



# Participation of Multifunctional RNA in Replication, Recombination and Regulation of Endogenous Plant Pararetroviruses (EPRVs)

Katja R. Richert-Pöggeler<sup>1\*</sup>, Kitty Vijverberg<sup>2,3</sup>, Osamah Alisawi<sup>4</sup>, Gilbert N. Chofong<sup>1</sup>, J. S. (Pat) Heslop-Harrison<sup>5,6</sup> and Trude Schwarzacher<sup>5,6</sup>

<sup>1</sup> Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Institute for Epidemiology and Pathogen Diagnostics, Braunschweig, Germany, <sup>2</sup> Naturalis Biodiversity Center, Evolutionary Ecology Group, Leiden, Netherlands, <sup>3</sup> Radboud University, Institute for Water and Wetland Research (IWWR), Nijmegen, Netherlands, <sup>4</sup> Department of Plant Protection, Faculty of Agriculture, University of Kufa, Najaf, Iraq, <sup>5</sup> Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom, <sup>6</sup> Key Laboratory of Plant Resources Conservation and Sustainable Utilization, Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

## OPEN ACCESS

### Edited by:

Jens Staal,  
Ghent University, Belgium

### Reviewed by:

Jie Cui,  
Institut Pasteur of Shanghai (CAS),  
China

Marco Catoni,  
University of Birmingham,  
United Kingdom

### \*Correspondence:

Katja R. Richert-Pöggeler  
katja.richert-poeggeler@  
julius-kuehn.de

### Specialty section:

This article was submitted to  
Plant Pathogen Interactions,  
a section of the journal  
Frontiers in Plant Science

**Received:** 31 March 2021

**Accepted:** 19 May 2021

**Published:** 21 June 2021

### Citation:

Richert-Pöggeler KR, Vijverberg K, Alisawi O, Chofong GN, Heslop-Harrison JS and Schwarzacher T (2021) Participation of Multifunctional RNA in Replication, Recombination and Regulation of Endogenous Plant Pararetroviruses (EPRVs). *Front. Plant Sci.* 12:689307. doi: 10.3389/fpls.2021.689307

Pararetroviruses, taxon *Caulimoviridae*, are typical of retroelements with reverse transcriptase and share a common origin with retroviruses and LTR retrotransposons, presumably dating back 1.6 billion years and illustrating the transition from an RNA to a DNA world. After transcription of the viral genome in the host nucleus, viral DNA synthesis occurs in the cytoplasm on the generated terminally redundant RNA including inter- and intra-molecule recombination steps rather than relying on nuclear DNA replication. RNA recombination events between an ancestral genomic retroelement with exogenous RNA viruses were seminal in pararetrovirus evolution resulting in horizontal transmission and episomal replication. Instead of active integration, pararetroviruses use the host DNA repair machinery to prevail in genomes of angiosperms, gymnosperms and ferns. Pararetrovirus integration – leading to Endogenous ParaRetroViruses, EPRVs – by illegitimate recombination can happen if their sequences instead of homologous host genomic sequences on the sister chromatid (during mitosis) or homologous chromosome (during meiosis) are used as template. Multiple layers of RNA interference exist regulating episomal and chromosomal forms of the pararetrovirus. Pararetroviruses have evolved suppressors against this plant defense in the arms race during co-evolution which can result in deregulation of plant genes. Small RNAs serve as signaling molecules for Transcriptional and Post-Transcriptional Gene Silencing (TGS, PTGS) pathways. Different populations of small RNAs comprising 21–24 nt and 18–30 nt in length have been reported for *Citrus*, *Fritillaria*, *Musa*, *Petunia*, *Solanum* and *Beta*. Recombination and RNA interference are driving forces for evolution and regulation of EPRVs.

**Keywords:** pararetroviruses, virus evolution, recombination, DNA repair, small RNAs, PTGS, TGS, disease resistance

## INTRODUCTION

Retroelements (class I transposable elements) can be considered to lie at the transition from an RNA to a DNA world. They are genomic DNA elements within all kingdoms, but employ reverse transcriptase for DNA synthesis using an RNA intermediate template (Xiong and Eickbush, 1990; Simon et al., 2008; Koonin et al., 2015). Besides intracellular forms, e.g., Long Terminal Repeat (LTR) retrotransposons, virion-forming elements exist that can leave the cell. These infectious retroelements developed two distinct strategies during adaptation and co-evolution with their respective hosts. Retroviruses infecting animal and human hosts encapsidate single stranded (ss) RNA, from which a linearized dsDNA molecule with LTRs is generated. Integration in the host genome by viral encoded integrase is obligatory to obtain full-length retroviral RNA (Krupovic et al., 2018). This replication scheme is also used by LTR retrotransposons, such as *Metaviridae* and *Pseudoviridae* also known as Ty3/Gypsy- and Ty1/Copia-elements, respectively. In plants, LTR retrotransposons are present. True retroviruses are lacking but plants have infective pararetroviruses (PRV), family *Caulimoviridae*. They encapsidate circular dsDNA that, after release and transport to the nucleus, forms a minichromosome allowing synthesis of terminal redundant viral RNA (Gronenborn, 1987; Hohn and Rothnie, 2013). Therefore, viral integration into the host chromosomal DNA is not required. In the cytoplasm, the viral RNA is reverse transcribed resulting in circular full-length viral DNA. Interestingly, PRV sequences can also be found inserted into plant genomic DNA and then become endogenous PRVs (EPRVs). Insertions usually comprise silenced, degenerated and/or fossil forms. Additionally, active, proliferating EPRV that can trigger virus infection exist in some hosts indicative for a recent invasion of the plant genome.

High throughput DNA and RNA sequencing, metagenomics, bioinformatics and palaeovirology have revealed the impact of viral retroelements on eukaryotic genomes and become important for understanding the origin of viruses. In cells, DNA polymerases secure the amplification of large genomes based on linear dsDNA molecules. However, relics of the “RNA world” can still be encountered such as ribosomal RNA, RNA splicing, telomerases and RNA-dependent RNA polymerases (RDR; Gilbert, 1986; Heslop-Harrison, 2000; Krupovic et al., 2018). Special attention should be paid to active, virion forming and/or infectious retroelements because of their impact on the host genomes, horizontal DNA transfer and triggering diseases. Sequence and structural analysis of the capsid proteins support the hypotheses that retroviruses, pararetroviruses and LTR retrotransposons share the same origin dating back 1.6 billion years (Krupovic and Koonin, 2017). They most likely evolved from a common ancestor that encoded the genes for a capsid protein, protease and reverse transcriptase including the RNase H domain.

