



OPEN ACCESS

EDITED BY

Hussain Touseef,
Matimate Agromart Pvt. Ltd., India

REVIEWED BY

Sudeep Bag,
The University of Georgia, United States
Adyatma Irawan Santosa,
Gadjah Mada University,
Indonesia

*CORRESPONDENCE

Xiaoman She
shexiaoman@gdppri.com
Zifu He
hezf@gdppri.com

[†]These authors have contributed equally to this work

SPECIALTY SECTION

This article was submitted to
Microbe and Virus Interactions with Plants,
a section of the journal
Frontiers in Microbiology

RECEIVED 18 August 2022

ACCEPTED 12 September 2022

PUBLISHED 26 September 2022

CITATION

Farooq T, Hussain MD, Shakeel MT, Riaz H,
Waheed U, Siddique M, Shahzadi I,
Aslam MN, Tang Y, She X and He Z (2022)
Global genetic diversity and evolutionary
patterns among *Potato leafroll virus*
populations.
Front. Microbiol. 13:1022016.
doi: 10.3389/fmicb.2022.1022016

COPYRIGHT

© 2022 Farooq, Hussain, Shakeel, Riaz,
Waheed, Siddique, Shahzadi, Aslam, Tang,
She and He. This is an open-access article
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Global genetic diversity and evolutionary patterns among *Potato leafroll virus* populations

Tahir Farooq^{1†}, Muhammad Dilshad Hussain^{2†},
Muhammad Taimoor Shakeel³, Hasan Riaz⁴,
Ummara Waheed⁵, Maria Siddique⁶, Irum Shahzadi⁷,
Muhammad Naveed Aslam³, Yafei Tang¹, Xiaoman She^{1*} and
Zifu He^{1*}

¹Guangdong Academy of Agricultural Sciences, Plant Protection Research Institute and Guangdong Provincial Key Laboratory of High Technology for Plant Protection, Guangzhou, China, ²State Key Laboratory for Agro-Biotechnology, and Ministry of Agriculture and Rural Affairs, Key Laboratory for Pest Monitoring and Green Management, Department of Plant Pathology, China Agricultural University, Beijing, China, ³Department of Plant Pathology, Faculty of Agriculture & Environment, The Islamia University of Bahawalpur, Bahawalpur, Pakistan, ⁴Institute of Plant Protection, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan, ⁵Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan, ⁶Department of Environmental Sciences, COMSATS University Islamabad, Abbottabad, Pakistan, ⁷Department of Biotechnology, COMSATS University Islamabad, Abbottabad, Pakistan

Potato leafroll virus (PLRV) is a widespread and one of the most damaging viral pathogens causing significant quantitative and qualitative losses in potato worldwide. The current knowledge of the geographical distribution, standing genetic diversity and the evolutionary patterns existing among global PLRV populations is limited. Here, we employed several bioinformatics tools and comprehensively analyzed the diversity, genomic variability, and the dynamics of key evolutionary factors governing the global spread of this viral pathogen. To date, a total of 84 full-genomic sequences of PLRV isolates have been reported from 22 countries with most genomes documented from Kenya. Among all PLRV-encoded major proteins, RTD and P0 displayed the highest level of nucleotide variability. The highest percentage of mutations were associated with RTD (38.81%) and P1 (31.66%) in the coding sequences. We detected a total of 10 significantly supported recombination events while the most frequently detected ones were associated with PLRV genome sequences reported from Kenya. Notably, the distribution patterns of recombination breakpoints across different genomic regions of PLRV isolates remained variable. Further analysis revealed that with exception of a few positively selected codons, a major part of the PLRV genome is evolving under strong purifying selection. Protein disorder prediction analysis revealed that CP-RTD had the highest percentage (48%) of disordered amino acids and the majority (27%) of disordered residues were positioned at the C-terminus. These findings will extend our current knowledge of the PLRV geographical prevalence, genetic diversity, and evolutionary factors that are presumably shaping the global spread and successful adaptation of PLRV as a destructive potato pathogen to geographically isolated regions of the world.

KEYWORDS

Potato leafroll virus, Polerovirus, phylogenetics, recombination, mutation, evolution, selection pressure, intrinsically disordered proteins

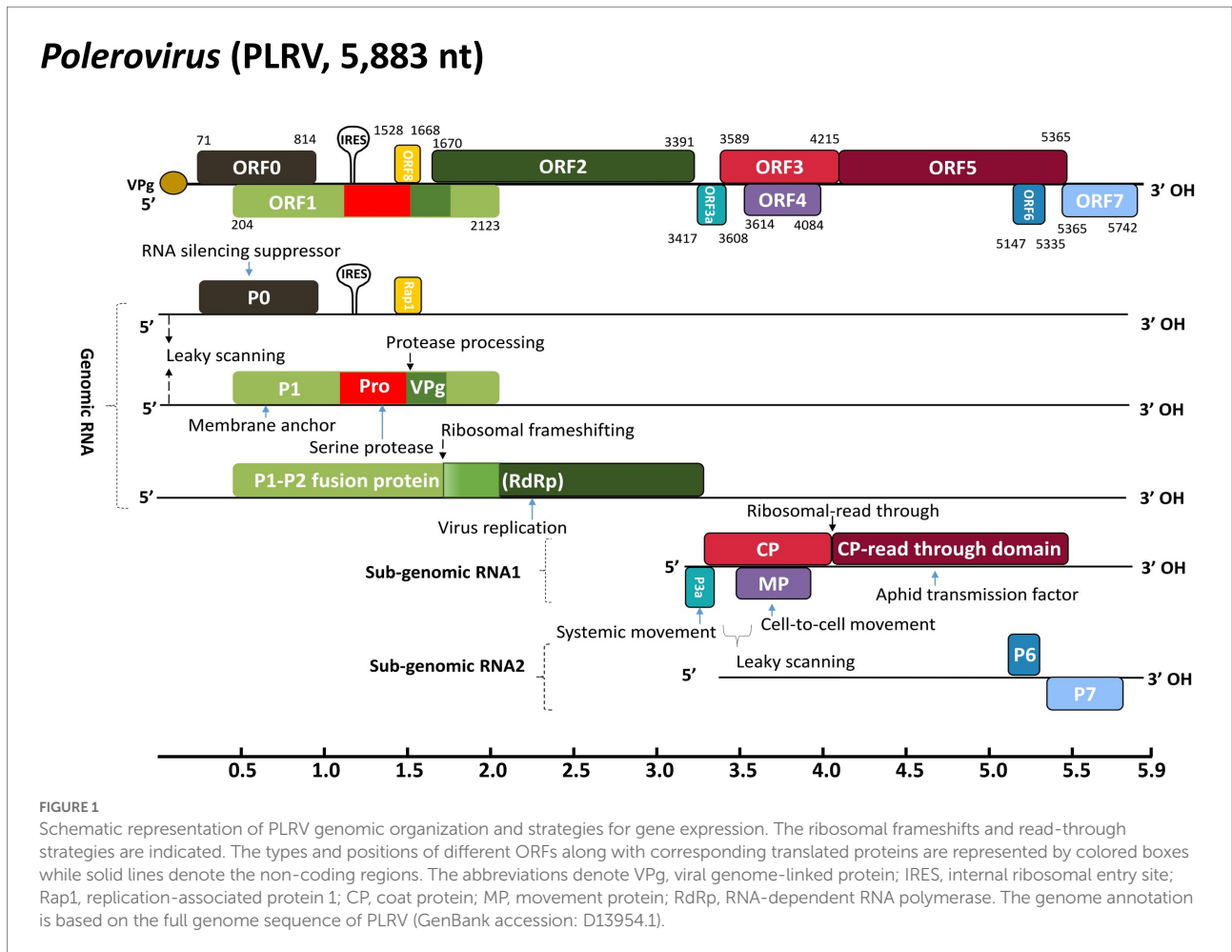
Introduction

The genetic diversity facilitates virus evolution and adaptation to new environments and regulates acute viral infections to species from all of life's domains being parasitized. Factually, massive diseases around the globe suspected to be caused by viruses have been documented for millennia (Wasik and Turner, 2013; Kumar et al., 2017; Gelbart et al., 2020; Rubio et al., 2020; Pascall et al., 2021; Harvey and Holmes, 2022). Potato leafroll Polerovirus (PLRV), from the genus *Polerovirus* and the family *Solemoviridae*, consists of a 5.3–5.7 kb long positive sense (+) monopartite single-stranded RNA [(+)ssRNA] genome with virus protein genome-linked (VPg) cap bounded at the 5' end and an OH group at the 3' end without poly(A) tail or tRNA-like organization (LaTourrette et al., 2021; Walker et al., 2021). Typically, the PLRV genome contains 6–7 overlapping open reading frames (ORFs), which are organized into genomic and sub-genomic RNAs (Figure 1; Krueger et al., 2013; Sömera et al., 2015; Barrios Barón et al., 2017). The RNA-dependent RNA polymerase (RdRp), expressed via ribosomal frameshifting, is a translational fusion of ORF1 that encodes P1 and ORF2, which encodes P2. Additionally, a Rap1 translation initiates through a peculiar internal ribosome entry site (IRES) around 1,500 nt downstream of the 5' end of the gRNA. From sub-genomic RNA 1 (sgRNA1), a capsid protein (CP), involved in virion formation, vector transmission, and virus movement, is encoded from ORF3 (Kaplan et al., 2007; Smirnova et al., 2015). Subsequently, ORF3 extends and ribosomes incorporate one amino acid and continue to translate ORF5 into a CP-read through domain (CP-RTD) as a fusion protein of the translational fusion of ORF3 and ORF5, which is involved in vector transmission and virus movement (Peter et al., 2008; Boissinot et al., 2014; Xu et al., 2018). Furthermore, leaky scanning of sgRNA1 results in the expression of P3a from ORF3a, reported for virus long-distance movement, and P4 from ORF4, a phloem restricted or cell-to-cell movement protein. In addition, PLRV encodes P6 (ORF6) and P7 (ORF7) from the sub-genomic RNA 2 (sgRNA2; Figure 1). The P7 protein plays the most important role in the enhancement of aphid fecundity (Patton et al., 2020). Interestingly, the characteristic feature of poleroviruses is the presence of ORF0 that encodes P0 protein, significantly involved in the viral suppressing of RNA silencing (VSR). The P0 with CP and VPg also contributes to vector specificity (Pazhouhandeh et al., 2006; Baumberger et al., 2007; Csorba et al., 2010; Patton et al., 2020). There are more than 50 viruses that infect potatoes (Kreuze et al., 2020); among them, PLRV is a major pathogen of potato (*Solanum tuberosum*) and transmitted by aphids (*Myzus persicae*) being dreadfully responsible for annual production losses of more

than 20 million tones globally (Kumari et al., 2020; Patton et al., 2020). It is ranked the second most prevalent pathogen in Asia, Africa, Europe, North America, and South America, threatening the sustainable food production system (CABI, 2010). In most cases, viral interaction may occur in the host as the result of mixed viral infection leading to disease severe epidemics. PLRV in association with potato potyvirus Y (PVY) causes additional losses to the marketable potato production industry (Palukaitis, 2012; Valkonen, 2015; Okeyo, 2017; Byarugaba et al., 2020).

