili Niu[1], Kunbo Wang[4] and Zhifang Li[5*]

[1]School of Public Health, Shanxi Medical University, Taiyuan, China, [2]Department of Epidemiology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3]The State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, School of Public Health, National Institute of Diagnostics and Vaccine Development in Infectious Diseases, Xiamen University, Xiamen, China, [4]Xiangya School of Public Health, Central South University, Changsha, China, [5]Department of Preventive Medicine, Changzhi Medical College, Changzhi, China

Objective: We conducted a meta-analysis to assess the efficacy of prophylactic human papillomavirus (HPV) vaccines against cervical cancer precursors and HPV persistent infection among Asian populations.

Methods: Randomized controlled clinical trials conducted in Asian countries were identified from three electronic databases (PubMed, EMBASE and the Cochrane Library). Publication retrieval was performed on September 1, 2022 and only those written in English were included. The data were analyzed with Cochrane Review Manager (version 5.3) and Stata/SE (15.1). Effect sizes were presented as risk ratios (RRs) and 95% confidence intervals (CIs).

Results: Ten articles were considered in the meta-analysis, without significant heterogeneity among them. The fixed-effect RRs and 95% CIs for cervical intraepithelial neoplasia grade 1 (CIN1+) and CIN2+ were 0.10 (0.05–0.21) and 0.11 (0.04–0.27), respectively. Positive effect of HPV vaccination on 6- and 12-month persistent infection were observed, with the respective pooled RRs of 0.05 (95% CI: 0.03–0.09) and 0.09 (95% CI: 0.05–0.15). HPV vaccination has a positive effect on the incidence of cytological abnormalities associated with HPV 16/18 (RR, 0.13; 95% CI (0.09–0.20)). Positive effects of HPV vaccination were also observed for HPV 16- and 18-specific immunogenicity (RR, 235.02; 95% CI (82.77–667.31) and RR, 98.24; 95% CI (50.36–191.67), respectively). Females receiving an initial vaccination showed significant decreased incidences of cervical intraepithelial neoplasia, HPV persistent infection and cytological abnormalities and a significantly higher antibody positive conversion rate compared with non-vaccination counterparts.

Conclusion: Prophylactic HPV vaccines are highly efficacious in preventing cervical cancer in Asian females. The government should accelerate the processes of vaccine introduction and vaccination implementation by prioritizing them in public health policies, which should be helpful to enhance Asian females' awareness of receiving HPV vaccination volitionally.

KEYWORDS
HPV vaccine, efficacy, Asia, cervical cancer, review manage

Introduction

Cervical cancer is a common malignant tumor among women in the world. According to the global cancer burden report 2020 released by the international agency for research on cancer of the World Health Organization, the number of new cases of cervical cancer in the world is as high as 604,127. In 2020, about 341,831 people died of this tumor, accounting for approximately 7.7% of all deaths caused by gynecological cancers (Sung et al., 2021). Currently, with the intensive implementation of cervical cancer prevention programs, the incidence of cervical cancer in developed countries such as US has decreased; however, in many low- and middle-income countries, the rate remains unchanged, or even shows a rising tendency. Even worse, globally, particularly in developing countries, the burden caused by cervical cancer may be greater than currently reported, considering that patients in rural areas often have no access to health care and therefore elude being reported (LaVigne et al., 2017).

Human papillomavirus (HPV) persistent infection is the main risk factor for cervical cancer (Hamborsky et al., 2015) and oropharyngeal cancer et al. malignant tumors (Tsentemeidou et al., 2021). The position paper of the World Health Organization (WHO) points out equivocally that HPV vaccination can effectively prevent the occurrence of HPV related diseases (World Health Organization, 2017). In 2019, the expert consensus on immune prevention of HPV related diseases such as cervical cancer clearly stated that primary prevention was the focus of cervical cancer prevention and control strategy (Vaccine and Immunization Branch of Chinese Preventive Medicine Association, 2019). In 2018, the WHO set a goal of global elimination of cervical cancer as a public health priblem by 2030, and "eliminate" has a specific definition: fewer than 4 new cases per 100,000 women per year. To achieve the 2030 elimination goal, the organization also proposed multi-stage implementation strategies, as follows: to provide 90% of school-age girls with HPV vaccines before an age of 15, to perform efficient cervical cancer screening for 70% of women aged between 35 years and 45 years, and to provide standardized treatment and management for 90% of women that are diagnosed with cervical cancer or precancerous lesions (World Health Organization, 2020). HPV vaccines (bivalent/tetravalent/9-valent HPV vaccines) have been widely used in men and women of school age to prevent related diseases caused by HPV infection. By 2019, HPV vaccination had been incorporated into national vaccination programs of 98 countries (Vaccines in National Immunization Programme, 2019). In the meantime, in these countries, clinical trials were conducted to uphold the programs as to the implementation of prophylactic HPV vaccination (Paavonen et al., 2009; Eriksson et al., 2013; Bonanni et al., 2015; Drolet et al., 2015). According to these trials, HPV vaccines successfully induce high levels of antiviral antibodies (Wheeler et al., 2008; Malagón et al., 2012; Naud et al., 2014; Schwarz et al., 2014), prevent the infection of HPV types targeted by vaccines (Wheeler et al., 2008; Bonanni et al., 2015), and mitigate the development of premalignant cervical intraepithelial neoplasia (CIN) and cervical cancer (Paavonen et al., 2009; FUTURE I/II Study Group et al., 2010).

To date, in Asia, a few countries have participated in HPV vaccination trials. Previously, a meta-analysis has reported the immunogenicity and safety of HPV vaccination in Asian people, but its efficacy has not been reported. Considering that such analysis is of great significance for Asians to enhance their awareness of receiving HPV vaccination volitionally, it is important to perform a systemic and discrete assessment of HPV vaccine efficacy for the Asian population.

This study investigated the efficacy of HPV vaccines in Asian countries by systematically reviewing available scientific evidence and conducting a meta-analysis of the related randomized controlled trials, with the more important aim to formulate the immunization strategy of HPV vaccination in developing countries in Asian, especially some countries without HPV vaccination or including it in the national immunization plan. Furthermore, the results of this study might provide a theory foundation for the direct introducing and licensing strategy of HPV vaccination without clinical trials in some Asian countries to ensure that more women could be protected as early as possible.

Materials and methods

Databases and search methods

Systematic searches of three electronic databases (PubMed, EMBASE and Cochrane Library) were conducted to identify reports of the randomized controlled clinical trials (RCTs) regarding the effect of HPV vaccination in Asian countries. The combined index terms were as follows: 'Human Papillomavirus' (HPV OR human papillomavirus OR HPV 16 OR HPV 18) AND 'HPV vaccine' AND 'efficacy' AND 'Asia'. This study focused on the efficacy profiles of the vaccination, and only studies conducted in Asia were included. Duplicate articles were excluded, and, additionally, non-RCT studies and those involving women in pregnancy were excluded. Studies involving

subjects vaccinated with therapeutic vaccines were excluded. Repeated cohorts of patients evaluated at different follow-up times were also excluded.

Data collection

All RCTs performed in Asian populations that provided data on the efficacy of HPV vaccination as the outcomes were included. We only included studies that provided the required information for each outcome. Databases released by 1 September 2022 were used, and we only included papers written in English.

Two investigators from our team assessed the studies independently, and any disagreement was discussed and solved with a third investigator. Data as to authors, the country, patient age, gender, funding sources, vaccination schedules, vaccine components, the mode of vaccine distribution, blinding, randomization and follow-up time were extracted from the included articles. The end points of efficacy were the incidence of HPV-16 or−18 associated cervical intraepithelial neoplasia (CIN), cervical cancer, cytological abnormalities and HPV persistent infection. Diseases were diagnosed by the pathology panel, and in the meantime, the HPV DNA type from the same sample was determined. Only studies where the participants had HPV seronegativity at the initial phase were included in seroconversion rate calculation.

The risk of bias of all studies was assessed based on the Cochrane collaboration's tool, which is specialized for assessing the risk of bias of randomized trials (Higgins et al., 2011). This tool consists of seven categories, i.e., random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attention bias), selective reporting (reporting bias) and other bias. We used 'low', 'high' and 'unclear' risk of bias to categorize these included trials. Irrespective of bias risk, all screened and selected eligible studies were included in the current meta-analysis.

Statistical analysis

Data were analyzed by Cochrane Review Manager version 5.3. Effect sizes were summarized as risk ratios (RRs) and the associated 95% confidence intervals (CIs). The RR was calculated based on the number of events, which included CIN, persistent infection and cytological abnormality. An RR value <1 suggested a preventive effect on a certain clinical endpoint. To deal with a possible heterogeneity problem, as a consequence of the differences in the methods and sample characteristics of these studies, we performed a heterogeneity test by assigning an I^2 score based on the Cochrane Q test result (Higgins et al., 2003) (this method presents a quantitative value of heterogeneity ranging from 0 to 100%, and according to the Cochrane recommendation, an I^2 value of 50% and above is considered to have a substantial heterogeneity, Higgins et al. (2022) and under such conditions, sensitivity analysis needs to be performed). When statistical homogeneity among the studies occurred ($p > 0.1$ and/or $I^2 < 50\%$), we used a fixed effects model for the meta-analysis; otherwise, a random effects model was employed. Sensitivity analyses were performed by eliminating one different trial each time, and statistics were recalculated.

Results

Article selection process

Study identification and selection was demonstrated in the flow diagram in Figure 1. From PubMed, EMBASE and the Cochrane Library, 120, 158 and 4 articles were identified, respectively. From these, 11 duplicated articles were removed, and 252 articles were then screened based on the title and abstract, most of which did not meet the inclusion criteria. A total of 19 full-text articles were considered to be eligible. Further identification excluded 9 articles due to the following reasons: not randomized controlled trials; not double-blind experiments; data unable to be extracted; no control group, repeated cohorts of patients evaluated at different follow-up times. Finally, 10 articles (Konno et al., 2010; Li et al., 2012; Konno et al., 2014; Wu et al., 2015; Zhu et al., 2017; Mikamo et al., 2019; Wei et al., 2019; Zhu et al., 2019; Qiao et al., 2020; Zhao et al., 2022) were introduced into the meta-analysis (Figure 1).

Study characteristics

The included studies are summarized in Table 1. Clinical trials of HPV vaccination in Asia were performed in two different countries, China and Japan. The age of the involved participants varied considerably, ranging from 9 years to 45 years. Among the 10 trials, the bivalent vaccine from GSK and Cecolin (containing HPV types 16 and 18) was used in seven trials(Konno et al., 2010, 2014; Wu et al., 2015; Zhu et al., 2017, 2019; Qiao et al., 2020; Zhao et al., 2022) and the quadrivalent vaccine from Merck (containing HPV types 6, 11, 16 and 18) in three trials (Li et al., 2012; Mikamo et al., 2019; Wei et al., 2019). All studies ($N = 10$) were subject to a blind and randomized control design. The majority of the studies ($N = 8$) only included women (Konno et al., 2010, 2014; Wu et al., 2015; Zhu et al., 2017, 2019; Wei et al., 2019; Qiao et al., 2020; Zhao et al., 2022).The follow-up times of these studies ranged from 15 to 90 months. Five studies included a placebo as the comparator (Li et al., 2012; Zhu et al., 2017, 2019; Mikamo et al., 2019; Wei et al., 2019), and five studies on the bivalent vaccine used the hepatitis virus vaccine as the comparator (Konno et al.,

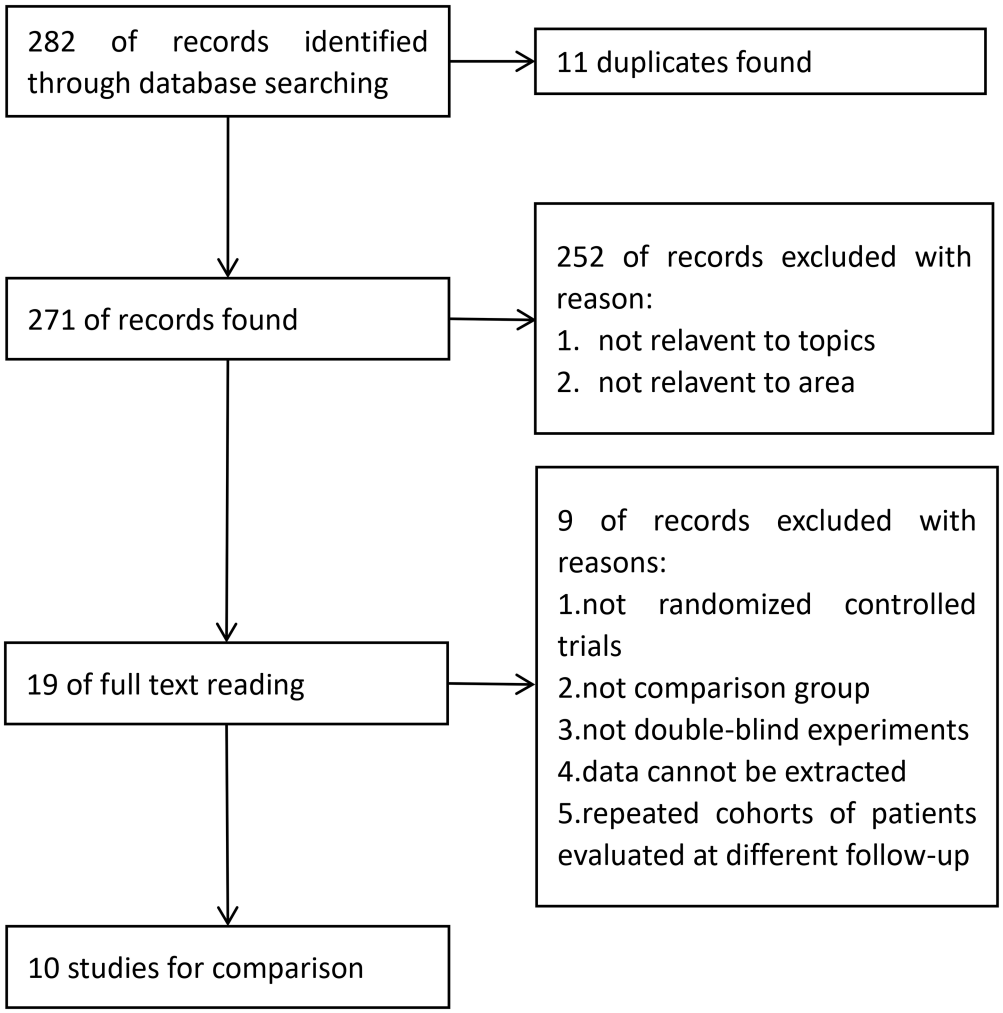

FIGURE 1
Flow diagram of trial selection in this study.

2010, 2014; Wu et al., 2015; Qiao et al., 2020; Zhao et al., 2022). Studies with the bivalent vaccine had implemented administration schedules of 0, 1 and 6 months (Konno et al., 2010, 2014; Wu et al., 2015; Zhu et al., 2017, 2019; Qiao et al., 2020; Zhao et al., 2022), and those with the quadrivalent vaccine of 0, 2 and 6 months (Li et al., 2012; Mikamo et al., 2019; Wei et al., 2019).

Assessment of the risk of bias of the included studies

Although all studies claimed that exact randomized controlled procedures were performed, only seven studies specified how the random sequences were generated (Wu et al., 2015; Zhu et al., 2017, 2019; Mikamo et al., 2019; Wei et al., 2019; Qiao et al., 2020; Zhao et al., 2022). Furthermore, only five studies explained in detail how the process of allocating each participant into the vaccinated or control group was blinded (Wu et al., 2015; Zhu et al., 2017, 2019; Qiao et al., 2020; Zhao et al., 2022). Consequently, most studies ($N = 5$) failed in explaining how participants and researchers were blinded(Konno et al., 2010, 2014; Li et al., 2012; Wu et al., 2015; Wei et al., 2019) or how the outcome assessment process was blinded ($N = 6$) (Konno et al., 2010, 2014; Li et al., 2012; Wu et al., 2015; Wei et al., 2019; Qiao et al., 2020). One study presented incomplete outcomes (Wei et al., 2019; Figure 2).

Outcomes

HPV vaccines and the incidences of CIN1+ and CIN2+

A preventive effect of HPV vaccination on the incidence of CIN1+ (4 RCTs: 15,717 participants; Figure 3) was observed, with

TABLE 1 Descriptive characteristics of the studies included in the review.