Present day retroelement repeats, originating from *Caulimoviridae*, *Metaviridae*, and *Pseudoviridae* occupy distinct niches in plant genomes (Krupovic et al., 2018; International Committee on Taxonomy of Viruses Executive Committee [ICTV], 2020), and can accumulate to high numbers in genomes

of angiosperms, gymnosperms and ferns (Becher et al., 2014; Geering et al., 2014; Diop et al., 2018; Gong and Han, 2018).

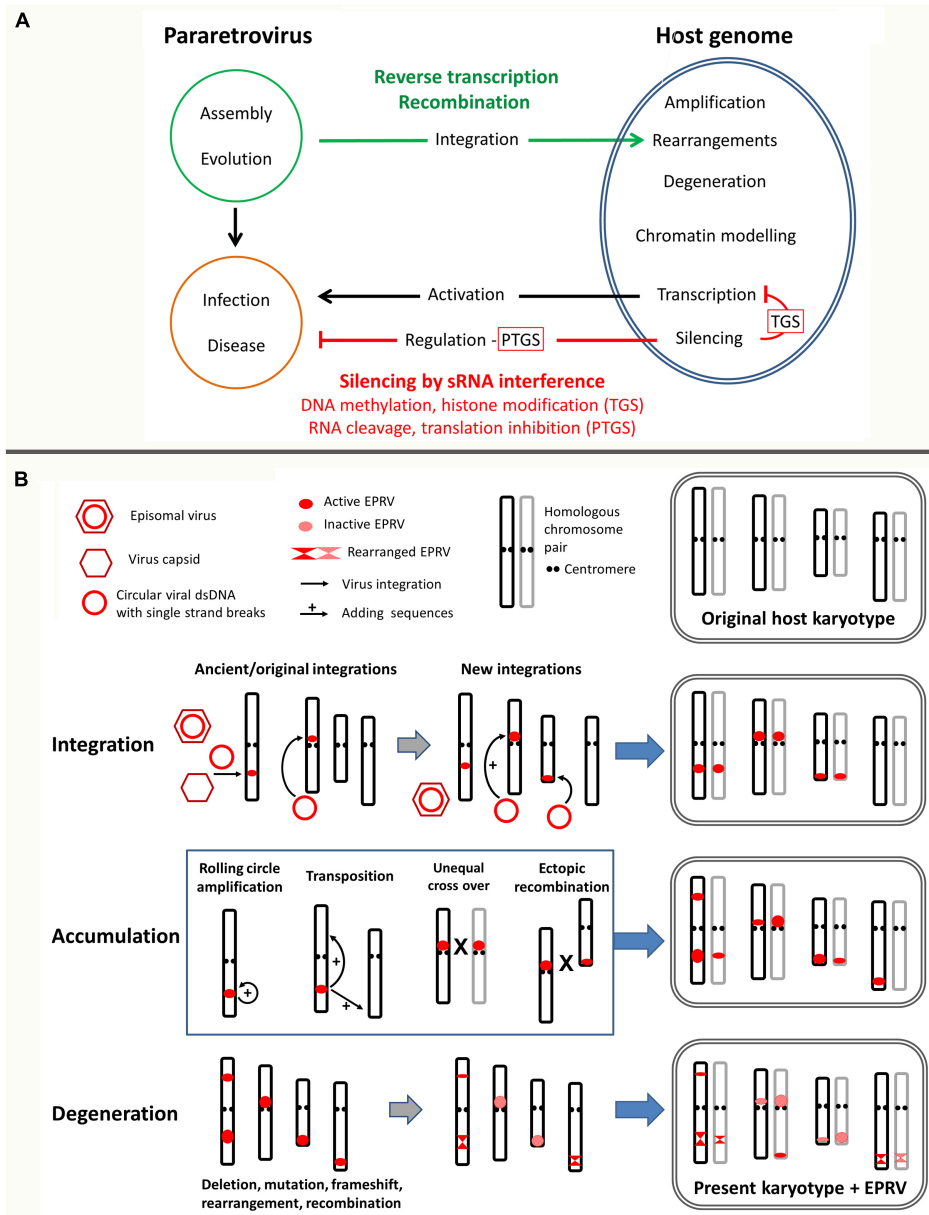
This review points out general principles identified in EPRV evolution, adaptation, function, defense and control (Figure 1A). Special focus will be given to mechanisms related to recombination, dsDNA break repair, and RNA interference, explaining the EPRV virosphere, the space in which EPRVs occur and which is influenced by them.

## ANCIENT INTRACELLULAR RECOMBINATION EVENTS

Ancient recombination events between proliferating LTR retrotransposons and co-infecting RNA viruses may have occurred in the ancestral plant cell. Indeed, analysis of encapsidated molecules of cucumber necrosis virus (*Tombusviridae*) identified also retrotransposon RNA in 0.4–1.3% of sequences isolated from virions (Ghoshal et al., 2015); such hetero-encapsidation can lead to recombination and formation of chimeric genomes. The actively transcribed *Hordeum vulgare* BARE-1 virus (*Pseudoviridae*) comprises 10% of the barley genome (Jääskeläinen et al., 2013). During its replication virus-like particles are formed that encapsidate its genomic RNA together with the reverse transcriptase, ribonuclease H and integrase (Chang et al., 2013). Since *Hordeum vulgare* BARE-1 virus lacks a movement protein for intercellular transport it stays intracellularly. Analogies regarding replication and recombination patterns of infectious retroelements have also been described for plus sense (+) ssRNA viruses (Ahluquist, 2006; Sztuba-Solińska et al., 2011; Tromas et al., 2014).