The frequent emergence of new viral diseases is primarily due to the ability of the viruses to rapidly evolve. Mostly, viruses retain genomic flexibility in order to adapt to various hosts and vectors (Rantalainen et al., 2011; Garcia-Ruiz, 2018; Nigam and Garcia-Ruiz, 2020; Rubio et al., 2020). Virus evolution and host adaptation are merely determined through genetic diversity among viral populations (Obenauer et al., 2006; Moury and Simon, 2011; Gelbart et al., 2020; Nigam and Garcia-Ruiz, 2020; Rubio et al., 2020; Farooq et al., 2021; LaTourrette et al., 2021). This viral evolutionary process, which is mediated by RNA recombination and preferential accumulation of mutations in certain portions of the genome, leads to the emergence of new viral strains with exceptional characteristics (Obenauer et al., 2006; Moury and Simon, 2011; Dombrovsky et al., 2013; Ibaba et al., 2017; Nigam and Garcia-Ruiz, 2020; LaTourrette et al., 2021). In general, poleroviruses evolve due to the most common genetic variations, which occur in an area encoding VPg, RdRp, and CP, and in a non-coding intergenic region between ORF2 and ORF3 at 5' UTR of the sub-genomic RNA 1 (Pagán and Holmes, 2010; Dombrovsky et al., 2013; Ndikumana et al., 2017; Kwak et al., 2018; LaTourrette et al., 2021). Previous studies have revealed that poleroviruses exhibit higher single-nucleotide polymorphisms (SNPs) on ORFs encoding P0, P1, and CP-RTD and lower SNPs between ORFs, encoding P2 through P4, with exception of a conserved region of P2 within the P1–P2 (Huang et al., 2005; Delfosse et al., 2021; LaTourrette et al., 2021).

Since the PLRV genome comprises overlapping ORFs, mutations may affect various proteins to mediate virus evolution. However, comprehensive knowledge regarding the geographical distribution, standing genetic diversity, and evolutionary patterns existing among global PLRV populations is currently not available, which is crucial for designing sustainable disease management strategies. To fulfill the knowledge gap, we vigorously analyzed the recent global biodiversity, genomic variations, evolutionary endpoints, and the patterns of disordered proteins to gain further insights into the genetic complexity and molecular variability among PLRV populations. To date, with the advances in next-generation sequencing technology, a total of 84 full-length genome sequences of PLRV isolates have been reported across 22 different countries.



PLRV isolates revealed a high level of genetic variability while the distribution patterns of recombination breakpoints across various genomic regions of PLRV isolates remained variable and most of its populations are evolving under strong purifying selection. These findings will expand our knowledge of PLRV geographical prevalence, genetic diversity, and evolutionary forces that are presumably governing the continual PLRV global spread and successful adaptation to different ecosystems.

Materials and methods

Acquisition of full-length PLRV genomic sequences

A total of 84 PLRV full-length genomic RNA sequences were retrieved from the GenBank database¹ on 20 July 2022. Detailed information about the attributes (accession, isolate, country of origin, host, and sequence length) of these sequences is given in [Supplementary Table S1](#). These 84 PLRV sequences were selected

for the subsequent analyses based on the criteria; partial sequences (<95% coverage of the reference genome), sequences with >2.5% of unknown characters, and sequences with <90% similarity with reference genome have been excluded.

Multiple sequence alignments (MSAs) and phylogenetic analysis

The MSAs were prepared by aligning 84 full-length genomic RNA sequences of globally reported PLRV isolates derived from the GenBank database ([Supplementary Table S1](#)), using the MUSCLE tool in the software Geneious Prime version 9.0.2. Likewise, alignments of the individual PLRV genes (*P0*, *P1*, *RdRp*, *CP*, *CP-RTD*, *MP*, and *RTD*) among corresponding genes of the globally reported PLRV isolates were performed. All alignments were manually analyzed and adjusted (when necessary) before proceeding to the subsequent analysis. The phylogenetic tree was constructed in the molecular evolutionary genetics analysis (MEGA-X) software by the maximum likelihood (ML) method with 1,000 bootstrap replicates ([Kumar et al., 2018](#)). The tree was visualized and annotated using iTOL ([Letunic and Bork, 2021](#)). Finally, the distribution and matrix of pairwise identities among

¹ www.ncbi.nlm.nih.gov

all PLRV isolates were determined using Sequence Demarcation Tool (SDT) v1.2 (Muhire et al., 2014).

Estimation of nucleotide diversity and haplotype variability indices

The nucleotide diversity or π (represented by the average pairwise number of nucleotide differences per site) was calculated using DnaSP V.5 (Librado and Rozas, 2009). The significant differences in the average nucleotide diversity among all PLRV sequences were estimated by calculation of their 95% bootstrap confidence intervals. A 100-nt sliding window with a step size of 10 nts across the full-length sequences of PLRV was implicated to calculate π . Additional population genetics-related parameters including the number of haplotypes (H), haplotype diversity (H_d), the number of polymorphic sites (S), Watterson's theta (θ), the total number of mutations (Eta), and Tajima's D were also estimated for PLRV genomes and individual coding sequences (P0, P1, RdRp, CP, CP-RTD, MP, and RTD) using DnaSP V.5 (Librado and Rozas, 2009).

Recombination analysis

The occurrence of recombination events across 84 full-length PLRV sequences was investigated by using several methods including Rdp, SisterScan, Bootscan, Chimera, Geneconv, maximum Chi-square, and 3Seq. The recombination analysis was implemented in the recombination detection program (RDP) V.4 (Martin et al., 2015). For all methods, alignments were performed with default settings. The p -values less than the Bonferroni-corrected cutoff (0.05) were used to infer the statistically significant results. The signals detected by at least four methods were regarded as reliable recombination events.

Analysis of positive and negative selection

The identification of potential positively and negatively selected sites in the coding sequences of P0, P1, RdRp, CP, CP-RTD, MP, and RTD was performed by using four distinct methods including single-likelihood ancestor counting (SLAC), partitioning for robust inference of selection, fixed-effects likelihood and random-effects likelihood (Scheffler et al., 2006). All these methods were employed in the adaptive evolutionary tool "Datamonkey" available online at www.datamonkey.org (Pond and Frost, 2005). To exclude the possibility of misleading results, the recombination breakpoints among all PLRV sequences (P0, P1, RdRp, CP, CP-RTD, MP, and RTD) were searched by the implementation of the Genetic Algorithm Recombination Detection (GARD) method (Kosakovsky Pond et al., 2006). Further, to analyze the mode and strength of natural selection pressure acting on the coding sequences of P0, P1, RdRp, CP,

CP-RTD, MP, and RTD, the ratio of non-synonymous to synonymous substitutions (dN/dS) was estimated by SLAC method based on the GARD-corrected phylogenetic trees.

Prediction of intrinsically disordered proteins

Several studies have been performed to predict and validate the disordered protein regions in the proteins of plant-infecting RNA viruses (Charon et al., 2016, 2018; Walter et al., 2019; Byrne et al., 2021; LaTourrette et al., 2021). In our study, the probability of protein disorder was predicted for P0, P1, RdRp, CP, CP-RTD, MP, and RTD proteins using the Protein DisOrder prediction System (PrDOS; Ishida and Kinoshita, 2007). PrDOS combines the disorder of homologous proteins and template and predicts residue disorder by a sliding window analysis of the target protein sequence. The designated PLRV reference genome sequence (NC_001747.1) was used for this purpose. The ordered and disordered protein regions were mapped and distinguished by different colors along with the prediction of the disorder probabilities. A default (0.5) threshold value indicating a false positive (FP) rate of 5% was used.

Results

Geographical prevalence and evolutionary relationships among PLRV isolates

To date, among a total of 84 full-length globally recognized PLRV isolates, 30 isolates have been reported from Kenya, followed by Germany (6), Colombia (5), India (5), China (4), Burundi (4), Peru (4), Canada (4), Bangladesh (3), France (3), Czech Republic (3), Egypt (2), United Kingdom (2), United States of America (1), Argentina (1), Cuba (1), Ireland (1), Spain (1), Australia (1), Netherlands (1), Poland (1) and Zimbabwe (1) (Figures 2A,B). Next, to determine the standing evolutionary relatedness across all PLRV isolates, molecular phylogenetic analysis was performed. The full-length genomic nucleotide-based phylogenetic tree was created using 84 PLRV isolates reported from 22 countries (Supplementary Table S1). The majority of them originated from Kenya, infecting *S. tuberosum*. The phylogenetic analysis revealed that PLRV genomes reported from Colombia exhibited diversity regarding their associated susceptible hosts. Among three isolates that were grouped, two isolates (MN125065.1 and MN125059.1) were associated with *S. quitoense*, while one isolate (MK116549.1) was reported to infect *S. phureja* (Figure 2C). Additionally, there were 14 isolates with missing information regarding the host plants. Of 30 PLRV isolates reported from Kenya, 21 isolates formed a distinct monophyletic group. Interestingly, two PLRV isolates (EU717546.1 and EU313202.2) from the Czech Republic were also included in this clade. Further, all PLRV isolates originating from China shared genetically distinct clades with isolates reported from France, Cuba, and Ireland. The PLRV isolates

originating from India shared clades with those reported from Germany and France indicating their evolutionary relatedness (Figure 2C). The nucleotide similarity index between PLRV isolates ranged between 92.5 and 100% with the lowest similarity (92.5%) observed between D13953.1 and AF453392.1 isolates originating from Australia and Peru, respectively, although the average percentage identity of aligned sequences of all these isolates was >97.7% (Supplementary Figure S2; Supplementary Table S2).

Analysis of genetic diversity among PLRV populations

To comprehend the exact pattern of virus evolution, the comparison of molecular genetic variability among viral populations is mandatory. Therefore, we conducted an in-depth analysis of genetic variability based on all datasets containing 84 isolates of PLRV genes (P0, P1, RdRp, CP, CP-RTD, MP, and RTD). The genetic diversity, determined for the complete genome analysis of PLRV populations, revealed that PLRV had a high number of mutations (Eta=29.23%) with a high haplotype diversity ($H_d=0.999$), an average nucleotide diversity ($\pi=0.02307$), significantly high negative Tajima's D value (-2.23905^{**}), and an average number of segregating/polymorphic sites ($\theta_w=0.05416$; Figure 3; Table 1).

The genetic diversity analysis of individual genes demonstrated that all genes were genetically variable with high numbers of polymorphisms and polymorphic sites and very high haplotype diversity as well as nucleotide diversity. Interestingly, the average number of mutations (Eta) was higher for RTD gene (38.81%), followed by P1 (31.66%), CP-RTD (30.93%), RdRp (29.24%), P0 (22.98%), CP (16.20%), and MP (13.15%; Figure 3A; Table 1). Similarly, RTD gene had the highest average pairwise nucleotide diversity ($\pi=0.0300$), followed by P0 ($\pi=0.0254$), CP-RTD ($\pi=0.0241$), P1 ($\pi=0.0217$), RdRp ($\pi=0.0191$), CP ($\pi=0.0117$), and MP ($\pi=0.0095$), respectively. The calculated average pairwise nucleotide diversity for the PLRV population was $\pi=0.0231$ (Figure 3B; Table 1).