Authors	Country	Funding source	Gender	Age in years	Vaccination schedule(s)	Follow-up time (months)	Vaccine component	Comparator	Mode of vaccine Distribution
Zhu et al. (2017)	China	GSK	Females	18–25y	3d(M0, 1, 6)	24	HPV16/18	Placebo	Clinical trial
Zhu et al. (2019)	China	GSK	Females	18–25y	3d(M0, 1, 6)	72	HPV16/18	Placebo	Clinical trial
Konno et al. (2010)	Japan	GSK	Females	20-25y	3d(M0, 1, 6)	48	HPV16/18	Hep A vaccine	Clinical trial
Konno et al. (2014)	Japan	GSK	Females	20-25y	3d(M0, 1, 6)	24	HPV16/18	Hep A vaccine	Clinical trial
Wei et al. (2019)	China	Merck & Co	Females	20–45y	3d(M0, 2, 6)	78	HPV6/11/16/18	Placebo	Clinical trial
Zhao et al. (2022)	China	Cecolin	Females	18–45y	3d(M0, 1, 6)	66	HPV16/18	Hep E vaccine	Clinical trial
Li et al. (2012)	China	Merck & Co	Females/ males	9-45y/9-15y	3d(M0, 2, 6)	90	HPV6/11/16/18	Placebo	Clinical trial
Wu et al. (2015)	China	GSK	Females	18–25y	3d(M0, 1, 6)	48	HPV16/18	Hep B vaccine	Clinical trial
Qiao et al. (2020)	China	Cecolin	Females	18–45y	3d(M0, 1, 6)	40	HPV16/18	Hep E vaccine	Clinical trial
Mikamo et al. (2019)	Japan	Merck & Co	Males	16-27y	3d(M0, 2, 6)	36	HPV6/11/16/18	placebo	Clinical trial

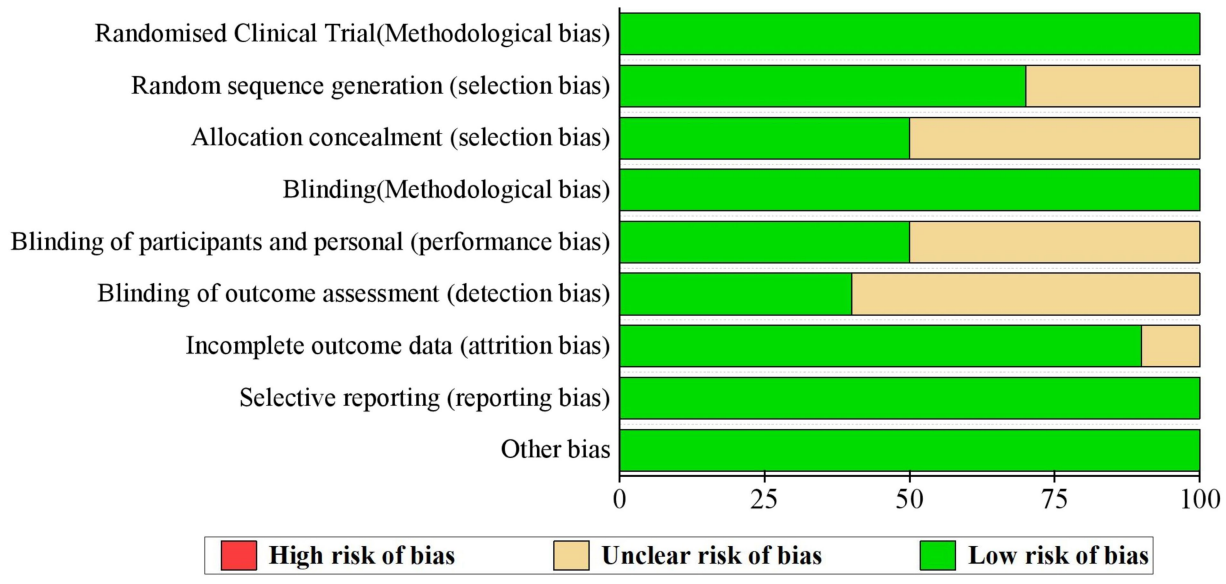

FIGURE 2
Summary of the risk of bias of the 10 studies.

a pooled RR of 0.08 (95% CI, 0.03–0.22); no significant heterogeneity was observed among the involved studies ($I^2 = 0\%$; $p = 0.70$).

As was expected, the incidence of CIN2+ (4 RCTs: 15,403 participants; Figure 4) also exhibited a statistically significant decrease after vaccination, with a pooled RR of 0.09 (95% CI:

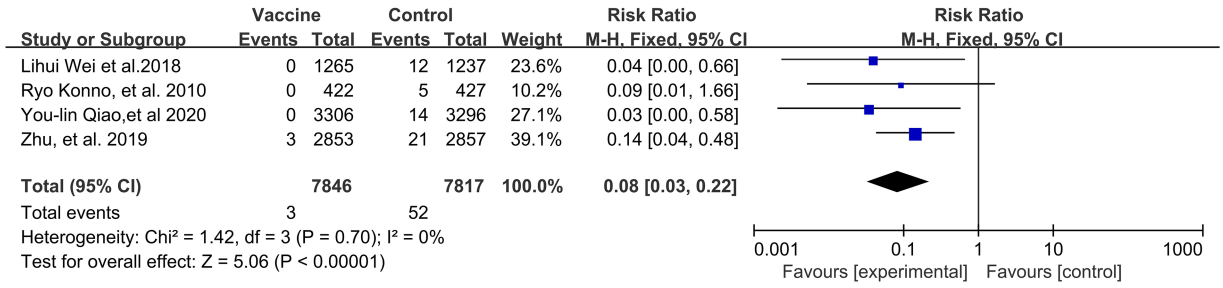

FIGURE 3
A forest plot of the meta-analysis of the incidence of HPV 16/18 associated CIN1+ after HPV vaccination.

FIGURE 4
A forest plot of the meta-analysis of the incidence of HPV 16/18 associated CIN2+ after HPV vaccination.

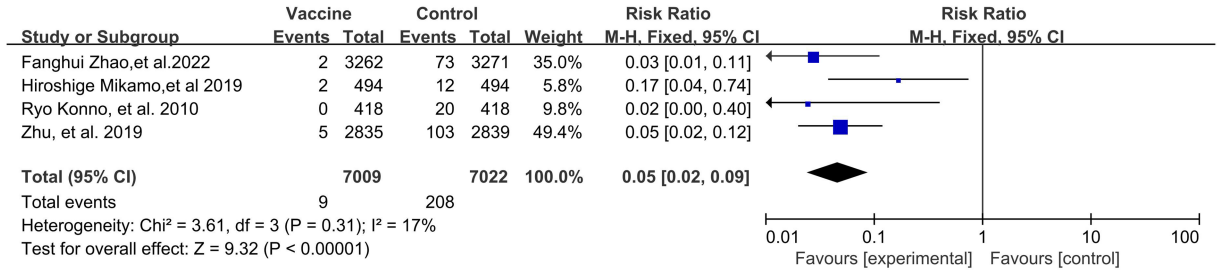

FIGURE 5
A forest plot of the meta-analysis of the incidence of HPV 16/18 associated 6-month persistent infection after HPV vaccination.

0.03–0.28). No significant heterogeneity was observed among the involved studies ($I^2 = 0\%$; $p = 0.77$).

HPV vaccines and the incidences of 6- and 12-month persistent infection

Preventive effects of HPV vaccination on the incidences of HPV 16/18 associated 6- (4 RCTs: 14,031 participants; Figure 5) and 12-month PI (3 RCTs: 6,783 participants; Figure 6) were observed, with the pooled RRs of 0.05 (95% CI: 0.02–0.09) and 0.08 (95% CI: 0.04–0.17), respectively. The respective heterogeneity test results were ($I^2 = 17\%$, $p = 0.31$) and ($I^2 = 0\%$, $p = 0.48$).

HPV vaccines and the incidence of cytological abnormalities (ASC-US+)

A preventive effect of HPV vaccination was observed on the incidence of ASC-US+ (3 RCTs: 9,085 participants; Figure 7), with a pooled RR of 0.14 (95% CI: 0.09–0.22). There was no significant heterogeneity among the studies ($I^2 = 0\%$; $p = 0.69$).

FIGURE 6
A forest plot of the meta-analysis of the incidence of HPV 16/18 associated 12-month persistent infection after HPV vaccination.

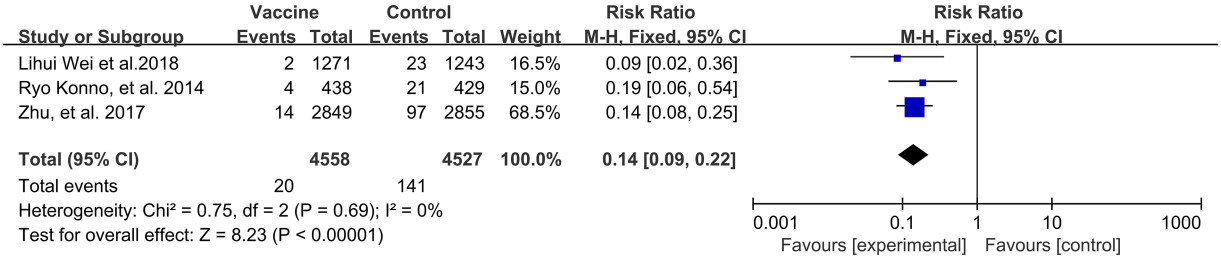

FIGURE 7
A forest plot of the meta-analysis of the incidence of HPV 16/18 associated cytological abnormality after HPV vaccination.

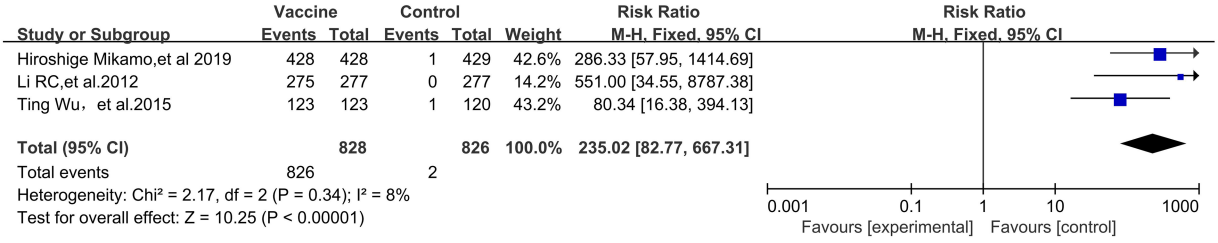

FIGURE 8
Comparison of the human palillomavirus type 16 (HPV 16) specific antibody conversion rate between the vaccinated group and the non-vaccination group in Asian populations.

HPV vaccines and immunogenicity

A preventive effect of HPV vaccination on HPV 16-specific immunogenicity (3 RCTs; 1,654 participants; Figure 8) was observed, with a pooled RR of 235.02 (95% CI: 82.77–667.31). No significant heterogeneity was observed among the studies (I^2 of 8%, $p = 0.34$).

HPV vaccination also showed a favorable effect on HPV 18-specific immunogenicity (RR, 98.24; 95% CI, 50.36–191.67) (3 RCTs: 1717 participants; Figure 9). No significant heterogeneity was observed among the involved studies ($I^2 = 0\%$, $p = 0.75$).

Sensitivity analyses

Sensitivity analyses were performed by eliminating one different trial each time, and statistics were recalculated. The results showed that after literature removal, no such statistical differences in the RR value as significance disappearance or even effect reversal were observed, which was indicative of stable results of this meta-analysis (Table 2).

Discussion

Cervical cancer has become a profound social and economic issue worldwide. To date, a quite large number of clinical trials, including vaccination-related trials, have been performed in developed countries such as European countries and United States, (Paavonen et al., 2009; Bonanni et al., 2015) by virtue of their well-established infrastructure and regulations. However, their trial results may not be directly applicable to Asian countries, where cervical cancer and HPV associated infectious diseases have posed

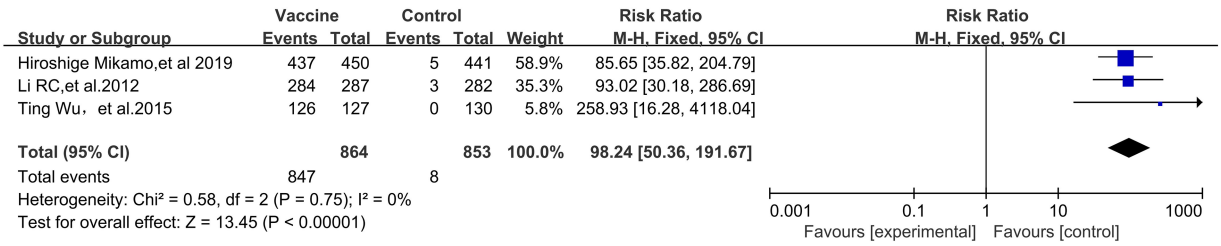

FIGURE 9
Comparison of the human palillomavirus type 18 (HPV 18) specific antibody conversion rate between the vaccinated group and the non-vaccination group in Asian populations.

TABLE 2 Sensitivity analysis results of the efficacy of HPV vaccine on cervical cancer among the Asian population.

Studies	Variation range of the RR value	Upper limit of the 95% CI value
HPV vaccine and the incidence of CIN1+	0.05–0.10	0.29
HPV vaccine and the incidence of CIN2+	0.06–0.13	0.43
HPV vaccine and the incidence of 6 PI	0.04–0.06	0.12
HPV vaccine and the incidence of 12 PI	0.06–0.10	0.26
HPV vaccine and the incidence of any cytological abnormality (ASC-US+)	0.13–0.15	0.31

serious threats to women's health, considering that vaccination effect may vary according to ethnic and social factors (Setiawan et al., 2017). Therefore, it is necessary, and also urgent, to assess the efficacy of HPV vaccination specific to Asian populations.

In this meta-analysis, a total of 10 articles were included. All these articles were high-quality as almost every article detailed the implementation processes for randomization, controlling and double blindness. In addition, all these studies had a large sample size. It is a long progression process from HPV infection to the development of cervical cancer. Therefore, we did not select cervical cancer as the primary endpoint for efficacy assessment. CIN1+ and CIN2+ are both precancerous lesions in relation to cervical cancer. Both the WHO and most trials (Malagón et al., 2012) have recommended that high-grade cervical lesions be the endpoints for prophylactic.

HPV vaccination efficacy assessment. Therefore, CIN1+ and CIN2+ were chosen as the assessed primary endpoints in this review. The meta-analysis showed that prophylactic HPV vaccination had satisfactory protective effect on precancerous lesions of cervical cancer, persistent infection and cytological abnormality, which were manifested by significant decrease in the incidences of CIN1+ (RR, 0.08; 95% CI, 0.03–0.22), CIN2+ (RR, 0. 09; 95% CI, 0.03–0.28), 6-month PI (RR, 0.05; 95% CI, 0.02–0.09), 12-month PI (RR, 0. 08; 95% CI, 0.04–0.17) and cytological abnormality (RR, 0.14; 95% CI, 0.09–0.22).

Whether prophylactic vaccines offer long-term protection remains an issue yet to be solved. In this study, six included trials offered a follow-up longer than 4 years (Konno et al., 2010; Li et al., 2012; Wu et al., 2015; Wei et al., 2019; Zhu et al., 2019; Zhao et al., 2022), and they all reported high sustained efficacy of HPV vaccination against HPV 16/18-assoicated CIN1 and CIN2. Previous reports have shown that HPV vaccine has a significant preventive effect on HPV 16/18 infection (Kudo et al., 2019; Sekine et al., 2020). Compared with unvaccinated populations, the incidence of cytological abnormalities, ASC-US or worse (ASC-US+), decreased by 24% in vaccinated populations (Ueda et al., 2018; Yagi et al., 2019). Future efficacy data from prophylactic vaccine trials with a longer-term follow-up are critical to fully explore the long-term efficacy of HPV vaccination.