## LEAVING THE CELL AND BECOMING A PARARETROVIRUS

Unequal recombination generates solo LTR footprints in the plant genome (Vitte and Panaud, 2003) that counteract bursts of retroelement replicative transposition. Such excised retrotransposon genomes might have initiated the transition to episomal replication as known from *Caulimoviridae*. An essential step toward adaptation for systemic spread within the plant host was the incorporation of genetic information for a movement protein (MP) to enable viruses to move between cells via plasmodesmata in the cell wall. Encounter of a LTR retrotransposon transcript with a + ssRNA virus encoding a 30 k MP in the cytoplasm most likely through recombination, formed chimeric molecules that gave rise to an ancestral form of *Caulimoviridae* (Koonin et al., 2015; Mushegian and Elena, 2015; Diop et al., 2018). Template switching between two RNA molecules during reverse transcription has been shown for retroviruses, LTR retrotransposons and is proposed for pararetroviruses (Froissart et al., 2005; Tromas et al., 2014; Sanchez et al., 2017). Recombination events are rare, but the rates amount to  $1.4 \times 10^{-5}$  events per site per generation for human immunodeficiency virus type 1 (Neher and Leitner, 2010) and



**FIGURE 1 |** Multiple layers of endogenous pararetrovirus (EPRV) host genome interaction. **(A)** Overview of involved pathways leading to EPRV invasion of and prevalence in genomes. The ancestral pararetrovirus (PRV) most likely assembled when transcripts of a LTR retrotransposon and an infecting RNA virus recombined allowing systemic plant infection and horizontal transmission. PRVs can become residents of the host genome using illegitimate recombination or reverse transcriptase and accumulate in clusters in certain genomic regions depending on the genomic context and chromatin environment of the host. Endogenous PRVs (EPRVs) usually stay silenced and/or degenerate by mutation, rearrangement or fragmentation (see under b), but sometimes activate. When still active and complete, EPRVs can escape the cellular surveillance system by producing transcripts with terminal repeats. Those serve as template for synthesis of full length PRV episomes that can trigger a viral infection and disease, the same as original episomal PRVs. Furthermore, pararetroviral encoded suppressor of gene silencing counteract the plant defense. EPRV silencing is regulated by RNA interference comprising transcriptional and post-transcriptional gene silencing (TGS and PTGS) as well as DNA methylation and histone modifications (see **Figure 2**). **(B)** Detailed view of EPRV interactions with chromosomes. Prevalence of EPRVs are influenced by three steps: integration, accumulation and degeneration. Four chromosomes and their homologs in lighter shade are shown. Size of the red circle/ellipsoid indicates the number of integrations. Steps of integration, accumulation and degeneration, here depicted separately, can happen simultaneously and over a long time period. The present karyotype shows significant changes in EPRV sequences from the ancient integrations; karyotype rearrangements also occur but are not shown. **Integration:** Tandem arrays and clusters are generated by simultaneous or subsequent integration and over time through backcrosses and selfing the two homologous chromosomes homogenize. **Accumulation:** Tandemly integrated EPRVs can amplify by several mechanisms shown in the middle box: rolling cycle amplification, transposition of newly synthesized copies using reverse transcription to sites on the same or different chromosome. Unequal recombination (cross overs) between sites on homologous chromosomes or ectopic recombination between heterologous chromosomes (shown as X) will both amplify and reduce the number of copies at given site. **Degeneration:** As soon as EPRV copies are integrated, the host genome acts by inactivation through epigenetic silencing mechanisms (see **Figure 2**), but also through sequence degeneration by mutation and deletions that cause frameshifts, fragmentations, rearrangements and recombination.

$4 \times 10^{-5}$  events per nucleotide site and replication cycle for the pararetrovirus cauliflower mosaic virus (Froissart et al., 2005).

The predecessors of extant members of *Caulimoviridae* might have originated from the pool of hybrid replicons. Integration loci and patterns are similar among LTR retrotransposons and EPRVs in petunia and banana (Richert-Pöggeler et al., 2003; Gayral et al., 2008). Furthermore, an integrase-like motif and quasi terminal repeats have been noticed in the *Petuvirus* petunia vein clearing virus (PVCV, Krupovic et al., 2018) that is also phylogenetically close to *Metaviridae* (Diop et al., 2018). Adaptation to episomal replication resulted in the loss of integrase function. The ancestral pararetrovirus probably gained further independence from vertical transmission by acquisition of a gene for vector transmission enabling dissemination between plants.

## INVASION OF GENOMES – ILLEGITIMATE RECOMBINATION

Pararetrovirus integration into plant chromosomes is supposed to occur mainly through illegitimate recombination (Liu et al., 2012; Geering et al., 2014). This can take place during somatic DNA repair or meiotic recombination. Both follow DNA double strand breaks (DSBs) that occur randomly, as a consequence of genotoxic agents or are strongly enhanced by the impact transcription has on replication fork progression (Aguilera and Gaillard, 2014; Knoll et al., 2014), or are induced deliberately during meiotic prophase (Schwarzacher, 2003). DSBs repair is vital to ensure the integrity of the genomes and is achieved by non-homologous end joining (NHEJ) or homologous recombination (HR) pathways (Knoll et al., 2014; Heyer, 2015). When single DSBs are present the NHEJ is usually reliable. In case of several DSBs occurring simultaneously or if larger stretches of DNA are missing, NHEJ leads to illegitimate rejoining, often associated with deletions, random translocations, as well as insertion of sequences from elsewhere (Knoll et al., 2014; Kowalczykowski, 2015). Accordingly, pararetrovirus integration can happen if their sequences are used as template instead of homologous host genomic sequences on the sister chromatid (during mitosis) or homologous chromosome (during meiosis). Virus integration occurs frequently in somatic cells (Gayral et al., 2008), but the manifestation of such an event in reproductive cells and thus in the progeny may be rare.

In the recombination pathway, following DSBs, a single strand (ssDNA) overhang is generated by strand resection that then attracts recombinases to search for homologous sequences (Pradillo et al., 2012). This ssDNA is essential for repair and can become quite long reaching 2–10 kb when less homologous ectopic sequences are involved (Chung et al., 2010). This leads to higher fidelity in one way, as it prevents recombination within short DNA repeats next to the break, and can lead to homology recognition over a larger DNA segment that avoids recombination between homologous chromosomes in polyploids (Sepsi and Schwarzacher, 2020). On the other hand, involvement of longer ssDNA in the homology search and slower repair kinetics (Chung et al., 2010) enhance recombination with sequences located further away from the break or with external sources such as viral DNA sequences. DNA-strand invasion

follows and generates a D-loop that promotes DNA strand annealing and depending on which ends are used gives rise to recombination, gene conversion or the *status quo*. Complete or partial ssDNA of pararetroviruses is present in infected plant nuclei (Hohn et al., 2008), and may indeed serve as templates for recombination or host repair machinery.

It is necessary that the DNA within the chromosome is accessible and able to unwind during repair and recombination (Stadler and Richly, 2017). Chromosome and chromatin configuration also influences where DNA breaks occur and are more frequent in particular regions (Dillon et al., 2013; Lawrence et al., 2017). ERPVs have been found accumulated in heterochromatin, in particular in AT-rich regions and next to TA dinucleotide-rich (*Oryza* sp.: Kunii et al., 2004; and Liu et al., 2012; various species: Geering et al., 2014, *Beta vulgaris*: Schmidt et al., 2021), but also next to retroelements and transposons (tomato: Staginnus et al., 2007; Petunia: Richert-Pöggeler et al., 2003; Schwarzacher et al., 2016). The latter would support an alternative mode of integration for pararetroviruses together with retrotransposons during reverse transcription when template switches can occur between viral RNA strands (Hohn, 1994).