Furthermore, we performed neutrality tests, such as Tajima's D and Fu and Li's D or Fu and Li's F. Fu and Li's D and Fu and Li's F values were significantly negative for the whole genome of PLRV and RdRP than other genes, such as P0, P1, CP-RTD, MP, and RTD (Table 1). Tajima's D value was also highly negative for RdRP (-2.34169), indicating the presence of polymorphic sites among this gene. Likewise, Tajima's D value was highly negative in P1 (-2.29075) as compared to CP-RTD (-2.10956) and RTD (-2.08312). Finally, Tajima's D values remained highly negative for CP (-2.13716) followed by a closer value in MP (-2.08740) trailed by P0 (-1.51524), whereas the PLRV population had highly negative Tajima's D

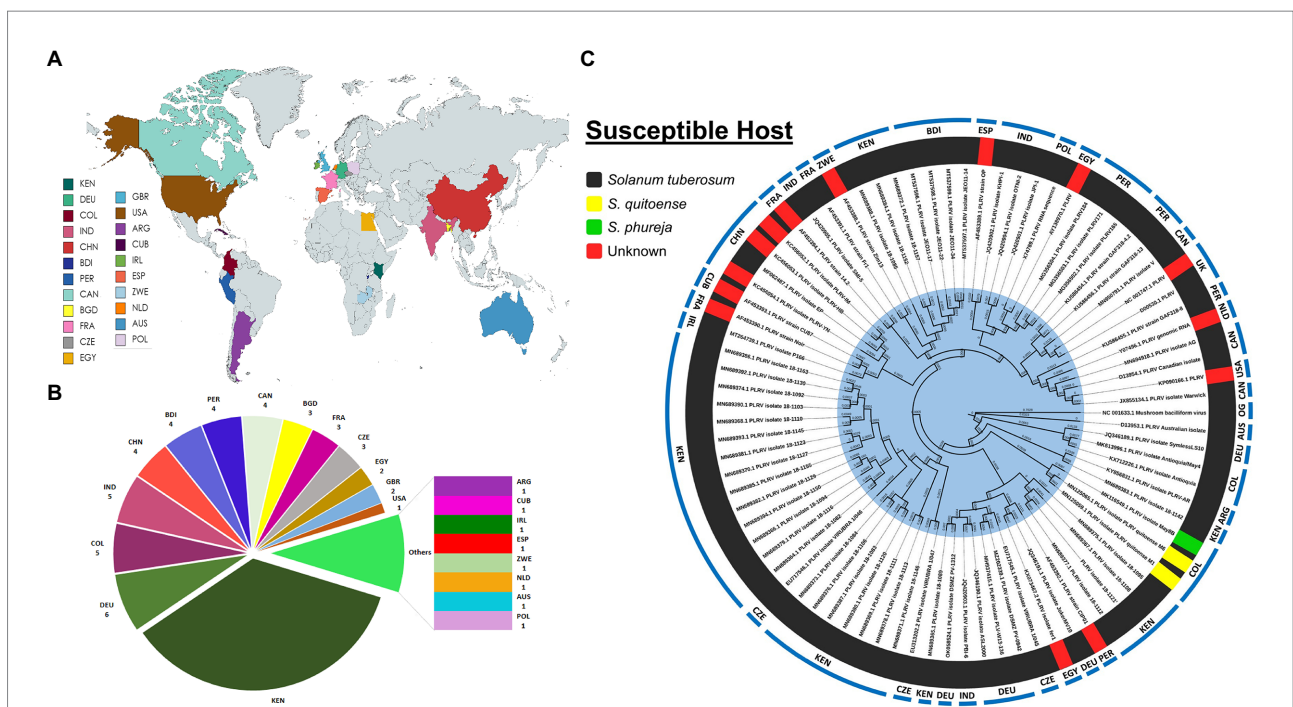


FIGURE 2 (A) Map indicating 22 countries with documented full-genomes of PLRV isolates. Abbreviations include KEN: Kenya, COL: Colombia, CHN: China, IND: India, CZE: Czech Republic, BDI: Burundi, DEU: Germany, PER: Peru, BGD: Bangladesh, FRA: France, EGY: Egypt, CAN: Canada, United States: United States, GBR: United Kingdom, ARG: Argentina, CUB: Cuba, IRL: Ireland, ESP: Spain, POL: Poland, NLD: Netherlands, AUS: Australia, ZWE: Zimbabwe and OG: Outgroup; (B) Pie chart showing the number of PLRV isolates reported from each country; (C) Phylogenetic tree was inferred from 84 full-genome PLRV sequences to investigate their evolutionary relatedness. The nucleotide alignments and construction of the tree were performed using MEGA-X. Bootstrap values indicate that the pattern of branching is supported by 1,000 replicates.

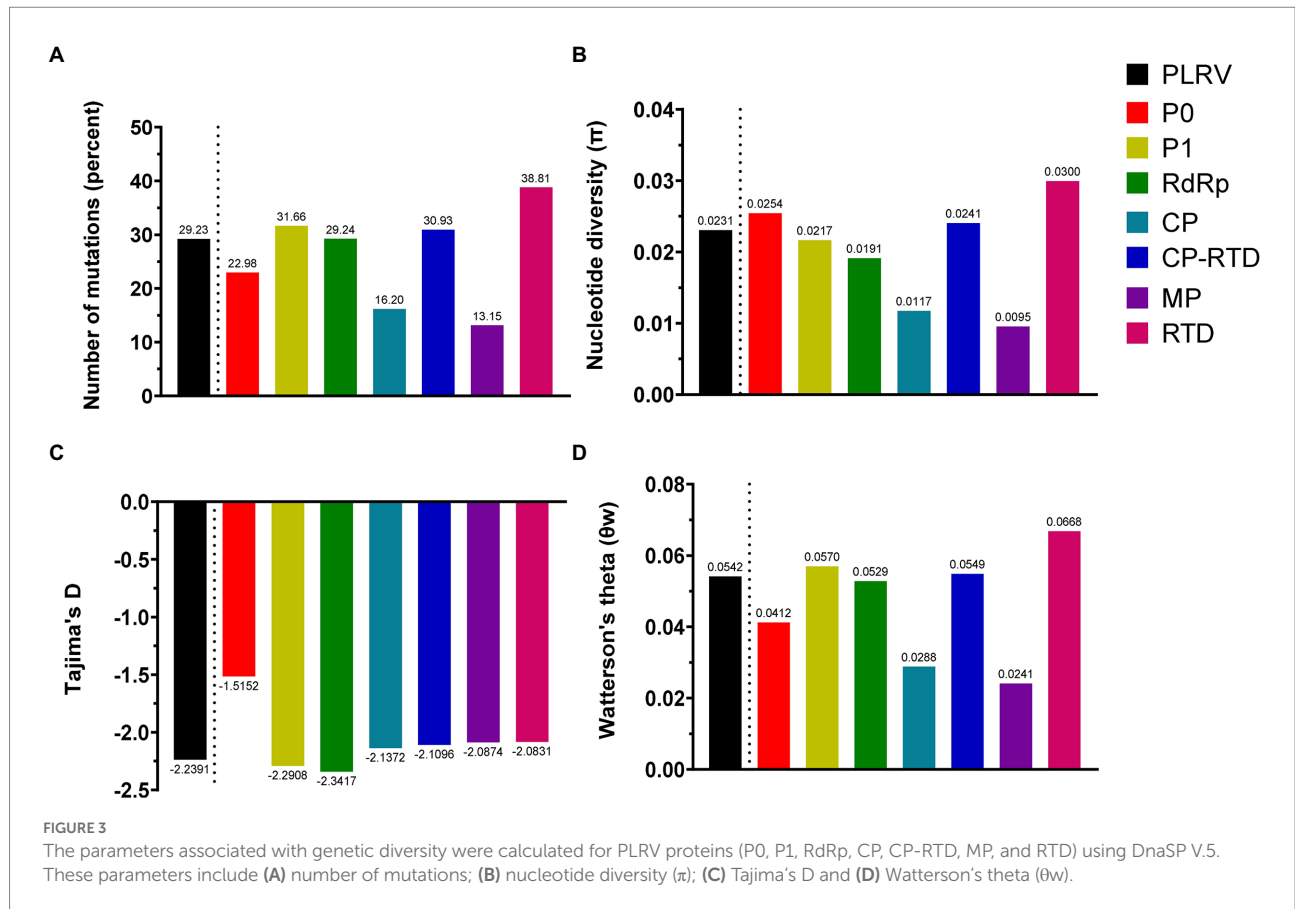


TABLE 1 Analysis of molecular diversity among full genomes and dynamically functional genes of PLRV global isolates.

Dataset	ORF	No. of sequences	No. of analyzed sites	S	H	H_d	π	θ_w		Eta	Neutrality tests		
								Per site	Per sequence		Tajima's D	Fu & Li's D	Fu & Li's F
PLRV	-	84	5,733	1,564	80	0.999	0.02307	0.05416	317.97386	1,676	-2.23905**	-3.83671**	-3.78200**
P0	P0	84	731	151	76	0.998	0.02543	0.04124	30.18751	168	-1.51524 ^{NS}	-3.81178**	-3.42221**
P1	P1	84	1838	524	78	0.998	0.02166	0.05697	104.76294	582	-2.29075**	-4.32812**	-4.15147**
P1-P2	RdRp	84	3,101	820	80	0.999	0.01912	0.05285	163.93846	907	-2.34169*	-4.29898**	-4.15429**
P3	CP	84	617	89	64	0.990	0.01173	0.02884	17.79264	100	-2.13716*	-3.90396**	-3.81299**
P3-P5	CP-RTD	84	2,124	583	79	0.999	0.02407	0.05487	116.55179	657	-2.10956*	-3.08402*	-3.20218**
P4	MP	84	456	55	51	0.962	0.00954	0.02411	10.99545	60	-2.08740*	-3.94768**	-3.83908**
P5	RTD	84	1,412	472	79	0.999	0.02995	0.06683	94.36097	534	-2.08312**	-2.82507*	-3.01130*

S, number of segregating (polymorphic) sites; H, number of haplotypes; Eta, the total number of mutations; H_b , haplotype diversity; θ_w , Watterson's theta and π , nucleotide diversity.

*Statistically significant ($p \leq 0.05$),

**Statistically significant ($p \leq 0.02$), and

^{NS}Statistically non-significant ($p > 0.10$).

value (-2.23905), showing the existence of extensive polymorphic sites in the PLRV population. These observations statistically remained significant, except for the observation of P0 (Figure 3C; Table 1).