Currently available prophylactic HPV vaccines offer protection against premalignant cervical disease by inducing and stimulating the expression of HPV16 and HPV18-specific antibodies. This meta-analysis showed that HPV vaccines were highly immunogenic; that is, they induced the expression of HPV16- and HPV18-specific antibodies in Asian populations. This finding was in perfect consistency with those reported in numerous studies that were conducted in western countries, including the US, European countries and Australia (Block et al., 2006; Muñoz et al., 2009; Einstein et al., 2014), as well as those conducted in other regions, such as Latin America (Perez et al., 2008) and Africa (Sow et al., 2013).

This study has the following limitations. It is a long progression process from HPV infection to cervical cancer and the confirmative evidence on how HPV vaccines reduce the incidence and mortality of cervical cancer remains unavailable at the present stage. This article mainly focused on the analysis of the RCTs conducted in Asian countries. The number of the included references was rather small. Therefore, future high-quality clinical trials with a large sample size and a longer-term follow-up remain to be conducted to further assess the

long-term efficacy of prophylactic HPV vaccines on cervical cancer. Additionally, to date, vaccination has not been included in immunization programs in most Asian countries and effective research object. Therefore, analysis of the effectiveness of vaccination in Asian countries, due to its lack of effective research subjects, brings a great deal of shortcomings in the conclusion of the study.

In summary, prophylactic HPV vaccination for cervical cancer is a prevention strategy full of challenges and hopes. This meta-analysis showed that prophylactic HPV vaccination had an effective preventing effect on HPV associated precancerous lesions. Although there is not more longer follow-up data of HPV vaccine from being on the market in 2006, but the Current data shows that HPV vaccine is an effective preventive measure against cervical cancer, and HPV Vaccine has been the main measure to the goal of global elimination of cervical cancer in 2030.In light with the results obtained in this meta-analysis, the government should accelerate the processes of vaccine introduction and vaccination implementation by prioritizing them in public health policies, and to enhance females' awareness of receiving HPV vaccination volitionally in Asian countries.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ZL designed this study. ZL and XR wrote the manuscript. XR, YH, BW, and XJ participated in study selection and data extraction. YH, BW, and MN performed statistical analysis. XR, YH, and KW reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank the data provided by the authors of included studies.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Block, S. L., Nolan, T., Sattler, C., Barr, E., Giacoletti, K. E. D., Marchant, C. D., et al. (2006). Comparison of the immunogenicity and reactogenicity of a prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in male and female adolescents and young adult women. *Pediatrics* 118, 2135–2145. doi: 10.1542/peds.2006-0461

Bonanni, P., Bechini, A., Donato, R., Capei, R., Sacco, C., Levi, M., et al. (2015). Human papilloma virus vaccination: impact and recommendations across the world. *Ther. Adv. Vacc.* 3, 3–12. doi: 10.1177/2051013614557476

Drolet, M., Bénard, É., Boily, M. C., Ali, H., Baandrup, L., Bauer, H., et al. (2015). Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect. Dis.* 15, 565–580. doi: 10.1016/S1473-3099(14)71073-4

Einstein, M. H., Takacs, P., Chatterjee, A., Sperling, R. S., Chakhtoura, N., Blatter, M. M., et al. (2014). Comparison of long-term immunogenicity and safety of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine and HPV-6/11/16/18 vaccine in healthy women aged 18-45 years: end-of-study analysis of a phase III randomized trial. *Hum. Vaccin. Immunother.* 10, 3435–3445. doi: 10.4161/hv.36121

Eriksson, T., Torvinen, S., Woodhall, S. C., Lehtinen, M., Apter, D., Harjula, K., et al. (2013). Impact of HPV16/18 vaccination on quality of life: a pilot study. *Eur. J. Contracept. Reprod. Health Care* 18, 364–371. doi: 10.3109/13625187.2013.801953

FUTURE I/II Study Group, Dillner, J., Kjaer, S. K., Wheeler, C. M., Sigurdsson, K., Iversen, O. E., et al. (2010). Four year efficacy of prophylactic human papillomavirus quadrivalent vaccine against low grade cervical, vulvar, and vaginal intraepithelial neoplasia and anogenital warts: randomised controlled trial. *BMJ* 341:3493. doi: 10.1136/bmj.c3493

Hamborsky, J., Kroger, A., and Wolfe, C.. (2015). *Epidemiology and Prevention of Vaccine-Preventable Diseases[R]. 13th ed.* Washington, DC: Public Health Foundation.

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Mohe, D., Oxman, A. D., et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928. doi: 10.1136/bmj.d5928

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. doi: 10.1136/bmj.327.7414.557

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., et al (2022). Cochrane handbook for systematic reviews of interventions version 6.3 Cochrane, 2022. Available at: https://training.cochrane.org/handbook (Accessed September 1, 2022).

Konno, R., Tamura, S., Dobbelaere, K., and Yoshikawa, H. (2010). Efficacy of human papillomavirus type 16/18 AS04-adjuvanted vaccine in Japanese women aged 20 to 25 years: final analysis of a phase 2 double-blind, randomized controlled trial. *Int. J. Gynecol. Cancer* 20, 847–855. doi: 10.1111/IGC.0b013e3181da2128

Konno, R., Yoshikawa, H., Okutani, M., Quint, W., V Suryakiran, P., Lin, L., et al. (2014). Efficacy of the human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical intraepithelial neoplasia and cervical infection in young Japanese women. *Hum. Vaccin. Immunother.* 10, 1781–1794. doi: 10.4161/hv.28712

Kudo, R., Yamaguchi, M., Sekine, M., Adachi, S., Ueda, Y., Miyagi, E., et al. (2019). Bivalent human papillomavirus vaccine effectiveness in a Japanese population: high vaccine-type-specific effectiveness and evidence of cross-protection. *J. Infect. Dis.* 219, 382–390. doi: 10.1093/infdis/jiy516

LaVigne, A. W., Triedman, S. A., Randall, T. C., Trimble, E. L., and Viswanathan, A. N. (2017). Cervical cancer in low and middle income countries: addressing barriers to radiotherapy delivery. Gynecol. *Oncol. Rep.* 22, 16–20. Published 2017 Sep 1. doi: 10.1016/j.gore.2017.08.004

Li, R., Li, Y., Radley, D., Liu, Y., Huang, T., Sings, H. L., et al. (2012). Safety and immunogenicity of a vaccine targeting human papillomavirus types 6, 11, 16 and

18: a randomized, double-blind, placebo-controlled trial in Chinese males and females. *Vaccine* 30, 4284–4291. doi: 10.1016/j.vaccine.2012.02.079

Malagón, T., Drolet, M., Boily, M. C., Franco, E. L., Jit, M., Brisson, J., et al. (2012). Cross-protective efficacy of two human papillomavirus vaccines: a systematic review and meta-analysis. *Lancet Infect. Dis.* 12, 781–789. doi: 10.1016/S1473-3099(12)70187-1

Mikamo, H., Yamagishi, Y., Murata, S., Yokokawa, R., Han, S. R., Wakana, A., et al. (2019). Efficacy, safety, and immunogenicity of a quadrivalent HPV vaccine in Japanese men: a randomized, phase 3, placebo-controlled study. *Vaccine* 37, 1651–1658. doi: 10.1016/j.vaccine.2019.01.069

Muñoz, N., Manalastas, R., Pitisuttithum, P., Tresukosol, D., Monsonego, J., Ault, K., et al. (2009). Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine in women aged 24-45 years: a randomised, double-blind trial. *Lancet* 373, 1949–1957. doi: 10.1016/S0140-6736(09)60691-7

Naud, P. S., Roteli-Martins, C. M., De Carvalho, N. S., Teixeira, J. C., de Borba, P. C., Sanchez, N., et al. (2014). Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine: final analysis of a long-term follow-up study up to 9.4 years post-vaccination. *Hum. Vaccin. Immunother.* 10, 2147–2162. doi: 10.4161/hv.29532

Paavonen, J., Naud, P., Salmerón, J., Wheeler, C. M., Chow, S. N., Apter, D., et al. (2009). Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet* 374, 301–314. doi: 10.1016/S0140-6736(09)61248-4

Perez, G., Lazcano-Ponce, E., Hernandez-Avila, M., García, P. J., Muñoz, N., Villa, L. L., et al. (2008). Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) L1 virus-like-particle vaccine in Latin American women. *Int. J. Cancer* 122, 1311–1318. doi: 10.1002/ijc.23260

Qiao, Y. L., Wu, T., Li, R. C., Hu, Y. M., Wei, L. H., Li, C. G., et al. (2020). Efficacy, safety, and immunogenicity of an Escherichia coli-produced bivalent human papillomavirus vaccine: an interim analysis of a randomized clinical trial. *J. Natl. Cancer Inst.* 112, 145–153. doi: 10.1093/jnci/djz074

Schwarz, T. F., Huang, L. M., Lin, T. Y., Wittermann, C., Panzer, F., Valencia, A., et al. (2014). Long-term immunogenicity and safety of the HPV-16/18 AS04-adjuvanted vaccine in 10- to 14-year-old girls: open 6-year follow-up of an initial observer-blinded, randomized trial. *Pediatr. Infect. Dis. J.* 33, 1255–1261. doi: 10.1097/INF.0000000000000460

Sekine, M., Yamaguchi, M., Kudo, R. J. B., Hanley, S., Hara, M., Adachi, S., et al. (2020). Epidemiologic profile of type-specific human papillomavirus infection after initiation of HPV vaccination. *Vaccines (Basel)* 8:425. Published 2020 Jul 29. doi: 10.3390/vaccines8030425

Setiawan, D., Luttjeboer, J., Pouwels, K. B., Wilschut, J. C., and Postma, M. J. (2017). Immunogenicity and safety of human papillomavirus (HPV) vaccination in Asian populations from six countries: a meta-analysis. *Jpn. J. Clin. Oncol.* 47, 265–276. doi: 10.1093/jjco/hyw192

Sow, P. S., Watson-Jones, D., Kiviat, N., Changalucha, J., Mbaye, K. D., Brown, J., et al. (2013). Safety and immunogenicity of human papillomavirus-16/18 AS04-adjuvanted vaccine: a randomized trial in 10-25-year-old HIV-Seronegative African girls and young women. *J. Infect. Dis.* 207, 1753–1763. doi: 10.1093/infdis/jis619

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Tsentemeidou, A., Fyrmpas, G., Stavrakas, M., Vlachtsis, K., Sotiriou, E., et al. (2021). Human papillomavirus vaccine to end Oropharyngeal cancer. A systematic review and meta-analysis. *Sex. Transm. Dis.* 48, 700–707. doi: 10.1097/OLQ.0000000000001405

Ueda, Y., Yagi, A., Nakayama, T., Hirai, K., Ikeda, S., Sekine, M., et al. (2018). Dynamic changes in Japan's prevalence of abnormal findings in cervical cervical cytology depending on birth year [published correction appears in Sci rep. 8(1):13384]. *Sci. Rep.* 8:5612. doi: 10.1038/s41598-018-23947-6

Vaccine and Immunization Branch of Chinese Preventive Medicine Association (2019). Expert consensus on immunoprophylaxis of human papillomavirus-related diseases [J]. *Chin. J. Prevent. Med.* 53, 761–803. doi: 10.3760/cma.j. issn. 0253-9624. 2019.08.001

Vaccines in National Immunization Programme. (2019). Available at: https://www. who.int/docs/default-source/documents/immunization/data/vaccine-intro-status. pdf? sfvrsn=bb 2857ec_2. (Accessed September 1, 2022).

Wei, L., Xie, X., Liu, J., Zhao, Y., Chen, W., Zhao, C., et al. (2019). Efficacy of quadrivalent human papillomavirus vaccine against persistent infection and genital disease in Chinese women: a randomized, placebo-controlled trial with 78-month follow-up. *Vaccine* 37, 3617–3624. doi: 10.1016/j.vaccine.2018.08.009

Wheeler, C. M., Bautista, O. M., Tomassini, J. E., Nelson, M., Sattler, C. A., Barr, E., et al. (2008). Safety and immunogenicity of co-administered quadrivalent human papillomavirus (HPV)-6/11/16/18 L1 virus-like particle (VLP) and hepatitis B (HBV) vaccines. *Vaccine* 26, 686–696. doi: 10.1016/j.vaccine.2007.11.043

World Health Organization (2017). Electronic address: sageexecsec@who.int. human papillomavirus vaccines: WHO position paper, may 2017-recommendations. *Vaccine* 35, 5753–5755. doi: 10.1016/j.vaccine.2017.05.069

World Health Organization. (2020) A global strategy for elimination of cervical cancer [EB/OL]. (2020-03-04) [2020-09-01]. Available at: https://www. who (Accessed November 17, 2020).

Wu, T., Hu, Y. M., Li, J., Chu, K., Huang, S. J., Zhao, H., et al. (2015). Immunogenicity and safety of an *E. coli*-produced bivalent human papillomavirus (type 16 and 18) vaccine: a randomized controlled phase 2 clinical trial. *Vaccine* 33, 3940–3946. doi: 10.1016/j.vaccine.2015.06.052

Yagi, A., Ueda, Y., Ikeda, S., Sekine, M., Nakayama, T., Miyagi, E., et al. (2019). Evaluation of future cervical cancer risk in Japan, based on birth year. *Vaccine* 37, 2889–2891. doi: 10.1016/j.vaccine.2019.04.044

Zhao, F. H., Wu, T., Hu, Y. M., Wei, L. H., Li, M. Q., Huang, W. J., et al. (2022). Efficacy, safety, and immunogenicity of an Escherichia coli-produced human papillomavirus (16 and 18) L1 virus-like-particle vaccine: end-of-study analysis of a phase 3, double-blind, randomised, controlled trial [published online ahead of print, 2022 Aug 26. *Lancet Infect. Dis.* S1473-3099, 00435–00432. doi: 10.1016/S1473-3099(22)00435-2

Zhu, F. C., Hu, S. Y., Hong, Y., Hu, Y. M., Zhang, X., Zhang, Y. J., et al. (2017). Efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine in Chinese women aged 18-25 years: event-triggered analysis of a randomized controlled trial. *Cancer Med.* 6, 12–25. doi: 10.1002/cam4.869

Zhu, F. C., Hu, S. Y., Hong, Y., Hu, Y. M., Zhang, X., Zhang, Y. J., et al. (2019). Efficacy, immunogenicity and safety of the AS04-HPV-16/18 vaccine in Chinese women aged 18-25 years: end-of-study results from a phase II/III, randomised, controlled trial. *Cancer Med.* 8, 6195–6211. doi: 10.1002/cam4.2399

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fmicb.2023.1163566

OPEN ACCESS

EDITED BY
Ana Grande-Pérez,
Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Spain

REVIEWED BY
Jeremy R. Thompson,
Plant Health & Environment Laboratories (MPI), New Zealand
Guillermo Dominguez Huerta,
The Ohio State University, United States

*CORRESPONDENCE
José Trinidad Ascencio-Ibáñez
✉ jtascenc@ncsu.edu

RECEIVED 10 February 2023
ACCEPTED 17 April 2023
PUBLISHED 25 May 2023

CITATION
Dye AE, Muga B, Mwangi J, Hoyer JS, Ly V, Rosado Y, Sharpee W, Mware B, Wambugu M, Labadie P, Deppong D, Jackai L, Jacobson A, Kennedy G, Ateka E, Duffy S, Hanley-Bowdoin L, Carbone I and Ascencio-Ibáñez JT (2023) Cassava begomovirus species diversity changes during plant vegetative cycles.
Front. Microbiol. 14:1163566.
doi: 10.3389/fmicb.2023.1163566

Cassava begomovirus species diversity changes during plant vegetative cycles

Anna E. Dye[1], Brenda Muga[2], Jenniffer Mwangi[2], J. Steen Hoyer[3], Vanessa Ly[4], Yamilex Rosado[4], William Sharpee[5,6], Benard Mware[5], Mary Wambugu[5], Paul Labadie[6], David Deppong[1], Louis Jackai[7], Alana Jacobson[8], George Kennedy[6], Elijah Ateka[2], Siobain Duffy[3], Linda Hanley-Bowdoin[1], Ignazio Carbone[9] and José Trinidad Ascencio-Ibáñez[4]*

[1]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, United States, [2]Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, [3]Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ, United States, [4]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, United States, [5]International Livestock Research Institute (ILRI), Nairobi, Kenya, [6]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, United States, [7]Department of Natural Resources and Environmental Design, North Carolina Agricultural and Technical State University, Greensboro, NC, United States, [8]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, United States, [9]Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, United States

Cassava is a root crop important for global food security and the third biggest source of calories on the African continent. Cassava production is threatened by Cassava mosaic disease (CMD), which is caused by a complex of single-stranded DNA viruses (family: *Geminiviridae*, genus: *Begomovirus*) that are transmitted by the sweet potato whitefly (*Bemisia tabaci*). Understanding the dynamics of different cassava mosaic begomovirus (CMB) species through time is important for contextualizing disease trends. Cassava plants with CMD symptoms were sampled in Lake Victoria and coastal regions of Kenya before transfer to a greenhouse setting and regular propagation. The field-collected and greenhouse samples were sequenced using Illumina short-read sequencing and analyzed on the Galaxy platform. In the field-collected samples, African cassava mosaic virus (ACMV), East African cassava mosaic virus (EACMV), East African cassava mosaic Kenya virus (EACMKV), and East African cassava mosaic virus-Uganda variant (EACMV-Ug) were detected in samples from the Lake Victoria region, while EACMV and East African mosaic Zanzibar virus (EACMZV) were found in the coastal region. Many of the field-collected samples had mixed infections of EACMV and another begomovirus. After 3 years of regrowth in the greenhouse, only EACMV-like viruses were detected in all samples. The results suggest that in these samples, EACMV becomes the dominant virus through vegetative propagation in a greenhouse. This differed from whitefly transmission results. Cassava plants were inoculated with ACMV and another EACMV-like virus, East African cassava mosaic Cameroon virus (EACMCV). Only ACMV was transmitted by whiteflies from these plants to recipient plants, as indicated by sequencing reads and copy number data. These results suggest that whitefly transmission and vegetative transmission lead to different outcomes for ACMV and EACMV-like viruses.