## PREVALENCE OF EPRV SEQUENCES

Often EPRVs form clusters and can occupy large parts of the genome (e.g., tobacco: Lockhart et al., 2000; Petunia, Richert-Pöggeler et al., 2003; rice: Liu et al., 2012; Citrinae spp.: Yu et al., 2019; *Beta vulgaris*: Schmidt et al., 2021) and could result from the simultaneous integration of several EPRV copies in tandem or nested, or from recombination of episomal viruses with already integrated sequences (Hohn et al., 2008). Amplifications of EPRVs within the host genome can further lead to the substantial amount of EPRVs found in many plant genomes. Several mechanisms could be involved even if they occur infrequently (**Figure 1B**): transposition similar to retroelements (Bennetzen, 2002), rolling circle amplification (e.g., Gayral et al., 2008), as well as unequal meiotic crossing-over of tandem arrays, or ectopic recombination between EPRV clusters on non-homologous chromosomes.

In order to control EPRVs, copies are frequently inactivated by sequence degeneration or fragmentation to render transcription of entire virus components ineffective (**Figure 1B**). Alternatively, epigenetic silencing through methylation and small RNAs (sRNAs) has been observed (Noreen et al., 2007; Staginnus et al., 2007; Schmidt et al., 2021, and see below). Heterochromatin that is generally transcriptionally inactive and shows low recombination rates (Heslop-Harrison and Schwarzacher, 2011) therefore can be viewed as safe havens for EPRVs (Schmidt et al., 2021) and similar to retroelements can influence the genome organization and recombination landscape (Kent et al., 2017).

## RNA INTERFERENCE AND EPRV CONTROL

Endogenous PRVs co-exist with exogenous virus(es) and viroid(s) in the same host. Like all viruses and viroid's,

EPRVs possess an RNA phase during their replication cycle (Baltimore, 1971; International Committee on Taxonomy of Viruses Executive Committee [ICTV], 2020). Activation has been reported for only a limited number of EPRVs and stresses, including from temperature, hybridization, age, or tissue culture, may activate expression of EPRVs (Ndowora et al., 1999; Lockhart et al., 2000; Richert-Pöggeler et al., 2003; Hansen et al., 2005; Kuriyama et al., 2020), as is also the case for retrotransposons (Grandbastien, 1998). Complete transcripts from genomic copies are produced if the EPRVs occur in tandem or via recombination of different segments (Richert-Pöggeler et al., 2003; Chabannes and Iskra-Caruana, 2013). (E)PRV RNA is multifunctional and serves as template for translation, for DNA-synthesis, and for RNA interference (RNAi) (Figure 2; Hohn and Rothnie, 2013; Pooggin, 2013; Prasad et al., 2019). RNAi is an evolutionary conserved, sequence-specific mechanism that regulates gene expression by employing transcriptional and post-transcriptional gene silencing (TGS and PTGS) strategies (Castel and Martienssen, 2013; Bond and Baulcombe, 2015) and is also involved in virus resistance (Wang et al., 2018). In TGS, transcription is prevented via RNA-directed DNA methylation (RdDM), while in PTGS, translation is disabled by cleavage of the transcript or translational inhibition. RNAi is a major player in antiviral defense since RNA dependent RNA polymerases (RDRs) enable a self-feeding process for systemic spreading of the RNAi signal (Obbard et al., 2009; Zhang et al., 2019; Jin et al., 2021). RNAi initiates on dsRNA molecules (Prasad et al., 2019) that are processed into small RNA (sRNA) duplexes of 21, 22 or 24 nucleotides long by plant *Dicer(-Like)* endoribonucleases (*DCLs*; Deleris et al., 2006). *DCL1* functions in the production of miRNAs from imperfect double strand stems of fold-back pre-miRNAs transcribed from (endogenous) MIR-loci (Bartel, 2004; Vijverberg et al., 2016). The role of miRNAs in virus defense is thought to be either direct, by targeting the viral RNA with the possibility of 2–3 mismatches, or indirect, by triggering the biogenesis of viral (v)siRNAs (Liu et al., 2017) or regulating plant defense genes (Carbonell and Carrington, 2015). *DCL3* and *DCL4* produce 24-nt and 21-nt small interfering RNAs (siRNAs) from perfect dsRNA templates, respectively, each with minor distinct catalytic profiles to ensure specificity (Nagano et al., 2014). Overlapping transcripts and self-complementary regions of (E)PRVs and other DNA viruses serve as sources of dsRNA templates for vsiRNA production (Figure 2; Prasad et al., 2019). The 24-nt (v)siRNAs function in TGS via RdDM to suppress the activation of transposons and pararetroviruses (Matzke and Mosher, 2014), while the 21-nt (v)siRNAs function in PTGS via sequence-specific cleavage of transcripts (Garcia-Ruiz et al., 2010). *DCL2* synthesizes stress-related natural-antisense-transcript (nat)-siRNAs and is thought to regulate the biogenesis of 22-nt vsiRNAs (Deleris et al., 2006). In *dcl4* mutants, *DCL2* is involved in increased 22-nt siRNAs production and alternate production of 22-nt siRNAs transacting (ta)-siRNA precursors (Zhang et al., 2019). The specific mode of action of 22-nt siRNAs is yet unclear, but evidence indicates that they mediate translational repression and are less effective in target cleavage (Wu et al., 2020). Recently also tRNAs have emerged as a source for small RNAs to suppress reverse transcriptase of (LTR-) retrotransposons (Schorn et al., 2017),