In addition, the average number of segregating/polymorphic sites (θ_w) was higher in RTD gene ($\theta_w=0.0668$) and P1 ($\theta_w=0.0570$) as compared to CP-RTD ($\theta_w=0.0549$), RdRp

($\theta_w=0.0529$), and P0 ($\theta_w=0.0412$). However, the lowest values were found in the CP ($\theta_w=0.0288$) and MP ($\theta_w=0.0241$), while the value of Watterson's theta for the full-length PLRV sequences was 0.0542 (Figure 3D; Table 1). These findings indicate that each gene has high genetic diversity displaying a high proportion of haplotypes, although P0 and CP-RTD appeared to be the most genetically diverse genes in the genome of PLRV.

Recombination is predominantly driving the genomic diversity of PLRV

To investigate the existence of recombination events among geographically isolated PLRV populations, we employed all seven methods, including RDP, GeneConv, Bootscan, MaxChi, Chimera, SisScan, and 3SEQ (Figure 4; Table 2). For all analyzed datasets, only the exceptional recombination events detected by at least four methods supported by a p -value of <0.001 were considered highly significant. The results showed that 63/84 isolates had detectable recombination events. Among them, the PLRV population that originated from Kenya was more likely to be prone to recombinational changes. For instance, a frequent recombination event was detected among 31 sequences of PLRV (p -value = 3.326×10^{-17}), and the predicted beginning and ending breakpoints were located at 1125–5015 nt of the genomic RNA, covering full sequences of RdRp, CP, and MP while partial sequences of P1 and RTD ORFs (Figure 4; Table 2). The major recombinant isolate (MN689367.1) associated with this event was reported from Kenya with KY856831.1 (Argentina) and NC001747.1 (United Kingdom) being major and minor parents, respectively. Furthermore, the second most frequently detected recombination event was found in 12 isolates with a high significance (p -value = 4.688×10^{-14}). The recombination breakpoints were distributed between 42 and 1,125 nt that covered a full portion of ORF0, encoding partial P0 and partial sequence of P1 ORF (Figure 4; Table 2). For this event, the major recombinant sequence was MN689365.1 (Kenya) with major (MN689381.1) and minor (KX712226.1) parents from Kenya and Colombia, respectively. Likewise, a third recombination event was found in 7 isolates (p -value = 2.057×10^{-04}) at 5745–2414 nt positions and covered a partial sequence of P7, 5' UTR, P0 (full), P1 (full), and RdRp (partial). The major recombinant isolate (MN689380.1) originated from Kenya while the major (MN689369.1) and minor (MN689384.1) parental sequences were reported from Kenya (Figure 4; Table 2). Notably, of ten significantly supported recombination events, five were associated with isolated reported from Kenya, while other recombinant isolates were reported from Australia (1), Germany (1), Argentina (1), Colombia (1), and China (1).

Colombian populations of PLRV were also detected to have recombinational changes. One out of five sequences was found to have recombination events (p -value = 1.014×10^{-10}) at 5856–751 nt positions, which partially mapped the dynamically functional P0 and P1 ORFs (Figure 4; Table 2). Moreover, one PLRV isolate (JQ346189.1) from Germany showed recombinational changes (p -value = 8.873×10^{-32}), and the predicted recombination breakpoints were located at 5793–2927 nt positions; mapping approximately half of the PLRV genome and including sub-genomic RNAs (Figure 4; Table 2). The analyses of recombination events and mapping of recombination breakpoints among the PLRV analyzed sequences signify that great diversity exists among viral populations, particularly at genomic and sub-genomic RNAs encoding various functional proteins. Our

results demonstrate that some of the genes have completely recombinant sequences, such as P0, P1, RdRp, CP, MP, and CP-RTD, while RTD and P6, and P7 have partial sequences under the effect of recombination.

Purifying selection pressure governs The evolution of PLRV

In order to attain a better understanding of the possible role of selection pressure in the evolution of PLRV, we compared the non-synonymous to synonymous substitutions (dN/dS) between the analyzed datasets, which included major functional proteins of PLRV, such as P0, P1, RdRp, CP, CP-RTD, MP, and RTD. Interestingly, the results indicated that the observed diverse variations on the genome of PLRV were being driven both by the positive and negative selection pressure as in addition to negative selection, several positively selected sites were also found among all PLRV genes with variable frequency (Figure 5). Particularly, both positively and negatively selected sites in the coding sequences of these genes were detected. Although most of the genes had their major part of tested sites under negative or purifying selection pressure (dN/dS <1), the impact of positive selection (dN/dS >1) and neutral selection (dN/dS = 1) could not be negated. However, the probability of positively and neutrally selected sites remained much lower than the negatively selected sites. Specifically, the proportion of negatively selected sites in P0 (13/239), P1 (24/592), RdRp (62/986), CP (9/204), CP-RTD (50/685), MP (5/147), and RTD (33/435) was higher compared to the positively selected sites observed in P0 (4/239), P1 (19/592), RdRp (34/986), CP-RTD (10/205), CP (1/204), MP (1/147), and RTD (19/435; Figure 5). On the other hand, P1, RdRp, CP, CP-RTD, and RTD proteins had very small and non-significant proportions of neutral selection (Figures 5B–E,G). However, P0 and MP showed no signs of neutral selection (Figures 5A,F). The considerably higher percentages of positively selected residues in RTD (4.36%) followed by RdRp (3.44%) indicate their functional importance in the relevant proteins. In conclusion, all tested proteins were observed to evolve under purifying and positive selection pressures with the former having a significant impact because the percentage of negatively selected sites was higher (Figure 5), compared with a significantly lower percentage of positive selection pressure found to variably affect all PLRV proteins (Figure 5).

Comparison of disordered residues among PLRV proteins

The intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) are well-known to participate in protein–protein interactions (PPIs) involving multiple or diverse binding partners. Additionally, these unstructured proteins

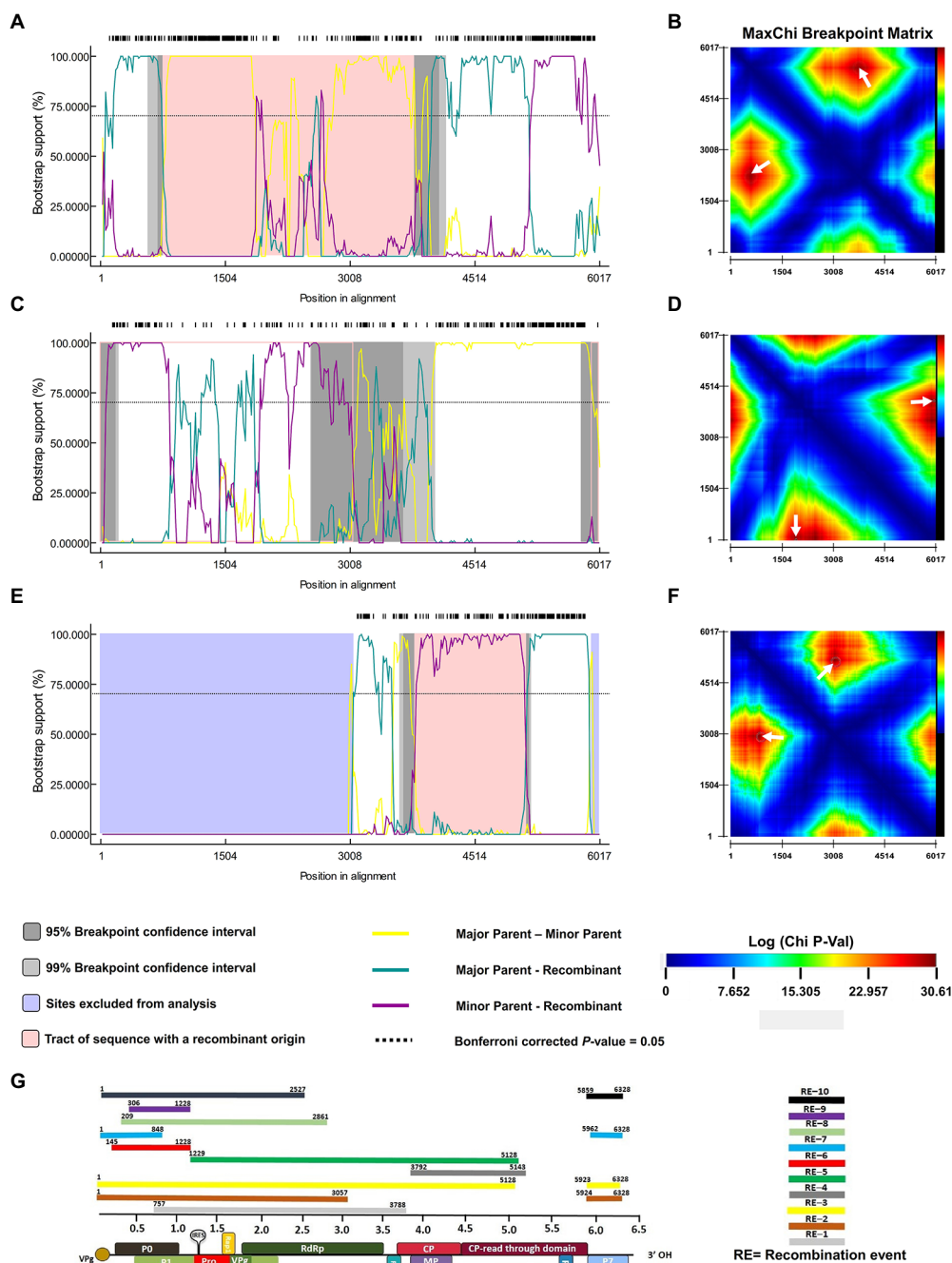


FIGURE 4

The most frequent recombination events in the PLRV genome (indicated by pink outline) were detected by using RDP V.4. Panels (A,C,E) represent the recombination events detected among 14, 8, and 2 isolates, respectively. The x-axis indicates the position in alignment while the y-axis denotes the percentage value for bootstrap support. Panels (B,D,F) represent the probabilities of recombination breakpoints best supported by *p*-values (displayed by the color key). Dark red peaks marked with white arrows indicate the statistically optimal position of recombination breakpoint pairs. Panel G is the schematic presentation of all significantly supported recombination events found with their corresponding position on the PLRV genome.

regulate several vital processes including the assembly of protein complexes, transcription, and translation (Uversky, 2002; Szilágyi et al., 2008). The results of protein disorder prediction showed that CP-RTD contained the highest percentage (48%) of disordered amino acids followed by CP (44%), MP (37%), P1 (35%), RdRp

(29%), and RTD (19%). Notably, P0 contained the lowest percentage (5.4%) of disordered residues (Supplementary Figure S1). A visual summary of the disorder probability and position in the target amino acid sequences is shown in Figure 6. We further compared the N- and C-terminals of all proteins to investigate the

TABLE 2 Recombination events with high significance as detected by RDP in globally-reported PLRV isolates.