KEYWORDS

plant virus, Cassava (*Manihot esculenta*), vegetative (asexual) propagation, vector transmission, whitefly (*Bemisia tabaci*)

1. Introduction

Cassava (*Manihot esculenta* Crantz) is a temperature-resilient and drought-resistant crop that is important to smallholder farmers. Storage roots are harvested for consumption and commercial applications and are important for both food and economic security in Africa. Sub-Saharan Africa produces over half of the cassava grown worldwide (Food Agriculture Organization of the United Nations, 2020). Production of cassava is expected to increase to meet the anticipated decrease in the production of maize and rice as temperature increases (MoALFI, 2019; Ray et al., 2019; Harvesters, 2021). Cassava is a promising crop, but its production is threatened by several viral diseases, including cassava mosaic disease (CMD)—a viral disease that is endemic across Africa and causes major crop losses (Legg et al., 2011; Rey and Vanderschuren, 2017). CMD is characterized by leaf yellowing, deformation, and stunting. These physiological changes impact storage root development and often cause severe reduction in size. Although economic losses caused by CMD have not been calculated on a regional scale since the early 2000s (Legg et al., 2006), recent estimates point to sustained crop losses in both Kenya and other East African countries (Arama et al., 2016; Tembo et al., 2017).

Cassava mosaic disease (CMD) is caused by a complex of single-stranded DNA viruses in the *Begomovirus* genus (family: *Geminiviridae*) (Patil and Fauquet, 2009). Cassava mosaic begomoviruses (CMBs) have bipartite genomes with DNA-A and DNA-B components. Both components are required to establish a systemic infection (Figure 1). The DNA-A component encodes for replication (Rep, REn), encapsidation (CP), and anti-host defense functions (TrAP, AV2, and AC4), whereas the DNA-B component encodes for two movement proteins—the nuclear shuttle protein (NSP) and the movement protein (MP) (Hanley-Bowdoin et al., 2013). The AC5 ORF is hypothesized to have anti-silencing functions (Wu et al., 2022). CMBs are transmitted by whiteflies in the *Bemisia tabaci* cryptic species complex in a persistent, non-propagative manner (Mugerwa et al., 2012). Begomoviruses have high rates of mutation and recombination, leading to intra-host diversity and the emergence of novel species of begomoviruses (Duffy and Holmes, 2008, 2009; Crespo-Bellido et al., 2021; Mishra et al., 2022).

To date, 11 cassava mosaic begomovirus (CMB) species have been identified, with nine of them found in Africa (Jacobson et al., 2018). These species reflect a combination of evolutionary forces including mutation and recombination in mixed infections. Six of the African viruses are known recombinants (Crespo-Bellido et al., 2021). For example, East African cassava mosaic Zanzibar virus (EACMZV) is thought to be a recombinant of East African cassava mosaic Kenya virus (EACMKV) and South African cassava mosaic virus (SACMV), while EACMKV likely originated via a recombination event involving East African cassava mosaic Cameroon virus (EACMCV), which is itself a recombinant of East African cassava mosaic virus (EACMV) and an unknown virus (Crespo-Bellido et al., 2021). Begomoviruses are classified as distinct species when their DNA-A components show <91% identity (Brown et al., 2015). Because individual CMB DNA-B components often co-infect with highly divergent DNA-A components (reassortant viruses), their sequences are not used for species classification.

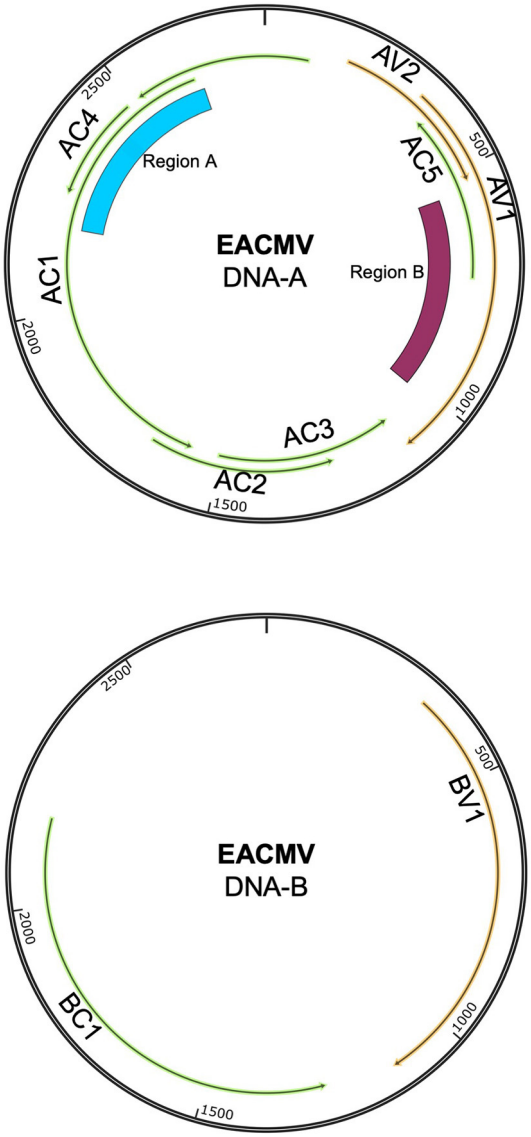

FIGURE 1
Genome map showing the DNA-A and DNA-B components of EACMV. Viral open-reading frames and genes are marked on both components. The ~400-bp region that distinguishes EACMV-like species is marked as Region A. This region overlaps with the AC4 and AC1 (Rep) open-reading frames. The area of recombination between ACMV and EACMV is marked as Region B. This region overlaps with the open-reading frames for AV1 (CP) and AC5. Genome maps were constructed in SnapGene.

African cassava mosaic virus (AMCV) and EACMV are thought to be the ancestral CMB species in sub-Saharan Africa (Jacobson et al., 2018). In the 1990s, a CMD pandemic spread from the Lake Victoria region in Uganda into central and eastern Africa (Legg and Thresh, 2000). The pandemic led to severe cassava crop

loss, with up to 100% yield loss. The pandemic was associated with three main factors: (1) a region of the ACMV coat protein (CP) recombined with EACMV DNA-A, resulting in a recombinant viral strain, EACMV-UG; (2) synergistic mixed infection with EACMV-UG and ACMV caused severe disease; and (3) the *B. tabaci* vector became superabundant at the wavefront of the pandemic (Zhou et al., 1997; Pita et al., 2001; Legg et al., 2006). The introduction of virus-resistant cassava cultivars and a reduction in whitefly populations have reduced disease impacts in the region, but CMD is still an important threat to cassava production (Were et al., 2021; Mwebaze et al., 2022). In addition, EACMV-UG continues to spread through Central Africa and toward West Africa, where it could negatively impact the large-scale cassava production in the region (Akinbade et al., 2010; Food Agriculture Organization of the United Nations, 2020; Mouketou et al., 2022).

Cassava mosaic begomoviruses (CMBs) are transmitted by vegetative propagation and whiteflies (Legg et al., 2015). This "vertical" transmission has the potential to transmit a diverse viral population during cycles of vegetative cutting and regrowth, while several filtering barriers in the whitefly body may reduce viral diversity during horizontal transmission. Thus, elucidating the effects of different transmission modes is important for understanding past viral pandemics and predicting future patterns of CMB spread. The exchange of infected cuttings leads to disease spread between farms and regions. It is estimated that 80% of disease transmission in Kenya is due to vegetative propagation (Mwatuni et al., 2015). Given the high frequency of vegetative transmission, it is essential to understand how this agricultural practice impacts changes in viral prevalence and species diversity. A few studies have evaluated the effect of vegetative transmission in other viral systems in potato and sugar cane (Sastry, 2013; Ranawaka et al., 2020). One study in cassava showed that CMB diversity increases through successive rounds of vegetative propagation (Aimone et al., 2021b). This study was performed under laboratory conditions using infectious clones of ACMV and EACMCV as the starting inoculum and, thus, may not fully represent the dynamics in field-infected plants.

The distributions of CMBs across different regions of sub-Saharan Africa have been characterized using field survey data (Ntawuruhunga et al., 2007; Chikoti et al., 2015; Harimalala et al., 2015; Tajebe et al., 2015; Doungous et al., 2022). Multiple studies have characterized the distribution of CMB species in Kenya during and after the pandemic (Sseruwagi et al., 2004; Mwatuni et al., 2015; Koima et al., 2018). Field studies provide important snapshot information, but they do not show how propagation affects viral species over time. In light of the laboratory studies showing that vegetative propagation impacts viral diversity (Aimone et al., 2021b), it is important to ask how vegetative transmission affects viral species presence in field-infected plants. Understanding of the effect of different transmission mechanisms on species presence (or species composition) can help inform management strategies and explain past disease trends.

2. Methods

2.1. Plant collection, growth, and symptoms

A CMD survey was conducted in coastal (semi-humid to semi-arid) and western (humid) regions of Kenya between June and September 2015. The cassava plants (unknown varieties), in this study, were from different fields separated by at least 10 km in the survey (Sseruwagi et al., 2004). The geo-coordinates (latitude, longitude, and altitude) of each sampling site were recorded using a Global Positioning System (GPS) receiver, GARMIN eTrex Legend (Garmin Ltd, Olathe, KS, USA). The plants were scored for CMD symptom severity at 3–6 months after field planting using a scale of 1 (no symptoms) to 5 (very severe symptoms) (Hahn et al., 1980). At the same time, lower shoot cuttings taken were propagated in an insect-proof greenhouse. The plants were watered daily, and pests were controlled by spraying at 2-week intervals using a broad-spectrum insecticide/miticide (Dynamec 1.8EC, Syngenta), as directed by the manufacturer to control mites, whiteflies, and mealy bugs. The plants were maintained in the greenhouse from 2015 to 2018 by cutting back at 3-month intervals, leaving two active buds for regrowth (Figure 2A). After 3 years of cutting and regrowth (10–13 cycles), the plants were photographed to record symptoms.

2.2. DNA isolation and PCR detection of CMBs

In 2015, total nucleic acid was isolated from the uppermost fully expanded leaves of the plants after 2 months in the greenhouse using a modified cetyltrimethylammonium bromide protocol (Lodhi et al., 1994). The 2015 samples were analyzed by PCR to confirm the presence of CMBs shortly after field collection (data not shown) and stored at −20°C. In 2018, after 3 years of repeated cycles of cutting and regrowth, total DNA was extracted from the third visible leaf relative to the apex of each plant using a Qiagen DNeasy Plant kit (Qiagen, Valencia, CA, USA).

Circular viral sequences in the 2015 and 2018 samples were subjected to linear amplification using an EquiPhi29 kit (Thermo Fisher Scientific, Waltham, MA), with 2 μl of total DNA as the input for rolling circle amplification (RCA) (Haible et al., 2006; Schubert et al., 2007; Jeske, 2009; Johne et al., 2009; Yang et al., 2014; Kathurima et al., 2016). PCR amplification of the RCA products was performed using begomovirus degenerate primers (RepMot: 5′GAGTCTAGAGGATANGTRAGGAAATARTTCTTGGC3′ and CPMot: 5′CGCGAATTCGACTGGACCTTACATGGNCCTTCAC 3′) (Ascencio-Ibanez et al., 2002) and HotStart Taq Polymerase (Qiagen, Valencia, CA). The reactions were performed for 35 cycles (denaturation: 30 s at 95°C, annealing: 1 min at 54°C, and elongation: 30 s at 72°C). Field- (2015) and greenhouse-collected (2018) samples were tested for the presence of begomoviruses. Four plants from the Lake Victoria region and four plants from the coastal region were selected for Illumina sequencing using the

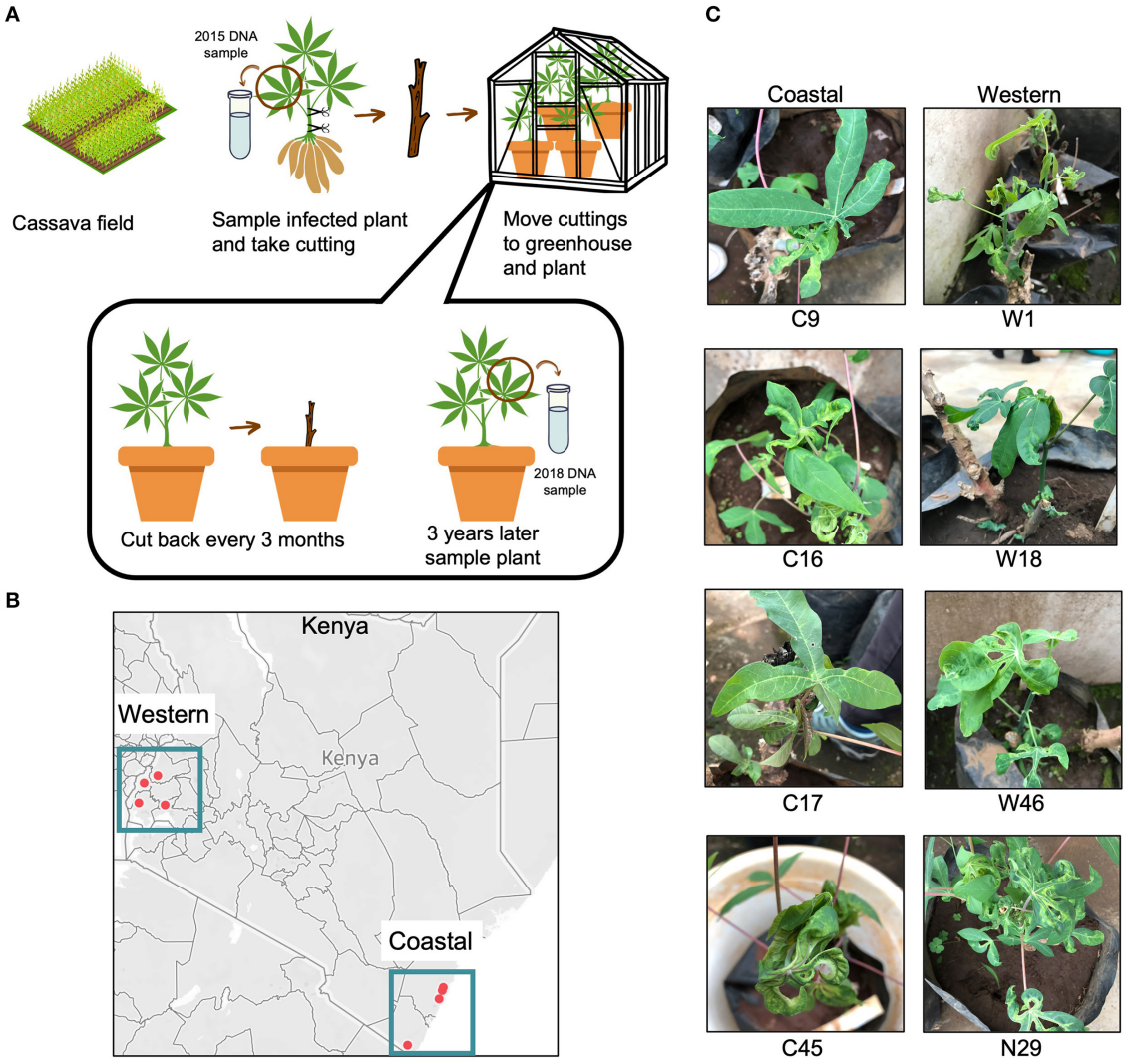

FIGURE 2
Cassava plants with cassava mosaic disease (CMD) collected from Kenyan fields. **(A)** Workflow showing the experimental design of the vegetative cutting experiment. The workflow was visualized using the Canva platform. **(B)** Points show collection locations of infected cassava in coastal and western areas of Kenya. Two points in the coastal region are close together and appear to overlap but are in two distinct field locations. The map was generated using Tableau. **(C)** Images of the field-collected cassava plants after 3 years of cutting in the greenhouse. Seven of the eight plants displayed severe CMD symptoms, while plant 17C showed no symptoms.

criteria that both the field and greenhouse samples for that plant tested positive for begomoviruses by PCR. One greenhouse sample did not test positive for begomoviruses (C17), but it was included to increase the number of samples for the coastal region. The selected samples were used for deep sequencing.