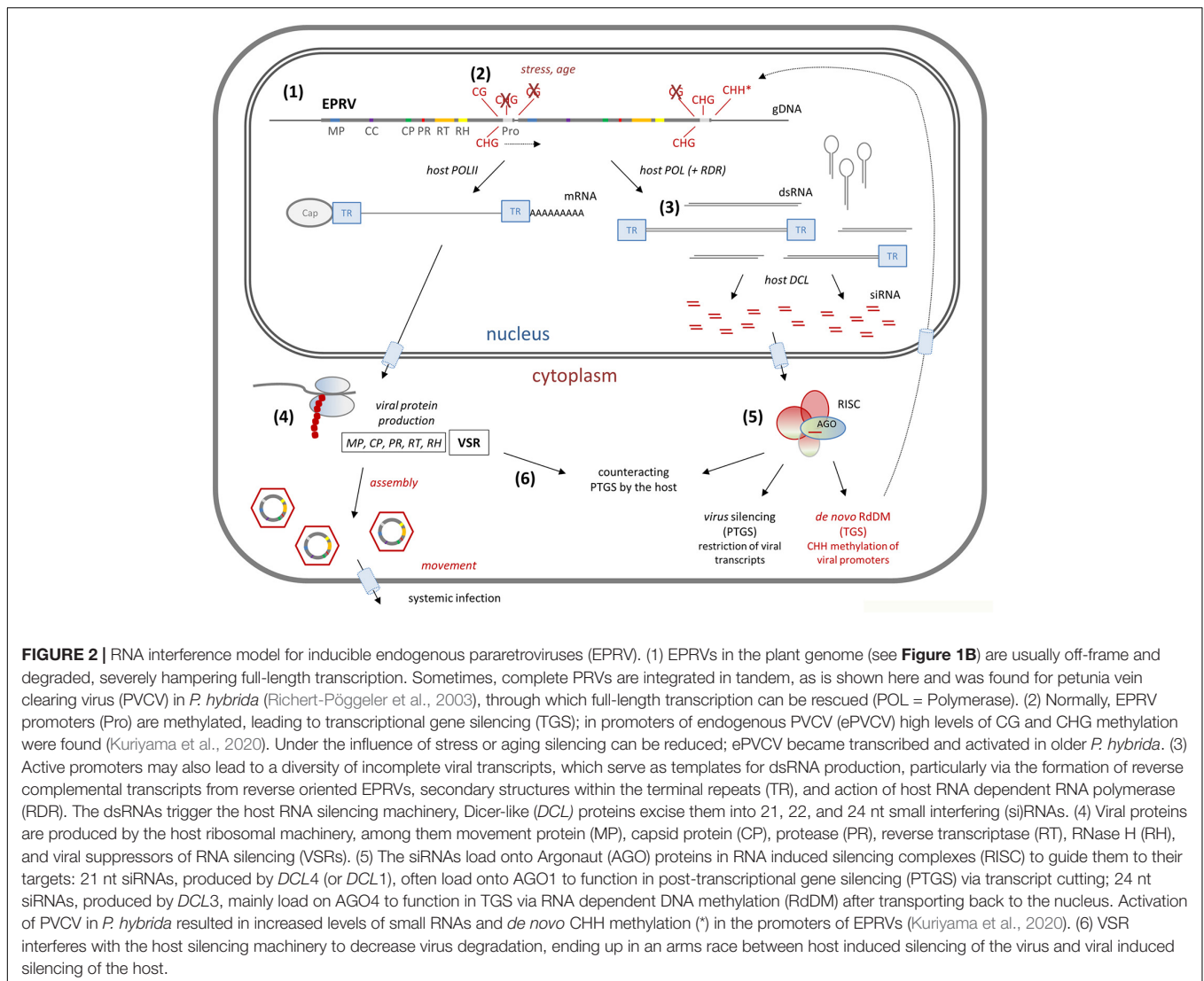
which may imply that they act similarly against (E)PRVs. All four *DCLs* generate DNA virus-derived 21-, 22-, and 24-nt small RNAs (Blevins et al., 2006). Each of the produced sRNA classes functions by guiding an RNA-induced silencing complex (RISC) to its (near) complementary target sequence after being loaded into an *ARGONAUTE* (*AGO*) protein (Figure 2; Rogers and Chen, 2013). Upon arrival, other proteins from the RISC act in the degradation of the transcript, repression of its translation or in methylation to suppress its expression.

Integrated copies of PVCV are associated with repressive H3K9 methylation (Noreen et al., 2007) and increased CG and CHG methylation in their promoter region (Figure 2; Kuriyama et al., 2020; also found in the related *Fritillaria imperialis* (*Fri*)EPRV, Becher et al., 2014). Endogenous TVCV-like sequences in *Solanum* species show CHH methylation (Staginnus et al., 2007), further supporting RdDM based transcriptional silencing of EPRVs, and *Fri*EPRV showed abundant 24 nt siRNAs, a hallmark of TGS by RdDM (Becher et al., 2014). In *Beta vulgaris*, sRNAs between 18 and 30 nt were found, indicating that florendovirus beet EPRVs, although not present as active forms in nature, are nevertheless silenced by TGS and PTGS (Schmidt et al., 2021). In citrus, *Cit*EPRV was assembled from 24 nt sRNAs, while other non-endogenous viral genomes in this study were assembled from 21–22 nt sRNAs (Barrero et al., 2017). Much higher sRNA coverage of *Cit*EPRV in symptomatic plants compared to asymptomatic plants, indicates that this EPRV can also be activated and become infectious despite the host defense (Matsumura et al., 2017). Higher levels of 21 nt and 22 nt compared to 24 nt sRNAs were found for episomal badnaviruses in *Musa acuminata* (Rajeswaran et al., 2014), supporting PTGS in the presences of active viruses. Kuriyama et al. (2020) demonstrated the interaction of endogenous PVCV in *P. hybrida* with EPRV expression and host defense during development: in younger plants, petal veins are white due to the silencing of the *Chalcone Synthase* gene *CHS-A* by PTGS and the promoter sequences of integrated PVCV by TGS. In older plants TGS seems less effective permitting episomal PVCV replication transiently. Associated activity of the identified Viral Suppressor of RNA silencing (VSR) counteracted the PTGS of *CHS-A* as seen in a changed color pattern of the flowers. With more viral transcripts available also derived 21-/22-nt and 24-nt siRNAs increased, reinitiating TGS and PTGS to terminate transcription from chromosomal as well as episomal PVCV DNA (Figure 2).

## OUTLOOK

The distribution of EPRVs observed in host genomes results from a combination of past virus integration, and the mechanisms of amplification or reduction of viral sequences integrated into the chromosomes over time (Figure 1B). EPRVs may thus be found at their initial position of invasion, or elsewhere in the genome, potentially with different (or no) site preferences.

New methodologies allow sequence analysis of ever smaller units such as virions or single cells. Thus, we can address: (i) which molecules are encapsidated?; (ii) are virions similar in their content?; (iii) what replicative and recombinogenic interactions of *Metaviridae* and EPRV occur within the cell?; (iv) which cells



carry active EPRVs within tissues?; (v) what are the initial landing sites of EPRVs within chromosomes and how do they spread?; and (vi) what pathways are involved in EPRV activation and silencing?