Recombination event	Sequences detected with recombination	Recombinant sequence ¹		Recombination breakpoints with (without) gaps		ORF (without) gaps	Parental sequences		Detection Methods ²	P-value ³
		Isolate	Country	Begin	End		Major	Minor		
1	1	DI13953.1	Australia	757 (653)	3,788 (3674)	P0, P1, RdRp	MK613996.1	Unknown	RGBMCS ₃	2.468 × 10 ⁻¹¹
2	1	JQ346189.1	Germany	5,924 (5793)	3,057 (2927)	P0, P1, RdRp	Unknown	JQ420903.1	RGBMCS ₃	8.873 × 10 ⁻³²
3	1	KY856831.1	Argentina	5,923 (5809)	5,128 (5014)	P0, P1, RdRp, MP, CP, RTD, P6, P7	JQ420903.1	JQ346189.1	RGBMCS ₃	3.024 × 10 ⁻¹⁹
4	1	MN689383.1	Kenya	3,792 (3680)	5,143 (5031)	MP, CP, RTD	MN689384.1	JQ346189.1	RGBMCS ₃	1.376 × 10 ⁻¹⁸
5	31	MN689367.1	Kenya	1,229 (1125)	5,128 (5015)	P1, RdRp, MP, CP, RTD	KY856831.1	NC001747.1	RMCS ₃	3.326 × 10 ⁻¹⁷
6	12	MN689365.1	Kenya	145 (42)	1,228 (1125)	P0, P1	MN689381.1	KX712226.1	RGBMCS ₃	4.688 × 10 ⁻¹⁴
7	3	KX712226.1	Colombia	5,962 (5856)	848 (751)	P0, P1	Unknown	MN125059.1	RGBMCS ₃	1.014 × 10 ⁻¹⁰
8	5	MF062487.1	China	209 (99)	2,861 (2742)	P0, P1, RdRp	KY856831.1	KC456053.1	RGMCS ₃	1.516 × 10 ⁻⁰⁴
9	1	MN689382.1	Kenya	306 (202)	1,228 (1124)	P0, P1	MN689381.1	JQ420905.1	MCS ₃	3.199 × 10 ⁻⁰³
10	7	MN689380.1	Kenya	5,859 (5745)	2,527 (2414)	P0, P1, RdRp, RTD, P7,	MN689369.1	MN689384.1	MCS ₃	2.057 × 10 ⁻⁰¹

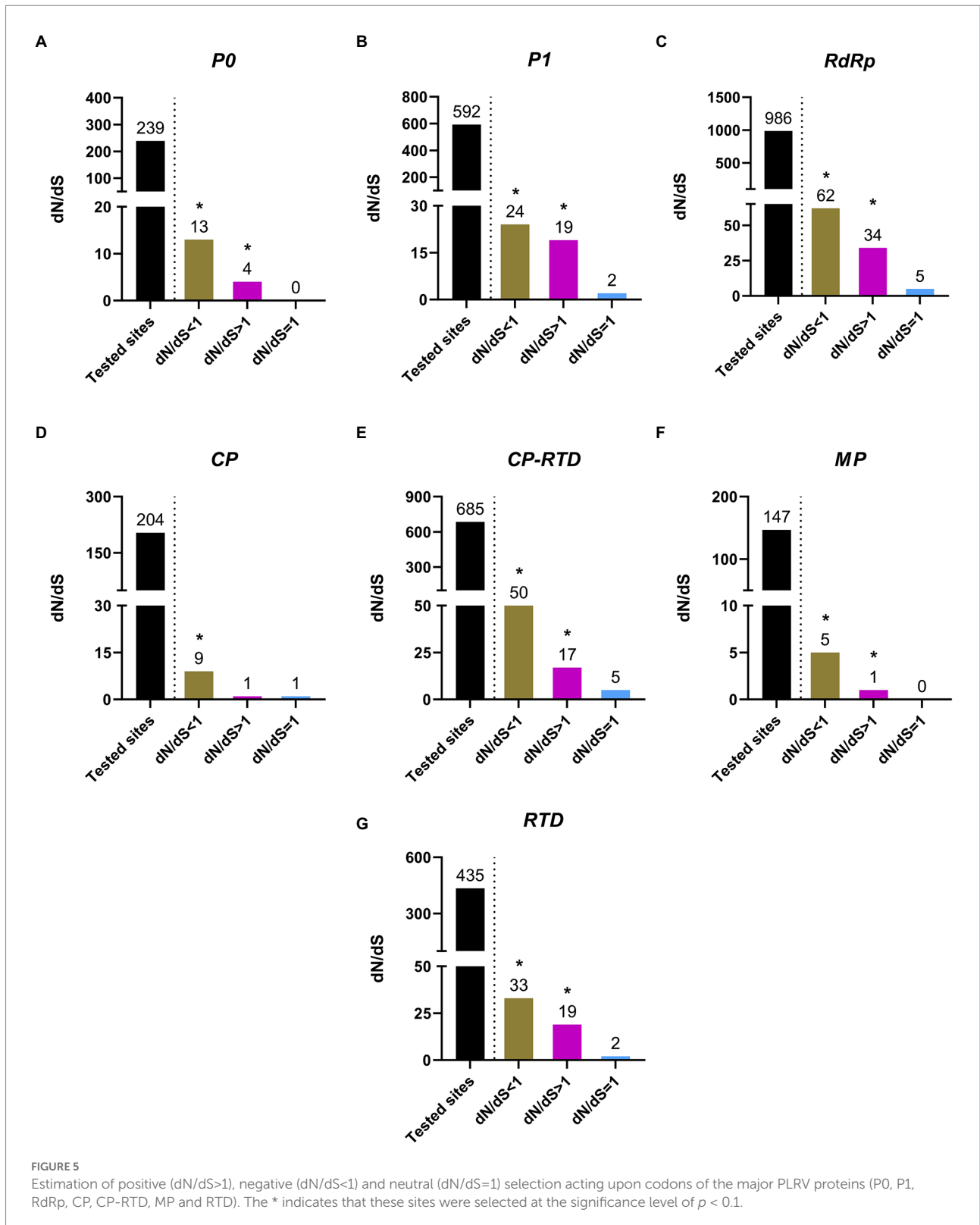
¹R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, CHIMERA; S, SisaScan; 3, 3SEQ. ²The described P-value corresponds to the calculated P-value for the event in question, detected by the program in bold and underlined.

patterns of disordered residues. Results demonstrated that of 5.4% disordered residues of P0, approximately half (2.5%) were confined to the N-terminus, while the rest (2.9%) were detected at the C-terminus of the protein (Figure 7A). Likewise, in P1, 31.3% of disordered residues were found at the 3' end, while only a small proportion (3.2%) of disordered amino acids was located at the 5' half of the protein (Figure 7B). Moreover, RdRp displayed a similar pattern where 16.1% disordered residues at the 3' end while 13% at the 5' half were located (Figure 7C). However, CP exhibited a contrasting pattern of distribution for the disordered residues. We found that a higher percentage (32.2) of disordered residues was located at the 5' region as compared to a lower percentage (12.2) at the 3' end (Figure 7D). Additionally, for CP-RTD and MP proteins, the disordered residues were 20.7 and 30.9% at the 5' half, while 27 and 6% were positioned at the 3' end, respectively (Figures 7E,F). Finally, RTD showed a pattern similar to P0 where 9.8 and 8.8% of disordered residues were located at the 5' and 3' halves, respectively (Figure 7G).

Discussion

In the present study, we performed a comprehensive analysis of the current geographical prevalence, genomic variations, recombination, and evolutionary endpoints associated with global PLRV populations. We also investigated and compared the disordered amino acids among all PLRV-encoded proteins (P0, P1, RdRp, CP, CP-RTD, MP, and RTD). We found that among full-length PLRV genomes reported from 22 countries, the PLRV isolates originating from Kenya displayed the highest genetic variability. The most significantly supported recombination event was also associated with PLRV isolates reported from Kenya. Notably, in the case of the observed type of infected host and disease prevalence, the influence of sampling biasness together with other factors cannot be ruled out (Lacroix et al., 2016). Further analysis at the individual gene level demonstrated that RTD and P1 contain the highest number of mutations. We found a considerable yet variable number of positively selected sites among all genes (P0, P1, RdRp, CP, CP-RTD, MP, and RTD). While a significant influence of purifying (negative) selection pressure seems to govern the evolution of the majority of PLRV proteins, a considerably higher percentage of positively selected amino acids in RdRp and RTD demonstrate the functional importance of these proteins. Additional results revealed that CP-RTD protein contains approximately half (48%) of amino acids identified as disordered. Given that previous studies on PLRV are either focused on one region and a single ORF (Zarghani et al., 2012) or include a few isolates (LaTourrette et al., 2021), our findings will provide detailed information on the aforementioned aspects of PLRV genetic diversity.

As mentioned earlier, PLRV belongs to polioviruses that are exclusively vectored by aphid species (Kaplan et al., 2007). During the insect-mediated transmission and infection processes, several hosts (insects and/or plants) and



environmental factors and their interactions impose certain evolutionary constraints on these viruses (Wan et al., 2015; Li et al., 2016; Nigam and Garcia-Ruiz, 2020). In the course of virus-host co-evolution, the viral and host-related factors

mainly regulate the compatible or incompatible interactions (Garcia-Ruiz, 2018). The viral factors are well-known to modulate the host physiology to facilitate the dissemination of virions through insect vector feeding (Mauck et al., 2014).