2.3. Library preparation and sequencing

Sequencing libraries were prepared using a method to enhance viral read counts for ssDNA virus genomes (Aimone et al., 2022). Two RCA reactions were performed for each sample using 2 μl of input DNA and the EquiPhi29 kit. After RCA, the products were end-repaired using Klenow and T4 DNA polymerase (New England BioLabs Inc., Ipswich, MA, USA), purified using 1.2× SPRI select beads (Beckman Coulter, Chaska, Minnesota, USA), and quantified using a Qubit fluorometer (Invitrogen, Waltham, Massachusetts). The RCA products from the two reactions were pooled after normalization of their concentrations and used to generate two libraries using an Illumina Nextera XT kit and IDT Unique Dual Indexes selection (Illumina, Inc., San Diego, California, USA; Integrated DNA Technologies, Coralville, IA, USA). The libraries were cleaned up using 0.8× SPRI select beads to remove library constructs with small inserts, normalized to 10 nM/library, and pooled for sequencing. Paired-end 150-bp reads were generated using the Illumina NovaSeq 6000 S4 platform for the eight sample pairs (field and greenhouse) with two technical replicates each for a total of 32 libraries.

2.4. Processing reads for reference-guided and *de novo* assemblies

Sequencing reads were processed using the ViralSeqMapping Pipeline on the Galaxy platform (Afgan et al., 2018; Aimone et al., 2022) (ViralSeq: https://cassavavirusevolution.vcl.ncsu.edu/). Sequencing adapters were trimmed, low-quality reads were discarded using Cutadapt (Martin, 2011), and reads were mapped to viral genomes using BWA-MEM (Li and Durbin, 2009; Li, 2013). Reads were mapped to the following reference sequences: EACMCV (monomer units in GenBank accessions MT856195.1 and MT856192.1), ACMV (MT858793.1 and MT858794.1), EACMV (MZ570970.1 and MZ570971.1), EACMV-UG (MK059418.1), EACMKV (AJ717572.1 and AJ704971.1), SACMV (AF155806.1 and AF155807.2), and EACMZV (AF422174.1 and AF422175.2). Reference sequences were chosen based on the CMBs previously recorded in Kenya and two CMBs (EACMCV and SACMV) that have not been recorded in Kenya as background controls. Mapped reads were sorted by coordinate order using SortSam in Picard tools (https://broadinstitute.github.io/picard/), PCR duplicate reads were removed using Picard MarkDuplicates, and read coverage was visualized using IGV (Robinson et al., 2011). Genomes were categorized as present with high amounts of virus in the sample, present with trace amounts of virus in the sample, or absent without detectable virus in the sample (Figure 3). Categories were determined by a consistent threshold of coverage across at least 95% of the DNA-A component for each species. 300× coverage was the threshold for high presence, 5× coverage was chosen as the threshold for trace amounts, and all samples below 5× coverage were treated as absent. Because of the recombinant nature of EACMV-like viruses, a 400-bp region that showed a high difference between EACMV, EACMKV, and EACMV was used to determine which EACMV-like virus was present.

De novo contigs were assembled from trimmed reads with MEGAHIT (Li et al., 2015), with the following arguments: –k-min 31 –min-count 3 –min-contig-len 500 –no-mercy. Contigs with begomovirus sequence similarity were identified using NCBI BLAST+ blastn on the Galaxy platform, with the reference sequences listed above as the subject sequences. The megablast setting was used, and the expectation value cutoff was set to 0.001. The contigs were analyzed further by querying the sequences using megablast against the entire BLAST database on the NCBI BLAST database (Sayers et al., 2021).

2.5. Consensus sequences

For each library, consensus sequences were generated using the "GetConsensus" pipeline on the Galaxy platform (GetConsensus: https://cassavavirusevolution.vcl.ncsu.edu/), which mapped reads to viruses present in the selected sample using BWA-MEM (Li and Durbin, 2009). Variants were called using Samtools mpileup and Varscan (Koboldt et al., 2012). Varscan was set to only include variants with over 50% prevalence in the sample and to include both indels and SNPs (Koboldt et al., 2012). The majority of variants were, then, used to create a consensus sequence using bcftools consensus. The pipeline ran the reads through mapping, mpileup generation, and variant calling in three successive rounds to ensure that all reads were appropriately mapped and checked for accuracy. Consensus sequences for individual viral components were generated for each technical replicate from both field- and greenhouse-collected samples. Consensus sequences for each component were aligned using MAFFT (Katoh and Standley, 2013) and implemented in the DeCIFR portal (https://tools.cifr.ncsu.edu/mafft). Multiple sequence alignments were visualized and compared for the similarity between replicates using the SnapGene software (Insightful Science; snapgene.com). Multiple sequence alignments were visualized with SnapGene software (Insightful Science; snapgene.com). Because of the levels of similarity between the technical replicates, consensus sequences from one replicate from each sample were used for subsequent analyses. Consensus sequences are available in Supplementary material "S1. Consensus sequences."

2.6. Phylogenetic trees

Consensus sequences were compiled into loci files and used to generate phylogenetic trees with the *de novo* tree inference tool (https://tools.cifr.ncsu.edu/denovo) in DeCIFR. The best-scoring maximum likelihood tree was based on 500 bootstrap replicates, and a GTRGAMMA model of evolution was performed using RAxML v8 (Stamatakis, 2014). Phylogenetic trees were inferred for each viral component separately. Trees were visualized using T-BAS v2.3 (https://tbas.cifr.ncsu.edu/) (Carbone et al., 2017, 2019).

2.7. Whitefly transmission

Cassava plants (cv. Kibandameno and cv. 60444) were co-inoculated with EACMCV (MT856195.1 and MT856192.1) (Fondong et al., 2000; Fondong and Chen, 2011; Hoyer et al., 2020) and ACMV (MT856193.2 and MT856194.1) infectious clone DNA (AddGene plasmids 159134 to 159137) by low-pressure biolistic bombardment (Aimone et al., 2022). ACMV and EACMCV produced in plants infected using viral clones can be acquired by whiteflies and subsequently transmitted into sucrose substrates and leaf disks (Kennedy et al., 2023). Three symptomatic cassava plants (two 60444 and one Kibandameno) were used as sources for three bioreplicates of whitefly transmission. One recipient plant per experiment was sequenced with the exception of the T8 lineage, for which two recipient plants were sequenced. The plants were placed in three separate insect cages with 400 non-viruliferous whiteflies for a 48-h acquisition access period (AAP).

Whiteflies were obtained from a colony of *B. tabaci* initiated from the offspring of 20 adults collected in 2016 from cassava fields in Kisumu county, Kenya, as described in the study by Kennedy et al. (2023). All founding adult whiteflies belonged to the SSA1-SG1 clade as determined by amplification of mtCOI using universal primers C1-2195 and L2-N-3014 (Simon et al., 1994) and published protocols (Boykin and De Barro, 2014). The colony was initially reared for at least two generations on eggplant (*Solanum melongena*), a non-host of CMBs, after which it was

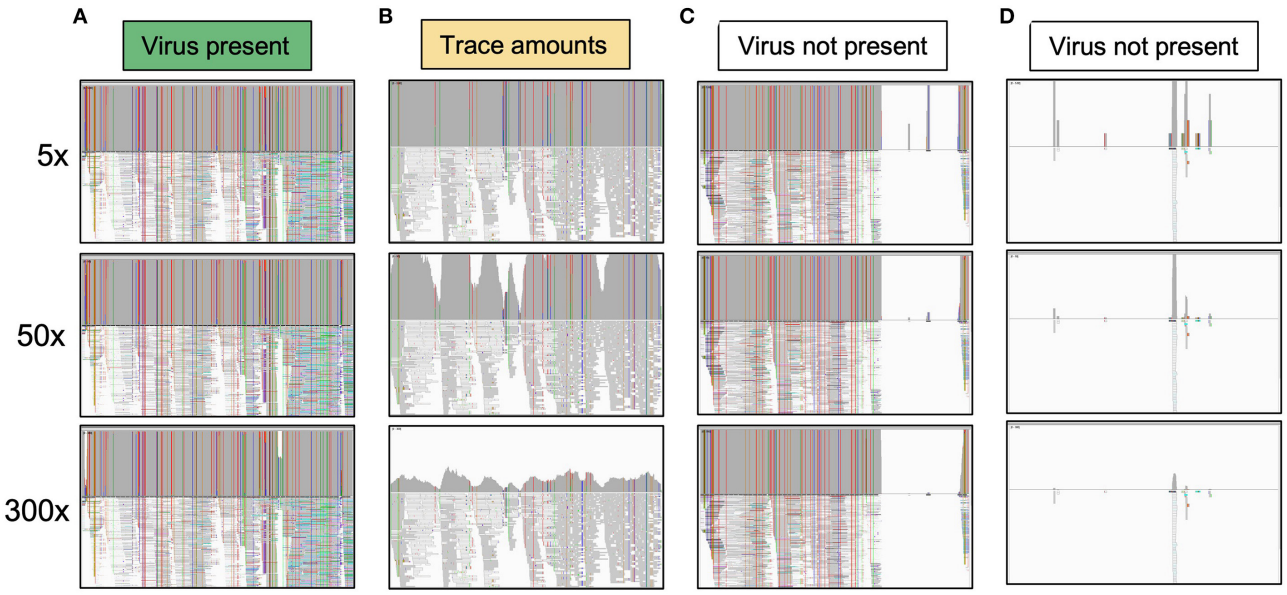

FIGURE 3
System for classifying the presence of viral species in a sample based on read coverage. Each category shows the IGV visualization of coverage across a viral genome component. **(A)** The virus is marked as present when it reaches 50× coverage. All viruses counted as present had at least 150× coverage, and all but one had >300× coverage. **(B)** An example of a virus present in trace amounts. The virus reaches the 5× threshold across over 95% of the genome component but does not meet the threshold for 50× coverage. **(C)** This example shows a virus that has high coverage for part of a genome component but not the entire genome component, indicating that the viral species queried is not present. Likely a related virus is present that shares some sequence identity for part of the genome. **(D)** This example shows that the virus is clearly not present. At any of the thresholds, the read coverage is not met across the genome component.

maintained on virus-free cassava (cv. Kibandameno) plants started in tissue culture.

Non-infected plants were moved into the insect cages to replace the infected plant for a 48-h inoculation access period (IAP) in complete darkness. At the end of the IAP, whiteflies were stored in 70% ethanol, and the recipient plants were treated with imidacloprid insecticide (Admire® Pro, Bayer CropScience). Leaf samples from the top three leaves were taken 28 days after the completion of the IAP. The sampled leaves emerged after whitefly feeding had finished. Total DNA was extracted using a Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA). Viral titers were quantified using quantitative PCR as previously described (Rajabu et al., 2018; Aimone et al., 2022). Samples were amplified in triplicate and compared with a standard curve. Titers are reported as viral copy number/ng of total DNA. DNA libraries were made for each source plant and recipient plant according to the method described above.

3. Results

3.1. Infected plant sample collection

Leaf samples and stem cuttings were collected from four areas of Kenya in 2015, in western and coastal Kenya (Figure 2B), and GPS locations were recorded as part of a larger study conducted by the Ateka research group. The 2015 DNA samples from the field-collected plants were retested in 2018 in end-point PCR assays using degenerate PCR primers. The expected 750-bp band was amplified from all eight samples, confirming the presence of CMB DNA-A at the time of field collection (Supplementary Figure S1A).

After 3 years of cutting back the plants every 3 months, leaf samples were collected from the eight plants again in 2018. The samples were tested for CMBs using degenerate begomovirus primers. In total, seven of the eight greenhouse samples tested positive for the presence of CMB DNA (Supplementary Figure S1B). The same seven plants exhibited leaf curling and mosaic patterning characteristic of CMD (Figure 2C). In contrast, coastal plant 17C showed no CMD symptoms and contained no detectable levels of CMB DNA-A in PCR assays. These results established that seven of the eight plants propagated for 3 years in the greenhouse were infected with at least one CMB, and only one plant, C17, appeared to have recovered from the infection.

3.2. Reference-guided viral genome assembly

Total DNA was used for next-generation sequencing of the eight 2015 and eight 2018 samples. Technical replicates were sequenced to address potential variability in the RCA and PCR amplification steps during library preparation (Aimone et al., 2022). The sequencing reads were mapped against CMB genomes and the cassava genome (NCBI assembly GCA_020916425.1), and the mapped reads from the technical libraries were combined. The mapping statistics are shown in Supplementary Table S1.

10.3389/fmicb.2023.1163566

TABLE 1 Viral read counts by sample from reference-guided assembly.

Area	Sample	Year	EACMV-like A	EACMV-like B	ACMV-A	ACMV-B
Coastal	C9	2015	101,627	106,795	208	100
		2018	1,256,012	465,484	92	42
	C16	2015	4,093,326	1,437,547	252	16
		2018	2,630,326	1,312,788	190	25
	C17	2015	4,330,732	1,734,657	394	30
		2018	5,531	1,132	0	0
	C45	2015	566,396	723,874	85	70
		2018	1,425,473	1,255,448	154	94
Western	W1	2015	9,039	5,132	326,607	52,980
		2018	3,491,183	3,048,044	1,155	470
	W18	2015	318,034	49,130	37,472	98,751
		2018	1,833,768	1,818,883	81	73
	W46	2015	3,750,519	1,546,347	411,696	453,304
		2018	1,881,391	2,420,126	72	44
	N29	2015	10,541	6,859	43,924	90,385
		2018	2,800,890	3,178,637	921	306

Green shows samples with consistent 300× coverage, indicating that the virus is present.
Yellow shows samples with coverage across the genome that does not reach 300× coverage, indicating that the virus could be present.
Cells with no highlighting show samples without consistent coverage across the genome, even at a 5× threshold, indicating that the virus is not present.

Viral genomes were assembled through reference-guided assembly of the sequencing reads. The reference genomes included EACMV, ACMV, EACMV-Ug, EACMKV, and EACMZV, all of which have been documented in Kenya (Mwatuni et al., 2015). The highest coverage was 9,000× for an EACMV-like virus and 4,500× for ACMV. We detected 23 virus instances using a threshold set at 300× coverage across at least 95% of the genome component (Figure 3A). Only one additional virus instance was detected when the coverage threshold was reduced to 50× (This instance was also detected at a 150× threshold). We assigned these 24 virus instances to the 'present' category (green cells in Table 1). We reduced the threshold to 10× and 5× coverage across 95% of the genome component to detect viruses that occur in trace amounts. The 10× and 5× thresholds uncovered one and eight additional instances, respectively. We assigned these nine instances that are not in the present category to the "trace" category (yellow cells in Table 1) (Figure 3B). The present and trace categories, which distinguish viruses occurring at high coverage vs. low coverage, were used to uncover changes in viral abundance between the 2015 and 2018 time points.