It will be essential to characterize the presence and consequences of EPRVs across all plant species, as is happening with endogenous retroviruses in animals. Solanaceae, and in particular *Petunia*, with a moderate diploid genome size of 1.4 Gb comprising mobile genetic elements such as DNA transposons, LTR retrotransposons and EPRVs, easy cultivation, seed or vegetative propagation, tissue culture and transformation, along with genomic and genetic resources, is an appropriate model to help answer the questions above. Genome editing using CRISPR/Cas will be useful to elucidate the functionality of EPRVs beyond a viral context.

The diversity in EPRV structure and in co-evolution with its host requires investigation of all pools of small RNA populations present in small RNA sequencing data. Bioinformatics will be an important tool to examine the arms race and evolution of associated regulatory mechanisms in the virus and host.

## AUTHOR CONTRIBUTIONS

KR-P in discussion with TS and KV conceived the review. KR-P wrote the abstract, introduction, outlook, and contributed to the other parts. KR-P, GC, and OA wrote the sections on PRV and EPRV evolution (1–3). TS, OA, and JH-H wrote the sections on recombination and prevalence of EPRVs in the genome (4, 5). KV wrote the section on RNAi with contributions of GC (6). TS, JH-H, and KV designed the figures. KR-P, JH-H, TS, KV, and GC edited the final manuscript. All authors contributed to writing, editing the manuscript including figures, and approved the version to be published.

## FUNDING

GC thanks the Philipp Schwartz Initiative of the Alexander von Humboldt Foundation for the financial support in frame of a Julius Kühn-Institut awarded Philipp Schwartz fellowship.

## REFERENCES

- Aguilera, A., and Gaillard, H. (2014). Transcription and Recombination: When RNA meets DNA. *Cold Spring Harb. Perspect. Biol.* 6:16543. doi: 10.1101/cshperspect.a016543
- Ahlquist, P. (2006). Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nat. Rev. Microbiol.* 4, 371–382. doi: 10.1038/nrmicro1389
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241.
- Barrero, R. A., Napier, K. R., Cunnington, J., Liefing, L., Keenan, S., Frampton, R. A., et al. (2017). An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. *BMC Bioinformatics* 18:26. doi: 10.1186/s12859-016-1428-4
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/s0092-8674(04)00045-5
- Becher, H., Ma, L., Kelly, L. J., Kovarik, A., Leitch, I. J., and Leitch, A. R. (2014). Endogenous pararetrovirus sequences associated with 24 nt small RNAs at the centromeres of *Fritillaria imperialis* L. (Liliaceae), a species with a giant genome. *Plant J.* 80, 823–833. doi: 10.1111/tjp.12673
- Bennetzen, J. L. (2002). Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115, 29–36. doi: 10.1023/A:1016015913350
- Blevins, T., Rajeswaran, R., Shivaprasad, P. V., Beknazariants, D., Si-Ammour, A., Park, H. S., et al. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* 34, 6233–6246. doi: 10.1093/nar/gkl886
- Bond, D. M., and Baulcombe, D. C. (2015). Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 112, 917–922. doi: 10.1073/pnas.1413053112
- Carbonell, A., and Carrington, J. C. (2015). Antiviral roles of plant ARGONAUTES. *Curr. Opin. Plant Biol.* 27, 111–117. doi: 10.1016/j.pbi.2015.06.013
- Castel, S. E., and Martienssen, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature* 14, 100–112.
- Chabannes, M., and Iskra-Caruana, M. L. (2013). Endogenous pararetroviruses—a reservoir of virus infection in plants. *Curr. Opin. Virol.* 3, 615–620. doi: 10.1016/j.coviro.2013.08.012
- Chang, W., Jääskeläinen, M., Li, S., and Schulman, A. H. (2013). BARE retrotransposons are translated and replicated via distinct RNA pools. *PLoS One* 8:1–12. doi: 10.1371/journal.pone.0072270
- Chung, W. H., Zhu, Z., Papusha, A., Malkova, A., and Ira, G. (2010). Defective resection at DNA double-strand breaks leads to de Novo telomere formation and enhances gene targeting. *PLoS Genet.* 6:24. doi: 10.1371/journal.pgen.1000948
- Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C., and Voinnet, O. (2006). Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* 313, 68–71. doi: 10.1126/science.1128214
- Dillon, L. W., Pierce, L. C. T., Ng, M. C. Y., and Wang, Y. H. (2013). Role of DNA secondary structures in fragile site breakage along human chromosome 10. *Hum. Mol. Genet.* 22, 1443–1456. doi: 10.1093/hmg/dd5561
- Diop, S. I., Geering, A. D. W., Alfama-Depauw, F., Loaec, M., Teycheney, P. Y., and Maumus, F. (2018). Tracheophyte genomes keep track of the deep evolution of the Caulimoviridae. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-017-16399-x
- Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., and Michalakakis, Y. (2005). Recombination every day: Abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biol.* 3:0389–0395. doi: 10.1371/journal.pbio.0030089