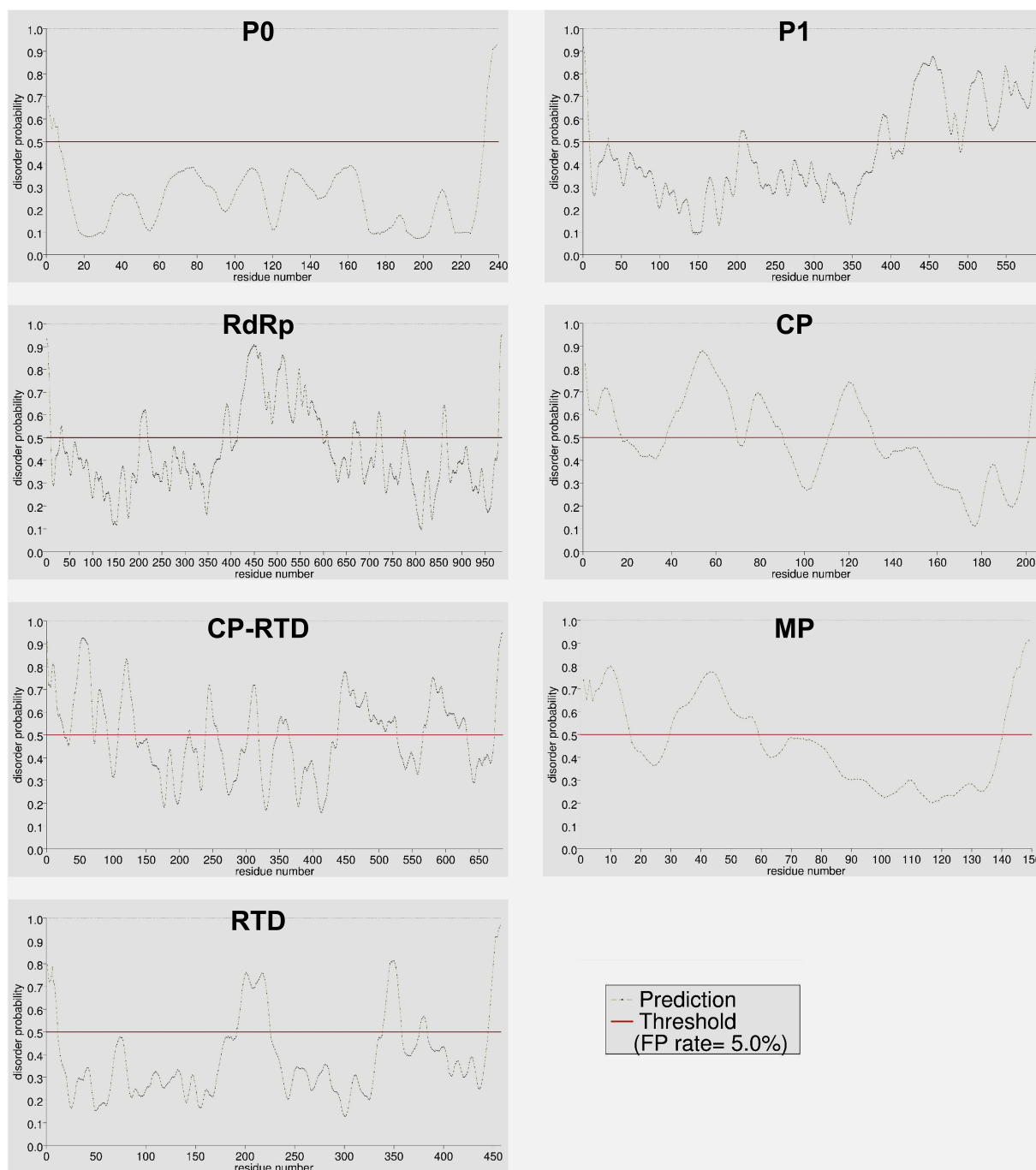
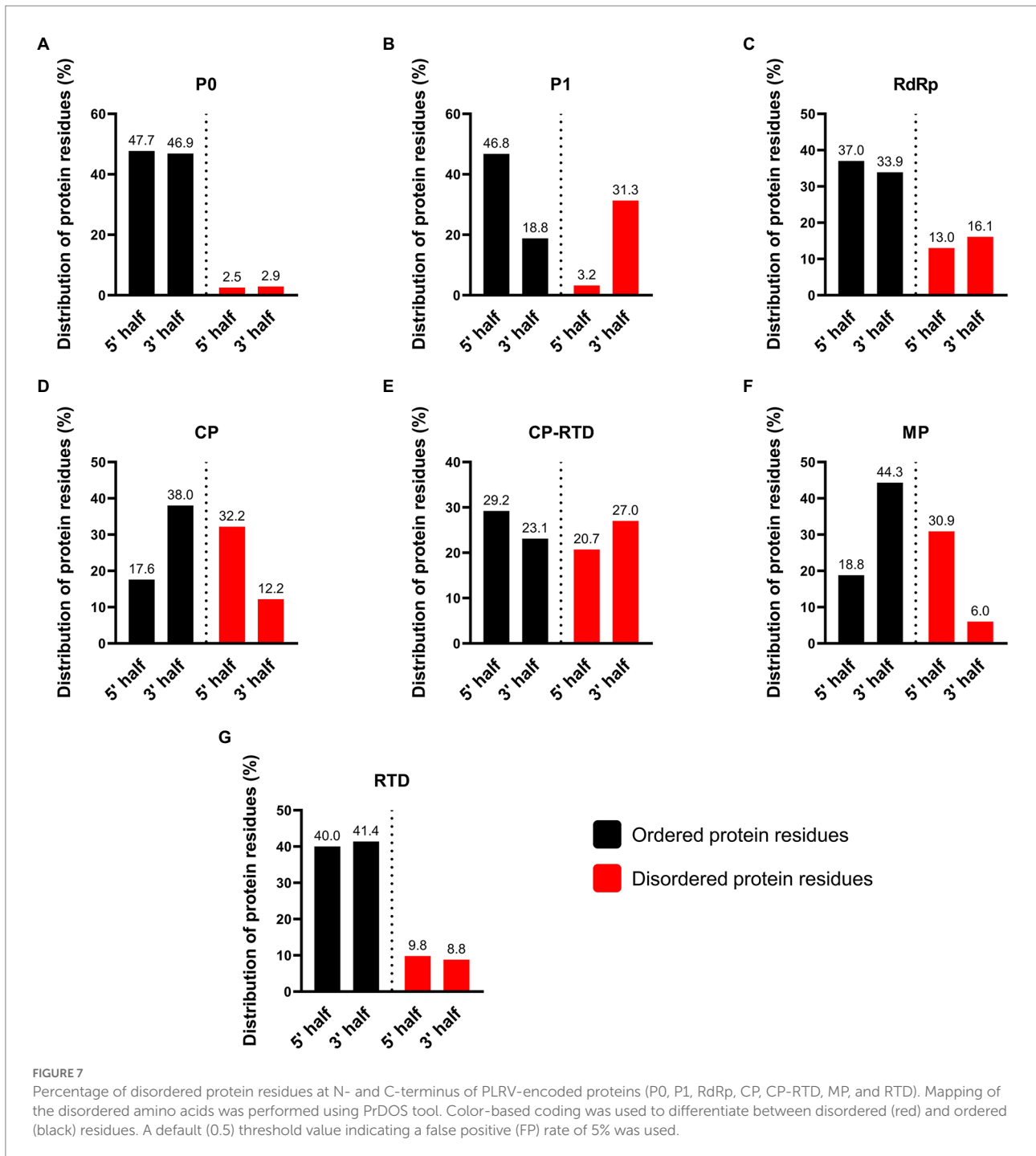


FIGURE 6
 Distribution of ordered and disordered residues across different proteins (P0, P1, RdRp, CP, CP-RTD, MP, and RTD) of PLRV. Color-based coding was used to differentiate between disordered (red) and ordered (black) residues. The horizontal red line denotes a default (0.5) threshold value representing a 5% false positive (FP) rate.

Additionally, climatic changes support the emergence of frequent viral epidemics by facilitating the vector populations into new and geographically isolated regions which ultimately expose new hosts to these viral pathogens (Trebicki, 2020). Thus, to reduce the chances of exclusion from the population during selection by the host (vector/plant) or the

environment, viruses must attain higher levels of fitness, and maintain balanced genomic flexibility and functionality.

The genome-wide profiling of variability among P0, P1, RdRp, CP-RTD, and RTD showed that RTD has the highest percentage of mutations followed by P1, CP-RTD, and RdRp. On the contrary, the accumulation of mutations remained lowest in the ORF



encoding MP. A similar trend was observed in the nucleotide diversity results where RTD displayed a higher value of P_i followed by P0 (Figure 3; Table 1). A recent study involving the combinatory analysis of PLRV and five other most variable poleroviruses demonstrates that P0 and CP-RT regions exhibited a high accumulation of nucleotide substitutions (LaTourrette et al., 2021). Further results showed that a highly negative value of Tajima's D was associated with all genes under study, indicating the presence of excessive low-frequency polymorphism and a

possible expansion in the population size after deviation from neutrality. Notably, different forms of mutations in the genomic RNA of viruses are introduced by RdRp during replication of the viral RNA (García-Arenal et al., 2001). Additionally, recombination events during the replication of viral RNA rapidly generate genetic diversity (García-Ruiz et al., 2018). Both interspecific and intraspecific forms of recombination are frequently found in poleroviruses (Pagán and Holmes, 2010). The mutated, new viral genomes derived from RNA recombination

might exert positive, deleterious, or neutral effects on the fitness of viruses, which consequently lead to the removal or fixation of these genomes in the viral populations (Garcia-Ruiz et al., 2018; Nigam and Garcia-Ruiz, 2020). These new genomes lead to the appearance of new strains/species of poleroviruses enabling them to cause infections in the new hosts (Ibaba et al., 2017). According to this model, it is reasonable to assume that the accumulation of mutations in the PLRV genomes is not random. Rather, preferential accumulation of mutations occurs in the proteins that are key determinants of vector transmission or host adaptation.

Recombination is the most commonly found phenomenon among +ssRNA viruses and facilitates genetic diversity *via* switching RNA genetic segments between viral isolates. This results in the introduction of new, resistance-breaking/virulent strains and host expansion (Nagy, 2008; Traoré et al., 2010; Garcia-Ruiz, 2018). In this study, we found a total of 10 recombination events spanning different regions of the PLRV genome (Table 2). The most frequently detected recombination event was associated with P0 while other recombination hotspots were also detected in other coding sequences including P1, RdRp, and CP-RTD, with variable levels of significance (Figure 4). The differential occurrence of recombination sites at the N- and/or C-terminus of these ORFs and overlapping regions suggest that possibly, different recombination-driven evolutionary histories are associated with these sites. Recent studies on poleroviruses have demonstrated that recombination breakpoints are located across the RdRp and RTD regions of *Brassica yellows virus* (BrYV; Umar et al., 2022) while recombination events in P0 and CP are hypothesized to possibly drive the evolution of *Turnip yellows virus* (TuYV; Umar et al., 2022). Likewise, studies have demonstrated that recombination events are associated with specific hotspots in the P2 and CP regions (Dombrovsky et al., 2013; Kwak et al., 2018). In the future, it will be worth studying how recombination-driven changes modulate the biological functions of these proteins.

RTD had the highest percentage of sites under negative selection followed by CP-RTD and RdRp, as compared to other genes. The highest number of positively selected sites were associated with RTD, followed by RdRp and P1. Noticeably, P0 and MP did not contain sites evolving under neutral selection pressure. Although, the ratios of negative selection associated with each ORF were variable, and there was a considerably high number of positively selected codons as well; it is evident that overall, purifying selection is the major evolutionary pressure acting on the PLRV genome (Figure 5). Likewise, findings from polerovirus-based studies (LaTourrette et al., 2021; Umar et al., 2022) suggest that the negative selection pressure acting on different ORFs of the viral genome might be essential to maintain the protein functionality, ultimately affecting the overall viral fitness.

The disordered protein regions, often referred to as IDPRs or IDPs lack a stable tertiary or three-dimensional structure and proper folding. Owing to their greater plasticity and flexibility, the IDPRs and IDPs are known to play vital roles in various biological functions (Wright and Dyson, 1999; Uversky, 2019). Some of the

important biological functions associated with IDPRs and IDPs are signaling, cell regulation, survival, differentiation, proliferation, and apoptosis (Kozłowski and Bujnicki, 2012; Katuwawala et al., 2019). While some of these proteins are assumed to participate in disease etiology and possibly represent novel drug targets (Hu et al., 2016). In viruses, IDPs are recognized to govern numerous functions including adaptation to the dynamic host-related environment, counteracting host-mediated defense mechanisms, and regulation of gene expression to facilitate viral replication in the host (Gitlin et al., 2014; Xue et al., 2014; Mishra et al., 2020). In the present study, we found that CP-RTD of the PLRV contains a high percentage (48%) of disordered amino acids (Figure 7). The identified disordered regions in CP-RTD encompass domains that essentially contribute to the virion formation, systemic movement of the virus, and aphid-mediated viral transmission (Peter et al., 2008). The presence of disordered regions in the CP-RTD of PLRV elaborates the effect of host-dependent mutations in this region (Peter et al., 2008) and it might also explain that why the viral genome is capable of tolerating high rates of mutation (Tokuriki et al., 2009; Sanjuán et al., 2010). Additional research is imperative to deeply understand the correlation of high mutation rates, accumulation of disordered regions, and rapid evolution of the viral genome.