If reads localized to some parts of a viral genome component but were absent in other parts, the virus was not called present even if there were high numbers of mapping reads (white cells in Table 1). Incomplete coverage across the genome component was observed primarily for recombinant viruses with DNA-A components that derived in part from EACMV (Figure 3C). This is illustrated for EACMV-UG, which is a recombinant between EACMV and ACMV and has a gap in read coverage in the AV1 gene of EACMV (Supplementary Figure S2) that matches ACMV AV1 reads. We also mapped the reads to SACMV and EACMCV but only detected mapping to regions shared by other CMBs and not across their entire genomes (data not shown), indicating that these viruses were not present in the Kenyan field samples. When few to no reads mapped to the viral component, it indicated that the virus was not present (Figure 3D).

We initially screened the samples for ACMV and EACMV-like viruses. Because EACMV-like viruses display high sequence similarity across large sections of their genomes, reads mapping to either their DNA-A or DNA-B components were combined (Table 1). All samples at both time points tested positive for EACMV-like DNA-A and DNA-B. The western samples were also positive for ACMV in the 2015 samples but not in the corresponding 2018 samples, indicating that ACMV could only be detected at trace levels after 3 years of greenhouse propagation. None of the plants taken from coastal fields had ACMV at either the 2015 or 2018 time points.

The EACMV-like species were distinguished by mapping to the unique sequence regions of each virus species (Table 2). Samples were also required to have continuous 300× coverage across the entire genome to be called positive for a given EACMV-like virus. EACMV, EACMZV, and EACMKV differ in a ca. 400-bp segment overlapping the AC4 and AC1 genes. All the 2015 coastal samples had >300× coverage of EACMV DNA-A and DNA-B. Three of the coastal samples (9C, 16C, and 45C) were positive for EACMV in 2018. Coastal sample 16C also had >300× for EACMZV in 2015, indicating that the plant was co-infected with EACMV and EACMZV, but we did not detect EACMZV in the 2018 16C sample. The 2018 17C sample had low read counts but consistent coverage for EACMV and EACMZV in 2018, suggesting that the viruses

TABLE 2 Viral reads mapping specifically to the 400 bp of difference between DNA-A components of EACMV, EACMKV, and EACMZV.

Area	Sample	Year	EACMV	EACMKV	EACMZV
Coast	9C	2015	15,855	4	16
		2018	1,126,388	7	28
	16C	2015	3,170,131	5	1,330,365
		2018	5,609,983	32	20
	17C	2015	3,364,091	6	62
		2018	179	36	588
	45C	2015	167,261	0	20
		2018	648,158	45	23
West	W1	2015	43	326	2
		2018	3,107,571	33	150
	W18	2015	58,464	85	81
		2018	876,022	41,021	13
	W46	2015	3,628,008	89	128
		2018	920,244	50,710	28
	29N	2015	1,222	0	3
		2018	1,240,786	27,938	45

Green shows samples in which the selected virus is present.
Yellow shows samples in which trace amounts of the selected virus are present.
Cells with no highlighting show samples without consistent coverage across the genome, even at a 5× threshold, indicating that the virus is not present.

TABLE 3 EACMV and recombinant EACMV-Ug.

Area	Sample	Year	EACMV	EACMV-Ug
Coast	9C	2015	19,285	16
		2018	208,399	108
	16C	2015	385,307	4
		2018	474,853	9
	17C	2015	324,008	73
		2018	1,066	0
	45C	2015	81,267	2
		2018	174,389	12
West	W1	2015	539	54,707
		2018	373,635	194
	W18	2015	576	59,946
		2018	211,792	0
	29N	2015	983	7,637
		2018	268,011	163
	W46	2015	816	216,159
		2018	195,366	2

Viral reads mapping to the recombinant region of ACMV and EACMV that defines the EACMV-Ug pandemic variant to distinguish which viruses are present. The green color shows samples with consistent 300× coverage, indicating that the virus is present. The coverage must cross the recombination points of the whole-genome component. The yellow color shows samples with coverage across the genome that does not reach the sufficient 300× coverage, indicating that trace amounts of virus are present. The coverage must cross the recombination points of the whole-genome component.
Green shows samples in which the selected virus is present.
Yellow shows samples in which trace amounts of the selected virus are present.
Cells with no highlighting show samples without consistent coverage across the genome, even at a 5× threshold, indicating that the virus is not present.

were present at very low levels. EACMKV was not detected in any coastal samples.

The EACMV patterns were more varied in the western samples. Two western samples (W18 and W46) were positive for EACMV in both the 2015 and 2018 time points. EACMV was not detected in the 2015 W1 sample but was present in the corresponding 2018 sample. EACMV was detected but did not reach the >300× coverage threshold in the 2015 N29 sample but met the threshold in 2018. EACMKV was detected in two 2015 western samples (W18 and W1), both below the 300× threshold, but reached the 300× threshold in three 2018 samples (W18, W46, and 29N). EACMKV was detected at the 5× coverage threshold in the 2015 W18 sample but reached the 300× threshold in corresponding 2018 sample. In contrast, EACMKV was also detected below the 300× threshold in the 2015 W1 sample and was not detected in the 2018 W1 sample. EACMZV was only detected in trace amounts in W46-2015.

The EACMV-Ug pandemic variant was distinguished from other EACMV-like viruses using a 465-bp region overlapping the AV1 gene that was derived from ACMV by recombination. No coastal samples showed evidence of the presence of EACMV-Ug at either time point. In contrast, EACMV-Ug was detected at >300× coverage in three of the four western samples. Two 2018 western samples (W1 and 29N) also showed trace coverage for EACMV-Ug. However, because ACMV is present in all the samples where the detection of EACMV-Ug was positive, reads from ACMV could have been binned and counted as EACMV-Ug. Similarly, given that there is a similarity between EACMV and EACMV-Ug over most of the genome, an EACMV infection might map to a large portion of the EACMV-Ug reference genome. To address this possibility, coverage across the full DNA segment was viewed in IGV for EACMV and EACMV-Ug (Supplementary Figure S2). Samples that showed a gap in coverage at the EACMV AV1 but full coverage for EACMV-Ug were counted as positive for EACMV-Ug. This analysis indicated that only the 2015 W18, W46, and 29N and 2018 N29 samples were positive for the EACMV-Ug variant, and that the 2018 W1 sample was infected with EACMV and not the Uganda variant (Table 3).

3.3. *De novo* genome assembly

De novo genome assemblies were constructed as an alternative to confirm species identified using reference-guided assembly. Full-length and partial-length assemblies were constructed (Table 4). Only the DNA-A components are described in the table because their sequence determines species identity. Full-length *de novo* assemblies are denoted with "F," and partial assemblies are denoted with "P." The number of each assembly is noted in parentheses. If more than 35 contigs with similarity to begomoviruses were assembled from a sample, the first 35 contigs were used for the analysis. If <35 contigs were assembled from a sample, all

TABLE 4 Viral species presence by *de novo* assemblies.

Area	Sample	Year	EACMV-A	EACMKV-A	EACMV-Ug-A	EACMZV-A	ACMV-A
Coast	C9	2015	F (1)				
		2018	F (1) P (14)				
	C16	2015	P (3)			P (6)	
		2018	P (2)			P (6)	
	C17	2015	F (1) P (11)			P (4)	
		2018	P (2)			F (1) P(1)	
	C45	2015	F (1) P(7)				
		2018	P (10)				
West	W1	2015		F (1)			F (1)
		2018	F (1) P (1)				F (1)
	W18	2015	F (1) P (3)		P (7)	P (2)	F (2) P (3)
		2018	F (2) P (5)	F (1)			
	W46	2015		P (1)	P (8)		P (12)
		2018	F (1) P (7)	P (3)			
	N29	2015	P (7)		P (3)		P (2)
		2018	F (1) P (3)	P (1)			

Green shows which component was marked as present via reference-guided assembly. Yellow shows samples which showed a possible positive for the viral segment via reference-guided assembly. Full and partial *de novo* contigs were assembled.

F, full length *de novo* assembly; P, partial length *de novo* assembly; (x), the number of contigs.

Green shows samples in which the selected virus is present.

Yellow shows samples in which trace amounts of the selected virus is present.

Cells with no highlighting show samples without consistent coverage across the genome, even at a 5× threshold, indicating that the virus is not present.

the contigs were used for the analysis. All *de novo* assemblies were required to contain sequences flanking at least one known recombination junction to ensure accurate identification. The table is color-coded green and yellow for comparison to the results obtained using reference-guided assembly. The *de novo* and reference-guided assemblies gave very similar results. *De novo* assemblies detected the same viral components identified as present in the reference-guided assembly in 23 of 24 instances (Table 4, green cells) and trace amounts in four out of nine instances (yellow cells). *De novo* sequences were identified when the reference-guided assembly was not present in four instances. *De novo* but not reference-guided assembly also detected EACMZV in the 2018 C16 sample, but in this case, EACMZV was detected in 2015 by both methods. The 2015 W18 sample was the only sample for which *de novo* assembly detected a DNA-A component (EACMZV) that was not identified by reference-guided assembly in 2015 or by either method in 2018.

We compared the number of samples by time and region that were positive for viral genome components by *de novo* and reference-guided assemblies (Figure 4). This analysis included both DNA-A and DNA-B components. Each bar shows the sample count of the different components with a longer bar indicating more samples in the category. The two methods gave similar results. Both methods detected more CMB species in the 2015 western samples than in the 2015 coastal samples. A decrease in the number of the different viral components was seen in the 2018 samples compared with the 2015 samples for both geographical regions, most likely reflecting a decrease in co-infections and a decline in viral species diversity over time. There is also evidence of reassortment between DNA-A and DNA-B components of different CMBs. For example, EACMV-A, EACMV-B, and EACMZV-A were present in coastal samples at both time points, but EACMZV-B was only in the 2015 samples, suggesting that EACMV-B is providing movement functions for EACMZV-A in 2018 (Bull et al., 2007; Briddon et al., 2010).

Viral consensus sequences were used to generate phylogenetic trees. Sequences were nearly identical between technical replicates for all viral components, with 43/50 sequences sharing >99.95% sequence identity (Supplementary Table S3). A maximum likelihood tree of EACMV DNA-A sequences showed that the Western samples grouped together both in 2015 and 2018 but showed no significant relationship between samples from the same plant for EACMV DNA-A alone (Supplementary Figure S3A). There were substantial numbers of SNPs between 2015 and 2018 EACMV DNA-A consensus sequences from the same plant (ranging from 71 to 105), so we deemphasized studying sequences as sample pairs and consider our data "collected from the field 2015" and "greenhouse data from 2018." We did not find strong evidence of horizontal transfer of sequences among the greenhouse plants. The only possible exception was W1, which did not have any detectable EACMV in 2015 and had an EACMV DNA-A sequence closely related to the W18 sequence in 2018. Phylogenetic trees of more limited data sets (EACMKV DNA-A and EACMZV DNA-A) also showed no close relationships between 2015 and 2018 samples from the same fields and no

phylogenetic relationships indicated horizontal transfer from one plant to another (Supplementary Figure S3B).

3.4. Whitefly transmission of CMBs

Our studies showed that when plants started with a co-infection consisting of ACMV and EACMV-like virus, only the EACMV-like virus was present at high levels after 3 years of vegetative cutting. Given that CMBs are also transmitted by silverleaf whiteflies, we asked if vector transmission of ACMV and EACMV-like viruses also shows a bias. For these studies, we used ACMV and East African cassava mosaic Cameroon virus (EACMCV) for whitefly transmission studies in a controlled environment (Figure 5A). EACMCV is likely a recombinant virus of EACMV and an unknown virus (Fondong et al., 2000; Crespo-Bellido et al., 2021), and its AV1 gene has 98.4% sequence similarity to that of EACMV. The AV1 gene encodes the coat protein (CP), the only viral protein that has been implicated in whitefly transmission (Briddon et al., 1990; Harrison et al., 2002; Pan et al., 2020). Hence, EACMCV is a good choice for studying the transmission of EACMV-like viruses.

Susceptible cassava cultivars (cv. Kibandameno and cv. 60444) were infected with EACMCV and ACMV by biolistic inoculation and used as source plants for transmission by SSA1-SG1 whiteflies, the predominant whitefly biotype in western Kenya. One bioreplicate used Kibandameno as the source plant and two bioreplicates used 60444 as the source plant. Whiteflies fed on source plants for a 48-h AAP and then were moved to virus-free recipient plants for a 48-h IAP. Viral titers in source plants and recipient plants were measured by qPCR. Source plant titers were above 10 million copies/ng of total DNA for the ACMV genome components and above 5 million for the EACMCV genome components just prior to their use for the transmission (Figure 5B). In the seven recipient plants, the ACMV titers were above 30 million viral copies/ng of total DNA, while EACMCV DNAs were not detectable above the healthy plant control at 28 days post-IAP. The ACMV DNA-A:B ratio in the recipient plants was 0.24, which was similar to the source plants. Sequencing read counts were consistent with the qPCR results (Figure 5C). Viral read counts mapping to ACMV-A and ACMV-B in the source and recipient plants were above 40,000. The source plants had over 50,000 reads mapping to EACMCV-A and EACMCV-B, but fewer than 50 reads were mapping to EACMV DNA components in the recipient plants. The results showed that ACMV but not EACMCV can establish a systemic infection when transmitted by whiteflies in this experiment. It is important to note that the whitefly colony used for the transmission studies was generated from *B. tabaci* SSA1-SG1 individuals collected from fields in western Kenya, and the results could differ for another whitefly biotype collected from a different region.

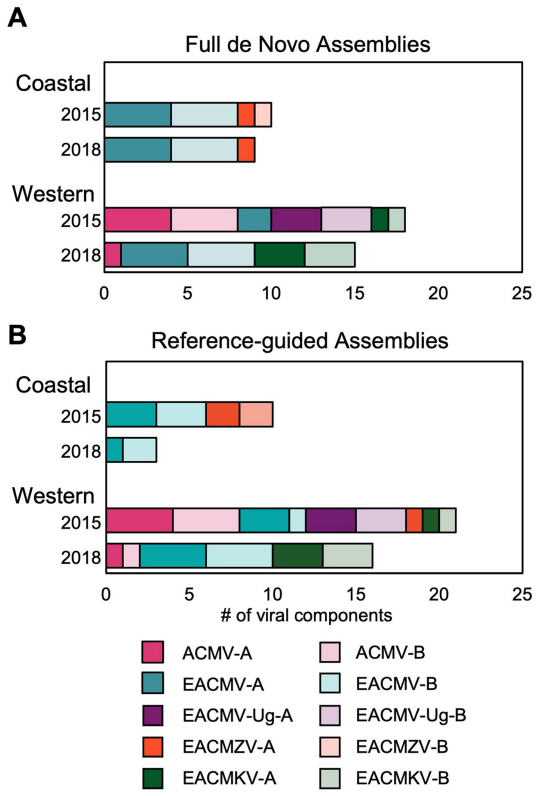

FIGURE 4
CMB sequence assemblies by region and year. **(A)** Full-length *de novo* genome assemblies generated with MEGAHIT are shown for the DNA-A and DNA-B components for ACMV, EACMV, EACMV-Ug, EACMZV, and EACMKV. The width for each component segment indicates the number of plants in that time point and region that have full-length *de novo* sequences. Wider segments correspond to more samples for the given component. **(B)** The number of viral components identified by the reference-guided assembly is shown using the same width parameters as in panel **(A)**.

4. Discussion

Cassava mosaic disease is transmitted by vegetative propagation and whitefly transmission. There is strong evidence that increased whitefly transmission was a key contributing factor to the severity and extent of the East Africa CMD pandemic in the 1990s (Legg et al., 2006). In contrast, recent surveys have suggested that vegetative propagation now accounts for the majority of CMB transmission in Kenya, and whitefly transmission is more likely to occur in western Kenya than in the coastal region (Mwatuni et al., 2015). However, it is not known whether the transmission mode impacts viral species diversity and influences CMB species on a regional scale.