- García-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Bremel, K. J., et al. (2010). *Arabidopsis* RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during turnip mosaic virus infection. *Plant Cell* 22, 481–496.
- Gayral, P., Noa-Carrazana, J.-C., Lescot, M., Lheureux, F., Lockhart, B. E. L., Matsumoto, T., et al. (2008). A single banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. *J. Virol.* 82, 6697–6710. doi: 10.1128/jvi.00212-08
- Geering, A. D. W., Maumus, F., Copetti, D., Choisine, N., Zwick, D. J., Zytnicki, M., et al. (2014). Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nat. Commun.* 5:6269. doi: 10.1038/ncomms6269
- Goshal, K., Theilmann, J., Reade, R., Maghodia, A., and Rochon, D. (2015). Encapsulation of host RNAs by cucumber necrosis virus coat protein during both agroinfiltration and infection. *J. Virol.* 89, 10748–10761. doi: 10.1128/jvi.01466-15
- Gilbert, W. (1986). The RNA world Superlattices point ahead. *Nature* 319:618.
- Gong, Z., and Han, G. (2018). Euphyllophyte paleoviruses illuminate hidden diversity and macroevolutionary mode of Caulimoviridae. *J. Virol.* 92, 1–13. doi: 10.1128/JVI.02043-17
- Grandbastien, M. A. (1998). Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* 3, 181–187.
- Gronenborn B. (1987). “The molecular biology of Cauliflower mosaic virus and its application as plant gene vector,” *Plant DNA Infectious Agents*, eds in T. Hohn, and J. Schell (Vienna: Springer). doi: 10.1007/978-3-7091-6977-3\_1
- Hansen, C. N., Harper, G., and Heslop-Harrison, J. S. (2005). Characterisation of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). *Cytogenet. Genome Res.* 110, 559–565.
- Heslop-Harrison, J. S. (2000). RNA, genes, genomes and chromosomes: repetitive DNA sequences in plants. *Chromosom. Today* 13, 45–56. doi: 10.1007/978-3-0348-8484-6\_4
- Heslop-Harrison, J. S. P., and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. *Plant J.* 66, 18–33. doi: 10.1111/j.1365-313X.2011.04544.x
- Heyer, W. D. (2015). Regulation of recombination and genomic maintenance. *Cold Spring Harb. Perspect. Biol.* 7:a016501. doi: 10.1101/cshperspect.a016501
- Hohn, T. (1994). “Recombination of a plant pararetrovirus: Cauliflower mosaic virus,” in *Homologous Recombination and Gene Silencing in Plants*, ed. J. Paszkowski (Dordrecht: Springer). doi: 10.1007/978-94-011-1094-5\_2
- Hohn, T., Richert-Pöggeler, K. R., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C. H., et al. (2008). “Evolution of integrated plant viruses,” in *Plant Virus Evolution* ed. M. J. Roossinck (Berlin: Springer), 2008, 53–81. doi: 10.1007/978-3-540-75763-4\_4
- Hohn, T., and Rothnie, H. (2013). Plant pararetroviruses: replication and expression. *Curr. Opin. Virol.* 3, 621–628. doi: 10.1016/j.coviro.2013.08.013
- International Committee on Taxonomy of Viruses Executive Committee [ICTV] (2020). The new scope of virus taxonomy. Partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol.* 5, 668–674. doi: 10.1038/s41564-020-0709-x
- Jääskeläinen, M., Chang, W., Moisy, C., and Schulman, A. H. (2013). Retrotransposon BARE displays strong tissue-specific differences in expression. *New Phytol.* 200, 1000–1008. doi: 10.1111/nph.12470
- Jin, Y., Zhao, J. H., and Guo, H. S. (2021). Recent advances in understanding plant antiviral RNAi and viral suppressors of RNAi. *Curr. Opin. Virol.* 46, 65–72. doi: 10.1016/j.coviro.2020.12.001
- Kent, T. V., Uzunović, J., and Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B Biol. Sci.* 372:2017. doi: 10.1098/rstb.2016.0458
- Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. *Chromosom. Res.* 22, 191–201. doi: 10.1007/s10577-014-9415-y
- Koonin, E. V., Dolja, V. V., and Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 47, 2–25. doi: 10.1016/j.viro.2015.02.039
- Kowalczykowski, S. C. (2015). An overview of the molecular mechanisms of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* 7:16410. doi: 10.1101/cshperspect.a016410
- Krupovic, M., Blomberg, J., Coffin, J. M., Dasgupta, I., Fan, H., Geering, A. D. W., et al. (2018). Ortervirales: New virus order unifying five families of reverse-transcribing viruses. *J. Virol.* 92, 1–5. doi: 10.1128/jvi.00515-18
- Krupovic, M., and Koonin, E. V. (2017). Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses, and metaviruses. *J. Virol.* 91, e210–e217. doi: 10.1128/JVI.00210-17
- Kunii, M., Kanda, M., Nagano, H., Uyeda, I., Kishima, Y., and Sano, Y. (2004). Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: Implications for integration and evolution. *BMC Genomics* 5:1–14. doi: 10.1186/1471-2164-5-80
- Kuriyama, K., Tabara, M., Moriyama, H., Kanazawa, A., Koiwa, H., Takahashi, H., et al. (2020). Disturbance of floral colour pattern by activation of an endogenous