The dynamic involvement of IDPs in the protein–protein interactions could lead to the expansion of the virus host range (Charon et al., 2018). Researchers have demonstrated that in the PVY genome, IDPs might be associated with virus adaptation either by reducing the fitness cost caused by resistance-breaking mutations or by larger exploration of the evolutionary pathways. Eventually, IDPs positively affect the adaptive capacity of RNA viruses (Charon et al., 2018). From an evolutionary point of view, intrinsically disordered regions (IDRs) are thought to have a higher mutational permissiveness than highly ordered regions (Lafforgue et al., 2022). Thus, IDRs-associated amino acid polymorphism could result in the emergence of adaptive solutions during the selection process (Lafforgue et al., 2022). It has been documented that during the evolution of potyviruses, the IDPRs tend to evolve faster than the ordered regions (Charon et al., 2016). Notably, another study involving an insect-infecting RNA virus (*Nodamura virus*, NoV) demonstrates that rapidly evolving IDRs might act as the reservoir for evolutionary innovation and play vital roles in virus adaptation to new environments (Gitlin et al., 2014). In this context, the IDPs/IDPRs are hypothesized to regulate the adaptive ability of PLRV either by introducing distinct evolutionary pathways or by minimizing the mutation-induced fitness penalty; however, further experiments are imperative to validate this hypothesis.

Conclusion

Our findings provide compelling evidence that global PLRV populations have high genetic diversity and the PLRV-encoded proteins are evolving both under positive and purifying selection with later having a more significant effect. The genome-wide profiling of

variability shows that high mutation and recombination are the main factors governing the rapid evolution of PLRV genomes. The presence of a significantly high number of disordered sites in the CP-RTD region might enable PLRV to attain efficient virion formation, systemic movement, and transmission by aphid vector. Previously less-known, these mechanisms presumably are the major determinants of PLRV adaptation to new environments by broadening the host range and pathogenicity levels. These results lay solid foundations for the planning and implementation of strategies aimed at the timely diagnosis and sustainable management of PLRV.

Data availability statement

Publicly available datasets were analyzed in this study. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary material](#).

Author contributions

TF and MH conceived the idea, acquired and analyzed data, and prepared the original draft. MS, HR, UW, MS, and IS analyzed data. MA, XS, and YT contributed to review and editing of the manuscript. ZH edited the manuscript, acquired funding, and supervised the study. All authors contributed to the article, and read and approved the submitted version.

Funding

This work was funded by Discipline Team Building Projects of Guangdong Academy of Agricultural Sciences in the 14th Five-Year Period (202105TD), the President Foundation of Guangdong

References

- Barrios Barón, M. P., Agrofoglio, Y. C., Delfosse, V. C., Nahiriñak, V., Gonzalez de Urreta, M., Almasia, N. I., et al. (2017). First complete genome sequence of potato Leafroll virus from Argentina. *Genome Announc.* 5, e00628–e00617. doi: 10.1128/genomeA.00628-17
- Baumberger, N., Tsai, C.-H., Lie, M., Havecker, E., and Baulcombe, D. C. (2007). The Ploverovirus silencing suppressor P0 targets Argonaute proteins for degradation. *Curr. Biol.* 17, 1609–1614. doi: 10.1016/j.cub.2007.08.039
- Boissinot, S., Erdinger, M., Monsion, B., Ziegler-Graff, V., and Brault, V. (2014). Both structural and non-structural forms of the Readthrough protein of cucurbit aphid-borne yellows virus are essential for efficient systemic infection of plants. *PLoS One* 9:e93448. doi: 10.1371/journal.pone.0093448
- Byarugaba, A. A., Mukasa, S. B., Barekye, A., and Rubaihayo, P. R. (2020). Interactive effects of potato virus Y and potato Leafroll virus infection on potato yields in Uganda. *Open Agriculture* 5, 726–739. doi: 10.1515/opag-2020-0073
- Byrne, M., Kashyap, A., Esquirol, L., Ranson, N., and Sainsbury, F. (2021). The structure of a plant-specific Partitivirus capsid reveals a unique coat protein domain architecture with an intrinsically disordered protrusion. *Communicat. Biol.* 4:1155. doi: 10.1038/s42003-021-02687-w
- CABI. (2010). *Invasive Species Compendium: Detailed Coverage of Invasive Species Threatening Livelihoods and the Environment Worldwide*. Cab International; London.
- Charon, J., Barra, A., Walter, J., Millot, P., Hébrard, E., Moury, B., et al. (2018). First experimental assessment of protein intrinsic disorder involvement in an Rna virus natural adaptive process. *Mol. Biol. Evol.* 35, 38–49. doi: 10.1093/molbev/msx249
- Charon, J., Theil, S., Nicaise, V., and Michon, T. (2016). Protein intrinsic disorder within the Potyvirus genus: from proteome-wide analysis to functional annotation. *Mol. Biosyst.* 12, 634–652. doi: 10.1039/C5MB00677E
- Csorba, T., Lózsá, R., Hutvágner, G., and Burgyán, J. (2010). Ploverovirus protein P0 prevents the assembly of small Rna-containing Risc complexes and leads to degradation of Argonaute1. *Plant J.* 62, 463–472. doi: 10.1111/j.1365-3113.2010.04163.x
- Delfosse, V. C., Barrios Barón, M. P., and Distéfano, A. J. (2021). What we know about Ploveroviruses: advances in understanding the functions of Ploverovirus proteins. *Plant Pathol.* 70, 1047–1061. doi: 10.1111/ppa.13368
- Dombrovsky, A., Glanz, E., Lachman, O., Sela, N., Doron-Faigenboim, A., and Antignus, Y. (2013). The complete genomic sequence of pepper yellow leaf curl virus (Pylcv) and its implications for our understanding of evolution dynamics in the genus Ploverovirus. *PLoS One* 8:e70722. Epub 2013/08/13. doi: 10.1371/journal.pone.0070722
- Farooq, T., Umar, M., She, X., Tang, Y., and He, Z. (2021). Molecular phylogenetics and evolutionary analysis of a highly recombinant begomovirus, Cotton leaf curl Multan virus, and associated satellites. *Virus Evolut.* 7:veab054. doi: 10.1093/ve/veab054

Academy of Agricultural Sciences, China (grant no: BZ202005), and the Project of Collaborative Innovation Center of GDAAS-XT202210.

Acknowledgments

The authors sincerely thank Muhammad Umar (University of Tasmania, Australia) for providing helpful suggestions during the preparation of 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

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.2022.1022016/full#supplementary-material>