Very few studies have investigated the impact of vegetative propagation on CMD. A recent laboratory study showed that CMB genome sequence diversity increases across multiple rounds of vegetative propagation (Aimone et al., 2021b) but could not address effects on CMB species diversity because the studies were initiated using infectious clones. In contrast, the studies reported here used viral inocula from field-infected plants to compare species diversity at the time of collection of field-infected cassava in western and coastal Kenya and after 3 years of vegetative cutting in a greenhouse. Our results suggest that the maintenance in the greenhouse favors the transmission of EACMV-like viruses, while ACMV is preferentially transmitted by whiteflies under

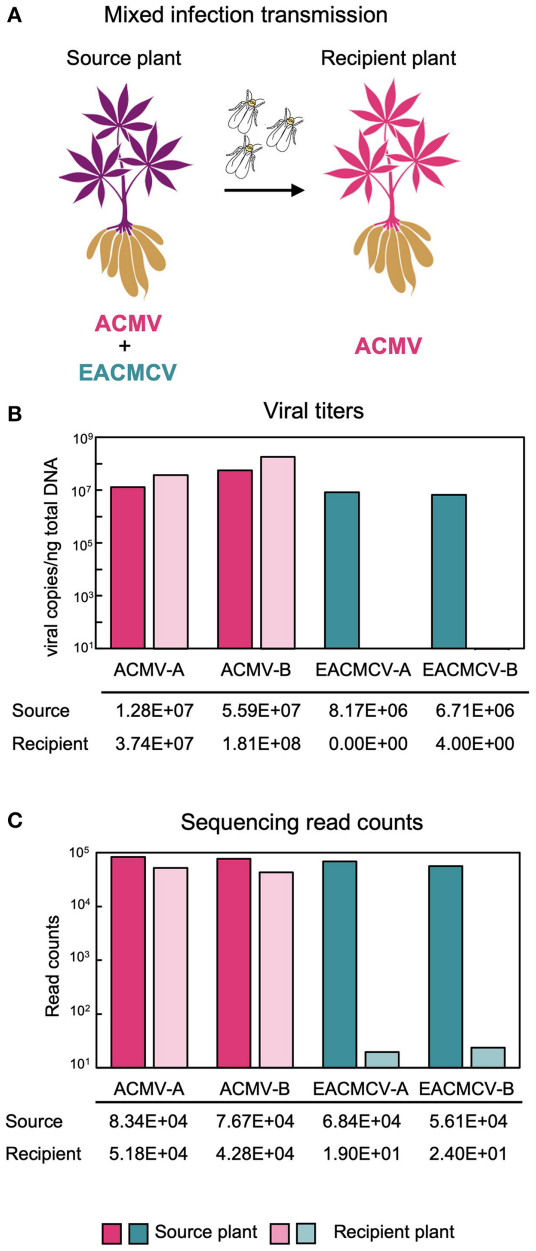

FIGURE 5
Whitefly transmission of ACMV and EACMCV. **(A)** Diagram of whitefly transmission showing that ACMV but not EACMCV is transmitted by whiteflies from cassava plants co-infected with both viruses in the laboratory. The diagram was generated using the Canva platform. **(B)** Average DNA copy numbers of ACMV and EACMCV in the source (three samples) and recipient (seven samples) plants. **(C)** Average NGS read counts by viral component from the same source and recipient plants.

controlled laboratory conditions using source plants inoculated with infectious viral clones.

Species diversity was richer in the western samples than in the coastal samples at the time of collection (2015) (Figure 4). EACMV was present in all samples at the coast but not in all samples from western Kenya. In contrast, ACMV was only detected in the western region. There were also differences in the EACMV-like species found in the western (EACMKV and EACMV-Ug) and the coastal areas (EACMZV). The viruses detected in the field-collected samples are consistent with other field studies in Kenya (Bull et al., 2006; Mwatuni et al., 2015; Ombiro, 2016; Koima et al., 2018; Were et al., 2021). After 3 years of maintaining the plants by vegetative cutting in a greenhouse, EACMV-like viruses became prominent in all plants, with EACMKV and EACMV in western samples and EACMV and EACMZV in coastal samples. In contrast, EACMV-Ug, with a recombinant coat protein region originating from ACMV, did not persist through vegetative propagation. ACMV also could not be detected in the western samples after vegetative propagation with the exception of two plants that had very low read counts for the virus.

The emergence of EACMV in the greenhouse over time could reflect whitefly transmission even though the plants underwent a consistent insecticide spraying regime. However, we think rampant vector transmission is unlikely because we did not detect convergence between the EACMV sequences in the western and coastal samples, which would have occurred if viruses had moved between plants. Our whitefly transmission results (Figure 5) also argue against efficient plant-to-plant transmission of EACMV in the greenhouse. Instead, we propose that EACMV and ACMV respond differently to vegetative propagation because of the distinct kinetics of their infection processes in cassava. It is possible that particular isolates emerge in the greenhouse due to the different selection pressures seen in the greenhouse compared with the field, as has been seen with TYLCV (Sánchez-Campos et al., 2018). ACMV develops symptoms and accumulates to high levels quickly after inoculation but then titers decrease and the plant recovers from ACMV symptoms, while EACMV-like viruses establish infection more slowly and do not recover over time (Vanitharani et al., 2004; Patil and Fauquet, 2009). As a consequence, EACMV, more than ACMV, has the potential to be maintained and become the predominant virus during repeated cycles of cassava regrowth.

We observed examples of a virus occurring in a 2018 sample but not in its corresponding 2015 sample. This was seen exclusively for the western samples and involved in the detection of EACMV or EACMKV. Part of our difficulty in detecting specific EACMV-like viruses is that they can form reassortants and function together in co-infections (Bull et al., 2007; De Bruyn et al., 2012). Specifically, EACMV and EACMKV have the same iteron sequences and can form reassortants (Argüello-Astorga et al., 1994; Argüello-Astorga and Ruiz-Medrano, 2001). EACMV-like viruses have very similar sequences due to recombination (Lefeuvre and Moriones, 2015; Crespo-Bellido et al., 2021) of different CMB species. The sequences of EACMV and EACMKV are highly similar for a large portion of the DNA-A component, only diverging significantly in the AC1 and AC2 genes. Thus, determining which reads map to EACMV or EACMKV can be challenging when using short-read sequencing and reference-guided assembly. This is particularly problematic when a plant contains two related viruses with large differences in their genome copy numbers, making it very difficult to detect the less abundant virus. We addressed this issue, in part, by using *de novo* assembly, which does not rely on mapping to reference genomes, to detect low-abundance species. Our results illustrate the importance of using a combination of reference-guided assembly and *de novo* assembly for the accurate identification of highly similar viral species. However, horizontal transfer (by whiteflies)

would be another explanation for the detection of a new species or a very distinct haplotype of the same species over the 3 years of vegetative cutting. We see only limited evidence of horizontal transfer of viruses (i.e., EACMV DNA-A in W1 in 2018 is closely related to the same in W18, Supplementary Figure S3A), but we observed more divergence among EACMV populations in 3 years than expected (Duffy and Holmes, 2009). Although there may be some confounding horizontal transfer of viruses in the experiment, our results indicate that EACMV-like viruses are favored by cycles of vegetative regrowth. The loss of ACMV and EACMV-Ug, which shares most of the ACMV coat protein that is essential for whitefly transmission (Briddon et al., 1990; Höfer et al., 1997; Harrison et al., 2002; Rana et al., 2016; Saurav et al., 2019; Pan et al., 2020) in the greenhouse, is consistent with a central role of vector transmission for maintenance of these viruses.

We hypothesize that the persistence of ACMV and EACMV-Ug in western Kenya is facilitated by vector transmission. This idea fits with trends observed in the East Africa CMD pandemic when high levels of ACMV and EACMV-Ug were accompanied by the emergence of a new super-abundant whitefly population (Colvin et al., 2004; Legg et al., 2006, 2014). It is also supported by the evidence from a 2015 field survey (Mwatuni et al., 2015) that found EACMV-like, ACMV, and EACMV-Ug viruses in western Kenya and symptoms indicative of both vegetative and whitefly transmission in infected plants. In contrast, the survey only detected EACMV and EACMZV and symptoms consistent with the vegetative transmission in coastal Kenya.

Whitefly density is likely not the only reason behind the differences in vector transmission between western and coastal Kenya. This idea is supported by the 2015 survey described above that showed that whitefly populations were high in both regions (Mwatuni et al., 2015). Whitefly diversity and differential ability to transmit viruses could be contributing factors to the difference in virus species by region and the incidence of vector transmission. Whiteflies are separated into genetic biotypes based on mtCOI sequences (Mugerwa et al., 2012, 2018; Manani et al., 2017). The SSA1-SG1 and SSA1-SG2 biotypes have been associated with severe CMD (Ndunguru et al., 2016; Aimone et al., 2021a). A recent study found SSA1-SG1, SSA1-SG2, and SSA2 biotypes in western Kenya and SSA2 and SSA1-SG3 biotypes in coastal Kenya (Munguti et al., 2021). Whitefly transmission is present in western Kenya, where SSA1-SG1 and SSA1-SG2 are present. A few fields with the SSA1-SG1 biotype have also been found in coastal Kenya in 2021, which may correlate with the first instance of AMCV in coastal Kenya (Munguti et al., 2021). There have not yet been comprehensive studies to show whether whitefly biotypes transmit CMBs or various CMB species differently, but other begomoviruses are known to be transmitted by different biotypes at different rates (Zhao et al., 2019; Chi et al., 2020; Fiallo-Olivé et al., 2020; Pan et al., 2020; Gautam et al., 2022). Other factors that could impact what CMBs are present and how they are transmitted include cassava cultivar differences by region, the adaptation to those improved varieties, and environmental factors.

These results also have implications for CMD management in East Africa. EACMV and EACMV-like viruses represent the greatest risk to the cassava seed systems that rely on vegetative propagation and ratooning to generate planting material for smallholder farmers (Ceballos et al., 2020). Regions with high pressure from SSA1-SG1 and SSA1-SG2 whiteflies are likely to be at greater risk from ACMV and EACMV-Ug. These observations can inform the models of CMD emergence and spread and help to develop better control methods.

Data availability statement

The data presented in the study are deposited in the NCBI Sequence Read Archive repository, BioProject PRJNA950083.

Author contributions

AD curated data, analyzed data, established analysis methodology, and wrote the manuscript. BMu and JM conceptualized the experiment, collected data, and established the methodology for vegetative transmission. JH helped conceptualize the vegetative experiments and completed some data analyses. YR and VL curated data and analyzed samples. WS, BMw, and MW conducted whitefly transmission experiments in Nairobi. PL curated data through library construction. LJ provided training and supervision for undergraduate trainees and developed the framework for the undergraduate research study with JA-I. GK and AJ developed the methodology, experimental conceptualization, and design of whitefly transmission experiments. SD assisted with bioinformatic analyses and manuscript editing. LH-B assisted with validation, manuscript preparation, supervision, and editing. EA organized the field collections and maintenance of plants in the greenhouse at JKUAT. IC assisted with bioinformatic analysis and resources. JA-I conceptualized the experiments and methodology, organized and supervised the specimen collection in Kenya, and supervised data analysis and manuscript preparation. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the USA National Science Foundation grant number OISE-1545553 to LH-B, SD, and GK and the Bill & Melinda Gates Foundation Grant #51466 to EA.

Acknowledgments

The authors would like to thank Emely Pacheco for preparing consensus sequences, Bria Massey, Ashley Yancey, and Ivori Schley for collecting samples from JKUAT, and Dr. Cyprian Rajabu for helping undergraduates with sample collection. We thank Rutgers OARC (Advanced Research and Computing) staff for the maintenance of the Amarel cluster, James White at NCSU for software support, and Mary Dallas for assistance with lab support. The authors also would like to thank Dr. Wellington Ekaya at BecA and other staff at JKUAT, BecA, and NCSU for logistic support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1163566/full#supplementary-material

References

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379

Aimone, C. D., De León, L., Dallas, M. M., Ndunguru, J., Ascencio-Ibáñez, J. T., Hanley-Bowdoin, L, et al. (2021a). A new type of satellite associated with cassava mosaic begomoviruses. *J. Virol.* 95, e00432-00421. doi: 10.1128/JVI.00432-21

Aimone, C. D., Hoyer, J. S., Dye, A. E., Deppong, D. O., Duffy, S., Carbone, I., et al. (2022). An experimental strategy for preparing circular ssDNA virus genomes for next-generation sequencing. *J. Virol. Methods* 300, 114405. doi: 10.1016/j.jviromet.2021.114405

Aimone, C. D., Lavington, E., Hoyer, J. S., Deppong, D. O., Mickelson-Young, L., Jacobson, A., et al. (2021b). Population diversity of cassava mosaic begomoviruses increases over the course of serial vegetative propagation. *J. Gen. Virol.* 102, 001622. doi: 10.1099/jgv.0.001622

Akinbade, S., Hanna, R., Nguenkam, A., Njukwe, E., Fotso, A., Doumtsop, A., et al. (2010). First report of the East African cassava mosaic virus-Uganda (EACMV-UG) infecting cassava (*Manihot esculenta*) in Cameroon. *New Dis. Rep.* 21, 2044-0588. doi: 10.5197/j.2044-0588.2010.021.022

Arama, P., Mulwa, R., Hillocks, R., Maruthi, M. N., Ogendo, J. O., Masinde, E. A., et al. (2016). Occurrence and estimated losses caused by cassava viruses in Migori County, Kenya. *Afr. J. Agric. Res.* 11, 2064–2074. doi: 10.5897/AJAR2016.10786

Argüello-Astorga, G., Guevara-Gonzalez, R., Herrera-Estrella, L., and Rivera-Bustamante, R. (1994). Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology* 203, 90–100. doi: 10.1006/viro.1994.1458

Argüello-Astorga, G., and Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch. Virol.* 146, 1465–1485. doi: 10.1007/s007050170072

Ascencio-Ibanez, J. T., Arguello-Astorga, G. R., Mendez-Lozano, J., and Rivera-Bustamante, R. F. (2002). First report of rhynchosia golden mosaic virus (RhGMV) infecting tobacco in Chiapas, Mexico. *Plant Dis.* 86, 692. doi: 10.1094/PDIS.2002.86.6.692C

Boykin, L. M., and De Barro, P. J. (2014). A practical guide to identifying members of the *Bemisia tabaci* species complex: and other morphologically identical species. *Front. Ecol. Evol.* 2, 45. doi: 10.3389/fevo.2014.00045

Briddon, R., Pinner, M., Stanley, J., and Markham, P. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology* 177, 85–94. doi: 10.1016/0042-6822(90)90462-Z

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-Ul-Rehman, M. S., and Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evol. Biol.* 10, 1–17. doi: 10.1186/1471-2148-10-97

Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C., et al. (2015). *Revision of Begomovirus Taxonomy Based on Pairwise Sequence Comparisons*. Cham: Springer. doi: 10.1007/s00705-015-2398-y

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., Stanley, J., et al. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *J. Gen. Virol.* 87, 3053–3065. doi: 10.1099/vir.0.82013-0

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., Stanley, J., et al. (2007). Infectivity, pseudorecombination and mutagenesis of Kenyan cassava mosaic begomoviruses. *J. Gen. Virol.* 88, 1624–1633. doi: 10.1099/vir.0.82662-0

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Kauff, F., et al. (2017). T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. *Bioinformatics* 33, 1160–1168. doi: 10.1093/bioinformatics/btw808

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Magain, N., et al. (2019). T-BAS version 2.1: Tree-Based Alignment Selector toolkit for evolutionary placement of DNA sequences and viewing alignments and specimen metadata on curated and custom trees. *Microbiol. Resour. Announc.* 8, e00328–00319. doi: 10.1128/MRA.00328-19

Ceballos, H., Rojanaridpiched, C., Phumichai, C., Becerra, L. A., Kittipadakul, P., Iglesias, C., et al. (2020). Excellence in cassava breeding: perspectives for the future. *Crop Breed. Genet. Genom.* 2, e200008. doi: 10.20900/cbgg20200008

Chi, Y., Pan, L.-L., Bouvaine, S., Fan, Y.-Y., Liu, Y.-Q., Liu, S.-S., et al. (2020). Differential transmission of Sri Lankan cassava mosaic virus by three cryptic species of the whitefly *Bemisia tabaci* complex. *Virology* 540, 141–149. doi: 10.1016/j.virol.2019.11.013

Chikoti, P. C., Tembo, M., Chisola, M., Ntawuruhungu, P., and Ndunguru, J. (2015). Status of cassava mosaic disease and whitefly population in Zambia. *Afr. J. Biotechnol.* 14, 2539–2546. doi: 10.5897/AJB2015.14757

Colvin, J., Omongo, C., and Maruthi, M. Otim-Nape, G., Thresh, J. (2004). Dual begomovirus infections and high *Bemisia tabaci* populations: two factors driving the spread of a cassava mosaic disease pandemic. *Plant Pathol.* 53, 577–584. doi: 10.1111/j.0032-0862.2004.01062.x

Crespo-Bellido, A., Hoyer, J. S., Dubey, D., Jeannot, R. B., and Duffy, S. (2021). Interspecies recombination has driven the macroevolution of cassava mosaic begomoviruses. *J. Virol.* 95, e00541-00521. doi: 10.1128/JVI.00541-21

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, G. W., et al. (2012). East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC Evol. Biol.* 12, 1–18. doi: 10.1186/1471-2148-12-228

Doungous, O., Masky, B., Levai, D. L., Bahoya, J. A., Minyaka, E., Mavoungou, J. F., et al. (2022). Cassava mosaic disease and its whitefly vector in Cameroon: incidence, severity and whitefly numbers from field surveys. *Crop Prot.* 158, 106017. doi: 10.1016/j.cropro.2022.106017

Duffy, S., and Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *J. Virol.* 82, 957–965. doi: 10.1128/JVI.01929-07

Duffy, S., and Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J. Gen. Virol.* 90, 1539. doi: 10.1099/vir.0.009266-0

Fiallo-Olivé, E., Pan, L.-L., Liu, S.-S., and Navas-Castillo, J. (2020). Transmission of begomoviruses and other whitefly-borne viruses: dependence on the vector species. *Phytopathology* 110, 10–17. doi: 10.1094/PHYTO-07-19-0273-FI

Fondong, V., Pita, J., Rey, M., De Kochko, A., Beachy, R. N., Fauquet, C., et al. (2000). Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *J. Gen. Virol.* 81, 287–297. doi: 10.1099/0022-1317-81-1-287

Fondong, V. N., and Chen, K. (2011). Genetic variability of East African cassava mosaic Cameroon virus under field and controlled environment conditions. *Virology* 413, 275–282. doi: 10.1016/j.virol.2011.02.024

Food and Agriculture Organization of the United Nations (2020). *FAOSTAT Statistical Database*. Rome: FAO.