- pararetrovirus, petunia vein clearing virus, in aged petunia plants. *Plant J.* 103, 497–511. doi: 10.1111/tj.14728
- Lawrence, E. J., Griffin, C. H., and Henderson, I. R. (2017). Modification of meiotic recombination by natural variation in plants. *J. Exp. Bot.* 68, 5471–5483. doi: 10.1093/jxb/erx306
- Liu, R., Koyanagi, K. O., Chen, S., and Kishima, Y. (2012). Evolutionary force of AT-rich repeats to trap genomic and episomal DNAs into the rice genome: Lessons from endogenous pararetrovirus. *Plant J.* 72, 817–828. doi: 10.1111/tj.12002
- Liu, S. R., Zhou, J. J., Hu, C. G., Wei, C. L., and Zhang, J. Z. (2017). MicroRNA-mediated gene silencing in plant defense and viral counter-defense. *Front. Microbiol.* 8:1801. doi: 10.3389/fmicb.2017.01801
- Lockhart, B. E., Menke, J., Dahal, G., and Olszewski, N. E. (2000). Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *J. Gen. Virol.* 81, 1579–1585. doi: 10.1099/0022-1317-81-6-1579
- Matsumura, E. E., Coletta-Filho, H. D., Nouri, S., Falk, B. W., Nerva, L., Oliveira, T. S., et al. (2017). Deep sequencing analysis of RNAs from Citrus plants grown in a Citrus Sudden Death-affected area reveals diverse known and putative novel viruses. *Viruses* 9:92. doi: 10.3390/v9040092
- Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408.
- Mushegian, A. R., and Elena, S. F. (2015). Evolution of plant virus movement proteins from the 30K superfamily and their homologs integrated in plant genomes. *Virology* 476, 304–315. doi: 10.1016/j.virol.2014.12.012
- Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H., and Fukuhara, T. (2014). Distinct substrate specificities of Arabidopsis *DCL3* and *DCL4*. *Nucleic Acids Res.* 42, 1845–1856. doi: 10.1093/nar/gkt1077
- Neher, R. A., and Leitner, T. (2010). Recombination rate and selection strength in HIV intrapatient evolution. *PLoS Comput. Biol.* 6:e1000660. doi: 10.1371/journal.pcbi.1000660
- Ndowora, T., Dahal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N. E., et al. (1999). Evidence that badnavirus infection in *Musa* can originate from integrated pararetroviral sequences. *Virology* 255, 214–320. doi: 10.1006/viro.1998.9582
- Noreen, F., Akbergenov, R., Hohn, T., and Richert-Pöggeler, K. R. (2007). Distinct expression of endogenous Petunia vein clearing virus and the DNA transposon *dTph1* in two Petunia hybrida lines is correlated with differences in histone modification and siRNA production. *Plant J.* 50, 219–229. doi: 10.1111/j.1365-313X.2007.03040.x
- Obbard, D. J., Gordon, K. H., Buck, A. H., and Jiggins, F. M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 99–115. doi: 10.1098/rstb.2008.0168
- Pooggin, M. M. (2013). How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *Int. J. Mol. Sci.* 14, 15233–15259. doi: 10.3390/ijms140815233
- Pradillo, M., López, E., Linacero, R., Romero, C., Cuñado, N., Sánchez-Morán, E., et al. (2012). Together yes, but not coupled: New insights into the roles of RAD51 and DMC1 in plant meiotic recombination. *Plant J.* 69, 921–933. doi: 10.1111/j.1365-313X.2011.04845.x
- Prasad, A., Sharma, N., Muthamilarasan, M., Rana, S., and Prasad, M. (2019). Recent advances in small RNA mediated plant-virus interactions. *Crit. Rev. Biotechnol.* 39, 587–601. doi: 10.1080/07388551.2019.1597830
- Rajeswaran, R., Golyaev, V., Seguin, J., Zvereva, A. S., Farinelli, L., and Pooggin, M. M. (2014). Interactions of Rice tungro bacilliform pararetrovirus and its protein P4 with plant RNA-silencing machinery. *Mol. Plant Microbe Interact.* 27, 1370–1378. doi: 10.1094/MPMI-07-14-0201-R
- Richert-Pöggeler, K. R., Noreen, F., Schwarzacher, T., Harper, G., and Hohn, T. (2003). Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. *EMBO J.* 22, 4836–4845. doi: 10.1093/emboj/cdg443
- Rogers, K., and Chen, X. (2013). MicroRNA biogenesis and turnover in plants. *Cold Spring Harb. Symp. Quant. Biol.* 77, 183–194.
- Sanchez, D. H., Gaubert, H., Drost, H. G., Zabet, N. R., and Paszkowski, J. (2017). High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nat. Commun.* 8:01374. doi: 10.1038/s41467-017-01374-x
- Schmidt, N., Seibt, K. M., Weber, B., Schwarzacher, T., Schmidt, T., and Heitkam, T. (2021). Broken, silent, and in hiding: Tamed endogenous pararetroviruses escape elimination from the genome of sugar beet (*Beta vulgaris*). *Ann. Bot.* 2021:mcab042. doi: 10.1093/aob/mcab042
- Schorn, A. J., Gutbrod, M. J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon control by tRNA-derived small RNAs. *Cell* 170, 61.e–71.e. doi: 10.1016/j.cell.2017.06.013
- Schwarzacher, T., Heslop-Harrison, J. S. P., and Richert-Pöggeler, K. R. (2016). Analysis of petunia vein clearing virus (PVCV) sequences, retroelements and tandem repeats in *Petunia axillaris* N and *P. inflata* S6. Supplementary Note 2 In: The Petunia Genome Consortium. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* 2, 1–9. doi: 10.1038/nplants.2016.74
- Schwarzacher, T. (2003). Meiosis, recombination and chromosomes: A review of gene isolation and fluorescent in situ hybridization data in plants. *J. Exp. Bot.* 54, 11–23. doi: 10.1093/jxb/erg042
- Sepsi, A., and Schwarzacher, T. (2020). Chromosome-nuclear envelope tethering—a process that orchestrates homologue pairing during plant meiosis? *J. Cell Sci.* 133, 1–13. doi: 10.1242/jcs.243667
- Simon, D. M., Clarke, N. A. C., McNeil, B. A., Johnson, I., Pantuso, D., Dai, L., et al. (2008). Group II introns in Eubacteria and Archaea: ORF-less introns and new varieties. *RNA* 14, 1704–1713. doi: 10.1261/rna.1056108
- Stadler, J., and Richly, H. (2017). Regulation of DNA repair mechanisms: How the chromatin environment regulates the DNA damage response. *Int. J. Mol. Sci.* 18:18081715. doi: 10.3390/ijms18081715
- Staginnus, C., Gregor, W., Mette, M. F., Chee, H. T., Borroto-Fernández, E. G., Da Câmara, et al. (2007). Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species 24. *BMC Plant Biol.* 7:1–16. doi: 10.1186/1471-2229-7-24
- Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., and Bujarski, J. J. (2011). RNA-RNA recombination in plant virus replication and evolution. *Annu. Rev. Phytopathol.* 49, 415–443. doi: 10.1146/annurev-phyto-072910-095351
- Tromas, N., Zwart, M. P., Maité, P., and Elena, S. F. (2014). Estimation of the in vivo recombination rate for a plant RNA virus. *J. Gen. Virol.* 95, 724–732. doi: 10.1099/vir.0.060822-0
- Vitte, C., and Panaud, O. (2003). Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* 20, 528–540. doi: 10.1093/molbev/msg055
- Vijverberg, K., D'Agostino, N., and Gerats, T. (2016). Identification of conserved miRNAs in *Petunia axillaris* and *P. inflata* young flower buds and their verification in the Petunia genome sequence. Supplementary Note 9 In: The Petunia Genome Consortium. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* 2, 1–9. doi: 10.1038/nplants.2016.74
- Wang, Z., Hardcastle, T. J., Pastor, A. C., Yip, W. H., Tang, S., and Baulcombe, D. C. (2018). A novel DCL2-dependent miRNA pathway in tomato affects susceptibility to RNA viruses. *Genes Dev.* 32, 1155–1160.
- Wu, H., Li, B., Iwakawa, H. O., Pan, Y., Tang, X., Ling-Hu, Q., et al. (2020). Plant 22-nt siRNAs mediate translational repression and stress adaptation. *Nature* 581, 89–93. doi: 10.1038/s41586-020-2231-y
- Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362. doi: 10.1002/j.1460-2075.1990.tb07536.x
- Yu, H., Wang, X., Lu, Z., Xu, Y., Deng, X., and Xu, Q. (2019). Endogenous pararetrovirus sequences are widely present in Citrinae genomes. *Virus Res.* 262, 48–53. doi: 10.1016/j.virusres.2018.05.018
- Zhang, B., Li, W., Zhang, J., Wang, L., and Wu, J. (2019). Roles of small RNAs in virus-plant interactions. *Viruses* 11:827. doi: 10.3390/v11090827

**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.

Copyright © 2021 Richert-Pöggeler, Vijverberg, Alisawi, Chofong, Heslop-Harrison and Schwarzacher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.