- García-Arenal, F., Fraile, A., and Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 39, 157–186. doi: 10.1146/annurev.phyto.39.1.157
- García-Ruiz, H. (2018). Susceptibility genes to plant viruses. *Viruses* 10:484. doi: 10.3390/v10090484
- García-Ruiz, H., Diaz, A., and Ahlquist, P. (2018). Intermolecular Rna recombination occurs at different frequencies in alternate forms of brome mosaic virus Rna replication compartments. *Viruses* 10:131. doi: 10.3390/v10030131
- Gelbart, M., Harari, S., Ya, B.-A., Kustin, T., Wolf, D., Mandelboim, M., et al. (2020). Drivers of within-host genetic diversity in acute infections of viruses. *PLoS Pathog.* 16:e1009029. doi: 10.1371/journal.ppat.1009029
- Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M., and Andino, R. (2014). Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog.* 10:e1004529. doi: 10.1371/journal.ppat.1004529
- Harvey, E., and Holmes, E. C. (2022). Diversity and evolution of the animal Virome. *Nat. Rev. Microbiol.* 20, 321–334. doi: 10.1038/s41579-021-00665-x
- Hu, G., Wu, Z., Wang, K., Uversky, V. N., and Kurgan, L. (2016). Untapped potential of disordered proteins in current Druggable human proteome. *Curr. Drug Targets* 17, 1198–1205. doi: 10.2174/1389450116666150722141119
- Huang, L., Naylor, M., Pallett, D., Reeves, J., Cooper, J., and Wang, H. (2005). The complete genome sequence, organization and affinities of carrot red leaf virus. *Arch. Virol.* 150, 1845–1855. doi: 10.1007/s00705-005-0537-6
- Ibaba, J. D., Laing, M. D., and Gubba, A. (2017). Pepo aphid-borne yellows virus: a new species in the genus Polorovirus. *Virus Genes* 53, 134–136. doi: 10.1007/s11262-016-1390-2
- Ishida, T., and Kinoshita, K. (2007). Prdos: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi: 10.1093/nar/gkm363
- Kaplan, I. B., Lee, L., Ripoll, D. R., Palukaitis, P., Gildow, F., and Gray, S. M. (2007). Point mutations in the potato Leafroll virus major capsid protein Alter Virion stability and aphid transmission. *J. Gen. Virol.* 88, 1821–1830. Epub 2007/05/09. doi: 10.1099/vir.0.82837-0
- Katuwawala, A., Oldfield, C. J., and Kurgan, L. (2019). Accuracy of protein-level disorder predictions. *Brief. Bioinform.* 21, 1509–1522. doi: 10.1093/bib/bbz100
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. (2006). Gard: a genetic algorithm for recombination detection. *Bioinformatics* 22, 3096–3098. doi: 10.1093/bioinformatics/btl474
- Kozłowski, L. P., and Bujnicki, J. M. (2012). Metadisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* 13:111. doi: 10.1186/1471-2105-13-111
- Kreuzer, F., Sousa-Dias, J., Jeevalatha, A., Figueria, A., Valkonen, J., and Jones, R. (2020). “Viral diseases in potato,” in *The Potato Crop. Its Agricultural, Nutritional and Social Contribution to Humankind*. eds. H. Campos and O. Ortiz (Cham: Springer), 389–430.
- Krueger, E. N., Beckett, R. J., Gray, S. M., and Miller, W. A. (2013). The complete nucleotide sequence of the genome of barley yellow dwarf virus-Rmv reveals it to be a new Polorovirus distantly related to other yellow dwarf viruses. *Front. Microbiol.* 4:205. doi: 10.3389/fmicb.2013.00205
- Kumar, A., Murthy, S., and Kapoor, A. (2017). Evolution of selective-sequencing approaches for virus discovery and Virome analysis. *Virus Res.* 239, 172–179. doi: 10.1016/j.virusres.2017.06.005
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., and Mega, X. (2018). Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kumari, P., Kumar, J., Kumar, R. R., Ansari, M., Rajani, K., Kumar, S., et al. (2020). Inhibition of potato Leafroll virus multiplication and systemic translocation by Sirna constructs against putative Atpase fold of movement protein. *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-78791-4
- Kwak, H.-R., Lee, H. J., Kim, E.-A., Seo, J.-K., Kim, C.-S., Lee, S. G., et al. (2018). Complete genome sequences and evolutionary analysis of cucurbit aphid-borne yellows virus isolates from melon in Korea. *Plant Pathol. J.* 34, 532–543. doi: 10.5423/ppj.Oa.03.2018.0049
- Lacroix, C., Renner, K., Cole, E., Seabloom, E. W., Borer, E. T., and Malmstrom, C. M. (2016). Methodological guidelines for accurate detection of viruses in wild plant species. *Appl. Environ. Microbiol.* 82, 1966–1975. doi: 10.1128/aem.03538-15
- Lafforgue, G., Michon, T., and Charon, J. (2022). Analysis of the contribution of intrinsic disorder in shaping Potyvirus genetic diversity. *bioRxiv*. doi: 10.1101/2022.05.13.491648
- LaTourrette, K., Holste, N. M., and García-Ruiz, H. (2021). Polorovirus Genomic Variation. *Virus Evol.* 7:2. doi: 10.1093/ve/veab102
- LaTourrette, K., Holste, N. M., Rodríguez-Peña, R., Leme, R. A., and García-Ruiz, H. (2021). Genome-wide variation in betacoronaviruses. *J. Virol.* 95:e0049621. doi: 10.1128/JVI.00496-21
- Letunic, I., and Bork, P. (2021). Interactive tree of life (ItoL) V5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Li, J., Zheng, H., Zhang, C., Han, K., Wang, S., Peng, J., et al. (2016). Different virus-derived Sirnas profiles between leaves and fruits in cucumber green mottle mosaic virus-infected *Lagenaria siceraria* plants. *Front. Microbiol.* 7:1797. doi: 10.3389/fmicb.2016.01797
- Librado, P., and Rozas, J. (2009). DnapS V5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). Rdp4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003
- Mauck, K. E., De Moraes, C. M., and Mescher, M. C. (2014). Biochemical and physiological mechanisms underlying effects of cucumber mosaic virus on host-plant traits that mediate transmission by aphid vectors. *Plant Cell Environ.* 37, 1427–1439. doi: 10.1111/pce.12249
- Mishra, P. M., Verma, N. C., Rao, C., Uversky, V. N., and Nandi, C. K. (2020). Intrinsically disordered proteins of viruses: involvement in the mechanism of cell regulation and pathogenesis. *Prog. Mol. Biol. Transl. Sci.* 174, 1–78. doi: 10.1016/b.smbts.2020.03.001
- Moury, B., and Simon, V. (2011). Dn/ds-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of potato virus Y. *Mol. Biol. Evol.* 28, 2707–2717. doi: 10.1093/molbev/msr105
- Muhire, B. M., Varsani, A., and Martin, D. P. (2014). Sdt: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277. doi: 10.1371/journal.pone.0108277
- Nagy, P. D. (2008). “Recombination in plant Rna viruses,” in *Plant Virus Evolution*. ed. M. J. Roossinck (Berlin, Heidelberg: Springer), 133–156.
- Ndikumana, I., Pinel-Galzi, A., Fargette, D., and Hébrard, E. (2017). Complete genome sequence of a new strain of Rice yellow mottle virus from Malawi, characterized by a recombinant Vpg protein. *Genome Announc.* 5, e01198–e01117. doi: 10.1128/genomeA.01198-17
- Nigam, D., and García-Ruiz, H. (2020). Variation profile of the Orthotospovirus genome. *Pathogens* 9:521. doi: 10.3390/pathogens9070521
- Obenauer, J. C., Denson, J., Mehta, P. K., Su, X., Mukatira, S., Finkelstein, D. B., et al. (2006). Large-scale sequence analysis of avian influenza isolates. *Science* 311, 1576–1580. doi: 10.1126/science.1121586
- Okeyo, O. G. (2017). Response of potato genotypes to virus infection and effectiveness of positive selection in Management of Seed Borne Potato Viruses. *J. Agri. Sci.* 10:71. doi: 10.5539/jas.v10n3p71
- Pagán, I., and Holmes, E. C. (2010). Long-term evolution of the Luteoviridae: time scale and mode of virus speciation. *J. Virol.* 84, 6177–6187. doi: 10.1128/jvi.02160-09
- Palukaitis, P. (2012). Resistance to viruses of potato and their vectors. *Plant Pathol. J.* 28, 248–258. doi: 10.5423/PPJ.RW.06.2012.0075
- Pascall, D. J., Tinsley, M. C., Clark, B. L., Obbard, D. J., and Wilfert, L. (2021). Virus prevalence and genetic diversity across a wild bumblebee community. *Front. Microbiol.* 12:650747. doi: 10.3389/fmicb.2021.650747
- Patton, M. F., Bak, A., Sayre, J. M., Heck, M. L., and Casteel, C. L. (2020). A Polorovirus, potato Leafroll virus, alters plant–vector interactions using three viral proteins. *Plant Cell Environ.* 43, 387–399. doi: 10.1111/pce.13684
- Pazhouhandeh, M., Dieterle, M., Marrocco, K., Lechner, E., Berry, B., Brault, V., et al. (2006). F-box-like domain in the Polorovirus protein P0 is required for silencing suppressor function. *Proc. Natl. Acad. Sci.* 103, 1994–1999. doi: 10.1073/pnas.0510784103
- Peter, K. A., Liang, D., Palukaitis, P., and Gray, S. M. (2008). Small deletions in the potato Leafroll virus Readthrough protein affect particle morphology, aphid transmission, virus movement and accumulation. *J. Gen. Virol.* 89, 2037–2045. Epub 2008/07/18. doi: 10.1099/vir.0.83625-0
- Pond, S. L., and Frost, S. D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533. doi: 10.1093/bioinformatics/bti320
- Rantalainen, K. I., Eskelin, K., Tompa, P., and Mäkinen, K. (2011). Structural flexibility allows the functional diversity of Potyvirus genome-linked protein Vpg. *J. Virol.* 85, 2449–2457. doi: 10.1128/jvi.02051-10
- Rubio, L., Galipienso, L., and Ferriol, I. (2020). Detection of plant viruses and disease management: relevance of genetic diversity and evolution. *Front. Plant Sci.* 11:1092. doi: 10.3389/fpls.2020.01092
- 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
- Scheffler, K., Martin, D. P., and Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics* 22, 2493–2499. doi: 10.1093/bioinformatics/btl427

- Smirnova, E., Firth, A. E., Miller, W. A., Scheidecker, D., Brault, V., Reinbold, C., et al. (2015). Discovery of a small non-Aug-initiated Orf in Poloroviruses and Luteoviruses that is required for long-distance movement. *PLoS Pathog.* 11:e1004868. doi: 10.1371/journal.ppat.1004868
- Sómera, M., Sarmiento, C., and Truve, E. (2015). Overview on Sobemoviruses and a proposal for the creation of the family Sobemoviridae. *Viruses* 7, 3076–3115. doi: 10.3390/v7062761
- Szilágyi, A., Györffy, D., and Závodszy, P. (2008). The Twilight Zone between Protein Order and Disorder. *Biophys. J.* 95, 1612–1626. doi: 10.1529/biophysj.108.131151
- Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N., and Tawfik, D. S. (2009). Do Viral Proteins Possess Unique Biophysical Features? *Trends Biochem. Sci.* 34, 53–59. doi: 10.1016/j.tibs.2008.10.009
- Traoré, O., Pinel-Galzi, A., Issaka, S., Poulicard, N., Aribi, J., Aké, S., et al. (2010). The adaptation of Rice yellow mottle virus to the Eif(Iso)4g-mediated Rice resistance. *Virology* 408, 103–108. doi: 10.1016/j.virol.2010.09.007
- Trebicki, P. (2020). Climate change and plant virus epidemiology. *Virus Res.* 286:198059. doi: 10.1016/j.virusres.2020.198059
- Umar, M., Farooq, T., Tegg, R. S., Thangavel, T., and Wilson, C. R. (2022). Genomic characterisation of an isolate of brassica yellows virus associated with brassica weed in Tasmania. *Plan. Theory* 11:884. doi: 10.3390/plants11070884
- Umar, M., Tegg, R. S., Farooq, T., Thangavel, T., and Wilson, C. R. (2022). Abundance of Poloroviruses within Tasmanian pea crops and surrounding weeds, and the genetic diversity of Tuyv isolates found. *Viruses* 14:1690. doi: 10.3390/v14081690
- Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* 269, 2–12. doi: 10.1046/j.0014-2956.2001.02649.x
- Uversky, V. N. (2019). Intrinsically disordered proteins and their “mysterious” (meta)physics. *Front. Phys.* 10:7. doi: 10.3389/fphy.2019.00010
- Valkonen, J. P. (2015). Elucidation of virus-host interactions to enhance resistance breeding for control of virus diseases in potato. *Breed. Sci.* 65, 69–76. doi: 10.1270/jsbbs.65.69
- Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Alfenas-Zerbini, P., et al. (2021). Changes to virus taxonomy and to the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2021). *Arch. Virol.* 166, 2633–2648. doi: 10.1007/s00705-021-05156-1
- Walter, J., Charon, J., Hu, Y., Lachat, J., Leger, T., Lafforgue, G., et al. (2019). Comparative analysis of mutational robustness of the intrinsically disordered viral protein Vpg and of its interactor Eif4e. *PLoS One* 14:e0211725. doi: 10.1371/journal.pone.0211725
- Wan, J., Cabanillas, D. G., Zheng, H., and Laliberté, J. F. (2015). Turnip mosaic virus moves systemically through both phloem and xylem as membrane-associated complexes. *Plant Physiol.* 167, 1374–1388. doi: 10.1104/pp.15.00097
- Wasik, B. R., and Turner, P. E. (2013). On the biological success of viruses. *Annu. Rev. Microbiol.* 67, 519–541. doi: 10.1146/annurev-micro-090110-102833
- Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331. doi: 10.1006/jmbi.1999.3110
- Xu, Y., Ju, H.-J., DeBlasio, S., Carino, E. J., Johnson, R., MacCoss, M. J., et al. (2018). A stem-loop structure in potato Leafroll virus open Reading frame 5 (Orf5) is essential for Readthrough translation of the coat protein Orf stop codon 700 bases upstream. *J. Virol.* 92, e01544–e01517. doi: 10.1128/JVI.01544-17
- Xue, B., Blocquel, D., Habchi, J., Uversky, A. V., Kurgan, L., Uversky, V. N., et al. (2014). Structural Disorder in Viral Proteins. *Chem. Rev.* 114, 6880–6911. doi: 10.1021/cr4005692
- Zarghani, S. N., Shams-Bakhsh, M., Zand, N., Sokhandan-Bashir, N., and Pazhouhandeh, M. (2012). Genetic analysis of Iranian population of potato Leafroll virus based on Orf0. *Virus Genes* 45, 567–574. doi: 10.1007/s11262-012-0804-z