Gautam, S., Mugerwa, H., Buck, J. W., Dutta, B., Coolong, T., Adkins, S., et al. (2022). Differential transmission of old and new world begomoviruses by middle East-Asia minor 1 (MEAM1) and Mediterranean (MED) cryptic species of *Bemisia tabaci*. *Viruses* 14, 1104. doi: 10.3390/v14051104

Hahn, S., Terry, E., and Leuschner, K. (1980). Breeding cassava for resistance to cassava mosaic disease. *Euphytica* 29, 673–683. doi: 10.1007/BF00023215

Haible, D., Kober, S., and Jeske, H. (2006). Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *J. Virol. Methods* 135, 9–16. doi: 10.1016/j.jviromet.2006.01.017

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., and Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol.* 11, 777–788. doi: 10.1038/nrmicro3117

Harimalala, M., Chiroleu, F., Giraud-Carrier, C., Hoareau, M., Zinga, I., Randriamampianina, J. A., et al. (2015). Molecular epidemiology of cassava mosaic disease in M adagascar. *Plant Pathol.* 64, 501–507. doi: 10.1111/ppa.12277

Harrison, B., Swanson, M., and Fargette, D. (2002). Begomovirus coat protein: serology, variation and functions. *Physiol. Mol. Plant Pathol.* 60, 257–271. doi: 10.1006/pmpp.2002.0404

Harvesters, A. (2021). *Kenya Plans to Increase Cassava Production as Cliamte Change hits Staple Crops*.

Höfer, P., Bedford, I. D., Markham, P. G., Jeske, H., and Frischmuth, T. (1997). Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. *Virology* 236, 288–295. doi: 10.1006/viro.1997.8751

Hoyer, J. S., Fondong, V. N., Dallas, M. M., Aimone, C. D., Deppong, D. O., Duffy, S., et al. (2020). Deeply sequenced infectious clones of key cassava begomovirus isolates from Cameroon. *Microbiol. Resour. Announc.* 9, e00802–00820. doi: 10.1128/MRA.00802-20

Jacobson, A. L., Duffy, S., and Sseruwagi, P. (2018). Whitefly-transmitted viruses threatening cassava production in Africa. *Curr. Opin. Virol.* 33, 167–176. doi: 10.1016/j.coviro.2018.08.016

Jeske, H. (2009). Geminiviruses. *Curr. Top. Microbiol. Immunol.* 331, 185–226. doi: 10.1007/978-3-540-70972-5_11

Johne, R., Müller, H., Rector, A., Van Ranst, M., and Stevens, H. (2009). Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends Microbiol.* 17, 205–211. doi: 10.1016/j.tim.2009.02.004

Kathurima, T., Ateka, E., Nyende, A., and Holton, T. (2016). The rolling circle amplification and next generation sequencing approaches reveal genome wide diversity of Kenyan cassava mosaic geminivirus. *Afr. J. Biotechnol.* 15, 2045–2052. doi: 10.5897/AJB2016.15357

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kennedy, G. G. S., Jacobson, W., Wambugu, A. L., Mware, M., and Hanley-Bowdoin, B. L. (2023). Genome formula changes during whitefly transmission of two bipartite cassava mosaic begomoviruses. *Res. Sq.* PREPINT (Version 1). doi: 10.21203/rs.3.rs-2684703/v1

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., Mclellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi: 10.1101/gr.129684.111

Koima, I. N., Orek, C. O., and Nguluu, S. N. (2018). Distribution of cassava mosaic and cassava brown streak diseases in agro-ecological zones of lower eastern Kenya. *Int. J. Innovative Sci. Res. Technol* 3, 391–399.

Lefeuvre, P., and Moriones, E. (2015). Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. *Curr. Opin. Virol.* 10, 14–19. doi: 10.1016/j.coviro.2014.12.005

Legg, J., Jeremiah, S., Obiero, H., Maruthi, M., Ndyetabula, I., Okao-Okuja, G., et al. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res.* 159, 161–170. doi: 10.1016/j.virusres.2011.04.018

Legg, J. P., Kumar, P. L., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., et al. (2015). Cassava virus diseases: biology, epidemiology, and management. *Adv. Virus Res.* 91, 85–142. doi: 10.1016/bs.aivir.2014.10.001

Legg, J. P., Owor, B., Sseruwagi, P., and Ndunguru, J. (2006). Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Adv. Virus Res.* 67, 355–418. doi: 10.1016/S0065-3527(06)67010-3

Legg, J. P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R., Bigirimana, S., et al. (2014). Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. *Virus Res.* 186, 61–75. doi: 10.1016/j.virusres.2013.11.018

Legg, J. P., and Thresh, J. M. (2000). Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. *Virus Res.* 71, 135–149. doi: 10.1016/S0168-1702(00)00194-5

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. [preprint]. doi: 10.48550/arXiv.1303.3997

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lodhi, M. A., Ye, G.-N., Weeden, N. F., and Reisch, B. I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. *Plant Mol. Biol. Rep.* 12, 6–13. doi: 10.1007/BF02668658

Manani, D. M., Ateka, E. M., Nyanjom, S. R., and Boykin, L. M. (2017). Phylogenetic relationships among whiteflies in the *Bemisia tabaci* (Gennadius) species complex from major cassava growing areas in Kenya. *Insects* 8, 25. doi: 10.3390/insects8010025

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Mishra, M., Verma, R. K., Pandey, V., Srivastava, A., Sharma, P., Gaur, R., et al. (2022). Role of diversity and recombination in the emergence of chilli leaf curl virus. *Pathogens* 11, 529. doi: 10.3390/pathogens11050529

MoALFI (2019). *National Root and Tuber Crops Development Strategy 2019-2022*. Nairobi: Ministry of Agriculture, Livestock, Fisheries and Irrigation.

Mouketou, A., Koumba, A., Gnacadja, C., Zinga-Koumba, C., Meye, C. A., Ovono, A., et al. (2022). Cassava mosaic disease incidence and severity and whitefly vector distribution in Gabon. *Afr. Crop Sci. J.* 30, 167–183. doi: 10.4314/acsj.v30i2.5

Mugerwa, H., Rey, M. E., Alicai, T., Ateka, E., Atuncha, H., Ndunguru, J., et al. (2012). Genetic diversity and geographic distribution of *Bemisia tabaci* (Gennadius) (Hemiptera: a leyrodidae) genotypes associated with cassava in East Africa. *Ecol. Evol.* 2, 2749–2762. doi: 10.1002/ece3.379

Mugerwa, H., Seal, S., Wang, H.-L., Patel, M. V., Kabaalu, R., Omongo, C. A., et al. (2018). African ancestry of New World, *Bemisia tabaci*-whitefly species. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-20956-3

Munguti, F. M., Kilalo, D. C., Nyaboga, E. N., Wosula, E. N., and Macharia, I. Mwango'mbe, A.W. (2021). Distribution and molecular diversity of whitefly species colonizing cassava in Kenya. *Insects* 12, 875. doi: 10.3390/insects12100875

Mwatuni, F., Ateka, E., Karanja, L., Mwaura, S., and Obare, I. (2015). Distribution of cassava mosaic geminiviruses and their associated DNA satellites in Kenya. *Am. J. Exp. Agric.* 9, 1–12. doi: 10.9734/AJEA/2015/18473

Mwebaze, P., Macfadyen, S., De Barro, P., Bua, A., Kalyebi, A., Tairo, J., et al. (2022). Impacts of Cassava Whitefly pests on the productivity of East and Central African smallholder farmers. *J. Dev. Agric. Econ.* 14, 60–78. doi: 10.5897/JDAE2022.1330

Ndunguru, J., De León, L., Doyle, C. D., Sseruwagi, P., Plata, G., Legg, J. P., et al. (2016). Two novel DNAs that enhance symptoms and overcome CMD2 resistance to cassava mosaic disease. *J. Virol.* 90, 4160–4173. doi: 10.1128/JVI.02834-15

Ntawuruhunga, P., Okao-Okuja, G., Bembe, A., Obambi, M., Mvila, J. A., Legg, J. P., et al. (2007). Incidence and severity of cassava mosaic disease in the Republic of Congo. *Afr. Crop Sci. J.* 15. doi: 10.4314/acsj.v15i1.54405

Ombiro, G. S. O. (2016). *Detection of Cassava Viruses from Elite Genotypes and Characterization of Cassava mosaic Begomoviruses from Farmers' Fields in Kenya*. Juja: JKUAT.

Pan, L.-L., Chi, Y., Liu, C., Fan, Y.-Y., and Liu, S.-S. (2020). Mutations in the coat protein of a begomovirus result in altered transmission by different species of whitefly vectors. *Virus Evol.* 6, veaa014. doi: 10.1093/ve/veaa014

Patil, B. L., and Fauquet, C. M. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Mol. Plant Pathol.* 10, 685–701. doi: 10.1111/j.1364-3703.2009.00559.x

Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S., Fauquet, C. M., et al. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J. Gen. Virol.* 82, 655–665. doi: 10.1099/0022-1317-82-3-655

Rajabu, C., Kennedy, G., Ndunguru, J., Ateka, E., Tairo, F., Hanley-Bowdoin, L., et al. (2018). Lanai: a small, fast growing tomato variety is an excellent model system for studying geminiviruses. *J. Virol. Methods* 256, 89–99. doi: 10.1016/j.jviromet.2018.03.002

Rana, V. S., Popli, S., Saurav, G. K., Raina, H. S., Chaubey, R., Ramamurthy, V., et al. (2016). A *Bemisia tabaci* midgut protein interacts with begomoviruses and plays a role in virus transmission. *Cell. Microbiol.* 18, 663–678. doi: 10.1111/cmi.12538

Ranawaka, B., Hayashi, S., Waterhouse, P. M., and De Felippes, F. F. (2020). *Homo sapiens*: the superspreader of plant viral diseases. *Viruses* 12, 1462. doi: 10.3390/v12121462

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., Chatterjee, S., et al. (2019). Climate change has likely already affected global food production. *PLoS ONE* 14, e0217148. doi: 10.1371/journal.pone.0217148

Rey, C., and Vanderschuren, H. (2017). Cassava mosaic and brown streak diseases: current perspectives and beyond. *Ann. Rev. Virol.* 4, 429–452. doi: 10.1146/annurev-virology-101416-041913

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Sánchez-Campos, S., Domínguez-Huerta, G., Díaz-Martínez, L., Tomás, D. M., Navas-Castillo, J., Moriones, E., et al. (2018). Differential shape of geminivirus mutant

spectra across cultivated and wild hosts with invariant viral consensus sequences. *Front. Plant Sci.* 9, 932. doi: 10.3389/fpls.2018.00932

Sastry, K. S. (2013). "Plant virus transmission through vegetative propagules (asexual reproduction)," in *Seed-borne Plant Virus Diseases* (Cham: Springer), 285–305. doi: 10.1007/978-81-322-0813-6_9

Saurav, G. K., Rana, V. S., Popli, S., Daimei, G., and Rajagopal, R. (2019). A thioredoxin-like protein of *Bemisia tabaci* interacts with coat protein of begomoviruses. *Virus Genes* 55, 356–367. doi: 10.1007/s11262-019-01657-z

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., et al. (2021). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 49, D10. doi: 10.1093/nar/gkaa892

Schubert, J., Habekuß, A., Kazmaier, K., and Jeske, H. (2007). Surveying cereal-infecting geminiviruses in Germany—diagnostics and direct sequencing using rolling circle amplification. *Virus Res.* 127, 61–70. doi: 10.1016/j.virusres.2007.03.018

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., et al. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87, 651–701. doi: 10.1093/aesa/87.6.651

Sseruwagi, P., Sserubombwe, W. S., Legg, J. P., Ndunguru, J., and Thresh, J. M. (2004). Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. *Virus Res.* 100, 129–142. doi: 10.1016/j.virusres.2003.12.021

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tajebe, L., Boni, S., Guastella, D., Cavalieri, V., Lund, O., Rugumamu, C., et al. (2015). Abundance, diversity and geographic distribution of cassava mosaic disease pandemic-associated *Bemisia tabaci* in Tanzania. *J. Appl. Entomol.* 139, 627–637. doi: 10.1111/jen.12197

Tembo, M., Mataa, M., Legg, J., Chikoti, P. C., and Ntawuruhunga, P. (2017). Cassava mosaic disease: incidence and yield performance of cassava cultivars in Zambia. *J. Plant Pathol.* 93, 681–689. Available online at: http://dx.doi.org/10.4454/jpp.v99i3.3955

Vanitharani, R., Chellappan, P., Pita, J. S., and Fauquet, C. M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *J. Virol.* 78, 9487–9498. doi: 10.1128/JVI.78.17.9487-9498.2004

Were, M. Ndong'a, M., Ogema, V., Mabele, A., Were, H. (2021). Diversity of cassava mosaic disease causal viruses in Kenya. *East Afr. Agric. For. J.* 85, 8–8.

Wu, H., Liu, M., Kang, B., Liu, L., Hong, N., Peng, B., et al. (2022). AC5 protein encoded by squash leaf curl China virus is an RNA silencing suppressor and a virulence determinant. *Front. Microbiol.* 13, 980147. doi: 10.3389/fmicb.2022.980147

Yang, X.-L., Zhou, M.-N., Qian, Y.-J., Xie, Y., and Zhou, X.-P. (2014). Molecular variability and evolution of a natural population of tomato yellow leaf curl virus in Shanghai, China. *J. Zhejiang Univ. Sci. B* 15, 133–142. doi: 10.1631/jzus.B1300110

Zhao, J., Chi, Y., Zhang, X.-J., Lei, T., Wang, X.-W., Liu, S.-S., et al. (2019). Comparative proteomic analysis provides new insight into differential transmission of two begomoviruses by a whitefly. *Virol. J.* 16, 1–12. doi: 10.1186/s12985-019-1138-4

Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J., et al. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J. Gen. Virol.* 78, 2101–2111. doi: 10.1099/0022-1317-78-8-2101

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact