


VNIVERSITAT DE VALÈNCIA

# Dynamics of RNA Virus Diversity and Host Transcriptomic Responses Across Time Scales:

## From Single Infection to Experimental Evolution

**María José Olmo Uceda**

Prof. Dr. Santiago F. Elena Fito *(Supervisor)*
Dr. Vicente Arnau Llombart *(Co-Supervisor)*

Doctoral Programme in Information Technologies, Communications and Computing
*Research line in Bioinformatics*

Escola Tècnica Superior d'Enginyeria

**May 2025**

# Dynamics of RNA Virus Diversity and Host Transcriptomic Responses Across Time Scales

## From Single Infection to Experimental Evolution

**María José Olmo Uceda**

This dissertation is submitted for the degree of
*Doctor of Philosophy*

Universitat de València May 2025

# INFORME FAVORABLE DIRECTORES DE TESIS

**Título de la Tesis:** *Dynamics of RNA Virus Diversity and Host Transcriptomic Responses Across Time Scales: From Single Infection to Experimental Evolution*

**Doctoranda:** Olmo Uceda, María José

**Director de tesis:** Elena Fito, Santiago Francisco

**Tutor de tesis:** Arnau Llombart, Vicente

**Co-director de tesis:** Arnau Llombart, Vicente

Los abajo firmantes, en su calidad director y codirector (y tutor) de la doctoranda María José Olmo Uceda, consideran que el trabajo realizado por la misma como parte de su tesis doctoral, así como todas las actividades formativas llevadas a cabo durante este período justifican sobradamente el depósito y defensa de la dicha tesis. Todos los trabajos realizados como parte de esta tesis han sido ya publicados en revistas listadas en el primer cuartil del *Science Citation Index*. Por lo tanto, emitimos un informe **FAVORABLE** para la realización del depósito y defensa de la tesis doctoral

Y para que así conste, firmamos la presente en Paterna en la fecha indicada,

Director de la tesis

Firmado digitalmente por ELENA FITO SANTIAGO FRANCISCO - DNI 25396106N
Fecha: 2025.05.13 11:42:03 +02'00'

Codirector/tutor de la tesis

Firmado por VICENTE ARNAU LLOMBART - NIF:***8241** el día 13/05/2025 con un certificado emitido por ACCVCA-120

Prof. Santiago F. Elena, PhD
E-mail: santiago.elena@cisc.es
Web: https://sfelenalab.csic.es
X and Bluesky: @SFElenaLab

i2sysbio

Parc Científic UV
Catedrático Agustín Escardino 9
46980 Paterna (València), Spain
Tel: +34 963 544 779

A mi padre
A mis madres

A Juanjo

A Marc

# Acknowledgements

Ninguna etapa de mi vida la he superado completamente sola, esta tesis no ha sido la excepción. Agradecer a todas las personas que me han acompañado en este camino sería una tarea interminable, pero quiero dedicar unas líneas a ello.

No podría haber llegado hasta aquí sin una educación pública, ni sin haber accedido a una beca predoctoral. Por eso, en primer lugar, gracias a nuestro estado del bienestar. Ojalá más personas con vocación científica puedan acceder a los recursos, el apoyo y las condiciones justas que les permitan desarrollar su potencial.

Gracias Santi, por confiar desde el principio y compartir tu ciencia conmigo. Has sido la persona que más me ha empujado a conocer nuevos espacios de conocimiento (y geográficos), a participar en proyectos apasionantes (y desafiantes) y a conocer a científiques brillantes (y aún mejores personas). A veces tienes el tacto de un estropajo pero es verdad que tu excelente gusto musical lo compensa todo. Aunque mi opinión sobre las jerarquías no ha cambiado, tu integridad científica y tu calidad humana me hacen sospechar que siempre te consideraré el jefe. Gracias por estos años de aprendizaje, ojalá muchos más.

Gracias Vicente, por acompañarme y orientarme desde el principio. Me has facilitado las cosas siempre que lo he necesitado. Tu disposición a escuchar y tu saber hacer me han impulsado a llegar aquí. Gracias por todo.

Gracias al EvolSysVir y al laboratorio 2.1 al completo. A los que lo forman ahora y a los que ya no, pero con los que he tenido la suerte de coincidir: Jose, Paqui, Christina, Silvia, Julia, Susana, Paula, Regis, Anamarija, Rubén, Ayoub, Ivair, Mónica, Miriam, Lara, Bea, Pablo, Noor, Tim, Celso, Mari Carmen, ... Compartir con vosotros

conversaciones científicas y no científicas ha sido un privilegio. Trabajar sabiendo que si levantaba la cabeza del ordenador iba a encontrar miradas de apoyo... eso no está escrito en la sección de métodos, pero es lo que me ha permitido conseguir muchos de los resultados. Sé que va a ser difícil encontrar un entorno de trabajo en el que me sienta tan en casa como con vosotros.

Gracias a las brillantes científicas que han compartido su trabajo conmigo, dejando que metiera mi ordenador en los datos que con tanta planificación y esfuerzo habíais conseguido: Silvia, Julia y Victoria. Esta tesis hubiera sido imposible sin vosotras. Sois la cabeza y las manos que hay detrás de esta ciencia. Gracias.

Y gracias en especial a mi preciosa y diversa tribu de predocs (y postdocs infiltrados): Izan, JC, Esme, Julia C., Victoria, Ana, Will, João, Dominik, Irene y Francesca. Sabéis lo que significáis para mi. Hemos reído y hemos llorado. Hemos viajado a sitios inesperados. Sabemos lo que es sentir el inicio de la primavera juntes ... Os debo palabras bonitas a cada uno de vosotros, porque la ciencia está muy bien, pero a veces, si no hubiera sido por vosotros, yo me hubiera vuelto a la música. Solo espero haberos aportado una pequeña parte de lo que vosotros me habéis aportado a mi. He sido feliz con vosotres y os quiero.

Hay gente que va por la vida regalando sabiduría y yo he tenido la suerte de toparme con alguna. Gracias a esas personas que, sin esperarlo, han pasado a formar parte de la amplia red que me ha sostenido durante este viaje.

Gracias a José Ángel, mi no director (aunque Christina llegara a pensar que lo era). Gracias por ser tan generoso con tu tiempo. Tu manera de acercarte a la ciencia, de querer entenderlo todo de esa forma tan directa que tienes, es una inspiración constante. Lo supe en la unidad de clase que recibí de ti durante el máster y lo he confirmado durante estos años. Gracias por tu apoyo.

Gràcies a Josep, que malgrat el seu horari, més que saturat i les seues noves funcions vitals, sempre ha trobat temps per a tirar-me el cable que necessitava.

Gràcies a la gent del CRM i voltants: a més de Josep, Lluís, Tomàs L., Tomàs A., David, Pau,... El poquet de temps que he pogut compartir amb vosaltres ha sigut del més estimulant. Si continue en la ciència (que voldria!) sé que buscaré trobar a gent en la qual puga replicar l'ambient de col · laboració que he viscut amb vosaltres.

Thanks to the Malik Lab: Harmit, Tamanash, María, Janet, Aida, Rechel, Peter, Phoebe, Ching-Ho, Pravrutha, Eva, Tiia, ... I shared with you some of the most pleaseant — but also the most difficult — times of my life. You showed me a kind of kindness I didn't expect, right when I needed it most. Thanks for everything.

Thanks to the rest of the scientists I've shared collaborations, meetings or discussions: Asher, Daria, Phuong Anh, Ron, ... I feel I've learned something from each of you. Thank you for sharing your time and your science with me.

Thanks to Alexandra Elbakyan. Without her work, this Thesis would have less references.

Por supuesto, gracias a mi gran y ruidosa familia, a mis hermanas (a la pequeña y a las mayores), a mis sobrinos, a mis suegros, a mi tía Vícen, ...

Gracias a mis padres, las mejores personas que he conocido y los fans más incondicionales que he tenido. Siempre he sentido vuestro amor y vuestro apoyo. Estos últimos años no han sido fáciles pero aún recuerdo vuestra alegría cuando os mandé la primera nómina del CSIC (y obviamente no fue por el salario). Sé que papá estaría orgulloso de verme acabar, a pesar de todo. Gracias mamá por enseñarme lo que es ser una mujer fuerte y generosa.

Gracias Juanjo, por ser el compañero más paciente, altruísta y el menos heterobásico con el que podía soñar. Tu contribución moral y logística en el desarrollo de mi yo adulto es incalculable. Gracias por todos estos años, por tu apoyo incondicional en todos, pero en este periodo en particular, y por la feliz co-autoría de nuestro hijo.

Gracias Marc, eres la persona más dulce, empática y divertida que conozco. Solo quiero que te dejemos un mundo en el que la ciencia esté por encima de los intereses económicos de unos pocos y en el que ser buena persona sea la norma. Queda mucho por hacer.

A vosotros os dedico esta tesis, porque sin vosotros no me hubiera atrevido a, con ya 30 años, empezar en la ciencia, pero mucho menos, a continuar hasta aquí.

Y, citando al gran Snoop Dogg (Esme, wink-wink),

*Last but not least, I wanna thank me [...]*
*I wanna thank me for tryna do more right things than wrong [...]*
*I wanna thank me for never quitting.*

# Abstract

Viruses are obligate intracellular symbionts that rely entirely on host cellular machinery to complete their life cycles. Despite their simplicity, they are among the most genetically diverse and evolutionarily successful biological entities on Earth. They infect all forms of life—from bacteria and archaea to plants and animals—and play pivotal roles in ecosystem dynamics, evolution, and global health. Among viruses, those with RNA genomes stand out due to their high mutation rates and effective population sizes, which enable rapid adaptation to new hosts and environments. These traits also make RNA viruses powerful model systems for studying evolution, population genetics, and host-pathogen interactions.

RNA viruses typically encode a small number of proteins, often using compact genomes of 1.8-30 kb. This minimalism forces them to rely on multifunctional proteins and extensive host manipulation. Their high error rates, resulting from the lack of proofreading capabilities in most of the RdRPs, generate a swarm of genetic variants during each infection cycle. This heterogeneous collection, known as a viral quasispecies, allows viruses to explore vast areas of sequence space and quickly respond to selective pressures. A particularly intriguing component of viral quasispecies are defective viral genomes (DVGs). DVGs arise during replication. Although they are typically replication-deficient on their own, DVGs can be replicated in trans by co-infecting helper viruses. Understanding how these populations evolve and adapt in different host contexts is central to virology, with implications for antiviral strategies, vaccine design, and our understanding of molecular evolution.

RNA viruses, despite their small genome sizes, behave as complex, evolving populations subject to mutation, recombination, selection, and drift. Their dynamics are influenced by within-host processes as well as between-host transmission. Different replication strategies, like stamping machine versus geometric replication, influence how mutations accumulate and spread. These factors collectively shape the evolutionary trajectory and phenotypic outcomes of infection.

Equally important are the responses mounted by host organisms. Over time, hosts have developed mechanisms to detect and eliminate viruses. Innate immunity provides a rapid antiviral defense. Despite taxonomic differences, many antiviral responses are conserved, especially RNA interference (RNAi). The study of host responses in model organisms provides valuable insights into the universal principles of antiviral defense.

In this thesis, I present a systems biology approach to investigate the dynamic interplay between viral evolution and host transcriptomic responses. Drawing on four published articles, this work integrates the analysis of RNA sequencing data across multiple host-virus systems and evolutionary

timescales to explore how viral populations adapt, and how hosts respond at the phenotypic and molecular levels. The thesis addresses two different scale times: (1) the long-term evolutionary trajectories of three (+)ssRNA viruses: TuMV in *A. thaliana*, and the betacoronaviruses MHV and HCoV-OC43 in mammalian cells— and (2) the short-term infection dynamics of the OrV in *C. elegans*. The chapters are also organized into two main axes: one focused on viral diversity; the other on host responses to infection.

This integrative and computational approach seeks an understanding of how viral populations evolve in different biological contexts, and how hosts reshape their transcriptomes in response to infection.

# Extended Abstract

**Chapter 1: Introduction**

Viruses are obligate intracellular symbionts that rely on host cellular machinery for replication. Despite their simplicity, they are genetically diverse and evolutionarily successful. RNA viruses, in particular, possess compact genomes, high mutation rates, and rapid replication cycles. These features allow them to quickly adapt, evade immunity, and cross species barriers, making them ideal models to study evolution, population genetics, and host–pathogen interactions.

With genomes ranging from 1.8 to 31 kb, RNA viruses often use overlapping reading frames and multifunctional proteins to maximize coding efficiency. Replication by RNA-dependent RNA polymerases (RdRPs), which lack proofreading activity, results in high mutation rates and the formation of *quasispecies*—a cloud of genetically related variants. This diversity enables the exploration of vast sequence space, providing the raw material for selection and rapid adaptation.

Population genetics helps explain how mutation, recombination, selection, and drift shape viral evolution. Replication modes—*geometric* or *stamping-machine*— influence how mutations accumulate. Geometric replication amplifies mutant genomes rapidly, whereas stamping-machine replication limits the spread of mutations. Host immune responses, replication bottlenecks, and within-host dynamics also influence evolutionary trajectories.

A striking feature of viral populations is the presence of *defective viral genomes* (DVGs), produced during replication. Although incapable of autonomous replication, DVGs can be replicated in trans by wild-type (WT) viruses. Some DVGs, known as *defective interfering particles* (DIPs), compete with WT viruses for replication resources, thereby interfering with their replication. First identified in influenza virus, DVGs are now recognized in many RNA viruses. Long considered replication byproducts,

DVGs are now understood to be functionally relevant. They can attenuate infections by reducing WT replication and by activating strong innate immune responses. Some DVGs may encode peptides or regulatory RNAs, suggesting broader functional potential. They also influence viral evolution: at high multiplicity of infection (MOI), interactions between WT and DVGs resemble social systems, where DVGs can act as *cheaters* exploiting WT replication. Yet DVGs can also support persistence or modulate host immunity, revealing a spectrum of cooperative and antagonistic roles. Transmission bottlenecks may purge or fix DVGs depending on selection and genetic drift. Thus, DVGs are integral to viral ecology and evolution. They modulate replication, immunity, and genetic diversity, challenging traditional views of RNA virus populations and highlighting the importance of population dynamics in viral adaptability.

Viral evolution is also driven by host defenses. Over time, hosts have developed mechanisms to detect and eliminate viruses. While innate immunity provides a rapid, nonspecific defense, adaptive immunity offers specificity and memory. Despite taxonomic differences, many antiviral responses are conserved, especially RNA interference (RNAi).

Model organisms help dissect these responses. In plants, *Arabidopsis thaliana* is a key model due to its genetic resources and short life cycle. Its antiviral strategies include RNA silencing, translational repression, and hormonal signaling involving salicylic acid, jasmonic acid, or ethylene. Dicer-like proteins and Argonaute effectors guide RNAi- based responses against viral RNAs.

Epigenetic regulation also plays a role. DNA methylation, often guided by small RNAs via the RNA-directed DNA methylation (RdDM) pathway, can silence viral sequences and shape immune memory. Yet how these modifications feed back into virus evolution remains underexplored.

In animals, *Caenorhabditis elegans* is another model system. Orsay virus (OrV), the only natural virus known to infect *C. elegans* causes mild symptoms, suggesting a long-standing coevolutionary relationship. Like plants, *C. elegans* must balance effective immune responses with the risk of self-damage—illustrating the trade-off between immunity and self-tolerance.

These models reveal how hosts detect viral threats, orchestrate immune responses, and evolve under pressure, offering insights into conserved antiviral strategies across life.

Virus–host interactions are fundamentally coevolutionary, shaped by reciprocal adaptations over time. Viruses evolve mechanisms to exploit host machinery, while hosts develop countermeasures, creating an evolutionary arms race. The *Red Queen*

hypothesis illustrates this dynamic: both must continually evolve to maintain relative fitness. While viruses mutate rapidly, their sequence space is constrained by functional demands. In contrast, host defenses often evolve under fewer constraints, allowing greater diversification [162].

Virulence–transmission trade-offs are central to virus evolution. High virulence may reduce host survival and limit viral spread, whereas lower virulence can enhance transmission. For instance, reduced virulence in HIV-1 correlates with increased transmissibility [12]. Anderson and May [6] demonstrated that, depending on host availability and transmission dynamics, optimal virulence may be low, high, or intermediate. Building on this, Ewald [50] proposed that transmission mode shapes virulence evolution. Vertical transmission (from parent to offspring) tends to favor low virulence or mutualism, while horizontal transmission (between individuals) often selects for increased virulence. Empirical data for analyze this theory exists, particularly in vector-borne and plant virus systems [57, 125, 60].

Viruses also coopt host factors to facilitate replication. In mammals, viral proteins often hijack components of innate immunity for their benefit [84]. In *Arabidopsis*, the TGA1 transcription factor—typically involved in antibacterial responses—interacts with the RdRP of TuMV to aid viral replication [110, 149]. Hosts can likewise coopt viral sequences. Over time, *endogenous viral elements* (EVEs) can become part of host genomes, contributing to immunity or novel functions such as placental development. These domesticated viral genes illustrate the evolutionary interplay between viruses and hosts [55].

**Chapter 2: Aims, structure of the Thesis, and other contributions**

In this Thesis, I present a systems biology framework to investigate the dynamic interplay between viral evolution and host transcriptomic responses. Drawing on four published articles, this work integrates RNA sequencing data across multiple host–virus systems and evolutionary timescales to explore how viral populations adapt, and how hosts respond at the phenotypic and molecular levels.

The Thesis addresses two complementary aspects of RNA virus biology: (1) the long-term evolutionary trajectories of three (+)ssRNA viruses—TuMV in *A. thaliana*, and the betacoronaviruses murine hepatitis virus (MHV) and human coronavirus OC43 (HCoV-OC43) in mammalian cells—and (2) the short-term infection dynamics of the OrV in *C. elegans*.

By combining experimental evolution, high-throughput sequencing, and computational modeling, the work provides a multiscale perspective on virus–host interactions, spanning from single infections to experimental evolution.

The studies are also organized into two main axes: one focused on viral diversity, quasispecies dynamics, and the emergence of defective viral genomes (DVGs); the other centered on host responses to infection. This integrative approach enables a comprehensive understanding of how viral populations evolve in different biological contexts, and how hosts reshape their transcriptomes in response to infection.

The general and specific objectives of this Thesis are:

- To apply a systems biology approach to the study of viral infections caused by positive-sense single-stranded RNA ((+)ssRNA) viruses, focusing on both viral dynamics and host responses in two distinct temporal contexts: short-term and long-term.
  - To study the process of TuMV adaptation to *A. thaliana* genotypes with defects in different epigenetic pathways.
  - To analyze the transcriptional response of *A. thaliana* genotypes with defects in epigenetic pathways infected with *A. thaliana*-naïve and well-adapted TuMV isolates.
  - To study the accumulation of DVG of two betacoronaviruses during their evolution in different types of susceptible mammalian cells.
  - To study the accumulation of OrV, its mutational spectrum, including DVGs, along the infection of *C. elegans*.
  - To map the transcriptomic responses of *C. elegans* during the process of an OrV infection.

The work in this Thesis would not have been possible without the collaboration with the scientists who performed all the experiments. By order of appearance of the published papers: Dr. Silvia Ambrós, Dr. Julia Hillung, and Dr. Victoria G. Castiglioni have been part of my development as a scientist and of the results presented here.

From my side, I performed all data analyses presented in this Thesis. This includes RNA-seq preprocessing and quality control, variant calling and viral diversity quantification (single nucleotide variants, nucleotide diversity measures, and DVG identification and characterization), host transcriptomic analysis (differential expression, dynamic expression clustering, functional enrichment, ...), most of the statistical analyses, and modeling of viral replication dynamics. Where appropriate, custom pipelines were developed or extended to meet the specific needs of each dataset. Most of

the code developed is openly available in the repositories referenced in the corresponding Methods sections of each chapter.

Some of the computational tools I developed and applied in this work—though not central to the Thesis—have led to independent publications, such as *DVGfinder* [123] and *SDgenerator* [120], which provide reusable resources for the study of DVGs in RNA-seq data. The rest of the contributions published during this time are listed in Chapter 2.

**Chapters 3 to 6: Published results**

Chapters 3, 4 and 5 correspond to three published papers and focus on the long-term evolutionary dynamics of viruses. Chapters 3 and 4 focus on the long-term evolution of an RNA virus infecting plant hosts with altered epigenetic regulators of defense responses. These chapters are based on an experimental evolution study of TuMV in *A. thaliana* mutants with defects in key epigenetic pathways.

**In Chapter 3**, published in *Evolution*, we investigated how an emerging RNA virus evolves, interacts, and adapts to populations of a novel host species with disrupted, epigenetically controlled defense mechanisms. Genes involved in epigenetic regulation in *Arabidopsis thaliana* can modulate the plant's susceptibility to viruses, but whether they also influence viral evolution remains unknown. Here, we investigated how host genes involved in epigenetic processes shape the phenotypic evolution of TuMV. We evolved TuMV lineages for 12 passages in four *A. thaliana* genotypes, each impaired in different components of two main epigenetic pathways: DNA methylation (mutants in *dcl2 dcl3 dcl4*, *ddm1* and *polv*) and histone modification (mutant in *jmj14*). In each passage, we recorded the area under the disease progression curve (AUDPS), infectivity (I), and viral load (VL), and analyzed the rates of phenotypic evolution. All three traits evolved significantly during the experiment, with AUDPS and I increasing in all genotypes, and VL exhibiting more heterogeneous dynamics. These traits evolved faster in genotypes with higher basal permissiveness. Phenotypic evolution converged across lineages, suggesting that host-imposed selection constraints limit the virus's adaptive landscape. Analysis of the viral diversity (high-throughput sequencing of the first (P1) and the last passage (P12) were performed) showed a net increase in genetic diversity across viral lineages. Analysis of viral genetic diversity within TuMV lineages (measured as APD) showed a significant overall increase after 12 passages ($\sim 30.2\% \pm 2.9$), with statistically significant differences among plant genotypes. The extent of diversity gain depended on both genotype and passage number, ranging from 20.0% in *ddm1* to 44.7% in *dcl2 dcl3 dcl4*. WT-evolved lineages showed the second lowest increase. Genomic distribution of diversity revealed an increase in segregating

sites from 518 (passage 1) to 1,206 (passage 12), with most being singletons (82.0% and 86.1% respectively). Nucleotide diversity within host genotypes increased 2.11-fold, and total sample diversity increased 2.23-fold. Interhost genotype diversity rose 48.3-fold, indicating strong genetic differentiation between host environments. The nucleotide differentiation coefficient (NST) also significantly increased, showing that initially homogeneous populations diverged due to host-specific evolution. Tajima's D test values were significantly negative after both passages ($D_1 = \breve{}2.457$, $D_{12} = \breve{}2.567$), suggesting purifying selection, rapid expansion, or segregating deleterious mutations. The high proportion of singletons supports population expansion during systemic infection. Cistron-level $d_N \breve{} d_S$ analysis indicated pervasive purifying selection in nine cistrons, becoming more intense by passage 12. NIb was especially affected, showing initial variability that was later homogenized (except in *polv*). Only VPg exhibited signs of diversifying selection, stronger early on and weakening over time. FEL codon analysis confirmed purifying selection at codons G3, Y25, and P78. Nine non-synonymous mutations were under positive selection. Codon 79 initially showed L79F and L79I variants that disappeared by passage 12. Eight other selected mutations clustered in VPg (aa 110–120), often reverting to ancestral alleles by passage 12. For example, codon 115 displayed seven different new alleles after passage 1 but mostly returned to ancestral N by the end. This transient diversification followed by reversion is consistent with leap-frog dynamics, suggesting reappearance of early genotypes over recent variants. Reversion is unlikely to be mutational; instead, persistence of ancestral variants at low frequency is a more parsimonious explanation.

No consistent association was found between fixed mutations and host genotypes, with the exception of R118H, present only in *ddm1*-evolved lineages and previously linked to resistance contexts. VPg, a known selection target in potyviruses, plays essential roles in viral replication, translation, and protein interactions. Its involvement in adaptation has been observed in multiple studies, especially regarding interactions with host eIF factors and resistance-breaking events.

The lack of fixed, host-specific mutations despite phenotypic divergence may seem surprising. This, combined with parallel changes across genotypes, suggests adaptation may reflect shared environmental pressures rather than specific host constraints. However, this interpretation may be reconciled by invoking low-frequency variants below detection thresholds that collectively determine phenotype, as described in other viral systems. Thus, viral quasispecies complexity, rather than consensus sequences, may drive host-specific adaptation under epigenetic pressure.

**In Chapter 4**, published in *BMC Genomics*, we investigated the transcriptomic responses of both mutant and wild-type *A. thaliana* genotypes to infection with ancestral or evolved TuMV isolates from the dataset studied in Chapter 3. Thee different contrast strategy were followed in the transcriptomics analysis. First, the differences between each mutant transcriptome and the WT in the same condition (mock, infected with naïve virus and infected with evolved virus) were investigated. In mock conditions, *dcl2 dcl3 dcl4* exhibited the largest number of differentially expressed genes (DEGs, 2073), followed by *ddm1* (809). Upon infection with the ancestral virus, *ddm1* showed the most DEGs (1267), and after infection with evolved viruses, *ddm1* and *polv* exhibited a rise in DEGs, while *jmj14* and *dcl2 dcl3 dcl4* showed decreases. A total of 74 DEGs were shared among all four mutant genotypes after evolved-virus infection, with most being underexpressed. The proportion of shared DEGs increased from 1% in mock conditions to 72% with evolved viruses. The transcriptomic perturbation varied among mutants, with *jmj14* showing the strongest selective pressure. Epigenetic mutants show changes in a core set of common genes. Six DEGs were overexpressed in all genotypes, independent of infection status, including genes involved in methionine biosynthesis and defense (MRD1), flower development (PI4KC3), and noncoding RNAs. These genes are regulated by RdDM. Additionally, five DEGs were shared by infected plants, irrespective of virus evolution, related to stress responses, meiosis, and methyltransferase MET2. Second, the impact of viral adaptation on plant responses was assessed by comparing transcriptomes of plants infected with evolved (P12) vs. ancestral (P1) viral lineages. Mutants like *jmj14* and *polv* exhibited more significant gene expression changes than WT plants, with altered distributions of activated and repressed genes. Functional enrichment analysis highlighted common pathways, such as salicylic acid response, activated by evolved viruses, and specific categories, like carotenoid biosynthesis, enriched in particular mutants like *ddm1* and *jmj14*. A core set of 1029 genes was differentially expressed across all plant genotypes, linked to TuMV adaptation. These genes showed consistent up- and down-regulation patterns, with specific exceptions in some genotypes. Functional analysis revealed enriched processes related to chitin response, salicylic acid signaling, and jasmonic acid metabolism. Third, by including the gene expression analysis between the plants infected with the naïve virus and mock conditions, we classified each gene into one of the nine basic expression profiles. All these analysis revealed both common targets of viral adaptation and genotype-specific transcriptional responses. As expected, immune-related genes were differentially expressed upon infection. However, additional functional categories were

also overrepresented, suggesting that viral adaptation is driven by factors beyond classical resistance genes.

Additionally, a prior study by Corrêa *et al.* showed that TEs in centromeric and pericentromeric regions were induced early in WT plants infected with evolved TuMV. These TEs were more prominently induced or repressed at later infection stages. In our study, we found a significant correlation between DEGs and TEs near their transcription start sites (TSS). While the proportion of TE-associated DEGs was slightly higher in all mutants than in WT, *ddm1* plants showed a significant increase in such DEGs. The data suggest that TEs near gene promoters are more likely to correlate with gene expression changes than those located within genes. Notably, *rpp7*, associated with disease resistance, had a high density of TEs within its sequence and was repressed in evolved-infected *ddm1* and *jmj14* plants. This gene is regulated by an epigenetic mechanism involving silencing of a "domesticated" TE within its intron. An enrichment analysis of TEs associated with DEGs responding to TuMV adaptation revealed a significant overrepresentation of class II elements (DNA transposons) and underrepresentation of class I elements (retrotransposons). Class I TEs, particularly Gypsy, were significantly less frequent across genotypes, except in *ddm1* plants where certain SINE and LINE types were enriched. On the other hand, class II transposons, especially Helitrons, were overrepresented, with a marked enrichment in both the promoter regions and coding sequences of DEGs, suggesting their potential role in epigenetic conflict and stress response pathways.

Finally, we explored how TuMV adaptation might alter the host-virus protein-protein interaction (PPI) network by integrating our transcriptomic results with existing PPI data. The results showed that in *ddm1* plants, there was a significant overrepresentation of activated interactors, especially for viral proteins like P3 and P3N-PIPO. VPg interacted with both repressed and activated proteins, some involved in translation initiation and chromatin remodeling, while the viral replicase NIb showed altered interactions with transcription factors and ubiquitin ligases. Notably, 6K1, a viroporin described as interactor of PAL1, a phenylalanine ammonia-lyase, suggesting a potential link to viral adaptation and the phenylpropanoid pathway, which plays a role in plant immunity. These findings point to important future directions in virus-host interaction research.

**Chapter 5**, published in *Viruses*, presents a separate experimental evolution study focused on the generation and accumulation of DVGs in two different betacoronviruses. We analyzed the long-term dynamics of DVGs during serial passages of the two aforementioned betacoronaviruses, HCoV-OC43 and MHV, in cell culture. The

combinations virus-studied were: HcoV-OC43 in BHK-21 and in HCT-8 cells and MHV in CCL-9.1 cells. Viruses were also evolved under high and low multiplicities of infection (MOI). RNA-seq was performed at passages 0, 11, 23, 35 and 47 passages. We *in silico* identified and characterize DVGs and analyzed their accumulation patterns. DVGs exhibited increased diversity and abundance over time, with deletions and insertions as the most common features in richness and abundance. In HCoV-OC43, DVG size distributions and dynamics differed between MOI conditions. Short and long DVGs dominated under low MOI, while additional size clusters emerged and became prevalent at high MOI suggesting a selection trend. For MHV, DVG size distribution was more uniform, with fewer pronounced differences between MOI treatments. We also identified hotspot regions for deletions in structural and accessory protein-coding regions, particularly under high MOI. E and Nsp12, the two viral proteins with viroporin potential, were identified as the most affected by deletions. Finally, we study the persistence of some DVGs, especially deletions and identify hotspots of persistente DVGs accumulate. The identification of the same persistent deletions in independent lineages reinforced the idea that some DVGs must be under positive selection. These results highlight the pervasive generation and evolution of DVGs, shaped by both virus-specific and environmental factors, and their potential role in modulating viral fitness and persistence.

Chapter 6 focus on the short-term dynamics of OrV infection on *C. elegans*. In Appendix A, there are additional results from Chapter 6, focused exclusively in the viral dynamics.

**In Chapter 6**, published in *Science Advances*, we examined the progression of OrV infection during larval development in *C. elegans*, aiming to understand how viral load and host responses evolve over time. OrV is the only known natural virus of *C. elegans*, and it generally has minimal impact on fitness due to the host's robust innate immune system. The progression of the infection was track from 0 to 44 hours post inoculation (hpi) with different methods (RT-qPCR, RNA-seq and smiFISH). We identified four distinct stages of infection, marked by dynamic changes in viral load, with RNA2 (capsid-encoding) peaking at the onset of viral egression. *(i)* A prereplication phase (0 to 6 hpi), *(ii)* an exponential viral replication that leads to a viral load peak (8 to 12 hpi), *(iii)* a plateau phase with moderate viral load levels (approximate half of the first pick)(14 to 36 hpi), and *(iv)* a phase of persistent residual infection or resolution of the infection, in which the viral load falls but remains above initial levels (38 to 44 hpi) . The transcriptomic analyses of twelve time points allow us to characterize the temporal shifts in the animal's gene expression. First, through

the identification of the differentially expressed genes at each stage, we pointed to a core set of 29 genes ubiquitous to the exponential viral replication and moderate stages (such as *B0507.6*, *col-133*, *ddn-1*, *droe-8*, *eol-1*, *fbxa-182* or *math-15*). Twenty-one of them belong to the intracellular pathogen response and others have been described previously as ZIP-1-dependent. The rest of the genes—*col-133*, *math-15*, *trpl-5*, *Y43F88.12* and *Y6G8.5*—had not been previously associated to the animal's response to OrV. A functional enrichment analysis on the up- and down-regulated genes per time point revealed two major waves of transcriptional activity, one between 12 and 18 hpi and another at 26 hpi, involving both gene ontologies and metabolic pathways. Early responses included up-regulation of RNA surveillance and axon guidance genes, while genes related to mitochondrial translation, cuticle formation, and fatty acid metabolism were down-regulated at distinct stages of infection. Second, to identify genes with expression dynamics significantly influenced by infection over time, we applied a likelihood ratio test (LRT) for the interaction condition and time. This approach uncovered 2,313 DEGs, which were grouped into clusters based on their temporal profiles, revealing distinct patterns of early and late regulation. Notably, 45 genes exhibited strong positive correlations ($r \geq 0.8$) with viral load, including known IPR and ZIP-1–dependent genes, while 13 others, such as *elo-9* and *clec-24*, showed anticorrelated patterns. Third, to distinguish OrV-specific responses from general responses to other pathogens, we compared our data with 56 transcriptomic datasets, including infections by Gram-positive and Gram-negative bacteria, fungi, and other OrV experiments. This comparison identified 188 genes uniquely altered by OrV and 208 genes commonly affected across all pathogen categories. OrV-specific genes showed diverse expression profiles and were enriched in cytoplasmic and mitochondrial components, amino acid biosynthesis, peptide-related processes, and the pentose phosphate pathway. Shared genes were enriched in the extracellular region, innate immune responses, collagen and cuticle functions, and lipid biosynthesis, and included 60% of IPR genes and numerous antimicrobial effectors. Finally, the experimental validation of some of the identified genes, allow us to suggest new antiviral (*lys-2*, *sago-2*) and proviral factors (*C53A5.9*, *clec-60*, *eol-1*, *ppw-1* and *vsra-1*).

In *Appendix A*, we show **additional results** from those published in **Chapter 6**. The extra results are focused on short-term evolution and population dynamics of OrV during the primary infection in *C. elegans*. OrV was found to follow a stamping machine replication model, with viral populations dominated by progeny from early templates. Viral diversity analysis revealed nucleotide diversity in the negative strand that was not transmitted to the positive strand, suggesting replication constraints

within host cells. Micro- to macro-evolutionary analysis highlighted the RdRP tail as a region under evolutionary conflict, potentially reflecting host-virus interaction pressures. We also report the first evidence of DVGs in OrV, identifying a recurrent deletion in RNA2 that may function as a subgenomic RNA producing a truncated free delta protein. Additionally, prediction of new small open reading frames (smORFs) revealed a novel peptide candidate, offering new insights into OrV's coding potential and evolutionary strategies.

Together, the studies in this Thesis make several original contributions:

- Demonstrating that viral adaptation is strongly conditioned by host genotype, particularly in the context of epigenetically regulated immune pathways.
- Providing a systems-level view of plant transcriptomic responses to viral evolution, revealing shifts not only in classical immune genes but also in metabolic and regulatory processes.
- Characterizing the evolution and accumulation patterns of DVGs in two betacoronaviruses in different contexts (cell types and MOI), and identifying how the different conditions shape their diversity while providing new ways to in silico analyze the DVG population.
- Describing the temporal dynamics of host transcriptional responses during OrV infection, including the identification of sustained and stage-specific responses.
- Characterizing for the first time the accumulation of genome and antigenome in the short-term dynamics of OrV and proposing the stamping machine model as the predominant replication mechanism.
- Presenting the first evidence for DVGs in OrV and suggesting the presence of a subgenomic RNA and a new peptide candidate.

These findings collectively support a model in which RNA virus evolution is closely intertwined with host regulatory mechanisms, and where the genetic diversity within viral populations—including defective genomic elements—plays a key role in determining the outcomes of infection. The integrative methodology used in this study offers a framework for future investigations at the intersection of evolutionary virology and host transcriptomic responses.

In **Chapter 7**, I present a **general discussion** exploring some of the results obtained. Such as the evolution of virulence, the existence of trade-offs in experimental evolution, or the role of DVGs in viral adaptation.

**Chapter 8: Conclusions**

*Conclusions on Long-Term Virus–Host Interactions:*

The long-term coevolutionary dynamics between viruses and their hosts revealed several critical insights. First, epigenetic regulation exerts a profound influence on the antiviral responses of *A. thaliana* and the evolution of turnip mosaic virus. TuMV evolved in mutant plants defective in DNA methylation and histone modification pathways showed increased viral accumulation and enhanced virulence. Notably, more permissive host genotypes tended to select for less virulent viral forms, whereas more resistant genotypes favored the evolution of more virulent strains.

In these epigenetically compromised backgrounds, TuMV exhibited accelerated evolution of phenotypic traits compared to those evolving in wild-type plants. A focal point of adaptation was the viral protein VPg, which emerged as a major target of selection regardless of the plant genotype. The flexible loop region of VPg, encompassing residues 79 to 120, accumulated mutations under positive selection pressure.

Importantly, viral adaptation had reciprocal effects on host gene expression. A core set of 1,029 genes displayed altered expression across all epigenetic mutants as well as wild-type plants following viral evolution. Genes with increased expression were mainly associated with salicylic acid (SA) and systemic acquired resistance (SAR), ethylene signaling, and responses to chitin, indicating an amplified defense response. In contrast, downregulated genes were involved in jasmonic acid (JA), amine and aromatic amino acid metabolism, and the suppression of phenylpropanoid biosynthesis—reflecting a costly trade-off between defense activation and plant growth, ultimately resulting in weaker plant phenotypes.

Furthermore, the nature of epigenetic modification influenced gene expression near transposable elements (TEs). Mutants in DNA modification pathways (*ddm1* and *polv*)—but not in histone methylation (*jmj14*)—affected gene expression in regions near TEs, particularly when TEs were located close to gene promoters. These associations suggest that promoter-proximal TEs are more likely to influence gene expression changes than those within gene bodies.

Integration of host transcriptomics and the virus–host protein–protein interaction network (PPIN) enabled the identification of 13 *Arabidopsis* genes critical for viral adaptation. Among these, the downregulation of *pal1* —which encodes the first enzyme in the phenylpropanoid synthesis pathway and physically interacts with the TuMV viroporin 6K1— by evolved viral strains may contribute to the observed weakened plant phenotype.

*Conclusions on the Accumulation and Diversity of DVGs in Coronaviruses:*

The accumulation and diversity of defective viral genomes (DVGs) during the evolution of human coronavirus OC43 (HCoV-OC43) in BHK-21 and HCT-8 cells, and of murine hepatitis virus (MHV) in CCL-9.1 cells, were found to depend on both cellular and viral factors. DVGs not only accumulated in both viruses but also evolved, with high multiplicity of infection (MOI) conditions promoting greater diversity and the persistence of specific DVG forms. This pattern suggests the presence of selection mechanisms favoring fitter DVGs that could impact viral evolution.

The most prevalent DVGs involved deletions, particularly within the *E* and *ns12.9* cistrons. These deletions may influence pathogenicity and were observed to arise independently across different lineages and cell types, indicating a potential adaptive advantage that may reflect a trade-off supporting host cell survival.

The evolution of DVGs was shaped by both viral genotype and host cell identity. HcoV-OC43 showed more defined and structured deletion patterns compared to MHV, emphasizing the importance of host-specific factors and supporting the use of multiple cell lines in future DVG research.

Persistent deletions likely undergo positive selection to modulate viral replication and pathogenesis, raising the possibility of exploiting DVGs as therapeutic agents. DVGs may inhibit wild-type virus replication or serve as broad-spectrum antivirals.

Moreover, high MOI conditions led to well-defined DVG accumulation patterns, particularly in deletion hotspots affecting structural and accessory protein-coding cistrons. In contrast, low MOI conditions resulted in more variability in DVG size distribution, underscoring MOI as a key factor in the dynamics of DVG evolution.

*Conclusions on the accumulation and diversity dynamics of OrV during the course of a primary infection in C. elegans and the transcriptomic responses of the animals:*

During a primary infection, OrV exhibits a stamping machine replication mode.

It produces a wide variety of DVGs, with deletions and single-nucleotide insertions being the most abundant. Specific deletion hotspots, as well as 5' and 3' cb structures, suggest that DVG generation is not entirely random but may be driven by as-yet unknown molecular mechanisms.

A persistent deletion in RNA2 is compatible with the transcription of a subgenomic RNA that may potentially lead to a truncated version of the free delta protein. This result, obtained *in silico*, requires experimental confirmation.

Additionally, we identified a putative peptide encoded by the antisense strand of RNA1. This putative product also awaits for experimental validation.

Evidence of selective pressure at both the micro- and macroevolutionary levels was found in the carboxy-terminal tail of the RdRP, highlighting this region as a possible key functional element in the OrV life cycle.

This work also provides the highest temporal resolution to date of the transcriptomic response of *C. elegans* to a primary OrV infection. By classifying genes based on their expression dynamics, we identified early-response genes, genes whose expression patterns closely correlated with viral load, and late-response genes associated with recovery. Comparisons with host responses to other virulent pathogens revealed a unique set of OrV-specific response genes.

Finally, experimental validation of several of these genes confirmed their functional roles, identifying new proviral factors (*C53A5.9*, *clec-60*, *eol-1*, *ppw-1*, and *vsra-1*) and antiviral factors (*lys-2* and *sago-2*) in the *C. elegans* response to OrV infection.

# Table of contents

**Abstract** **xi**

**Extended Abstract** **xiii**

**List of figures** **xxix**

**List of tables** **xxxi**

**1 Introduction** **1**

1.1 Introduction to viruses, RNA viruses and DVGs . . . . . 1

1.1.1 RNA viruses . . . . . 2

1.1.2 Defective viral genomes . . . . . 5

1.2 RNA viruses as evolving populations . . . . . 7

1.2.1 Viral Evolution . . . . . 7

1.2.2 Viral Diversity . . . . . 7

1.2.3 Population dynamics in viral evolution . . . . . 8

1.3 Host Responses to viral infections . . . . . 13

1.3.1 Host immunity . . . . . 13

1.3.2 *A. thaliana* immunity . . . . . 14

1.3.3 *C. elegans* immunity . . . . . 18

1.4 Virus-host interactions . . . . . 23

1.4.1 Host factors required for TuMV cycle . . . . . 23

1.4.2 Host factors required for OrV cycle . . . . . 28

1.4.3 Virus-host coevolution . . . . . 28

**2 Aims, Structure of the Thesis and Other Contributions 33**
2.1 Aims . . . 33
2.2 Structure of this Thesis . . . 34
2.3 Other Contributions . . . 36

**3 Phenotypic and genomic changes during turnip mosaic virus adaptation to *Arabidopsis thaliana* mutants lacking epigenetic regulatory factors 39**

**4 Transcriptomic insights into the epigenetic modulation of turnip mosaic virus evolution in *Arabidopsis thaliana* 57**

**5 Accumulation dynamics of defective genomes during experimental evolution of two betacoronaviruses 77**

**6 Story of an infection: viral dynamics and host responses in the *Caenorhabditis elegans*-Orsay virus pathosystem 107**

**7 General Discussion 125**

**8 Conclusions 131**

**References 135**

**Resum Ampliat 151**

**Appendix A: Extra Results to Chapter 6 167**

**Appendix B: Supplementary material from published papers 225**
.1 Supplementary to Chapter 3 . . . 225
.2 Supplementary to Chapter 4 . . . 234
.3 Supplementary to Chapter 5 . . . 248
.4 Supplementary to Chapter 6 . . . 253

# List of figures


1.1 ICTV MSL40.v1 *Orthornavirae* Kingdom classification . . . . . 2
1.2 Virus genomes . . . . . 3
1.3 Types of substitutions . . . . . 6
1.4 *A. thaliana* immune responses to TuMV infection . . . . . 21
1.5 *C. elegans* known immune responses to OrV infection . . . . . 22
1.6 TuMV-*A. thaliana* PPIN . . . . . 25
1.7 "Now, here, you see, it takes all the running you can do, to keep in the same place." . . . . . 29
1.8 Continuum mutualism-parasitism hypothesis . . . . . 30

2.1 Published results classification and chapter correspondence . . . . . 34

# List of tables

1.1 Classification of the viruses studied in this thesis, in order of appearance in the resulsts chapters. Common names are TuMV, HCoV-OC43, MHV and OrV. . . . . 4

1.2 Host factors with physical interactions with TuMV proteins. Effect is "P" when a proviral or "A" an antiviral effect has been described in the reference in *Ref.* column. Methods used to identify the physical interaction (PI) include yeast two-hybrid (Y2H), bimolecular fluorescence complementation (BiFC), and enzyme-linked immunosorbent assay (ELISA). PI tested in [110] are identified with a "Yes" and with an asterisk ("*") the PI not contemplated in the PPIN from [110]. . . . . 26

# CHAPTER 1

# Introduction

> Nothing is yours. It is to use. It is to share. If you will not share it, you cannot use it.
>
> Ursula K. Le Guin, *The Dispossessed, 1974*

## 1.1 Introduction to viruses, RNA viruses and Defective Viral Genomes

Viruses are the most abundant entities on Earth. They are obligate intracellular parasites founded infecting all kingdoms of life in all ecosystems studied. Outside of the hosts, virus exists as virions or viral particles. Virions consist of a genome encapsidated and protected by a protein capside. Some of them are additionally enveloped by a host lipidic layer.

In 1971 David Baltimore classified viruses in seven groups depending on the type of nucleoacid that encodes its genome (DNA/RNA, single (ss) or doble (ds) stranded) and the way the viral mRNA is produced. The DNA viruses groups including: I (dsDNA), II (ssDNA); the RNA viruses groups are: III (dsRNA), IV ((+)ssRNA), V ((-)ssRNA); and the retro-transcribing (RT) viruses belonging to groups: VI (ssRNA-RT) and VII (dsDNA-RT) [108].

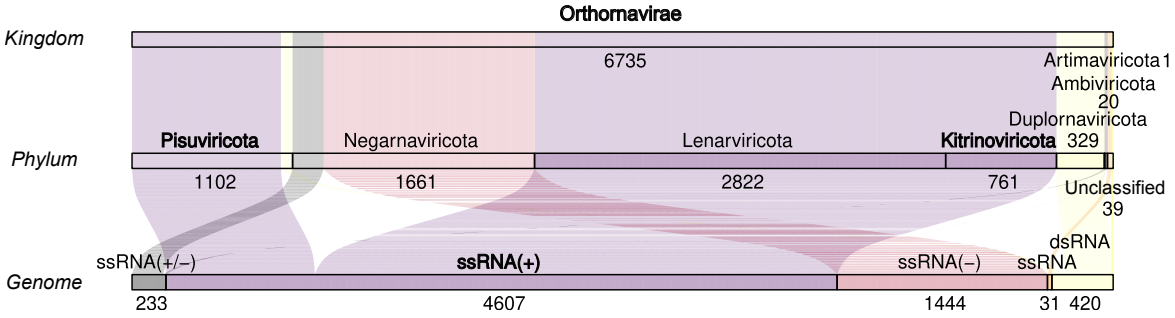



Fig. 1.1 ICTV MSL40.v1 *Orthornavirae* Kingdom classification. *Orthornavirae* is divided in 7 phyla with different RNA genome architectures. Viruses from *Pisuviricota* (TuMV, HCoV-OC43 and MHV) and *Kitrinoviricota* (OrV), all with a (+)ssRNA genome, are the subject of this thesis. Numbers indicate the viruses in the clade.

In contrast with Baltimore's genomic classification, the International Committee on Taxonomy of Viruses (ICTV) maintains a viral classification based on phylogeny. Actually, ICTV consider seven great clades (realms) which include all the known viruses according to their common evolution, grouping them by similarities on their genes and proteins. *Adnaviria*, *Duplodnaviria*, *Monodnaviria*, *Riboviria*, *Ribozyviria*, *Singelaviria*, and *Varidnaviria* are these realms, and RNA viruses are distributed in the *Riboviria* and *Ribozyviria* realms. From broader to more specific taxa, the actual ICTV classification comprises: the 7 realms listed above, 11 kingdoms, 22 phyla, 4 subphyla, 49 classes, 93 orders, 12 suborders, 368 families, 213 subfamilies, 3769 genera, 86 subgenera and 16215 species of virus (Master Species List 2024, MSL40.v1).

### 1.1.1 RNA viruses

Viruses with RNA genome are classified into the *Riboviria* and *Ribozyviria* realms. Most of the known animal and plant pathogenic viruses are classified into the *Riboviria* realm.

ICTV divides *Riboviria* in two Kingdoms: *Orthornavirae* and *Pararnavirae*. Members of the *Pararnavirae* kingdom encode an RNA-dependent DNA polymerase or reverse transcriptase (RdDP) while viruses in *Orthornavirae* encode an RNA-dependent RNA polymerase (RdRP). A few viruses of the *Orthornavirae* kingdom, specifically from the *Pisuviricota* (turnip mosaic virus (TuMV), human coronavirus OC43 (HCoV-OC43) and murine hepatitis virus (MHV)) and from the *Kitrinoviricota* (Orsay virus (OrV)), will be the subjects of this thesis.

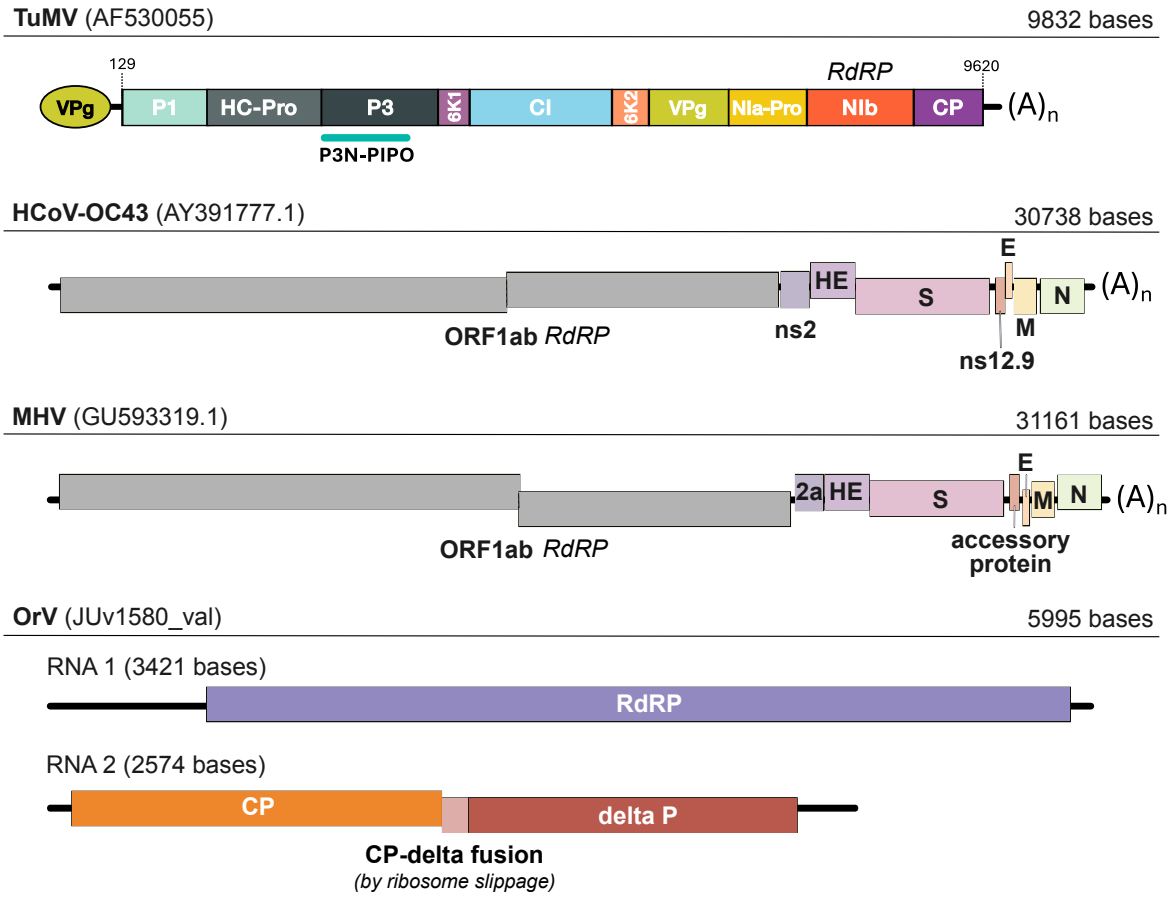



Fig. 1.2 Genome organization of the four viruses studied in this thesis. Regular name with the reference used between parentheses and the number of total bases. Each protein or cistron (in the case of polyproteins) is represented with colors. All the viruses have an RdRP. Genome sizes range from the bi-segmented OrV with the smallest segment having 2.5 kb to MHV with a more than 31 kb genome.

| Specie acronym | Genus | Family | Order | Class | Phylum |
|---|---|---|---|---|---|
| *Potyvirus rapae* | *Potyvirus* | *Potyviridae* | *Patatavirales* | *Stelpaviricetes* | *Pisuviricota* |
| *Betacoronavirus 1* | *Betacoronavirus* | *Coronaviridae* | *Nidovirales* | *Pisoniviricetes* | *Pisuviricota* |
| *Betacoronavirus muris* | *Betacoronavirus* | *Coronaviridae* | *Nidovirales* | *Pisoniviricetes* | *Pisuviricota* |
| *Orsay nodavirus (inc.sed.)* | Unclassified | *Nodaviridae* | *Nodamuvirales* | *Magsaviricetes* | *Kitrinoviricota* |

Table 1.1 Classification of the viruses studied in this thesis, in order of appearance in the resulsts chapters. Common names are TuMV, HCoV-OC43, MHV and OrV.

As introduced above, all of them have a (+)ssRNA genome (Fig. 1.1) and encode for an RdRP. Table 1.1 shows their taxonomic details and their genomes are represented in Fig. 1.2, in order of appearance.

### Positive single-strand RNA viruses viral cycle

The viral cycle of (+)ssRNA viruses involves common key steps to produce new virions. The general process can be described as follows:

- **Attachment and entry:** The virus attaches to specific receptors on the host cell surface. Following binding, the virus enters the host cell either by endocytosis or direct fusion with the plasma membrane [13].
- **Disassembly or uncoating:** Once inside the host cell cytoplasm, the viral capsid is uncoated, releasing the viral RNA genome.
- **Primary translation:** (+)RNA is translated by host ribosomes to form protein(s) directly or indirectly involved in viral RNA replication. At least the RdRp. In other cases, as in TuMV, a complete polyprotein is translated and then processed into their mature proteins.
- **Transcription:** The viral RdRp uses the (+)RNA genome as a template and synthesizes a complementary (-)RNA strand antigenome. Some viruses as HCoV-OC43 and MHV, produce subgenomic RNAs (sgRNAs), smaller copies of the genome coding for proteins different from the RdRP [144, 187].
- **Replication:** The (-)RNA strand then serves as a template for generating additional (+)RNA genomes that will constitute the progeny.

  Transcription and replication occur, depending on the virus, in the cytoplasm or in membrane invaginations called replication organelles [34, 7].

- **Secondary translation:** Some viruses as alphaviruses and coronaviruses translate their structural proteins or the proteins involved in movement (as in TuMV) in a second round of translation. Others can produce alternative products by ribosome frameshifts [30].
- **Assembly:** The new synthesized viral proteins and the (+)RNA genomes are assembled into new virions. In some viruses these processes occur at the membranes of the endoplasmic reticulum and/or Golgi complex (*e.g.*, coronaviruses). In other viruses as alphaviruses, assembly occurs in the cytoplasm.
- **Egression:** Newly formed virions are transported to the cell surface, sometimes in vesicles. These virions are then released either by cell lysis or via exocytosis, allowing the virions to infect new cells [124].

### 1.1.2 Defective viral genomes

Viruses produce varying copies of themselves during their replication. Point mutations and hypermutations, short and large deletions, insertions or duplications, and genome rearrangements with change of the copy sense are among the plethora of varying products that constitute the viral population.

A mutant viral copy that requires the presence of a helper virus to complete its viral cycle is referred to as a defective viral genome (DVG). When the DVG is able to encapsidate, it forms a defective particle (DP). If the presence of the DP interferes with the dynamics of the wild-type virus (WT), it is considered a defective interfering particle (DIP).

DIPs were first described in influenza A virus by Preben Von Magnus and Svend Gard in 1947 [177]. In 1970, Alice Huang and David Baltimore proposed that DIPs play a major part in the progression of the infection [73]. More recently, Marco Vignuzzi and Carolina López classified the roles of the DVGs in pathogenesis in three categories: (*i*) interference with standard viral replication, (*ii*) immunostimulation, and (*iii*) establishment of viral persistence [173]. The first two roles can be potentially exploited for therapeutic purposes, and an increasing number of applied studies have tried the use of DIPs as antiviral drugs [136, 193, 190] and as vaccines or coadyuvants of vaccines [33, 104].

In parallel, different methods for identifying and characterizing DVGs *in silico* [137, 10, 15, 158, 123, 120] and for modeling their dynamics within the viral population have been proposed [148, 107, 119, 105].

DVGs have been reported in all types of RNA viruses although not all DVG types have to be present in the same virus (see review [173] for a comprehensive list). They are diverse in nature, but we consider four basic types of DVG with genomic structural changes (Fig. 1.3): (1) deletions (or delVGs) are produced when the RdRP detaches from the template (breakpoint position, BP) and reattaches in a downstream locus (reinitiation site, RI) without copying the part in between. (2) Insertions or duplications have from one to several additional nucleotides (the RI is upstream of the BP so the region in between is duplicated), and (3) 5'- and (4) 3'-copybacks or snapbacks have duplicated the 5' or the 3' extremes, respectively, as a complementary strand. They would be produced when the RdRP detaches in the BP and then reattaches in the nascent strand that is then copied as complementary.

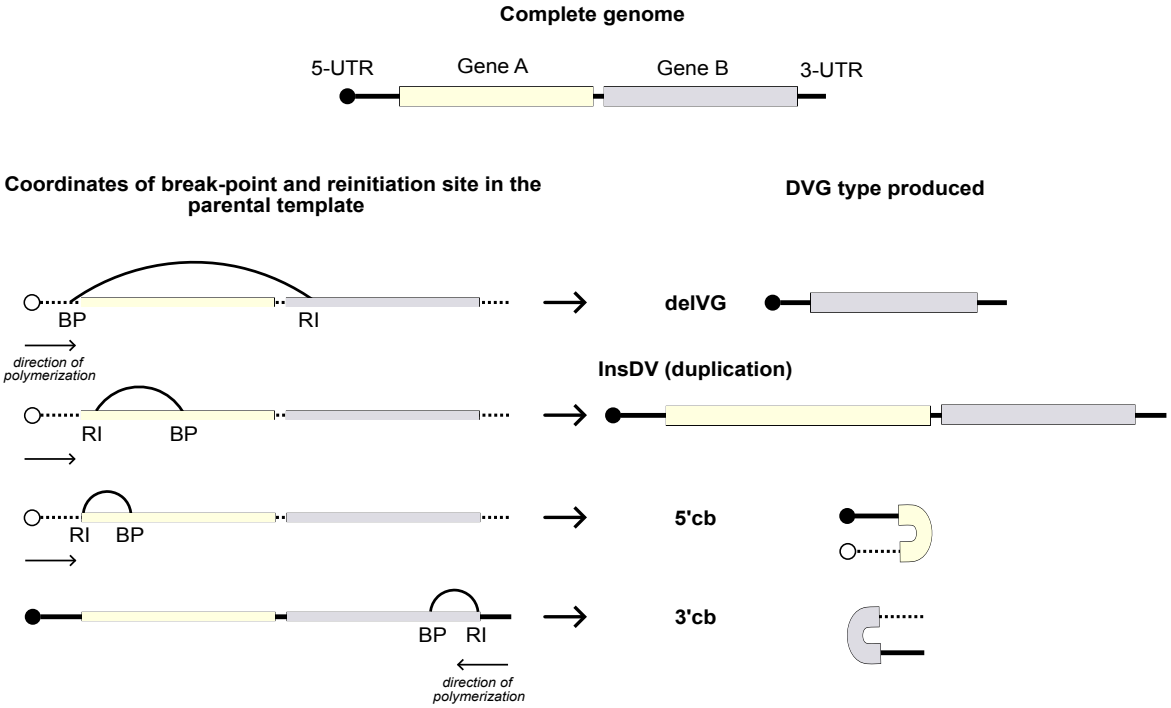



Fig. 1.3 Types of DVGs. An example of deletion, insertion, 5'copy-back and 3'copy-back produced during the discontinuous copy of the parental genome. Points where the RdRP releases (BP) and re-attaches (RI) are marked in the parental genome (left), and the DVG product is depicted on the right. Circles on the left of the genomes are to facilitate the identification of the 5 UTR. Dotted lines represent the antigenome (complementary sequence of genome).

DVGs, as part of the viral quasispecies, are subjects of the forces that shape the viral outcome. Whether as cheaters or as cooperators of the WT virus, their importance in the progression of viral infections makes their study and characterization essential.

## 1.2 RNA viruses as evolving populations

I am rooted, but I flow.

Virginia Wolf, *The Waves, 1931*

### 1.2.1 Viral Evolution

Viral evolution refers to the genetic changes that occur in the genome of a virus over time, regardless of the scale of time or the transient nature of the change [38]. It is shaped by both deterministic and stochastic processes. The former includes natural selection, which acts on viral variants that confer a competitive or transmission advantage, making them more prevalent in the population. Stochastic processes, on the other hand, encompass genetic drift due to random fluctuations in allele frequencies and bottlenecks, which can lead to significant genetic shifts independent of selection pressures [38, 127].

In the next two sections, we will delve deeper into the mechanisms that generate viral diversity and into some population dynamics concepts.

### 1.2.2 Viral Diversity

Viral diversity refers to the genetic variability within a viral population. This diversity includes nucleotide substitutions and DVGs [41, 40]. Some of these events, particularly those leading to DVGs, are crucial in shaping the viral population's adaptive potential.

Quantifying viral diversity is essential for understanding viral evolution and adaptation processes. Advancements in next-generation sequencing (NGS) technologies have enabled high-resolution analysis of viral populations, allowing for the detection of minor variants that can contribute to viral fitness, pathogenicity, and resistance to therapeutic interventions [151, 174].

***Mechanisms of Genetic Variation***

Several mechanisms that introduce genetic variation in other forms of life also operate in RNA viruses: mutation, recombination, and genome segment reassortment.

***Mutation*** A mutation is a change in the viral genome nucleotide sequence that arises due to errors during the replication of the parental genome. RNA viruses tend to have the highest mutation rates among all living beings, ranging from $10^{-6}$ to $10^{-4}$ mutations per nucleotide per cell infection. This elevated mutation rate is due to the

use of RdRP for genome replication, which, with some exceptions (*e.g.*, coronaviruses), lacks proofreading abilities [128].

*Types of Mutation:*

- *Transitions* refer to the change between purines ($A \leftrightarrow G$) or between pyrimidines ($U \leftrightarrow C$).
- *Transversions* are the substitutions of a purine by a pyrimidine and *vice versa* [40].
- *Insertions and deletions (indels)* are the addition or subtraction of nucleotides in the genome.

*Mutation rate* refers to the frequency with which a mutation event occurs during genome replication. This biochemical event depends on factors such as the replication enzyme, template, and environmental conditions (*e.g.*, temperature), but it is independent of the relative fitness of the produced genomes. In contrast, *mutation frequency* refers to the proportion of mutated sequences relative to the consensus sequence in a viral population, and it is influenced by the relative fitness of the genomes carrying the mutations [40].

***Recombination*** Recombination is the process by which new viral genomes are created by the covalent linkage of genetic material from multiple parental genomes. This process can be classified into two types: (1) *replicative recombination*, which occurs during genome replication when the polymerase switches from one parental genome to another, and (2) *non-replicative recombination*, where there is no polymerase involved, but the genomes are joined after being broken [91, 72].

***Genome Segment Reassortment*** Genome reassortment occurs in RNA viruses with segmented genomes. This process involves the exchange of complete genome segments between two or more different parental viruses, which results in the generation of novel viral strains [127, 39]. Reassortment plays a key role in the emergence of new viral variants that may exhibit altered pathogenicity or immune escape mechanisms, presenting challenges for public health.

### 1.2.3 Population dynamics in viral evolution

Within an infected cell, viruses exist as heterogeneous populations known as **quasispecies**, a cloud of closely related viral genomes connected by mutations [42]. The term quasispecies refers to the idea that viral populations are not composed of identical genomes, but rather of a diverse array of closely related genetic variants.

Quasispecies are highly dynamic and fluctuating populations. The quasispecies structure allows viruses to adapt rapidly to new selective pressures, such as changes in host immune responses or the presence of antiviral treatments. The evolution of quasispecies populations is shaped by both deterministic forces (natural selection) and stochastic forces (genetic drift) [42, 122]. The prevalence of one or the other type of force is highly influenced by the **effective population size** ($N_e$), which refers to the number of individuals contributing offspring to the next generation. In a theoretical viral population with an infinite size $N_e$, mainly selection would drive the evolution of the population, while the effect of stochastic processes are more important in populations with small $N_e$ [117].

### Deterministic factors

These factors arise from systematic forces within the viral population and the host environment, mainly driven by natural selection pressures, as first outlined by Charles Darwin in On the Origin of Species in 1859 [32]. Darwin proposed that traits that enhance survival and reproduction become more common in a population over generations. In the context of viral evolution, these forces act on genetic changes that result in phenotypic variations, ultimately influencing viral fitness and adaptation.

- Positive Selection: *Increase frequency*
  Variants that confer a fitness advantage, such as more rapid replication, immune evasion, or adaptation to specific host niches, are more likely to increase in frequency.

- Negative Selection: *Decrease frequency*
  Variants that are deleterious to viral replication, encapsidation, survival... (*e.g.*, those that are easily recognized by the immune system, reduce replication, or cannot be encapsidated) will be selected against and decrease in frequency.

- Cooperative Interactions: *Increase frequency*
  When two or more variants cooperate, the overall fitness of the viral population can increase. This positive effect can be more generalized, and viruses can benefit from coinfection [143].

- Antagonistic Pleiotropy: *Decrease frequency*
  A mutation can affect more than one phenotypic trait. Variants that have trade-offs (where a mutation that confers a benefit in one area results in a cost in another) may see their frequency decrease over time. Adaptation to a specific

niche, which improves fitness in that environment but reduces fitness in others, is a classic example of antagonistic pleiotropy, where a mutation's benefits in one context come with costs in another [130].

- Genetic Hitchhiking: *Increase / Decrease frequency*
  A neutral or mildly deleterious mutation can increase in frequency if it is genetically linked to a beneficial mutation or reduce in frequency if it is genetically linked to a negatively selected one.

- Clonal Interference: *Decrease frequency*
  In asexual populations, beneficial mutations compete for fixation, often hindering each other. This leads to the sequential fixation of advantageous mutations, impacting the rate of adaptive evolution. Even beneficial variants may see reduced frequency because they cannot fully outcompete other variants with different advantageous mutations. Clonal interference tends to decrease the rate of frequency increase of all competing variants, as they interfere with one another's growth, resulting in slower rates of fixation [61, 114].

### Stochastic Factors

These factors are driven by random events rather than predictable selection pressures, making them especially important in small viral populations where chance plays a stronger role in determining variant frequencies.

Genetic Drift: *Increase / Decrease frequency*

In small viral populations, random fluctuations in allele frequencies (genetic drift) can cause neutral or slightly deleterious variants to increase in frequency by chance. On the flip side, genetic drift can also lead to the random loss of variants, especially those that are rare or have a small fitness advantage. Thus, genetic drift can both increase and decrease the frequency of viral variants, depending on the random fluctuations within the population [182]. Bottlenecks, founder effects or Muller's ratcher are all distinct views of the genetic drift.

- Bottlenecks and founder effect: *Decrease / Increase frequency*
  Bottlenecks occur when a viral population experiences a sharp reduction in size, such as during transmission. In such cases, only a small subset of variants survives the bottleneck. Variants not represented in the small surviving pool may be lost entirely, leading to a decrease in their frequency and reduced diversity within the viral population. However, others can potentially increase their frequency purely by chance in a founder effect.

- Muller's ratchet: *Increase frequency*
  Muller's ratchet refers to the irreversible accumulation of deleterious mutations in a small or asexual population, where there is no mechanism for genetic recombination to purge these mutations [118]. The term was coined by [52] and it has been proposed to be an important reason for the evolution of sexual reproduction. Many viruses, especially RNA viruses, have high mutation rates, leading to the accumulation of harmful mutations. Over time, this can reduce the overall fitness of the viral population, leading to a "ratcheting" process, where viral genomes become progressively more flawed, possibly resulting in the extinction of less-fit viral strains. Repeated bottlenecks often lead to loss of fitness due to Muller's ratchet effect [26, 45].

### Virus-virus relationships

Viral evolution is highly impacted by virus-virus interactions. These relations exist between members of the same quasispecies, between the same virus but not produced in the same cell, and also between different viruses. Competition or interference was, as far as we know, the first behavior described within viruses [177, 188, 37]. Superinfection exclusion (SIE) is one extreme phenomenon where the infection of a host by a second virus is prevented by the presence of a first virus, usually due to immune system activation or resource limitation. This type of competition leads to the dominance of the first virus and inhibits the replication of subsequent infections by different viral strains [75]. More recently, cooperation, competition, and cheating behaviors have been described as shaping the outcome of viral infections [36, 35, 97, 98, 143]. In [36], the authors proposed the term **sociovirology**.

The *Prisoner's Dilemma* (PD), a game theory model, has been used to explain the evolution of viral strategies like cooperation and competition. In this model, cooperators produce public goods, benefiting both themselves and other viruses, while defectors exploit these goods without contributing. In [167] described that although cooperation could lead to better overall fitness, defectors can outcompete cooperators. Authors studied RNA phage $\phi 6$ at high and low MOIs. At high MOI, multiple viral genomes infected the same host, leading to initial fitness increases followed by a decline as defectors take over, fitting the PD model. In contrast, at low MOI, phages reproduce mainly as clones, preventing defectors and allowing cooperation to persist [167]. Later, [168] expanded on this, showing that at high MOI, defectors gain a competitive advantage, but this lowers overall fitness. When viruses were allowed to reproduce clonally, cooperation could evolve, leading to a mix of cooperators and

selfish variants. This showed that the fitness outcomes in viral populations depend on the ecological conditions and can change over time [168]. In [27], authors took this further by modeling the evolution of cooperation in RNA viruses, focusing on gene products (public goods) shared within the host cell. They conclude that if the trade-off between replication and gene product production was linear, defectors would dominate. However, when the trade-off is nonlinear, viruses evolve into ultra-defectors and ultra-cooperators, creating a *snowdrift game scenario* that maintains high levels of cooperation [27].

## 1.3 Host Responses to viral infections

My silences had not protected me.
Your silence will not protect you.

Audre Lorde, *Sister outsiders, 1984*

### 1.3.1 Host immunity

All living organisms have a set of tools collectively called immune system. These tools have evolved to defend the living systems from external (bacteria, fungi, viruses...) and internal threats.

The traditional view of immunity posits that immune defenses are mounted in two consecutive steps: an innate, nonspecific response that acts immediately upon infection and the adaptive response that is slower but highly specific [142].

Only vertebrates have adaptive immunity; the rest of the organisms rely on their innate immunity. But this does not mean that plants or invertebrates lack specific responses to pathogens or are not able to mount a long-term defense. In fact, there is increasing evidence of both aspects [142, 176, 83, 165, 92]. Immune memory, as the ability of the immune system to respond more rapidly and effectively to pathogens that have been encountered previously, is not exclusive of the adaptive immunity.

*Innate immune memory*

On the one hand, trained innate immunity (TRIM) would be a primitive form of adaptation in host defense, resulting from the remodeling of the chromatin structure [121, 132, 69]. Priming the defenses of plants to enhance their responses against biotic and abiotic stresses has been a key concept in crop protection [111] and the study of epigenetics has been nurtured in part for its agricultural applications.

On the other hand, RNA-based memory illustrates a convergent evolution example found in all taxa. In bacteria and archaea, the CRISPR-Cas system provides the memory of past viral infections. In invertebrates, the best-known immune defensive mechanism against viruses is RNA interference (RNAi) by small interfering RNAs and PIWI-interacting RNAs (piRNAs). This mechanism can provide long-term responses, and it has been found that endogenous viral sequences integrated into the insect genome, as an album of past viral infections, would function as templates for the production of piRNAs [181, 53].

*Inducible responses*

Innate immunity includes constitutive and inducible responses. The first serves as a continuous physiological barrier, while inducible responses are activated after the detection of pathogens or the damage caused by the pathogens. Inducible responses can be highly efficient but are costly and commonly associated with a growth and developmental penalty, so organisms have evolved fine-tuned mechanisms for the activation and deactivation of these responses [178, 59, 142].

The physiological response could be divided into three processes [142]:
(1) The recognition of the pathogen,
(2) the processing of the information via activation of the corresponding signaling pathway and
(3) the expression of immune effector molecules to eliminate the pathogen or alleviate their damaging effect.

Many of these mechanisms are highly conserved across multicellular eukaryotic organisms [17], with some strategies also shared with bacteria and archaea [96, 150]. Research on innate immunity in organisms that depend solely on this defense strategy to counter pathogen invasion has significantly deepened our understanding of the host-pathogen interaction, offering considerable translational potential.
In this thesis we have studied the responses against virus infection of two model organisms: *A. thaliana* and *C. elegans*.

### 1.3.2 *A. thaliana* immunity

*Arabidopsis thaliana* (L.) Heynh is probably the most studied plant model. *A. thaliana* belongs to the *Brassica* genus, known for its agricultural and horticultural crops importance. Broccoli, cauliflower, turnip or mustard are some of the edible species of this genus.

Plants are sessile organisms. They cannot escape from biotic or abiotic stresses. Instead, they have evolved efficient response mechanisms to face environmental threats. The multilayered response of plants is based on preformed physical barriers, RNA silencing, pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI), and effector-triggered immunity (ETI), which activate downstream defense responses. The epigenetic mechanisms participate directly in the regulation of immune responses and in plant immune memory [131, 132]. The effect of the immune response can be local (reduced to the damaged cell) or systemic (preparing other tissues for an anticipated response). Identification of genetic associations in response to TuMV infection has been addresed in a recent genome-wide association study (GWAS) [18].

A summary of some of the mechanism and elements involved in the pathogen response are listed below and in Fig. 1.4 is represented the *A. thaliana* antiviral immunity.

**Perception of pathogens.** Plants perceive pathogens through outside and inside cell receptors. The Pattern Recognition Receptors (PRRs), located on their cell membranes, detect conserved pathogen molecules known as Pathogen-Associated Molecular Patterns (PAMPs) or Damage-Associated Molecular Patterns (DAMPs). Examples of PAMPs include bacterial *flagellin* and fungal *chitin*. This recognition activates the first line of defense, **PAMP-Triggered Immunity (PTI)**, which initiates a cascade of signaling events.

Inside the cell, the pathogen effectors can be recognized by NLRs, members of the primary family of plant resistance (R) proteins. This recognigion leds to the activation of the **Effector-Triggered Inmunity (ETI)**. *A. thaliana* has more than 150 NLR-coding genes [172]. Leucine-Rich Repeat (LRR) are important structural motifs for protein-protein interactions, with a conserved region of 20-29 that includes a 11-residue consensus sequence (LxxLxLxxN/CxL, with x = any residue, and $L \in \{L, V, I, F\}$). This motifs are present in both PRRs and NLRs [197].

Inside the cell, viral-derived dsRNA is recognized and targeted by the **RNAi machinery**. Dicer-like RNases (DCLs) cleave viral dsRNA molecules into smaller RNAs which are charged onto Argonaute proteins (AGO), and incorporated in the RNA-inducer silencing complex (RISC) to guide the post-transcriptional silencing or the translational supression of the targeted viral RNA [106]. Additionally, RNA methylation (specifically the N6-methyladenosine (m6A)), has been reported in the genome of bromoviruses infecting *A. thaliana*[131]. Amplification of silencing rsponse involves plant RNA-dependent RNA polymerases (RDRs) which produce additional dsRNA from the virus [14]. *A. thaliana* has four DCLs, ten AGO and six RDRs [106].

**Signaling Pathways** Once the pathogen is recognized, signaling cascades are activated, which include $Ca^{++}$ **influx** and activation of calcium-dependent protein kinases (CDPKs), the production of **ROS**, and the activation of **MAPK (mitogen-Activated Protein Kinase)** pathways.

PRR and NLR signalling co-regulate immune responses and converge at multiple points but ETI has a stronger amplitude and duration of the response. Some of the overlapping immune responses of PTI and ETI include ROS (reactive oxygen species) and NO burst, $Ca^{++}$ influx, MAPK activation, phytohormone production and transcription reprogramming [194]. If ETI exceeds certain threshold activation, a programmed cell death named as hypersensitive response (HR) takes place. The

crosstalk between PTI and ETI and their quantitative output was represented by the 'zigzag' model [81].

Plants are able to signal not only inside the affected cell but at longer distances. The primary infection can induce a broader-spectrum resistance in distal tissues of the plant named systemic adquired resistance (SAR) that involves the rapid generation and transportation of mobil signals preparing the rest of the plant for the stress [82]. Some signalling molecules can spread throughout the plant vascular system while others can be volatilized to the air, even reaching nearby plants. Additionally, mycorrhizal fungi could interconnect individual plants to form common mycorrhizal networks (CMNs). These networks can also transport signals that "advise" neighbouring plants [154].

*Inter-tissue communication:* The infected cell, signals neighboring cells to reinforce defenses. ROS, NO, some phytohormones and volatile organic compounds (VOC) act as messengers to propagate the defense response within the plant [86, 145]. sRNAs can also move between neighboring cells throughout plasmodesmata, and be loaded in the phloem reaching distal parts of the plant [194, 102]. This would potentially trigger immunity in the recipient tissues but, as far as we know, it has not being studied in the fight agains viral infections (Fig. 1.4).

*Inter-plant communication:* VOC produced by infected plant are released to the air and can be captured by nearby healthy plants, primming them for a more quickly defense response [8]. Viruses can also exploit the VOCs emission. Some VOCs attrack aphids to the infected plants, facilitating the viral spread by these vectors [20] (Fig. 1.4).

**Phytohormones in immunity**

Phytohormones play a central role in regulating plant immune responses by orchestrating the perception of pathogens and activating downstream signaling pathways. The timming at which the hormones are released is key with cases of opposite outcome depending on the stage of the infection. The major phytohormones involved in antiviral immunity are three chloroplast-derived hormones:

- **Salicylic acid (SA):** Phenolic compound essential for establishment of local and systemic resistance. SA biosynthesis and signalling are activated by ETI. SA is responsible for the activation of **SAR** in distal tissues [2] (Fig. 1.4).
- **Jasmonic acid (JA):** Oxygenated fatty acid (oxylipin) involved in resistance to necrotrophic pathogens and insect infestation. Together with ethylene, JA regulates induced systemic resistance (ISR), which is invoked by non-pathogenic

microbes, such as rhizobacteria[2]. TuMV P1 protein is a suppressor of JA biosynthesis in *Nicotiana benthamiana* [77].

- **Abscisic acid (ABA):** Isoterpenoid involved in responses to abiotic and biotic stresses. At the early stages of infection, it mediates the stomatal closure against invaders and induces callose deposition if the pathogen evades the first line of defense. If activated at later stages, ABA can suppress ROS induction and SA or JA signaling transduction, thereby negating defenses controlled by these two pathways [2].

  ABA has a defensive role against viruses through inhibition of the basic $\beta - 1,3-$glucanase, which is responsible for the degradation of $beta - 1,3-$glucan (callose) [2], and by strengthening RNAi by promoting expression of AGO1, AGO2, and AGO3. Some viruses target the ABA pathway, increasing the plant susceptibility [89]. Exogenous application of ABA induces plant resistance to some viruses.

**Expression of immune effectors**

Once the plant has perceived the pathogen and initiated signaling pathways, it activates the expression of defense-related genes. These include pathogenesis-related proteins (PR proteins) with enzymatic activity, such as PR1 (antifungal and antiviral activity), PR2 ($\beta$-1,3-glucanases; antibacterial, antigungal, and antiviral activity), and PR10 (ribonuclease; programmed cell death during HR; antibacterial, antifungal, antinematode, and antiviral activity; also salt and cold stress tolerance) [43].

Cell wall fortification is another essential response, where the plant deposits lignin and callose to reinforce cell walls and prevent pathogen penetration. Callose is also deposited in plasmodesmata to minimize viral movement. It has been shown that TuMV NIa-Pro suppresses callose deposition [22].

**Epigenetic regulation of immunity and immune memory**

Epigenetic mechanisms, including DNA methylation, histone modifications, and small RNAs, play a critical role in regulating the immune response in plants. These mechanisms also provide immune memory, allowing the plant to mount faster and more efficient responses to repeated pathogen attacks. Through RNA silencing, small RNAs can regulate the expression of immune-related genes, and histone modifications can alter chromatin structure to maintain a heightened state of defense [131, 132, 70].

### 1.3.3 *C. elegans* immunity

*Caenorhabditis elegans* (Maupas, 1900) is a nematode from the *Rhabditidae* family used as an experimental model in the study of innate immunity [68]. There is a stage-dependent factor of immunity, with larval-stage animals having stronger responses and better survival than adults [83]. The defense of these animals relies on three main axis:

- 1) **Physical barriers**. *C. elegans* has estructural elements that make difficult the entry and colonization of pathogens from the more external to the interior of the cells: The worms have a strong cuticle made of collagen and chitin that constitutes their exoskeleton. They also have a pharyngeal grinder that destroys the pathogens taken during feeding. Both cuticle and pharyngeal grinder are replaced at each molting [156]. The intestinal terminal web is composed of actin and intermediate filaments that likely act as a physical barrier for OrV entry [198].

- 2) **Avoidance behavior**. *C. elegans*, can learn to avoid pathogens. Their ability to avoid pathogenic bacteria such as *Pseudomonas aeruginosa* and *Serratia marcescens* has been documented [196, 129]. Some studies suggest that the learned avoidance behavior of *P. aeruginosa* can be transmitted vertically for up to four generations [115]. However, difficulties in replicating these results have led to a reconsideration of the findings [58]. Transgenerational inheritance of this behavior may involve piRNAs, the TGF-$\beta$ pathway, and ASI, the two sensory neurons of the ciliated endings in the amphid sensillum [115]. Apart from avoiding bacteria, *C. elegans* males showed a preferential mating with non-infected hermaphrodites rather than OrV-infected individuals [170].

  Stimuli that promote pathogen avoidance include those caused by the alteration of aerotaxis, intestinal bloating, metabolites, transgenerational epigenetic inheritance, and crosstalk with innate immunity [99]. This information is modulated by the nervous system [179]. Additionally, these animals secrete a complex cocktail of small chemicals, called ascaroside pheromones, that affect many aspects of the *C. elegans* population dynamics, such as the dauer formation [88], the mating behavior or aggregation [29, 192]. Some of these pheromones have been associated with avoidance behavior [138, 29], but as far as we know, the relation of ascaroside secretion and the avoidance of pathogens has not been studied yet.

- 3) **Physiological response**. *C. elegans* possesses an inducible system involving multiple signaling cascades. p38MAPK, FSHR-1, ZIP-2, STAT, insulin signaling,

TFG-$\beta$, Wnt/Hox, ERK MAP, UPR and autophagy pathways are some of the major pathways involved in the immune system. But it lacks some components of the innate immunity present in other invertebrates, such as the NF-$\kappa$B transcription factor or specialized immune cells such as macrophage-like hemocytes found in Drosophila. Antimicrobial peptides (AMPs), lysozymes, lectines, caenopores, or ROS are some of the effector molecules involved in the immune response [153, 160].

After infection with OrV, there is a multilayer response that involves the targeting and degradation of both viral components, the viral RNA (vRNA) and the viral proteins. These functions are carried out by the RNA interference pathway, the urydilation pathways, the ubiquitin-mediated response, and the activation of transcriptional responses [95, 90, 51, 126, 155, 109, 94]. Fig. 1.5, modified from a recent review of the *C. elegans* intracellular pathogen response, summarizes the host and viral known components in the antiviral response. Within the activated genes, there are a set of 80 genes found activated for OrV and some species of microsporidia, another intracellular pathogen that affects the nematode. This was called the Intracellular Pathogen Response (IPR) and includes several *pals* genes and components of the ubiquitination machinery [134, 66, 126].

These three axes play in a fine orchestration. The physical barriers act constitutively, but there are evidences of changes in the external cuticle and in the remodeling of the intestinal terminal web based on the detection of Orsay virus [198]. And the research of host defense responses in relation to the nervous system is gaining attention [183, 179]. *C. elegans* has been a model for the study of neuroimmunity, and components of the nervous system are key in the pathogen detection, in the modulation of the responses, and in the inter-tissue signaling [83]. In this new paradigm, pathogens are detected by either the innate immune system, the nervous system, or both, which then triggers the innate immune response and/or behavioral changes [103].

In *C. elegans* the transgenerational inheritance of environmental responses is transmitted in the form of RdRP-amplified sRNAs. In [133] can be found a list of the factors involved; among them, the argonaute PPW-1 played an important role and would be studied in our work [24]. The innate immunity can also be primed, and there are several examples of immune memory within-generation and in the protection of the offspring [157, 184, 53].

Finally, biotic and abiotic environmental factors highly affect the viral infection outcome. In the case of the biotic factors, [67] found that some of the naturally

associated bacteria induced resistance to OrV infection even when the animals lacked DRH-1, key in the RNAi.

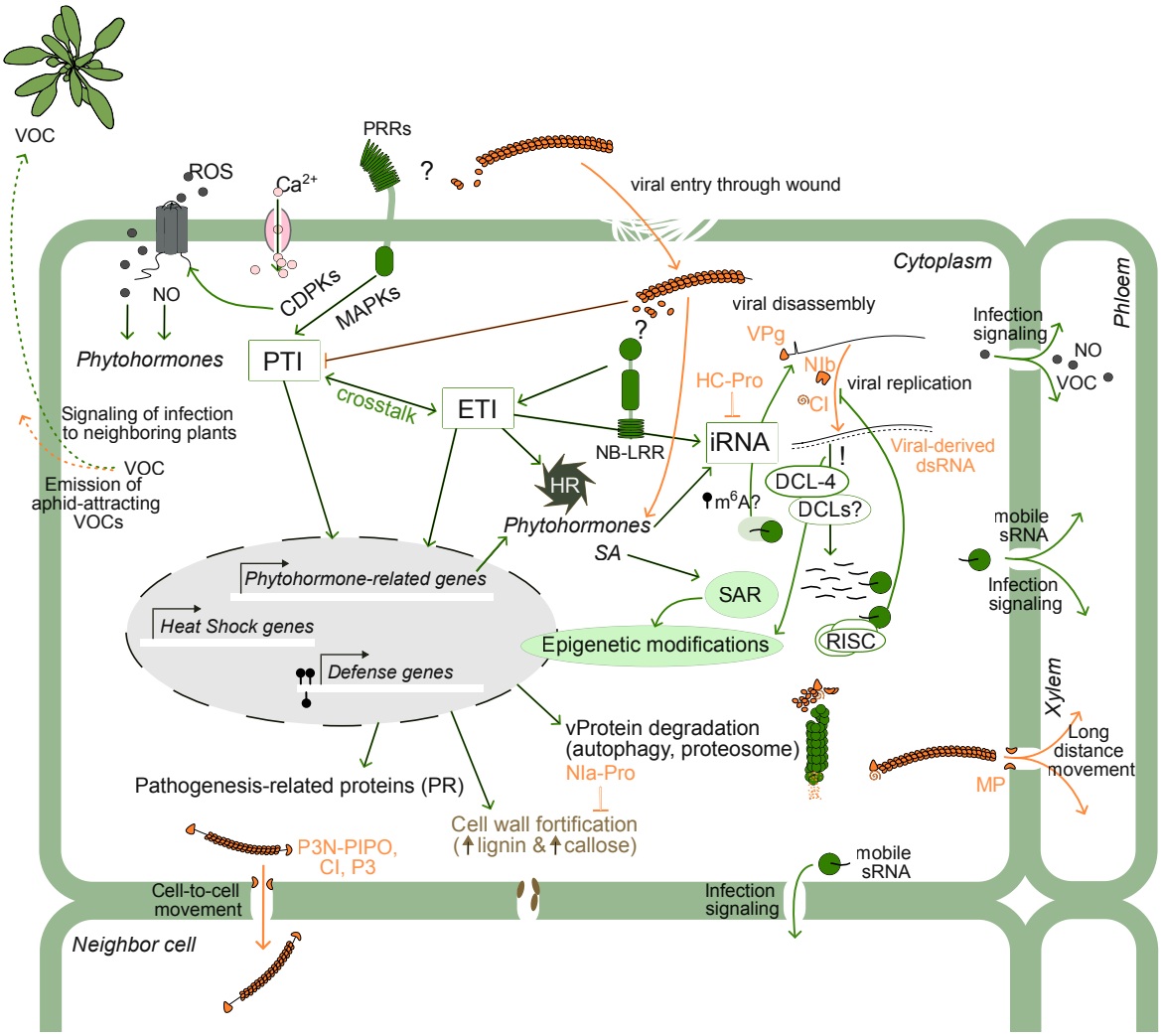



Fig. 1.4 ***A. thaliana* immune responses to TuMV infection**. TuMV can entry a cell by a mechanical wound, throughout the aphid stylet injection or through the plasmodesmata from an already infected cell. Other pathogens can be detected outside the cell with the PPRs activating then the PTI an/or inside the cell by the R proteins as NB-LRR receptors, activating the ETI. TuMV dsRNA intermediates of replication are detected by the antiviral RNAi machinery. The activation of the immunity can led to a local transcriptional reprogramming, and to a more systemic response in space and time (green background), through the communication of the infection state to distal parts of the plant (systemic acquired resistance or SAR) and through the induction of a memory state by epigenetic modifications. If the immune activation overcomes certain threshold, the cell can induce a hypersensitive response (HR) that includes programmed cell death. Plant elements and relations in green, virus elements and relations of activation (arrow-end) or inhibition (plain-end) in orange. MP = viral proteins involved in movement (P3N-PIPO, CI, P3). VPg and CI are also important for stabilization of viral ribonucleoprotein complexes. *Abbrev:* ROS = reactive oxygen species, PRRs = pattern recognition receptors, NB-LRR = Nucleotide-binding site, Leucine-rich repeat containing proteins, CDPKs = calcium-dependent protein kinases, MAPKs = Mitogen activated protein kinases, PTI = pathogen-associated molecular pattern-triggered immunity, ETI = effector-triggered immunity RNAi = interference RNA, HR = hipersensible response, SA = salicylic acid, SAR = systemic adquired response, VOC = volatile organic compounds.

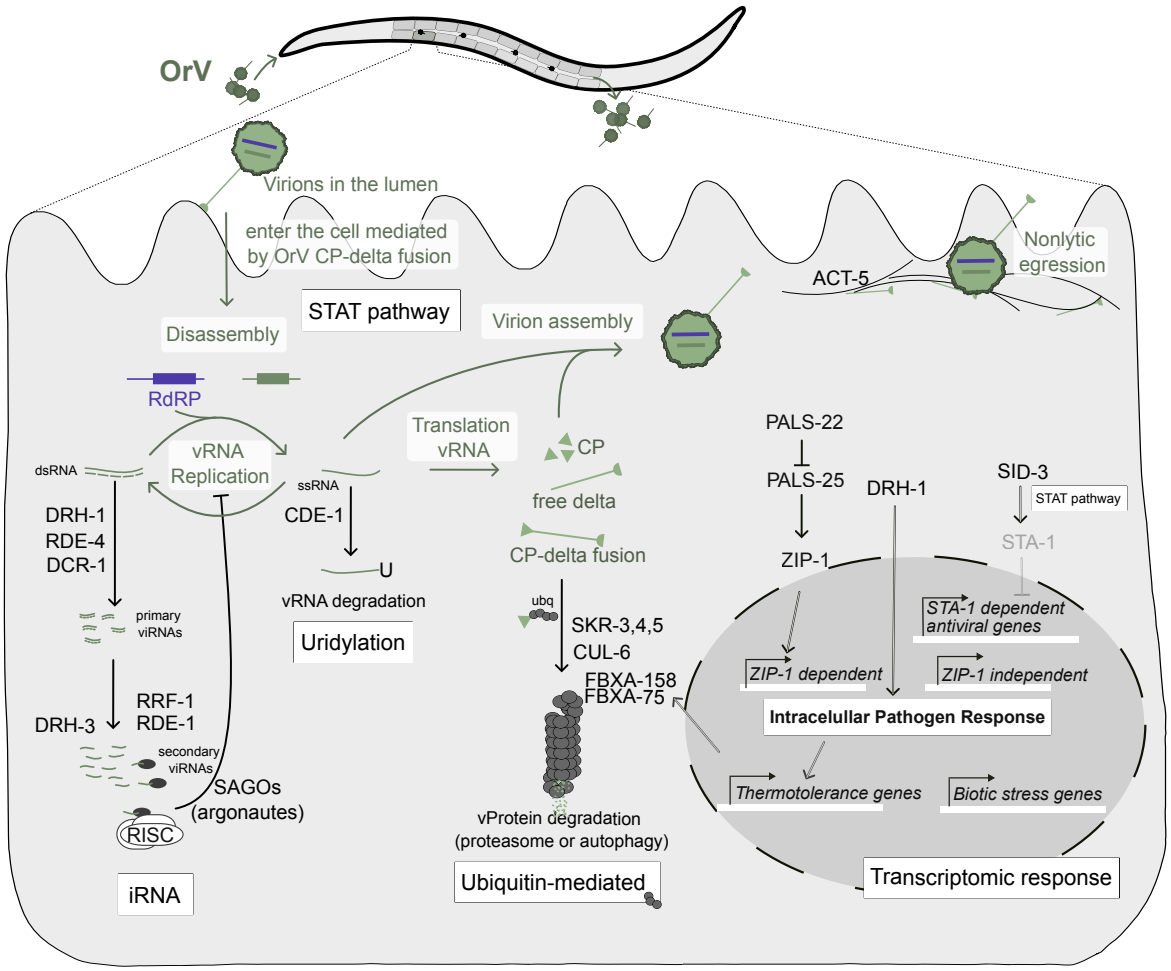



Fig. 1.5 *C. elegans* known immune responses to OrV infection. Figure modified from [66]. OrV follows an orofecal route. Virions enter the intestinal epithelial cells mediated by the CP-*delta* fusion protein. If there is a host receptor involved, it is not known yet. [160] suggests the STAT signaling pathway would be activated with the OrV entry, and SID-3 would mediate to activate the STA-1-dependent antiviral gene expresion. Once OrV is disassembled, the RdRP is translated and starts the transcription. The production of viral dsRNA is detected by DRH-1, starting the RNAi antiviral response and the activation of the IPR. The viral ssRNA is targeted for degradation through uridylation of the 3' end by CDE-1. The IPR response involves the activation of genes that confer thermotolerance (*cul-6*, *skr-2,3,4*, ...). The ubiquitination (ubq) of the viral proteins marks them for degradation by the proteasome or via autophagy. The new virions egress the cell in a nonlytic way mediated by free *delta* protein. The known antiviral responses, the pathways, and some of the host factors involved in the antiviral response are represented in black (with a white background for antiviral responses). The OrV processes and elements are in green, except for the RdRP that is in purple.

## 1.4 Virus-host interactions

> All that you touch you change.
> All that you change changes you.
> The only lasting truth is change.
>
> Octavia Butler, *Parable of the Sower, 1993*

The study of the virus-host interactions is essential for understanding how viruses exploit the host cell resources and for finding possible antiviral targets. There are several experimental approaches to identify and characterize these relationships and can be divided into genetic functional screenings and physical protein interaction screenings. A good overview of the methods can be found in [141, 140].

Viruses use their own proteins but also need to hijack the cell system to complete their cycles. Furthermore, as we briefly reviewed in the previous section, the host responds to the invasion with their own tools. The nature of these interactions can be protein-protein (PPI), protein-RNA, or RNA-RNA but PPI have been, probably because for the available techniques, the most well studied.

The host proteins that have a positive effect on the virus cycle are called *proviral factors*, while the ones limiting the virus survival or replication are termed *restriction factors* or *antiviral factors* [141]. These relations are not static, rather, they are the subject of evolution. Both hosts and viruses are continually adapting in response to one another while they are influenced by a broader environmental context and the relationships with other organisms within the ecosystem.

### 1.4.1 Host factors required for TuMV cycle

TuMV-TuMV and TuMV-*A.thaliana* physical interactions (specifically PPI) have been study predominantly with yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) methods [16]. A high-throughout Y2H screening has identified 378 novel protein-protein interactions between TuMV and *A. thaliana*[110]. This PPI network (PPIN), added to previously known data, resulted in a 394 virus-host and virus-virus PPIN, with NIb being the most connected viral protein (259 physical interactions), followed by CI (60) and VPg (51) (Fig. 1.6, in dark orange the virus-virus and in black the virus-plant interactions). This PPIN is not a fixed photo, and some mutations or posttranslational modifications in the interacting surfaces of the

viral and host proteins can potentiate or create new associations not contemplated here or debilitate or even remove others.

Some of the *A. thaliana* interacting proteins, such as RH8, RH9, RHD3, SCE1, or SUMO3, have been identified as proviral for TuMV [74, 100, 191, 116, 191, 110, 28] (Table 1.2, in orange in Fig. 1.6). The TuMV genome lacks the 5' cap structure, but VPg mimics them to allow the cap-eIF4E/eIF(iso)E mechanism [76]. In addition, AT1G30070, that interacts with NIb, HC-Pro, and CI, has been identified as antiviral (Table 1.2, in orange in Fig. 1.6). AT1G30070 may be involved in $Ca^{++}$-dependent ubiquitination and subsequent proteasomal degradation of target proteins and probably serves as a molecular bridge in ubiquitin E3 complexes. The protein localizes in the cytoplasm and in the nucleus [163], both cellular compartments where the interactions could occur. Recently, eIF4E1 has been identified as a restriction factor by enabling the plant to maintain its normal translation activity and preventing the accumulation of virus proteins [195].

Other proviral factors without a direct viral protein interaction have been also reported. SYP71 is an SNARE protein essential for the TuMV infection by mediating the fusion of the virus-induce vesicles with the chloroplast. SYP71 does not directly interact with TuMV proteins but with anothe host protein, VAP27-3 that directly interacts with the TuMV-6K2 [180] (Fig. 1.6, Table 1.2). SEC22, SYTA, and VTI11 are other proteins essential for TuMV infection through their role in the vesicle maturation and viral cellular movement [185].

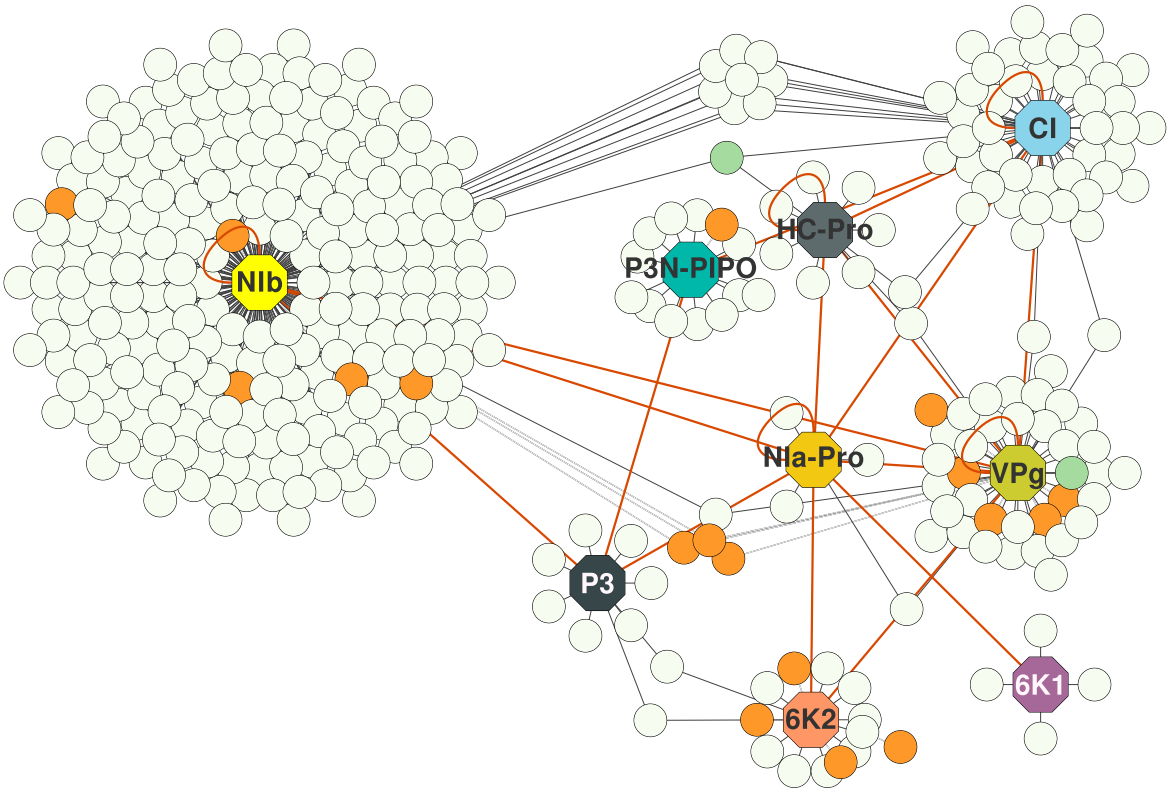



Fig. 1.6 TuMV-*A. thaliana* PPIN

.

Table 1.2 Host factors with physical interactions with TuMV proteins. Effect is "P" when a proviral or "A" an antiviral effect has been described in the reference in *Ref.* column. Methods used to identify the physical interaction (PI) include yeast two-hybrid (Y2H), bimolecular fluorescence complementation (BiFC), and enzyme-linked immunosorbent assay (ELISA). PI tested in [110] are identified with a "Yes" and with an asterisk ("*") the PI not contemplated in the PPIN from [110].

| **Viral protein** | **Host protein** | **Ath gene** | **Effect** | **Method** | **Ref.** | **Val. in [110]** | **Not in [110] PPIN** |
|---|---|---|---|---|---|---|---|
| HC-Pro | AT1G30070 | AT1G30070 | A | Y2H | [110] | Yes | |
| P3N-PIPO | PCAP1 | AT4G20260 | P | Y2H | [175] | Yes | |
| P3N-PIPO | PIP1-3 | AT1G01620 | | Y2H | [110] | Yes | |
| CI | AT1G30070 | AT1G30070 | A | Y2H | [110] | Yes | |
| 6K2 | Sec24a | AT3G07100 | P | Y2H | [80] | | |
| 6K2 | VSR4 | AT2G14720 | P | Y2H, BiFC | [186] | | * |
| 6K2 | RHD3 | AT3G13870 | P | Y2H, BiFC | [116] | | * |
| 6K2 | VAP27-3 | AT3G60600 | | Y2H, BiFC | [180] | | * |
| VPg | CNX1 | AT5G61790 | P | Y2H | [110] | Yes | |
| VPg | PAB2 | AT4G34110 | P | ELISA | [46] | | |
| VPg | PAB4 | AT2G23350 | P | ELISA | [46] | | |
| VPg | PAB8 | AT1G49760 | P | ELISA | [46] | | |
| VPg | eIF4E1, CUM1 | AT4G18040 | A | Y2H, ELISA | [195] | | |
| VPg | eIF(iso)4E, LSP1 | AT5G35620 | P | Y2H | [49] | | |
| VPg | eIF4G, CUM2 | AT3G60240 | | Y2H | [49] | | |
| VPg | eIF(iso)4G1 | AT5G57870 | | Y2H | [49] | | |
| VPg | eIF(iso)4G2 | AT2G24050 | | Y2H | [49] | | |
| Vpg | OBE1 | AT3G07780.1 | P* | Y2H | [47, 110] | Yes | |
| VPg | OBE2 | AT5G48160 | P | vY2H | [47] | | |

**Table 1.2 Continued from previous page**

| Viral protein | Host protein | Ath gene | Effect | Method | Ref. | Val. in [110] | Not in [110] PPIN |
|---|---|---|---|---|---|---|---|
| VPg | OBE3 | AT1G14740 | | Y2H | [47] | | |
| VPg | OBE4 | AT3G63500 | | Y2H | [47] | | |
| VPg | eEF1A | AT3G58470 | | ELISA | [164] | | |
| VPg | RH8 | AT4G00660 | P | Y2H, BiFC | [74] | | * |
| NIa-Pro | PAB2 | AT4G34110 | P | ELISA | [46] | | |
| NIa-Pro | PAB4 | AT2G23350 | P | ELISA | [46] | | |
| NIa-Pro | PAB8 | AT1G49760 | P | ELISA | [46] | | |
| NIb | AT1G30070 | AT1G30070 | A | Y2H | [110] | Yes | |
| NIb | PAB2 | AT4G34110 | P | Y2H | [46] | | |
| NIb | PAB4 | AT2G23350 | P | Y2H | [46] | | |
| NIb | PAB8 | AT1G49760 | P | Y2H | [46] | | |
| NIb | SCE1 | AT3G57870 | P | Y2H, BiFC | [191, 110] | Yes | |
| NIb | HSP70 | AT3G09440 | | vY2H | [159] | | |
| NIb | SUMO3 | AT5G55170 | P | v Y2H | [28, 110] | Yes | |
| NIb | RH9, MPH1 | AT3G22310 | P | Y2H, BiFC | [100] | | * |
| NIb | eEF1A | AT3G58470 | | ELISA | [164] | | * |
| NIb | TGA1 | AT5G65210 | P | Y2H | [110] | Yes | |
| NIb | NPR1-1 | AT1G64280.1 | P | Y2H | [110] | Yes | |
| NIb | RP40 | AT1G72370 | | Y2H | [110] | Yes | |

### 1.4.2 Host factors required for OrV cycle

Some of the last reviews on OrV-*C.elegans* pathosystem cover the known host factors that OrV needs to complete its cycle [66, 189].

As far as we know, all the interactions detected so far, have been identify with genetic interaction methods but not with physical interaction methods.

A non-exhaustive list of them include:

- The inositol phospholipid-binding domain and the F-actin-binding domain from *hipr-1* is essential for the early stages, probably for the pre-replication stage [79].
- Genes involved in signaling and actin nucleation: *sid-3*, *viro-2* and *nck-1* [78] and *drl-1* [139] are necessary in early stages of the viral cycle.
- The argonaute ALG-1 is essential for OrV replication and its function is independent from its splicing activity. Its interactor in the RISC, AIN-1, is also important for OrV replication [31].
- Suppresor and negative regulator of IPR genes: *pals-22* [135], *pnp-1* [161].
- Genes necessary to mantain the lipidic content: *sbp-1*, *fat-6*, *fat-7*, *elo-5* and *elo-6* [21]
- Genes necessary to mantain a endoplasmic reticulum homeostasis: *sel-11* and *ire-1* [48].

### 1.4.3 Virus-host coevolution

Virus-host relationships are not static but are subject to continuous evolution. Viruses, as obligatory intracellular parasites, rely on host cells to replicate, and they have evolved mechanisms to manipulate host processes for their benefit. Over time, both viruses and hosts have adapted to each another, creating a coevolutionary arms race. As hosts evolve defensive mechanisms, viruses mutate rapidly to evade these responses, often exploiting new genetic variations that improve them. Simultaneously, the host evolves new countermeasures, resulting in a dynamic interaction between the two that drives their continuous mutual adaptation.

This process can be understood through the **Red Queen Hypothesis**, which posits that organisms must perpetually evolve to maintain their fitness relative to other coevolving organisms, including pathogens and competitors. In the context of virus-host interactions, this hypothesis suggests that hosts and viruses must continuously adapt

to one another in a never-ending evolutionary "arms race" to survive. The Red Queen Hypothesis was first introduced by Leigh Van Valen in 1973 to explain the extintion law proposed in its work analyzing the extintion rate in the fossil record [171].

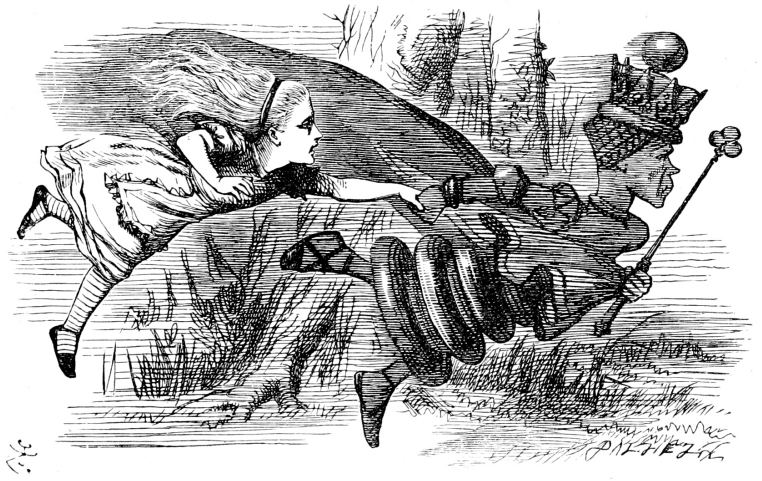


Fig. 1.7 “Now, here, you see, it takes all the running you can do, to keep in the same place.” Lewis Carroll, in *Through the Looking-Glass and What Alice Found There*, 1871. Illustration by John Tenniel.

Despite viruses have a substantial advantage due to its high mutation rates and large population sizes, deep mutation scanning (DMS) analyses suggest the evolutionary landscape may favor the host. Viruses are highly adaptable and evolve faster, but their mutational space is more restricted because their functional constrains, while hosts may have a greater and less fragile mutational space to explore [162].

One key aspect of virus-host coevolution is the existence (or not) of a **trade-off between virulence and transmission**. Under the conventional wisdom of virulence evolution, viruses that evolve greater virulence (causing severe disease in the host) often experience reduced transmission rates because they harm their host too quickly. On the other hand, viruses that evolve to be less virulent, typically improve host survival, which may enhance their transmission potential. This trade-off has been observed in human immunodeficiency virus type 1 (HIV-1), for which attenuation of virulence was associated with higher rates of transmission [12]. But these evolutionary trajectory to a less pathogenic virus-host relation is not only possible outcome. Back in 1982, R.M. Anderson and R. M. May [6] conclude that,

> *The complicated interplay between virulence and transmissibility of parasites leaves room for many revolutionary paths to be followed, with many endpoints.*

Pointing that the coevolutionary path of any given host-parasite relationship is ultimately determined by how the virulence and the production of transmission stages of the parasite are interconnected. Depending on the nature of this connection, the coevolutionary outcome could lead to either zero virulence, very high virulence, or

intermediate levels of virulence [6]. In 1987, Paul W. Ewald proposed the **parasitism-mutualism continuum hypothesis**, which stated that the evolution of a symbiont—toward either increased virulence (parasitism) or a mutually beneficial relationship (mutualism), with an intermediate stage of commensalism (non-pathogenic behavior)—depends on the pathogen's mode of transmission [50]. This hypothesis states that only a purely vertical transmission (through parentals to progeny) would select for mutualist symbionts while only a strict horizontal transmission (between organisms) would select for more virulent ones [50]. This hypothesis has been empirically tested in virus and vector-borne plants with positive and negative results [57, 125, 60].

On the one hand, Kover and Clay (1998), using a plant-fungus pathosystem, described that when a trade-off exists between virulence and vertical transmission, natural selection may favor increased virulence and horizontal transmission [87]. On the other hand, studies have shown that viral evolutionary trajectories from parasitism to mutualism do exist [147] and can be favored under stress conditions [64, 65], highlighting the need to understand host-guest relationships within the broader context of their environmental circumstance, including both abiotic factors and other biotic interactions.

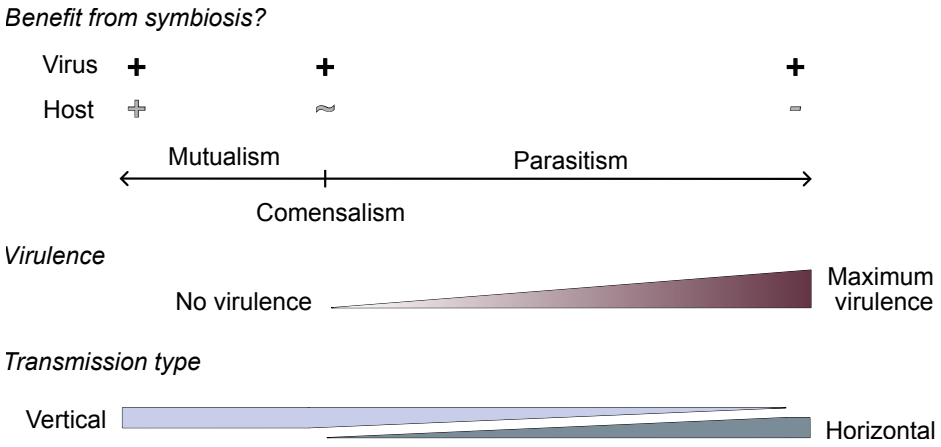



Fig. 1.8 Continuum mutualism-parasitism hypothesis.

.

Finally, other important consequence (or cause) of virus-host co-evolution is **cooptation**. Viruses needs host machinery in practically all the stages of the viral cycle and examples of cooptation have been review along this document (see also [146]). Examples of viruses hijacking canonically antiviral effectors and repurposing them for proviral functions have been identified in mammals. These examples include innate immunity effectors [84]. In [110], Arabidopsis TGA1 was identified as a host proviral factor for TuMV interacting with the viral RdRP NIb, while the same protein was previously reported as antibacterial in the infection with *Pseudomonas syringae*[149].

Conversely, hosts have coopted viral elements for their own benefit. Throughout evolutionary history, integrated viruses have played a role in shaping host genomes and providing adaptive advantages. As an example, the endogenous viral elements (EVEs) found in many eukaryotic genomes are remnants of ancient viral infections that have been incorporated into the host genome. These viral sequences have provided important functions, such as influencing immune responses or facilitating the development of placental tissues in mammals. EVEs cooption has benefit various host functions, but it appears to particularly have enhance immunity and antiviral defense [55].

# CHAPTER 2

# Aims, Structure of the Thesis and Other Contributions

## 2.1 Aims

The general and specific objectives of this Thesis are:

- To apply a systems biology approach to the study of viral infections caused by positive-sense single-stranded RNA ((+)ssRNA) viruses, focusing on both viral dynamics and host responses in two distinct temporal contexts: short-term and long-term.
  - To study the process of TuMV adaptation to *A. thaliana* genotypes with defects in different epigenetic pathways.
  - To analyze the transcriptional response of *A. thaliana* genotypes with defects in epigenetic pathways infected with *A. thaliana*-naïve and well-adapted TuMV isolates.
  - To study the accumulation of DVG of two betacoronaviruses during their evolution in different types of susceptible mammalian cells.
  - To study the accumulation of OrV, its mutational spectrum, including DVGs, along the infection of *C. elegans*.
  - To map the transcriptomic responses of *C. elegans* during the process of an OrV infection.

## 2.2 Structure of this Thesis

This thesis is presented as a compendium of four published articles.

With these four studies, we present a systems biology analysis of the dynamics of the viral changes and the transcriptomic responses elicited in their hosts. This involves the analysis of RNA sequencing data from different pathosystems in two distinct evolutionary scenarios, based on the temporal scale of the involved processes:

(1) the **short-term** dynamics of OrV infection in *C. elegans*, and

(2) the **long-term** evolution of three (+)ssRNA viruses: on the one hand, the experimental evolution of TuMV in *A. thaliana*, and on the other hand, the evolution of two betacoronaviruses (MHV and HCoV-OC43) in three mammalian cell types.

Additionally, based on the main focus of each study, the works can be categorized into two groups: (A) host responses-focused, and (B) viral diversity-focused.

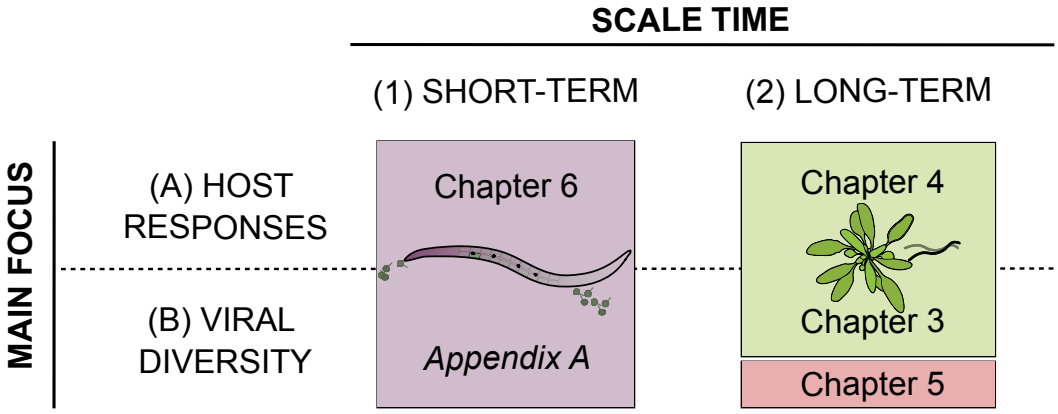



Fig. 2.1 Results classification with chapter correspondence.

Long-term viral-host interactions results:

1. **Chapter 3**: B.2. Ambrós S.*, **Olmo-Uceda M.J.***, Corrêa R.L., Elena S.F. (2024). Phenotypic and genomic changes during turnip mosaic virus adaptation to *Arabidopsis thaliana* mutants lacking epigenetic regulatory factors. *Evolution* **78**, 69–85. https://doi.org/10.1093/evolut/qpad192. [**B.2**]

   Bioinformatic analysis: RNA-seq data curation, viral diversity analysis, VPg structure prediction, and all the visualizations. Edition and revision of the manuscript.

2. **Chapter 4**: **Olmo-Uceda M.J.***, Ambros S.*, Correa R.L., Elena S.F. (2024). Transcriptomic insights into the epigenetic modulation of turnip mosaic virus

evolution in *Arabidopsis thaliana*. *BMC Genomics* **25**, 897. https://doi.org/10.1186/s12864-024-10798-x [**A.2**]

Conceptualization of all the analysis performed. Bioinformatic analysis: RNA-seq data curation, transcriptomic analsysis, transposable elements analysis, PPIN-related analysis and all the visualizations. Statistical analysis. Writing of the first draft together with Santiago F. Elena. Edition and revision of the manuscript.

3. **Chapter 5**: Hillung J.*, **Olmo-Uceda M.J.***, Muñoz-Sánchez J.C., Elena S.F. (2024). Accumulation dynamics of defective genomes during experimental evolution of two betacoronaviruses. *Viruses* **16**, 644. https://doi.org/10.3390/v16040644. [**B.2**]

   Conceptualization of all the analysis performed. Bioinformatic analysis: DVG analysis and all the visualizations. Most of the statistical analysis. Writing of the first draft together with Julia Hillung. Edition and revision of the manuscript.

Short-term viral-host interactions results:

1. **Chapter 6**: Castiglioni V.G.*, **Olmo-Uceda M.J.***, Villena-Giménez A., Muñoz-Sánchez J.C., Legarda E.G., Elena S.F. (2024). Story of an infection: Viral dynamics and host responses in the *Caenorhabditis elegans*-Orsay virus pathosystem. *Science Advances* **10**, adn5945. https://doi/10.1126/sciadv.adn5945 [**A.1**]

   Conceptualization of the project. Conceptualization of all the bioinformatic analysis performed. Bioinformatic analysis: RNA-seq data curation, transcriptomic analysis, comparative analysis, and visualizations. Most of the statistical analysis. Writing of the first draft together with Victoria G. Castiglioni. Edition and revision of the manuscript.

2. Results extra to Chapter 6 in **Appendix A**: **Olmo-Uceda M.J.** Dynamics of within-host replication and generation of diversity during single infections of Orsay virus in *Caenorhabiditis elegans*. [**A.2**]

* *Equal contribution.*

Fig. 2.1 shows the correspondence between the different results and the chapters constituting this Thesis. Associated to the Chapter 6, there are additional results on Appendix A, focused on the OrV accumulation and diversity dynamics.

## 2.3 Other Contributions

During the development of my PhD, I have led and participated in other projects; some of them are reflected in the following publications:

- **Olmo-Uceda, M.J.**, Muñoz-Sánchez J.C., Lasso-Giraldo W., Arnau V., Díaz-Villanueva W., Elena S.F. (2022). DVGfinder: A metasearch tool for identifying defective viral genomes in RNA-Seq data. *Viruses* **14**, 1114. doi: 10.3390/v14051114.

- Castiglioni V.G., **Olmo-Uceda, M.J.**, Martín S., Félix M.A., González R., Elena S.F.. (2024). Experimental evolution of an RNA virus in *Caenorhabditis elegans*. *Infection, Genetics and Evolution* **123**, 105623. doi: 10.1016/j.meegid.2024.105623.

- Muñoz-Sánchez J.C., **Olmo-Uceda, M.J.**, Oteo J.Á., Elena S.F. (2024). Quantifying defective and wild-type viruses from high-throughput RNA sequencing. *Bioinformatics* **40**, btae651. doi: 10.1093/bioinformatics/btae651.

- Muñoz-Sánchez J.C., Lázaro J.T., Hillung J., **Olmo-Uceda M.J.**, Sardanyés J., Elena S.F. (2025). Quasineutral multistability in an epidemiological-like model for defective-helper betacoronavirus infection in cell cultures. *Applied Mathematical Modelling* **137**, 115673. doi: 10.1016/j.apm.2024.115673.

- Hillung J., Lázaro J.T., Muñoz-Sánchez J.C., **Olmo-Uceda M.J.**, Sardanyés J., Elena S.F. (2024). Decay of HCoV-OC43 infectivity is lower in cell debris-containing media than in fresh culture media. *MicroPublications Biology* **16**, 001092. 10.17912/micropub.biology.001092. doi: 10.17912/micropub.biology.001092.

- Budzyńska D., Minicka J., **Olmo-Uceda M.J.**, Elena S.F., Hasiów-Jaroszewska B. (2024). Population dynamics of defective viral genomes of tomato black ring virus during host-to-host transmission. *Journal of Virology* **98**, e0124424. doi: 10.1128/jvi.01244-24.

- Silva, J.M.F., **Olmo-Uceda M.J.**, Morley, V.J., Turner, P.E., Elena, S.F. (2025). Taylor's power law rules the dynamics of allele frequencies during viral evolution in response to host changes. *bioRxiv* doi: 10.1101/2024.12.27.630372.

- Elena, S.F., **Olmo-Uceda M.J.**, Ambrós, S. (2025). The pace of ecological succession among host genotypes determines the virulence and the dynamics of genome evolution in a plant RNA virus. *bioRxiv* doi: 10.1101/2025.0327.645716.

As part of my fellowship –*Formación de Profesorado Universitario*, Ministerio de Ciencia, Innovación y Universidades, Gobierno de España– I have taught 100 hours of Fundamentals of Programming in the Data Science Degree (Dept. de Informatics, *Universitat de València*), during the academic years 2021-22, 2022-23, and 2023-24.

Additionally, I enjoyed a short stay in the Santa Fe Institute in November 2022. I have participated in two international collaborations: (1) with the Laboratory of Plant Molecular Biology & Virology from the National Taiwan University, led by Prof. Shih-Shun Lin; and (2) a stay in the laboratory of Prof. Harmit Malik at the Fred Hutchinson Cancer Research Institute during the summer of 2024. Finally, I have participated in a national collaboration with the Mathematical And Computational Biology from *Centre de Reserca Matemàtica*, led by Dr. Josep Sardanyés Cayuela.

# CHAPTER 3

## Phenotypic and genomic changes during turnip mosaic virus adaptation to *Arabidopsis thaliana* mutants lacking epigenetic regulatory factors

*Evolution*, 2024, **78(1)**, 69–85
https://doi.org/10.1093/evolut/qpad192
Advance access publication 28 October 2023
**Original Article**

# Phenotypic and genomic changes during *Turnip mosaic virus* adaptation to *Arabidopsis thaliana* mutants lacking epigenetic regulatory factors

**Silvia Ambrós[1], María J. Olmo-Uceda[1], Régis L. Corrêa[1,2], Santiago F. Elena[1,3]**

[1]Instituto de Biología Integrativa de Sistemas (I²SysBio), CSIC-Universitat de València, Valencia, Spain
[2]Department of Genetics, Universidade Federal do Rio de Janeiro, Rio do Janeiro, Brazil
[3]Santa Fe Institute, Santa Fe, NM, United States

Corresponding author: Instituto de Biología Integrativa de Sistemas (I2SysBio), CSIC-Universitat de València, Catedrático Agustín Escardino 9, Paterna, 46980 Valencia, Spain. Email: santiago.elena@csic.es
S.A. and M.J.O.U. contributed equally to this study.

## Abstract

In this study, we investigated how an emerging RNA virus evolves, interacts, and adapts to populations of a novel host species with defects in epigenetically controlled plant defense mechanisms. Mutations in epigenetic regulatory pathways would exert different effects on defense-response genes but also induce large-scale alterations in cellular physiology and homeostasis. To test whether these effects condition the emergence and subsequent adaptation of a viral pathogen, we have evolved five independent lineages of a naive turnip mosaic virus (TuMV) strain in a set of *Arabidopsis thaliana* genotypes carrying mutations that influence important elements of two main epigenetic pathways and compare the results with those obtained for viral lineages evolved in wild-type plants. All evolved lineages showed adaptation to the lack of epigenetically regulated responses through significant increases in infectivity, virulence, and viral load although the magnitude of the improvements strongly depended on the plant genotype. In early passages, these traits evolved more rapidly, but the rate of evolution flattened out in later ones. Viral load was positively correlated with different measures of virulence, though the strength of the associations changed from the ancestral to the evolved viruses. High-throughput sequencing was used to evaluate the viral diversity of each lineage, as well as characterizing the nature of fixed mutations, evolutionary convergences, and potential targets of TuMV adaptation. Within each lineage, we observed a net increase in genome-wide genetic diversity, with some instances where nonsynonymous alleles experienced a transient rise in abundance before being displaced by the ancestral allele. In agreement with previous studies, viral VPg protein has been shown as a key player in the adaptation process, even though no obvious association between fixed alleles and host genotype was found.

**Keywords:** DNA methylation, experimental evolution, histone modification, virus evolution, virus–host interactions

Plants and viruses interact, triggering defense and counter-defense mechanisms that frequently lead to coevolution and reciprocal adaptation. A variety of plant immune processes participate in antiviral defense (Soosaar et al., 2005; Zhou & Zhang, 2020). These antiviral factors can be broadly divided into two groups: (a) basal, which are preexisting and restrict within-cell propagation and cell-to-cell spread and (b) inducible, which are triggered following infection and inhibit systemic movement and replication. Inducible mechanisms involve genes whose expression results in a broad-scale change in plant physiology via a variety of signaling pathways, as opposed to basal mechanisms, corresponding to alleles of cellular genes whose products interact inefficiently with viral factors (Carr et al., 2010). These changes include local cell apoptosis (Loebenstein, 2009), upregulation of nonspecific responses against pathogens throughout the entire plant (systemic acquired and induced resistances) (Carr et al., 2010; Kachroo et al., 2006), and the activation of the RNA-silencing-based resistance, which seems to play a role both in basal and inducible mechanisms (Carr et al., 2010; López-Gomollon & Baulcombe, 2022; Voinnet, 2001). RNA-based immunity is triggered by the recognition and degradation of double-stranded RNAs (dsRNA). Viral dsRNAs produced during replication are degraded by Dicer-like (DCL) proteins into small RNAs (sRNA) that are loaded into Argonaute (AGO) proteins and used as guide to repress complementary RNAs (Voinnet, 2001). The silencing response can be enhanced by RNA-dependent RNA polymerases (RDR), which produce extra dsRNAs from targets (Borges & Martienssen, 2015).

The RNA-mediated mechanisms employed by plants to defend against viruses are a component of a larger and evolutionarily conserved system that governs gene expression and manages transposable elements (TE) via the addition of epigenetic modifications to DNA or histones (Hung & Slotkin, 2021). DNA methylation can be observed in all sequence contexts in eukaryotes (CG, CHG, and CHH; with H ≠ G). RNA-directed DNA methylation (RdDM) is a well-studied mechanism that uses sRNAs produced from the borders of TEs to restrict their expression, influencing the expression of neighboring genes (Liu et al., 2022). RdDM plays an important role in determining heterochromatin boundaries,

Received May 22, 2023; revisions received September 22, 2023; accepted October 27, 2023
Associate Editor: Brandon Ogbunugafor; ; Handling Editor: Tracey Chapman
© The Author(s) 2023. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE).
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

coordinating the actions of components such as methyltransferase domains rearranged methylase 2 (DRM2), RNA polymerase IV (POLIV), and RNA polymerase V (POLV) (Böhmdorfer et al., 2016). Alternative noncanonical pathways (e.g., POLII-derived mRNAs, RDR6-derived dsRNAs, or DCL-independent mechanisms) feed into the canonical RdDM pathway (Cuerda-Gil & Slotkin, 2016). Gene expression and TE mobilization can also be controlled via sRNA-independent mechanisms in which epigenetic marks are copied during replication by maintenance DNA methyltransferases such as the plant-specific chromomethylase 3 and require chromatin remodelers such as decreased DNA methylation 1 (DDM1) (Bond & Baulcombe, 2014). The introduction of changes in histones is commonly related to the targets of DNA methylation marks. Histone H3 trimethylation of lysine 9 (H3K9m3), for example, has been linked to TE repression, whereas H3K4m3 has been linked to gene expression stimulation. DNA methylation and histone modification are both reversible markers. The proteins repressor of silencing 1 (ROS1), increase in bonsai methylation 1 (IBM1), and jumonji14 (JMJ14), for example, are involved in the demethylation of DNA, H3K9, and H3K4, respectively (Gong et al., 2002; Lu et al., 2010; Saze et al., 2008). The chromatin environment within or surrounding TEs and genes is altered under stressful situations or in epigenetically deficient mutants, influencing their expression (Lloyd & Lister, 2021).

The link between the loss of DNA methylation factors and resistance or susceptibility to bacteria and fungi has been well established (e.g., Dowen et al., 2012; Le et al., 2014; López et al., 2011; López Sánchez et al., 2016; Luna et al., 2012; Yu et al., 2013). However, the role of RNA virus infection in DNA methylation responses had received less attention. Using turnip crinkle virus (Diezma-Navas et al., 2019), two tobamoviruses (Leone et al., 2020), and turnip mosaic virus (TuMV) (Corrêa et al., 2020), it has been shown that *Arabidopsis thaliana* mutants for various RdDM factors, chromatin remodelers, and histone modifiers were more or less susceptible to infections depending on the presence of repressive marks. Furthermore, evolution experiments in mutants for innate immunity pathways or basal and inducible defense pathways have shown that viral populations quickly respond to the novel selective pressures imposed by mutations in the RdRM genes (Mongelli et al., 2022; Navarro et al., 2022), suggesting a direct role of these genes in the infection cycle of RNA viruses.

In this work, we aim to explore the role that epigenetic regulation of defense responses to biotic stresses exerts in the emergence of a novel viral pathogen that spills over from its reservoir host. We hypothesize that given the expected large-scale effects of mutations in epigenetic regulatory pathways in plant defenses and cellular homeostasis, virus evolution and adaptation to the novel host will be affected to some extent. This impact can be either direct, if epigenetic factors directly interact with virus factors, or indirect if epigenetic regulation modifies the expression of plant resistance genes that ultimately interact with viral factors or induces a strong physiological dysregulation that represents a different cellular environment for the virus. We leveraged the knowledge of *A. thaliana* epigenetic regulation mechanisms, the intrinsic variability and evolvability of the TuMV mutant swarm, and the great power of high-throughput RNA-Seq to study the impact of epigenetic regulation of host responses to the adaptation of this viral pathogen to a novel host species upon spilling over from its reservoir host. The results from TuMV evolution experiments in plants defective for several epigenetic pathways show that the host genotype affects viral genetic diversity independent of the pathway being disrupted, and it is followed by an increase in virulence along evolutionary passages.

## Materials and methods

### Plants, viruses, and growth conditions

A collection of 20 different *A. thaliana* (L.) Heynh mutants of the Col-0 accession have been used (Table 1). This collection covers a number of single, double, and even triple mutants affecting different genes involved in canonical and noncanonical RdDM, sRNA-independent DNA methylation, and histone modification. In all experiments, plants were maintained in a BSL2 climatic chamber under a photoperiod of 8 h light (LED tubes at PAR 90–100 μmol/m²/s) at 24 °C and 16 h dark at 20 °C and 40% relative humidity in a mixture of 50% DSM WNR1 R73454 substrate (Kekkilä Professional, Vantaa, Finland), 25% grade 3 vermiculite and 25% 3–6 mm perlite. Pest management was performed by the introduction of *Stratiolaelaps scimitus* and *Steinernema feltiae* (Koppert Co., Málaga, Spain).

Certain TuMV strains show high prevalence in wild *A. thaliana* populations (Pagán et al., 2010). However, the YC5 strain used in this study was originally isolated from calla lily plants (Chen et al., 2003), and it is naive to *A. thaliana*. The viral stock used in all the experiments was obtained by inoculating the plasmid p35STunos containing a cDNA of the YC5 genome (GenBank AF30055.2) under the control of the 35S promoter and the *nos* terminator (Chen et al., 2003) in *Nicotiana benthamiana* Domin plants as described in Navarro et al. (2022).

### Phenotyping infection

Five different disease-related traits were measured for each plant (20 for preliminary experiments and 10 for evolution experiments) during 14 days postinoculation (dpi) (17 dpi in the case of the second block of preliminary experiments): (a) Infectivity, as the frequency of symptomatic plants out of the number of inoculated plants at 14 dpi. (b) Area under the disease progress stairs (*AUDPS*) summarizes in a single figure the temporal dynamics at which symptoms first appear (Simko & Piepho, 2012); $AUDPS \in [0, T]$, where $T$ is the total number of days in the assay. (c) Average severity of symptoms at 14 dpi, evaluated on a discrete scale ranging from zero for noninfection or asymptomatic infections to five for plants showing generalized necrosis and wilting (Butković et al., 2021). (d) Area under the symptoms intensity progression stair (*AUSIPS*) summarizes both the severity of symptoms and their temporal dynamics; $AUSIPS \in [0, 5T]$. And (e) viral load, measured by absolute reverse transcription real-time quantitative polymerase chain reaction (RT-qPCR) as the number of viral genomes per ng of total RNA in the plants as described below. Viral load was estimated after passages 1, 4, 7, 9, 10, and 12 with threefold replication.

Plant genotypes were clustered according to their phenotypic similarity in response to TuMV infection using the *k*-means method. Goodness of fit of models with an increasing number of clusters was evaluated using the minimum

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

**Table 1.** Different *A. thaliana* genotypes used in this study. Highlighted in gray are those used for the evolution experiments

| Genotype (allele) | Gene name | Affected pathway and expected phenotype | Reference |
|---|---|---|---|
| Wild-type (WT) | | | |
| *ago4-2* | *AGO4* | Canonical and DCL-independent RdDM initiation.<br>Reduced non-CpG methylation and gene activation in RdDM targets. | Zilberman et al. (2003) |
| *atx1-2* | *ATX1* | H3K4 histone methyltransferase.<br>Repression of ATX1 targets, including pathogen- and disease-resistance encoding genes. | Alvarez-Venegas et al. (2003) |
| *cmt3-11t* | *CMT3* | Maintenance CHG methylation.<br>CHG siRNA-independent DNA methylation lost. | Lindroth et al. (2001) |
| *dcl2-1 dcl3-1 dcl4-2t* | *DCL2*, *DCL3*, *DCL4* | siRNAs production.<br>All DCL-dependent RdDM and RNAi pathways compromised. | Bouché et al. (2006) |
| *ddm1-2* | *DDM1* | Chromatin remodeling and stability.<br>All siRNA-independent methylation is compromised in repeated sequences. | Jeddeloh et al. (1999) |
| *ddm1-2 polIV* | *DDM1*, *NRPD1a* | Chromatin remodeling and stability and canonical RdDM.<br>Only POLII initiation in TE induced situation. | McCue et al. (2012) |
| *ddm1-2 rdr6-15* | *DDM1*, *RDR6* | Chromatin remodeling and stability and amplification of siRNAs.<br>POLIV-RDR2 pathways working in TE induced situation. | McCue et al. (2012) |
| *ddm1-2 polIV rdr6-15* | *DDM1*, *NRPD1a*, *RDR6* | Chromatin remodeling and stability, canonical RdDM and siRNA amplification<br>POLIV and RDR6-dependent RdDM pathways compromised. | Nuthikattu et al. (2013) |
| *drm1-2 drm2-2* | *DRM1*, *DRM2* | Canonical and noncanonical RdDM.<br>siRNA-dependent CHH lost. | Cao & Jacobsen (2002) |
| *drm1-2 drm2-1 cmt3-7* | *DRM1*, *DRM2*, *CMT3* | Canonical and noncanonical RdDM.<br>siRNA-independent and -dependent CHG/CHH lost. | Chan et al. (2006) |
| *hda6-6* | *HDA6* | Histone deacetylase.<br>Increased histone acetylation in HDA6 targets (genes and repeats). | Earley et al. (2006) |
| *ibm1-4* | *IBM1* | H3K9 histone demethylase.<br>H3K9me2 and CHG DNA hypermethylation in a large number of genes. | Saze et al. (2008) |
| *jmj14-1* | *JMJ14* | H3K4 histone demethylase.<br>Increased H3K4me3 and gene activation of JMJ14 targets. | Searle et al. (2010) |
| *polIV* (*nrpd1a-3*) | *NRPD1a* | Canonical and NERD[a] RdDM.<br>Loss of 24 nt siRNAs associated with de novo methylation of repeats. | Herr et al. (2005) |
| *polV (SALK_017795)* | *NRPE1* | Canonical, NERD and RDR6 RdDM.<br>Derepression of TEs mostly from intergenic regions. | Pontier et al. (2005) |
| *rdr1-1 rdr2-1 rdr6-15* | *RDR1*, *RDR2*, *RDR6* | RNA silencing, tasiRNA, Canonical RdDM and RDR6-RdDM.<br>Most RNA silencing and RdDM pathways compromised. | Garcia-Ruiz et al. (2010) |
| *rdr2-1* | *RDR2* | Canonical RdDM.<br>Loss of 24 nt siRNA amplification associated with de novo methylation of repeats. | Xie et al. (2004) |
| *rdr6-15* | *RDR6* | RNA silencing, tasiRNA and RDR6-RdDM.<br>Unable to amplify sRNAs. POLIV-RDR2 pathway still working. | Mourrain et al. (2000) |
| *ros1-4* | *ROS1* | DNA demethylase. Spread of repressive marks from TEs to nearby genes. | Gong et al. (2002) |

Note: [a]NERD: Needed for RDR2-independent DNA methylation. Binds to H3K4 and triggers siRNA-dependent DNA methylation (Pontier et al., 2012).

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

value of the Bayes information criterion (BIC) and the associated Akaike's weights.

## Experimental evolution

Five TuMV lineages were evolved during 12 consecutive serial passages in each one of the five selected genotypes. To begin the evolution experiment, 10 *A. thaliana* plants per lineage and genotype were inoculated as described above with the virus stock from *N. benthamiana*. After plants showed symptoms of infection, they were pooled and frozen with liquid $N_2$. Frozen plant tissue was homogenized using a Mixer Mill MM400 (Retsch GmbH, Haan, Germany). For inoculations, 0.1 g were diluted in 1-ml inoculation buffer (50 mM phosphate buffer pH 7.0, 3% PEG6000, 10% Carborundum) and 5 µl rubbed onto two leaves per plant. Plants were all inoculated at growth stage 3.5 (Boyes et al., 2001). This synchronization ensures that all individual hosts are at the same phenological stage when inoculated. The next passages were made by harvesting the symptomatic plants at 12 dpi, preparing the infectious sap, and inoculating it into a new batch of 10 plants.

Following the evolution experiment, a cross-infection experiment using all 24 surviving viral lineages was conducted in a single experimental block using 10 plants from each combination of the five plant genotypes utilized in the evolution experiment. Each of the 1,200 plants was monitored throughout 14 dpi, and the severity of the symptoms for each viral lineage–host genotype combination recorded.

## Statistical analyses of disease-related phenotypic traits

Prior to any further analysis, infectivity data were probit-transformed to ensure normality. Transformed infectivity, *AUDPS*, symptoms severity, and *AUSIPS* data were fitted to a MANCOVA model with plant genotype ($G$) as main factor, independent evolution lineages ($L$) nested within $G$, and passage ($t$) as a covariable:

$$\begin{aligned}\mathbf{X}_{ijk}(t) \sim{}& \bar{\mathbf{x}} + t + G_i + L(G)_{ij} \\ &+ (t \times G)_i + [t \times L(G)]_{ij} + \varepsilon_{ijkt}\end{aligned} \quad (1)$$

where $\mathbf{X}_{ijk}(t)$ is the vector of disease traits observed at time $t$, for an individual infected plant $k$ of evolutionary lineage $j$ of genotype $i$, $\bar{\mathbf{x}}$ represents the vector of grand mean values and $\boldsymbol{\varepsilon}_{ijkt}$ the vector of errors assumed to be Gaussian. The significance of each factor was tested using the Wilk's $\Lambda$ summary statistic. Log-viral load data were fitted to an ANCOVA model with an identical structure to equation 1.

Rates of phenotypic evolution for each disease trait were estimated by fitting the corresponding time series to the first-order ARIMA(1,0,0) model:

$$D(t) - \rho_1 D(t-1) = D(0) + \nu_D t + \varepsilon_t \quad (2)$$

where $D(k)$ represents any of the four disease traits at passage $k$, $\rho_1$ measures the degree of self-similarity in the times-series data, $\varepsilon_t$ represents the sampling error at passage $t$, and $\nu_D$ the linear dependency of trait $D$ with $t$, that is, the rate of phenotypic evolution. Estimated rates of phenotypic evolution were fitted to a MANOVA with the following structure:

$$\mathbf{V}_{ijk} \sim \bar{\mathbf{v}} + P_i + G(P)_{ij} + \varepsilon_{ijk} \quad (3)$$

where $\mathbf{V}_{ijk}$ is the vector of rates of phenotypic evolution, $\bar{\mathbf{v}}$ the vector of grand mean values and $\boldsymbol{\varepsilon}_{ijk}$ the vector of Gaussian errors. $P$ stands for the epigenetic pathway and $G$ is defined as in equation 1.

Cross-infection severity data were fitted to a repeated measures ANOVA in which local host ($LH$) and test host ($TH$) were treated as orthogonal factors, independent evolution lineages ($L$) were nested within $LH$ and $t$ was treated as the within-plant factor:

$$\begin{aligned}S_{ijklt} \sim{}& \bar{s} + t + t \times LH_i + t \times TH_j \\ &+ t \times L(LH)_{ik} + t \times (LH \times TH)_{ij} \\ &+ t \times (TH \times L(LH))_{ijk} + \xi_{ijklt} \\ &+ LH_i + TH_j + L(LH)_{ik} + (LH \times TH)_{ij} \\ &+ (TH \times L(LH))_{ijk} + \varepsilon_{ijkl}\end{aligned} \quad (4)$$

where all terms involving $t$ evaluate the within-plant variance, $\bar{s}$ refers to the grand mean value of symptoms severity, $\xi_{ijklt}$ to the error among observations for a given plant, and $\varepsilon_{ijkl}$ the error among plants, both assumed to be Gaussian.

In all models, the type III sum of squares was used to partition total variance among factors. The magnitude of the effects was evaluated using the $\eta_P^2$ statistic. Values $\eta_P^2 \geq 15\%$ are considered as large effects. All these statistical analyses were done with SPSS version 28.0.1.0 (IBM, Armonk, NY, USA).

## Total RNA extractions, quantification of viral load, and preparation of samples for RNA-Seq

Pools were made of 10 infected symptomatic plants per lineage, genotype, and serial passage, frozen with liquid $N_2$ and preserved at −80 °C until homogenized into fine powder. Aliquots of ~0.1 g were used for total RNA (RNAt) extractions with the Agilent Plant RNA isolation Mini kit (Agilent Technologies, Santa Clara, CA, USA). Three aliquots of RNAt per sample were separated, and their concentrations were adjusted to 50 ng/µl.

Viral load was quantified in each aliquot by RT-qPCR using standard curves and specific forward F117_F and reverse F118_R primers, which amplify sequence 9,320–9,492 at the very 3′ end of the CP gene, in an ABI StepOne Plus Real-time PCR System (Applied Biosystems, Foster City, CA, USA) as described elsewhere (Corrêa et al., 2020; Navarro et al., 2022). This choice of primers ensures that only full positive-sense genomes will be amplified. For each sample, three technical replicates were quantified. Results were analyzed using the StepOne software 2.2.2 (Applied Biosystems).

For RNA-Seq, RNAt from 70- to 90-mg infected and healthy plants was extracted using the GeneJET Plant RNA Purification Mini Kit (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. RNA quality was checked with NanoDrop One (Thermo Fisher Scientific) and agarose gel electrophoresis, and integrity and purity were tested with Bioanalyzer 2100 (Agilent Technologies). Library preparation and Illumina sequencing were done by Novogene Europe using a NovaSeq 6000 platform and an Lnc-stranded mRNA-Seq library method, ribosomal RNA depletion and directional library preparation, 150 paired end, and 6 Gb raw data per sample. Novogene did a quality check of the libraries using a Qubit 4 Fluorometer (Thermo Fisher Scientific), qPCR for quantification, and bioanalyzer for size distribution detection.

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

### RNA-Seq data processing and measures of viral diversity

The quality of the Fastq files was assessed with FASTQC (Andrews, 2010) and MultiQC (Ewels et al., 2016) and pre-processed as paired reads with BBDuk (https://sourceforge.net/projects/bbmap/). Adapters were removed, the first ten 5′ nucleotides of each read were cut, and the 3′ end sequences with average quality below 10 were removed. Processed reads shorter than 80 nucleotides were discarded. The parameter values were set to: *ktrim* = r, *k* = 31, *mink* = 11, *qtrim* = r, *trimq* = 10, *maq* = 5, and *forcetrimleft* = 10, and *minlength* = 80. The processed files were mapped to TuMV YC5 genome with the BWA-MEM algorithm (Li, 2013). Resulting SAM files were binarized and sorted with SAMtools (Danecek et al., 2021) and duplicates marked with the MarkDuplicates method of GATK version 4.2.2.0 (McKenna et al., 2010).

For TuMV variants calling, the viral assemblies without PCR duplications were used as input for LoFreq (Wilm et al., 2012). Nucleotide counts per position were obtained from the same processed BAM files with pysamstats version 1.1.2 (https://hpc.nih.gov/apps/pysamstats.html). These counts files were used to calculate two diversity metrics: (a) Shannon's entropy per nucleotide site and (b) Shao et al. (2014) pairwise alignment positional nucleotide counting method which allows the calculation of average pairwise distance (APD):

$$\mathrm{APD} = \frac{2}{L}\sum_{i=1}^{L}\frac{A_i(C_i+G_i+U_i)+C_i(G_i+U_i)+G_iU_i}{N_i(N_i-1)}$$

where $L$ is the length of the reference genome, $N_i$ the number of total nucleotides, and $A_i$, $C_i$, $G_i$, and $U_i$ the counts of each base at site $i$. APD data were analyzed using a two-way nonparametric aligned rank transform (ART) ANOVA (Wobbrock et al., 2011), with passage and plant genotype as orthogonal factors. In this case, the type I sum of squares was used. This test was done using the R package ARTool version 0.11.1 (https://cran.r-project.org/web/packages/ARTool/index.html).

Viral consensus sequences were obtained for each lineage after the first and last passages using the consensus function in SAMtools and used to evaluate the possible role of selection in the observed patterns of molecular diversification among lineages evolved in the same and different host genotypes. Four different approaches have been taken. First, we evaluated Nei's coefficient of nucleotide differentiation $N_{ST}$ (Nei, 1982). Second, Tajima's $D$ test of selection (Tajima, 1989) was computed, and its significance was evaluated using DnaSP version 6.12.03 (Rozas et al., 2017). Third, to evaluate the sign and strength of selection on each of the 10 cistrons in the large ORF, we obtained the average difference among nonsynonymous and synonymous substitution rates, $d_N$–$d_S$, using the modified Nei–Gojobori method. These computations were done in MEGA 11 (Tamura et al., 2021); in each analysis, the nucleotide substitution model with the lowest BIC was chosen. Standard errors were estimated by 1,000 bootstrap resampling. Finally, we performed a codon selection analysis using the fixed effects likelihood (FEL) method (Kosakovsky Pond & Frost, 2005), as implemented in the datamonkey.org server, and with a cutoff significance of $p = .05$.

The VPg structure was predicted using AlphaFold version 2.3.2 (Collaboratory version: https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb#scrollTo=pc5-mbsX9PZC) (Jumper et al., 2021) and visualized with UCSF ChimeraX version 1.2 (Pettersen et al., 2021).

## Results and Discussion

The role of epigenetics in short-term adaptation and long-term evolution has attracted attention as details of the diverse molecular mechanisms involved have been revealed (Ashe et al., 2021; Klironomos et al., 2013; Mendizabal et al., 2014; Rapp & Wendel, 2005; Richards et al., 2010; Stajic & Jensen, 2021; Stajic et al., 2019). Epigenetic information is more labile than hard-encoded DNA one, which implies higher rates of genomic modifications that translate into more phenotypic variance upon which selection may operate facilitating adaptation to sudden challenges (Banta & Richards, 2018; Richards et al., 2010). A prototypical example of fast environmental change is created by infection by prevalent parasites (Gómez-Díaz et al., 2012; Kasuga & Gijzen, 2013). Of particular interest, viruses have been shown to manipulate the epigenetic pathways during infection. For example, the NS1 protein of influenza A virus contains an amino acid sequence that mimics the host histone H3K4 to hijack transcription elongation factors (Marazzi et al., 2012). Mutant *A. thaliana* plants having loss or gain of epigenetic marks were, respectively, more tolerant and susceptible to TuMV, while infection of wild-type (WT) plants was associated with changes in DNA methylation patterns in protein-coding genes (Corrêa et al., 2020). Together, these observations suggest that epigenetic regulation of host genes impacts viral replication and, therefore, would exert a selective pressure upon emerging viral populations that spilled over from a reservoir host into a novel one. To test this hypothesis, we have explored the impact that mutations in components of *A. thaliana* epigenetic regulatory pathways have on an emerging viral pathogen, a TuMV strain isolated from calla lily.

### Classification of plant genotypes based on the TuMV infection phenotype

In the first study, we evaluated the effect of mutations in different epigenetic pathways (Table 1) on TuMV infectivity (Supplementary Figure 1A), symptom severity (Supplementary Figure 1B), and virus accumulation (Supplementary Figure 1C).

Fitting the infectivity data in Supplementary Figure 1A to a GLM, with plant genotype and days postinoculation (dpi) as orthogonal factors with a binomial distribution and probit link function, shows significant differences among plant genotypes that change in magnitude along dpi (test of interaction term: $\chi^2 = 450.331$, 323 df, $p < .001$, $\eta^2_P = 31.2\%$). Symptom severity curves (Supplementary Figure 1B) were fitted to a repeated measures ANOVA with dpi as an intraindividual factor and plant genotype as an interindividual orthogonal factor. Highly significant differences exist among plant genotypes that depend on the precise time in which individuals are observed (test of interaction term: $F_{247,4927} = 12.394$, $p < .001$, $\eta^2_P = 38.3\%$). Finally, viral load data (Supplementary Figure 1C) were fitted to a GLM with plant genotype and dpi as orthogonal factors with a Gamma distribution and log-link function. Viral load also varies among plant genotypes in a time-dependent manner (test of interaction term: $\chi^2 = 589.111$, 24 df, $p < .001$, $\eta^2_P = 33.9\%$). Henceforth, we conclude that significant differences exist among plant epigenetic mutants in their phenotypic response to TuMV infection, ranging from

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

the fast dynamics observed in *rdr1 rdr2 rdr6* to the slow one seen in *ddm1 polIV*.

Since performing evolution experiments in all the genotypes listed in Table 1 would be practically unmanageable, we decided to select genotypes based on the similarity of their responses to infection with the naive TuMV strain. For this purpose, the mutant plants were classified on the basis of their phenotypic responses to infection using a *k*-means clustering method. The smallest BIC = 33.950 (Akaike's weight 60.22%) was found for $k = 4$ phenogroups (Supplementary Table 1). The four phenogroups contained a heterogeneous mixture of genotypes. Phenogroup I includes mutants involved in siRNA-dependent RdDM (*ago4*, *ddm1*, *ddm1 polIV*, *ddm1 polIV rdr6*, *ddm1 rdr6*, *polV*, and *rdr6*) and histone modification (*hda6* and *jmj14*) pathways (Table 1). Phenogroup II was formed by a functionally diverse group of mutants: *drm1 drm2*, *polIV*, and *rdr2* affect the siRNA-dependent RdRM, *ibm1* is a H3K9 histone demethylase, and *ros1* is a DNA demethylase (Table 1). Phenogroup III was formed by *dcl2 dcl3 dcl4*, *drm1 drm2 cmt3*, and *rdr1 rdr2 rdr6*. This phenogroup shows functional coherence, since all these mutants affect different steps of the RNA-silencing and siRNA-dependent RdDM pathways (Table 1). Finally, phenogroup IV included *atx1* (H3K4 histone methyltransferase) and *cmt3* (unable to maintain siRNA-independent CHG methylation) (Table 1).

With these findings in hand, plus additional considerations about (a) the timing of different mutants to reach phenological growth stage 3.5 and the necessity of having all plants ready at the same time for the next passage and (b) the phenotype of certain mutants that made it hard to unequivocally distinguish symptoms of infection, we decided to use three mutants involved in siRNA-dependent RdDM (*dcl2 dcl3 dcl4*, *ddm1*, and *polV* from phenogroups III, I, and I, respectively) and *jmj14* (from phenogroup I) involved in histone demethylation, as hosts for the evolution experiments (Supplementary Figure 2). WT plants were included to set the baseline of virus evolution in a novel host with fully operational epigenetic pathways. Notice that the WT response to infection was similar to that of Phenogroup III.

## Evolution of disease-related traits and viral load

The evolution of the four disease-related traits evaluated for each lineage and host genotype is shown in Figure 1. Though time series are noisy, in all cases a clear increasing trend has been observed (i.e., infections became more virulent), with the exception of lineage *dcl2 dcl3 dcl4*/L3 that got extinct at passage eight. The data were fitted to equation 1 (Supplementary Table 2). All factors in the model, except lineage, were highly significant (in all cases $p \leq .005$), though the magnitude of the effects varied widely. Net differences among plant genotypes showed the smallest effect ($\eta^2_P = 3.1\%$), while the largest effect was associated with differences among evolutionary passages ($\eta^2_P = 51.5\%$). Interestingly, the differences among viral lineages evolved in different plant genotypes varied in a moderate extent along passages ($\eta^2_P = 6.1\%$). Diversification among viral lineages evolved in the same plant host genotype follows divergent temporal trajectories ($\eta^2_P = 19.0\%$).

The last row in Figure 1 shows the evolution of viral load. Overall, it shows significant change along the evolution experiment, with the exception of the above-mentioned extinct lineage *dcl2 dcl3 dcl4*/L3 and lineage *polV*/L2, which shows a net decrease. Fitting log-viral load data to an ANCOVA similar to equation 1 (Supplementary Table 3) identifies the same significant factor than for the disease-related traits (all $p < .001$): Plant genotype had a very large effect ($\eta^2_P = 77.7\%$), log-viral load changes along evolution passages also in a large magnitude ($\eta^2_P = 27.9\%$), and the magnitude of the observed passage-effect depended on the plant genotype ($\eta^2_P = 14.9\%$), with lineages evolved in a particular plant genotype also differing in their temporal dynamics ($\eta^2_P = 55.0\%$). As for the disease-related traits, no net differences exist among lineages evolved in the same plant genotype.

The evolutionary trajectories of the three disease-related traits show a good degree of parallelism, with a sudden reduction after the first passage that is rapidly recovered and further increased after four passages. Thereafter, the rate of evolution slows down to a plateau value. In qualitative terms, this dynamic profile is equivalent for all the host genotype although quantitative differences exist (see below). Such pattern of phenotypic evolution has commonly been observed in evolution experiments with microorganisms (Ebert, 1998; Elena & Lenski, 2003), and in particular in this experimental pathosystem (Navarro et al., 2022) and can be explained by lineages reaching a (local) maximum in a phenotypic landscape. Interestingly, this observation suggests the existence of a maximum threshold value for the virulence that can be achieved by TuMV in this host. By contrast, neither infectivity nor viral load experienced such a biphasic shape. Infectivity remained more or less constant and close to the maximum value along the evolution experiment, suggesting either (a) no genetic variation in the viral population exists that controls this trait or (b) in our infection-saturated regime, the effect of genetic variation affecting infectivity might simply not be detectable. Viral load experienced a significant reduction after the first passage in all host genotypes and lineages to remain systematically lower than the ancestral strains until the end of the experiment, with minor yet significant increases for all lineages evolved in *jmj14*. Nonetheless, as shown in the insets of Figure 2, an overall net increase from the third passage onwards.

Recent studies have linked the evolution of virulence with epigenetic regulation of the pathogen's genome (Gómez-Díaz et al., 2012; Kasuga & Gijzen, 2013). Epigenetic regulation of microbial virulence factors would allow pathogens to quickly accommodate their gene expression to new host species or genotypes. This mechanism would be of interest for pathogens with complex genomes such as protozoa, fungi, bacteria, or even large DNA viruses that alternate latent with lytic replication cycles (e.g., herpes simplex [Singh & Tscharke, 2020] or Epstein-Barr [Murata et al., 2021]). However, direct epigenetic regulation of lytic RNA viruses' gene expression seems a less efficient mechanism. Instead, RNA viruses manipulate the epigenetic marks of their hosts to improve their own replication (Corrêa et al., 2020; Diezma-Navas et al., 2019; Leone et al., 2020; Marazzi et al., 2012). However, at this point, it is important to reemphasize that our experimental design does not allow us to distinguish between three possible hypotheses to explain virus adaptation to the epigenetic mutant hosts: (a) epigenetic host factors directly interact with TuMV factors, (b) epigenetic factors alter the expression of other plant resistance genes that act on TuMV factors, and (c) given the large-scale effects of epigenetic regulations, epigenetic mutants present a dysfunctional physiology that itself drives virus evolution. Therefore, being conservative, we can only state that epigenetic regulation affects viral evolution, without going into the possible mechanisms by which this occurs.

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

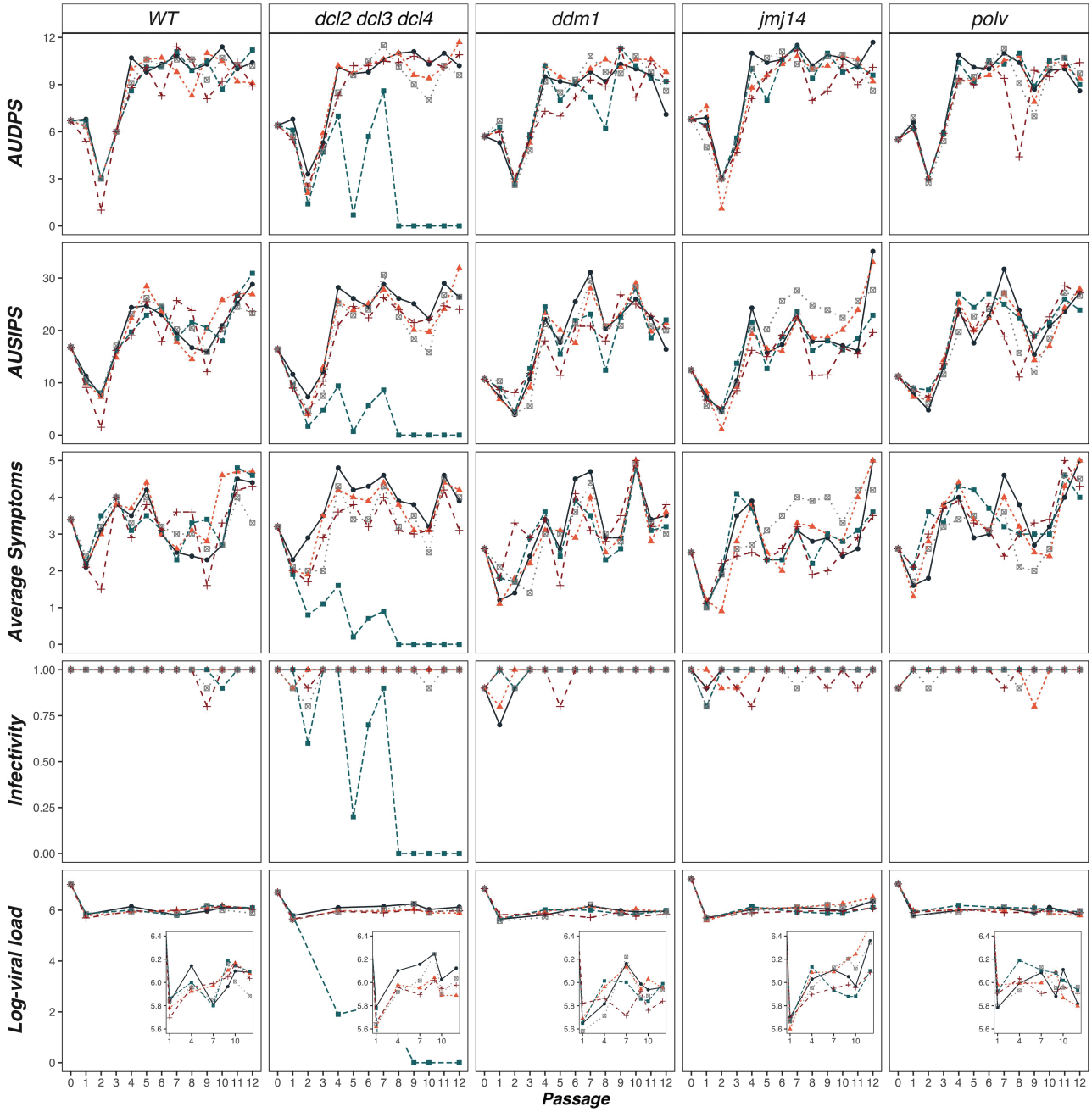



**Figure 1.** Time series of the disease-related traits along the evolution experiment for five lineages per plant genotype. Lineages are represented by different colors and symbols. First row: *AUDPS*. Second row: *AUSIPS*. Third row: average symptoms severity measured 14 dpi. Fourth row: infectivity measured at the end of each evolutionary passage (14 dpi). Fifth row: log-viral load (genomes per ng of total RNA); insets zoom-in from passages 1–12 to better appreciate the dynamics.

## Evolution of the relationship between TuMV accumulation and virulence

Models of virulence evolution typically assume that pathogens' replication within the host is what causes disease severity (Lenski & May, 1994) and hence a positive association must exist between virulence and accumulation. However, experimental evidence to support this assumption is rather limited for plant-virus pathosystems (Agudelo-Romero et al., 2008; Doumayrou et al., 2012; Pagán et al., 2007). To test this hypothesis, we examined the relationship between log-viral load and the four virulence-related traits. Partial correlation coefficients (controlling for plant genotypes) between log-viral load and the four disease-related traits were computed at different time points during the course of evolution (Supplementary Figure 3). In summary, at the beginning of the experiment, viral accumulation was correlated with infectivity ($r_p$ = .720, 22 df, $p$ < .001) and *AUDPS* ($r_p$ = .698, 22 df, $p$ < .001), but not with average symptoms severity ($r_p$ = .145, 22 df, $p$ = .499) nor *AUSIPS* ($r_p$ = .207, 22 df, $p$ = .331). In other words, for the ancestral TuMV isolate, viral accumulation was associated with traits related to the appearance of symptoms, but not with their severity. After 12 passages of evolution, however, all four disease-related traits were strongly associated with viral load: infectivity ($r_p$ = .990, 22

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

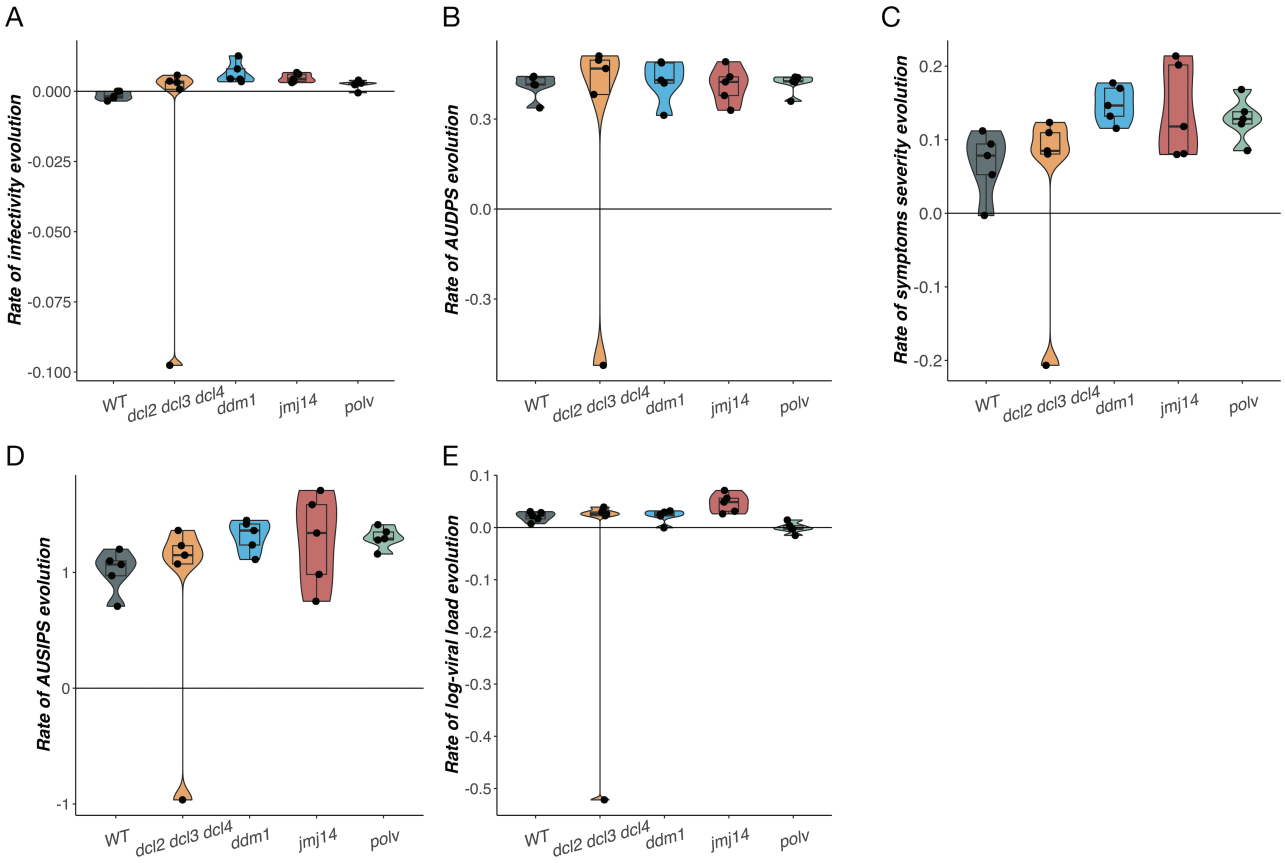



**Figure 2.** Rates of evolution for the five disease-related traits measured. Mutants *dcl2 dcl3 dcl4* and *polv* affect RdDM, mutant *ddm1* affects sRNA-independent DNA methylation, while mutant *jmj14* affects histone methylation pathway. (A) Rate of infectivity evolution, (B) rate of *AUDPS* evolution, (C) rate of severity of symptoms evolution, (D) rate of *AUSIPS* evolution, and (E) rate of log-average viral load evolution. Each lineage is represented by a point; extinct lineage *dcl2 dcl3 dcl4*/L3 is also included.

df, $p < .001$; representing 37.6% increase in the strength of the association), *AUDPS* ($r_p = .885$, 22 df, $p < .001$; 26.8% stronger association), symptoms severity ($r_p = .813$, 22 df, $p < .001$; 460.0% stronger), and *AUSIPS* ($r_p = .785$, 22 df, $p < .001$; 279.0% stronger).

Most evolution experiments with pathogenic microorganisms show that a selective advantage exists in serial passage experiments for the parasite strain with the highest numerical representation in the transferred inoculum. Within-host host competition and selection for a faster parasite growth rate could be the driving mechanism (Ebert, 1998). Our results are relevant in two aspects to this question. First, we found that virus accumulation was associated with increased transmissibility (i.e., infectivity and *AUDPS*) and that the association became stronger as the lineages adapted to their local hosts. Second, the observation that a significant association between virus accumulation and virulence-related traits (symptoms severity and *AUSIPS*) was built as viral lineages adapted to their local host genotypes suggests that virulence may result in enhanced TuMV replication in the absence of epigenetically regulated antiviral responses. Interestingly, evolution experiments with other potyviruses, tobacco etch virus (TEV), in different natural accessions of *A. thaliana* with allelic variation in the *Restricted TEV Movement* (*RTM*) gene, and therefore in their susceptibility to TEV, found significant increases in virus accumulation without changes in virulence (measured as reduction in weight due to infection) (Hillung et al., 2014). This discrepancy between two closely related viruses and different genetic variants of the same host suggests that the association between viral load and virulence shall depend on the specific combination of virus and host genotypes. In particular, virulence would not depend on viral load if (a) the extent of damage is not proportional to the number of viral particles, as in the case of hypersensitive responses (Morel & Dangl, 1997), (b) expressing resistance pathways is costly (Heidel et al., 2004; Traw et al., 2007), (c) allocating resources to defenses detracts from vegetative growth or reproductive effort (Heil, 2001; Pagán et al., 2008), or (d) the relationship between virus growth rate and virulence is non-monotonic and growth rate itself depends on the metabolic efficiency of the replication process (Lindsay et al., 2023). Ultimately, while there is often a relationship between viral load and symptom severity, it is just one piece of the puzzle and does not tell the whole history. The interplay of various factors, including the individual's immune response, the characteristics of the virus, and other physiological status-related factors (e.g., other biotic and abiotic stresses), all contribute to the overall picture in highly complex and unpredictable manners.

## Host epigenetic pathways determine the rates of TuMV phenotypic evolution

To further evaluate the role of mutations in epigenetic pathways in virus evolution, we calculated the rates of phenotypic evolution by fitting the time-series data shown in Figure 1 to equation 2.

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

Using the rates of evolutionary change estimated for each lineage as independent replicates, we fitted equation 3. The average rates of phenotypic evolution are displayed in Figure 2 and the corresponding MANOVA in Supplementary Table 4. First, significant overall differences between epigenetic pathways exist of very large magnitude ($\eta^2_P = 56.5\%$). Lineages evolved in WT plants typically, but not always, exhibit the slowest rates of phenotypic evolution. On average, viral lineages evolved in the *ddm1* RdDM mutant and the histone modification mutant *jmj14* exhibit faster rates of phenotypic change than those in evolved the other mutant genotypes (Figure 2). Furthermore, lineages evolved in *jmj14* show the fastest rate of evolution of viral load. Hence, histone modifications seem to exert a stronger selective pressure upon evolving viral populations than changes in DNA methylation profiles in terms of viral replication and accumulation. Second, mutants within pathways show only marginally significant differences in rates of phenotypic evolution, despite being of large magnitude ($\eta^2_P = 37.8\%$). Regardless of the disease phenotype being assessed, viral lineages evolved in *ddm1*, which exhibit compromised siRNA-independent methylation, exhibit the quickest rates of phenotypic change, followed by those evolved in the RdDM mutants *polv* and *dcl2 dcl3 dcl4* (Figure 2). Notably, the rates of evolution of the five traits are strongly and positively correlated (partial correlation controlling for plant genotype: $r_p \geq .750$, 22 df, $p < .001$).

Previous studies with TEV and TuMV have shown that the rate of evolution inversely correlates with the degree of susceptibility to infection, with faster phenotypic evolution taking place in less susceptible hosts (González et al., 2019; Hillung et al., 2014; Navarro et al., 2022). These previous studies included *A. thaliana* genotypes that differed in mutations affecting specific disease signaling pathways or well-known resistance genes. Here we have extended these observations to the case of mutations in essential components of epigenetically mediated regulatory pathways. On the one side, our finding that mutations in the H3K4 histone demethylase JMJ14 are associated with the fastest rates of evolution for virus replication and accumulation, but not in symptoms severity, is consistent with the fact that the *jmj14* mutant shows increased H3K4me3 and activation of JMJ14 targets, thus resulting in enhanced expression of resistance genes (Corrêa et al., 2020). On the other side, *dcl2 dcl3 dcl4* plants show compromised DCL-dependent RdDM and RNAi pathways, being unable to mount an efficient siRNA-mediated antiviral response and becoming highly permissive to viral infection. As a consequence of weaker selection, rates of TuMV evolution measured in these plants are the slowest among all host genotypes tested. Notice, however, that this conclusion is mostly driven by the extinction of lineage *dcl2 dcl3 dcl4*/L3. If this lineage is not included in the calculations, *dcl2 dcl3 dcl4* is no longer the most permissive genotype, but *polv* occupies that position, with *dcl2 dcl3 dcl4* now being the second most permissive.

In agreement with the results described here for a plant pathosystem, Mongelli et al. (2022) have done evolution experiments with drosophila C virus (DCV) infecting eight *Drosophila melanogaster* genotypes that carried mutations affecting innate immunity antiviral defense pathways. First, they also found that the extent and evolutionary dynamics of viral accumulation and infectivity depended on the particular host genotype, with a clear tendency to increase viral load in immunity-deficient flies. Second, in agreement with our observations, DCV results also highlight pleiotropy (viral genotype-by-host genotype interaction) as a major determinant of viral evolution. Indeed, they clearly show that the fitness effect of mutations that already existed in the standing genetic variation of the stock DCV population was strongly dependent on both, the fly's genetic background (the same mutation being beneficial or deleterious in different backgrounds) and the virus' genetic background (epistatic interactions with other mutations in the same haplotype).

## Analysis of cross-infection data

The gene-for-gene and matching-allele models represent the two ends of a continuum of possible outcomes of host–pathogen interaction (Agrawal & Lively, 2002). With a pure gene-for-gene interaction, the susceptible host types are expected to disappear and the resistant types are expected to dominate the population. Vice versa, the most virulent virus allele would become fixed at the cost of milder ones. However, constitutive activation of defenses is known to be costly for *A. thaliana* (e.g., SA-related defense responses pay a fitness cost in the absence of pathogens; Traw et al., 2007) and high virulence usually comes with a cost in terms of pathogen's transmission (Acevedo et al., 2019). Hence, a pure gene-for-gene strategy seems unlikely to be achieved. By contrast, with a pure matching-allele interaction, negative frequency-dependent selection emerges, such that rare *A. thaliana* resistance alleles have advantage and, as a result, a genetic polymorphism shall be maintained (Brown & Tellier, 2011; Schmid-Hempel, 2013). Symptom progression was evaluated for each one of the 24 evolved TuMV lineages across all five host genotypes. To test the extent of the specificity of TuMV adaptation to their local host genotypes, we fitted the cross-infection data to Equation 4. Focusing on the analyses of the interplant factors (Table 2), a first highly significant effect of large magnitude ($\eta^2_P = 28.4\%$) has been observed among different local host genotypes (*LH* in equation 4), supporting the idea that mutations affecting different epigenetic regulatory pathways determine the evolutionary fate of viruses. A second highly significant effect, yet of moderate magnitude ($\eta^2_P = 12.7\%$), has been observed for the genotype in which viral lineages have been tested (*TH* in equation 4). This describes the effect of differences among plant genotypes irrespective of the infecting viral lineage and its past evolutionary history. The interaction between these two main factors (*TH* × *LH* in equation 4) is also highly significant, suggesting that disease progression associated with the infection of a virus evolved in a given local host depends on the test host genotype; however, the magnitude of this effect is small ($\eta^2_P = 7.6\%$). A fourth highly significant effect, though of small magnitude ($\eta^2_P = 5.9\%$), has been observed among viral lineages evolved in the same local host genotype (*L*(*LH*) in equation 4), which suggests that differences among lineages that have evolved in the same host genotype are minor compared to differences among lineages evolved in different local hosts. In this sense, these significant effects confirm that infection with some viral lineages generally results in more severe disease in a local host-dependent manner. Finally, the highly significant and of large magnitude ($\eta^2_P = 28.2\%$) effect associated with the viral lineage by host genotype interaction (*TH* × *L*(*LH*) in equation 4) confirms that the outcome of infection ultimately depends on the particular combination of host and viral genotypes. The analyses of the intraplant factors drive to identical conclusions and confirm that the severity of symptoms increases along time (*t* in equation 4; $\eta^2_P = 95.5\%$).

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

**Table 2.** Results of fitting the severity of symptoms progression data to the repeated measures ANOVA described in equation (4).

| Effect | *SS*[a] | df | *F* | *p* | $\eta_P^2$ | 1–*β* |
|---|---|---|---|---|---|---|
| Intraplant tests[b] | | | | | | |
| dpi (*t*) | 16,642.836 | 5.276 | 8169.327 | <.001 | 0.883 | 1 |
| dpi by local host ($t \times LH$) | 296.353 | 21.102 | 36.367 | <.001 | 0.119 | 1 |
| dpi by test host ($t \times TH$) | 125.540 | 21.102 | 15.406 | <.001 | 0.054 | 1 |
| dpi by local host by test host ($t \times LH \times TH$) | 161.892 | 84.407 | 4.967 | <.001 | 0.069 | 1 |
| dpi by lineage within local host ($t \times L(LH)$) | 97.434 | 105.509 | 2.391 | <.001 | 0.042 | 1 |
| dpi by test host by lineage within local host ($t \times TH \times L(LH)$) | 536.702 | 395.659 | 3.513 | <.001 | 0.196 | 1 |
| Error within dpi | 2,200.213 | 5,697.490 | | | | |
| Interplant tests | | | | | | |
| Intersection ($\Sigma$) | 2,0676.442 | 1 | 22,744.816 | <.001 | 0.955 | 1 |
| Local host (*LH*) | 388.727 | 4 | 106.903 | <.001 | 0.284 | 1 |
| Test host (*TH*) | 142.222 | 4 | 39.112 | <.001 | 0.127 | 1 |
| Local host by test host ($LH \times TH$) | 80.348 | 16 | 5.524 | <.001 | 0.076 | 1 |
| Lineage within local host ($L(LH)$) | 61.478 | 20 | 3.381 | <.001 | 0.059 | 1 |
| Test host by lineage within local host ($TH \times L(LH)$) | 385.996 | 75 | 5.662 | <.001 | 0.282 | 1 |
| Error | 981.787 | 1,080 | | | | |

Note: [a]Type III sum of squares.
[b]Greenhouse–Geisser correction for data sphericity ($\varepsilon = 0.3768$).

These results align with those by González et al. (2019), Hillung et al. (2014), and Navarro et al. (2022), all pointing toward the evolution of specialist and generalist viral lineages depending on whether they evolved in more permissive and resistant host genotypes (significant $TH \times LH$ and $TH \times L(LH)$ in equation 4). Indeed, these studies have also shown that more permissive hosts *dcl2 dcl3 dcl4* were selected for more specialized viruses, while more restrictive hosts *jmj14* were selected for more generalist viruses, hence matching the expectation from the gene-for-gene model. It is interesting to note that in all these *in planta* evolution experiments, as mentioned above, the manual transmission of a large inoculum minimizes, if not completely eliminates, the trade-off between transmission and virulence, so that virulence can evolve without the constraint imposed by a need to maintain an active host as a source of inoculum. Therefore, observation of the evolution of a gene-for-gene system should be based primarily, if not solely, on the cost of expression of plant antiviral defense systems and not on selection on viral populations.

## Genomic changes in evolved TuMV lineages

Figure 3A shows the change in viral genetic diversity within lineages, measured as *APD*, experienced by each lineage in the course of evolution. In all cases, a significant increase has been observed ($F_{1,19} = 165.788$, $p < .001$), being on average 30.2% ±2.9 larger after 12 passages. More interestingly, significant differences in the amount of genetic diversity accumulated in the lineages exist among plant genotypes ($F_{4,19} = 3.369$, $p = .030$). Indeed, the magnitude of the differences among lineages evolved in different plant genotypes depending on the passage being characterized ($F_{4,19} = 3.884$, $p = .018$). The average magnitude of *APD* increases ranged from the 20.0% ±4.8 observed for lineages evolved in *ddm1* to the 44.7% ±5.0 observed among lineages evolved in *dcl2 dcl3 dcl4*, with lineages evolved in WT plants showing the second smaller increase in genetic variability.

Figure 3B summarizes the distribution of genetic diversity along the genomes of the evolved lineages. For the shake of a better visualization, only single-nucleotide variants with an allele frequency > 10% are shown, although all variants identified (Supplementary Figure 3) would be used in the analyses below. The total number of segregating sites after passage 1 was 518, rising up to 1,206 (2.33-fold increase) by passage 12. Interestingly, most of the segregating sites at both times were singletons (425 after passage 1 and 1,038 after passage 12). Treating each independent lineage as an observation and each host genotype as a subpopulation, the average nucleotide diversity within-host genotypes observed was $\pi_S(1) = (3.869 \pm 0.426) \times 10^{-3}$, increasing to $\pi_S(12) = (8.173 \pm 0.060) \times 10^{-3}$ (2.11 times more variability). The nucleotide diversity for the entire sample was $\pi_T(1) = (3.879 \pm 0.409) \times 10^{-3}$ and increased to $\pi_T(12) = (8.639 \pm 0.637) \times 10^{-3}$ (2.23 times more). Hence, the estimated interhost genotype nucleotide diversities at both time points were $\delta_{ST}(1) = (0.965 \pm 4.137) \times 10^{-5}$ and $\delta_{ST}(12) = (4.658 \pm 0.870) \times 10^{-4}$, which represents a 48.3-fold increase in differences among viral lineages evolved in different host genotypes. These figures can be combined to compute the coefficient of nucleotide differentiation (Nei, 1982) among host genotypes at the beginning of the experiment ($N_{ST}(1) = [0.249 \pm 1.054] \times 10^{-2}$, $z = 0.236$, $p = .407$) and at the end ($N_{ST}(12) = [5.392 \pm 0.087] \times 10^{-2}$, $z = 61.977$, $p < .001$). Hence, we conclude that genetically homogeneous founding viral populations had generated significant differentiation after 12 passages of evolution in different host genotypes (significant difference between both $N_{ST}$ values: $z = 4.862$, $p < .001$).

## Assessing the role of selection upon TuMV genome

To assess whether selection played a role in virus genetic differentiation among plant genotypes, we performed a Tajima's *D* test (Tajima, 1989) on the consensus sequences per sample and found it was significantly negative ($D = -2.457$, $p = .006$) after passage 1, becoming further more by the end

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

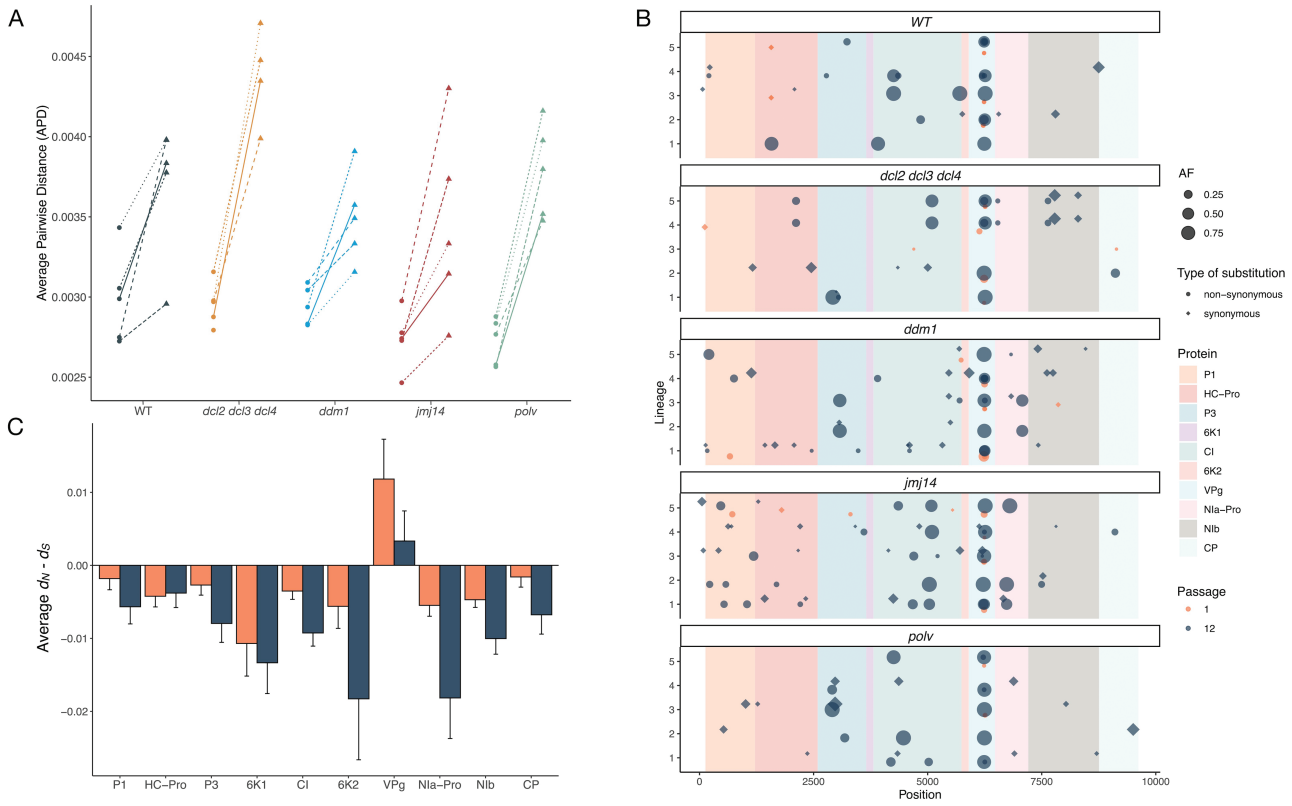



**Figure 3.** (A) Observed mean change in viral diversity between the first and last evolution passages. Diversity is measured as average pairwise *p*-distances (*APD*). Different lineages are represented by different line types. *APD* at passage 1 is represented by solid circles while solid triangles represent *APD* after passage 12. Different colors correspond to different plant genotypes. (B) Distribution of variability along TuMV genome observed for each lineage and host genotype. In blue, SNVs observed after passage 12; in orange, variants observed after passage 1. The size of the balls is proportional to allele frequency (>10% in all cases). (C) Average $d_N$–$d_S$ per cistron. Orange bars represent passage 1, and dark blue bars represent passage 12. Error bars correspond to ±1 *SD*.

of the experiment ($D = -2.567$, $p < .001$). These negative $D$ values are compatible with the action of purifying selection, the presence of segregating slightly deleterious mutations or fast population expansions (Yang, 2006). It is expected that in fast-expanding populations, many new mutations may rise in frequency, thus being observed as singletons in each lineage. Singletons inflate the number of segregating sites and thus cause $D < 0$. Indeed, this was the case here: 82.0% of the observed variable sites were singletons after passage 1, and 86.1% after passage 12, thus the observed pattern of molecular diversity among lineages evolved in the same and in different host genotypes was compatible with the fixation of mutations during the exponential growth experienced by viral populations during the systemic colonization of plants after each passage.

The three above-mentioned explanations for a negative $D$ value are not mutually exclusive; hence, room for natural selection still exists. Therefore, we sought to evaluate the role of selection at the cistron level. To do so, we computed the average $d_N$–$d_S$ for each cistron in the main ORF at both passages (Figure 3C). Overall, nine cistrons were, on average, under purifying selection, which became more intense by the end of the evolution experiment. This is particularly relevant for NIb (Supplementary Figure 4), which shows two highly variable regions in passage 1, one of which is systematically homogenized, in all host genotypes except in *polv*, by the end of the experiment. The only cistron that showed signatures of positive diversifying selection was *VPg*, though the strength of this selection also varied along the evolution experiment, being stronger at the beginning and weaker by the end. To further expand this observation, we performed a FEL codon-based selection analysis (Figure 4A). Codons encoding for G3, Y25, and P78 have been found to be under significant purifying selection. The mutation at Y25 was observed at the two sequenced time points, while the other two mutations were seen only in passage 12. Nine nonsynonymous mutations have been diagnosed as under significant positive diversifying selection. In codon 79, mutations L79F and L79I were observed in different lineages, regardless of the host genotype, after the first passage. However, by the end of the experiment, only the ancestral allele L was retained in all 24 lineages. The other eight nonsynonymous mutations all appeared in a specific domain of VPg that encompasses amino acids 110–120 (Figure 4B). All these polymorphisms show a remarkably similar pattern: most lineages show alternative alleles after the first passage of evolution. However, after the last passage, many of these mutations had reverted to the corresponding ancestral alleles. Let us take codon 115 as an illustrative example (Figure 4A), in all cases, after passage 1, all lineages showed different new alleles (e.g., N115A [in 4 lineages], N115D [1], N115G [12], N115L [1], N115P [2], N115Q [1], and N115R [4]). However, by the end of the experiment, the ancestral allele N was fixed back in a majority of lineages (13), while the rest shared some new alleles (*e.g.*, N115D [*ddm1*/L1 and WT/L1], N115G [*ddm1* lineages L2 and L3], and N115S [*jmj14* lineages L1 and L4,

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

| | | Passage | | Mutated codons (*P*-value) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 12 | G3 (0.016) | Y25 (0.021) | P78 (0.028) | L79 (0.039) | G100 (0.005) | D110 (0.004) | L112 (0.030) | D113 (0.026) | P114 (0.010) | N115 (0.006) | R118 (0.025) | N120 (0.046) |
| *dcl2 dcl3 dcl4* | *L1* | • | | G | Y | P | F | N | N | L | S | L | R | N | D |
| | | | • | G | Y | P | L | G | D | L | Y | L | N | R | N |
| | *L2* | • | | G | Y | P | F | N | D | F | S | L | Q | Y | A |
| | | | • | G | Y | P | L | G | D | L | G | P | N | R | N |
| | *L3* | • | | G | Y | P | F | N | N | L | S | Q | G | N | V |
| | | | •† | | | | | | | | | | | | |
| | *L4* | • | | G | Y | P | I | G | D | F | T | Q | G | N | D |
| | | | • | G | Y | P | L | G | D | L | D | P | N | R | N |
| | *L5* | • | | G | Y | P | F | D | D | L | S | I | G | Y | A |
| | | | • | G | Y | P | L | G | D | L | D | Q | N | R | N |
| *ddm1* | *L1* | • | | G | Y | P | L | G | D | L | S | L | G | Y | D |
| | | | • | G | Y | P | L | G | D | L | D | Q | D | C | N |
| | *L2* | • | | G | Y | P | F | N | I | F | S | L | A | Y | D |
| | | | • | G | Y | P | L | G | D | L | D | P | G | H | N |
| | *L3* | • | | G | Y | P | I | S | D | L | S | L | G | C | D |
| | | | • | G | Y | P | L | G | D | L | D | P | G | H | D |
| | *L4* | • | | G | Y | P | I | S | D | L | S | L | L | Y | D |
| | | | • | G | Y | P | L | G | D | L | G | P | N | H | N |
| | *L5* | • | | G | Y | P | F | G | V | A | T | L | A | Y | D |
| | | | • | G | Y | P | L | G | D | L | N | P | N | R | N |
| *jmj14* | *L1* | • | | G | Y | P | F | I | D | L | S | Q | G | Y | D |
| | | | • | G | Y | P | L | G | D | L | D | P | S | R | N |
| | *L2* | • | | G | Y | P | F | N | D | L | S | Q | P | Y | D |
| | | | • | G | Y | P | L | G | D | L | D | P | N | R | N |
| | *L3* | • | | G | Y | P | F | D | N | L | S | L | G | Y | D |
| | | | • | G | Y | P | L | R | D | L | Y | P | N | R | N |
| | *L4* | • | | G | Y | P | L | N | D | L | S | I | G | N | D |
| | | | • | G | Y | P | L | G | D | L | D | P | S | R | N |
| | *L5* | • | | G | Y | P | F | N | D | M | S | L | D | C | D |
| | | | • | G | Y | P | L | S | D | L | D | P | N | R | N |
| *polv* | *L1* | • | | G | Y | P | F | N | D | L | S | L | P | Y | D |
| | | | • | G | Y | P | L | G | D | L | G | Q | N | L | N |
| | *L2* | • | | G | Y | P | F | N | D | L | T | L | R | Y | D |
| | | | • | G | Y | P | L | V | D | L | D | P | S | R | N |
| | *L3* | • | | G | Y | P | L | F | G | F | S | L | G | Y | D |
| | | | • | G | Y | P | L | G | D | L | D | P | S | R | N |
| | *L4* | • | | G | Y | P | I | N | V | L | S | L | G | Y | D |
| | | | • | G | Y | P | L | G | D | L | G | P | S | L | N |
| | *L5* | • | | G | Y | P | F | I | V | L | S | L | A | Y | D |
| | | | • | G | Y | P | L | G | D | L | N | P | N | R | N |
| *WT* | *L1* | • | | G | Y | P | F | S | E | L | S | Q | G | N | D |
| | | | • | G | Y | P | L | G | D | L | S | P | D | R | N |
| | *L2* | • | | G | Y | P | I | I | V | L | S | L | G | N | D |
| | | | • | G | Y | P | L | G | D | L | G | P | N | C | N |
| | *L3* | • | | G | Y | P | F | N | E | L | T | L | R | Y | A |
| | | | • | G | Y | P | L | G | D | L | D | P | S | R | N |
| | *L4* | • | | G | Y | P | F | N | E | L | S | L | A | Y | V |
| | | | • | G | Y | P | L | S | D | L | D | P | S | R | N |
| | *L5* | • | | G | Y | P | L | N | E | L | T | L | R | C | A |
| | | | • | G | Y | P | L | G | D | L | G | L | N | R | N |

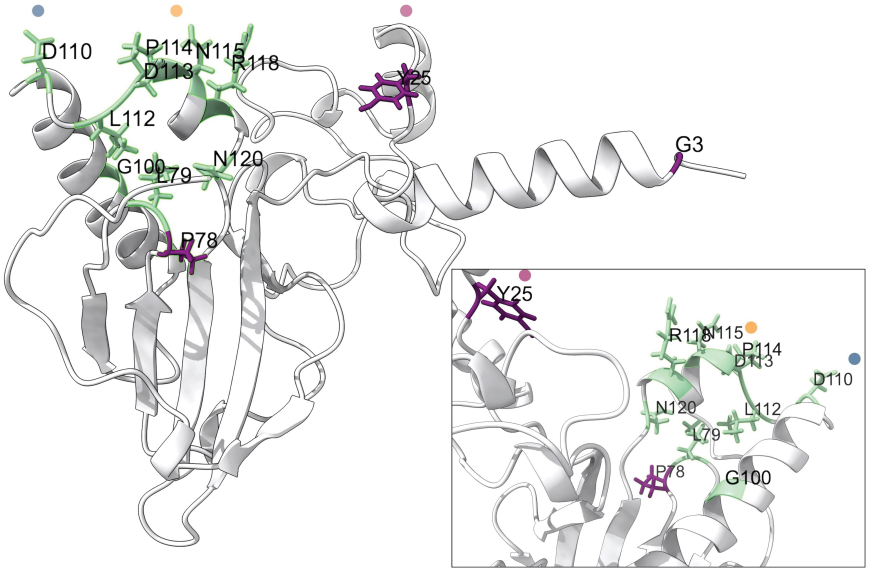



**Figure 4.** (A) Results of the FEL analysis showing positions at the *VPg* cistron under significant negative stabilizing selection (purple) and positive diversifying selection (green). Rows correspond to the different lineages are paired by sequenced passages. Each cell contains the observed amino acid. The header shows the reference amino acid, and the probability of accepting the neutral evolution hypothesis provided by FEL. (B) VPg structure predicted with AlphaFold. The position of the amino acids under selection is shown with the same colors as in (A) on the protein structure. The inset shows the domain encompassing amino acids 110–120. The blue, yellow, and purple dots have been added to facilitate orientation between the two views.

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

*polv* lineages L2, L3 and L4, and WT lineages L3 and L4]). Together, this observation of a transient rise in the frequency of novel alleles followed by a reversal to ancestral ones is in agreement with the expectation of pervasive leap-frog events (Gerrish & Lenski, 1998). According to this effect, the majority of genotypes at some time point are less closely related to the immediately preceding majority genotype than to an earlier one. Given that reversion mutations are unlikely, the permanence of ancestral alleles at very low frequency right after the switch from WT to mutant host genotypes seems a more parsimonious explanation for the leap-frogs.

No obvious association between the alleles fixed at the end of the evolution experiment and host genotypes exists, except perhaps for allele R118H (keeping a positively charged side chain and neutral functional effect), which has been found only in three lineages evolved in *ddm1*. Interestingly, this mutation was also observed by Navarro et al. (2022) associated with TuMV lineages evolved in jasmonate-insensitive mutant plants, showing an enhanced systemic acquired resistance.

*VPg* has been commonly found as a target of selection in evolution experiments of potyviruses in *A. thaliana* (Agudelo-Romero et al., 2008; Gallois et al., 2010; González et al., 2019, 2021; Hillung et al., 2014; Melero et al., 2023; Navarro et al., 2022). VPg plays many essential roles in genome transcription (it is linked to the 5′-end of the viral genome and provides the hydroxyl group that primes the synthesis of the complementary strains by the viral RdRp; Murphy et al., 1996), translation (directly interacts with the eukaryotic initiation factors eIF(iso)4E and eIF(iso)4G; Léonard et al., 2000), and interacts with all other viral proteins and some of the host cell proteins (Martínez et al., 2023). In addition, the role of VPg as a target of selection in crops carrying resistance genes has been amply demonstrated. One of the most extensively used resistance genes against potato virus Y (PVY) in commercial pepper cultivars is *pvr2*, which has many different alleles (Charron et al., 2008; Nicaise et al., 2007). The *pvr2* locus encodes for the eIF4E factor that must interact properly with VPg to ensure PVY replication in pepper. Interestingly, all the resistance-breaking viral isolates found so far contain mutations in the *VPg* cistron (Ayme et al., 2006; Duprat et al., 2002; Moury et al., 2007).

Some readers will certainly be surprised by this lack of correlation between the phenotypic changes observed and the fixation of specific mutations in each host. This, together with the observed parallelism between lineages evolved in different host genotypes, might lead one to hypothesize that all changes in viral lineages simply reflect adaptation to host species and environmental conditions. Certainly, this is a very plausible hypothesis that we cannot rule out with the present data. How can we reconcile our interpretation of the phenotypic changes being associated with specific selective constraints imposed by differences in epigenetic regulations and the lack of fixation of host-specific mutations in the viral lineages (except those evolved in *ddm1*)? One possibility is the presence of a plethora of mutations at very low frequency (below the LowFreq detection threshold), part of the quasispecies' mutant spectrum, and whose collective action determines the observed phenotypes, as previously described for other viral systems (Chumakov et al., 1991; Pfeiffer & Kirkegaard, 2006; Sanz-Ramos et al., 2008; Vignuzzi et al., 2006). In other words, the complexity of the viral quasispecies determines phenotypes that cannot be linked to consensus sequences.

## Conclusions

We have shown that epigenetic regulatory pathways not only have an effect on the antiviral responses of hosts and that viruses interfere with such defenses but also that epigenetic regulation impacts the evolution of RNA viruses, either by directly interacting with host factors, by modulating the expression of other resistance genes, or by inducing a fundamentally different cellular environment compared with WT plants. Previous studies have shown that mutations in resistance/susceptibility recessive genes and dominant genes involved in defense responses to infection modulate the evolution of viral populations. A common observation of these experiments is that more permissive hosts select for less virulent pathogens, while more virulent pathogens are selected in more resistant hosts. Here we extend these previous results to the case of epigenetic regulation of defense genes. In particular, we describe how mutations in the two main epigenetic pathways (i.e., DNA methylation and histone modification) condition the evolution of an emerging plant RNA virus. Parallelizing previous results, more resistant genotypes, in this particular case a mutant affecting histone modification, were selected for more virulent viruses, while a mutant unable to produce siRNAs and thus particularly sensitive to infection, was selected for less virulent viral strains. Likewise, we confirm that VPg is a main target of selection also in the context of a potyviral interplay with the host epigenetic regulation, which suggests a possible role of this multifunctional viral protein in manipulating unknown aspects of the host response to infection. In a follow-up article, we will tackle this issue by deeply characterizing host genotype-specific transcriptomic responses to infection with the ancestral and evolved TuMV lineages.

## Supplementary material

Supplementary material is available online at *Evolution* (https://academic.oup.com/evolut)

## Data availability

Raw Illumina RNA-seq data generated for this study are available in NCBI SRA under BioProject accession PRJNA974369. All R scripts developed in this work are available at https://github.com/MJmaolu/evolution_TuMV_in_A.thaliana_Epigenetic_Mutants/tree/main. Raw fitness and virus diversity data can be downloaded from https://git.csic.es/sfelenalab/tumv-evolution-in-mutants-epigenetic-pathways.

## Author contributions

S.F.E. conceived the study. S.A. performed the experiments. M.J.O.U., R.L.C., and S.F.E. analyzed the data. S.F.E. drafted the initial version of the manuscript. All authors contributed to later versions of the manuscript.

## Funding

This work was supported by grant PID2022-136912NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF a way of making Europe," and by grant CIPROM/2022/59 funded by Generalitat Valenciana.

*Conflict of interest*: The authors declare no conflict of interest.

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

## Acknowledgments

We thank F. de la Iglesia and P. Agudo for excellent technical support and the EvolSysVir lab members for their comments and fruitful discussions.

## References

Acevedo, M. A., Dillemuth, F. P., Flick, A. J., Faldyn, M. J., & Elderd, B. D. (2019). Virulence-driven trade-offs in disease transmission: A meta-analysis. *Evolution*, *73*(4), 636–647. https://doi.org/10.1111/evo.13692

Agrawal, A., & Lively, C. (2002). Infection genetics: Gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research*, *4*(1), 79–90.

Agudelo-Romero, P., de la Iglesia, F., & Elena, S. F. (2008). The pleiotropic cost of host-specilization in tobacco etch potyvirus. *Infection Genetics and Evolution*, *8*(6), 806–814. https://doi.org/10.1016/j.meegid.2008.07.010

Álvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U., & Avramova, Z. (2003). ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Current Biology*, *13*(8), 627–637. https://doi.org/10.1016/s0960-9822(03)00243-4

Andrews, S. (2010). *FASTQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ashe, A., Colot, V., & Olroyd, B. P. (2021). How does epigenetics influence the course of evolution? *Philosophical Transactions of the Royal Society B*, *376*, 20200111.

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadoeuf, J., Palloix, A., & Moury, B. (2006). Different mutations in the genome-linked protein VPg of potato virus Y confer virulence on the *pvr23* resistance in pepper. *Molecular Plant-Microbe Interactions*, *19*(5), 557–563. https://doi.org/10.1094/MPMI-19-0557

Banta, J. A., & Richards, C. L. (2018). Quantitative epigenetics and evolution. *Heredity*, *121*(3), 210–224. https://doi.org/10.1038/s41437-018-0114-x

Böhmdorfer, G., Sethuraman, S., Rowley, M. J., Krzyszton, M., Rothi, M. H., Bouzit, L., & Wierzbicki, A. T. (2016). Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *eLife*, *5*, e19092. https://doi.org/10.7554/eLife.19092

Bond, D. M., & Baulcombe, D. C. (2014). Small RNAs and heritable epigenetic variation in plants. *Trends in Cell Biology*, *24*(2), 100–107. https://doi.org/10.1016/j.tcb.2013.08.001

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*, *16*(12), 727–741. https://doi.org/10.1038/nrm4085

Bouché, N., Lauressergues, D., Gasciolli, V., & Vaucheret, H. (2006). An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO Journal*, *25*(14), 3347–3356. https://doi.org/10.1038/sj.emboj.7601217

Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, M. J., Hoffman, N. E., Davis, K. R., & Görlach, J. (2001). Growth stage-based phenotypic analysis of *Arabidopsis*: A model for high throughput functional genomics in plants. *Plant Cell*, *13*, 1499–1510.

Brown, J. K. M., & Tellier, A. (2011). Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, *49*, 345–367. https://doi.org/10.1146/annurev-phyto-072910-095301

Butković, A., González, R., Rivarez, M. P. S., & Elena, S. F. (2021). A genome-wide association study identifies *Arabidopsis thaliana* genes that contribute to differences in the outcome of infection with two turnip mosaic potyvirus strains that differ in their evolutionary history and degree of host specialization. *Virus Evolution*, *7*(2), veab063. https://doi.org/10.1093/ve/veab063

Cao, X., & Jacobsen, S. E. (2002). Role of the Arabidopsis *DRM* methyltransferases in de novo DNA methylation and gene silencing. *Current Biology*, *12*(13), 1138–1144. https://doi.org/10.1016/s0960-9822(02)00925-9

Carr, J. P., Lewsey, M. G., & Palukaitis, P. (2010). Signaling in induced resistance. *Advances in Virus Research*, *76*, 57–121. https://doi.org/10.1016/S0065-3527(10)76003-6

Chan, S. W. L., Henderson, I. R., Zhang, X., Shah, G., Chien, J. S. C., & Jacobsen, S. E. (2006). RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*. *PLoS Genetics*, *2*(6), e83. https://doi.org/10.1371/journal.pgen.0020083

Charron, C., Nicolaï, M., Gallois, J. L., Robaglia, C., Moury, B., Palloix, A., & Caranta, C. (2008). Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. *Plant Journal*, *54*(1), 56–68. https://doi.org/10.1111/j.1365-313X.2008.03407.x

Chen, C. C., Chao, C. H., Chen, C. C., Yeh, S. D., Tsai, H. T., & Chang, C. A. (2003). Identification of turnip mosaic virus isolates causing yellow stripe and spot on calla lily. *Plant Disease*, *87*(8), 901–905. https://doi.org/10.1094/PDIS.2003.87.8.901

Chumakov, K. M., Powers, L. B., Noonan, K. E., Roninson, I. B., & Levenbook, I. S. (1991). Correlation between amount of virus with altered nucleotide sequence and the monkey test for acceptability of oral poliovirus vaccine. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(1), 199–203. https://doi.org/10.1073/pnas.88.1.199

Corrêa, R. L., Sanz-Carbonell, A., Kogej, Z., Müller, S. Y., Ambrós, S., López-Gomollón, S., Gómez, G., Baulcombe, D. C., & Elena, S. F. (2020). Viral fitness determines the magnitude of transcriptomic and epigenomic reprograming of defense responses in plans. *Molecular Biology and Evolution*, *37*(7), 1866–1881. https://doi.org/10.1093/molbev/msaa091

Cuerda-Gil, D., & Slotkin, R. K. (2016). Non-canonical RNA-directed DNA methylation. *Nature Plants*, *2*(11), 16163. https://doi.org/10.1038/nplants.2016.163

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*, giab008.

Diezma-Navas, L., Pérez-González, A., Artaza, H., Alonso, L., Caro, E., Llave, C., & Ruiz-Ferrer, V. (2019). Crosstalk between epigenetic silencing and infection by tobacco rattle virus in *Arabidopsis*. *Molecular Plant Pathology*, *20*(10), 1439–1452. https://doi.org/10.1111/mpp.12850

Doumayrou, J., Avellan, A., Froissart, R., & Michalakis, Y. (2012). An experimental test of the transmission-virulence trade-off hypothesis in a plant virus. *Evolution*, *67*(2), 477–486. https://doi.org/10.1111/j.1558-5646.2012.01780.x

Dowen, R. H, Pelizzola, M., Schmitz, R. J., Lister, R., Dowen, J. M., Nery, J. R., Dixon, J. E., & Ecker, J. R. (2012). Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, E2183–E2191.

Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K. S., & Robaglia, C. (2002). The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant Journal*, *32*(6), 927–934. https://doi.org/10.1046/j.1365-313x.2002.01481.x

Earley, K., Lawrence, R. J., Pontes, O., Reuther, R., Enciso, A. J., Silva, M., Neves, N., Gross, M., Viegas, W., & Pikaard, C. S. (2006). Erasure of histone acetylation by Arabidopsis *HDA6* mediates large-scale gene silencing in nucleolar dominance. *Genes & Development*, *20*(10), 1283–1293. https://doi.org/10.1101/gad.1417706

Ebert, D. (1998). Experimental evolution of parasites. *Science*, *282*(5393), 1432–1435. https://doi.org/10.1126/science.282.5393.1432

Elena, S. F., & Lenski, R. E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, *4*(6), 457–469. https://doi.org/10.1038/nrg1088

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

Gallois, J. L., Charron, C., Sánchez, F., Pagny, G., Houvenaghel, M. C., Moretti, A., Ponz, F., Revers, F., Caranta, C., & German-Retana, S. (2010). Single amino acid changes in the turnip mosaic virus viral genome-linked protein (VPg) confer virulence towards *Arabidopsis thaliana* mutants knocked out for eukaryotic initiation factors eIF(iso)4E and eIF(iso)4G. *Journal of General Virology*, *91*, 288–293.

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J., & Carrington, J. C. (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*(2), 481–496. https://doi.org/10.1105/tpc.109.073056

Gerrish, P. J., & Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, *102–103*(1–6), 127–144.

Gómez-Díaz, E., Jordà, M., Peinado, M. A., & Rivero, A. (2012). Epigenetics of host-pathogen interactions: The road ahead and the road behind. *PLoS Pathogens*, *8*(11), e1003007. https://doi.org/10.1371/journal.ppat.1003007

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldán-Arjona, T., David, L., & Zhu, J. K. (2002). *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell*, *111*(6), 803–814. https://doi.org/10.1016/s0092-8674(02)01133-9

González, R., Butković, A., & Elena, S. F. (2019). Role of host genetic diversity for susceptibility-to-infection in the evolution of virulence of a plant virus. *Virus Evolution*, *5*(2), vez024. https://doi.org/10.1093/ve/vez024

González, R., Butković, A., Escaray, F. J., Martínez-Latorre, J., Melero, I., Pérez-Parets, E., Gómez-Cadenas, A., Carrasco, P., & Elena, S. F. (2021). Plant virus evolution under strong drought conditions results in a transition from parasitism to mutualism. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(6), e2020990118. https://doi.org/10.1073/pnas.2020990118

Heidel, A. J., Clarke, J. D., Antonovics, J., & Dong, X. (2004). Fitness costs of mutations affecting the systemic acquired resistance pathway in *Arabidopsis thaliana*. *Genetics*, *168*(4), 2197–2206. https://doi.org/10.1534/genetics.104.032193

Heil, M. (2001). The ecological concept of costs of induced systemic resistance (ISR). *European Journal of Plant Pathology*, *107*, 137–146.

Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*, *308*(5718), 118–120. https://doi.org/10.1126/science.1106910

Hillung, J., Cuevas, J. M., Valverde, S., & Elena, S. F. (2014). Experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. *Evolution*, *68*(9), 2467–2480. https://doi.org/10.1111/evo.12458

Hung, Y. H., & Slotkin, R. K. (2021). The initiation of RNA interference (RNAi) in plants. *Current Opinion in Plant Biology*, *61*, 102014. https://doi.org/10.1016/j.pbi.2021.102014

Jeddeloh, J. A., Stokes, T. L., & Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genetics*, *22*(1), 94–97. https://doi.org/10.1038/8803

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kachroo, P., Chandra-Shekara, A. C., & Klessig, D. F. (2006). Plant signal transduction and defense against viral pathogens. *Advances in Virus Research*, *66*, 161–191. https://doi.org/10.1016/S0065-3527(06)66004-1

Kasuga, T., & Gijzen, M. (2013). Epigenetics and the evolution of virulence. *Trends in Microbiology*, *21*(11), 575–582. https://doi.org/10.1016/j.tim.2013.09.003

Klironomos, F. D., Berg, J., & Collins, S. (2013). How epigenetic mutations can affect genetic evolution: Model and mechanisms. *Bioessays*, *35*(6), 571–578. https://doi.org/10.1002/bies.201200169

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Le, T. N., Schumann, U., Smith, N. A., Tiwari, S., Au, P. C. K., Zhu, Q. H., Taylor, J. M., Kazan, K., Llewellyn, D. J., Zhang, R., Dennis, E. S., & Wang, M. -B. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. *Genome Biology*, *15*(9), 458. https://doi.org/10.1186/s13059-014-0458-3

Lenski, R. E., & May, R. M. (1994). The evolution of virulence in parasites and pathogens: Reconciliation between two competing hypotheses. *Journal of Theoretical Biology*, *169*(3), 253–265. https://doi.org/10.1006/jtbi.1994.1146

Léonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M. G., & Laliberté, J. F. (2000). Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *Journal of Virology*, *74*(17), 7730–7737. https://doi.org/10.1128/jvi.74.17.7730-7737.2000

Leone, M., Zavallo, D., Venturuzzi, A., & Asurmendi, S. (2020). RdDM pathway components differentially modulate *Tobamovirus* symptom development. *Plant Molecular Biology*, *104*(4-5), 467–481. https://doi.org/10.1007/s11103-020-01051-6

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. http://arxiv.org/abs/1303.3997

Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., & Jacobsen, S. E. (2001). Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. *Science*, *292*(5524), 2077–2080. https://doi.org/10.1126/science.1059745

Liu, P., Cuerda-Gil, D., Shahid, S., & Slotkin, R. K. (2022). The epigenetic control of the transposable element life cycle in plant genomes and beyond. *Annual Review of Genetics*, *56*, 63–87. https://doi.org/10.1146/annurev-genet-072920-015534

Lloyd, J. P. B., & Lister, R. (2021). Epigenome plasticity in plants. *Nature Reviews Genetics*, *23*(1), 55–68. https://doi.org/10.1038/s41576-021-00407-y

Loebenstein, G. (2009). Local lesions and induced resistance. *Advances in Virus Research*, *75*, 73–117. https://doi.org/10.1016/S0065-3527(09)07503-4

López, A., Ramírez, V., García-Andrade, J., Flors, V., & Vera, P. (2011). The RNA silencing enzyme RNA polymerase V is required for plant immunity. *PLoS Genetics*, *7*(12), e1002434. https://doi.org/10.1371/journal.pgen.1002434

López Sánchez, A., Stassen, J. H. M., Furci, L., Smith, L. M., & Ton, J. (2016). The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant Journal*, *88*(3), 361–374. https://doi.org/10.1111/tpj.13252

López-Gomollon, S., & Baulcombe, D. C. (2022). Roles of RNA silencing in viral and non-viral plant immunity and in the crosstalk between disease resistance systems. *Nature Reviews Molecular Cell Biology*, *23*(10), 645–662. https://doi.org/10.1038/s41580-022-00496-5

Lu, F., Cui, X., Zhang, S., Liu, C., & Cao, X. (2010). JMJ14 is an H3K4 demethylase regulating flowering time in *Arabidopsis*. *Cell Research*, *20*(3), 387–390. https://doi.org/10.1038/cr.2010.27

Luna, E., Bruce, T. J. A., Roberts, M. R., Flors, V., & Ton, J. (2012). Next-generation systemic acquired resistance. *Plant Physiology*, *158*(2), 844–853. https://doi.org/10.1104/pp.111.187468

Lyndsay, R. J., Holder, P. J., Talbot, N. J., & Gudelj, I. (2023). Metabolic efficiency reshapes the seminal relationship between pathogen growth rate and virulence. *Ecology Letters*, *26*(6), 896–907. https://doi.org/10.1111/ele.14218

Marazzi, I., Ho, J. S. Y., Kim, J., Manicassamy, B., Dewall, S., Albrecht, R. A., Seibert, C. W., Schaefer, U., Jeffrey, K. L., Prinjha, R. K., Lee, K., García-Sastre, A., Roeder, R. G., & Tarakhovsky, A. (2012). Suppression of the antiviral response by an influenza histone mimic. *Nature*, *483*, 428–433.

Martínez, F., Carrasco, J. L., Toft, C., Hillung, J., Giménez-Santamarina, S., Yenush, L., Rodrigo, G., & Elena, S. F. (2023). A binary

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

interaction map between turnip mosaic virus and *Arabidopsis thaliana* proteomes. *Communications Biology*, *6*(1), 28. https://doi.org/10.1038/s42003-023-04427-8

McCue, A. D., Nithikattu, S., Reeder, S. H., & Slotkin, R. K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genetics*, *8*, e1002474.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Melero, I., González, R., & Elena, S. F. (2023). Host developmental stages shape the evolution of a plant RNA virus. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *378*(1873), 20220005. https://doi.org/10.1098/rstb.2022.0005

Mendizabal, I., Keller, T. E., Zeng, J., & Yi, S. V. (2014). Epigenetics and evolution. *Integrative and Comparative Biology*, *54*(1), 31–42. https://doi.org/10.1093/icb/icu040

Mongelli, V., Lequime, S., Kousathanas, A., Gausson, V., Blanc, H., Nigg, J., Quintana-Murci, L., Elena, S. F., & Saleh, M. -C. (2022). Innate immunity pathways act synergistically to constrain RNA virus evolution in *Drosophila melanogaster*. *Nature Ecology and Evolution*, *6*(5), 565–578. https://doi.org/10.1038/s41559-022-01697-z

Morel, J. B., & Dangl, J. L. (1997). The hypersenstivie response and the induction of cell death in plants. *Cell Death and Differentiation*, *4*(8), 671–683. https://doi.org/10.1038/sj.cdd.4400309

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N., Rémoué, K., Sanial, M., Vo, T. A., & Vaucheret, H. (2000). Arabidopsis *SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, *101*(5), 533–542. https://doi.org/10.1016/s0092-8674(00)80863-6

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A., & Jacquemond, M. (2007). Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Molecular Plant-Microbe Interactions*, *17*(3), 322–329. https://doi.org/10.1094/mpmi.2004.17.3.322

Murata, T., Sugimoto, A., Inagaki, T., Yanagi, Y., Watanabe, T., Sato, Y., & Kimura, H. (2021). Molecular basis of Epstein-Barr virus latency establishment and lytic reactivation. *Viruses*, *13*(12), 2344. https://doi.org/10.3390/v13122344

Murphy, J. F., Klein, P. G., Hunt, A. G., & Shaw, J. G. (1996). Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. *Virology*, *220*(2), 535–538. https://doi.org/10.1006/viro.1996.0344

Navarro, R., Ambrós, S., Butković, A., Carrasco, J. L., González, R., Martínez, F., Wu, B., & Elena, S. F. (2022). Defects in plant immunity modulate the rates and patterns of RNA virus evolution. *Virus Evolution*, *8*(2), veac059. https://doi.org/10.1093/ve/veac059

Nei, M. (1982). Evolution of human races at the gene level. In Bonné-Tamir, B. (Ed.), *Human genetics, part A: The unfolding genome* (pp. 167–181). Alan R. Liss.

Nicaise, V., Gallois, J. L., Chafiai, F., Allen, L. M., Schurdi-Levraud, V., Browning, K. S., Candresse, T., Caranta, C., Le Gall, O., & German-Retana, S. (2007). Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in *Arabidopsis thaliana*. *FEBS Letters*, *581*(5), 1041–1046. https://doi.org/10.1016/j.febslet.2007.02.007

Nuthikattu, S., McCue, A. D., Panda, K., Fultz, D., DeFraia, C., Thomas, E. N., & Slotkin, R. K. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiology*, *162*(1), 116–131. https://doi.org/10.1104/pp.113.216481

Pagán, I., Alonso-Blanco, C., & García-Arenal, F. (2007). The relationship of within-host multiplication and virulence in a plant-virus system. *PLoS One*, *2*(8), e786. https://doi.org/10.1371/journal.pone.0000786

Pagán, I., Alonso-Blanco, C., & García-Arenal, F. (2008). Host responses in life-history traits and tolerance to virus infection in *Arabidopsis thaliana*. *PLoS Pathogens*, *4*(8), e1000124. https://doi.org/10.1371/journal.ppat.1000124

Pagán, I., Fraile, A., Fernández-Fueyo, E., Montes, N., Alonso-Blanco, C., & García-Arenal, F. (2010). *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. *Philosophical Transactions of the Royal Society B*, *365*, 1938–1995.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Pfeiffer, J. L., & Kirkegaard, K. (2006). Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 5520–5525.

Pontier, D., Picart, C., Roudier, F., García, D., Lahmy, S., Azevedo, J., Alart, E., Laudié, M., Karlowski, W. M., Cooke, R., Colot, V., Voinnet, O., & Lagrange, T. (2012). NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*. *Molecular Cell*, *48*(1), 121–132. https://doi.org/10.1016/j.molcel.2012.07.027

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M. A., Lerbs-Mache, S., Colot, V., & Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes & Development*, *19*(17), 2030–2040. https://doi.org/10.1101/gad.348405

Rapp, R. A., & Wendel, J. E. (2005). Epigenetics and plant evolution. *New Phytologist*, *168*, 81–91.

Richard, C. L., Bossdorf, O., & Pigliucci, M. (2010). What role does heritable epigenetic variation play in phenotypic evolution? *BioScience*, *60*, 232–237.

Rozas, J., Ferrer-Mata, A., Sánchez-del Barrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Osins, S. E., & Sánchez-Gracia, A. (2017). DnaSP v6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.

Sanz-Ramos, M., Díaz-San Segundo, F., Escarmís, C., Domingo, E., & Sevilla, N. (2008). Hidden virulence determinants in a viral quasispecies *in vivo*. *Journal of Virology*, *82*(21), 10465–10476. https://doi.org/10.1128/JVI.00825-08

Saze, H., Shiraihi, A., Miura, A., & Kakutani, T. (2008). Control of genic DNA methylation by a jmjC domain containing protein in *Arabidopsis thaliana*. *Science*, *319*, 462–465.

Schmid-Hempel, P. (2013). Host-parasite genetics. In *Evolutionary parasitology: The integrated study of infections, immunology, ecology, and genetics* (pp. 259–272). Oxford University Press.

Searle, I. R., Pontes, O., Melnyk, C. W., Smith, L. M., & Baulcombe, D. C. (2010). a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in Arabidopsis. *Genes Develop*, *24*, 986–991.

Shao, W., Kearney, M. F., Boltz, V. F., Spindler, J. E., Mellors, J. W., Maldarelli, F., & Coffin, J. M. (2014). PAPNC, a novel method to calculate nucleotide diversity from large scale next generation sequencing data. *Journal of Virological Methods*, *203*, 73–80. https://doi.org/10.1016/j.jviromet.2014.03.008

Simko, I., & Piepho, H. (2012). The area under the disease progress stairs: Calculation, advantage, and application. *Phytopathology*, *102*(4), 381–389. https://doi.org/10.1094/PHYTO-07-11-0216

Singh, N., & Tscharke, D. C. (2020). Herpes simplex virus latency is noisier the closer we look. *Journal of Virology*, *94*(4), e01701–e01719. https://doi.org/10.1128/JVI.01701-19

Soosaar, J. L. M., Burch-Smith, T. M., & Dinesh-Kumar, S. P. (2005). Mechanisms of plant resistance to viruses. *Nature Reviews Microbiology*, *3*(10), 789–798. https://doi.org/10.1038/nrmicro1239

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

Stajic, D., & Jansen, L. E. T. (2021). Empirical evidence for epigenetic inheritance driving evolutionary adaptation. *Philosophical Transactions of the Royal Society B*, *376*, 20201421.

Stajic, D., Perfeito, L., & Jansen, L. E. T. (2019). Epigenetic gene silencing alters the mechanisms and rate of evolutionary adaptation. *Nature Ecology and Evolution*, *3*, 4981–4498.

Tajima, F. (1989). Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tamura, K., Stecher, G., & Kumar, S. (2021). Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Traw, M. B., Kniskern, J. M., & Bergelson, J. (2007). SAR increases fitness of *Arabidopsis thaliana* in presence of natural bacterial pathogens. *Evolution*, *61*(10), 2444–2449. https://doi.org/10.1111/j.1558-5646.2007.00211.x

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, *439*(7074), 344–348. https://doi.org/10.1038/nature04388

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends in Genetics*, *17*(8), 449–459. https://doi.org/10.1016/s0168-9525(01)02367-8

Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L., & Nagarajan, N. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201. https://doi.org/10.1093/nar/gks918

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analysis using only ANOVA procedures. In *Proceedings of the. ACM Conference on Human Factors in Computing Systems* (pp. 143–146). ACM Press.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E., & Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology*, *2*(5), E104. https://doi.org/10.1371/journal.pbio.0020104

Yang. Z. (2006) *Computational molecular evolution* (pp. 262–264). Oxford University Press.

Yu, A., Lepère, G., Jay, F., Wang, J., Bapaume, L., Wang, Y., Abraham, A. -L., Penterman, J., Fischer, R. L., Voinnet, O., & Navarro, L. (2013). Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), 2389–2394. https://doi.org/10.1073/pnas.1211757110

Zilberman, D., Cao, X., & Jacobsen, S. E. (2003). *ARGONAUTE4* control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, *299*(5607), 716–719. https://doi.org/10.1126/science.1079695

Zou, J. M., & Zhang, Y. (2020). Plant immunity: Danger perception and signaling. *Cell*, *181*, 978–989.

Downloaded from https://academic.oup.com/evolut/article/78/1/69/7331433 by guest on 03 March 2025

# CHAPTER 4

## Transcriptomic insights into the epigenetic modulation of turnip mosaic virus evolution in *Arabidopsis thaliana*

BMC Genomics

RESEARCH

Open Access

# Transcriptomic insights into the epigenetic modulation of turnip mosaic virus evolution in *Arabidopsis thaliana*

María J. Olmo-Uceda[1†], Silvia Ambrós[1†], Régis L. Corrêa[1,2] and Santiago F. Elena[1,3*]

## Abstract

**Background** Plant-virus interaction models propose that a virus's ability to infect a host genotype depends on the compatibility between virulence and resistance genes. Recently, we conducted an evolution experiment in which lineages of turnip mosaic virus (TuMV) were passaged in *Arabidopsis thaliana* genotypes carrying mutations in components of the DNA methylation and the histone demethylation epigenetic pathways. All evolved lineages increased infectivity, virulence and viral load in a host genotype-dependent manner.

**Results** To better understand the underlying reasons for these evolved relationships, we delved into the transcriptomic responses of mutant and WT plant genotypes in mock conditions and infected with either the ancestral or evolved viruses. Such a comparison allowed us to classify every gene into nine basic expression profiles. Regarding the targets of viral adaptation, our analyses allowed the identification of common viral targets as well as host genotype-specific genes and categories of biological processes. As expected, immune response-related genes were found to be altered upon infection. However, we also noticed the pervasive over-representation of other functional groups, suggesting that viral adaptation was not solely driven by the level of expression of plant resistance genes. In addition, a significant association between the presence of transposable elements within or upstream the differentially expressed genes was observed. Finally, integration of transcriptomic data into a virus-host protein-protein interaction network highlighted the most impactful interactions.

**Conclusions** These findings shed extra light on the complex dynamics between plants and viruses, indicating that viral infectivity depends on various factors beyond just the plant's resistance genes.

**Keywords** Adaptation, Epigenetics, Experimental evolution, Gene expression, *Potyvirus rapae*, Systems biology, Virus evolution, Virus-host interactions

## Introduction

Plants and viruses engage in complex interactions that trigger defense and counter-defense mechanisms, often leading to coevolution and reciprocal adaptation. Plant immune processes play a crucial role in antiviral defense [1, 2]. These antiviral factors fall into two categories: basal, which are preexisting and limit viral spread within and between cells, and inducible, which activate upon infection and prevent systemic movement and replication. Inducible mechanisms involve genes that induce broad-scale changes in plant physiology through various

†María J. Olmo-Uceda and Silvia Ambrós contributed equally to this work.

*Correspondence:
Santiago F. Elena
santiago.elena@csic.es
[1] Instituto de Biología Integrativa de Sistemas (I 2 SysBio), CSIC-Universitat de València, Catedrático Agustín Escardino 9, Paterna, Valencia 46980, Spain
[2] Departmento de Genética, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[3] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

signaling pathways, while basal mechanisms involve cell protein alleles whose interaction with viral factors is altered [3]. These changes include local cell death [4], upregulation of nonspecific pathogen responses throughout the plant (systemic acquired resistance -SAR- and induced systemic resistance -ISR-) [4, 5], and activation of RNA-silencing-based resistance, involved in both basal and inducible mechanisms [4, 6, 7].

RNA-based immunity starts by recognizing and degrading double-stranded RNAs (dsRNA). During viral replication, Dicer-like (DCL) proteins cut viral dsRNAs into small RNAs (sRNA), which are then loaded into Argonaute (AGO) proteins to silence complementary RNAs [7]. The *Arabidopsis thaliana* genome has four DCL genes, with most viral siRNAs depending on the DCL4 protein. However, strong antiviral defense may also require the hierarchical activity of DCL2 and DCL3 [8]. Enhancing the silencing response involves RNA-dependent RNA polymerases (RDR), which produce additional dsRNAs from the target [9].

These RNA-mediated defenses are part of a broader and evolutionarily conserved system that regulates gene expression and controls transposable elements (TE) through epigenetic modifications of DNA or histones [10]. DNA methylation occurs in all sequence contexts (CG, CHG, and CHH; with H different from G) and is initiated by RNA-directed DNA methylation (RdDM). This process uses sRNAs from the borders of TEs to limit their expression and affect neighboring gene expression [11]. RdDM involves several components, including DCL3, METHYLTRANSFERASE DOMAINS REARRANGED METHYLASE 2 (DRM2), RNA polymerase IV (POLIV), and RNA polymerase V (POLV) [12]. Non-canonical pathways also contribute to the RdDM pathway, such as DCL-independent mechanisms, POLII-derived mRNAs, or RDR6-derived dsRNAs [13]. Epigenetic marks are copied during replication by maintenance DNA methyltransferases like the plant-specific CHROMOMETHYLASE 3, and chromatin remodelers like DECREASED DNA METHYLATION 1 (DDM1) are needed to control gene expression and TE mobilization in sRNA-independent ways [14]. Histone modification is linked to DNA methylation marks. For instance, histone H3 trimethylation of lysine 9 (H3K9m3) is associated with TE repression, while H3K4m3 promotes gene expression. Both DNA methylation and histone modifications are reversible, with proteins like REPRESSOR OF SILENCING 1 (ROS1), INCREASE IN BONSAI METHYLATION 1 (IBM1), and JUMONJI14 (JMJ14) involved in demethylating DNA, H3K9, and H3K4, respectively [15–17]. Under stressful conditions or in epigenetically-deficient mutants, the chromatin environment around TEs and genes can change, affecting their expression [18].

The connection between the loss of DNA methylation factors and resistance or susceptibility to RNA virus infection has received limited attention. Research on turnip crinkle virus [19], two tobamoviruses [20], and turnip mosaic virus (TuMV) [21] has shown that *A. thaliana* mutants for various RdDM factors, chromatin remodelers, and histone modifiers exhibit varying susceptibility to infections based on the presence of repressive marks. Additionally, evolution experiments with mutants for innate immunity pathways and basal and inducible defense pathways have shown that viral populations rapidly adapt to the new selective pressures imposed by mutations in RdDM genes [22, 23] and histone modification genes [23], indicating these genes play a direct role in the infection cycle of RNA viruses.

All these studies require having an adequate virus-host experimental pathosystem. Members of the *Potyvirus* genus are the largest group of plant RNA viruses that infect a wide range of agronomic species and cause significant economic losses. Extensive studies have tackled the multifunctional roles of potyviral proteins and their interplay with host factors [24, 25]. TuMV is a prototypical member of the genus that infects a broad range of vegetable crops species. From a scientific point of view, TuMV has two particularly relevant properties: (*i*) its wide host range that includes both mono and dicotyledons and (*ii*) being the potyvirus with the highest incidence in wild populations of *A. thaliana* [26]. Therefore, the pathosystem composed by TuMV and *A. thaliana* has been the focus of extensive scientific studies, including the evolutionary ones led by our group.

A common observation in evolution experiments with potyviruses is that the transcriptomic response of *A. thaliana* changes as viral lineages adapt better to the new host [21, 27–29]. To investigate the role of epigenetic regulation of host genes involved in defense responses in the probability of an RNA virus establishing successful infections after spilling over from a reservoir to a naïve host, Ambrós et al. [23] evolved TuMV lineages in *A. thaliana* plants with mutations in DNA methylation and histone demethylation pathways. They found that evolved viral lineages became more infectious, virulent, and reached higher viral loads. Interestingly, the extent of these phenotypic changes depended on the mutated epigenetic pathway. For instance, *jmj14* mutants with affected histone methylation selected for more virulent viruses, while mutants unable to produce siRNAs (*dcl2 dcl3 dcl4*) and thus more susceptible to infection, selected for less virulent viral strains. In terms of genomic changes, all cistrons were shown to be under purifying selection, particularly strong upon the *NIb* cistron encoding the viral replicase. In contrast, the VPg protein was shown to be the most polymorphic one, with nine amino acid

substitutions, all affecting a particular domain in the protein, being under positive selection (see Fig. 4 in [23]). All but one of these fixed mutations was host genotype-specific, rising independently in different hosts. Mutation R118H was the only one that found specifically associated to evolution in *ddm1* plants. We hypothesize that given the expected large-scale effects of mutations in epigenetic regulatory pathways on plant defenses and cellular homeostasis, virus evolution and adaptation to the new host will be affected. This impact can be direct, if epigenetic factors directly interact with virus factors, or indirect, if epigenetic regulation changes the expression of plant resistance genes that ultimately interact with viral factors or cause a significant physiological imbalance, creating a different cellular environment for the virus. To further understand the role of DNA methylation and histone modification pathways in TuMV adaptation, we examined differences in whole-genome transcriptomic responses to infection with the ancestral TuMV and derived viral lineages in their corresponding local host genotypes.

## Methods

### Plants, virus and experimental evolution

For full details of the evolution experiment, refer to [23]. Briefly, the study used wild-type A. *thaliana* accession Col-0 and four mutants: *dcl2 dcl3 dcl4*, *ddm1*, *polv* (all affecting RdDM), and *jmj14* (a histone demethylase). All plants were grown in a climate-controlled chamber with an 8-hour light period (LED lights at PAR 90–100 µmol $m^{-2}$ $s^{-1}$) at 24 °C, and a 16-hour dark period at 20 °C with 40% relative humidity. The soil mix was 50% DSM WNR1 R73454 substrate (Kekkilä Professional, Vantaa, Finland), 25% grade 3 vermiculite, and 25% 3–6 mm perlite. Pest management included introducing *Stratiolaelaps scimitus* and *Steinernema feltiae* (Koppert Co., Málaga, Spain).

TuMV infectious sap was prepared from TuMV-infected *Nicotiana benthamiana* plants inoculated with the plasmid p35STunos, containing a cDNA of the TuMV genome (GenBank AF30055.2) as described in [21]. This sequence corresponds to the YC5 strain from calla lily [30]. Symptomatic tissues were pooled, frozen with liquid nitrogen, and homogenized using a Mixer Mill MM400 (Retsch GmbH, Haan, Germany).

Five TuMV lineages were evolved over twelve consecutive serial passages in each of the five genotypes (Fig. 1). To begin the experiment, ten *A. thaliana* plants per lineage and genotype were rubbed gently on two leaves with 5 µL of inoculum [0.1 g of stock diluted in 1 mL inoculation buffer (50 mM phosphate buffer pH 7.0, 3% PEG6000, 10% Carborundum)]. Plants were inoculated at growth stage 3.5 on the Boyes et al. scale to ensure phenological synchronization [31]. Subsequent passages were performed by harvesting symptomatic plants at 14 days post-inoculation (dpi), preparing infectious sap as described, and inoculating a new batch of ten healthy plants.

### Total RNA extractions and preparation of samples for HTS

Pools of ten infected symptomatic plants per lineage, genotype, and serial passage were frozen with liquid nitrogen and stored at − 80 °C until homogenized into a fine powder. Aliquots of ~ 0.1 g were used for total RNA extractions using the Agilent Plant RNA Isolation Mini Kit (Agilent Technologies, Santa Clara, CA, USA). Three aliquots of total RNA per sample were separated and adjusted to 50 ng/µL.

For RNA-Seq, RNA from 70 to 90 mg of both infected and healthy plants was extracted using the GeneJET Plant RNA Purification Mini Kit (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. RNA quality was checked with a NanoDrop One (Thermo Fisher Scientific) and agarose gel electrophoresis. RNA integrity and purity were verified using a Bioanalyzer 2100 (Agilent Technologies). Library preparation and Illumina sequencing were performed by Novogene Europe using a NovaSeq 6000 platform and a Lnc-stranded mRNA-Seq library method. This included ribosomal RNA depletion and directional library preparation, 150 paired-end sequencing, and 6 Gb raw data per sample. Novogene conducted quality checks on the libraries using a Qubit 4 Fluorometer (Thermo Fisher Scientific), qPCR for quantification, and a Bioanalyzer for size distribution detection.

### HTS data processing

The quality of the Fastq files was assessed using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC [32]. Paired reads were preprocessed with BBDuk (https://sourceforge.net/projects/bbmap/). During preprocessing, adapters were removed, the first ten 5' nucleotides of each read were trimmed, and sequences from the 3' end with an average quality below ten were cut off. Reads shorter than 80 nucleotides after processing were discarded. The parameters used were: *ktrim* = r, *k* = 31, *mink* = 11, *qtrim* = r, *trimq* = 10, *maq* = 5, *forcetrimleft* = 10, and *minlength* = 80.

The processed reads were then mapped to the TuMV isolate YC5 using the BWA-MEM algorithm [33]. The resulting SAM files were converted to binary and sorted with SAMtools [34]. Duplicates were marked using the MarkDuplicates method in GATK version 4.2.2.0 [35].

Next, the reads were mapped to the *A. thaliana* genome (TAIR10) using STAR version 2.7.9a [36] with the 'alignReads' mode and 'GeneCounts' quantification mode. These quantification results were used for transcriptome

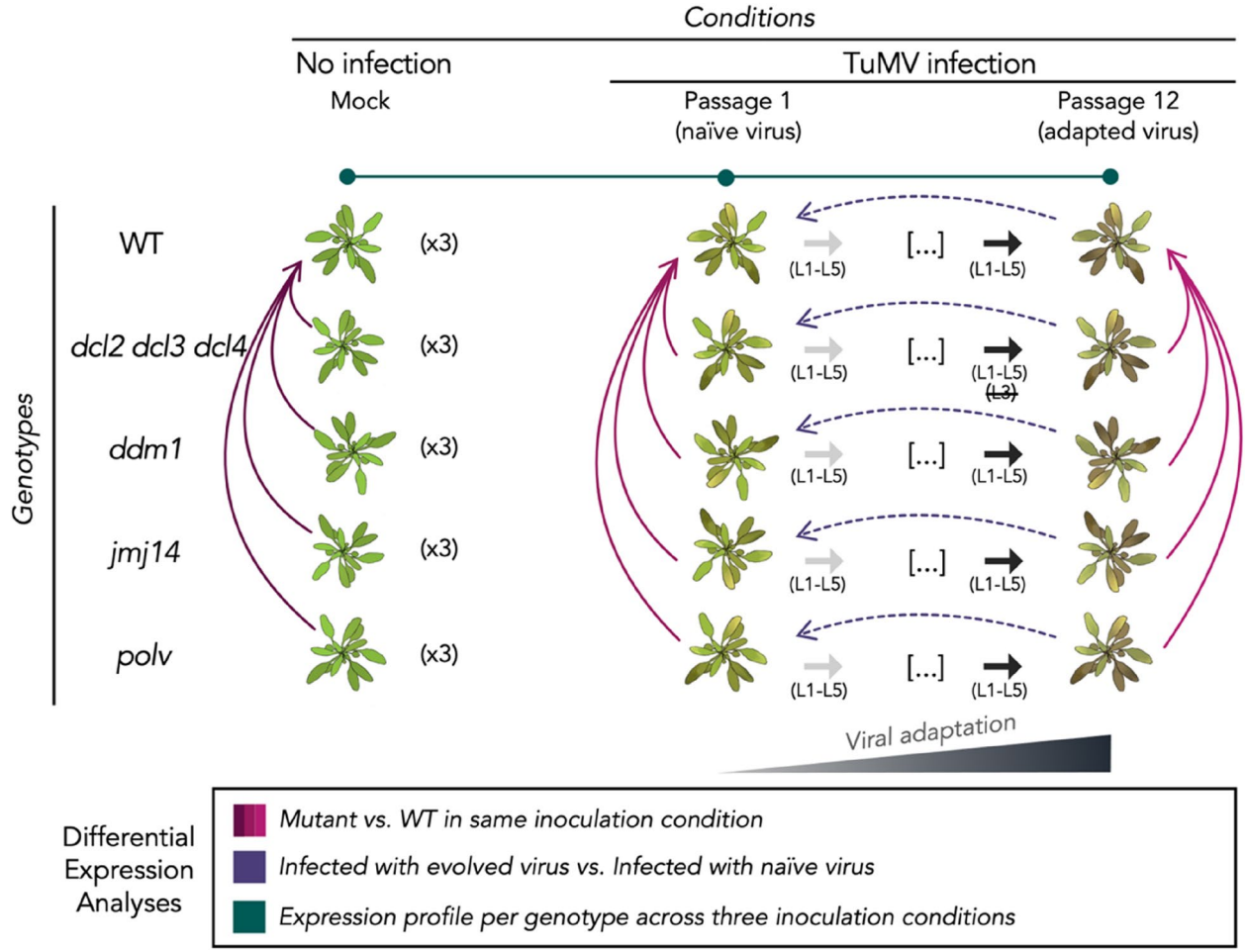



**Fig. 1** Schematic representation of the evolution experiment and the three different contrasts performed on the transcriptomic data. Considering three inoculation conditions: mock (no infection), infection with the ancestral virus (P1) and infection with the evolved viral lineages (P12), the five different plant genotypes and the five biological replicates (L1 – L5), three blocks of differential expression analyses were performed (represented by arrows, with the arrowhead pointing to the baseline of the contrast). Bourgogne: differences between each mutant and the WT in each condition. Purple: differences between the infection with adapted and the naïve virus for each genotype. The lineages were used as block variables. Green: the differential expression analyses between infection with naïve and mock and the differences between adapted and naïve infection (purple lines) led us to classify each gene expression in nine different profiles

analysis. The resulting SAM files were again converted to binary and sorted with SAMtools [34].

All R and shell scripts created for this project are available at https://github.com/MJmaolu/evolution_TuMV_in_A.thaliana_Epigenetic_Mutants/tree/main.

#### Host gene expression and functional annotation

The counts matrix was constructed from the STAR results (reversed-stranded column). Differential expression analyses were performed using DESeq2 v1.36.0 [37]. Genes were considered significant if they had an adjusted $P < 0.05$ (Benjamini and Hochberg method) and a $|\log_2 FC| > 1$. This double criterion for defining differentially expressed genes (DEGs), the number of replicates per treatment (five) and our systems-wide description of the data (rather than focusing in particular genes) makes unnecessary further validation of expression levels by RT-qPCR [38, 39]. The first set of analyses compared the responses of each mutant genotype to the wild-type (WT) plants. This was done separately for mock-inoculated samples, samples inoculated with the ancestral virus, and samples inoculated with the evolved virus (bourgogne lines in Fig. 1). The second set of analyses compared samples infected with the evolved virus to those infected with the ancestral virus, treating samples infected by the same viral lineage as paired samples (purple lines in Fig. 1). The third set of analyses compared the differences between naïve

infection and mock infection for each genotype (green lines in Fig. 1).

Functional enrichment analyses were performed using STRINGdb v2.8.4 [40] and clusterProfiler v4.4.4 [41].

All DEGs were screened for annotated transposable elements (TEs) within their open reading frames (ORFs) or up to one thousand nucleotides upstream of their promoters. This was done using the GenomicRanges package (v1.48.0) [42] and the *A. thaliana* transposon annotation file from TEtranscripts (https://labshare.cshl.edu/shares/mhammelllab/www-data/TEtranscripts/TE_GTF/, last accessed 11/09/2024). Functional enrichment analysis of the resulting genes was performed as previously described.

All processes were conducted using custom scripts in R version 4.2.1. These scripts, along with session information, are available in the mentioned github repository.

#### Classification of gene expression across mock-ancestral viruses-evolved viruses into nine basic profiles

All the genes were classified in the nine possible profiles (Fig. 3 A) depending on the change in the expression between the three inoculation conditions. For evaluating the direction of changes (mock vs. ancestral and ancestral vs. evolved) the results obtained with DESeq2 for each genotype were used. For each of both changes, cases were assigned into three categories: (*i*) increasing expression if adjusted $P < 0.05$ and $\log_2 FC > 1$, (*ii*) decreasing expression if adjusted $P < 0.05$ and $\log_2 FC < -1$ and (*iii*) non-affected otherwise.

#### PPI network and viral protein enrichment

The *A. thaliana* – TuMV PPI published by Martínez et al. [43] was used. The DEGs selected in the contrast evolved vs. ancestral (purple lines in Fig. 1) were represented in the PPI. Node sizes indicate the number of genotypes where the gene is significantly altered (adjusted $P < 0.05$) with a $|\log_2 FC| > 1$. The network was generated with Cytoscape version 3.10.0 [44]. We used the number of genes activated and repressed by the adapted viral lineages as a proxy of the probabilities for gene activation and repression in each host genotype. With these probabilities, we calculated the expected number of interactors altered for each viral protein (Fig. 7 C) and used a $\chi^2$ homogeneity test to evaluate whether some of the viral proteins were enriched or depleted in selected DEGs. Significant changes were considered when $P < 0.05$.

## Results and discussion

Figure 1 illustrates the experimental setup for the evolution study, highlighting specific contrasts in transcriptomic profiles. RNA-Seq analysis was performed on plants from four mutant lines and WT, infected with viral strains isolated at passages 1 (P1, the ancestral stage) and 12 (P12, the evolved lineages). Additionally, transcriptomes of mock-inoculated plants from these five genotypes were also analyzed. The *dcl2 dcl3 dcl4* mutant shows dysregulation of RNA silencing mechanisms, affecting antiviral siRNAs, RdDM, and endogenous siRNAs produced by RDR6-generated dsRNA. This mutant is notably susceptible to virus infection [8]. The *ddm1* and *polv* mutants exhibit altered DNA methylation mechanisms. DDM1 acts as a master regulator of transposable elements (TEs), and its deficiency leads to a loss of RdDM-independent DNA methylation and increased transposition events [45]. POLV plays a crucial role in RdDM by initiating CHH methylation on both DNA strands, specifically targeting short TEs and the edges of longer ones [12]. In *polv* mutants, there is a significant reduction in non-CG methylation across the genome, accompanied by disruption of various stress-related genes [46, 47]. The *jmj14* mutant shows elevated levels of H3K4 activation marks at its target sites [48].

#### Summary of differentially expressed genes amongst infection treatments and plant genotypes

First, we analyzed the differences in the transcriptomes of the mutants with respect to the WT in each inoculation condition (Fig. 2 and Supplemental Fig. S1). These contrasts represent a first approach to discover differences in the way evolution has changed the interaction of evolved viruses with their host genotypes deficient in epigenetic components relative to their interaction with WT plants that are fully competent in all epigenetic regulation pathways. Figure 2 A shows the distribution of the number of differentially expressed genes (DEGs) for each of the four mutant plant genotypes compared to WT plants, non-inoculated. This distribution sets the baseline for the overall effect of the different mutations in plant transcriptomes in absence of infection. The plant genotype showing the largest amount of genotype-specific DEGs was *dcl2 dcl3 dcl4* (2073), followed by *ddm1* (809), *polv* (462), and *jmj14* (258) (Supplemental Fig. S1B). Indeed, the number of genes altered in common by the three mutants affecting DNA methylation-related pathways was 313, 21.32% larger than the number of genes altered by the histone modification mutant. One-hundred sixty DEGs were in common for the four infected mutants. Remarkably, these genes were grouped into two large, well-balanced, clusters of over- and under-expressed genes highly consistent across the four mutant plants (Fig. 2 A) with only two genes presenting a different expression pattern between the mutants (left panel in Supplemental Fig. S1A).

Second, we sought to evaluate the differences with WT in the plants infected with the ancestral P1 viral

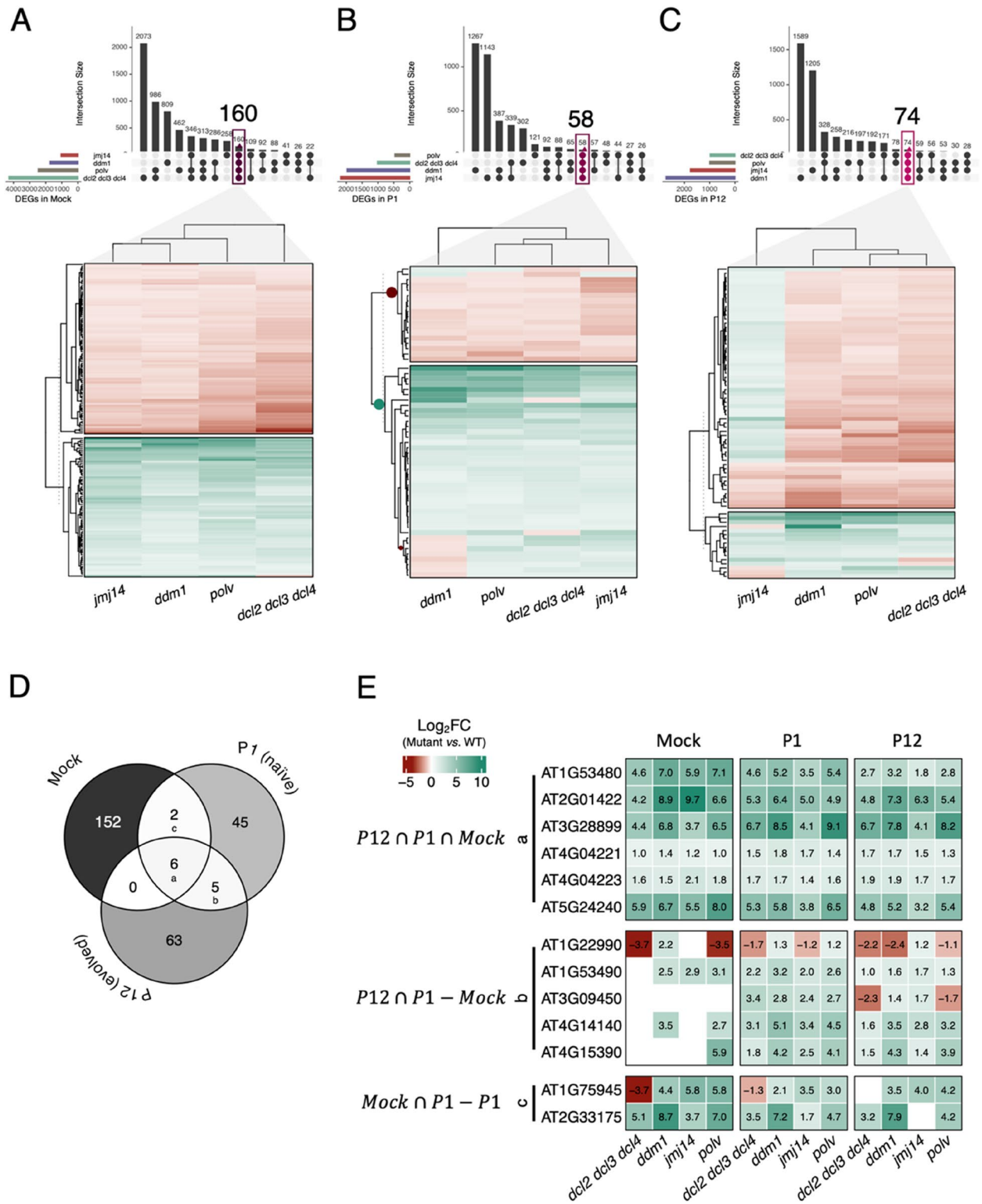



**Fig. 2** Distribution of differentially expressed genes (DEGs) compared to WT per plant genotype for mock-inoculated plants **A** plants inoculated with the ancestral virus **B** and plants inoculated with the evolved lineages. **C** Expression clustering for DEGs shared by all mutant plants are indicated in each case. **D** Distribution of DEGs shared by all the mutants (core genes) in mock-inoculated and plants infected with the ancestral (P1) and evolved (P12) viruses. **E** Gene expression ($\log_2 FC$ represent mutant vs. WT expression by inoculation condition) of the three overlapping sets of core genes in (D), only significant DEGs (adjusted $P < 0.05$) with $|\log_2 FC| > 1$ expression levels are shown, rest in white

populations (Fig. 2 B). The distribution of mutant-specific DEGs has significantly changed relative to the one observed for mock-inoculated plants ($\chi^2 = 3828.937$, 14 d.f., $P < 0.001$): now *ddm1* plants show the largest number of DEGs (1267), followed by *jmj14* (1143), *dcl2 dcl3 dcl4* (302), and *polv* (121). The subset of DEGs in common for the three infected DNA methylation-related mutants was now 27, which represents 97.64% less than the number of DEGs affected by the histone modification infected *jmj14* mutant. A total of 58 DEGs were shared by all four mutant plants after infection. These DEGs group in two major clusters, with an excess of overexpressed genes, although *ddm1* plants showed a group of eight underexpressed DEGs that were overexpressed in all the other mutants (Fig. 2 B, intermediate panel in Supplemental Fig. S1A).

Third, we compared the number of DEGs in plants infected with the evolved lineages (Fig. 2 C). In this case, the relevant comparison is between the number of DEGs observed for each plant genotype and ancestral and evolved viruses. Significant differences have been generated during the course of evolution ($\chi^2 = 733.467$, 14 d.f., $P < 0.001$), mostly driven by an increase in the relative number of *ddm1* (percentage deviation 4.9%) and *polv* (20.8%) specific DEGs along with a reduction in the relative numbers of *jmj14* (3.3%) and *dcl2 dcl3 dcl4* (26.7%). In this case, 74 DEGs were shared by all four infected mutant plant genotypes. These DEGs are clustered into two main categories: 17 over- and 57 under-expressed. However, the behavior of infected *jmj14* plants was remarkably different from the other mutants, with most of these DEGs being overexpressed (Fig. 2 C, right panel in Supplemental Fig. S1A). Overall, the percentage of DEGs shared by all mutants increased from the 1% in mock, to 21% in plants infected with the ancestral viruses and to 72% in plants inoculated with the evolved ones (Supplemental Fig. S1B), being the difference highly significant ($\chi^2 = 109.360$, 2 d.f., $P < 0.001$).

Finally, we evaluate the number of mutant-specific responses integrated with the degree of transcriptome perturbation, observing a differential pattern among mutants (Supplemental Fig. S1B). The direction of change, however, differed among genotypes: (*i*) the level of perturbation is higher in mock conditions than in infection for *dcl2 dcl3 dcl4* and *polv*, the most permissive genotypes [23] but not in the rest. (*ii*) The percentage of mutant-specific DEGs was higher in *ddm1* and *jmj14* mutants during infection (above 50%). And (*iii*) the percentage of mutant-specific DEGs seems to stabilize in *ddm1* (from 46% of specificity in mock conditions to 64% and 58% in evolved viral lineages) while continues growing in *jmj14* (from 24% in mock to 53% in ancestral and 68% in evolved viruses). In agreement with the observation that TuMV rate of evolution was faster in *jmj14* [23], these results confirm the stronger selective pressure that this mutant imposes on the virus.

For the sake of simplicity, hereafter we will focus on a reduced core of DEGs shared by all plant genotypes: six among all three treatments, five between infected plants but not with mock-inoculated plants and two that are shared between mock-inoculated and ancestral-inoculated plants (Fig. 2 D-E).

#### Epigenetic mutants exhibit changes in a reduced set of common genes

Figure 2 D illustrates the overlapping among the 13 core DEGs. The six overexpressed DEGs shared by all plant genotypes, regardless of their infection status, are shown in Fig. 2 E(a). These six loci (Supplemental Table S1) can be seen as nonspecific responses of plants, regardless of their epigenetic context, to the inoculation process and growth conditions. This list includes *METHIONINE OVERACCUMULATION 1 (MTO1) RESPONDING DOWN 1* (*MRD1*), a gene that encodes for a protein involved in down-regulation of methionine biosynthesis, and is essential for salicylic acid (SA)-mediated defense [49], a ribosomal protein L34e superfamily member (AT3G28899), the ubiquitin family protein gene *1-PHOSPHATIDYLINOSITOL 4-KINASE C3* (*PI4KC3*), involved in regulation of flower development and responses to abscisic acid [50] and three noncoding RNAs (AT2G01422, AT4G04221 and AT4G04223). Interestingly, *MRD1*, *PI4KC3*, AT2G01422, AT3G28899, and AT4G04223 have previously been shown to be directly regulated by RdDM [46, 51, 52].

Figure 2 E(b) shows the expression pattern of the five core DEGs shared only by infected plants irrespective of virus' evolution status (Supplemental Table S1). These five DEGs would represent shared responses to infection. This list includes genes related to stresses (AT1G22990, AT3G09450 and AT4G15390), meiosis (AT1G53490) and the methyltransferase *MET2*. AT1G53490, AT4G15390 and *MET2* are consistently induced upon infection with both the ancestral and evolved viruses in all plant genotypes. Changes of expression direction due to virus evolution is observed for the gene AT1G22990 in *ddm1*, *jmj14* and *polv*, indicating a potential role in the adaptation process in these particular genotypes. The same holds true for the gene AT3G09450 in the DNA methylation-related *dcl2 dcl3 dcl4* and *polv* mutants. Among the five genes, one (AT1G53490) is predicted to be directly regulated by POLV [46].

Two core DEGs are shared by mock- and ancestral-inoculated plants, but not with the evolved-inoculated ones (Fig. 2 D; Supplemental Table S1). Therefore, these two DEGs can be seen as either targets or drivers of

TuMV adaptation to specific mutant plant genotypes. AT1G75945 and AT2G33175 both encode for mitochondrial membrane-associated hypothetical proteins that both were overexpressed compared to WT plants, except in *dcl2 dcl3 dcl4* plants, in which the first one was underexpressed (Fig. 2 E(c)). Notably, the ~4-fold increase in expression observed in *polv* plants agrees with the previous observation by Corrêa et al. [46].

#### Impact of viral adaptation into plant responses

Once the differences between mutants and WT plants in their response to infection where explored, we focused on the expression changes elicited by the adaptation of the viral lineages to their corresponding local host genotypes [Fig. 1 purple lines: infected with evolved viral lineages (P12) vs. infected with their corresponding ancestors (P1)]. In this context, by up-regulated we meant significantly more expressed in plants infected with the evolved lineages with respect to plants infected with their corresponding ancestors; by down-regulated we meant the opposite. To simplify wording, we will refer to these genes as activated or repressed by the viral adaptation. The total number of virus-adaptation DEGs was larger in *jmj14* and *polv* than in WT (Fig. 3 A) while, at the other extreme, *dcl2 dcl3 dcl4* and *ddm1* showed a similar degree of disruption than WT plants (Fig. 3 A). However, it is noticeable that in all four mutants, the distributions of differentially activated and repressed genes were different than observed in WT plants (Fisher's exact tests, $P < 0.001$). Overall, in all mutant plant genotypes, the number of repressed genes was larger than in WT plants. The situation was more variable for activated genes, with *dcl2 dcl3 dcl4*, *ddm1* and *polv* showing less than WT plants but *jmj14* showing more.

A functional enrichment analysis (GO: biological process) of the DEGs separated by its regulation revealed some categories common to all the genotypes (Supplemental Fig. S2). In the one hand, SAR, response to SA, response to hypoxia, or response to chitin were activated in the evolved viruses, while amine metabolic process, cellular response to hypoxia and to cold, and cellular response to jasmonic acid stimulus were enriched in the group of DEGs repressed in the evolved viruses. In the other hand, some categories were found only in specific genotypes, e.g., cutin biosynthetic process, isoprenoid and tetrapyrrole metabolic processes were enriched within the DEGs activated by the viral adaptation in WT but not in the mutants. *ddm1* presented the largest number of specific terms within the DEGs activated in evolved lineages (e.g., carotenoid related processes, cellular homeostasis or ketone biosynthetic processes) while *jmj14* was the genotype with more specific terms within the DEGs repressed (e.g., cell wall biogenesis, sulfur compound biosynthetic process or tropism between the specifically enriched terms) (Fig. 3 B, Supplemental Fig. S2).

#### A core of genes associated with differences in TuMV adaptation was found differentially expressed in all plant genotypes

From a total of 9409 DEGs, we found a set of 1029 genes whose expression was significantly affected by the degree of TuMV adaptation, independently of the plant genotype in which evolution was carried on (Fig. 3 C). This core of genes showed a consistent profile of up- (48.0%) and down-regulation (51.5%) across all plant genotypes, except for five genes (Fig. 3 D). These exceptions were (Fig. 3 D, Table S2): AT1G69480 and AT4G09820 both specifically repressed in *dcl2 dcl3 dcl4* but activated in all other genotypes, AT2G40610 and AT5G65730 only repressed in *jmj14* but activated in the rest of genotypes, and AT4G25480 repressed in *dcl2 dcl3 dcl4*, *jmj14* and *polv* but activated in WT and *ddm1* plants.

The GO biological processes associated with the set of genes that increase their expression with virus adaptation in all genotypes include response to chitin, response to SA, SAR, response to ethylene signaling pathway and phosphorelay signal transduction system (Fig. 3 E and Supplemental Fig. S2). Among the shared genes whose expression has been repressed by viral adaptation, processes related with amine and aromatic amino acid family biosynthesis, jasmonic acid metabolic process, auxin metabolic processes, circadian rhythm, response to blue light, or response to toxic substances were observed (Fig. 3 E and Supplemental Fig. S2). Categories such as cellular response to hypoxia, cellular response to ethylene stimulus, response to cold, or response to red or far-red light appeared enriched both in activated and repressed genes (Fig. 3 E and Supplemental Fig. S2).

#### Classification of genes according to their profile of gene expression in non-infected plantsvs. plants infected with the ancestral or the evolved viruses

In order to get a wider picture of the effect of infection with viral lineages at different degrees of adaptation to their local host genotypes, we considered the DEGs across the three inoculation conditions (Fig. 1 green lines) using the following approach. According to the inoculum type, i.e., mock, ancestral and evolved viruses, genes can be classified into nine basic profiles based on the direction of the changes between each pair of conditions, as illustrated in Fig. 4 A. Figure 4 B shows the clustering of the five plant genotypes according to the similarity between the genes classified into each of the nine profiles. DNA methylation mutants *dcl2 dcl3 dcl4*, *ddm1* and *polv* cluster together (*polv* share 81% with *dcl2*

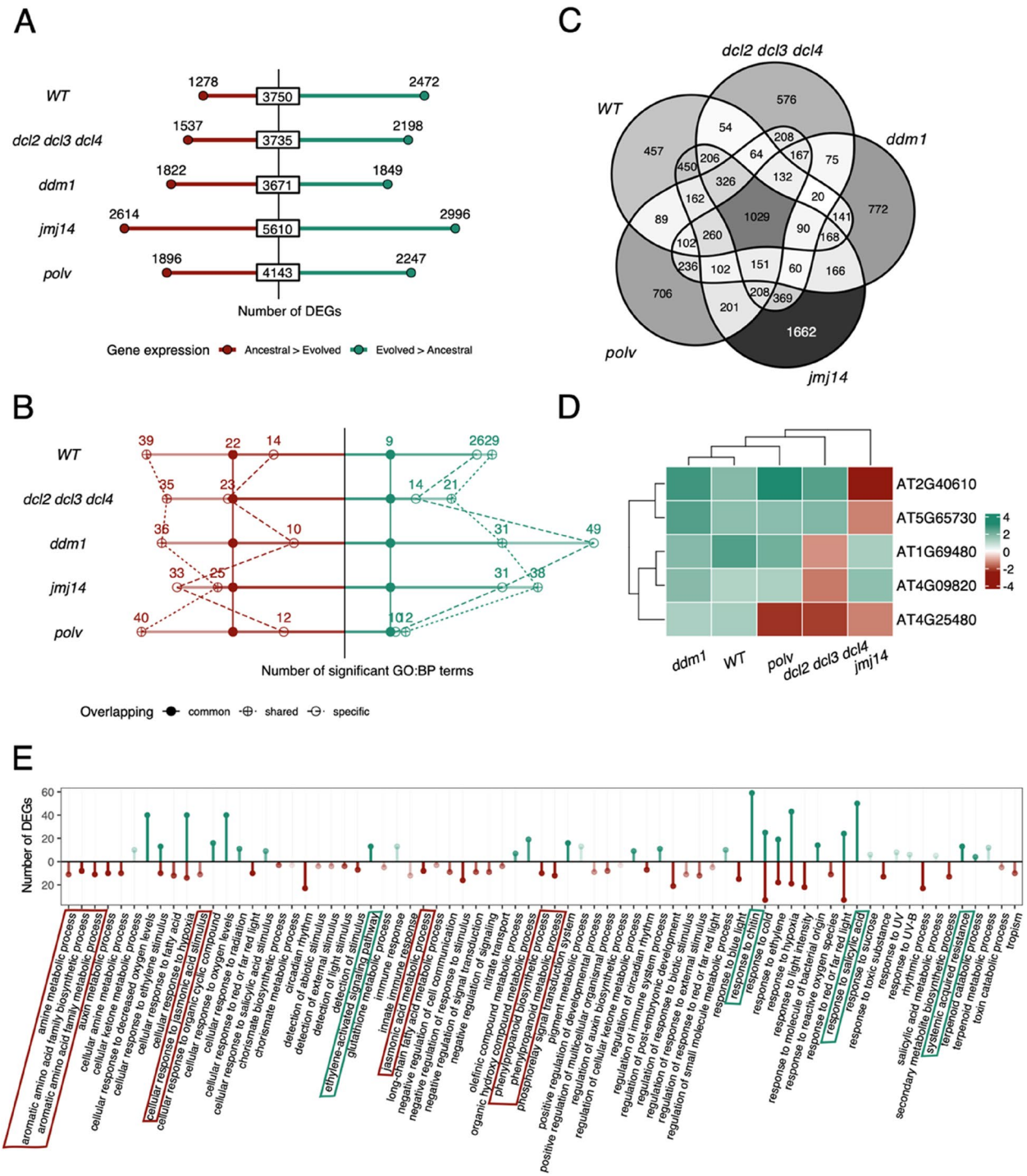



**Fig. 3** Characterization of DEGs between plants infected with the ancestral and evolved viruses, per plant genotype. **A** Number of repressed (red) and activated (green) DEGs elicited by viral adaptation in each plant genotype. **B** Number of GO biological processes terms within the repressed (red) and activated (green) DEGs enriched per plant genotype, indicating the number of common to all plant genotypes (solid symbols), shared by more than one plant genotype (crossed symbols), and specific of a single plant genotype (open symbols). **C** Distribution of DEGs across plant genotypes. **D** Only five of the 1029 DEGs common to all genotypes, present an opposite gene expression. The color scale represents the level of expression in $\log_2 FC$ as in (A) and (B). **E** GO biological processes terms from DEGS in common indicated in panel (C). In green, enriched categories within the activated DEGs, in red enriched categories within the repressed DEGs (adjusted $P < 0.05$

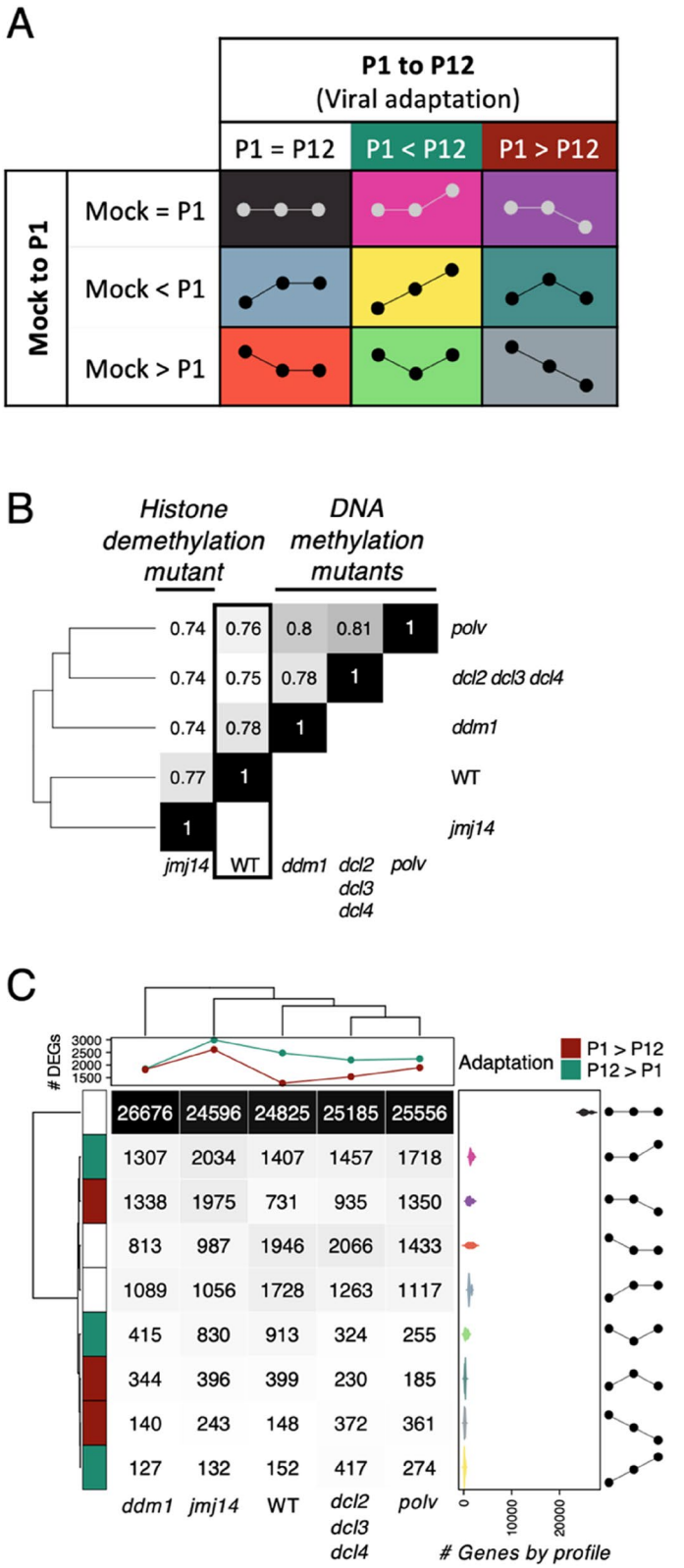



**Fig. 4** Classification of gene expression profiles according to the nine possible patterns of variation across conditions. **A** Classification profiles based on the direction of changes between the three inoculation conditions, mock-inoculated, ancestral (P1) virus-infected and evolved (P12) viruses-infected. **B** Hierarchical clustering of plant genotypes based on the similarity amongst their classification profiles in panel (A). **C** Abundance of each profile per plant genotype

*dcl3 dcl4* and an 80% with *ddm1*) while the histone demethylation mutant *jmj14* shows a more dissimilar distribution of profiles.

Figure 4 C shows all DEGs found on each of the plant genotypes classified into each of the nine possible expression patterns when comparing between mock-inoculated, ancestral- and evolved-infected plants. By far, in all five genotypes the most common pattern corresponded to genes that showed no differences between mock-inoculated and virus-infected plants, followed in WT by genes that responded to infection in the same magnitude (up or down) both for the ancestral and the evolved viruses. The most interesting cases are precisely those in which the pattern showed differences between WT and mutant plants inoculated with ancestral and evolved viruses. Fisher's exact tests were used to evaluate pairwise differences between the patterns of DEGs shown for WT and each of the four mutant plant genotypes. In all four cases, the number of DEGs up or down found in the ancestral and the evolved inoculated mutant plants was different than observed for WT plants (Fisher's exact test, $P < 0.001$ in all comparisons). Indeed, the number of up- and down-regulated DEGs due to virus evolution relative to the number of DEGs not associated to virus evolution (Fig. 4 C) is 2.75-fold largest for *jmj14* followed by 1.93-fold for *ddm1*, while the smallest difference was found for WT plants (1.02-fold). This observation further supports that TuMV adaptation to histone demethylation mutant plants resulted in a stronger genome-wide alteration in gene expression than adaptation to WT plants.

Next, we sought to functionally categorize the DEGs specific to each of the nine patterns in Fig. 4 C. Supplemental Fig. S3, present the results of the functional enrichment GO analyses per pattern for each plant genotype. For example, comparing across plant genotypes, a pervasively enriched GO term was photosynthesis processes though classified into different profiles: downregulated upon infection in *dcl2 dcl3 dcl4*, *polv* and WT, downregulated only by evolved infection in *jmj14* and upregulated in evolved-infected *ddm1* plants. Responses to chitin also showed an upregulation pattern for *ddm1* plants infected with the evolved viral lineages but not in the other genotypes. As an additional example of highly variable profiles, responses to SA signaling were always upregulated upon infection with evolved viruses, except in *ddm1* plants, and with *polv* plants showing upregulation also when infected with the ancestral viruses.

To delve deeper into functional responses potentially influenced by virus adaptation, we specifically examined each plant genotype and focused on differences between virus ancestral- and evolved-infected plants (Supplemental Fig. S4). In WT plants, only responses to cold were significantly altered by both types of viral lineages. Processes related to $O_2$ levels and SA were overrepresented in evolved-infected plants, while processes related to membrane transport and organization, endocytosis, rRNA processing, and cytoplasmic translation were underrepresented in evolved-infected plants. In *dcl2 dcl3 dcl4*, responses to chitin, and hypoxia were specifically overrepresented in evolved-infected plants, while processes related to circadian rhythms and the transition from vegetative to reproductive phases in meristems were underrepresented in the same plants. This observation aligns with the impact of TuMV infection on the transition from vegetative growth to reproduction, resulting in plant castration [53]. *ddm1* plants exhibited the shortest list of significant GO terms, with only photosynthesis being shared by three of the expression profiles, two of which were overrepresented in evolved-infected plants compared to their ancestral counterparts. The histone modification-deficient *jmj14* plants displayed a diverse array of GO terms across the nine expression patterns. For example, responses to chitin, hypoxia, SA, jasmonic acid, and SAR were overrepresented in evolved-infected plants compared to ancestral-infected ones. In contrast, photosynthesis, response to cold, biogenesis of ribonucleoprotein complexes, rRNA processing and metabolism, and, notably, brassinosteroid-mediated signaling were underrepresented among evolved-infected plants. Finally, in *polv* plants, plastids organization and microtubule-related processes were specifically overrepresented in evolved-infected plants, while endomembrane system organization, Golgi vesicle transport, and processes related to the transition from vegetative to reproductive phases in meristems and regulation of multicellular differentiation were significantly underrepresented among evolved-infected plants compared to ancestral-infected ones. In short, an analysis of the intersections by profile shows that, avoiding the no-changes profile, the pattern with more genes in common among genotypes was represented by the repression in plants infected with ancestral viruses but not change with evolution (242 genes in common) while the less convergent was the upregulation through all conditions (6 genes in common) (Supplemental Fig. S5).

### DNA but not histone methylation affects the expression of genes nearby TEs

In a prior study, Corrêa et al. [21] observed that in WT plants infected with evolved TuMV lineages, but not with their ancestral counterparts, TEs concentrated in centromeric and pericentromeric regions were induced early after infection. At later stages of infection, both induction and repression of different TE families were observed. Subsequently, Corrêa et al. [46] demonstrated that POLV target genes had a higher percentage of TE marks around

their transcription start sites (TSS) than other genes. In contrast, genes regulated by JMJ14 showed a lower level of marks around their TSS. These observations suggest that genes found to be upregulated in infected DNA methylation mutants, especially in *ddm1*, but less in *jmj14* plants, are likely to be controlled by TE-related mechanisms. To further investigate this hypothesis, we sought differences in the number of DEGs with a TE within 1 Kbp upstream of their TSS between virus ancestral- and evolved-infected plants (Fig. 5 A). First, a highly significant correlation exists between the total number of DEGs and the number of DEGs in the proximity of TEs (Fig. 5 A: Pearson's $r=0.990$, 3 d.f., $P=0.001$). The proportion of DEGs associated with TE was slightly higher in all the mutants than in WT plants, but this difference was significantly enriched only in *ddm1* plants (Fig. 5 A: Fisher's exact test, $P=0.045$). However, when comparing the TE-related DEGs between the three DNA methylation mutants and the histone modification mutant *jmj14*, also *ddm1* showed a significant augment in the number of TE-associated DEGs responding differentially to ancestral and evolved viral populations (Fig. 5 A: Fisher exact test, $P=0.036$). The observed differences in the number of DEGs close to TEs in *ddm1* plants could be attributed to the higher basal instability of these genetic elements and the stronger transcriptome disturbance caused by the evolved virus (Supplemental Fig. S1B) in this genotype compared to the others.

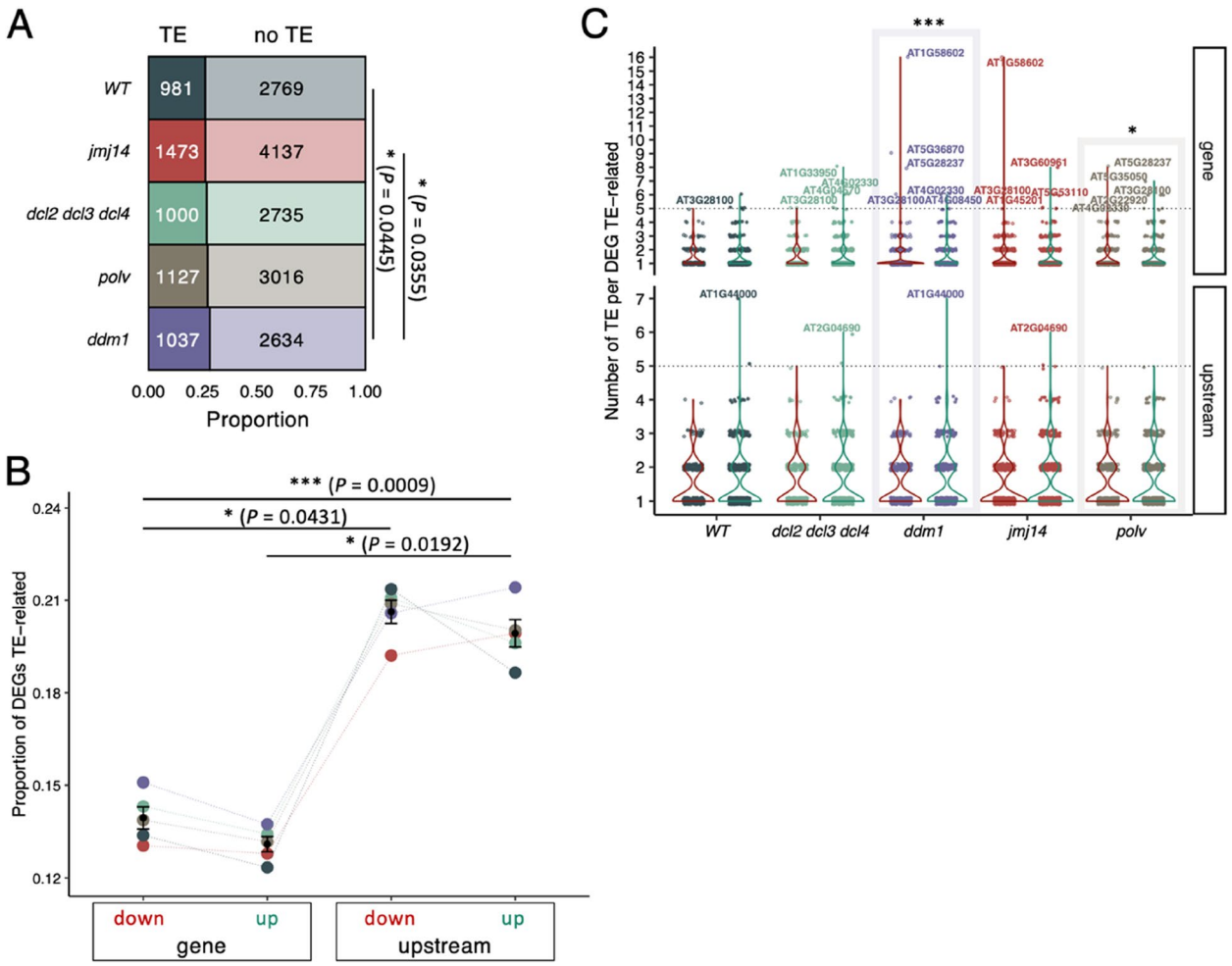



**Fig. 5** Analysis of DEGs (contrasting evolved vs. ancestral viruses) in the context of their association to TEs. **A** Proportion of DEGs in close proximity to TEs in each plant genotype. *ddm1* present a higher number of DEGs TE-associated than WT and *jmj14*. **B** Interaction between the position of TEs in genes (inside or within 1 Kbp upstream of their TSS) and the direction of regulation (upregulated or downregulated in the evolved lineages). Values are normalized by the total number of DEGs with the same locus and regulation. *P* values of Kruskall-Wallis rank sum test are indicated over the significant contrasts. **C** Distribution of DEGs related to the presence of TEs. Violin plots color represents DEGs-TE related within the activated (green) and repressed (red) ones. *ddm1* and *polv* showed significant differences in the distribution of its TEs depending on the localization and the expression where the TEs are associated (Kruskal-Wallis, $P<0.001$ and $P=0.001$, respectively)

Subsequently, we categorized the DEGs close to TEs based on their regulation (upregulated or downregulated in the evolved lineages) and the position of the TEs in the gene (whether they were located inside or within 1 Kbp upstream TSS) (Fig. 5 B). Significant differences exist between the number of DEGs containing a TE inside or upstream the gene (Kruskal-Wallis test: $\chi^2 = 6.818$, 1 d.f., $P < 0.001$), the sign of the regulation of the TE-related DEGs (Kruskal-Wallis test: $\chi^2 = 4.811$, 1 d.f., $P = 0.028$) and the interaction between both TE location and the sign of regulation (Kruskal-Wallis test: $\chi^2 = 16.006$, 3 d.f., $P = 0.001$). A Dunn *post hoc* pairwise comparisons test found significant differences between the location of TEs (in down-regulated, $P = 0.043$; in up-regulated, $P = 0.019$) but the regulation component only was significant between different locus ($P < 0.001$) (Fig. 5 B). No significant differences between genotypes were found in any group. The data suggest that TEs near gene promoters are more likely to correlate with changes in gene expression than those within genes.

Additionally, we checked the number of TEs associated with each TE-related DEG. This data revealed that although the proportion of DEGs having TEs in proximity of TSS is higher than the ones having TEs inside the genes (Fig. 5 B), the total counts of TEs in these same DEGs is higher inside than upstream the genes (Fig. 5 C: Wilcoxon rank test with continuity correction, $P < 0.001$). After separating by sign of regulation and localization, *ddm1* (Kruskal-Wallis, $P < 0.001$) and *polv* (Kruskal-Wallis, $P = 0.001$) presented differences within its four possible combinations (Fig. 5 C). The DEG with the highest density of TEs found inside its sequence (16 of both classes of TE) was *RECOGNITION OF PERONOSPORA PARASITICA 7* (*RPP7*), a gene that encodes an LRR and NB-ARC domains-containing disease resistance protein (Fig. 5 C) [54]. *RPP7* appeared as repressed by the evolved viruses adapted in *ddm1* and *jmj14* mutant plants. This gene is known to be regulated by a particularly interesting epigenetics mechanism in which a "domesticated" TE located in its first intron must be properly silenced in order to produce full-length transcripts [55]. Out of these 16 TEs inside *RPP7*, half belonged to the MuDR superfamily (Fig. 6) and more precisely to the VANDAL transposons. VANDAL transposons have coevolved mechanisms to escape the epigenetic silencing [56]. It is noteworthy that in *jmj14* and WT plants both activated and repressed genes were enriched in MuDR inside coding regions (Fig. 6). However, in *polv* plants only activated genes were enriched in MuDR while in *ddm1* plants, MuDR enrichment was only for repressed genes.

### Class II TEs are significantly enriched among Arabidopsis genes responding to TuMV adaptation while class I TEs were significantly deprived

An enrichment analysis of the types of TEs associated with the DEGs associated with virus adaptation revealed a general pattern of overrepresentation of class II elements (DNA transposons) and underrepresentation of class I ones (retrotransposons). A detailed analysis by superfamilies of TEs is shown in Fig. 6. The distribution of TE families across all the genome was used as the null model (expected values) and the proportion found at each condition (observed values) was tested with a $\chi^2$ goodness-of-fit. Briefly, from class I, Gypsy appears less than expected ($P < 0.001$), no matter the sign of regulation, localization relative to the gene coding sequence, or the plant genotype. An interesting exception is *ddm1* plants, where ATSNE2 (SINE type) and L1 (LINE type) were significantly enriched within the activated genes (Fig. 6: $P < 0.001$ and $P = 0.008$, respectively).

In contrast with this deprivation in class I TEs, a general enrichment in class II transposons was observed in all four plant genotypes, both upstream the TSS and inside the DEGs ($P < 0.001$). Members of the Helitron superfamily were the most abundantly over-represented type (Fig. 6). Helitrons (alike ssDNA viruses, bacteria and plasmids) use rolling circle replication for its propagation. This mechanism, in contrast with the cut-and-paste strategy of other DNA transposons, allows the retention of the original TE while the daughter TEs is copied. This mechanism allows TEs to capture gene fragments that can derive in epigenetic conflict [57, 58]. Cases of epigenetic conflict in the stress response context driven by Helitrons have been previously reported. For example, ATREP2 TEs have been found enriched in genes regulating ISR responses to herbivores [57, 59].

A GO enrichment analysis of the biological processes of these TE-proximal DEGs (Supplemental Fig. S7A) showed that the response to SA appeared overrepresented in all five plant genotypes. The same occurs with various categories related to hypoxia and the response to chitin, represented in genes related to both classes of TE. Responses to ethylene and to cold were also overrepresented in at least four of the five plant genotypes (Supplemental Fig. S7A). *ddm1* plants presented the largest number of terms enriched within the DEG TE-related elicited by the viral adaptation and none within the repressed ones (Supplemental Fig. S7B). Its more specific categories are related to photosynthesis. In contrast, *polv* present the largest number of enriched terms within the TE-related DEGs repressed in the evolved viruses (Supplemental Fig. S7B). In this case, genes are related to class I TE and all the functional categories are related to metabolic processes (Supplemental Fig. S7A).

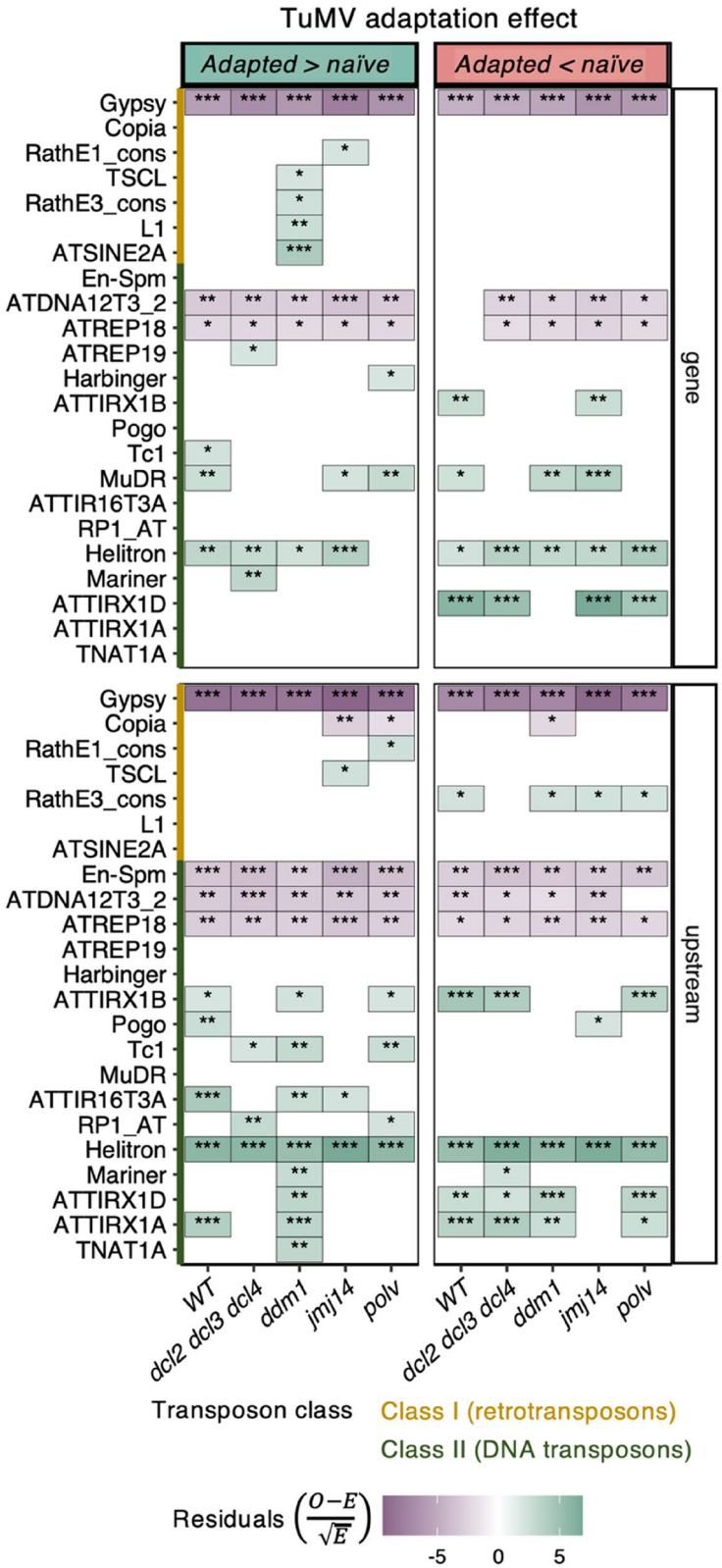



**Fig. 6** Analysis of the TE families linked DEGs associated to differences between ancestral and evolved viral isolates. Asterisks indicate significance levels in a $\chi^2$ goodness-of-fit test: $^{*}P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$

### Putative alteration of the TuMV-*A. thaliana* protein-protein interaction network due to virus adaptation

The way in which virus and host interact changes with the degree of virus' adaptation, which indeed depends on the fine tuning in the interactions between viral and host factors [27–29]. In order to explore the extent to which the target genes of TuMV adaptation could be perturbing the virus-host interaction network, we integrated the transcriptomic results described above with the protein-protein interaction (PPI) data in [43].

Figure 4 A depicts expression profiles for DEGs whose proteins have been demonstrated to interact with viral proteins. The nine derived profiles differ significantly from the genome-wide expression in four host genotypes, except for the *polv* mutant [$\chi^2$ homogeneity tests; *dcl2 dcl3 dcl4* ($P<0.001$), *ddm1* ($P=0.011$), *jmj14* ($P=0.002$), *polv* ($P=0.271$), and WT ($P=0.045$)] (Fig. 7 A). Focusing on the DEGs showing differences between virus ancestral- and evolved-infected plants, we found in infected *ddm1* and WT plants more PPI-associated DEGs activated within the viral adaptation, while *dcl2 dcl3 dcl4* and *jmj14* were enriched in more repressed genes (Fig. 7 B). We used the number of genes activated and repressed by the evolved viral lineages as a proxy of the probabilities for gene activation and repression in each host genotype. With these probabilities, we calculated the expected number of interactors altered for each viral protein (Fig. 7 C) and used a $\chi^2$ homogeneity test to evaluate whether some of the viral proteins were enriched or depleted in selected DEGs. P3 and P3N-PIPO showed more interactions than expected with proteins encoded by the overexpressed genes in *ddm1* ($\chi^2$ homogeneity tests, $P<0.001$ for both proteins), an over-representation of repressed genes was found for 6K1 in WT ($P=0.031$), for CI and NIb in *dcl2 dcl3 dcl4* ($P=0.029$ and $P=0.007$, respectively), as well as for CI in *jmj14* ($P=0.021$). Despite this, *ddm1* is the only genotype with significantly more upregulated interactors than expected by share chance.

Finally, Supplemental Fig. S8, maps the effects of viral adaptation in the host transcriptome over the host-virus PPI network. Our previous work showed that TuMV VPg is a target of virus adaptation to different genotypes of *A. thaliana*, to its susceptibility to infection or to environmental conditions through an increase in its fitness and/or virulence [22, 23, 53, 60]. In the present PPI network analysis, five repressed and four activated interactors were found for VPg, although none of them was shared by all plant genotypes. It is noteworthy that four out of the five repressed interactors displayed binding functions, highlighting the *FAS1* subunit of the chromatin assembly factor 1 (*CAF-1*), that has a remarkable histone chaperone function whose defect induced an altered plant epigenetic landscape [61], and the *eIFIso(4G2)* translation initiation factor. It is well known that potyviruses require translation initiation proteins of the eIF4F/eIF4E/eIF4G complex including the scaffolding eIFIso(4G2) protein. Recently, it has been showed a connection between the levels/degradation of some of these translation initiation proteins and the increase or restriction of TuMV accumulation in *A. thaliana* infected plants [62]. Regarding the VPg induced interactors it highlights a HSP40/DNAJ chaperone which plays a role in viral pathogenicity.

By contrast, it should be noted that thirteen altered interactors were found in the five plant genotypes (Supplemental Table S3). Viral replicase NIb had five repressed (with diverse functions, but remarkably a methyltransferase superfamily member) and four activated interactors (including transcription factor MYB75 and E3 ubiquitin ligase), multifunctional CI two activated (E3 ubiquitin ligase and an Agenet domain-containing protein involved in chromatin remodeling) and one repressed (Zinc finger FYVE domain containing protein), and 6K1 a repressed one, *PHENYLALANINE AMMONIA-LYASE 1* (*PAL1*). A further exploration of these extensive results would be beyond the scope of this article and the focus of future work, but we refer readers to the information shown in Supplemental Table S2. However, in this complex context we wanted to highlight the significant virus-host interaction involving 6K1 and PAL1. It is relevant to note that 6K1 is the only viral protein for which only one interactor has been found, shared by all genotypes, for which an unexpected biological function

(See figure on next page.)

**Fig. 7** Interactors in the host-virus PPI network are altered by the adaptation of the virus. **A** Distribution of the PPI-related genes in the nine profiles classification developed in this work. The expression profiles (Fig. 4 A) of the genes involved in the PPI, were found to significantly differ from the genome-wide expression in four host genotypes, except in the case of *polv* mutant [$\chi^2$ homogeneity tests; *dcl2 dcl3 dcl4* ($P<0.001$), *ddm1* ($P=0.011$), *jmj14* ($P=0.002$), *polv* ($P=0.271$) and WT ($P=0.045$)]. **B** Gene expression of DEGs PPI-associated. Asterisks indicate significant changes (adjusted $P<0.05$), when $|\log_2 FC| > 1$, bolded. The lateral bars represent the number of genotypes with the interactor significantly affected (adjusted $P<0.05$ and $|\log_2 FC| > 1$). **C** Observed (bars) and expected (points) host interactors for each viral protein were tested to search for significant differences using a $\chi^2$ homogeneity test. Asterisks indicate significance level: $^{*}P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$. P1: ancestral viruses, P12: evolved viruses

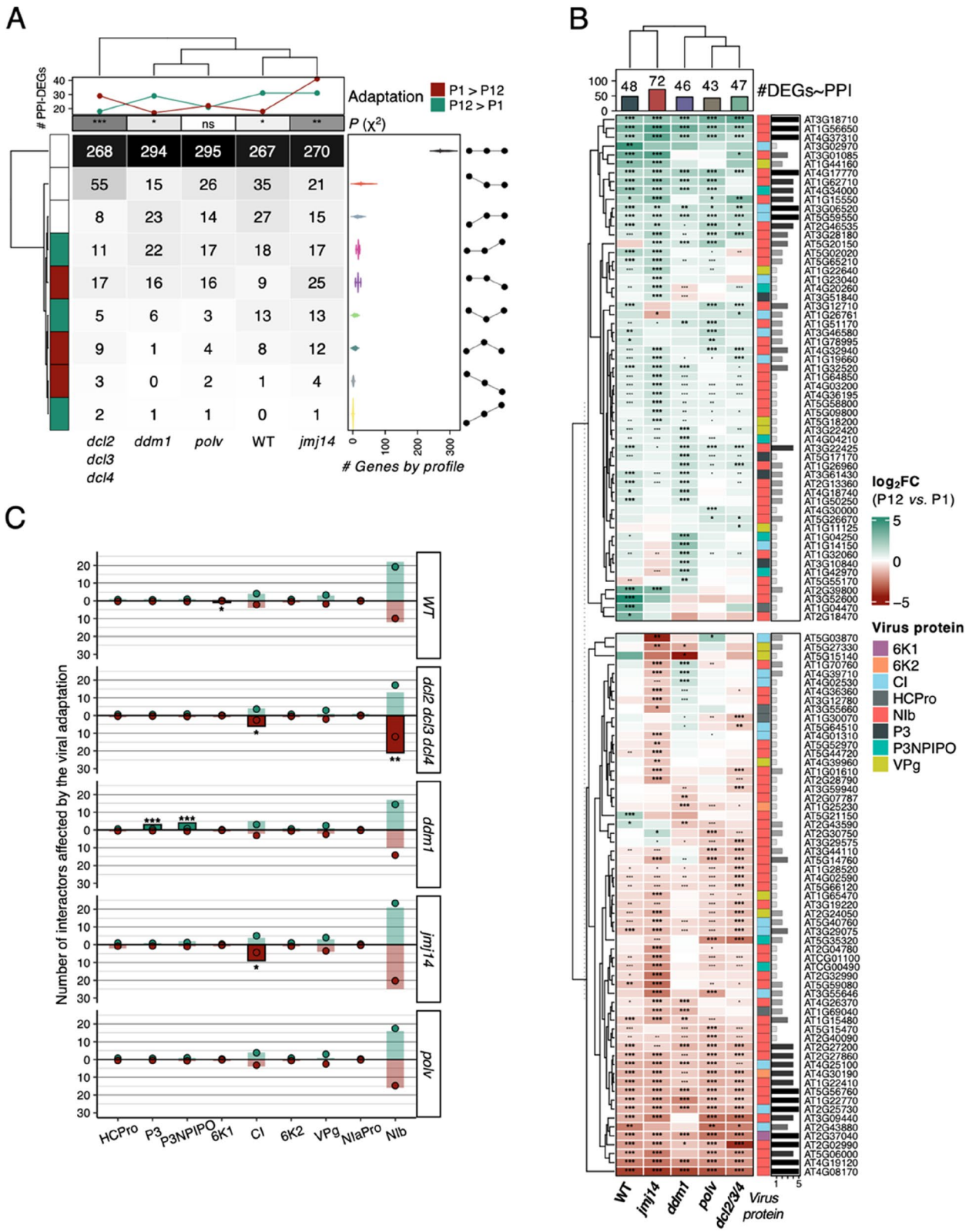



**Fig. 7** (See legend on previous page.)

has been reported recently [63]. Pentamers of 6K1 had been demonstrated to be a functional viroporin [63]. Viroporins are small hydrophobic proteins that insert in membranes modifying its permeability and are essential for viral pathogenicity, affecting inflammasome formation, inducing apoptosis and autophagy, and evading immune responses [64]. Their role in viral pathogenicity and inflammatory responses has been assessed in several viruses [65], as well as their potential as drug targets [66]. 6K1, contrary to the majority of the known viroporins, also plays an essential role in the viral replication cycle [67]. 6K1 interactor PAL1 is a phenylalanine ammonia-lyase that acts upstream of the phenylpropanoid pathway. This pathway was found to be consistently repressed in plants infected with the evolved TuMV populations (Fig. 3 and Supplemental Fig. S3), thus suggesting its importance in viral adaptation. Among other diverse roles, this pathway has been shown to be redundant in synthetizing SA and to be involved in resistance to fungal [68] and viral [69] pathogens. Concretely, the overexpression of *PAL1* in cassava increases resistance to cassava brown streak virus, another member of the *Potyviridae* family [69]. Despite the character of the interaction 6K1 - PAL1 remains unknown, the repression of *PAL1* expression by the evolved lineages in the five plant genotypes suggests *PAL1* as a promising candidate for future studies.

### Concluding remarks

Evolution profoundly alters how plant viruses interact with their hosts [21, 27–29, 60]. Building on earlier studies [23], our comprehensive analysis of large transcriptomic datasets reveals that changes in epigenetic regulatory pathways play a crucial role in shaping the evolution of RNA viruses. We identified a core set of genes that influence TuMV adaptation in *A. thaliana*, independent of the specific genetic makeup of the plant's epigenetic pathways. Additionally, we found many genes whose impact on viral adaptation varies depending on the host genotype. Our results highlight the significant role of the JMJ14 histone demethylase in driving TuMV evolution. Furthermore, our analysis uncovered a strong association between genes regulated by epigenetic marks and the presence of TEs located within or near genes involved in TuMV adaptation, particularly evident in *ddm1* mutant plants. By integrating transcriptomic data with PPIs in the same pathosystem, we identified additional host genes critical for TuMV adaptation. Notably, our study emphasizes the importance of the interaction between viroporin 6K1 and PAL1 in facilitating virus adaptation.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10798-x.

Supplementary Material 1.

#### Acknowledgements

We thank Francisca de la Iglesia and Paula Agudo for excellent technical support and the EvolSysVir lab members for comments and fruitful discussions. Computations were performed on the Garnatxa HPC cluster at the Institute for Integrative Systems Biology (I2SysBio).

#### Authors' contributions

S.F.E. conceived the experiments. S.A. and R.L.C. performed the experiments. M.J.O-U., R.L.C., S.F.E. conceived the bioinformatic analysis. M.J.O-U. wrote the code, performed all bioinformatic analyses and create the visualizations. M.J.O-U. and S.F.E. performed the statistical analyses. S.F.E. and M.J.O-U. wrote the first version of the manuscript. All authors edited and approved the latest version of the manuscript.

#### Funding

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported by grants PID2022-136912NB-I00 funded by MCIN/AEI/https://doi.org/10.13039/501100011033 and by "ERDF a way of making Europe" (S.F.E), and CIPROM/2022/59 (S.F.E.) and CIDEGENT/2021/030 (R.L.C.) funded by Generalitat Valenciana. M.J.O.-U. was supported by grant FPU2019/05246 funded by MCIN/AEI/https://doi.org/10.13039/501100011033 and by "ESF investing in your future".

#### Availability of data and materials

Raw Illumina RNA-Seq data generated for this study are available in NCBI SRA under BioProject accession PRJNA974369.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 2 July 2024 Accepted: 12 September 2024
Published online: 30 September 2024

#### References

1. Soosaar JLM, Burch-Smith TM, Dinesh-Kumar SP. Mechanisms of plant resistance to viruses. Nat Rev Microbiol. 2005;3:789–99.
2. Zhou JM, Zhang Y. Plant immunity: danger perception and signaling. Cell. 2020;181:978–89.
3. Carr JP, Lewsey MG, Palukaitis P. Signaling in induced resistance. Adv Virus Res. 2010;76:57–121.
4. Loebenstein G. Local lesions and induced resistance. Adv Virus Res. 2009;75:73–117.
5. Kachroo P, Chandra-Shekara AC, Klessig DF. Plant signal transduction and defense against viral pathogens. Adv Virus Res. 2006;66:161–91.
6. López-Gomollon S, Baulcombe DC. Roles of RNA silencing in viral and non-viral plant immunity and in the crosstalk between disease resistance systems. Nat Rev Mol Cell Biol. 2022;23:645–62.
7. Voinnet O. RNA silencing as a plant immune system against viruses. Trends Genet. 2001;17:449–59.

8. García-Ruiz H, Takeda A, Chapman EJ, Sullivan CM, Fahlgren N, Brempelis KJ, Carrington JC. Arabidopsis RNA-dependent RNA polymerase and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during *turnip mosaic virus* infection. Plant J. 2010;22:481–96.
9. Borges F, Martienssen RA. The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol. 2015;16:727–41.
10. Hung YH, Slotkin RK. The initiation of RNA interference (RNAi) in plants. Curr Opin Plant Biol. 2021;61:102014.
11. Liu P, Cuerda-Gil D, Shahid S, Slotkin RK. The epigenetic control of the transposable element life cycle in plant genomes and beyond. Annu Rev Genet. 2022;56:63–87.
12. Böhmdorfer G, Sethuraman S, Rowley MJ, Krzyszton M, Rothi MH, Bouzit L, Wierzbicki AT. Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. eLife. 2016;5: e19092.
13. Cuerda-Gil D, Slotkin RK. Non-canonical RNA-directed DNA methylation. Nat Plants. 2016;2:16163.
14. Bond DM, Baulcombe DC. Small RNAs and heriTable epigenetic variation in plants. Trends Cell Biol. 2014;24:100–7.
15. Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK. *ROS1*, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell. 2002;111:803–14.
16. Saze H, Shiraihi A, Miura A, Kakutani T. Control of genic DNA methylation by a jmjC domain containing protein in *Arabidopsis thaliana*. Science. 2008;319:462–5.
17. Lu F, Cui X, Zhang S, Liu C, Cao X. JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis. Cell Res. 2010;20:387–90.
18. Lloyd JPB, Lister R. Epigenome plasticity in plants. Nat Rev Genet. 2021;23:55–68.
19. Diezma-Navas L, Pérez-González A, Artaza H, Alonso L, Caro E, Llave C, Ruiz-Ferrer V. Crosstalk between epigenetic silencing and infection by tobacco rattle virus in Arabidopsis. Mol Plant Pathol. 2019;20:1439–52.
20. Leone M, Zavallo D, Venturuzzi A, Asurmendi S. RdDM pathway components differentially modulate *Tobamovirus* symptom development. Plant Mol Biol. 2020;104:467–81.
21. Corrêa RL, Sanz-Carbonell A, Kogej Z, Müller SY, Ambrós S, López-Gomollón S, Gómez G, Baulcombe DC, Elena SF. Viral fitness determines the magnitude of transcriptomic and epigenomic reprograming of defense responses in plants. Mol Biol Evol. 2020;37:1866–81.
22. Navarro R, Ambrós S, Butković A, Carrasco JL, González R, Martínez F, Wu B, Elena SF. Defects in plant immunity modulate the rates and patterns of RNA virus evolution. Virus Evol. 2022;8:veac059.
23. Ambrós S, Olmo-Uceda MJ, Corrêa RL, Elena SF. Phenotypic and genomic changes during *turnip mosaic virus* adaptation to *Arabidopsis thaliana* mutants lacking epigenetic regulatory factors. Evolution. 2024;78:69–85.
24. Ivanov KI, Eskelin K, Lõhmus A, Mäkinen K. Molecular and cellular mechanisms underlying potyvirus infection. J Gen Virol. 2014;95:1415–29.
25. Revers F, García JA. Molecular biology of potyviruses. Adv Virus Res. 2015;92:101–99.
26. Pagán I, Fraile A, Fernández-Fueyo E, Montes N, Alonso-Blanco C, García-Arenal F. *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. Philos Trans R Soc B. 2010;365:1983–95.
27. Agudelo-Romero P, Carbonell P, Pérez-Amador MA, Elena SF. Virus adaptation by manipulation of host's gene expression. PLoS ONE. 2008;3:e2397.
28. Cervera H, Ambrós S, Bernet GP, Rodrigo G, Elena SF. Viral fitness correlates with the magnitude and direction of the perturbation induced in the host's transcriptome: the tobacco etch potyvirus-tobacco case study. Mol Biol Evol. 2018;35:1599–615.
29. Hillung J, García-García F, Dopazo J, Cuevas JM, Elena SF. The transcriptomics o fan experimentally evolved plant-virus interaction. Sci Rep. 2016;6: 24901.
30. Chen CC, Chao CH, Chen CC, Ye SD, Tsai HT, Chang CA. Identification of turnip mosaic virus isolates causing yellow stripe and spot on calla lily. Plant Dis. 2003;87:901–5.
31. Boyes DC, Zayed AM, Ascenzi R, McCaskill MJ, Hoffman NE, Davis KR, Görlach J. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. Plant Cell. 2001;13:1499–510.
32. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.
33. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013, 1303.3997v2. https://doi.org/10.48550/arXiv.1303.3997.
34. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10: giab008.
35. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
37. Love MI, Huber W, Anders S. Moderated estimation of old change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
38. Coenye T. Do results obtained with RNA-sequencing require independent verification? Biofilm. 2021;3: 100043.
39. Everaert C, Luypaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, Mestdagh P. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci Rep. 2017;7:1559.
40. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurements sets. Nucleic Acids Res. 2021;49:D605-12.
41. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation. 2021;2:100141.
42. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9: e1003118.
43. Martínez F, Carrasco JL, Toft C, Hillung J, Giménez-Santamarina S, Yenush L, Rodrigo G, Elena SF. A binary interaction map between turnip mosaic virus and *Arabidopsis thaliana* proteomes. Commun Biol. 2023;6:28.
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
45. Zemach A, Ki MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. The Arabidopsis nucleosome remodeler DDM1 allows DNA methylatransferases to access H1-containing heterochromatin. Cell. 2013;153:193–205.
46. Corrêa RL, Kutnjak D, Ambrós S, Bustos M, Elen SF. Identification of epigenetically regulated genes involved in plant-virus interaction and their role in virus-triggered induced resistance. BMC Plant Biol. 2024;24:172.
47. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013;152:352–64.
48. Greenberg MVC, Deleris A, Hale CJ, Liu A, Feng S, Jacobsen SE. Interplay between active chromatin marks and RNA-directed DNA methylation in *Arabidopsis thaliana*. PLoS Genet. 2013;9: e1003946.
49. Singh A, Sharma A, Singh N, Nandi AK. MTO1-RESPONDING DOWN 1 (MRD1) is a transcriptional target of OZF1 for promoting salicylic acid-mediated defense in Arabidopsis. Plant Cell Rep. 2022;41:1319–28.
50. Akhter S, Uddin MN, Jeong IS, Kim DW, Liu XM, Bahk JD. Role of AtPI4K-gamma3, a type II phosphoinositide 4-kinase, in abiotic stress responses and floral transition. Plant Biotechnol J. 2015;14:215–30.
51. Blevins T, Pontvianne F, Cocklin R, Podicheti R, Chandrasekhara C, Yerneni S, Braun C, Lee B, Rusch D, Mockaitis K, Tang H, Pikaard CS. A two-step process for epigenetic inheritance in Arabidopsis. Mol Cell. 2014;54:30–42.
52. Kurihara Y, Matsui A, Kawashima M, Kaminuma E, Ishida J, Morosawa T, Mochizuki Y, Kobayashi N, Toyoda T, Shinozaki K, Seki M. Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. Biochem Biophys Res Commun. 2008;376:553–7.
53. Melero I, González R, Elena SF. Host developmental stages shape the evolution of a plant RNA virus. Philos Trans R Soc B. 2023;378:20220005.
54. Liu J, Kim BM, Kaneko Y, Inukai T, Masuta C. Identification of the TuNI gene causing systemic necrosis in Arabidopsis ecotype L*er* infected with *turnip mosaic virus* and characterization of its expression. J Gen Plant Pathol. 2015;81:180–91.

55. Tsuchiya T, Eulgem T. An alternative polyadenylation mechanism coopted to the Arabidopsis *RPP7* gene through intron retrotransposon domestication. Proc Natl Acad Sci USA. 2013;110:E3535-43.
56. Sasaki T, Ro K, Caillieux E, Manabe R, Bohl-Viallefond G, Baduel P, Colot V, Kakutani T, Quadrana L. Fast co-evolution of anti-silencing systems shapes the invasiveness of Mu-like DNA transposons in eudicots. EMBO J. 2022;41:e110070.
57. Barro-Trastoy D, Köhler C. Helitrons: genomic parasites that generate developmental novelties. Trends Genet. 2024;40:437–48.
58. Coates BS. Horizontal transfer of a non-autonomous helitron among insect and viral genomes. BMC Genomics. 2015;16:137.
59. Wilkinson SW, Hannan Parker A, Muench A, Wilson RS, Hooshmand K, Henderson M, Moffat EK, Rocha PSCF, Hipperson H, Stassen JHM, López Sánchez A, Fomsgaard IS, Krokene P, Mageroy MH, Ton J. Long-lasting memory of jasmonic acid-dependent immunity requires DNA demethylation and ARGONAUTE1. Nat Plants. 2023;9:81–95.
60. Carrasco JL, Ambrós S, Gutiérrez PA, Elena SF. Adaptation of turnip mosaic virus to *Arabidopsis thaliana* involves rewiring of VPg-host proteome interactions. Virus Evol. 2024;10:veae055.
61. Franek M, Nešpor Dadejová M, Pírek P, Kryštofová K, Dobisová T, Zdráhal Z, Dvořáčková M, Lochmonová G. Histone chaperone deficiency in Arabidopsis plants triggers adaptive epigenetic changes in histone variants and modifications. Mol Cell Proteom. 2024;23:100795.
62. Zafirov D, Giovinazzo N, Lecampion C, Field B, Ducassou JN, Couté Y, Browning KS, Robaglia C, Gallois JL. Arabidopsis eIF4E1 protects the translational machinery during TuMV infection and restricts virus accumulation. PLoS Pathog. 2023;20:e1011417.
63. Chai M, Li L, Li Y, Yang Y, Wang Y, Jiang X, Luan Y, Li F, Cui H, Wang A, Xiang W, Wu X, Cheng X. The 6-kilodalton peptide 1 in plant viruses of the family *Potyviridae* is a viroporin. Proc Natl Acad Sci USA. 2024;121:e2401748121.
64. Xia X, Cheng Z, Wang M, Ou X, Sun D, Mao S, Huang J, Yan Q, Wu Y, Chen S, Zhang S, Zhu D, Jia R, Liu M, Zhao XX, Gao Q, Tian B. Functions of viroporins in the viral life cycle and their regulation of host cell responses. Front Immunol. 2022;13:890549.
65. Nieva JL, Madan V, Carrasco L, Viroporins. Structure and biological functions. Nat Rev Microbiol. 2012;10:563–74.
66. To J, Surya W, Torres J. Targeting the channel activity of viroporins. Adv Prot Chem Struct Biol. 2016;104:307–55.
67. Cui H, Wang A. Plum pox virus 6K1 protein is required for viral replication and targets the viral replication complex at the early stage of infection. J Virol. 2016;90:5119–31.
68. Ferrari S, Plotnikova JM, de Lorenzo G, Ausubel FM. Arabidopsis local resistance to Botrytis cinerea involves salicylic acid and camalexin and requires EDS4 and PAD2, but not SID2, EDS5 or PAD4. Plant J. 2003;35:193–205.
69. Kavil S, Otti G, Bouvaine S, Armitage A, Maruthi MN. *PAL1* gene of the phenylpropanoid pathway increases resistance to the *Cassava brown streak virus* in cassava. Virol J. 2021;18:184.


## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# CHAPTER 5

## Accumulation dynamics of defective genomes during experimental evolution of two betacoronaviruses

viruses

MDPI

*Article*

# Accumulation Dynamics of Defective Genomes during Experimental Evolution of Two Betacoronaviruses

**Julia Hillung [1,†], María J. Olmo-Uceda [1,†], Juan C. Muñoz-Sánchez [1] and Santiago F. Elena [1,2,*]**

[1] Instituto de Biología Integrativa de Sistemas (I2SysBio), CSIC-UV, Catedrático Agustín Escardino Benlloch 9, 46980 Paterna, Valencia, Spain; julia.hillung@csic.es (J.H.); mariajose.olmo@csic.es (M.J.O.-U.); jc.munoz@csic.es (J.C.M.-S.)
[2] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
* Correspondence: santiago.elena@csic.es; Tel.: +34-963-544-779
† These authors contributed equally to this work.

**Abstract:** Virus-encoded replicases often generate aberrant RNA genomes, known as defective viral genomes (DVGs). When co-infected with a helper virus providing necessary proteins, DVGs can multiply and spread. While DVGs depend on the helper virus for propagation, they can in some cases disrupt infectious virus replication, impact immune responses, and affect viral persistence or evolution. Understanding the dynamics of DVGs alongside standard viral genomes during infection remains unclear. To address this, we conducted a long-term experimental evolution of two betacoronaviruses, the human coronavirus OC43 (HCoV-OC43) and the murine hepatitis virus (MHV), in cell culture at both high and low multiplicities of infection (MOI). We then performed RNA-seq at regular time intervals, reconstructed DVGs, and analyzed their accumulation dynamics. Our findings indicate that DVGs evolved to exhibit greater diversity and abundance, with deletions and insertions being the most common types. Notably, some high MOI deletions showed very limited temporary existence, while others became prevalent over time. We observed differences in DVG abundance between high and low MOI conditions in HCoV-OC43 samples. The size distribution of HCoV-OC43 genomes with deletions differed between high and low MOI passages. In low MOI lineages, short and long DVGs were the most common, with an additional cluster in high MOI lineages which became more prevalent along evolutionary time. MHV also showed variations in DVG size distribution at different MOI conditions, though they were less pronounced compared to HCoV-OC43, suggesting a more random distribution of DVG sizes. We identified hotspot regions for deletions that evolved at a high MOI, primarily within cistrons encoding structural and accessory proteins. In conclusion, our study illustrates the widespread formation of DVGs during betacoronavirus evolution, influenced by MOI and cell- and virus-specific factors.

**Keywords:** defective genomes; evolutionary dynamics; experimental evolution; mutant swarm; virus evolution

check for updates

**Citation:** Hillung, J.; Olmo-Uceda, M.J.; Muñoz-Sánchez, J.C.; Elena, S.F. Accumulation Dynamics of Defective Genomes during Experimental Evolution of Two Betacoronaviruses. *Viruses* **2024**, *16*, 644. https://doi.org/10.3390/v16040644

Academic Editor: Eric O. Freed

Received: 21 March 2024
Revised: 18 April 2024
Accepted: 19 April 2024
Published: 20 April 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Owing to their high mutation rates, rapid replication, and large population sizes, RNA viruses quickly adapt to novel hosts and trigger epidemics by crossing species barriers [1–3]. However, many mutations during infection lead to aberrant genomes that cannot complete the infectious cycle. In vitro, these incomplete genomes, known as defective viral genomes (DVGs), accumulate in virus populations when the virus is passaged repeatedly at a high multiplicity of infection (MOI) [4–6]. Some DVGs can be encapsidated and transmitted. When co-infecting with a wild-type virus, in some cases, DVGs may interfere with virus replication [7–9]. In this case, the term defective interfering particle (DIP) is used to refer to encapsidated DVGs that negatively impact wild-type virus replication. This interference occurs on multiple levels, including resource competition, immune response activation, and sometimes, the survival of infected cells [5,8–10]. Although specifics of DIP generation

and regulation in RNA viruses are still not fully understood, their role in modulating viral infections is well-established [11,12].

Deletions are the most common type of DVGs among RNA viruses [13,14]. They are thought to form through homologous recombination due to observed homology in specific regions and RNA structures [15–17]. Copy-back (cb) and snap-back (sb) genomes are also prevalent DVG classes in negative-sense RNA viruses. These genomes have loops with complementary ends [18,19].

Coronaviruses have the largest known RNA virus genomes, ~30 kb in length [20]. They encode nonstructural proteins involved in viral RNA synthesis and interact with host cell functions. Replication generates a set of subgenomic mRNAs, all sharing common 5′ leader and a 3′ terminal sequences [21]. Specific regions, such as the leader sequence and a portion of the replicase gene, contribute to efficient replication and transmission. Additionally, *cis*-acting elements like the transcriptional regulatory sequences (TRS) are crucial for transcription [22]. In most described coronavirus DIPs, the combination of both internal and terminal genome sequences enables the formation of secondary and higher-order structures essential for their function as replication signals [23]. Deletion mapping of murine hepatitis virus (MHV) DIPs reveals the requirement of an internal and discontinuous sequence for replication [24]. The 3′ proximal coding regions, including the *N* cistron, can tolerate substantial alterations, suggesting they are not part of the 3′ *cis*-acting elements [25–28]. Efficient genome packaging into virions requires specific RNA signals, primarily located at the terminal sequences [29–33].

To investigate the dynamics of DVG accumulation in betacoronaviruses infecting highly susceptible cells, we conducted an evolution experiment by performing undiluted passages of (*i*) the human coronavirus OC43 (HCoV-OC43) in baby hamster kidney cells (BHK-21) and human large intestine carcinoma cells (HCT-8), and (*ii*) of MHV in murine liver cells (CCL-9.1). We anticipated that a high MOI would promote the accumulation of DVGs over passages due to frequent co-infection with wild-type viruses. As a control, we also carried out parallel evolution experiments at a low MOI. We replicated both MOI conditions and virus–cell type combinations three times. After experimental evolution, we employed high-throughput RNA-seq and bioinformatic tools to reconstruct DVG populations in the ancestral virus and in samples at four equidistant evolutionary passages. We did not specifically evaluate whether the accumulated DVGs acted as DIP, but focused on describing their temporal dynamics and potential mechanisms of generation.

## 2. Materials and Methods

### *2.1. Viruses and Cells*

Human large intestine carcinoma cells (HCT-8; ATCC® CCL-244TM) and murine liver cells NCTC 1469 (CCL-9.1; ATCC® CCL-9.1TM) were obtained from the American Type Culture Collection (ATCC). Baby hamster kidney cells (BHK-21) were kindly provided by Dr. R. Sanjuán ($I^2$SysBio, Valencia, Spain). BHK-21 cells and HCT-8 cells were cultured at 37 °C in DMEM (Gibco-Invitrogen, Grand Island, NY, USA) supplemented with 0.22% (*w*/*v*) sodium bicarbonate (Sigma Aldrich, Saint Louis, MO, USA), sodium pyruvate (Sigma Aldrich), 10% fetal bovine serum (FBS), 1× penicillin-streptomycin (Gibco-Invitrogen), 1× amphotericin B (Gibco-Invitrogen), and non-essential amino acids using standard laboratory procedures. CCL-9.1 cells were cultured at 37 °C in DMEM media supplemented with 0.22% (*w*/*v*) sodium bicarbonate (Sigma Aldrich), sodium pyruvate (Sigma Aldrich), 10% horse serum, 1× penicillin-streptomycin, 1× amphotericin B, and non-essential amino acids. All cell lines were routinely tested for mycoplasma contamination.

The viruses used in this study were HCoV-OC43 (ATCC® VR-1558) and MHV (ATCC® VR-766), both obtained from ATCC. The maintenance medium for HCoV-OC43 was DMEM supplemented with 0.22% (*w*/*v*) sodium bicarbonate (Sigma Aldrich), sodium pyruvate (Sigma Aldrich), 2% FBS, 1× penicillin-streptomycin (Gibco-Invitrogen), and 1× amphotericin B (Gibco-Invitrogen). Cell culture media of CCL-9.1 was also used as maintenance medium by infection with MHV.

### 2.2. Experimental Evolution

To generate the initial virus inoculum, BHK-21 and HCT-8 cells were plated one day before and grew to 70–80% confluency. Afterward, the cell layers were infected with the corresponding virus at a MOI of 0.01. HCoV-OC43 was grown for 4 d at 33 °C, and MHV was grown for 20 h at 37 °C. The supernatants from infected cells were then collected and used as the infectious stock. For successive passages, all cell types were grown in 6-well plates to 70–80% confluency and infected with 250 µL of undiluted viral inoculum from the previous passage. HCoV-OC43 was passaged 47 times in BHK-21 cells and 31 times in HCT-8 cells, while MHV was passaged 19 times in CCL-9.1 cells.

The MOI was not maintained the same throughout all passages; instead, at high MOI, we infected every passage without dilution to get the highest MOI possible. For low MOI samples, the inocula were diluted to avoid infecting each cell with more than one infectious particle. Table S1 shows the range of MOIs for each virus and cell type.

### 2.3. Viral Load Quantification

The quantity of infectious virus in the supernatant was determined after each passage. To determine the infectious viral particles for HCoV-OC43, plaque assays were performed in BHK-21. The procedure involved seeding the cells into 6-well plates and incubating them for 24 h until reaching 80% confluency. Serial dilutions of the virus were prepared in DMEM, and 250 µL of each dilution was used to infect the wells of the 6-well plate for 90 min at 33 °C and 5% $CO_2$ with swinging every 15 min. A negative control consisting of only DMEM media without virus was included for comparison. After the infection period, the cells were covered with 2 mL of media, which consisted of a 50:50 mix of 2× DMEM supplemented with 2% FBS and 2% agar. The agar–DMEM layer was overlaid with liquid DMEM supplemented with 1% FBS as it solidified. The plates were then incubated at 33 °C and 5% $CO_2$ for 5 d. After incubation, cells were fixed using 10% formaldehyde and plugs were removed. Monolayers were stained with 2% crystal violet in 10% formaldehyde, washed with tap water, and plaques counted to determine the plaque-forming units (PFU) per mL of inoculum.

The infectious viral particles for MHV were determined by $TCID_{50}$. Cells were seeded into 96-well plates and incubated for 24 h to reach 90% confluence. Viral transfer samples were diluted in complete DMEM. Once diluted, 125 µL of the dilutions for each sample were used to infect the wells of the 96-well plate. A negative control consisting of only complete DMEM was also included. The plates were then incubated at 37 °C for 2 d. After incubation, the wells were inspected under a light microscope for the presence of visual cytopathic effect, and the $TCID_{50}$/mL was calculated using the Reed and Muench method [34]. For comparison purposes with plaque assay results, $TCID_{50}$/mL values were converted into PFU/mL by multiplying by $\ln(0.5) \approx 0.693$.

### 2.4. Preparation of Viral RNA

Samples from four equidistant time series of each evolutionary lineage were collected to extract RNA using the NZY Viral RNA Isolation kit (NZYTech, Lisboa, Portugal) following the manufacturer's instructions. After extraction, the RNA was purified and concentrated using the RNA Clean&Concentrator-5 Kit (Zymo Research, Irvine, CA, USA).

### 2.5. High-Throughput RNA Sequencing

Libraries for high-throughput genome sequencing (HTS) were prepared by Novogene UK Company Ltd., Cambridge, UK (www.novogene.com; accessed on 17 April 2024) using at least 200 ng of total RNA or a larger amount if available. The ribosomal RNAs (rRNA) from both eukaryotes and prokaryotes were depleted from the total RNA samples. The remaining RNAs were fragmented into 250–300 bp fragments and reverse-transcribed into double-stranded cDNAs, followed by end repair, A tailing, and adapter ligation. After fragment size selection and PCR amplification, the metatranscriptome library was checked for quality, and sequencing was conducted on the Illumina NovaSeq6000 platform

(Illumina, San Diego, CA, USA). The raw data obtained from the samples ranged between 3 GB and 15 GB, providing ~64–124,981 × viral nucleotide mean depth, with a median of 13,362.52. The sequencing was not strand-specific.

### *2.6. Processing of HTS Reads and Identification of DVGs*

Raw paired reads from Illumina metagenomic data from each sample were merged into a single interleaved file using the reformat.sh script from the BBMap/BBTool package (https://github.com/BioInfoTools/BBMap/blob/master/sh/reformat.sh (accessed on 17 April 2024)). The interleaved reads were then processed using the BBduk.sh tool to remove adapter sequences, clean, and trim the reads to obtain high-quality nucleotides at both ends (https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh (accessed on 17 April 2024)). Next, the remaining reads were mapped to the corresponding host genomes: the hamster genome for BHK-21 cell lineages, the human genome for HCT-8, and the mouse genome for CCL-9.1, using the BWA-MEM algorithm [35]. Reads that mapped with the host genomes were excluded from further analysis. The resulting reads were analyzed using the metasearch tool DVGfinder to identify and classify DVG events. Predicted DVG species sizes were calculated based on the breaking or start points (BP) and rejoining or end points (RP) by the program [36]. Reference genomes used were ATCC VR-1558 for HCoV-OC43 and GU593319.1 for MHV.

The raw sequencing data were deposited in GenBank (NCBI) under BioProject number PRJNA1021788.

### *2.7. Processing and Diversity Measures of DVG*

Only the results from DVGfinder Filtered mode were used. In the case where both possible DVG directions were present in the original results, the sum of their abundance values was used, and BP and RP lost their original meaning. The canonic sgRNAs were identified and discarded from the rest of the analysis. A deletion-type DVG was identified as sgRNAs when its start coincides with the TRS-leader and its end with one of the canonical TRS-backs; both coordinates have a flexible margin of ten bases.

Richness was understood as the number of different DVGs relative to the total number of mapped viral reads of the group (sample or lineage, depending on the case) per $10^5$ (reads per hundred thousand; RPHT). Abundance was used as a measure of the number of reads supporting the DVG, also expressed as RPHT.

### *2.8. Identification of Transmitted DVGs*

To study the possible transmission of DVGs between passages, each DVG was labeled as persistent if the exact same event (DVG type with same BP and RP coordinates) was identified in a previous passage of the same lineage (including the stock).

### *2.9. Characterization of Deletion Type DVGs*

All the DVGs with deletions bigger than three bases were analyzed from different perspectives: their length, their location in the genome, their distance to the nearest TRS, their nucleotide composition, and their secondary structure. The cistrons that were more affected by deletions were also identified. Smaller DVGs were discarded to minimize known Illumina error profiles [37].

#### 2.9.1. Deletion Size Distribution

The distribution of the reconstructed size of deletion type DVG was calculated considering the abundance of each species in the sample. For a better visualization of the frequency coordinates, only deletions bigger than 50 bases were represented. The size of the deleted fragments was analyzed by a Lomb–Scargle periodogram method [38].

#### 2.9.2. Distribution of Deletion Breaking (BP) and Rejoining (RP) Points in the Viral Genomes

For every cell type, a histogram of the start and end position of deletions was represented. The data was generated with the method hist from the base package graphics, setting the breaks argument to the quotient of the viral length between 50.

#### 2.9.3. K-Mer Analysis

The *k*-mer composition of the sequences ten bases pre- and post- BP and RP of deletions was compared with the *k*-mer composition of the complete genome ($k$ = 1 to 3) with a $\chi^2$ test (chisq.test from base package stats, with rescale.p = T). The *k*-mer composition was analyzed with the oligonucleotideFrequency method from Biostrings v2.66.0 [39].

#### 2.9.4. Minimum Free Energy (MFE) around Junctions

The MFE of the structures predicted with RNAfold v2.5.1 [40] for 50-base sliding windows were used to perform a Spearman correlation analysis with the frequencies for 50-base windows of being start or end coordinates of a deletion calculated previously. RNA2Drawer [41] was used for representation of the obtained structures.

#### 2.9.5. Deletions per Base Calculation

To evaluate the viral cistrons more affected by deletions, the findOverlaps method from GenomicRanges v1.50.2 [42] was used in a homemade script in R version 4.3.2. The number of deletions was relativized to the cistron length.

### *2.10. Other Statistical Analyses*

Except otherwise indicated, statistical analyses were done with SPSS version 28.0.1.0 (IBM, Armonk, NY, USA).

Viral load time series for individual lineages were fitted to the general ARIMA($p$, $d$, $q$) model:

$$\left(1 - \sum_{i=1}^{p} \rho_i L^i\right)(1-L)^d V_t = \left(1 + \sum_{j=1}^{q} \mu_j L^j\right)\varepsilon_t \quad (1)$$

where $V_t$ represents the viral load at passage $t$; $L$ is the lag operator such that $L^n V_t = V_{t-n}$; $\rho_i$ are the parameters of the autoregressive part of the model with $p \geq 0$ being the order of the autoregressive model (number of time lags); $\mu_i$ are the parameters of the moving average part with $q \geq 0$ being the order of the moving-average model; $d \geq 0$ is the differencing order, representing the number of times the time series is differenced to achieve stationarity; and $\varepsilon_t$ are the error terms assumed to be independent identically distributed N(0, $\sigma^2$). ARIMA fitting and model selection were done using the package forecast in R version 4.3.2 in RStudio version 2023.12.1+402. Estimated slope values were further analyzed by runs tests or pairwise Mann–Whitney tests.

Diversity and abundance of DVG classes were analyzed using the Scheirer–Ray–Hare (SRH) non-parametric two-way analysis of variance. In all tests done, the specific factors analyzed (i.e., virus species, cell type, or MOI) were orthogonal to the class of DVG (i.e., deletions, insertions, 3′, and 5′ cb).

Other specific statistical tests are presented in the text as needed.

### *2.11. Code Implementation*

All the code developed for this work is accessible in https://github.com/MJmaolu/AccumulationDynamicsDVGs/tree/main (accessed on 17 April 2024). Intensive computations were run on the HPC cluster Garnatxa at I2SysBio (CSIC-UV). The R scripts were run in Rstudio version 4.2.1 (23 June 2022).

## 3. Results

### *3.1. Experimental Evolution of Betacoronavirus DVGs Populations*

Viral load was evaluated at each passage, and dilutions were adjusted to ensure that, on median, MOIs remained within two wide but disjoint intervals that can broadly be defined as higher and lower MOIs, respectively (Table S1). Notice that what is defined as high or low MOIs depends on the particular combination of virus and cell type, and is not an absolute but a relative term. What is relevant is that, in all cases, the $\log_2$-fold difference between high and low MOIs was consistently in the range of 10–20.

The time-series data for viral load are shown in Figure 1. Individual lineage series were fitted to ARIMA models and the best-fitting one was determined according to the minimum BIC criterium (Table S2). In each model, the slope parameter defines the rate of evolution per passage. The average rates of evolution for each combination of factors are also displayed in Table S2. In the case of HCoV-OC43 in BHK-21 cells, the average rate was significantly positive at a high MOI (runs test, $p < 0.001$) but not different from zero at a low MOI (runs test, $p = 1.000$). On average, rates were 13% faster at a high MOI. In the case of HC-OC43 in the HCT-8 cell, both rates of evolution were significantly negative (runs tests, $p < 0.001$ in both cases). Rates were, on average, 47% faster at a high MOI, although the difference was not significant (Mann–Whitney, $p = 0.719$). Finally, MHV in CCL-9.1 cells showed no difference from zero at a high MOI (runs test, $p = 1.000$) but were significantly negative at a low MOI (runs test, $p < 0.001$), declining 567% faster at a low MOI (Mann–Whitney, $p < 0.001$).

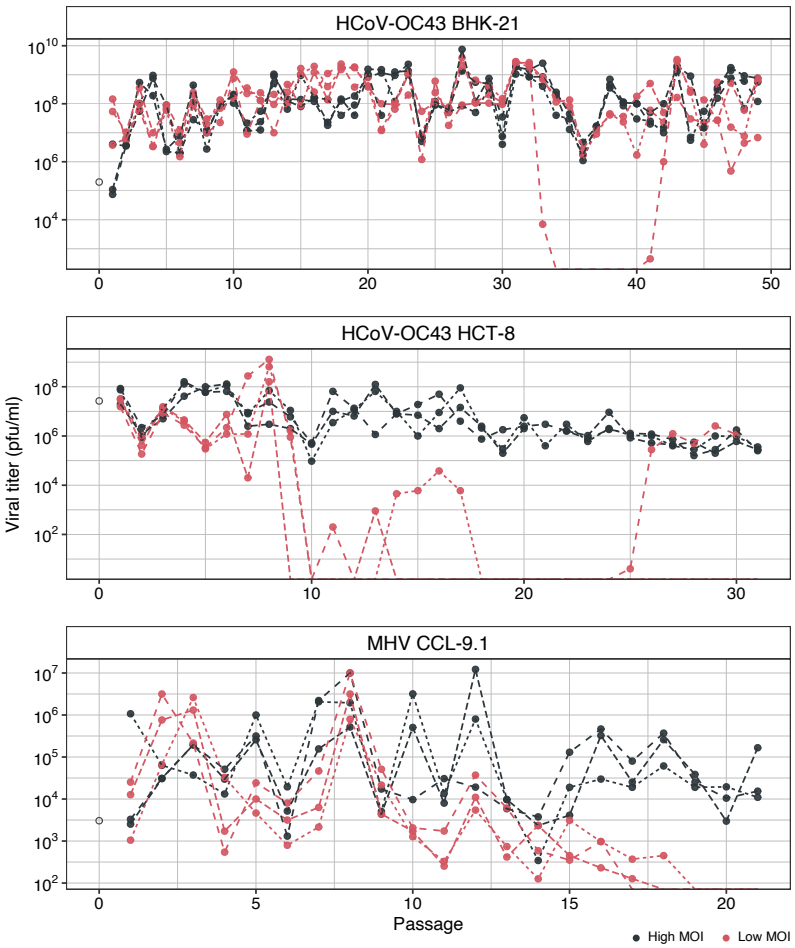



**Figure 1.** Evolution of viral loads at high and low MOIs. (**Upper panel**) HCoV-OC43 in BHK-21. (**Middle panel**) HCoV-OC43 in HCT-8. (**Lower panel**) MHV in CCL-9.1. Black dots and lines: high MOI; red: low MOI. Empty dot: ancestral virus (stock).

### *3.2. Evolution of DVG Richness and Abundance Is Cell Type and Virus Dependent*

Since the sequencing was not strand-specific, in the following analyses we will not distinguish between DVGs found in the negative- or positive-sense RNA strains. In addition, prior to all the evolutionary relevant analyses, canonical subgenomic RNAs (sgRNAs) were filtered out by excluding any reconstructed DVGs whose junctions agreed with the expected coordinates of sgRNAs within a range of $\pm 10$ nucleotides. The percentages of sgRNAs observed in each evolving lineage are shown in Figure S1. Interestingly, these percentages were affected by MOI, virus species, and host cell type. In the case of HCoV-OC43, 51.35% less sgRNAs relative to total DVGs were observed at a high MOI than at a low MOI in BHK-21. This reduction was 41.44% in HCT-8. These observations suggest that, for a given amount of sgRNAs, more DVGs were accumulating at high MOIs. In contrast, the trend was the opposite for MHV in CCL-9.1: 19.26% more sgRNAs per total DVGs were present at high MOIs than at low MOIs.

Next, we examined the dynamics of DVG generation and accumulation during the evolution experiments. Figure 2 shows two different measures of DVG diversity over time. Firstly, the number of unique DVGs was expressed as RPHT (Figure 2A), which represents diversity in terms of the different DVG types observed. Secondly, the total RPHT per DVGs (Figure 2B) was a measure of the total amount of DVGs relative to absolute viral accumulation. In all lineages, a common trend was observed: an increase in DVG richness from the initial to intermediate stages of viral evolution (Figure 2A), with no significant effect of MOI (Wilcoxon paired-samples tests, $p \geq 0.382$ in all three cases). Throughout the course of the evolution experiment, some disparities were noted between parallel replicates, but the overall pattern for HCoV-OC43 in both cell types was a reduction in richness after the initial surge, maintaining a relatively stable level until the end of the experiment, with no discernible differences observed between both cell types. For MHV at a high MOI, DVG richness increased up to passage nine and then remained stable until the end of the experiment. However, at a low MOI, MHV displayed differences between parallel lineages, with the number of DVG reads either increasing or decreasing after passage nine, depending on the lineage. The shifts in richness at different passages were more remarkable for HCoV-OC43 in BHK-21 and for MHV in CCL-9.1 cells.

Furthermore, the abundance of DVG counts increased to varying degrees at the onset of evolution in all lineages (Figure 2B). After the initial passages in low MOI lineages, the abundance of detected DVGs tended to decline, while in high MOI lineages, DVG abundance remained relatively steady or even slightly increased. Interestingly, notable disparities were noted in the accumulation of DVGs between high and low MOIs for HCoV-OC43 in both cell types (Wilcoxon paired-samples tests, $p = 0.022$ and $p = 0.041$, respectively), with DVGs consistently being more abundant in lineages evolved at a high MOI. In contrast, no significant differences in terms of the total number of accumulated DVGs were observed for MHV lineages evolved at high and low MOIs (Wilcoxon paired-samples test, $p = 1.000$).

Next, we tested whether DVG richness and abundance were correlated, that is, whether more abundance would contain more different DVG types. The alternative being abundance resulting from the accumulation of a few different DVGs. Spearman correlations were significant for HCoV-OC43 infecting BHK-21 at a low MOI ($r_S = 0.882$, 13 d.f., $p = 0.001$) as well as for MHV at both MOIs (low: $r_S = 0.900$, 10 d.f., $p = 0.002$; high: $r_S = 0.776$, 13 d.f., $p = 0.005$) but not in the other three cases. These observations suggest that viral populations containing more DVGs may also contain more diverse types.

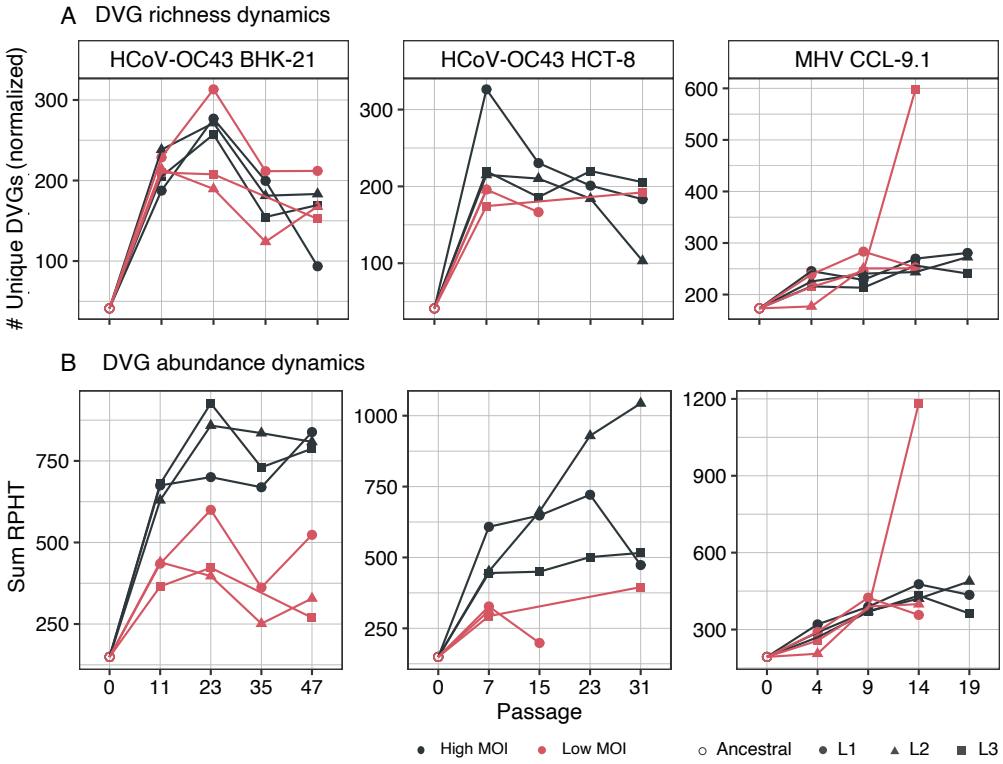



**Figure 2.** Evolution of richness and abundance of DVGs. (**A**) DVGs richness, as the number of unique DVGs relative to the total viral RPHT. (**B**) DVGs abundance, estimated as the total number of DVG read counts relative to the total viral reads in the sample, also RPHT. Black color represents high MOI while red represents low MOI. Different lineages are presented by different symbols. When a lineage was extinct, no point is shown.

### *3.3. Deletions Are the Most Pervasively Reconstructed Type of DVG*

DVG classes considered in the reconstruction include deletions, insertions, and cb or sb genomes (hereafter all referred to as cb) at 5′ or 3′ ends. Richness was evaluated as in Section 3.2 but now distinguishing between the four types of DVGs; estimates are shown in Figure 3A. Firstly, we asked whether the two viruses show differences in the number of unique DVGs types. A Scheirer–Ray–Hare (SRH) two-way non-parametric ANOVA test showed highly significant differences between the two viruses ($p < 0.001$), with HCoV-OC43 generating, on average, 47% more different DVGs than MHV. Likewise, the overall distribution of different DVGs also differed among the four classes of DVGs, with deletions being the most abundant type, followed by insertions and with 3′ and 5′ cb being the least common ($p < 0.001$). Interestingly, these differences among the four classes were consistent for the two viruses ($p = 0.442$). Secondly, we interrogated the HCoV-OC43 data for differences among cell types. In this case, the SRH test found no differences in the number of different DVG types among the two cells ($p = 0.093$), although the distribution of the four DVG types was strongly biased by deletions and insertions ($p < 0.001$). Thirdly, we tested whether MOI had an overall effect on the number of different DVGs generated. In this case, the SRH test also found significant differences ($p < 0.001$) between lineages evolved at high and low MOIs, with the former containing 74% more variants than the latter. The rank-order of the four classes of DVGs was not affected by MOI ($p = 0.894$).

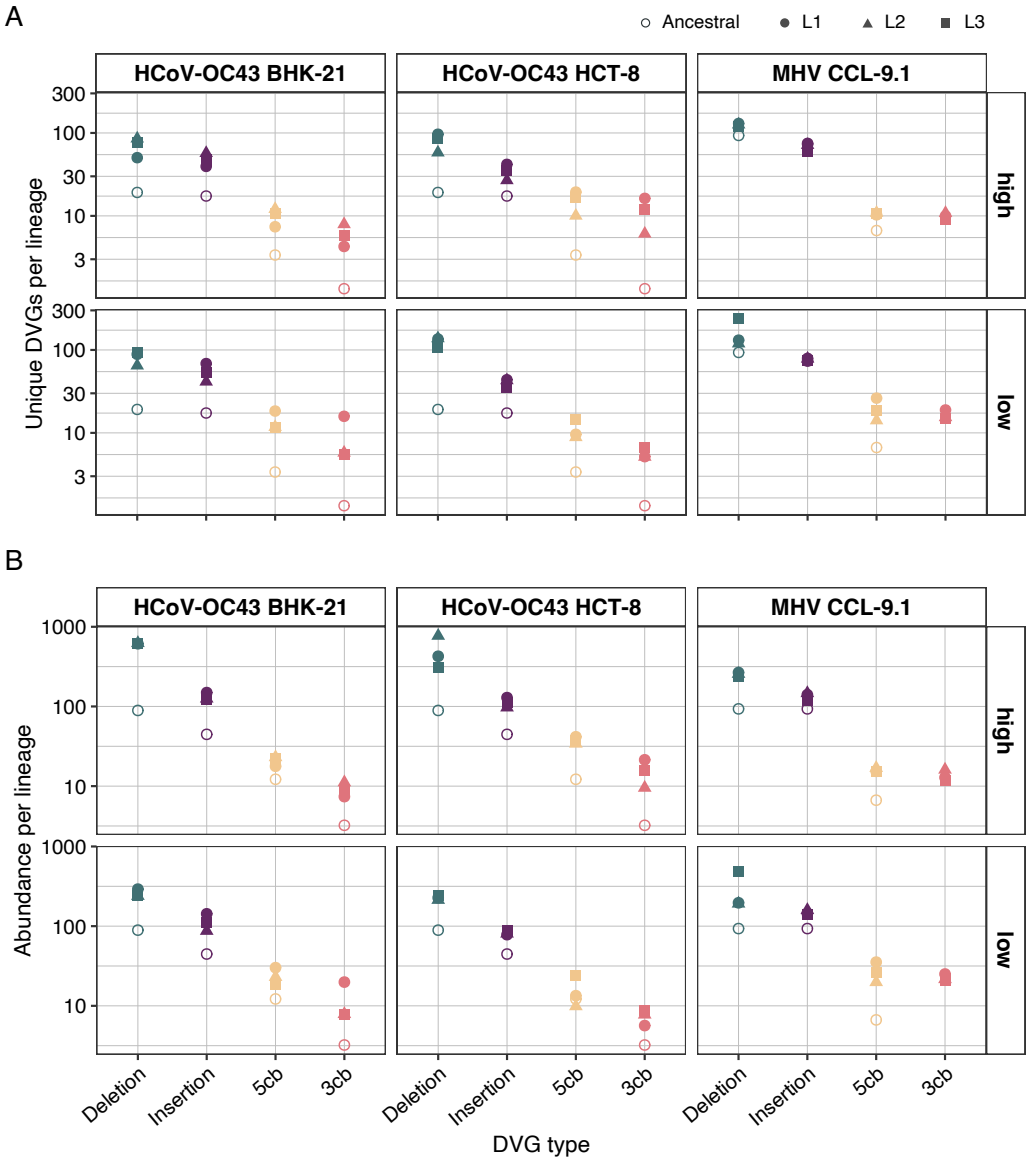



**Figure 3.** Diversity of DVG classes detected in each lineage of experimental evolution. (**A**) Richness represented as the normalized number of unique DVGs present in each lineage. (**B**) Abundance represented as the normalized number of reads supporting the DVG founded in each lineage. Empty points represent the stock sample (passage 0), the values for the stock in HCoV-OC43 are the same for BHK-21 and HCT-8 at both MOIs. The MHV ancestral is the same for both MOIs in CCL-9.1.

Abundance was evaluated as in Section 3.2 but now distinguishing between DVG types; estimates are shown in Figure 3B. Following the same logic as in the previous paragraph, we firstly sought for an effect of virus species on DVG abundance. In this case, the SRH test found no significant differences among the two viruses in the abundance of DVGs ($p = 0.575$). Overall, deletions were the most variable class, followed by insertions and 5′ and 3′ cb (these two at approximately the same variability) ($p < 0.001$); these differences were not affected by the viral species ($p = 0.093$). Secondly, we found no significant differences among the two cell types in the abundance of DVGs generated from the HCoV-OC43 genome ($p = 0.835$), nor an effect on the distribution of the four classes of DVGs ($p = 0.994$). Finally, differences in MOI had an overall effect on DVGs abundance, being 35% more abundant in lineages evolved at a high MOI ($p < 0.001$), although this effect did not translate into a change in the proportion of the four classes of DVGs ($p = 0.856$).

### *3.4. The Distribution of Deletion Sizes Suggests That Purifying Selection upon DVG Size Depends on Particular Virus–Cell Type Combinations*

Next, we honed in on deletions for our analysis, as they were the most common type of DVGs. We examined the sizes of defective genomes with deletions at each sampled time point. One-base deletions were predominant in all the lineages. Thus, to uncover the pattern of remaining deletion sizes, we filtered out DVGs with deletions shorter than 50 nucleotides.

The evolution of the distribution of deletion sizes, accounting for their corresponding abundance, is shown in Figure 4A. The size distribution of HCoV-OC43 genomes with deletions differed between high and low MOI passages in both cell types, with bigger deletions being more common at a high MOI (Figure 4A, black bars). Notably, certain deletion sizes at a high MOI became predominant throughout the evolution (BHK-21: ~12.4 kb, ~22.5–30.7 kb; HCT-8: ~13.6 kb). HCoV-OC43 evolved to cell-specific size distributions, although some lengths were common to both cell types. At a low MOI (red bars in Figure 4A; for a more detailed presentation, see Figure S2), large deletion sizes ~30–31 kb are predominant for HCoV-OC43, followed by ~18 kb in BHK-21 and ~18–20 kb in HCT-8 cells. The 30,698-nucleotide-long DVG is widely present across all the lineages and MOIs of HCoV-OC43 infected cells but it was also present in the viral stock.

For MHV, there were also differences in the distribution of deletion sizes between high and low MOIs, although it is less obvious than what was shown for HCoV-OC43. The size of the most frequent deletions at a high MOI were $\geq$17 kb (Figure 4A, black bars), with varying maxima to different time points along viral evolution. The last passage showed a clear predominance of the longest deletions (~31.1 kb, 19.1 kb). However, the distribution of deletion sizes at a low MOI lacked a clear abundance pattern (Figure S2).

From these analyses, we can conclude that although the generation of short deletions is more likely, the occurrence of very long deletions is also common. The size distribution of deletions is highly dynamic and underwent changes throughout the course of the evolution experiment, with certain lengths being pervasively favored. This dynamism is influenced by the combination of viral species and cell type, with the viral species being the most influential factor.

Aguilar Rangel et al. [14] proposed the idea that the triplet periodicity of the genetic code must place a constraint on deletions in coding sequences. To test this hypothesis, we evaluated whether a characteristic multiplicity existed in the distribution of deletion size. A Lomb–Scargle periodogram (an adaptation of the classic Fourier transformation for unevenly sampled data [38]) of deletion sizes revealed a robust tri-nucleotide periodicity in the cistrons (Figure 4B). HCoV-OC43 data in both cell sizes and MOIs show a characteristic frequency multiple of three (a strong signal at period three and a weaker one at period six). No obvious periodicity has been observed for MHV at a high MOI but a weak signal at period three and a stronger one at period six appears to exist (Figure 4B). This analysis suggests that deletions more likely to disrupt the translation reading frame did not accumulate as much as those that keep it, suggesting that DVG size was evolving under a purifying selection.

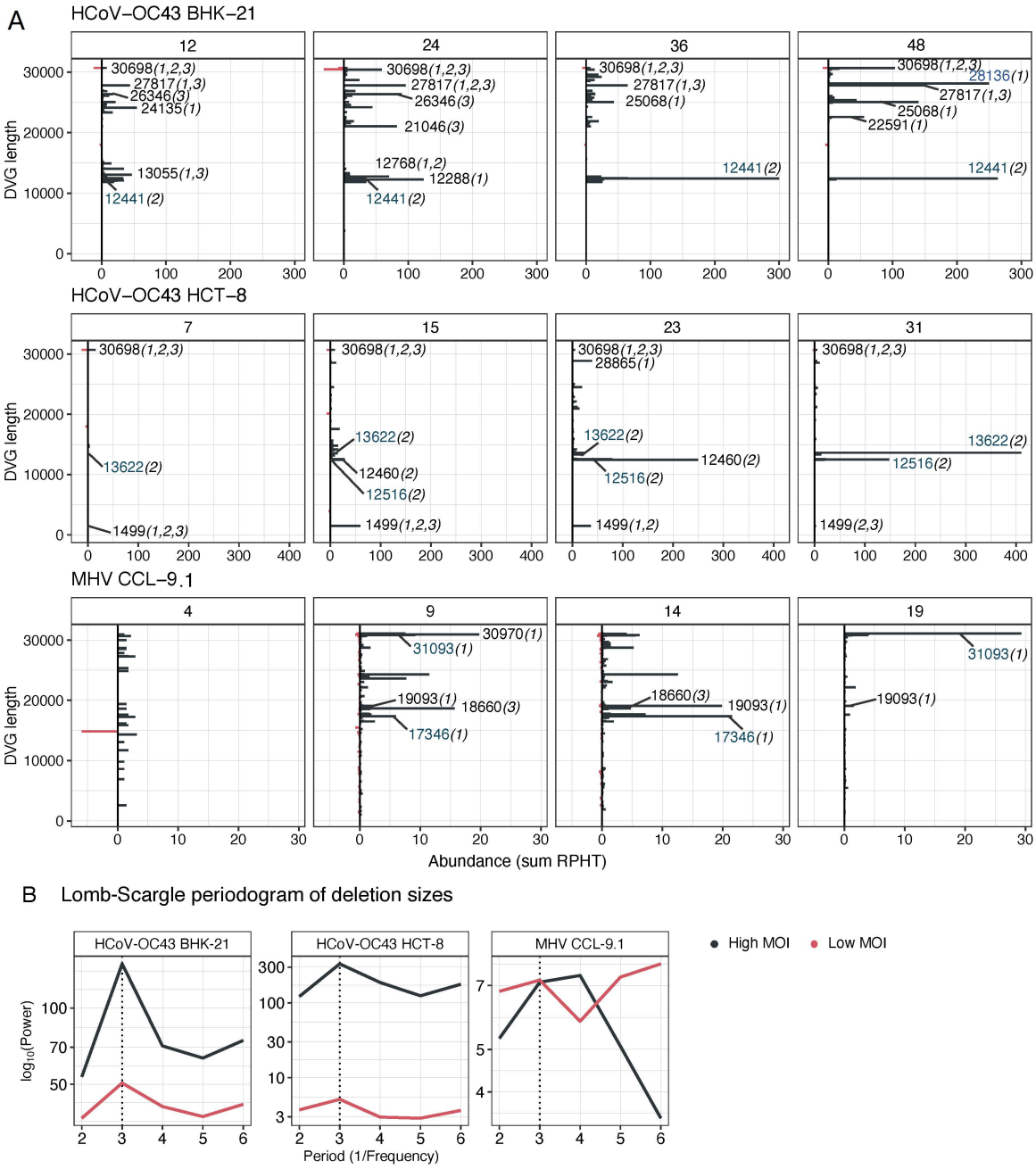



**Figure 4.** Distribution of deletion sizes. (**A**) Changes in the size distribution of deletions along the evolution experiments. Size distribution of genomes with deletions across different evolutionary time points (upper box of each panel, passage numbers), taking into account the variability and abundances of the reconstructed deletions. Replicates of lineages were combined into a single diagram. Numbers indicate some predominant sizes, between parenthesis the lineages in which they were found. Blue numbers highlight the lengths that gain predominance with evolution. Genomes with deletions of fewer than 50 nucleotides were excluded from the visualization. (**B**) Lomb–Scargle periodogram of deletion sizes. All deletions > 3 nucleotides were included in the frequency analysis.

### *3.5. Factors That Influence the Occurrence of Deletions*

Next, we conducted an analysis to determine if specific domains of the viral genome were more prone to act as BP and RP for deletions owing to their sequence or involvement in structural elements. Firstly, we hypothesized that the emergence of deletions could be linked to the very same mechanisms involved in sgRNA formation. These mechanisms in the coronavirus genome involve TRS. To test this possibility, we calculated the frequency of

each genomic position as BP or RP of a deletion (Figure 5A). Deletions shorter than three nucleotides were excluded hereafter based on the consideration that very small deletions, such as sequencing errors [37], must follow a different distribution compared to larger, biologically relevant events. At this stage, we want to restress that canonical sgRNAs were also filtered out before the analysis to avoid spurious results.

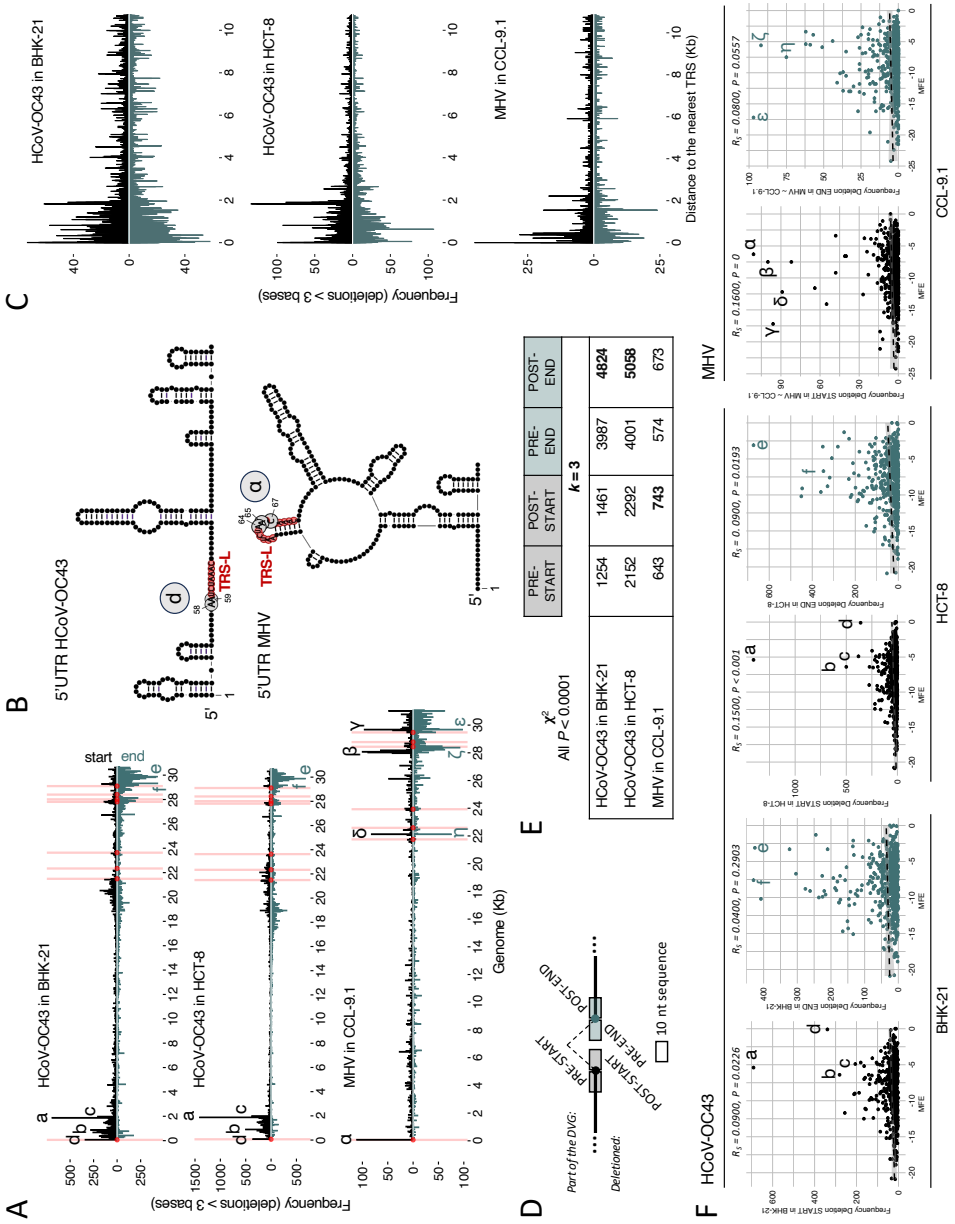



**Figure 5.** Genomic determinants of DVG formation. (**A**) Frequencies of BP and RP of deletions > 3 at each genome positions (in kb). Each bar of the histogram represents a 50-base window. The upper part of the diagram indicates start points, while the lower part does so for end points. Specific TRS positions are highlighted with pink bars and with circles at the base of the start and end points. The most frequently affected areas are indicated with letters a-f for HCoV-OC43 and α–η for MHV. HCoV-OC43 ranges: a = 1850–1900; b = 850–900; c = 1950–2000; d = 50–100; e = 29,850–29,900; and f = 29,800–29,850. MHV ranges: α = 50–100; β = 28,050–28,100; γ = 29,650–29,670; δ = 28,000–28,050, ε = 29,650–29,700, ζ = 28,350–28,400, and η = 2100–22,150. (**B**) 5′UTR structure prediction and localization of one of the most frequent start points (d in HCoV-OC43 and α in MHV) of deletions; TRS-L highlighted in red. (**C**) The distance in nucleotides from each deletion's BP (above) and RP (below) to the nearest TRS (in kb). (**D**) Schematic representation of DVG junction with pre- and post-

BP and RP sequences as studied in the *k*-mer analysis. (**E**) Results from the $\chi^2$ test of 3-mer distribution between the sequences around the deletion compared to those found in the complete genome. Sequences of 10 nucleotides before the BP (pre-BP), 10 after the RP (post-RP), 10 after the BP (post-BP), and 10 before the RP (post-RP) were analyzed. All the results are significant ($p < 0.001$), with the highest $\chi^2$ value of each case bolded. (**F**) Correlation between frequency of BP (black) or RP (green) points with the minimum free energy (MFE) predicted for the 50-base window secondary structure.

In HCoV-OC43, BP accumulated within the first 2 kb of the genome, while the more frequent RP accumulated within the last 2 kb (Figure 5A), with additional likely sites around 415 nucleotides from the 3′ end in BHK-21 and 765 and 1215 in HCT-8. Interestingly, the most likely BP d is two nucleotides upstream from the TRS-L and a few nucleotides before a large structural hairpin in 5′UTR (Figure 5B). This observation was consistent for both cell types. In the case of MHV, the most likely BP was labeled as α, and it lay in the middle of TRS-L and the loop of a hairpin structure (Figure 5B). The density of potential BP and RP in MHV was also high at the last third of the genome for BP and RP, but highly frequent sites were also found in the TRS-L sequence (67 nucleotides), at 970 and at 5975 nucleotides for BP and at 996 and 5996 nucleotides for RP (Figure 5A).

To further explore the proximity of BP and RP to canonical the breakpoints for sgRNA generation (i.e., TRS), we calculated the distance in nucleotides from each deletion's BP (above) and RP (below) to it nearest TRS (Figure 5C). The frequency distribution at various distances from TRS (~0–11 kb) appeared relatively symmetric for both BP and RP (Figure 5C). The most frequent recombination points in all cell types, but particularly in BHK-21, were concentrated within a range of 2 kb around TRS (where the a–c hotspots were also concentrated; Figure 5A), showing a consistent decrease in frequencies with increasing distance from TRS (Figure 5C).

Next, we explored whether the likelihood of being BP and RP can be attributed to specific sequences. The richness in A/U sequences around breakpoints was documented previously for the influenza virus [43] and previous studies with coronaviruses pointed to the UUG triplet as the preferred sequence for spanning junction start positions and a significant A increase in the end positions [44]. Following that, we analyzed the distribution of nucleotides (*k*-mer analysis) of the surroundings of the recombination points (Figure 5D). Figure 5E shows the results of the $\chi^2$ test supporting distinct 3-mer distributions between sequences around the deletion coordinates and those found in the entire genome ($p < 0.001$ in all cases). Analysis of shorter *k*-mers produced identical results. In addition to all nucleotide distributions being significantly different at the recombination coordinates, we observed in HCoV-OC43 that, irrespective of the cell type, the most notable changes in the *k*-mer distribution were observed in the sequences surrounding the RP. In MHV, the highest $\chi^2$ values were in the sequences downstream of the BP (Figure 5D).

Inspired by the results shown in Figure 5B associating highly frequent BP with RNA secondary structures, we tested whether the likelihood of being BP or RP could be associated with the complexity of secondary structures. We assumed that the lower the minimum free energy (MFE), the more complex the RNA structure [45], and the more likely the RNA polymerase will jump out of the template and rejoin somewhere else. As shown in Figure 5F, very small yet significant Spearman correlations were observed between the frequency of BP and MFE for both viruses and, in the case of HCoV-OC43, both cell types (in all cases, $p \leq 0.023$). Regarding RP, a significant correlation was only found for HCoV-OC43 in HCT-8 ($p = 0.019$). Qualitatively, regions with the highest deletion frequency tended to have intermediate MFE values. This suggests that deletions occurred more frequently in structured regions, but not necessarily in highly complex structures.

### *3.6. E and Nsp12, Two Viral Proteins with Viroporin Potential, Are the Most Affected by Deletions*

Now our focus has moved on to exploring whether deletions evenly affected all cistrons or were preferentially found in certain cistrons. To achieve this, we calculated the density of deletions per cistron by dividing the number of deletions found in a given cistron

by its length. Data are shown in Figure 6. These data were then fitted to a generalized linear mixed model (GLMM) in which cell type, MOI, and cistron were treated as fixed effects and lineage (nested factor within the interaction of the fixed factors) and passage (within-individual repeated measures) as random ones, and a Gaussian distribution and identity link function were used. Focusing on HCoV-OC43, no differences among cell types (main or in combination with other fixed factors) were found ($\chi^2 \leq 1.971$, $p \geq 0.923$). MOI had a net effect, with density across cistrons being 178% higher at a high MOI ($\chi^2 = 55.961$, 1 d.f., $p < 0.001$). Also, significant overall differences existed among passages ($\chi^2 = 36.509$, 6 d.f., $p < 0.001$), with the density of deletions across cistrons increasing with evolutionary time. Highly significant differences were found among cistrons ($\chi^2 = 126.321$, 7 d.f., $p < 0.001$), with ORF1ab and S showing the lowest density of deletions whilst E and ns12.9 showed the highest. Most interestingly, a significant interaction between MOI and cistron was observed ($\chi^2 = 36.298$, 7 d.f., $p < 0.001$), with a reduction in the magnitude of the differences between cistrons at a low MOI.

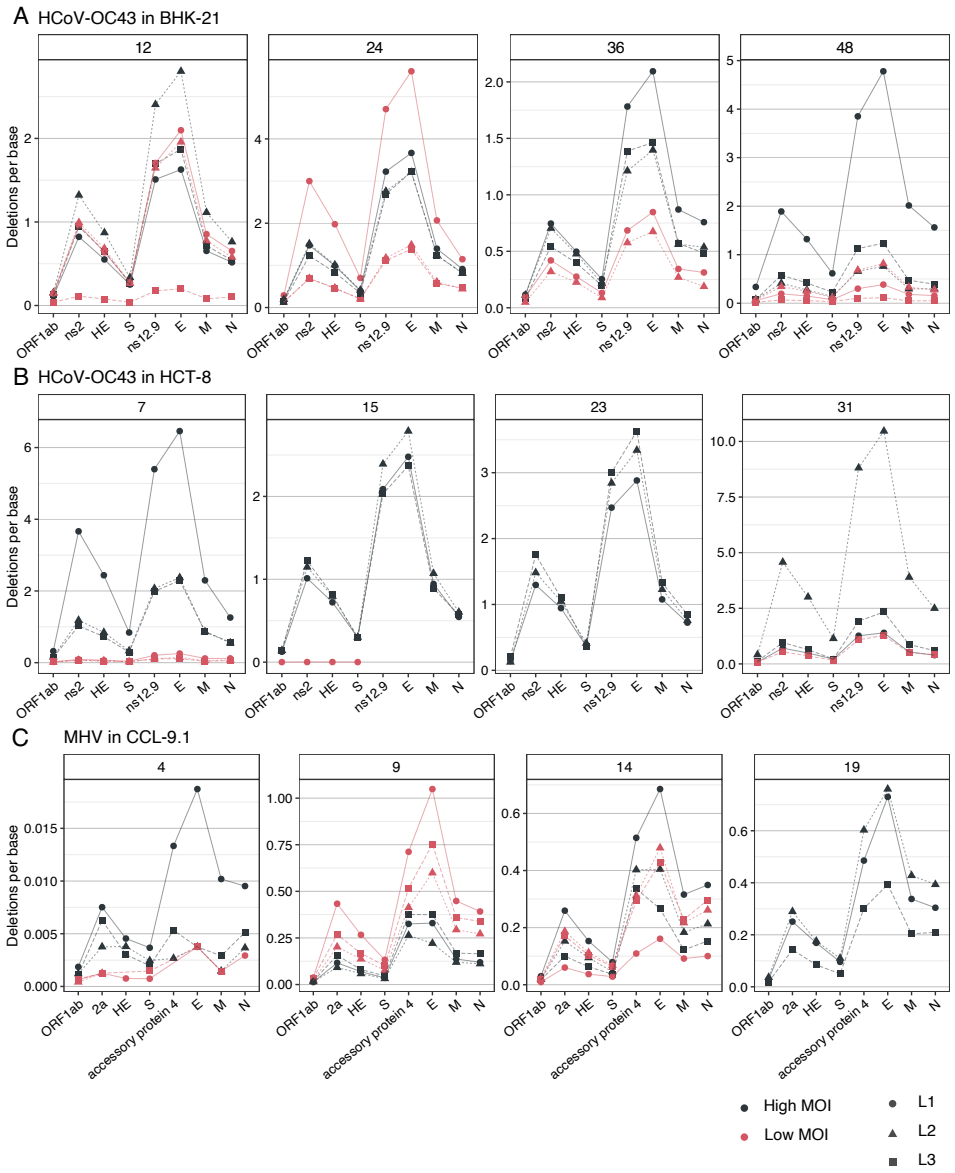



**Figure 6.** Observed frequency of deletions per cistron per base. Each panel corresponds to the indicated passage in the evolution experiment. (**A**) HCoV-OC43 evolved in BHK-21 cells. (**B**) HCoV-OC43 evolved in HCT-8 cells. (**C**) MHV evolved in CCL-9.1 cells.

The results for MHV were somehow different. A significant increase in the overall deletion density along evolutionary time ($\chi^2$ = 88.730, 3 d.f., $p < 0.001$) was observed. Significant differences among cistrons existed ($\chi^2$ = 75.268, 5 d.f., $p < 0.001$), again with ORF1ab and S showing the lower deletion density and the auxiliary protein 4 and E showing the highest. All other factors and combinations were not significant.

### *3.7. A Subset of DVGs Persists along the Evolution Experiment*

Next, we examined the temporal changes in the proportions of persistent DVGs during the experimental evolution. DVGs qualified as de novo if they were observed only in a single passage, while they were qualified as persistent if, after the first detection, they were observed at subsequent passages of the same lineage. For each virus species and cell type independently, data in Figure 7A were fitted to a GLMM with MOI and type of DVG treated as fixed effects and lineage (nested factor within the interaction of the fixed factors) and passage (within-individual repeated measures) as random ones, and a Gamma distribution and log link function. In the case of HCoV-OC43 in BHK-21, as the number of passages increases, the percentage of persistent DVGs rose consistently in BHK-21 cells for all DVGs types ($\chi^2$ = 18.830, 3 d.f., $p < 0.001$) regardless of the MOI ($\chi^2$ = 4.872, 3 d.f., $p$ = 0.181). Interestingly, a highly significant overall difference was found among DVG types ($\chi^2$ = 515.509, 3 d.f., $p < 0.001$), with deletions and insertions being the most common type of persistent DVGs (consistent with the results shown in Section 3.3). Yet, the rate at which DVGs accumulated was the same for all types ($\chi^2$ = 8.859, 9 d.f., $p$ = 0.450). Another interesting result of this analysis was the significant interaction between DVG type and MOI ($\chi^2$ = 13.705, 3 d.f., $p$ = 0.003), due to a disproportionate larger reduction (57.68%) of deletions at a low MOI (5.31% for 3′ cb, 4.39% for 5′ cb, and 22.33% for insertions). However, the pattern observed for HCoV-OC43 in HCT-8 cells differed from that in BHK-21 cells. There was a net overall effect of MOI ($\chi^2$ = 14.602, 1 d.f., $p < 0.001$), with 65.16% more DVGs accumulating at a high MOI, being that this effect also dependent on the evolutionary time ($\chi^2$ = 6.332, 2 d.f., $p$ = 0.042), with an increasing trend for all types of DVGs at a high MOI but mostly declining at a low MOI.

The case of MHV showed similarities and differences with HCoV-OC43 (Figure 7A) as all the terms in the GLMM were highly significant. Among the similarities, the proportion of persistent DVGs increased with the number of passages ($\chi^2$ = 72.091, 3 d.f., $p < 0.001$), although this increase depended both on the type of DVG and on MOI ($\chi^2$ = 85.477, 6 d.f., $p < 0.001$). Also, overall, the abundance of DVGs depended on their type, with deletions being the most common followed by insertion, 5′, and 3′ cbs ($\chi^2$ = 253.615, 3 d.f., $p < 0.001$). The most remarkable difference was that the overall effect of MOI went in the opposite direction: 24.15% more accumulation at a low MOI. The increase in DVG abundance at a low MOI might seem counterintuitive, but various factors, such as the possibility of secondary infections at a low MOI (especially for MHV with a short infection cycle), could contribute to co-infection and the eventual transmission of DVGs.

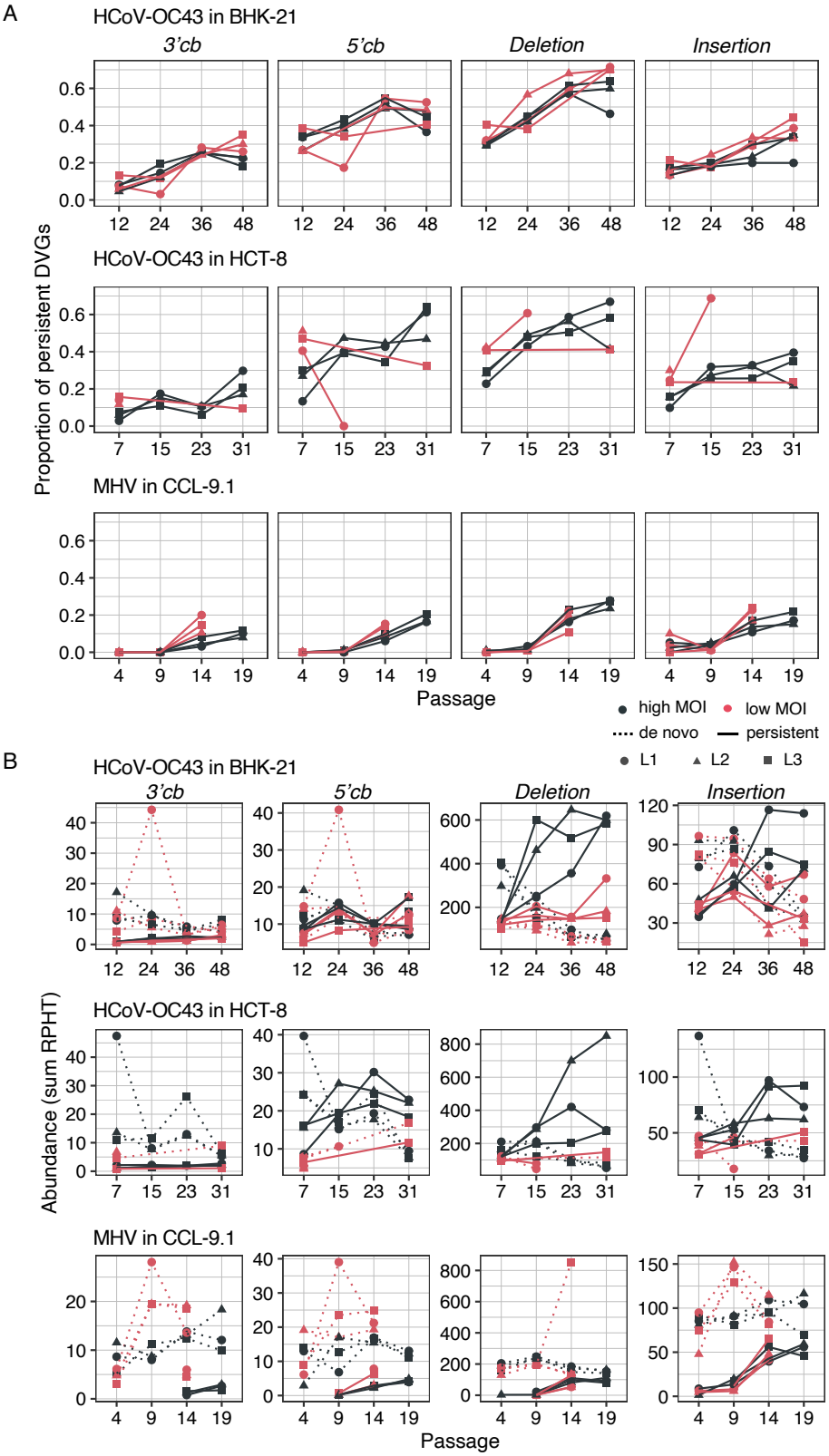



**Figure 7.** Transmission of persistent DVGs. (**A**) Proportion of persistent DVGs at different time points of experimental evolution, separated by DVG type. The proportion was calculated as the ratio of DVGs classified as persistent among all DVGs in the sample. (**B**) Abundance of persistent (solid lines) and de novo generated (dotted lines) DVGs at high (black) and low (red) MOIs, separated by DVG type. Evolutionary lineages are represented by shapes. RPHT: DVG reads per hundred thousand viral reads.

We considered that the most parsimonious explanation for the temporal persistence of DVGs was that they were transmitted across passages. This assumption was supported by the occurrence of deletions being not random but associated with specific regions of the viral genome, suggestive of some sort of selection at play. However, we could not rule out the possibility that some deletions might pervasively emerge de novo at the same coordinates. To test this possibility, we calculated the relative abundance of reads categorized as persistent or as de novo at the various time points, distinguishing between different DVG types and MOI condition (Figure 7B). This distinction was added as an additional fixed factor to the GLMM above and its effect on the abundance of the different types of DVGs was evaluated. In the case of HCoV-OC43 in BHK-21 cells, this factor was significant by itself ($\chi^2$ = 26.159, 1 d.f., $p < 0.001$) as well as in interaction with all the other factors (in all cases, $\chi^2 \geq 6.613$, $p \leq 0.010$). Three particularly relevant cases were the following: (*i*) the interaction of this factor with passage number ($\chi^2$ = 26.159, 1 d.f., $p < 0.001$); while de abundance of the de novo DVGs decreased, on average, 54.17% along the evolution experiment, transmitted DVGs had increased 114.36%. (*ii*) The interaction with MOI ($\chi^2$ = 6.613, 1 d.f., $p = 0.010$). At a high MOI, transmitted DVGs were 46.01% more abundant than at a low MOI, while de novo ones were only 19.87% more, suggesting that a high MOI favors the transmission of already existing DVGs. (*iii*) The interaction with DVG type ($\chi^2$ = 250.006, 3 d.f., $p < 0.001$). While deletion DVGs were 146.50% more abundant among the transmitted class than among de novo class, all other types were less abundant in the transmitted class (79.74% for 3′ cb, 5.99% for 5′ cb, and 4.86% for insertions). When analyzing the patterns in HCT-8 cells, the results were not fully congruent. For example, the interaction with MOI was not significant ($\chi^2$ = 0.403, 1 d.f., $p = 0.525$). The situation with MHV was similar to HCoV-OC43 in HCT-8 and we are not discussing it in detail.

### *3.8. Hotspots of BP and RP for Persistent DVGs*

In Sections 3.5 and 3.6, we demonstrated that recombination events leading to deleted genomes were not randomly distributed across the genome. Here, we explored whether specific hotspots were overrepresented among persistent DVGs, potentially linked to their viability in *trans*-complementation with the wild-type helper virus. Figure 8 illustrates the coordinates of BP and RP for persistent DVGs across all evolutionary lineages. As a reference, Figure 8A,C show schematic representations of the viral genomes. Overall, the most frequent coordinates in HCoV-OC43 infecting BHK-21 (Figure 8B top) aligned with the findings shown in Figure 5A. BPs were notably concentrated within the first 2 kb at the 5′ end of the genome, with another densely populated region between positions 20 and 22 kb. Notably, BPs exhibit minimal variability, while RPs were more diverse, spanning positions from 17–30 kb. This was particularly evident for the two vertical lines around positions 50 and 2000, corresponding to the previously characterized a and d hotspots (Figure 5A). At a low MOI, we observed fewer hotspots, with a high density within the first 2 kb, similar to a high MOI (Figure 8B top). However, the high-density points in the last third of the genome at a high MOI were absent in the low MOI samples. In lineages evolved in HCT-8 (Figure 8B bottom), persistent deletions condensed at the same hotspots as in BHK-21. A notable difference was observed: in HCT-8, the variability of BP increased at specific positions, forming a range of BPs spanning positions from 0 to 5 kb and another range around positions 20 to 24 kb (Figure 8B bottom). These hotspots seemed to be specific to HCoV-OC43 in HCT-8 cells. At a low MOI, the limited number of persistent DVGs hampered pattern recognition, but the identified spots aligned well with those seen at a high MOI.

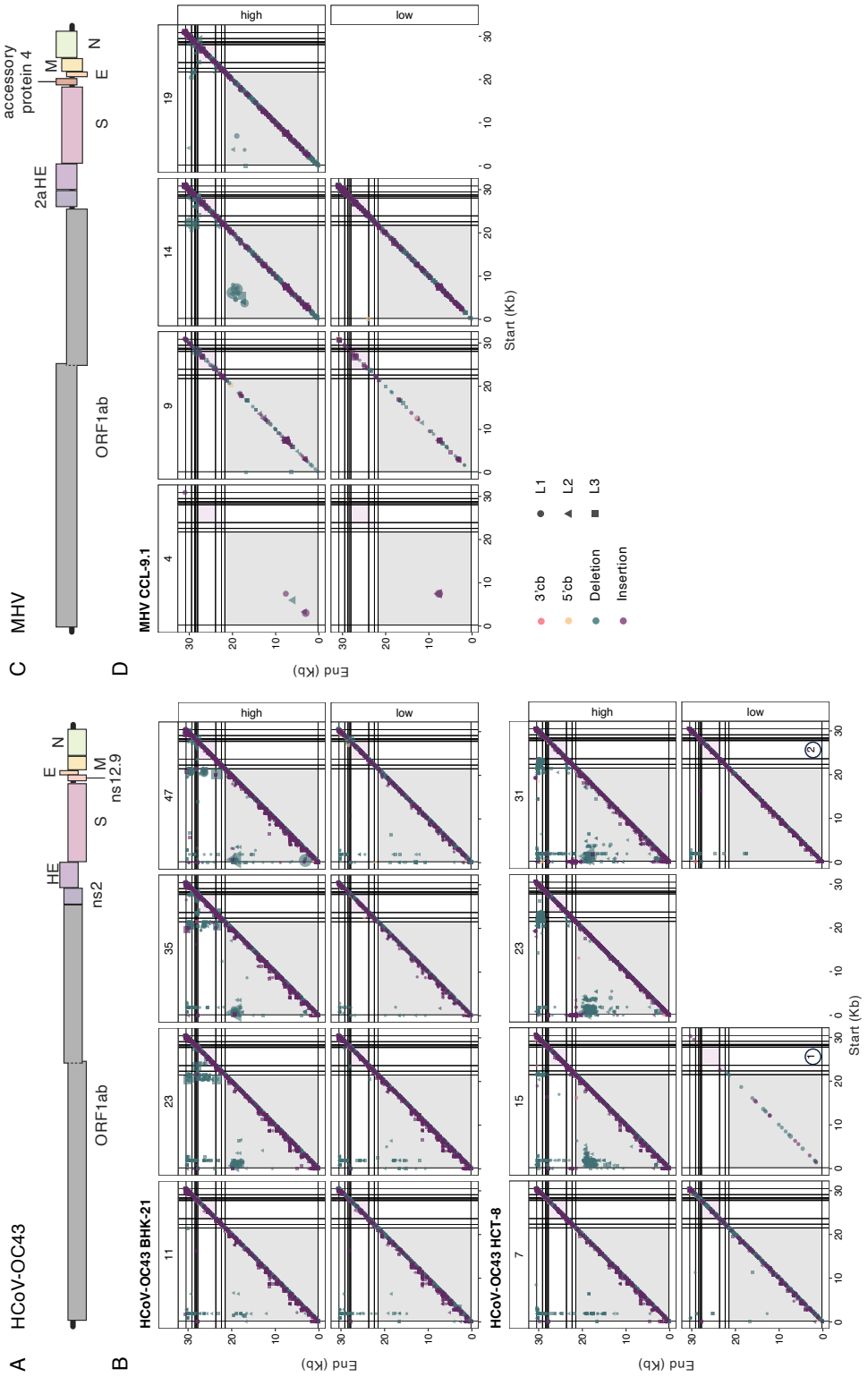



**Figure 8.** Distribution of persistent DVGs BP and RP. (**A**) Schematic representation of HCoV-OC43 genome. (**B**) Coordinates found for HCoV-OC43, at each passage characterized, in both cell types. (**C**) Schematic representation of MHV genome. (**D**) As in (**B**) but for MHV. Coordinates from parallel lineages were joined into the same plot and represented by different symbols. Background colors indicate the cistrons as in (**A**,**C**).

The number of spots for persistent DVGs was small for MHV at both high and low MOIs (Figure 8D). However, two regions with accumulated coordinates for DVGs were

observed at later evolutionary passages. These hotspots spanned from positions 3 to 9 kb and from 16 to 22 kb.

### *3.9. Independently Evolved Lineages Show Convergent Deletions, Reinforcing the Idea That Some DVGs Must Be under Positive Selection*

Next, we aimed to assess the level of reproducibility in the generation of DVGs amongst lineages of the same virus evolved in the same cell host type. Figure 9 depicts the counts of unique deletions (>3 nucleotides) shared among the three independent lineages. A question that immediately arose was whether the count of shared unique DVGs among lineages differed from what would be expected based on the composition of each individual lineage. To illustrate this assessment, we focused on HCoV-OC43 in BHK-21 at a high MOI (Figure 9A). The total count in lineage 1 was 4920 + 116 + 140 + 126 = 5302, 3831 for lineage 2, and 3465 for lineage 3. The count of unique DVGs observed exclusively in lineage 1 was 4920. The expected counts for this group can be calculated as $(4920/5302)\cdot(1 - 3532/3831)\cdot(1 - 3142/3465)\cdot(5302 + 3831 + 3465) \approx 85$. Using the same rationale, the expected counts could be computed for the other six groups displayed in the panel (BHK-21 low: ≈53; HCT-8 high: ≈248; HCT-8 low: ≈13; CCL-9.1 high: ≈14; CCL-9.1 low: ≈2). Applying a goodness-of-fit test, we discovered that observed counts significantly deviated from the expected values ($\chi^2$ = 586,779.739, 6 d.f., $p < 0.001$). This deviation was primarily due to the deficit in counts in the three pairwise comparisons and especially in counts in the group shared by all three lineages. This finding suggested that most of the observed deletion DVGs were produced in a lineage-specific manner, while very few were produced in an almost deterministic manner and shared by all lineages. Similar qualitative outcomes were obtained for HCoV-OC43 in other scenarios (Figure 9A,B), with a very large excess of linage-specific unique DVGs ($\chi^2 \geq$ 15,940.544, 6 d.f., $p < 0.001$). The same results were found for MHV at both MOIs ($\chi^2 \geq$ 71,675.329, 6 d.f., $p < 0.001$).

Noticeably, despite their scarceness, for both viruses and regardless of cell types, the count of shared DVGs among the three lineages consistently remained higher at a high MOI (126 vs. 47 and 308 vs. 12, for HCoV-OC43 in BHK-21 and HCT-8, respectively; and 20 vs. 2 for MHV).

Subsequently, to further characterize the few common deletions shared by the three parallel lineages, we examined their coordinates. In HCoV-OC43 samples (Figure 9A), the coordinates of deletion BP and RP paralleled with the coordinates of the persistent deletions shown in Figure 8B. Again, we observed that the most frequent breakpoints were at the 5′ end of the genome for both cell types and MOIs (within the first 2 kb) and in the last third part of the 5′ genome (around 21 kb) appeared at a high MOI, here more specifically in the HCT-8 samples. Similar to the case of persistent DVGs, we observed a limited number of deletions of BP but highly variable RP.

Finally, we examined the impact of cell type on the generation of common deletions among HCoV-OC43 lineages (Figure 10). Consistent with the probabilistic computations detailed above, we observed a substantial deviation between the actual counts in each of the three categories and the expected count distribution for both MOIs ($\chi^2 \geq$ 28,669.853, 2 d.f., $p < 0.001$). This effect primarily stemmed from a shortage of shared DVGs between the two cell types, suggesting that the spectrum of DVGs generated by HCoV-OC43 was influenced, to some extent, by the specific cell in which it was replicating. Notably, this deficit was more pronounced for lineages evolved at a high MOI. Additionally, when comparing the count numbers in the three groups for lineages evolved at high and low MOIs, we discovered a highly significant distinction (homogeneity $\chi^2$ = 4228.490, 2 d.f., $p < 0.001$). Specifically, at a high MOI, lineages evolved in HCT-8 cells contained proportionally more unique DVGs than lineages evolved in BHK-21 cells, whereas the situation was reversed at a low MOI (Figure 10A).

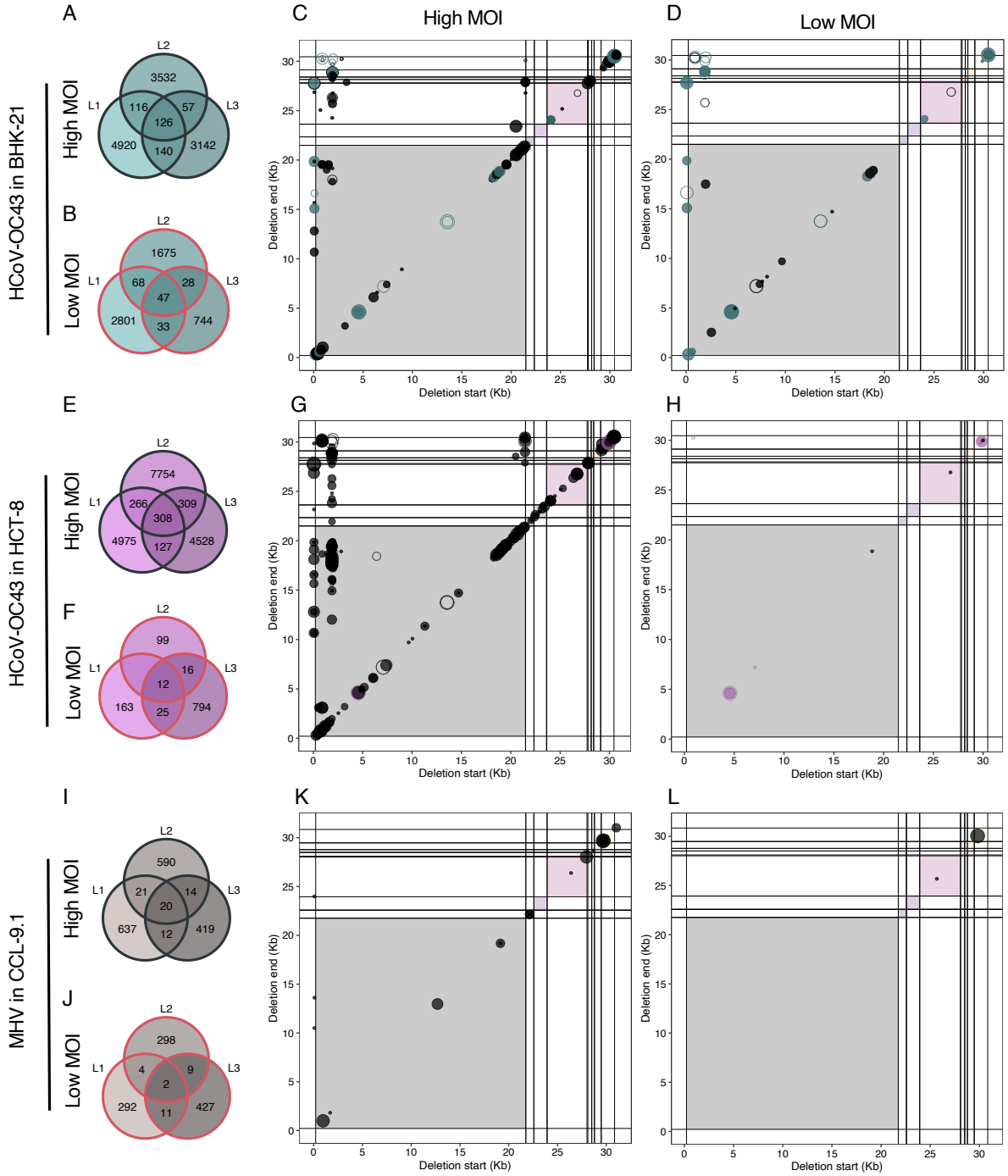



**Figure 9.** Convergencies between evolutionary lineages. Intersection size of deletions founded at lineages of HCoV-OC43 evolved in BHK-21 at (**A**) high and (**B**) low MOIs. BP and RP coordinates of common deletions shared by all the HCoV-OC43 lineages evolved in BHK-21cells at (**C**) high and (**D**) low MOIs. Intersection size of deletions founded at lineages of HCoV-OC43 evolved in HCT-8 at (**E**) high and (**F**) low MOIs. BP and RP coordinates of common deletions of parallel lineages of HCoV-OC43 in HCT-8 at (**G**) high and (**H**) low MOIs. Intersection size of deletions founded at lineages of MHV evolved in CCL-9.1 at (**I**) high and (**J**) low MOIs. BP and RP coordinates of common deletions of parallel lineages of HCoV-OC43 in HCT-8 at (**K**) high and (**L**) low MOIs. For all the scatter plots, the empty points indicate the deletion was already found in the ancestral virus; if the dot color is different from black, the deletion was convergent also between both MOIs. Background colors represent cistrons as in Figure 8A for HCoV-OC43 and Figure 8B for MHV. Symbols' size indicates the number of samples where the deletions were found.

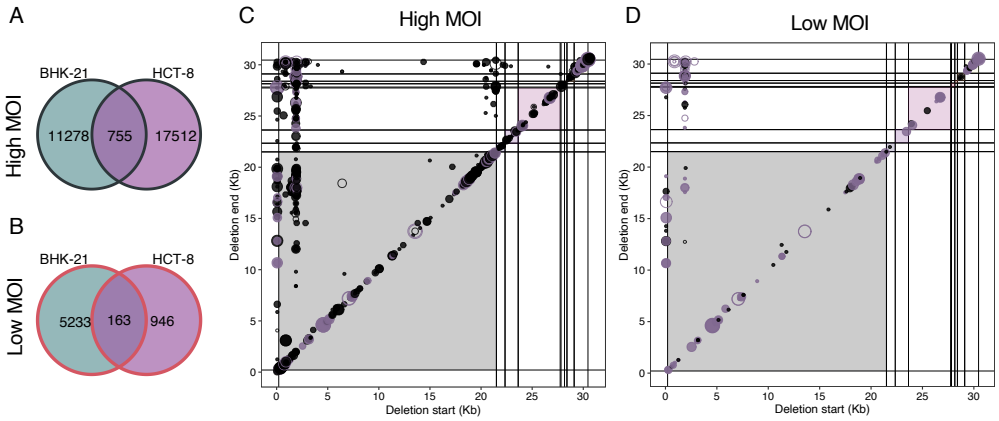



**Figure 10.** HCoV-OC43 deletion DVGs convergent among both cell types. (**A**,**B**) Intersection sizes at (**A**) high and (**B**) low MOIs. Hotspots of common deletions among cell types (intersection set in (**A**,**B**)) at (**C**) high and (**D**) low MOIs. Background colors indicate viral cistrons (as in Figure 8A), the size of the points indicate the number of samples where the deletion is present. Empty circles indicate that the deletions were already present in the ancestral stock and purple and black colors represent, respectively, common to both MOIs or unique to the MOI.

It was noteworthy that 114 out of the 163 convergent deletions identified under low MOI conditions were also present in the high MOI set. More strictly, 60 deletions were shared across all lineages in both cell types at a high MOI and six at a low MOI. Additionally, the pattern of deletion coordinates shared between the two host cell types (Figure 10C,D) was similar to the pattern of common deletions among parallel lineages of the same cell type (Figure 9A) and also to the patterns of persistent deletions (Figure 8B). This consistency could be taken as further evidence that not all deletions were random occurrences, but were regulated by similar molecular mechanisms or selective pressures.

## 4. Discussion

During replication, viruses, especially RNA viruses, spontaneously generate varying versions of their genomes, including hypermutated genomes, insertions, deletions, and reorganizations (5′ and 3′ cb and sb). These unusual genomes cannot complete a whole viral cycle on their own but can play crucial roles in infection outcomes and viral evolution [11,12,14,46]. Understanding the dynamics of DVG formation and accumulation can provide insights into the underlying mechanisms of their appearance and function. To investigate this in betacoronaviruses, we conducted evolution experiments with HCoV-OC43 and MHV.

DVGs exhibited an overall increase in both richness and abundance in all evolving lineages. In HCoV-OC43, DVG richness initially spiked out but later stabilized, suggesting the existence of a diversity ceiling. However, DVG abundance at a high MOI remained consistently high in advanced passages. At a high MOI, HCoV-OC43 DVGs showed higher variability and abundance compared to a low MOI, likely due to potential *trans*-complementation and other sort of interactions between DVGs and helper viruses. Our observations support the hypothesis that at a high MOI, DVGs in HCoV-OC43 were not only generated de novo after each passage but some might also be selected, favoring the accumulation of the fittest DVGs while leading to a decrease in less fit variants. Indeed, at a high MOI, a much larger proportion of persistent DVGs was observed compared to a low MOI. Transmission at a low MOI likely occurred through secondary infections, as uninfected cells remained available after the first round of infection. Additionally, transmission could be facilitated by collective viral infection [47–49] or cell fusion [50–53]. Under these conditions, a DVG could co-infect with a helper virus and be *trans*-complemented for successful propagation, even at an initially low MOI; despite this, foundational effects could be more accused at a low MOI and, in some of the lineages, the viral load was undetectable.

An overall decrease in viral load in most cases could be explained by two alternative hypotheses. Firstly, it could be attributed to the progressive accumulation of DVGs that interfere with the replication of wild-type viruses [6,14,54,55]. This effect would be stronger at a high MOI, which seems not to be a general case in Figure 1. Secondly, it could be a consequence of strong transmission bottlenecks that turn on Muller's ratchet, leading to the fixation of deleterious mutations and, consequently, fitness declines, as is well established for other RNA viruses [56–60]. To further explore the role of DVGs in these dynamics, we performed RNA-seq from the supernatants of infected cells from each lineage at four equidistant evolutionary time points, as well as from the corresponding ancestral viruses, in order to reconstruct the DVG populations at each sample. Significant DVG accumulation was observed in both cell lines with HCoV-OC43; in BHK-21 cells, there was also a general increase with small fluctuations in infectious viral load. The positive trends for both full and defective viruses suggest that an interfering dynamic was not necessarily occurring. These dynamics reinforce the idea of DVGs as triggers and/or reservoirs of viral diversity, providing an adaptative advantage [12,44,61]. However, in HCT-8 cells, viral loads showed an overall decreasing trend. This could be attributed to technical limitations in testing viral loads, as it had to be assessed in BHK-21 cells due to the inability to form plaques in HCT-8 cells. Viral adaptation to HCT-8 cells during the passages might have resulted in fitness tradeoffs in the BHK-21 cells used for plaquing. Further studies are needed to investigate the potential role of DVG–complete virus cooperation at this stage.

In contrast to HCoV-OC43, the richness and abundance of DVGs in MHV were positively correlated and increased throughout evolution, while the overall trend for viral load indicated a decrease, leading to undetectable levels in some MOI lineages. We proposed two non-mutually exclusive hypotheses to explain the observed fluctuations in viral load, one calling for the increasing accumulation of DVGs and another one associating the effect with reductions in the inoculum size and the subsequent onset of Muller's ratchet [56–60]. Our finding of an accumulation of diverse populations of DVGs in evolving lineages, in both viruses but particularly in MHV, gives support to the first possibility, although does not rule out the second hypothesis, especially at a low MOI.

The MHV strain that we used in this experiment was interferon-sensitive. Although MHV has a short infection cycle and should not be notably affected by interferon-based cell defense [62], we cannot completely rule out the possibility that interplay between the virus and the cells' immune response may have impacted the generation and propagation of DVGs, contributing too to the divergent outcomes observed amongst the two betacoronaviruses. The interferon response has been identified as key in infections in which the presence of DVGs impacts the severity of symptoms [50], or in which infection does not resolve but becomes chronic [12,63,64].

The choice of cell lines for the HCoV-OC43 evolution experiments was primarily motivated by their high susceptibility and ability to produce high viral titers, which was essential for virus passaging at a high MOI. As a result, the outcome of viral DVG evolution in both cell lines showed similarities, indicating that they may not have imposed largely different constraints for viral adaptation. In retrospect, the use of more diverse cell lines with varying properties and susceptibilities might have provided additional insights into the impact of host factors on DVG evolution. Nevertheless, the results obtained from the selected cell lines still offered valuable information on DVG dynamics and accumulation in the context of HCoV-OC43 evolution. Future studies could consider employing a broader range of cell lines to gain a more comprehensive understanding of how host factors influence the evolution of DVGs. Indeed, Aguilar Rangel et al. [14] have shown that the Dengue virus (DENV) generates different spectra of DVGs in human and mosquito cells. Likewise, Hasiów-Jaroszewska et al. [65] also described variation in the composition of DVGs of tomato black ring virus in different host species. We could also compare our results for MHV in CCL-9.1 with those obtained in DBT-9 cells by Gribble et al. [44] and identify common hotspots of recombination, as well as differences.

Deletions were the most variable and abundant type of reconstructed DVGs, an observation that is consistent with previous studies with other positive-sense single-stranded RNA viruses: Aguilar Rangel et al. [14], in the cases of poliovirus and DENV; Zhou et al. [50], in the case of SARS-CoV-2; and Jaworski and Routh [66] for Flock House virus. For this reason, we conducted a more detailed analysis of this specific DVG type. The analysis, focusing on deletions longer than three nucleotides, distinguishes patterns in genome sizes that dominated in abundance. Over the course of HCoV-OC43 evolution at a high MOI, these patterns were refined within a specific size range, giving rise to dynamic peaks of high abundance. This dynamic behavior within a certain size range suggests adaptive evolution. In contrast, the accumulation pattern of certain sizes at a low MOI was not prominent, and, notably, MHV did not exhibit a clear pattern of size accumulation. This lack of a discernible pattern may indicate different evolutionary dynamics or constraints under a low MOI. Furthermore, short genomes with deletions (<2 kb) were particularly prevalent in the HCT-8 cell lineage. This observation underscores the potential influence of host-specific factors on the evolution dynamics of viral genomes with deletions.

Furthermore, we observed that the cistrons *E* and *ns12.9* (or its syntenic *accessory protein 4* in MHV) were more impacted by deletions, while *ORF1ab* and *S* were less affected at both MOIs for both viruses. Coronaviruses *E* cistron encodes for a small structural protein involved in multiple stages of the replicative cycle; it participates in viral assembly, virion release, and pathogenesis [67]. In MHV, E is not essential for virus replication and a Δ*E* mutant showed a lower growth rate and a lower infectious titer than the wild-type virus, but it still was capable of producing viable particles [68]. *E* is highly expressed in infected cells, but only a small part is incorporated into the virion envelope. E also can form homotypic interactions, oligomerizing to generate viroporins, ion-channel proteins associated with pathogenicity. The HCoV-OC43 ns12.9 protein is associated with virus morphogenesis and pathogenesis by viroporin formation. The HCoV-OC43-Δ*ns12.9* was defective in growth in vitro and reduced inflammation and virulence in the brains of infected mice, confirming its influence in virus pathogenesis [69]. Some studies had associated the inhibition of the E viroporins with a reduction in pathogenicity, highlighting its therapeutical value to face coronavirus infections [67], while others point towards different viroporins as potential targets for antiviral drugs [69]. The identification of *E*, *ns12.9*, and *accessory protein 4* as the most prone cistrons for deletions suggests the tantalizing possibility that DVGs in these cistrons may be selected for as a viral strategy to modulate virulence.

Furthermore, the E protein may interact with the host's immune system, implying that deletions in these regions could impact the virus' ability to replicate and evade immune responses. Our evolution experiment took place in cell culture and not in a whole organism with a complete immune system. The independent emergence of deletions of equivalent parts of the genome in different viruses and cell types suggests that the shift from the whole organism to cell culture was a more substantial landmark than the differences among various cell culture types. Additionally, these deletions were accumulated already in the first stages of the evolution, which suggests that the adaptive advantage of these variants was significant. However, the underlying mechanisms could be much more complex, potentially involving specific viral sequences and structures. The analysis of the most frequent sites for BP and RP, along with their surrounding sequences, suggests that the mechanism of DVG generation is not random and is influenced, in part, by the genome sequence but further by the resulting secondary RNA structure.

It is noteworthy that some hotspots of persistent deletions of BPs are in the first third of the viral genome, a region generally assumed to be crucial for viral replication. We identified a site close to the TRS-L as one of the most frequent BPs, suggesting that the RNA structures needed for replication would be intact. We should not rule out the effect that successive deletions have on the increase in frequency in some spots, or a special case of the latter, that sgRNAs are targets of deletions. Unfortunately, the sequencing technology used for this work limits our ability to investigate this further.

We have also paid attention to identifying deletions that consistently appeared over the course of evolution. In the case of HCoV-OC43, some of these persistent deletions likely represent instances where deletions appeared de novo multiple times in the same lineage, because our results also indicated that the process of DVG generation was not completely random. Notably, the pattern of coordinates for persistent deletions highlighted the existence of convergent deletions between parallel lineages and even between different cell types. This suggests that a subset of deletions arose independently of the host and lineage. Another portion of pervasive but not de novo generated deletions was likely composed of deletions that were transmitted along passages. When comparing samples at high vs. low MOIs, we observed more persistent deletions under high MOI conditions, suggesting that the virus, given the opportunity for trans-complementation, exhibits an increased tendency for supporting persistent deletions. In the case of MHV, most DVGs were not transmitted but de novo generated within each passage, not being under specific selection pressure, consistent with the less defined and more variable size distribution. This suggests, in general, that the dynamics of DVG accumulation differ among two closely related betacoronavirus species, with potentially distinct mechanisms governing their appearance, abundance, and fitness [70,71].

Furthermore, our findings indicated that the dynamics of persistent deletions in HCoV-OC43 at a high MOI maintained hotspots of deletion coordinates, becoming more abundant with passages. Certain genomic regions seemed more susceptible to deletions than others [24], with these hotspots showing an increased abundance during experimental evolution at a high MOI. In both viruses, regions susceptible to deletions were detected, and the abundance of deletions in these hotspots increased with the progress of experimental evolution at a high MOI. Despite a larger genome size being a disadvantage for replication competition, longer DVGs lacking structural proteins may have the advantage of being more readily complemented in trans and transmitted. Zhou et al. [50] also found hotspots of BPs and RPs in the genome of SARS-CoV-2 in samples from patients, which suggest that our in vitro results might reflect relevant processes in vivo.

The dynamics of deletions in parallel lineages showed a remarkable degree of convergence, despite the majority of shared deletions between the lineages were very short, potentially classified as sequencing errors [37]. After removing these short ones, we still found a number of DVGs of various sizes shared among lineages evolved in the same host. This finding prompted distinct patterns of deletion evolution within different cell types, combining stochastic generation of deletions within each passage and the potential persistence of positively selected deletions shared by different lineages. The observation that some deletions can be positively selected to modulate virus replication and pathogenesis, for instance, by altering coding reading frames and generating novel proteins [14], opens tantalizing possibilities for the development of DVGs with antiviral activity. In general, short DIPs, and not long DVG, have been considered for their interfering activity over wild-type viruses [72], although results by Levi et al. [73] suggest that long DVGs can be effective and broad-spectrum antiviral for alphaviruses.

## 5. Conclusions

In conclusion, the buildup of DVGs seems to be governed by dynamic processes that set a cap on the diversity of DVGs capable of accumulating in each sample. This restriction is affected by both the passage number and, in specific instances, the MOI. Throughout evolution at a high MOI, a selective process appears to take place, giving preference to certain DVGs for accumulation while others diminish over successive passages. This observation implies that DVGs possess the ability to transmit under high MOI conditions in a virus- and cell-type-dependent manner, resulting in their enduring presence and potential influence on viral evolution.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v16040644/s1, Figure S1: Percentage of sgRNAs over the total

number of deletions observed per sample. Thick lines are only shown for illustrating the average trend over time; Figure S2: Distribution of deletion type DVGs at a low MOI. To facilitate visualization only deletions bigger than 50 nucleotides are shown. The number represent the exact length of the DVG and in parenthesis the lineages where they were founded; Table S1: Statistics describing the variation in MOI for each experimental evolution lineage. Median values ± IQR are shown. Last column shows the fold difference between high and low MOIs (averaging lineages); Table S2: Fitting of viral load data to the best-fitting ARIMA($p$, $d$, $q$) model based on the minimum BIC criterium. Errors represent ±1 SE.

**Author Contributions:** Conceptualization, S.F.E.; methodology, J.H., M.J.O.-U. and J.C.M.-S.; software, M.J.O.-U.; validation, J.H., M.J.O.-U., J.C.M.-S. and S.F.E.; formal analysis, J.H., M.J.O.-U., J.C.M.-S. and S.F.E.; investigation, J.H., M.J.O.-U., J.C.M.-S. and S.F.E.; data curation, J.H., M.J.O.-U., J.C.M.-S. and S.F.E.; writing—original draft preparation, J.H. and M.J.O.-U.; writing—review and editing, S.F.E.; visualization, J.H. and M.J.O.-U.; supervision, S.F.E.; project administration, S.F.E.; funding acquisition, S.F.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by CSIC PTI Salud Global grant 202020E153, by grants SGL2021-03-009 and SGL2021-03-052 from European Union NextGenerationEU/PRTR through the CSIC Global Health Platform established by EU Council Regulation 2020/2094, and by grant PDC2022-133020-I00 from MCIN/AEI/10.13039/501100011033 and "European Union NextGenerationEU/PRTR". M.J.O-U. was supported by grant FPU2019/05246 funded by MCIN/AEI/10.13039/501100011033 and "ESF investing in your future". J.C.M-S. was supported by grant ACIF/2021/296 from Generalitat Valenciana.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Excel files with viral titers and dilution schemes are available at https://git.csic.es/sfelenalab/dvg-accumulation-in-betacoronaviruses-experimental-evolution (accessed on 17 April 2024).

**Acknowledgments:** We thank Paula Agudo, Juan V. Bou, José L. Carrasco, Francisca de la Iglesia, J. Tomás Lázaro, María C. Marqués, Josep Sardanyés, and Ernesto Segredo-Otero for help, assistance and fruitful discussions.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Duffy, S.; Shackelton, L.A.; Holmes, E.C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* **2008**, *9*, 267–276. [CrossRef] [PubMed]
2. Sanjuán, R.; Nebot, M.R.; Chirico, N.; Mansky, L.M.; Belshaw, R. Viral mutation rates. *J. Virol.* **2010**, *84*, 9733–9748. [CrossRef] [PubMed]
3. Belshaw, R.; Sanjuán, R.; Pybus, O.G. Viral mutation and substitution: Units and levels. *Curr. Opin. Virol.* **2011**, *1*, 430–435. [CrossRef] [PubMed]
4. Stampfer, M.; Baltimore, D.; Huang, A.S. Absence of interference during high-multiplicity infection by clonally purified vesicular stomatitis virus. *J. Virol.* **1971**, *7*, 409–411. [CrossRef]
5. Huang, A.S. Defective interfering viruses. *Annu. Rev. Microbiol.* **1973**, *27*, 101–118. [CrossRef] [PubMed]
6. Felt, S.A.; Achouri, E.; Faber, S.R.; López, C.B. Accumulation of copy-back viral genomes during respiratory syncytial virus infection is preceded by diversification of the copy-back viral genome population followed by selection. *Virus Evol.* **2022**, *8*, veac091. [CrossRef]
7. Barrett, A.D.T.; Dimmock, N.J. Modulation of Semliki Forest virus-induced infection of mice by defective-interfering virus. *J. Infect. Dis.* **1984**, *150*, 98–104. [CrossRef] [PubMed]
8. Roux, L.; Simon, A.E.; Holland, J.J. Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. *Adv. Virus Res.* **1991**, *40*, 181–211. [PubMed]
9. Smith, C.M.; Scott, P.D.; O'Callaghan, C.; Easton, A.J.; Dimmock, N.J. Defective interfering influenza RNA inhibits infectious influenza virus replication in human respiratory tract cells: A potential new human antiviral. *Viruses* **2016**, *8*, 237. [CrossRef]
10. Huang, A.S.; Baltimore, D. Defective viral particles and viral disease processes. *Nature* **1970**, *226*, 325–327. [CrossRef]
11. Rezelj, V.V.; Levi, L.I.; Vignuzzi, M. The defective component of viral populations. *Curr. Opin. Virol.* **2018**, *33*, 74–80. [CrossRef] [PubMed]

12. Vignuzzi, M.; López, C. Defective viral genomes are key drivers of the virus-host interaction. *Nat. Microbiol.* **2019**, *4*, 1075–1087. [CrossRef] [PubMed]
13. Di Gioacchino, A.; Legendre, R.; Rahou, Y.; Najburg, V.; Charneu, P.; Greenbaum, B.D.; Tangy, F.; van der Werf, S.; Cocco, S.; Komarova, A.V. sgDI-tector: Defective interfering viral genome bioinformatics for detection of coronavirus subgenomic RNAs. *RNA* **2022**, *28*, 277–289. [CrossRef]
14. Aguilar Rangel, M.; Dolan, P.T.; Taguwa, S.; Xiao, Y.; Andino, R.; Frydman, J. High-resolution mapping reveals the mechanism and contribution of genome insertions and deletions to RNA virus evolution. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2304667120. [CrossRef]
15. Jennings, P.A.; Finch, J.T.; Robertson, J.S. Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* **1983**, *34*, 619–627. [CrossRef] [PubMed]
16. Saira, K.; Lin, X.; DePasse, J.V.; Halpin, R.; Twaddle, A.; Stockwell, T.; Angus, B.; Cozzi-Lepri, A.; Delfino, M.; Dugan, V.; et al. Sequence analysis of in vivo defective interfering-like RNA of influenza a H1N1 pandemic virus. *J. Virol.* **2013**, *87*, 8064–8074. [CrossRef]
17. Poirier, E.Z.; Mounce, B.C.; Rozen-Gagnon, K.; Hooikaas, P.J.; Stapleford, K.A.; Moratorio, G.; Vignuzzi, M. Low-fidelity polymerases of alphaviruses recombine at higher rates to overproduce defective interfering particles. *J. Virol.* **2016**, *90*, 2446–2454. [CrossRef]
18. Li, T.; Pattnaik, A.K. Replication signals in the genome of vesicular stomatitis virus and its defective interfering particles: Identification of a sequence element that enhances DI RNA replication. *Virology* **1997**, *232*, 248–259. [CrossRef] [PubMed]
19. Sun, Y.; Kim, E.J.; Felt, S.A.; Taylor, L.J.; Agarwal, D.; Grant, G.R.; López, C.B. A specific sequence in the genome of respiratory syncytial virus regulates the generation of copy-back defective viral genomes. *PLoS Pathog.* **2019**, *15*, e1007707.
20. Brian, D.A.; Baric, R.S. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **2005**, *287*, 1–30.
21. Yang, D.; Leibowitz, J.L. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus Res.* **2015**, *206*, 120–133. [CrossRef] [PubMed]
22. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The architecture of SARS-CoV-2 transcriptome. *Cell* **2020**, *181*, 914–921. [CrossRef]
23. Repass, J.F.; Makino, S. Importance of the positive-strand RNA secondary structure of a murine coronavirus defective interfering RNA internal replication signal in positive-strand RNA synthesis. *J. Virol.* **1988**, *72*, 7926–7933. [CrossRef] [PubMed]
24. Lin, Y.J.; Lai, M.M. Deletion mapping of a mouse hepatitis virus defective interfering RNA reveals the requirement of an internal and discontiguous sequence for replication. *J. Virol.* **1993**, *67*, 6110–6118. [CrossRef] [PubMed]
25. Liu, Q.; Johnson, R.F.; Leibowitz, J.L. Secondary structural elements within the 3′ untranslated region of mouse hepatitis virus strain JHM genomic RNA. *J. Virol.* **2001**, *75*, 12105–12113. [CrossRef] [PubMed]
26. De Haan, C.A.M.; Volders, H.; Koetzner, C.A.; Masters, P.S.; Rottier, P.J.M. Coronaviruses maintain viability despite dramatic rearrangements of the strictly conserved genome organization. *J. Virol.* **2002**, *76*, 12491–12502. [CrossRef] [PubMed]
27. Goebel, S.J.; Taylor, J.; Masters, P.S. The 3′ cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J. Virol.* **2004**, *78*, 7846–7851. [CrossRef] [PubMed]
28. Liu, P.; Li, L.; Millership, J.J.; Kang, H.; Leibowitz, J.L.; Giedroc, D.P. A U-turn motif-containing stem-loop in the coronavirus 5′ untranslated region plays a functional role in replication. *RNA* **2007**, *13*, 763–780. [CrossRef] [PubMed]
29. Sethna, P.B.; Hung, S.L.; Brian, D.A. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons'. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5626–5630. [CrossRef]
30. Hofmann, M.A.; Sethna, P.B.; Brian, D.A. Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. *J. Virol.* **1990**, *64*, 4108–4114. [CrossRef]
31. Van der Most, R.G.; Bredenbeek, P.J.; Spaan, W.J.M. Domain at the 3′ end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. *J. Virol.* **1991**, *65*, 3219–3226. [CrossRef] [PubMed]
32. Fosmire, J.A.; Hwang, K.; Makino, S. Identification and characterization of a coronavirus packaging signal. *J. Virol.* **1992**, *66*, 3522–3530. [CrossRef] [PubMed]
33. Zhao, X.; Shaw, K.; Cavanagh, D. Presence of subgenomic mRNAs in virions of coronavirus IBV. *Virology* **1993**, *196*, 172–178. [CrossRef] [PubMed]
34. Reed, L.J.; Muench, H. A simple method of estimating fifty per cent endpoints. *Am. J. Epidemiol.* **1938**, *27*, 493–497. [CrossRef]
35. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**. [CrossRef]
36. Olmo-Uceda, M.J.; Muñoz-Sánchez, J.C.; Lasso-Giraldo, W.; Arnau, V.; Díaz-Villanueva, W.; Elena, S.F. DVGfinder: A metasearch tool for identifying defective viral genomes in RNA-Seq data. *Viruses* **2022**, *14*, 1114. [CrossRef] [PubMed]
37. Schirmer, M.; D'Amore, R.; Ijaz, U.Z.; Hall, N.; Quince, C. Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinform.* **2016**, *17*, 215. [CrossRef] [PubMed]
38. VanderPlas, J.T. Understanding the Lomb-Scargle periodogram. *Astrophys. J. Suppl. Ser.* **2018**, *236*, 16. [CrossRef]
39. Pagès, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. Biostrings: Efficient Manipulation of Biological Strings. R Package Version 2.70.1. 2024. Available online: https://bioconductor.org/packages/Biostrings (accessed on 20 March 2024).
40. Lorenz, R.; Hofacker, I.L.; Stadler, P.F. RNA folding with hard and soft constraints. *Algorithms Mol. Biol.* **2016**, *11*, 8. [CrossRef]
41. Johnson, P.Z.; Kasprzak, W.K.; Shapiro, B.A.; Simon, A.E. RNA2Drawer: Geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. *RNA Biol.* **2019**, *16*, 1667–1671. [CrossRef]

42. Lawrence, M.; Huber, W.; Pagès, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.T.; Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **2013**, *9*, e1003118. [CrossRef] [PubMed]
43. Elshina, E.; te Velthuis, A.J.W. The influenza virus RNA polymerase as an innate immune agonist and antagonist. *Cell. Mol. Life Sci.* **2021**, *78*, 7237–7256. [CrossRef] [PubMed]
44. Gribble, J.; Stevens, L.J.; Agostini, M.L.; Anderson-Daniels, J.; Chappell, J.D.; Lu, X.; Pruijssers, A.J.; Routh, A.L.; Denison, M.R. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **2021**, *17*, e1009226. [CrossRef] [PubMed]
45. Zuker, M. On finding all suboptimal foldings of an RNA molecule. *Science* **1989**, *244*, 48–52. [CrossRef] [PubMed]
46. Elena, S.F. The role of indels in evolution and pathogenicity of RNA viruses. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2310785120. [CrossRef] [PubMed]
47. Maeda, J.; Maeda, A.; Makino, S. Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells. *Virology* **1999**, *263*, 265–272. [CrossRef]
48. Cuevas, J.M.; Durán-Moreno, M.; Sanjuán, R. Multi-virion infectious units arise from free viral particles in an enveloped virus. *Nat. Microbiol.* **2017**, *2*, 17078. [CrossRef]
49. Andreu-Moreno, I.; Sanjuán, R. Collective infection of cells by viral aggregates promotes early viral proliferation and reveals a cellular-level Allee effect. *Curr. Biol.* **2018**, *28*, 3212–3219. [CrossRef]
50. Zhou, T.; Gilliam, N.J.; Li, S.; Spandau, S.; Osborn, R.M.; Connor, S.; Anderson, C.S.; Mariani, T.J.; Thakar, J.; Dewhurst, S.; et al. Generation and functional analysis of defective viral genomes during SARS-CoV-2 infection. *mBio* **2023**, *14*, e00250-23. [CrossRef]
51. Rosell, A.; Havervall, S.; von Meijenfeldt, F.; Hisada, Y.; Aguilera, K.; Grover, S.P.; Lisman, T.; Mackman, N.; Thålin, C. Patients with COVID-19 have elevated levels of circulating extracellular vesicle tissue factor activity that is associated with severity and mortality. *Arterioscler. Thromb. Vasc. Biol.* **2021**, *41*, 878–882. [CrossRef]
52. Zavan, B.; Crupi, M.J.F.; Ilkow, C.S.; Bell, J.C.; Boulton, S. Extracellular vesicles and viruses: Two intertwined entities. *Int. J. Mol. Sci.* **2023**, *24*, 1036. [CrossRef] [PubMed]
53. Buchrieser, J.; Dufloo, J.; Hubert, M.; Monel, B.; Planas, D.; Rajah, M.M.; Planchais, C.; Porrot, F.; Guivel-Benhassine, F.; van der Werf, S.; et al. Syncytia formation by SARS-CoV-2-infected cells. *EMBO J.* **2020**, *39*, e106267. [CrossRef] [PubMed]
54. Leroy, H.; Han, M.; Woottum, M.; Bracq, L.; Bouchet, J.; Xie, M.; Benichou, S. Virus-mediated cell-cell fusion. *Int. J. Mol. Sci.* **2020**, *21*, 9644. [CrossRef] [PubMed]
55. Rajah, M.M.; Bernier, A.; Buchrieser, J.; Schwartz, O. The mechanism and consequences of SARS-CoV-2 spike-mediated fusion and syncytia formation. *J. Mol. Biol.* **2022**, *434*, 167280. [CrossRef] [PubMed]
56. Clarke, D.K.; Duarte, E.A.; Moya, A.; Elena, S.F.; Domingo, E.; Holland, J.J. Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* **1993**, *67*, 222–228. [CrossRef] [PubMed]
57. Duarte, E.A.; Clarke, D.K.; Moya, A.; Elena, S.F.; Domingo, E.; Holland, J.J. Many-trillionfold amplification of single RNA virus particles fails to overcome the Muller's ratchet effect. *J. Virol.* **1993**, *67*, 3620–3623. [CrossRef] [PubMed]
58. Novella, I.S.; Elena, S.F.; Moya, A.; Domingo, E.; Holland, J.J. Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *J. Virol.* **1995**, *69*, 2869–2872. [CrossRef]
59. Elena, S.F.; González-Candelas, F.; Novella, I.S.; Duarte, E.A.; Clarke, D.K.; Domingo, E.; Holland, J.J.; Moya, A. Evolution of fitness in experimental populations of vesicular stomatitis virus. *Genetics* **1996**, *142*, 673–679.
60. De la Iglesia, F.; Elena, S.F. Fitness declines in tobacco etch virus upon serial bottleneck transfers. *J. Virol.* **2008**, *81*, 4941–4947. [CrossRef]
61. Lin, C.H.; Tam, H.M.H.; Yang, C.Y.; Hsieh, F.C.; Wang, J.L.; Yang, C.C.; Hsu, H.W.; Liu, H.P.; Wu, H.Y. Evolution of the coronavirus spike protein in the full-length genome and defective viral genome under diverse selection pressures. *J. Gen. Virol.* **2023**, *104*, 001920. [CrossRef]
62. Kyuwa, S.; Sugiura, Y. Role of cytotoxic T lymphocytes and interferon-γ in coronavirus infection: Lessons from murine coronavirus infections in mice. *J. Vet. Med. Sci.* **2020**, *82*, 1410–1414. [CrossRef] [PubMed]
63. Mura, M.; Combredet, C.; Najburg, V.; Sánchez David, R.Y.; Tangy, F.; Komarova, A.V. Nonencapsidated 5′ copy-back defective interfering genomes produced by recombinant measles viruses are recognized by RIG-I and LGP2 but not MDA5. *J. Virol.* **2017**, *91*, e00643-17. [CrossRef]
64. Wignall-Fleming, E.B.; Vasou, A.; Young, D.; Short, J.A.L.; Hughes, D.J.; Goodbourn, S.; Randall, R.E. Innate intracellular antiviral responses restrict the amplification of defective virus genomes of parainfluenza virus 5′. *J. Virol.* **2020**, *94*, e00246-20. [CrossRef] [PubMed]
65. Hasiów-Jaroszewska, B.; Minicka, J.; Zarzyńska-Nowak, A.; Budzyńska, D.; Elena, S.F. Defective RNA particles derived from tomato black ring virus genome interfere with the replication of parental virus. *Virus Res.* **2018**, *250*, 87–94. [CrossRef] [PubMed]
66. Jaworski, E.; Routh, A. Parallel ClickSeq and nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS Pathog.* **2017**, *13*, e1006365. [CrossRef] [PubMed]
67. Schoeman, D.; Fielding, B.C. Coronavirus envelope protein: Current knowledge. *Virol. J.* **2019**, *16*, 69. [CrossRef] [PubMed]
68. Kuo, L.; Masters, P.S. The small envelope protein E is not essential for murine coronavirus replication. *J. Virol.* **2003**, *77*, 4597–4608. [CrossRef] [PubMed]
69. Zhang, R.; Wang, K.; Ping, X.; Yu, W.; Qian, Z.; Xiong, S.; Sun, B. The ns12.9 accessory protein of human coronavirus OC43 is a viroporin involved in virion morphogenesis and pathogenesis. *J. Virol.* **2015**, *89*, 11383–11395. [CrossRef] [PubMed]

70. De Groot, R.J.; van der Most, R.G.; Spaan, W.J.M. The fitness of defective interfering murine coronavirus DI-a and its derivatives is decreased by nonsense and frameshift mutations. *J. Virol.* **1992**, *66*, 5898–5905. [CrossRef]
71. Brian, D.A.; Spaan, W.J.M. Recombination and coronavirus defective interfering RNAs. *Semmin. Virol.* **1997**, *8*, 101–111. [CrossRef]
72. Notton, T.; Sardanyés, J.; Weinberger, A.D.; Weinberger, L.S. The case of transmissible antivirals to control population-wide infectious disease. *Trends Biotechnol.* **2014**, *32*, 400–405. [CrossRef] [PubMed]
73. Levi, L.I.; Rezelj, V.V.; Henrion-Lacritick, A.; Erazo, D.; Boussier, J.; Vallet, T.; Bernhauerová, V.; Suzuki, Y.; Carrau, L.; Weger-Lucarelli, J.; et al. Defective viral genomes from chikungunya virus are broad-spectrum antivirals and prevent virus dissemination in mosquitoes. *PLoS Pathog.* **2021**, *17*, e1009110. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

# CHAPTER 6

## Story of an infection: viral dynamics and host responses in the *Caenorhabditis elegans*-Orsay virus pathosystem

Check for updates

SCIENCE ADVANCES | RESEARCH ARTICLE

VIROLOGY

# Story of an infection: Viral dynamics and host responses in the *Caenorhabditis elegans*–Orsay virus pathosystem

**Victoria G. Castiglioni[1]*†, María J. Olmo-Uceda[1]†, Ana Villena-Giménez[1], Juan C. Muñoz-Sánchez[1], Esmeralda G. Legarda[1], Santiago F. Elena[1,2]***

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

**Orsay virus (OrV) is the only known natural virus affecting *Caenorhabditis elegans*, with minimal impact on the animal's fitness due to its robust innate immune response. This study aimed to understand the interactions between *C. elegans* and OrV by tracking the infection's progression during larval development. Four distinct stages of infection were identified on the basis of viral load, with a peak in capsid-encoding RNA2 coinciding with the first signs of viral egression. Transcriptomic analysis revealed temporal changes in gene expression and functions induced by the infection. A specific set of up-regulated genes remained active throughout the infection, and genes correlated and anticorrelated with virus accumulation were identified. Responses to OrV mirrored reactions to other biotic stressors, distinguishing between virus-specific responses and broader immune responses. Moreover, mutants of early response genes and defense-related processes showed altered viral load progression, uncovering additional players in the antiviral defense response.**

## INTRODUCTION

Viruses, the most abundant biological entities on Earth, infect all known living organisms (*1*, *2*). From bacteria to eukaryotes, most organisms have immune systems to defend themselves against viruses. Viruses, in turn, adapt to their hosts by evolving strategies to evade or neutralize the host's defenses. This ongoing arms race between viruses and their hosts has driven the development of complex interactions and defense mechanisms (*3*, *4*).

The nematode *Caenorhabditis elegans* Maupas lacks adaptive immunity (*5*). Instead, it uses the innate immune system to defend itself against viruses and other microbes, such as its natural pathogen Orsay virus (OrV) (*6*). OrV has a positive-sense, bisegmented RNA genome of ~6.3 kb, which encodes four proteins: an RNA-dependent RNA polymerase (RdRP), a viral capsid protein (CP), protein δ, and a protein fusion of CP-δ (*6*, *7*). Free δ mediates nonlytic viral release (*8*), and the CP-δ fusion is incorporated into the virions as pentameric head fibers (*9*, *10*).

OrV replicates in the intestinal epithelia of *C. elegans* (*6*, *11*), which consists of 20 single-layer cells surrounding the central lumen. Infection of OrV is restricted to the anterior intestinal cells and, after cell egression, results in an orofecal route of transmission (*6*). The infection results in enlarged intestinal lumens and subcellular structural changes. However, the infection does not affect animal life span, suggesting that the innate immune system is able to effectively combat the virus (*12*).

Several conserved antiviral defense mechanisms that *C. elegans* uses against OrV infection have been described, including (i) the RNA interference pathway, which degrades viral RNA and triggers antiviral gene expression (*12*–*14*); (ii) viral RNA uridylation, which targets viral RNA to degradation (*15*); and (iii) expression of antiviral genes, which include protein ubiquitination for targeted degradation and the activation of the intracellular pathogen response (IPR), which is a common response to intracellular pathogens and stresses (*13*, *16*–*18*). In addition, several studies have described pathways necessary for OrV replication, such as zinc and the lipid metabolism (*19*), and cytoskeleton rearrangements (*20*, *21*). The latter have also been reported to have antiviral functions, as have chromatin remodelers (*21*).

Studies of OrV infection have allowed us to better understand virus-host interaction mechanisms. However, there is little known on the progression of the viral infection and in the dynamic interaction between *C. elegans* development and OrV during the course of an infection. To shed light in these matters, we have analyzed the progression of the infection and the host response in a time-resolved manner in the mildly resistant wild-type nematode strain N2. By examining the progression of the viral load, we observed that the infection could be divided into four stages, noted that the ratio of the two genomic RNAs varied during the different phases of the infection, and reported signs of viral egression after the first peak, suggesting that a viral generation would be complete. By looking at the changes in the transcriptome along the four infection stages, we pinpointed a set of genes up-regulated throughout the infection process, highlighted genes most correlated and anticorrelated with virus accumulation, and saw responses to OrV that are frequently linked to other biotic stressors and a unique response to OrV, which differed from the response to other biotic stresses. Last, we identified early responding genes and described proviral and antiviral genes.

## RESULTS

### Dynamics of OrV accumulation during larval development

Wild-type (N2) animals are mildly resistant to OrV infection (*6*). However, inoculations with high doses of OrV have been previously shown to produce high infection rates in N2 (*22*). Here, we used a highly concentrated stock and observed high infection rates in the N2-background strain ERT54, which contains a fluorescent reporter that gets activated upon OrV infection (fig. S1A) (*16*). The use of this highly concentrated stock enabled the study of a synchronous infection in wild-type populations during larval development.

[1]Institute of Integrative Systems Biology (I2SysBio), CSIC-Universitat de València, Paterna, 46980 Valencia, Spain. [2]Santa Fe Institute, Santa Fe, NM 87501, USA.
*Corresponding author. Email: v.g.castiglioni@csic.es (V.G.C.); santiago.elena@csic.es (S.F.E.)
†These authors contributed equally to this work.

Downloaded from https://www.science.org on March 12, 2025

We first followed the progression of the viral load in a synchronized population from hatching to the end of larval development (L4). For this purpose, we inoculated animals at the time of hatching and took samples at increasing hours post-inoculation (hpi) for quantitative reverse transcription polymerase chain reaction (RT-qPCR), RNA sequencing (RNA-seq), and single-molecule inexpensive RNA fluorescence in situ hybridization (smiFISH) analyses (Fig. 1A). Three biological replicates, with 700 animals for RT-qPCR and smiFISH or 1200 animals for RNA-seq, were done for each time point. Viral counts analyzed by RT-qPCRs targeted RNA2 were determined using the standard curve method and were used as a proxy for viral load. OrV reads counted by RNA-seq were relativized to the total reads in each sample and were also used as a proxy for viral load. Both analyses indicated that the progression of the viral load could be significantly divided into four different phases (Fig. 1B; best isotonic regression model $R = 0.7562$, $F_{4,211} = 70.4516$, $P < 0.0001$): (i) a prereplication phase in which the viral load is indistinguishable from background levels (0 to 6 hpi), (ii) an exponential viral replication leading to a viral load peak (8 to 12 hpi), (iii) a phase of moderate viral load levels (14 to 36 hpi), and (iv) a phase of persistent residual infection or resolution of the infection, in which the viral load came close to the initial background levels (38 to 44 hpi). RT-qPCR and RNA-seq quantifications of OrV showed a high correlation [$r_s = 0.9720$, 10 degrees of freedom (df), $P < 0.0001$], confirming that both methods are reliable estimates of viral load.

Furthermore, we sought to characterize the percentage of infected animals throughout larval development. We inoculated animals at the time of hatching and took sequential samples for smiFISH analysis (*23*). RNA1 and RNA2 were labeled independently with 22 and 18 smiFISH probes, respectively. We observed that all the animals that expressed RNA2 also expressed RNA1, but the reverse was not necessarily true. RNA1 encodes the RdRP; therefore, it would be reasonable that its levels would rise faster than RNA2 during early infection. Accumulation of RdRP has been shown to precede the accumulation of the CPs (*11*). We observed that the incidence of infection in the animal population recapitulated the same four phases described above: (i) a prereplication phase in which the percentage of infected animals is close to zero and the smiFISH signal was hardly above background (4 hpi); (ii) an exponential increase in the percentage of infected animals in which the intensity of signal quickly increased, leading to a peak in infection (6 to 14 hpi); (iii) a phase of moderate stable signal (18 to 32 hpi); and (iv) a phase of persistent residual infection in which the percentage of infected animals was notably reduced and the intensity of signal was close to background levels but still detectable in puncta (36 to 44 hpi) (Fig. 1C). Furthermore, the highest percentage of infected animals was observed earlier than the viral load peak, at 8 and 12 hpi, respectively, which may suggest that some animals are either able to resolve the infection quickly or not able to develop a productive infection.

We next asked whether the accumulation of RNA1 and RNA2 undergoes a shift in gene expression from early to late replication cycle stages. We hypothesized that accumulation of RNA1, which codes for the RdRP, would be higher at the beginning of the viral replication cycle when viral replication is higher, while levels of structural proteins from RNA2 would became more abundant toward the end of the viral replication cycle as virions are being assembled (*6*). For this purpose, we looked at both RNA1 and RNA2 labeling with smiFISH and at viral counts relativized to fragment size from RNA-seq. At the initial phases of the infection, a higher percentage of animals accumulate RNA1 in comparison with RNA2, indicating that higher RNA1 levels are needed before transcription of RNA2 (Fig. 1, C and D). Consistently, RNA-seq data indicated that the highest RNA1 to RNA2 ratio is at the beginning of the infection. However, RNA2 was always more abundant than RNA1, with an average $2.40 \pm 1.62$ RNA2 to RNA1 ratio throughout all time points and a ratio of $1.81 \pm 0.43$ RNA2 to RNA1 in our inoculum (Fig. 1E). The ratio shows a dynamic behavior, with higher values coinciding with the viral load peaks and the highest ratio coinciding with the second viral peak, indicating that RNA2 is more abundant before the end of the viral replication cycle and consistent with the need of RNA2-encoded products for encapsidation (Fig. 1E).

Together, our data show that the infection is highly dynamic, with the viral load, the percentage of infected worms, and the RNA2 to RNA1 ratios varying rapidly and going through four distinct infection phases from the beginning to the end of the larval development.

## Spatial progression of infection

Next, we characterized the spatial distribution of viral RNAs in the animal. In the initial infection phases, viral particles were only observed inside intestinal cells (Figs. 1, C and D, and 2, A and B). However, at 14 hpi, viral RNA was first seen in the intestinal lumen, indicating a productive release of viral particles. By 24 hpi, most animals showed a strong signal in the lumen (Fig. 2, A and B). These findings suggest that RNA1 is prioritized for translation and transcription after viral entry. Around 6 hours after RdRP translation (*11*), RNA2 levels begin to rise, reaching their peak at 12 hpi. Following this peak, we notice a decline in RNA2 levels, along with a substantial relocation of OrV to the lumen, signifying egression. Overall, these results imply that a viral replication cycle might conclude in as little as 14 hpi.

OrV has been described to colonize the anterior intestinal cells (*11*), but whether a late infection can also colonize posterior intestinal cells or whether the number of infected cells increases over time is still unknown. To answer these questions, we quantified the relative infected area per animal at 10 and 24 hpi (Fig. 2C). We did not observe a significant difference in the relative infected area at 10 versus 24 hpi (Mann-Whitney test, $P = 0.3713$), suggesting that the infection does not spread within the animal from the first viral load peak to the second viral load peak. We also quantified the total number of infected cells per infected animal during time and did not observe a significant difference at any time point (Kruskal-Wallis test, $P = 0.3239$; Fig. 2D), indicating that in most animals, throughout the viral replication cycle, OrV only colonizes, on median, one to two intestinal cells.

Together, these results show that posterior colonization of the intestine does not take place within the observed time span and that a viral replication cycle might conclude in as little as 14 hpi.

## *C. elegans* dynamic transcriptional response to OrV infection

To better understand the complexity of the host's responses to a viral infection, we next focused on the transcriptomic response of *C. elegans* to OrV through time. For this purpose, we compared the transcriptome of control and infected animals at 12 time points from 0 to 44 hpi (Fig. 1A and table S1) by performing a differential expression analysis (DEA) by time point, which revealed a complex

Downloaded from https://www.science.org on March 12, 2025

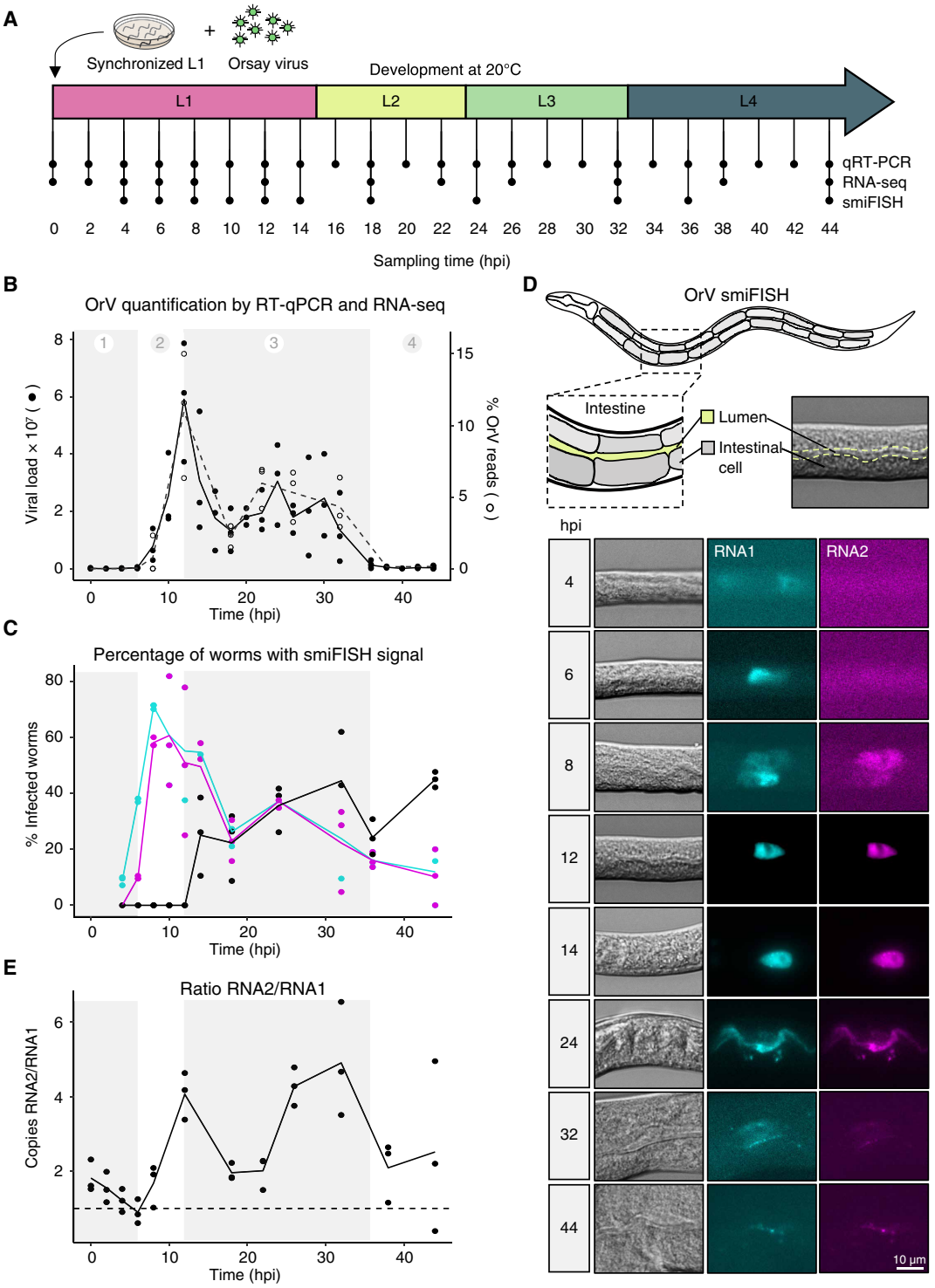



**Fig. 1. Dynamic viral accumulation during larval development.** (**A**) Experimental design. Synchronized recently hatched N2 populations were inoculated with $2.8 \times 10^9$ copies of OrV and grown at 20°C. Samples were analyzed at the indicated time points by qRT-PCR, RNA-seq, or smiFISH. (**B**) Viral load determined by RT-qPCR (left axis, solid circles, and solid line) and percentage of viral reads from the total RNA determined by RNA-seq (right axis, empty circles, and dashed line) over time. The viral load by RT-qPCR was defined as copies of RNA2/10 ng of total RNA. Lines represent the mean; $n = 3$ replicates of 700 or 1200 animals, respectively, per time point. The different stages of the infection, in gray shading, are described in the main text. (**C**) Percentage of animals with cellular RNA1 (cyan), RNA2 (magenta), or with luminal expression of either RNA1 or RNA2 (black) over time as determined by smiFISH. Lines represent the mean; $n = 3$ replicates per time point. (**D**) Representative smiFISH images of inoculated animals at the indicated time points. Left: differential interference contrast (DIC); middle: RNA1 in cyan; right: RNA2 in magenta. (**E**) Ratio of RNA1/RNA2 over time previous relativization to fragment length of each genome determined by RNA-seq; $n = 3$ replicates per time point. See also fig. S1.

Downloaded from https://www.science.org on March 12, 2025

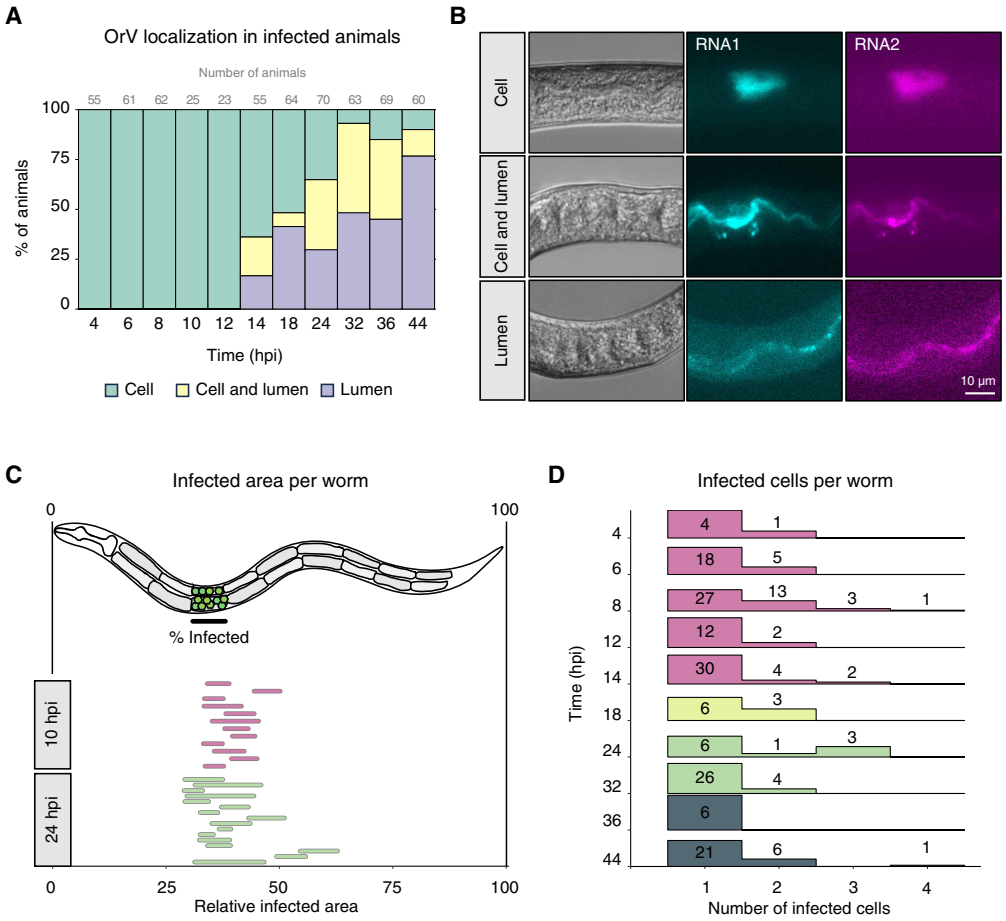



**Fig. 2. Spatial progression of the infection.** (**A**) Percentage of animals with OrV signal in intestinal cells, intestinal cells and luminal space, or luminal space during development as determined by smiFISH. (**B**) Representative smiFISH images of animals with OrV in intestinal cells, intestinal cells and luminal space, or luminal space at 24 hpi. (**C**) Infected area per animal relative to the animal length as determined by smiFISH at 10 (pink; $n = 12$ animals) and 24 (green; $n = 16$ animals). (**D**) Number of infected cells per infected animal over time as determined by smiFISH. Density plot represents the relative abundance of the number of infected cells within each time point. Colors indicate the developmental stage at the time of analysis, as defined in Fig. 1A.

landscape (Fig. 3). This is to date the highest resolution time-series dataset produced in this pathosystem. Our goal was to provide the most sensitive landscape of transcriptional responses, so we considered as differentially expressed genes (DEGs) all genes with a $P < 0.05$, without filtering by magnitude of change. We found a significant correlation ($r_s = 0.9000$, 9 df, $P = 0.0002$) between the number of DEGs (Fig. 3A) and OrV accumulation at each time point (Fig. 1B) with a predominance of up-regulated genes in the first 18 hpi and a more symmetric number of up- and down-regulated responses from 22 hpi on, although the strongest changes still correspond to overexpressed genes (Fig. 3, A and C). Filters on the magnitude of change revealed similar results on the correlation test (DEGs with $|\log_2 FC| \geq 1$: $r_s = 0.8975$, 9 df, $P = 0.0002$; $|\log_2 FC| \geq 2$: $r_s = 0.8909$, 9 df, $P = 0.0004$). Moreover, we identified a group of 29 up-regulated DEGs at all time points with high viral loads (i.e., ≥0.5% of the total sample reads) (Fig. 3B and fig. S3). All these DEGs were up-regulated between 6 to at least 32 hpi; *col-133* was up-regulated at every time point analyzed (Fig. 3C and fig. S3B). Twenty-one of these 29 genes—including *eol-1*, *ddn-1*, and the *pals* genes—were already characterized as part of the IPR (*18*). Among the rest, *math-39* and *Y105C5A.9* were identified as dependent of the Zinc-induced protein-1 (ZIP-1) transcription factor highly associated with IPR (*24*) and positive modulator of nonsense-mediated mRNA decay (*25*). *col-133*, *math-15*, *trpl-5*, *Y43F8B.12*, and *Y6G8.5* had not been previously associated to the animal's response to OrV. A smaller group—formed by *col-133*, *ddn-1*, *eol-1*, *fbxa-182*, *math-15*, *pals-33*, and *pals-37*—was differentially expressed from 6 hpi until the end of the time course (Fig. 3, B and C). Considering only genes with $|\log_2 FC| \geq 1$, the ubiquitous sets are reduced to 15 and 6, respectively (Fig. 3B, genes remaining highlighted in yellow), and the novel players *math-15* and *Y43F8B.12* remain in the selection.

We then performed a functional enrichment analysis (FEA). The FEA was carried out on the up- and down-regulated genes separately, and, contrary to what happened with DEGs, the number of significant categories in the down-regulated responses was bigger than in the up-regulated ones (Fig. 3D). Specifically, we found a first strong response between 12 and 18 hpi in the three Gene Ontologies (GOs) [biological process (BP), molecular function (MF), and cellular component (CC)] and in the metabolic data from Kyoto Encyclopedia of Genes and Genomes (KEGG) and Wikipathways and a second massive response around 26 hpi in all the databases explored (Fig. 3D; complete results in table S2).

Downloaded from https://www.science.org on March 12, 2025

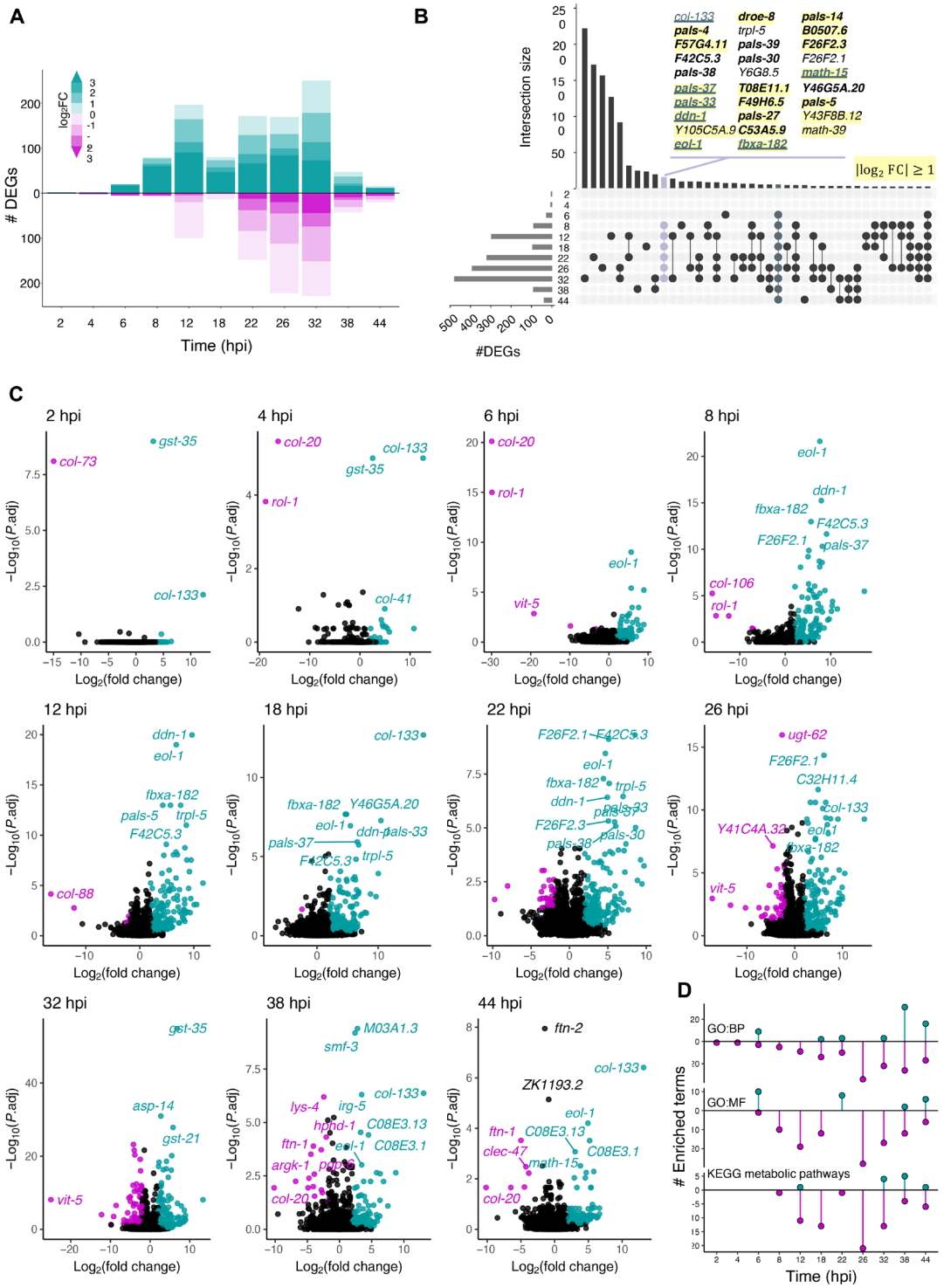



**Fig. 3. *C. elegans* transcriptomic responses across the OrV infection process.** (**A**) Number of significant DEGs (adjusted $P \leq 0.0499$) up- (cyan) and down-regulated (magenta) at each point analyzed. The different intensities indicate the $\log_2$ fold change (FC) value used as filter. (**B**) Upset plot showing the intersections between the DEGS at each time. The points indicate the sets that are counted in the vertical bars. The horizontal bars indicate the total number of DEGs at each time. We found 29 genes ubiquitous to the exponential viral replication and moderate stages (*B0507.6*, *C53A5.9*, *col-133*, *ddn-1*, *droe-8*, *eol-1*, *F26F2.1*, *F26F2.3*, *F42C5.3*, *F49H6.5*, *F57G4.11*, *fbxa-182*, *math-15*, *math-39*, *pals-4*, *pals-5*, *pals-14*, *pals-27*, *pals-30*, *pals-33*, *pals-37*, *pals-38*, *pals-39*, *T08E11.1*, *trpl-5*, *Y105C5A.9*, *Y6G8.5*, *Y46G5A.20*, and *Y43F8B.12*). Seven of them (underlined, dark blue) were persistent from 6 hpi to the end of the infection process. Within the genes with $|\log_2 FC| \geq 1$, the ubiquitous sets contain 15 and 6 genes, respectively (highlighted in yellow). Twenty-one of the ubiquitous genes belong to the IPR (bold). (**C**) Differentially expressed genes at each time point. Cyan and magenta represent up-regulated and down-regulated genes (adjusted $P < 0.05$ and $|\log_2 FC| \geq 1$, respectively). (**D**) Number of significant enriched terms (adjusted $P \leq 0.05$) per time. The FEA was performed on the up and down DEGs separately. GOs biological process (GO:BP) and molecular function (GO:MF) as well as KEGG metabolic pathways for *C. elegans* are shown. See also figs. S2 and S3.

Downloaded from https://www.science.org on March 12, 2025

When focusing on the up-regulated responses in the early stages of infection, at 6 hpi, the following GO:BPs were enriched: nervous system development, axon extension involved in axon guidance, nuclear RNA surveillance, and regulation of transcription DNA templated (fig. S2A); and the GO:MFs related with 5′,3′-exonuclease activity and RNAII polymerase transcription factor activities (fig. S2B).

When focusing on the early down-regulated genes, we observed an earlier enrichment in significant terms. Within the first 2 hpi, there was a down-regulation of genes involved in extracellular matrix organization; within 6 hpi in genes involved in collagen and cuticle development (fig. S2A) and lipase activity (fig. S2B); and within 8 hpi in genes related to helicase activities and uridine diphosphate (UDP)–glycosyltransferase activities (fig. S2B). When analyzing the DEGs using KEGG pathways, we noticed an enrichment in other types of O-glycan biosynthesis at 8 hpi (fig. S2C), a down-regulation in mitochondrial translation and proteolysis at 12 hpi (fig. S2A), and significant effects upon peroxisome, lysosome, metabolism of sphingolipids, pentose phosphate pathway, and ABC transporters (fig. S2C).

During the moderate phase, we observed up-regulated genes related with the defense response to bacterium, the regulation of tube size and epithelial cell differentiation (fig. S2A), and polyadenylate [poly(A)]– and poly(U)-RNA binding and receptor tyrosine kinase binding (fig. S2B). In contrast, the down-regulated genes were associated with the control of proteolysis by protein deneddylation (removal of ubiquitin-like protein), regulation of transport and stress response to $Cd^{++}$, as well as fatty acid elongation (fig. S2A). Towards the end of this phase, genes related to autophagy were also down-regulated.

During the persistent residual infection phase, there was an increase of significant terms among the overexpressed genes, including genes related with the regulation of mitogen-activated protein kinase cascade, Rho protein signaling transduction, mitochondrial unfolded protein response ($UPR^{mt}$), and acetyl-CoA metabolic processes (fig. S2A). Genes related with the glutathione metabolism were enriched from 38 hpi (fig. S2C), and some categories previously identified as down-regulated were up-regulated in this phase, including fatty acid elongation and very long fatty acid metabolic process (fig. S2A), pentose phosphate pathway, and glycolysis/gluconeogenesis (fig. S2C).

## DEGs that correlate or anticorrelate with the virus accumulation

Taking advantage of the time-course experiment design, we performed a likelihood ratio test (LRT) to identify the genes with an expression profile altered by the infection (interaction between condition and time). This contrast revealed a total of 2313 DEGs (adjusted $P$ value $\leq$ 0.0499; Fig. 4A and table S3). For exploratory reasons, the genes were clustered by their profiles on 32 different groups (fig. S4). When focusing in the expression pattern of the 150 genes with the lowest adjusted $P$ value (the most significant genes from hereon) (Fig. 4B), we noticed that they could be grouped into five different profiles (Fig. 4, C to G), distinguishing between early up-regulation (Fig. 4C), late up-regulation (Fig. 4, D and E), and late down-regulation (Fig. 4, F and G). Some of the genes highlighted above (*ddn-1*, *eol-1*, and *ZC196.3*), and multiple *pals* (*pals-1*, *-3*, *-4*, *-5*, *-27*, *-30*, *-32*, *-37*, *-38*, and *-39*) share the early up-regulation profile (Fig. 4C).

Once the genes with a differential expression pattern due to infection were identified, we searched among them for those that would present a differential expression pattern correlated with the dynamics of OrV accumulation. A total of 45 genes showed a highly positive correlation ($r \geq 0.8011$, 9 df, $P \leq 0.0030$) between their differential expression and OrV accumulation (Fig. 4H), while 13 presented an anticorrelated profile ($r \leq -0.8012$, 9 df, $P \leq 0.0030$) (Fig. 4I). Among the genes with a positive correlation, two had a $r \geq 0.9148$ (9 df, $P < 0.0001$) *bris-1*, enriched in neurons and predicted to enable guanyl-nucleotide exchange factor activity and to be involved in regulation of ADP-ribosylation factor (ARF) protein signal transduction, and *F46A8.13*, a long noncoding RNA also enriched in neurons. In contrast, *elo-9*, a protein with fatty acid elongase activity, had an $r = -0.9131$ (9 df, $P \leq 0.0001$). Some of the genes with a positive correlation were already associated with OrV response: *Y39G8B.5* and *ZC196.3* are cataloged as part of the IPR (*18*), and *C17E4.6*, *pes-4*, *pxf-1*, *taf-6.2*, *T07A5.7*, and *Y39G8B.5* are ZIP-1 dependent (*24*). Moreover, *clec-24*, a C-type lectin predicted to be associated with the membrane and an antimicrobial effector, displays a differential expression pattern with a negative correlation with virus accumulation. Among the aforementioned, *T07A5.7*, *Y39G8B.5*, and *ZC196.3* show the expression profile in Fig. 4C, while *acs-5* is a characteristic representant of the expression profile shown in Fig. 4G.

Together, we found a group of 45 genes with a differential expression profile that highly correlates with OrV accumulation ($r \geq 0.8$), with IPR genes such as *ZC196.3*, and 13 genes with a negative correlation ($r \leq -0.8$), such as *tpi-1*.

## Shared responses with other infections

After characterizing the animal's responses to OrV infection on a temporal scale, we wanted to tease apart those specific to OrV infection from those common to other biotic stresses. For this purpose, we took advantage of the transcriptomic results available at WormExp 2.0 Microbes database (*26*) and the results of four more experiments performed with OrV, *Nematocida parisii* and *Myzocytiopsis humicola* (*13*, *27*). A total of 56 experiments were included in this analysis (fig. S5 and table S4), comparing the results obtained in wild-type N2 animals upon infection with OrV ($n = 12$), Gram-negative ($n = 21$) and Gram-positive ($n = 18$) bacteria, and fungi ($n = 5$). With this analysis, we found 188 protein coding genes associated only to the OrV transcriptomic response (OrV-specific genes from now) and 208 genes affected by all pathogens categories (Fig. 5A). Differences in the developmental stage of the nematode at the time of inoculation, the sampling time since infection, the transcriptomic technique used, and the transcriptomic analysis itself likely contribute to the differences, but our goal here was to sketch a broad picture of comparisons that could open new avenues for further analysis.

We noticed a variety of gene expression profiles within the OrV-specific responding genes (Fig. 5B). A functional enrichment of the OrV-specific genes highlighted the cytoplasm and the mitochondria components and cellular metabolic processes. Amino acid biosynthesis and peptide-related categories were also significantly enriched, as well as the pentose phosphate pathway (fig. S6B and complete results in table S6). Some of the OrV-specific genes—*akir-1*, *ash-2*, *pxf-1*, *smu-2*, *taf-6.2, unc-57*, *W04A8.1*, *Y65B4A.4*, and *ztf-6*—are among those positively correlated to virus accumulation (Fig. 4H and fig. S6A, blue border), while *F46B6.6* and *tpi-1*

Downloaded from https://www.science.org on March 12, 2025

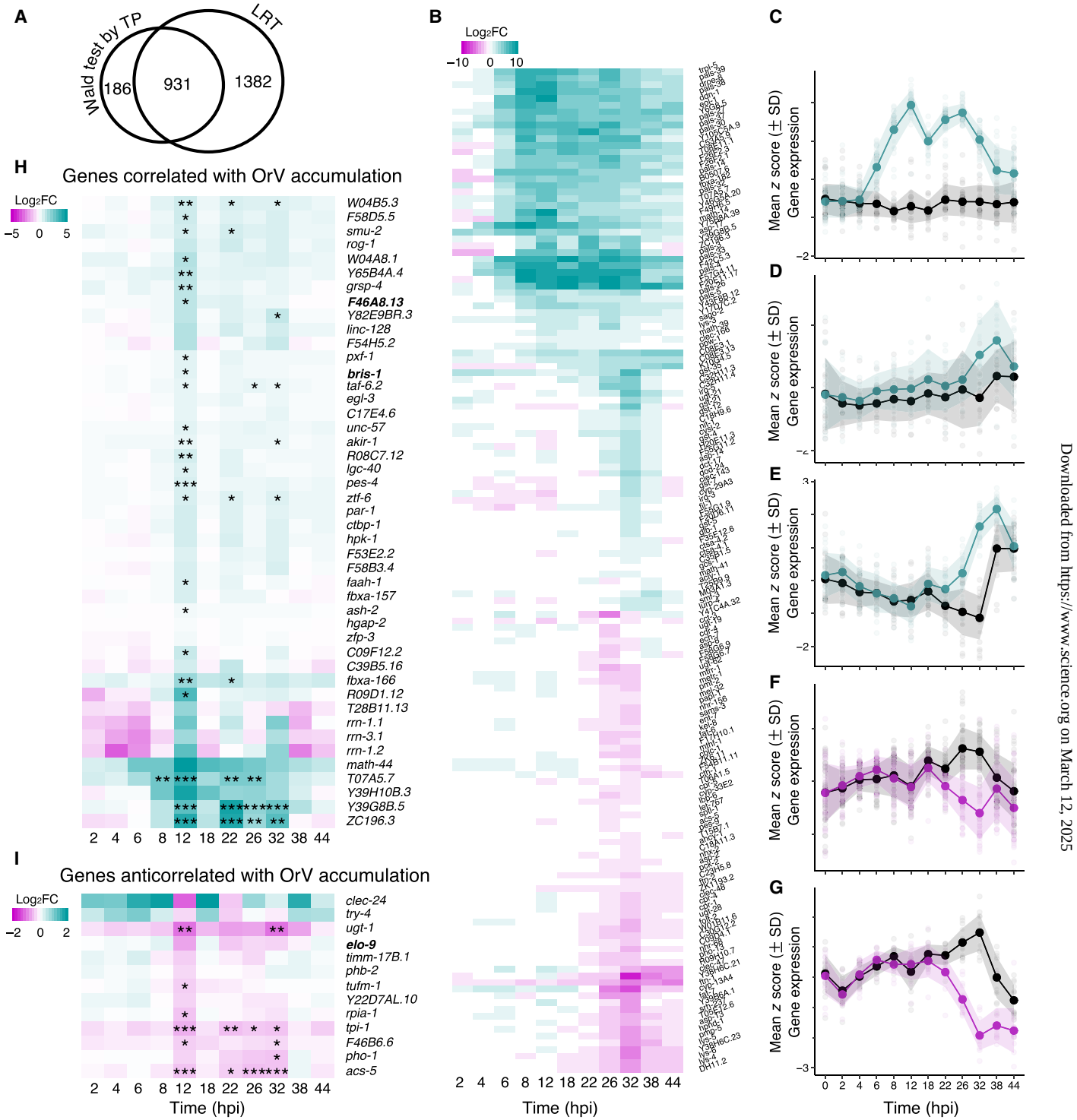



**Fig. 4. Differential expression profiles.** (**A**) Total number of DEGs obtained by DEA at individual time point (TP). Wald test and LRT were used to obtain those with an expression profile affected by the infection process. (**B**) Differential expression profile of the 150 most significant genes (smaller *P*-adjusted values) detected with the LRT. (**C** to **G**) Representative expression profiles of the 150 most significant genes in (B). The mean *z* score (±1 SD) of each cluster is represented by the line and the shadow, respectively. Lighter points represent the individual expression values for each gene with black representing the controls. Cyan or magenta represents the infected group, and the general up- or down-regulation of the cluster, respectively. (**H**) Expression profile of DEGS correlated ($r \geq 0.8011$) with viral accumulation. In bold, genes with $r \geq 0.9148$. (**I**) Expression profile of DEGs anticorrelated ($r \leq -0.8012$) with viral accumulation. In bold, $r = -0.9131$. Asterisks indicate adjusted significance level at each time point: $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, and $^{***}P \leq 0.001$. See also fig. S4.

Downloaded from https://www.science.org on March 12, 2025

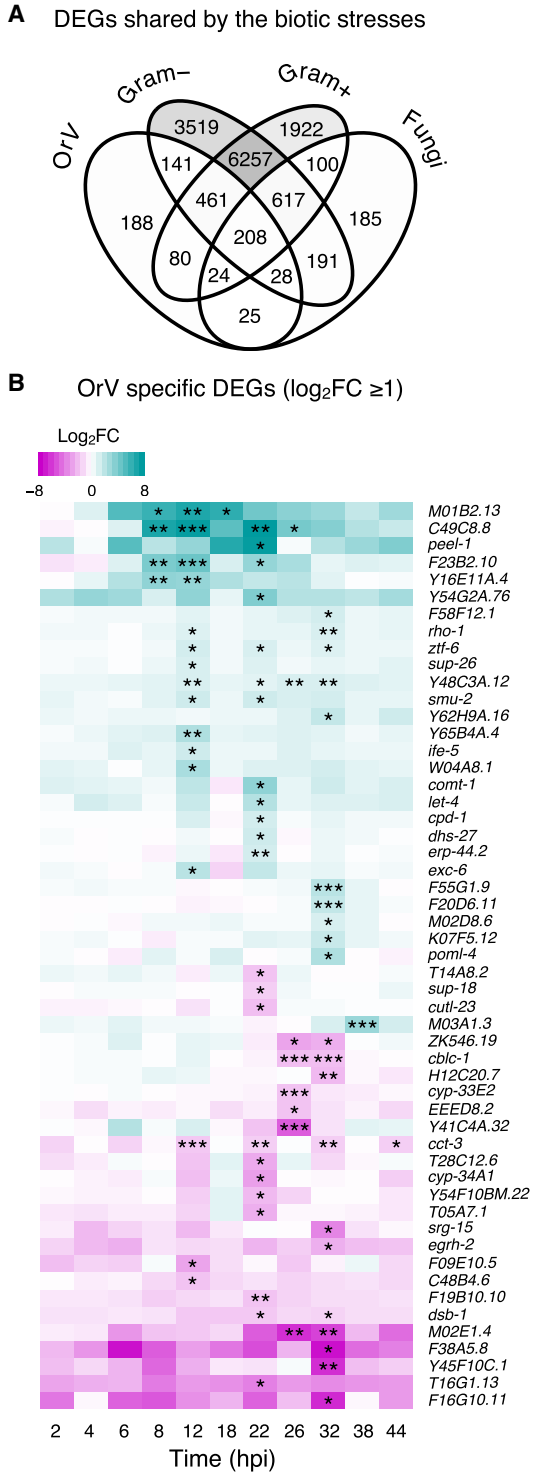



**Fig. 5. Comparison with other biotic stresses.** (**A**) Intersections between the different categories of stresses included in the analysis. A gene was included in a biotic stress category when was detected as DEG in any experiment belonging to the pathogen category. Only protein coding genes were considered. There are a group of 188 genes in the OrV response that do not appear in the other experiments included (experiments in N2 were compared). Genes (208) are common to all the categories studied. Greater gray intensity reflects greater overlap size. (**B**) Expression profile of the OrV-specific DEGs with a $|\log_2 FC| > 1$ for at least one time point. Asterisks indicate adjusted significance level at each time point: $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$. See also figs. S8 to S10.

are anticorrelated to OrV viral load (Fig. 4H and fig. S6A, burgundy border). The topology analysis also revealed *tpi-1*, a triose phosphate isomerase involved in several metabolic pathways, as the most connected gene in the StringDB network (fig. S6A).

Next, we asked whether these specific responses to viral infections were shared with other species of *Caenorhabditis*. To this end, we compared the DEGs upon infection of *Caenorhabditis briggsae* with Santeuil virus (*13*) with our results. We found that ~32% of the genes reported in (*13*) were also affected in this OrV–*C. elegans* experiment (table S4). From them, only *C49C8.8* (WBGene000425 in *C. briggsae*) was differentially expressed upon infection with both viruses but not with the other biotic stresses (table S5). *C49C8.8*, which is overexpressed at 8 hpi and has an expression pattern resembling the damping waves profile (fig. S4, cluster 8), appears to be the most specific to both virus responses.

Within the set of genes affected by all pathogens (Fig. 5A), we found 23 of the previously identified as ubiquitous genes (*B0507.6*, *C53A5.9*, *ddn-1*, *droe-8*, *F26F2.1*, *F26F2.3*, *F42C5.3*, *F49H6.5*, *fbxa-182*, *math-15*, *pals-4*, *pals-5*, *pals-14*, *pals-27*, *pals-30*, *pals-33*, *pals-37*, *pals-38*, *pals-39*, *T08E11.1*, *trpl-5*, *Y105C5A.9*, and *Y46G5A.20*). In addition, a functional enrichment of this general responses highlighted the extracellular region, the innate immune responses, collagen- and cuticle-related genes, and metabolic pathways related with the biosynthesis of lipids (fig. S7 and table S7).

Next, we focused on three specific immune-related categories: IPR response, antimicrobial effectors, and lipid metabolism. We found 65 of the 80 IPR genes within our DEGs (adjusted $P \leq 0.0468$), with activation responses earlier than previously documented (fig. S8A) (*28*), but more surprisingly, 60% of IPR genes were altered in all pathogen categories (fig. S7A, purple border). More specifically, some of the IPR genes were involved in the response to extracellular pathogens: 52.5% of the IPR genes were differentially expressed at 12 hpi with *Bacillus thurigiensis* and 48.8% at 24 hpi with *Enterococcus faecalis* (fig. S8B) (*26*), suggesting that genes involve in the IPR could play a role in the response to a wider range of pathogens.

We also noticed more antimicrobial effectors involved in antiviral responses than previously reported (*29*): Several C-type lectins were up- and down-regulated (fig. S9A) (*30*); the lysozyme *lys-3* was up-regulated at 18 and 26 hpi; *lys-2* at 32 hpi; and *lys-4*, *lys-5*, and *lys-6* were down-regulated from 26 hpi. Moreover, the antimicrobial effectors *clec-10*, *clec-45*, *clec-51*, *clec-53*, *clec-67*, *clec-70*, *clec-72*, *clec-80*, *clec-82*, *clec-166*, *clec-227*, *cnc-4*, *lys-2*, *lys-5*, and *lys-6* were described as differentially expressed upon OrV, Gram-positive and Gram-negative bacteria, and fungi infections (fig. S9B).

The lipid metabolism plays key roles in a variety of diseases, whether pathogenic in origin or not (*31*). The feedback between the OrV replication and lipid metabolism has been already studied (*19*). Here, we found almost 35% of the genes associated with lipid metabolism (GO0006629) in our DEGs list (fig. S8A). Of these, *hyl-1*, *pcyt-2.1*, *psd-1*, *scp-1*, and *T26C12.1* were among the OrV-specific genes, and *hyl-1* and *scp-1*, which were significantly up-regulated at 18 and 22 hpi, respectively, are predicted to be regulatory genes. A general trend toward down-regulation was observed, as previously described (*19*). However, the biosynthesis of unsaturated fatty acids was one of the most significant categories enriched in the common genes (fig. S7). Notably, the metabolism of lipids is also important in the responses to other biotic stresses (fig. S8B), and *dld-1*, *fat-2*, *fat-5*, *fat-6*, and *sptl-2* are part of the common response to all pathogens tested so far.

Downloaded from https://www.science.org on March 12, 2025

Together, in these comparisons, we have identified a group of 208 genes altered in all pathogen categories where the extracellular compartment and the lipid metabolism are revealed as enriched, which includes 60% of the IPR genes. In contrast, a set of 188 genes were differentially expressed only in response to OrV, with a predominance of the cytoplasm and mitochondria components and amino acid– and peptide-related categories. The repertoire of antimicrobial effectors used by the animal against OrV infection has been also expanded.

### Pro- and anti-viral functions of DEGs

We lastly sought to perform an initial characterization of the roles of some of the DEGs upon OrV infection. We selected the secondary argonautes *ppw-1*, *sago-2*, and *vsra-1*, which were differentially up-regulated early in infection (fig. S9A). We also selected *lys-2*, a lysozyme involved in the defense response to bacteria and fungi, and the antimicrobial peptide *clec-60* due to their roles in the defense response against microorganisms (fig. S9A). In addition, we selected *gst-35*, a glutathione transferase that may have roles in detoxification and cell signaling; *eol-1*, an RNA pyrophosphohydrolase involved in the regulation of olfactory learning; and *C53A5.9*, a paralog of influenza virus NS1a binding protein homolog because they were significantly up-regulated during early infection (2 to 6 hpi) (fig. S9A). Last, we included *drh-1* and *pals-22* mutants as controls for sensitivity and resistance, respectively (*12*, *18*).

To assess the role of these proteins upon OrV infection, we studied whether the absence of these proteins would affect viral accumulation and infectivity. We first examined the impact of these proteins on OrV infectivity by looking at three key factors: the percentage of infected animals (Fig. 6, A and B), the RNA2 signal-positive animals (Fig. 6C), and the number of infected cells at 12 hpi (Fig. 6D), which aligns with the initial peak of viral load in wild-type animals (Fig. 1B). For infectivity, the fit of the data to a logistic regression revealed significant differences among genotypes ($\chi^2 = 394.9264$, 10 df, $P < 0.0001$). Further Bonferroni sequential post hoc test identified five distinct groups (adjusted $P \leq 0.0395$). Mutants *clec-60*, *gst-35*, *lys-2*, and *sago-2* showed no significant difference from the wild type. By contrast, *pals-22*(*jy3*) and *ppw-1*(*pk2505*) formed a group with an ~80% reduction in infectivity, while *eol-1*(*tm6609*) and *vsra-1*(*tm1637*) had a significant but smaller reduction (33.92 and 50.47%, respectively). Conversely, *drh-1*(*ok3495*) was 17.96% more susceptible to infection.

Regarding the RNA2-positive animals, significant differences were found among genotypes (logistic regression: $\chi^2 = 176.7448$, 10 df, $P < 0.0001$; Fig. 6C). Three significant groups emerged (post hoc–adjusted $P \leq 0.0473$). *pals-22*(*jy3*) had an 88.36% lower accumulation of RNA2 compared to wild type relative to RNA1. *ppw-1*(*pk2505*) and *vsra-1*(*tm1637*) showed a moderate reduction (~36%). The rest of the mutants showed no significant difference from wild type.

We then analyzed the distribution of infected cells per animal (Fig. 6D). Because of the non-normal distribution and heteroscedasticity of the data, a Kruskal-Wallis test was used, showing significant differences among genotypes ($\chi^2 = 184.4403$, 10 df, $P < 0.0001$). Games-Howell post hoc analysis revealed three distinct groups (adjusted $P \leq 0.0473$). At the one side, *pals-22* (*jy3*) had only one infected cell per animal, significantly less than the wild type. At the other side, *lys-2*(*tm2398*) and *drh-1*(*ok3495*) had roughly 53% more infected cells than wild-type animals.

Together, these results suggest that LYS-2 and DRH-1 prevent multicellular OrV colonization and PALS-22, PPW-1, and VSRA-1, which had lower infectivity; a reduction in RNA2 expression relative to RNA1; and fewer infected cells per animal, hinder OrV amplification early in viral replication. The other mutants behaved similar to the wild type in these analyses.

To enhance our understanding of these proteins' roles in OrV infection, we proceeded to examine the viral accumulation in the mutants. We monitored the progression of viral load in synchronized mutant populations from hatching to the end of larval development (L4 stage). OrV inoculation took place upon hatching, and samples were collected at various hpi for RT-qPCR analysis (Fig. 6E). Five replicates with 700 animals each were used per condition and time point. The data from the accumulation curves were fitted to a generalized linear model with a gamma distribution and log-link function. Animal genotype and hpi were considered as model factors. We specifically looked for differences in the temporal dynamics of OrV accumulation between mutant and wild-type animals and conducted sequential Bonferroni post hoc pairwise tests. We observed significantly higher OrV accumulation in *lys-2*(*tm2398*) and *sago-2*(*tm894*) ($P < 0.0001$), with *lys-2*(*tm2398*) showing a fourfold increase in viral load at 10 hpi, and *sago-2*(*tm894*) exhibiting a 1.6-fold increase at 12 hpi. This suggests that these genes may play a direct role in the antiviral response. Conversely, we noted significantly lower OrV accumulation in the other secondary argonautes, *ppw-1*(*pk2505*) and *vsra-1*(*tm1637*) ($P \leq 0.0001$), with viral load peaks at 18 hpi. This trend was also observed in the antimicrobial peptide *clec-60*(*tm2319*) ($P < 0.0001$), along with a disrupted OrV accumulation pattern in *C53A5.9*(*tm10539*) and *eol-1*(*tm6609*) ($P \leq 0.0009$). Last, we did not detect differential OrV accumulation in *gst-35*(*esv1*) ($P = 0.2109$), suggesting that its altered expression upon infection may be an indirect effect unrelated to the defense response. However, we cannot rule out the possibility that redundancy within the glutathione transferase family might be masking a potential role in OrV infection.

From these experiments we conclude that *C53A5.9*, *clec-60*, *eol-1*, *ppw-1*, and *vsra-1* had proviral functions, while *lys-2* and *sago-2* appear to have antiviral functions. *Gst-35* appears to be either redundant or not involved in the defense response. However, more experiments would be necessary to clarify the role of any of these proteins in OrV infection.

## DISCUSSION

### OrV dynamics along larval development

In this study, we followed the progression of a viral infection within a population, which was synchronously infected. Regardless of the quantification method used, we observed that the infection was highly dynamic and followed four phases: (i) a prereplication phase in which the viral load and percentage of infected animals remained low and RNA1 was preferentially transcribed, which lasted ~6 hours; (ii) an exponential viral replication phase in which the viral load and the number of infected animals rapidly increased, spanning from 8 to 14 hpi; (iii) a phase of moderate replication in which the percentage of infected animals and the viral load remained fairly constant, from 16 to 36 hpi; and (iv) a phase of persistent residual infection from 38 to 44 hpi (Fig. 7). The ratio of RNA2 to RNA1 was dynamic throughout the infection and started with ~2 RNA2 molecules per RNA1 molecule, which opens the possibility that more RNA2 gets

Downloaded from https://www.science.org on March 12, 2025

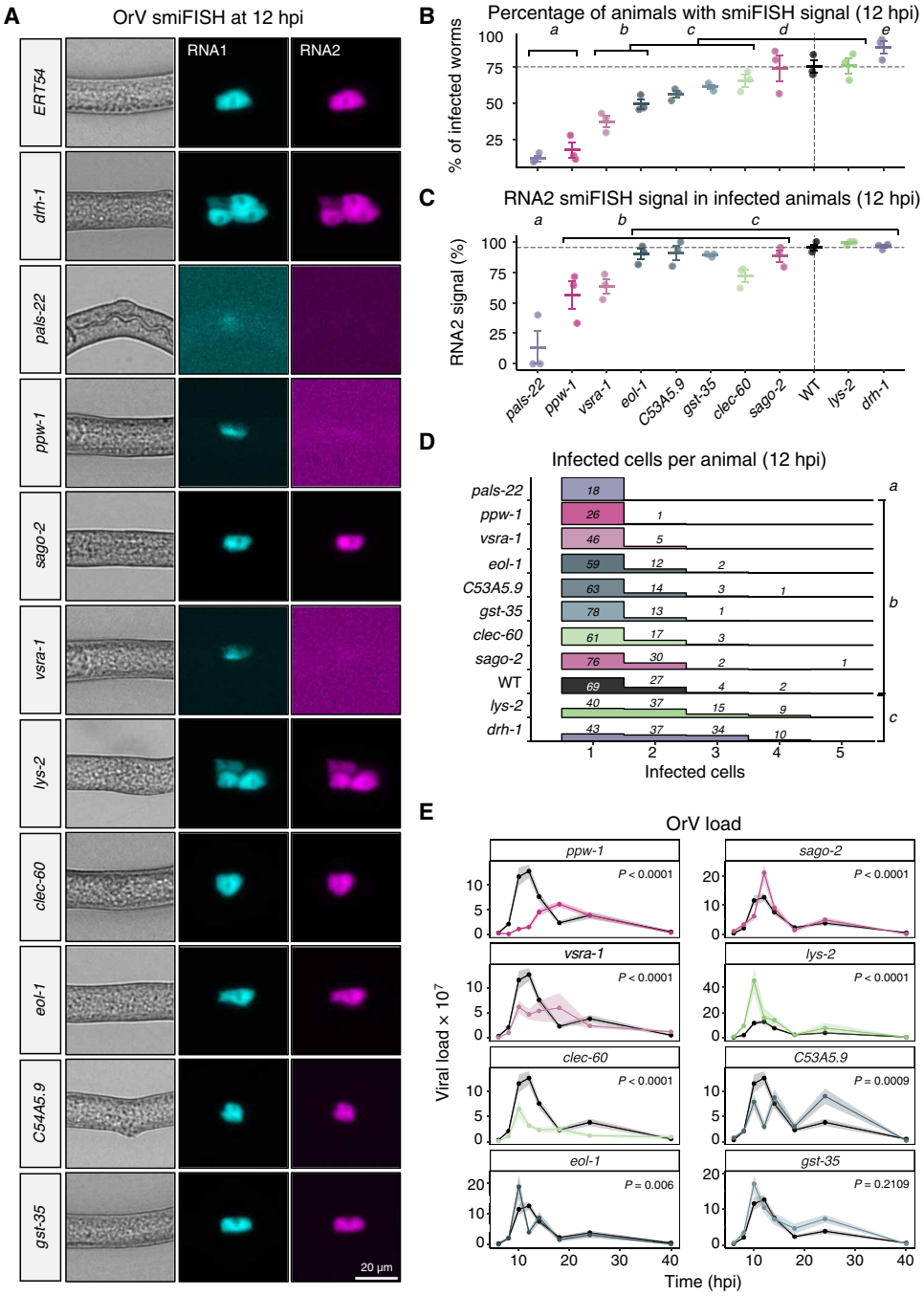



**Fig. 6. OrV dynamics in mutants of identified DEG.** (**A**) Representative smiFISH images of wild-type, *clec-60*, *C53A5.9*, *drh-1*, *eol-1*, *gst-35*, *lys-2*, *pals-22*, *ppw-1*, *sago-2*, and *vsra-1* mutants at 12 hpi. (**B**) Percentage of infected animals at 12 hpi as determined by smiFISH. $n = 49 \pm 5$ per point. Error bars represent ±1 SD. (**C**) Ratio of RNA2 to RNA1 positive animals determined by smiFISH in the indicated mutants. $n = 27 \pm 12$ per point. Error bars represent ±1 SD. (**D**) Number of infected cells per infected animal over time as determined by smiFISH. Density plot represents the relative abundance of the number of infected cells within each mutant. (**E**) Viral load (±1 SEM; $n = 5$ replicates per time point) determined by RT-qPCR at 6, 8, 10, 12, 14, 18, 24, and 44 hpi in *clec-60*, *C53A5.9*, *eol-1*, *gst-35*, *lys-2*, *ppw-1*, *sago-2*, and *vsra-1* mutants (colored lines and dots) as well as wild type (black lines and dots). *P* values correspond to the result of the sequential Bonferroni post hoc pairwise test comparing the temporal dynamics of OrV accumulation on each mutant with the wild-type animals. See also fig. S9.

Downloaded from https://www.science.org on March 12, 2025

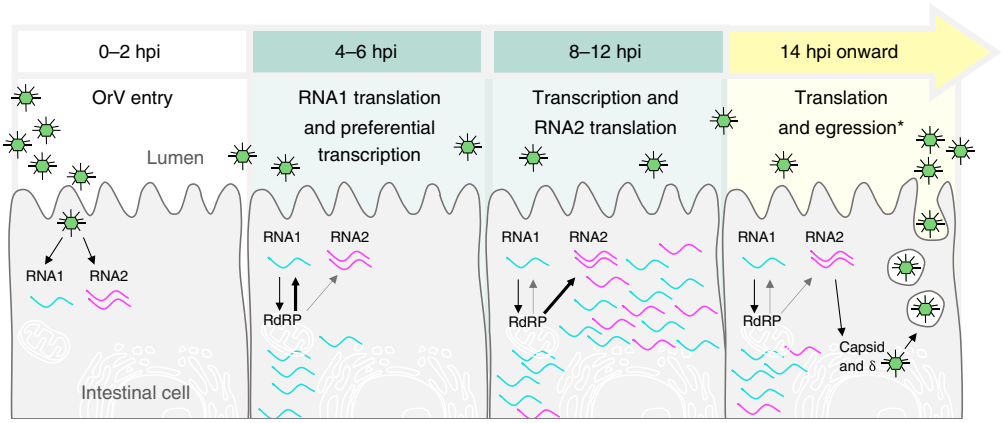



**Fig. 7. Model representation of the OrV infection cycle.**

encapsulated in each virion. Moreover, RNA1 preferentially accumulated during the initial part of the exponential viral replication phase. During the second part of the exponential viral replication phase, RNA2 levels increased up to four times more than RNA1 levels, suggesting that the structural proteins encoded in RNA2 were needed to complete the viral replication cycle and egression. In agreement with these results, we started observing OrV smiFISH signal in the lumen at 14 hpi, which allowed us to determine the beginning of egression at around 14 hpi. Together, our data suggests that OrV accumulation during larval development is highly dynamic and that a viral replication cycle would be complete at 14 hpi, which would be consistent with the increase in RNA2 levels leading up to 12 hpi and the decrease in viral load after 12 hpi (Fig. 7).

Moreover, we observed that the number of infected cells per animal remained close to one throughout the rest of the infection, suggesting that the increase in viral load we observed by RT-qPCR and RNA-seq is due to intracellular viral replication, possibly in the same infected cells. The ability of neighboring cells toward off the virus raises the possibility that the infected cells share some diffusible signal with their neighbors to induce an antiviral state, as has been observed in mammalian cells (*32*). In *lys-2* and *drh-1* mutants, the number of infected cells was increased, as was the viral load, suggesting that the increase in viral load may be due to an increased number of cells and that LYS-2 and DRH-1 prevent multicellular colonization. Another member of the RNAi pathway, RDE-1, is also necessary to prevent multicellular colonization (*33*), hinting that DRH-1 may prevent multicellular colonization due to its role in the RNA interference (RNAi) pathway and not through activating the IPR response. Last, the viral load in *lys-2*(*tm2398*) was reduced to wild-type levels after the initial peak, suggesting that *lys-2* is involved in preventing multicellular colonization but, unlike *drh-1*, is not involved in intracellular defense.

We also observed that the relative infected area per animal did not increase over time, which may be due to the differential localization of antiviral genes within the intestine. It has been recently reported that intestinal cross-linking collagens, which are restricted to the posterior intestinal cells, mediate antiviral immunity in the intestine (*21*), with intestinal RNAi of different collagens resulting in increased viral loads. The increase in viral load upon RNAi of these collagens may thus be due to an increased number of infected cells, as we observed in the *lys-2* and *drh-1* mutants.

By analyzing the transcriptomic response during the whole span of larval development, we have been able to characterize temporal responses to OrV infection. We identified a group of genes upregulated through the whole infection and genes whose expression profiles were correlated or anticorrelated with viral accumulation. A comparative analysis of the responses elicited by other biotic pathogens in wild-type animals teased apart OrV-specific genes from defense genes responding in a non-pathogen–specific manner.

Last, the developmental stage of the host has an impact on the response to viral infections (*34*). Our experiments were performed in L1 larvae, unlike a previous study in which the transcriptomic response to OrV in wild-type animals was analyzed in L3 larvae (*13*). At 12 hpi, we see some key differences, such as the absence of molting-related genes in the L3-infected dataset. This is probably because, in their case, the L3 molt had already passed, and in our case, the 12 hpi was close to the L1 molt. Taking into consideration the developmental timing of the animals in future experiments will help us understand the effect of the viral infection in development and distinguish specific responses to OrV.

## DEGs with pro- and antiviral roles

Our transcriptomic analysis provides, to our knowledge, the best transcriptomic temporal resolution of an OrV infection process. In an effort to characterize some of the genes that are differentially expressed in OrV infection, we followed the progression of the infection in eight mutants of differentially expressed genes upon OrV. One mutant per gene of interest was used in this analysis in an effort to try to perform an initial characterization of the function of each of these genes in OrV infection. A more in-depth analysis, with independently isolated mutants of these genes, would be necessary to improve our understanding of the function of these genes. Of the eight mutants analyzed, the following findings were of particular interest:

We observed that *eol-1*, a gene related with the nervous system, plays an important role from the first stages of the infection: The absence of EOL-1, which is related to olfactory learning, has an impact on the viral load profile, with a higher maximum viral load than the control at an earlier time point. Moreover, a subsequent decline in viral load suggests an earlier activation of the defense response, as the infection process continues similarly than that of wild-type animals. The involvement of *eol-1* in the response to the virus is unclear, but our results suggest that it contributes to shape the process, possibly due to erratic behavior that fails to avoid the virus.

Our analysis of *clec-60*, which is involved in antimicrobial defense against *Staphylococcus aureus* and *Microbacterium nematophilum*

Downloaded from https://www.science.org on March 12, 2025

(*35*, *36*), indicates that *clec-60* is also involved in viral infection in *C. elegans* and may have proviral roles, as *clec-60*(*tm2319*) had lower viral loads. Similarly to what was observed in other RNA viruses, C-type lectins could be involved in OrV entry into host cells (*37*, *38*). However, further studies on the role of *clec-60* and other C-type lectins in OrV infection would be necessary to clarify this.

Analysis of the lysozyme *lys-2* indicated that this gene was up-regulated within the first 2 hpi. *lys-2* was previously described to be regulated by all microbes; it was exposed to except by OrV and *Yersinia pseudotuberculosis* (*29*). Our results indicate that *lys-2* is also regulated by OrV. We did not observe a higher percentage of infected animals in *lys-2*(*tm2398*) when compared to wild-type animals, possibly due to the highly concentrated viral stock used in these experiments. However, we observed a significant increase in *lys-2*(*tm2398*) viral load upon OrV infection—with a viral peak four times higher than in wild-type animals—suggesting that *lys-2* has antiviral effects in *C. elegans*. Moreover, we observed a notable increase in the number of infected cells at 12 hpi, suggesting that *lys-2* prevents multicellular colonization. The decrease in viral load in *lys-2* mutants after the viral load peak suggests that the intracellular defense response, which is able to clear OrV after colonization, is not affected in *lys-2* mutants, suggesting that the antiviral properties of *lys-2* rely in the prevention of multicellular colonization and not in intracellular defense mechanisms.

Last, we studied the role of the Argonautes (AGOs), which were differentially expressed upon OrV infection (SAGO-2, PPW-1, and VSRA-1), all of which were overexpressed. Analysis of the AGO mutants upon OrV infection suggested that these AGOs have different functions during the infection. *Sago-2*(*tm894*) showed a higher viral load coinciding with the 12-hpi viral load peak, confirming that SAGO-2 is necessary in the defense response against OrV. We did not observe a higher percentage of infected animals in *sago-2*(*tm894*) when compared to wild-type animals, possibly due to the saturating viral stock used in these experiments. In contrast, *ppw-1*(*pk2505*) and *vsra-1*(*tm1637*) showed reduced viral loads, percentage of infected animals, and RNA2 to RNA1 ratios. These results suggest that PPW-1 and VSRA-1 are necessary for viral replication and that viral replication in these mutants is stalled after the beginning of RNA1 replication but before RNA2 replication, similarly to what was observed with ALG-1 (*39*), PALS-22 (*18*), and *Nicotiana benthamiana* AGO1 (*40*). However, it could also be that the loss of PPW-1 and VSRA-1 cause a misregulation of the immune system, which benefits the host and results in reduced OrV levels. PPW-1 regulates immune-responsive genes (*41*). Although the function of PPW-1 and SAGO-2 seems notably different in regard to the defense response to OrV, both proteins have 98.65% amino acid similarity.

### Influence of OrV on lipid metabolism and the UPR$^{mt}$

Our study unveiled a dynamic pattern of immune, stress, and metabolic responses throughout the OrV infection process. While extensive research on the impact of viral infections on metabolism has been conducted in mammals (*42*), studies in *C. elegans* have predominantly focused on bacterial infections (*43*).

Viral infections affect the lipid metabolism in several pathosystems: Some viruses trigger the accumulation of lipid droplets, while others reduce the lipid storage (*44*). OrV infection leads to a reduction in lipid accumulation, although OrV itself relies on lipids for effective infection (*19*). Our data revealed a general down-regulation of lipid metabolism genes from 22 hpi, with down-regulation of some responses as early as 12 hpi. Given that wild-type animals are capable of reducing the viral load, the reduction of the lipids could be modulated by the virus—as it is the case in other pathosystems (*45*–*47*)—or could be part of an active response of the animal. Certainly, the fatty acid oleate, generated by *fat-6* and *fat-7*, plays a pivotal role in activating the innate system of *C. elegans* (*48*). This significance is underscored by the inclusion of *fat-2*, *fat-6*, or *fat-7* as some of the DEGs shared by biotic stresses in all the categories tested (in addition to *acox-1.5*, *dld-1*, *fat-5*, *gba-4*, *sptl-2*, and *Y49E10.18*). This implies that *C. elegans* might have the capacity to adjust its lipid metabolism as a component of an engaged antiviral defense. Nevertheless, additional research is warranted to delve deeper into this phenomenon and gain a more comprehensive understanding.

Our study highlights the substantial impact of viral infection on mitochondria, a crucial cellular component for maintaining balance through pathways such as the unfolded protein response (UPR$^{mt}$). This activation is vital for managing stress (*49*, *50*). Under normal conditions, cytosolic activating transcription factor associated with stress-1 (ATFS-1) is transported into the mitochondria via the translocase of the outer membrane (TOM) complex for degradation. If this translocation fails, then ATFS-1 enters the nucleus, triggering the expression of genes associated with mitochondrial chaperones, proteolysis, glycolysis, and innate immunity, such as *lys-2* (*51*), which we found to have antiviral effects. Although we observed an enrichment of UPR$^{mt}$ at 38 hpi, coinciding with the onset of the residual infection phase in which viral load decreased significantly, several indications point to an earlier pathway activation: (i) Genes linked to mitochondrial translation were down-regulated at 12 hpi; (ii) components of the TOM complex, *tomm-22* and *tomm-70*, were also down-regulated at 12 and 26 hpi, respectively, hindering ATFS-1 translocation; and (iii) ZIP-3, known to inhibit UPR$^{mt}$ activation (*49*), displayed reduced expression (cluster 5 in fig. S4). In contrast, ZIP-1, identified as a key regulator in the IPR (*24*), showed an up-regulated expression profile (cluster 9 in fig. S4). The expression patterns of both bZIP transcription family members align with the activation of immune genes through both pathways.

### Concluding remark

*C. elegans* stress responses often share common ground, and a single stressor can trigger various pathways. Consider, for instance, a proteotoxic stress that not only kickstarts the innate immunity but also sparks a heat shock response and activates the UPR$^{mt}$. Moreover, the developmental stage of the host can influence the nature of these responses. While it may sometimes seem like untangling a tightly woven knot, our time-series work has shed a little light into the intricate sequences that dictate how the animal tackles an OrV infection at each stage of the process, making it the highest resolution transcriptome of the process to date. Our results also highlight the importance of studying early responses to infection, which are very different from those of advanced infection stages. This research offers a fresh perspective to appreciate the nuanced strategies these tiny creatures use in their battle against viral intrusion.

## MATERIALS AND METHODS

### *C. elegans* strain maintenance

All *C. elegans* strains used in this study are derived from the N2 Bristol strain unless specified and are listed in table S8. All strains were maintained at 20°C on nematode growth medium (NGM) plates seeded with *Escherichia coli* OP50 bacteria under standard conditions (*52*, *53*). Mutants and transgenic strains were backcrossed three times into the N2 strain used in the laboratory unless already backcrossed into N2.

Downloaded from https://www.science.org on March 12, 2025

*clec-60*(*tm2319*), *ppw-1*(*pk2505*), *sago-2*(*tm894*), and *vsra-1*(*tm1637*) were ordered from the Caenorhabditis Genetics Center, while *C53A5.9*(*tm10539*), *eol-1*(*tm6609*), and *lys-2*(*tm2398*) were ordered from the National Bioresource Project and backcrossed three times to a wild-type N2 background. *Drh-1*(*ok3495*) was a gift from M.A. Félix, and *pals-22*(*jy3*) was from E.R. Troemel. A *gst-35* mutant, *gst-35*(*esv1*), was generated using CRISPR-Cas9 and resulted in a 478–base pair (bp) deletion, which included part of the first exon and a splice junction, resulting in a truncated protein that only maintained homology to the first three amino acids (fig. S11, B to E).

To achieve synchronized animal populations, plates containing eggs were meticulously rinsed with M9 buffer to eliminate larvae and adult animals while retaining the eggs. After a 1-hour interval, plates were subjected to another M9 buffer wash to gather the larvae that had hatched during that period.

### OrV stock preparation, quantification, and inoculation

JU2624 or SFE2 animals coming from one freshly starved plate were inoculated with OrV by soaking in viral stock kindly provided by M.A. Félix (*6*) for 1 hour. The volume was then divided among 30 9-cm NGM plates and allowed to grow for 5 days until starved. Animals were then resuspended in 15 ml of M9 (0.22 M $KH_2PO_4$, 0.42 M $Na_2HPO_4$, 0.85 M NaCl, and 0.001 M $MgSO_4$), let stand for 15 min at room temperature, vortexed, and centrifuged for 2 min at 400*g*. The supernatant was centrifuged twice at 21,000*g* for 5 min and then passed through a 0.2-μm filter.

RNA of the resulting viral stock was extracted using the Viral RNA Isolation Kit (NYZtech). The concentration of viral RNA was then determined by RT-qPCR (details below) and normalized using a standard curve. Primers used can be found in table S9.

For the standard curve, cDNA of OrV was obtained using Accuscript High Fidelity Reverse Transcriptase (Agilent) and reverse primers at the 3′ end of the virus. Approximately 1000 bp of the 3′ end of RNA1 and RNA2 were amplified using forward primers containing 20 bp coding the T7 promoter and DreamTaq DNA Polymerase (Thermo Fisher Scientific). The PCR products were gel purified using MSB Spin PCRapace (Invitek Molecular), and an in vitro transcription was performed using T7 Polymerase (Merck). The remaining DNA was then degraded using deoxyribonuclease I (Life Technologies). RNA concentration was determined by NanoDrop (Thermo Fisher Scientific), and the number of molecules per microliter was determined using the online tool EndMemo RNA Copy Number Calculator (www.endmemo.com/bio/dnacopynum.php). Primers used can be found in table S9.

Synchronized animal populations were inoculated with $2.8 \times 10^9$ copies of OrV by pipetting the viral stock on top of the bacterial lawn containing the animals. Seven hundred animals per plate were grown for RT-qPCRs and smiFISH and 1200 animals per plate for the transcriptomic analysis.

### RNA extractions and RT-qPCRs

Inoculated and control animals were collected at the designated times with phosphate-buffered saline (PBS)–0.05% Tween. Samples were centrifuged for 2 min at 1350 rpm, and the supernatant was discarded. Another two wash steps were performed before freezing the samples in liquid nitrogen. A total of 500 μl of TRIzol (Invitrogen) was added to the animal pellet, and the pellet was disrupted by following five cycles of freeze thawing and five cycles of 30 s of vortex followed by 30 s of rest. A total of 100 μl of chloroform was then added, and the tubes were shaken for 15 s and let rest for 2 min. Samples were centrifuged for 15 min at 11,000*g* at 4°C, and the top layer containing the RNA was then mixed with the same volume of 100% ethanol. The sample was then loaded into RNA Clean & Concentrator columns (Zymo Research), and the rest of the protocol was followed according to the manufacturer's instructions.

RT-qPCRs were performed using Power SYBR Green PCR Master Mix (Applied Biosystems) on an ABI StepOne Plus Real-time PCR System (Applied Biosystems). A total of 10 ng of total RNA was loaded, and a standard curve was used for OrV quantifications. Primers used for RT-qPCRs can be found in table S9.

### Total RNA extraction and preparation for RNA-seq

Library preparation and Illumina sequencing was done by Novogene Europe (www.novogene.com) using a NovaSeq 6000 platform and a Lnc-stranded mRNA-seq library method, ribosomal RNA depletion, directional library preparation, 150 paired end, and 6-Gb raw data per sample. Novogene checked the quality of the libraries using a Qubit 4 Fluorometer (Thermo Fisher Scientific), qPCR for quantification, and Bioanalyzer for size distribution detection.

### RNA-seq data preprocessing

The quality of the Fastq files was assessed with FASTQC (*54*) and MultiQC (*55*) and preprocessed as paired reads with BBDuk (https://sourceforge.net/projects/bbmap/). Briefly, adapters were eliminated, the first 10 5′ nucleotides of each read were cut, and the 3′ end sequences with average quality below 10 were removed. Processed reads shorter than 60 nucleotides were discarded. (The parameter values were set to: *ktrim* = r, *k* = 31, *mink* = 11, *qtrim* = rl, *trimq* = 10, *maq* = 5, and *forcetrimleft* = 10, and *minlength* = 60).

The resultant processed files were use in the next steps. The N2 Bristol (Wbcel235, GCF_000002985.6 assembly) genome was used as reference to the alignment with STAR version 2.7.9a (*56*) using the run mode "alignReads" and the quantification mode "GeneCounts." Quantification results were used for the transcriptome analysis.

### Differential expression analysis and FEA

DESEq2 v1.38.3 (*57*) was used for the DEA using test = "LRT" (complete design = condition + time + condition:time, reduced design = condition + time). The DEGs by time point were extracted with test = "Wald" on each result (time). DEGs were considered significant with Benjamini and Hochberg adjusted $P \leq 0.05$ values. In most heatmaps, asterisks indicate significance level by time point.

The FEAs were performed using the AnimalEnrichr site with the enrichr package v3.2 (*58*). Only terms with adjusted $P \leq 0.05$ were considered.

### Correlation with virus accumulation

A Pearson's correlation between the differential expression of the DEGs ($log_2$FC by time point) and the virus accumulation (percentage of viral reads in sample) was performed with cor.test function from stats package v.4.2.1. Only significant DEGs by LRT were evaluated. Genes with an $r \geq 0.8$ were considered correlated, and those with an $r \leq -0.8$ were considered to be anticorrelated.

### Network functional enrichment and visualization

The network of the genes common to all biotic stresses and the specific genes in the OrV response was created with Cytoscape v3.10.0 (*59*) selecting the option "Full STRING network". The enrichment

Downloaded from https://www.science.org on March 12, 2025

and visualization of the categories were performed with the STRING plugin. Only the singletons with annotations were shown in the final networks.

### smiFISH

The smiFISH protocol was adapted from previous studies (*23*, *60*). Inoculated and control animals were washed with PBS–0.05% Tween and centrifuged for 2 min at 1350 rpm four times, discarding the supernatants. A total of 800 μl of Bouins fixation mix (400 μl of Bouins fix, 400 μl of methanol, 10 μl of β-mercaptoethanol) were added, and the sample was incubated at room temperature for 30 min in a rotatory shaker. Samples were then frozen in liquid nitrogen and kept at −80°C overnight. The following day, samples were gently shaken at 4°C for 30 min and washed four times with borate Triton solution (20 mM $H_3BO_3$, 10 mM NaOH, and 0.5% Triton) and five times with borate Triton β-mercaptoethanol solution (20 mM $H_3BO_3$, 10 mM NaOH, 0.5% Triton, and 2% β-mercaptoethanol), leaving the borate Triton β-mercaptoethanol incubate for 1 hour between each wash in the last three washes. The sample was then incubated for 5 min with wash buffer A (1 ml of Stellaris RNA FISH wash buffer A, 1 ml of deionized formamide, and 3 ml of $H_2O$).

Forty probes against OrV were designed using Oligostan (*60*), 22 against RNA1, and 18 against RNA2. The sequences of the probes can be found in table S9. The probes were annealed as follows: 2 μL of 0.83 μM probe set, 1 μl of 50 μM fluorescent labeled amplification product (FLAP) label, 1 μl of NEB3, and 6 μl of diethyl pyrocarbonate $H_2O$ were incubated at 85°C for 3 min, 65°C for 3 min, and 25°C for 5 min. The FLAP labels contained CAL Fluor 610 or Quasar 670 modifications at both the 5′ and 3′ ends of the following sequence: 5′-AATGCATGTCGACGAGGTCCGAGTGT (Biosearch Technologies).

One microliter of annealed probe solution targeting RNA1 and 1 μl of annealed probe solution targeting RNA2 were then mixed with 98 μl of hybridization buffer (100 μl of Stellaris smiFISH hybridization buffer and 25 μl of deionized formamide). A total of 100 μl of hybridization buffer containing the smiFISH probes were then added to the samples, and the samples were incubated overnight at 37°C in a rotary shaker. The following day, samples were washed with wash buffer A2 (1 ml of Stellaris RNA FISH wash buffer A and 4 ml of $H_2O$), incubated for 30 min in wash buffer A2 at 37°C, and then incubated for 30 min in wash buffer A2 containing 25 ng of 4′,6-diamidino-2-phenylindole (DAPI) at 37°C. Samples were then incubated 5 min in Stellaris RNA FISH wash buffer B, centrifuged, and resuspended in a small volume of Stellaris RNA FISH wash buffer B. A total of 0.1 ng of DAPI was added to each sample, and samples were mounted using *N*-propyl gallate mounting medium. Samples were imaged using Leica Dmi8 microscope with Leica DFC9000 GTC scientific complementary metal-oxide semiconductor camera and objective HC PL APO 40×/0.95 CORR PH2. Images were analyzed and processed using ImageJ (Fiji) (*61*). smiFISH and DAPI signal were used to determine the number of infected cells in each animal, the percentage of animals with RNA2 versus RNA1, and the localization of the signal.

### Percentage of infected animals

To determine the percentage of infected animals, the presence or absence of smiFISH, *pals5p::GFP*, or *lys-3p::GFP* signal was used to determine the infection status of each animal. From the total number of animals analyzed, the percentage of infected animals was calculated.

### Clustering of gene expression profiles

All the significant genes (adjusted $P \leq 0.05$) identified with the LRT were clustered using the *hclust* method from stats package v.4.2.1. A co-occurrence matrix was generated from the results through cutting the tree from 2 to 100 groups (cutree, $k = 2{:}100$). Only the genes that belonged to the same cluster in at least the 50% of the $k$ were treated as connected for the adjacency matrix. To identify the modules, we used the method clusters from the igraph package v1.5.0 (*62*), keeping those formed by a minimum of five genes.

### Comparison with other biotic stresses

The Microbes datasets from WormExp v2.0 (*26*) was used for the comparison with other biotic stresses with the restriction that only experiments performed on wild-type N2 strain were considered. Data from four additional experiments in which N2 strain was used and at least three biological replicates were done were also included: infection of L3 nematodes at 12 hpi with OrV (L3.O.12) and with *Nycteris parisii* (L3.Np.12) (*13*) and infection in L4 nematodes with *M. humicola* at 12 (L4.Mh.12) and 24 hpi (L4.Mh.24) (*27*). Our significant DEGs by time point (adjusted $P \leq 0.0499$) were used in the comparison. A total of 56 experiments were included, and only protein coding genes were considered (table S4). All pathogens were classified into the following categories: virus, Gram-negative and Gram-positive bacteria, and fungi. A gene was considered common when identified as DEGs in at least one experiment of each pathogen category and specific to OrV when it was described only in OrV infection experiments.

### Lipids metabolism, antimicrobial effectors, IPR, and *zip-1*–dependent genes

*C. elegans* genes related to the lipid metabolism (GO0006629) where downloaded from Wormbase database (*63*). IPR genes were extracted from table S5 in (*18*) and ZIP-1–dependent genes from data S3 in (*24*).

### *gst-35* CRISPR knockout

The *gst-35*(*esv1*) deletion was generated by imprecise repair of a CRISPR-Cas9–induced DNA double-strand break and selected using the *dpy-10* co-CRISPR approach (*64*). The fusion was generated using plasmid-based expression of Cas9 and single guide RNAs (sgRNAs). Two pairs of sgRNA plasmids were used to target the 5′ and 3′ ends of the *gst-35* open-reading frame and generated by ligation of annealed oligo pairs into the pU6::sgRNA expression vector pJJR50 (Addgene) as previously described (*65*). The sgRNA targeting the *dpy-10* locus (*64*) was cloned into the pU6::sgRNA expression vector pJJR50 as previously described (*65*). The injection mix was prepared in MilliQ $H_2O$ and contained Peft-3::Cas9 (60 ng/ml; Addgene) and each sgRNA plasmid (45 ng/ml). Young adult hermaphrodites were injected in the germline. For selection of edited genomes, injected animals were transferred to individual culture plates, incubated for 3 to 4 days at 20°C, and 96 nontransgenic F1 animals (wild type, Dpy, or Rol) from two plates containing high numbers of Dpy and Rol animals were selected and transferred to individual NGM plates. After laying eggs, F1 animals were lysed and genotyped by PCR using two primers flanking the *gst-35* open-reading frame. Sanger sequencing was used to determine the precise molecular lesion (Eurofins Genomics). The sequences were targeted, and the primers used can be found in table S9.

Downloaded from https://www.science.org on March 12, 2025

## Supplementary Materials

**The PDF file includes:**
Figs. S1 to S11
Table S8
Legends for tables S1 to S7 and S9
References

**Other Supplementary Material for this manuscript includes the following:**
Tables S1 to S7 and S9

## REFERENCES AND NOTES


1. C. A. Suttle, Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
2. R. Solé, S. F. Elena, *Viruses as Complex Adaptive Systems* (Primers in Complex Systems, Princeton Univ. Press, 2019).
3. A. X. Han, S. P. J. de Jong, C. A. Russell, Co-evolution of immunity and seasonal influence viruses. *Nat. Rev. Microbiol.* **21**, 805–817 (2023).
4. K. B. G. Scholthof, The past is present: Coevolution of viruses and host resistance within geographic centers of plant diversity. *Annu. Rev. Phythopahtol.* **61**, 119–136 (2023).
5. I. Engelmann, N. Pujol, Innate immunity in *C. elegans*, in *Invertebrate Immunity*, *Advances in Experimental Medicine and Biology*, K. Söderhäll, Ed. (Springer, 2010), pp. 105–121.
6. M. A. Félix, A. Ashe, J. Piffaretti, G. Wu, I. Nuez, T. Bélicard, Y. Jiang, G. Zhao, C. J. Franz, L. D. Goldstein, M. Sanroman, E. A. Miska, D. Wang, Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLOS Biol.* **9**, e1000586 (2011).
7. H. Jiang, C. J. Franz, G. Wu, H. Renshaw, G. Zhao, A. E. Firth, D. Wang, Orsay virus utilizes ribosomal frameshifting to express a novel protein that is incorporated into virions. *Virology* **450-451**, 213–221 (2014).
8. W. Yuan, Y. Zhou, Y. Fan, Y. J. Tao, W. Zhong, Orsay δ protein is required for nonlytic viral egress. *J. Virol.* **92**, e00745–18 (2018).
9. Y. Fan, Y. R. Guo, W. Yuan, Y. Zhou, M. V. Holt, T. Wang, B. Demeler, N. L. Young, W. Zhong, Y. J. Tao, Structure of a pentameric virion-associated fiber with a potential role in Orsay virus entry to host cells. *PLOS Pathog.* **13**, e1006231 (2017).
10. Y. R. Guo, Y. Fan, Y. Zhou, M. Jin, J. L. Zhang, H. Jiang, M. V. Holt, T. Wang, N. L. Young, D. Wang, W. Zhong, Y. J. Tao, Orsay virus CP-δ adopts a novel β-bracelet structural fold and incorporates into virions as a head fiber. *J. Virol.* **94**, e01560–20 (2020).
11. C. J. Franz, H. Renshaw, L. Frezal, Y. Jiang, M. A. Félix, D. Wang, Orsay, Santeuil and Le Blanc viruses primarily infect intestinal cells in *Caenorhabditis* nematodes. *Virology* **448**, 255–264 (2014).
12. A. Ashe, T. Bélicard, J. Le Pen, P. Jarkies, L. Frézal, N. J. Lehrbach, M. A. Félix, E. A. Miska, A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *eLife* **2**, e00994 (2013).
13. K. Chen, C. J. Franz, H. Jiang, Y. Jiang, D. Wang, An evolutionarily conserved transcriptional response to viral infection in *Caenorhabditis* nematodes. *BMC Genomics* **18**, 303 (2017).
14. J. N. Sowa, H. Jiang, L. Somasundaram, E. Tecle, G. Xu, D. Wang, E. R. Troemel, The *Caenorhabditis elegans* RIG-I homolog DRH-1 mediates the intracellular pathogen response upon viral infection. *J. Virol.* **94**, e01173–19 (2020).
15. J. Le Pen, H. Jiang, T. Di Domenico, E. Kneuss, J. Kosałka, C. Leung, M. Morgan, C. Much, K. L. M. Rudolph, A. J. Enright, D. O'Carroll, D. Wang, E. A. Miska, Terminal uridylyltransferases target RNA viruses as part of the innate immune system. *Nat. Struct. Mol. Biol.* **25**, 778–786 (2018).
16. M. A. Bakowski, C. A. Desjardins, M. G. Smelkinson, T. A. Dunbar, I. F. Lopez-Moyado, S. A. Rifkin, C. A. Cuomo, E. R. Troemel, Ubiquitin-mediated response to microsporidia and virus infection in *C. elegans*. *PLOS Pathog.* **10**, e1004200 (2014).
17. K. C. Reddy, T. Dror, J. N. Sowa, J. Panek, K. Chen, E. S. Lim, D. Wang, E. R. Troemel, An intracellular pathogen response pathway promotes proteostasis in *C. elegans*. *Curr. Biol.* **27**, 3544–3553.e5 (2017).
18. K. C. Reddy, T. Dror, R. S. Underwood, G. A. Osman, C. R. Elder, C. A. Desjardins, C. A. Cuomo, M. Barkoulas, E. R. Troemel, Antagonistic paralogs control a switch between growth and pathogen resistance in *C. elegans*. *PLOS Pathog.* **15**, e1007528 (2019).
19. L. A. Casorla-Perez, R. Guennoun, C. Cubillas, B. Peng, K. Kornfeld, D. Wang, Orsay virus infection of *Caenorhabditis elegans* is modulated by zinc and dependent on lipids. *J. Virol.* **96**, e01211–e01222 (2022).
20. H. Jiang, K. Chen, L. E. Sandoval, C. Leung, D. Wang, An evolutionarily conserved pathway essential for Orsay virus infection of *Caenorhabditis elegans*. *mBio* **8**, e00940–17 (2017).
21. Y. Zhou, W. Zhong, Y. J. Tao, Collagen and actin network mediate antiviral immunity against Orsay in *C. elegans* intestinal cells. *PLOS Pathog.* **20**, e1011366 (2024).
22. M. G. Sterken, L. B. Snoek, K. J. Bosman, J. Daamen, J. A. G. Riksen, J. Bakker, G. P. Pijlman, J. E. Kammenga, A heritable antiviral RNAi response limits Orsay virus infection in *Caenorhabditis elegans* N2. *PLOS ONE* **9**, e89760 (2014).
23. D. M. Parker, L. P. Winkenbach, A. Parker, S. Boyson, E. O. Nishimura, Improved methods for single-molecule fluorescence *in situ* hybridization and immunofluorescence in *Caenorhabditis elegans* embryos. *Curr. Protoc.* **1**, e299 (2021).
24. V. Lažetić, F. Wu, L. B. Cohen, K. C. Reddy, Y. T. Chang, S. S. Gang, G. Bhabha, E. R. Troemel, The transcription factor ZIP-1 promotes resistance to intracellular infection in *Caenorhabditis elegans*. *Nat. Commun.* **13**, 17 (2022).
25. E. J. E. Kim, H. G. Son, H. E. H. Park, Y. Jung, S. Kwon, S. J. V. Lee, *Caenorhabditis elegans algn-2* is critical for longevity conferred by enhanced nonsense-mediated mRNA decay. *iScience* **23**, 101713 (2020).
26. W. Yang, K. Dierking, H. Schulenburg, AnimalExp: A web-based application for a *Caenorhabditis elegans*-specific gene expression enrichment analysis. *Bioinformatics* **32**, 943–945 (2016).
27. G. A. Osman, M. K. Fasseas, S. L. Koneru, C. L. Essmann, K. Kyrou, M. A. Srinivasan, G. Zhang, P. Sarkies, M. A. Félix, M. Barkoulas, Natural infection of *C. elegans* by an oomycete reveals a new pathogen-specific immune response. *Curr. Biol.* **28**, 640–648.e5 (2018).
28. R. González, M. A. Félix, *Caenorhabditis elegans* immune responses to microsporidia and viruses. *Dev. Comp. Immunol.* **154**, 105148 (2024).
29. K. Dierking, W. Yang, H. Schulenburg, Antimicrobial effectors in the nematode *Caenorhabditis elegans*: An outgroup to the Arthropoda. *Philos. Trans. R. Soc. B* **371**, 20150299 (2016).
30. B. Pees, W. Yang, A. Zárate-Potes, H. Schulenburg, K. Dierking, High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates. *J. Innate Immun.* **8**, 129–142 (2015).
31. H. Yoon, J. L. Shaw, M. C. Haigis, A. Greka, Lipid metabolism in sickness and in health: Emerging regulators of lipotoxicity. *Mol. Cell* **81**, 3708–3730 (2021).
32. B. Song, X. Sheng, J. L. Justice, K. K. Lum, P. J. Metzger, K. C. Cook, J. C. Kostas, I. M. Cristea, Intercellular communication within the virus microenvironment affects the susceptibility of cells to secondary viral infections. *Sci. Adv.* **9**, eadg3433 (2023).
33. Y. Guo, Z. Xun, S. R. Coffman, F. Chen, The Shift of the intestinal microbiome in the innate immunity-deficient mutant *rde-1* strain of *C. elegans* upon Orsay virus infection. *Front. Microbiol.* **8**, 933 (2017).
34. I. Melero, R. González, S. F. Elena, Host developmental stages shape the evolution of a plant RNA virus. *Philos. Trans. R. Soc. B* **378**, 20220005 (2023).
35. J. E. Irazoqui, E. R. Troemel, R. L. Feinbaum, L. G. Luhachack, B. O. Cezairliyan, F. M. Ausubel, Distinct pathogenesis and host responses during infection of *C. elegans* by *P. aeruginosa* and *S. aureus*. *PLOS Pathog.* **6**, e1000982 (2010).
36. D. O'Rourke, D. Baban, M. Demidova, R. Mott, J. Hodgkin, Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. *Genome Res.* **16**, 1005–1016 (2006).
37. T. B. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, J. Middel, I. L. Cornelissen, H. S. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor, Y. van Kooyk, DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587–597 (2000).
38. Z. S. Yang, C. Y. Lin, M. B. Khan, M. C. Hsu, W. Assavalapsakul, A. Thitithanyanont, S. F. Wang, Understanding the role of galectins toward influenza A virus infection. *Expert Opin. Ther. Targets* **27**, 927–937 (2023).
39. C. Cubillas, L. E. Sandoval Del Prado, S. Goldacker, C. Fujii, A. N. Pinski, J. Zielke, D. Wang, The *alg-1* gene is necessary for Orsay virus replication in *Caenorhabditis elegans*. *J. Virol.* **97**, e0006523 (2023).
40. M. Pollari, S. De, A. Wang, K. Mäkinen, The potyviral silencing suppressor HCPro recruits and employs host ARGONAUTE1 in pro-viral functions. *PLOS Pathog.* **16**, e1008965 (2020).
41. U. Seroussi, A. Lugowski, L. Wadi, R. X. Lao, A. R. Willis, W. Zhao, A. E. Sundby, A. G. Charlesworth, A. W. Reinke, J. M. Claycomb, A comprehensive survey of *C. elegans* argonaute proteins reveals organism-wide gene regulatory networks and functions. *eLife* **12**, e83853 (2023).
42. J. Li, Y. Wang, H. Deng, S. Li, H. J. Qiu, Cellular metabolism hijacked by viruses for immunoevasion: Potential antiviral targets. *Front. Immunol.* **14**, 1228811 (2023).
43. S. M. Anderson, R. Pukkila-Worley, Immunometabolism in *Caenorhabditis elegans*. *PLOS Pathog.* **16**, e1008897 (2020).
44. Y. Qu, W. Wang, M. Z. X. Xiao, Y. Zheng, Q. Liang, The interplay between lipid droplets and virus infection. *J. Med. Virol.* **95**, e28967 (2023).
45. N. S. Heaton, G. Randall, Multifaceted roles for lipids in viral infection. *Trends Microbiol.* **19**, 368–375 (2011).
46. E. Ketter, G. Randall, Virus impact on lipids and membranes. *Annu. Rev. Virol.* **6**, 319–340 (2019).
47. L. Herrera-Moro Huitron, L. A. De Jesús-González, M. Martínez-Castillo, J. M. Ulloa-Aguilar, C. Cabello-Gutierrez, C. Helguera-Repetto, J. Garcia-Cordero, M. León Juárez, Multifaceted nature of lipid droplets in viral interactions and pathogenesis. *Microorganisms* **11**, 1851 (2023).
48. S. M. Anderson, H. K. Cheesman, N. D. Peterson, J. E. Salisbury, A. A. Soukas, R. Pukkila-Worley, The fatty acid oleate is required for innate immune activation and pathogen defense in *Caenorhabditis elegans*. *PLOS Pathog.* **15**, e1007893 (2019).


Downloaded from https://www.science.org on March 12, 2025

49. P. Deng, N. U. Naresh, Y. Du, L. T. Lamech, J. Yu, L. J. Zhu, R. Pukkila-Worley, C. M. Haynes, Mitochondrial UPR repression during *Pseudomonas aeruginosa* infection requires the bZIP protein ZIP-3. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 6146–6151 (2019).
50. M. Kumar, S. Sharma, S. Mazumder, Role of UPR$^{mt}$ and mitochondrial dynamics in host immunity: it takes two to tango. *Front. Cell. Infect. Microbiol.* **13**, 1135203 (2023).
51. C. N. Martineau, N. V. Kirienko, N. Pujol, Innate immunity in *C. elegans*. *Curr. Top. Develop. Biol.* **144**, 309–351 (2021).
52. S. Brenner, The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
53. T. Stiernagle, Maintenance of C. elegans (WormBook: The Online Review of C. Elegans Biology, 2006); http://www.wormbook.org/.
54. S. Andrews, A quality control tool for high throughput sequence data (2010); http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
55. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
56. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
58. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
59. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
60. N. Tsanov, A. Samacoits, R. Chouaib, A. M. Traboulsi, T. Gostan, C. Weber, C. Zimmer, K. Zibara, T. Walter, M. Peter, E. Bertrand, F. Mueller, smiFISH and FISH-quant—A flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).
61. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
62. G. Csardi, T. Nepusz, The igraph software package for complex network research. *Complex Syst.* **1695**, 1–9 (2006).
63. P. Davis, M. Zarowiecki, V. Arnaboldi, A. Becerra, S. Cain, J. Chan, W. J. Chen, J. Cho, E. da Veiga Beltrame, S. Diamantakis, S. Gao, D. Grigoriadis, C. A. Grove, T. W. Harris, R. Kishore, T. Le, R. Y. N. Lee, M. Luypaert, H. M. Müller, C. Nakamura, P. Nuin, M. Paulini, M. Quinton-Tulloch, D. Raciti, F. H. Rodgers, M. Russell, G. Schindelman, A. Singh, T. Stickland, K. van Auken, Q. Wang, G. Williams, A. J. Wright, K. Yook, M. Berriman, K. L. Howe, T. Schedl, L. Stein, P. W. Sternberg, WormBase in 2022-data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics* **220**, iyac003 (2022).
64. J. A. Arribere, R. T. Bell, B. X. H. Fu, K. L. Artiles, P. S. Hartman, A. Z. Fire, Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**, 837–846 (2014).
65. S. Waaijers, J. Muñoz, C. Berends, J. J. Ramalho, S. S. Goerdayal, T. Y. Low, A. D. Zoumaro-Djayoon, M. Hoffmann, T. Koorman, R. P. Tas, M. Harterink, S. Seelk, J. Kerver, C. C. Hoogenraad, O. Bossinger, B. Tursun, S. van den Heuvel, A. J. R. Heck, M. Boxem, A tissue-specific protein purification approach in *Caenorhabditis elegans* identifies novel interaction partners of DLG-1/Discs large. *BMC Biol.* **14**, 66 (2016).
66. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).


**Acknowledgments:** We thank F.de la Iglesia and P. Agudo for excellent technical support and the members of the EvolSysVir laboratory for valuable comments and fruitful discussions. We thank N. Flames for giving us access to a microinjector and V. Portegijs for assistance with the smiFISH protocol. We thank M.A. Félix for strains JU2624 and RB2519 as well as for OrV JUv1580 and E.R. Troemel for strains ERT54 and ERT288. We also thank Wormbase (*63*). Some strains were provided by the Caenorhabditis Genetics Center, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). Computations were performed on the HPC cluster Garnatxa at the Institute for Integrative Systems Biology (I2SysBio). **Funding:** This work was supported by grants PID2022-136912NB-I00 funded by MCIU/AEI/10.13039/501100011033 and by "ERDF a way of making Europe", as well as CIPROM/2022/59 funded by Generalitat Valenciana to S.F.E. V.G.C. was supported by grant FJC2021-047264-I funded by MCIU/AEI/10.13039/501100011033 and by NextGenerationEU/PRTR. M.J.O.-U. was supported by grant FPU2019/05246 funded by MCIU/AEI/10.13039/501100011033 and by "ESF investing in your future". A.V.-G. was supported by the European Space Agency contract 4000135960/21/NL/GLC/my to S.F.E., J.C.M.-S. was supported by grant ACIF/2021/296 from Generalitat Valenciana. E.G.L. was supported by grant FPU21/00410 funded by MCIN/AEI/10.13039/501100011033 and by "ESF invest in your future".
**Author contributions:** V.G.C.: Writing—original draft, conceptualization, investigation, writing—review and editing, methodology, resources, funding acquisition, validation, supervision, project administration, and visualization. M.J.O.-U.: Writing—original draft, conceptualization, investigation, writing—review and editing, methodology, data curation, validation, formal analysis, software, project administration, and visualization. A.V.-G.: Investigation and writing—review and editing. J.C.M.-S.: Conceptualization, writing—review and editing, methodology, and visualization. E.G.L.: Validation. S.F.E.: Writing—original draft, conceptualization, writing—review and editing, methodology, funding acquisition, data curation, validation, supervision, formal analysis, and project administration. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** RNA-seq data were deposited in the NCBI SRA: PRJNA1037048. All the code generated for this study are accessible in Zenodo at 10.5281/zenodo.12620790. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 18 December 2023
Accepted 23 August 2024
Published 27 September 2024
10.1126/sciadv.adn5945

Downloaded from https://www.science.org on March 12, 2025

—

In Appendix A, I show **extra results** to this Chapter. The results are focused on the dynamics of the viral accumulation, the viral diversity (including the identification and characterization of DVGs) and provide new insights into the the OrV transcription and replication processes.

—

# CHAPTER 7

# General Discussion

Octavia Butler wrote in her *Parable of the Sower*: "The only lasting truth is change" [19]. This idea applies to all aspects of life, not just to a (unfortunately not so) dystopian novel. Without change, the appearance of life would not have been possible. Therefore, at least as living organism as us, *the only lasting truth is evolution*.

Charles Darwin proposed in 1859 that all organisms arise and develop through natural selection of small heritable changes that confer survival and reproductive advantage to individuals [32]. His theory provided us with a framework for studying evolution. Viruses, especially RNA viruses, due to their high mutation rates and large population sizes, have become an ideal system for studying evolutionary trajectories. However, in our anthropocentric world, the study of viruses likely would not have received such abundant resources if some of them were not direct threats to humans or human interests (such as our farm animals and plants).

The findings presented in this thesis integrate research on the evolution of RNA viruses and the transcriptomic responses elicited in their hosts, focusing on the dynamics or changes in viral diversity and in host (immune) responses. These works emphasize the role of diversity, DVGs, replication dynamics, epigenetic factors, and the broader immune responses in shaping virus-host interactions at different scale times. Each Chapter has its own discussion section that I will not repeat here. Nevertheless, below I present a more general discussion.

—

Since viruses depend on their hosts for survival, there is a general intuition that they would evolve to minimize harm to the host while still preserving their ability to infect new, susceptible individuals. This pattern does appear to hold true in some systems. Furthermore, there are growing evidences indicating that some viruses enhance their host's tolerance to abiotic stresses such as drought, extreme temperatures or excess of light irradiation [64].

The prediction of the evolution of virulence has been a hot topic in evolutionary ecology, epidemiology and parasitology (including all types of parasites), however, the exact factors that drive the evolution of a virus to act as a parasite, a commensalist or a mutualist are still not entirely understood. In 1982, Anderson and May pointed that the endpoint of the coevolution of a parasite would depend on the interplay between virulence and transmissibility. The mutualism-parasitism continuum hypothesis proposed by Ewald in 1987 stated that the main responsible of the outcome would be the mode of transmission [50]. According to this hypothesis, viruses transmitted vertically (from parent to offspring) are more likely to evolve toward a *mutualistic* relationship, where virus and host benefit from one another, as the virus depends on the host's survival for its own propagation, and as more offspring produce the host, the better for the virus. On the other hand, viruses that are transmitted horizontally (between individuals not necessarily related) are more likely to evolve toward a parasitic relationship, where the virus may exploit the host to maximize its reproduction and so asseverate their transmission at the cost of damaging the host. This hypothesis is useful in some contexts but relies on one critical assumption; the existence of a trade-off between virulence and the pathogen fitness [3].

If such a trade-off exists, it is expected that viruses evolve towards a balance between maximizing their ability to spread (transmission) and minimizing the damage they cause to the host (virulence). This trade-off has been experimentally identified in natural populations: myxomatosis virus infecting wild rabbit populations in Australia [6], HIV-1 infections (Amsterdam seroconverters Cohort and Rakai Community Cohort Study from Uganda, respectively) [56, 12] and in laboratory conditions: phage $\phi 1$ [113] and cauliflower mosaic virus (CaMV)-*Brassica rapa* using the aphid *Myzus persicae* to quantify CaMV transmission [44]. But this trade-off seems not to be universally applicable.

($i$) Some *experimental designs can remove the virulence-transmission trade-off.* Our studies on TuMV evolution would be an example of this case. In nature, TuMV is transmitted by aphids but in our experimental conditions [5] the vector transmission is removed from the picture and substituted by the mechanical transmission of a

large viral inoculum. We were able to determine that virulence (using the level of disease symptoms as a proxy) increased in all the plant genotypes, concluding that the removal of the virulence-transmission constrain allowed to the virus to evolve without the need to maintain an active host as a source of inoculum. But not only that constrain was remove. Through simplifying the pathosystem (by removing the vector) the genetic space the virus has been able to explore was also expanded. This broader landscape could enabled the emergence of genetic solutions—such those driving to an increased virulence and viral accumulation—that may not have been accessible under the selective pressures imposed by vector-mediated transmission. In fact, several studies have shown that small mutations or deletions, can disrupt vector transmission [166, 169], highlighting the narrow evolutionary space viruses must navigate when relying on a vector. In this context, vector-mediated transmission functions as a significant evolutionary bottleneck, limiting the range of viable viral strategies [54].

(*ii*) As pointed above, in natural conditions, *other virulence or non-virulence-related trade-offs can constrain the evolutionary trajectory of the virus.* The recovery rate ($\nu$) was included in Anderson and May model of the basic reproductive number ($R_0$, as a proxy of the viral fitness) [6]. In their definition, $R_0 = \frac{\beta N}{\alpha b \nu}$, they included transmission rate ($\beta$) and host density ($N$) in the numerator, and virulence ($\alpha$), background mortality ($b$), and recovery rate ($\nu$) in the denominator, with $\beta, \alpha$ and $\nu$ treated as independent factors. With this definition in hand, the trade-off emerges naturally. In [1], authors made the assumption that recovery and transmission rates are both function of the virulence ($\beta(\alpha)$, $\nu(\alpha)$). If $\nu(\alpha)$, hosts may recover more quickly from less virulent strains than from highly virulent ones, making the recovery rate a negative function of virulence. These authors also summarize some empirical evidences supporting their assumption [1]. The original continuum hypothesis focused on the transmission–virulence trade-off, but left the recovery or clearance rate out of consideration [50]. If the recovery rate is also linked to virulence, the hypothesis predicts that transmission rate will increase with virulence up to a point, but will eventually slow down as the cost of higher virulence outweighs its benefits. Even more interesting, the virulence-clearance trade-off highlights the importance of the host responses to the infection on the outcome. The clearance rate depends on the ability of the host to remove the virus, and this ability can change through the life of the host by several reasons. Developmental stages [112] or how prepared is the immune system to recognize and fight against the viral infection are some of them. Clearance-transmission trade-offs on DENV highlights the dependence of the virulence evolution on the epidemiological context. Primary DENV infections can result in severe disease but the majority of

them are asymptomatic or only cause dengue fever. By contrast, secondary infections have a higher risk of developing into a severe hemorrhagic disease. The antibodies raised during the primary infection facilitate viral entry into host target cell, and this enhances the viral replication by 10-fold or more with respect the primary replication [4, 63]. In addition, an over-response of the immune system is what causes the damage to the host tissue [11]. Post-secondary infections have a better prognosis, not so good as primary but with less hospitalization cases than secondary infections [11]. We have seen in [5] that more challenging genetic systems selected for more virulent viruses while more permissive host genotypes (for example, lacking components of the RNAi) select for less virulent ones. While the hosts in our system do not coevolve with the virus, the Red Queen Hypothesis still provides a useful lens: the virus is engaged in a one-sided evolutionary chase, constantly adapting to overcome the host responses. In this context, more efficient immune systems act as stronger selective environments, pushing the virus toward more aggressive strategies to maintain its fitness. In these experiments, the dependency of the viral evolution outcome to the completeness of the immune system confirmed once more the importance of host immune responses as part of the trade-offs the virus face. Simplifying experimental systems is often essential—either to make them experimentally accessible or to reveal dynamics that are hiden in more complex settings. Still, it is important to recognize that such simplifications remove key natural constraints, and that the outcomes observed under these oversimplified conditions may differ from those in the natural biological context.

We have studied the intra-host viral dynamics of OrV during the infection of *C. elegans*. Although much remains to be understood about this pathosystem, it is not too far-fetched to consider it as a potential example of commensalistic symbiosis. Back in 1904, Smith wrote:

> *The less complete the adaptation, the more virulent the disease produced. The final outcome is a harmless disease of little or no fatality, unless other parasites complicate the invasion* [152].

Despite relying solely on its innate immunity, *C. elegans* mounts an efficient antiviral response, achieving near-complete clearance of OrV in approximately 40 hours post infection. Interestingly, recent studies suggest that OrV can also establish persistent infections [25]. Results in the short-term diversity dynamics presented in this Thesis highlighted that OrV produces a plethora of small frequency variants, including DVGs, while the quasispecie remains dominated by the ancestral genome. These findings are consistent with the apparent molecular stasis observed in its experimental evolution

[23], and support the idea of OrV as a well-adapted virus to its host, possibly explaining the limited genetic space it occupies.

With the work presented in this Thesis [24], we have also contributed to disentangling the host responses specific to OrV from those triggered by more virulent parasites. If OrV is indeed a well-adapted symbiont, then comparing the host's response to it with responses to harmful pathogens can reveal how immune systems discriminate between different types of infections. Moreover, identifying conditions that lead to increased viral virulence or enhanced host susceptibility would make this pathosystem an even more powerful model for studying the balance between tolerance, resistance, and pathogenicity.

Finally, DVGs have traditionally been studied by their interference with wild-type viruses, and there are several mathematical models and empirical data supporting their role as "cheaters" [9, 85, 119, 93, 177, 97, 62]. However, some of our findings suggest that increased accumulation of DVGs does not necessarily correlate with reduced viral accumulation or transmission. In the experimental evolution of HCoV-OC43 in BHK-21 cells, both viral titers and DVG loads increased, indicating that in specific contexts, DVGs could improve the overall viral performance [71]. Beyond their competitive or cooperative interactions, DVGs may serve as reservoirs of genetic diversity [174], preserving sequences that would otherwise be eliminated by purifying selection, and potentially facilitating saltatory evolutionary changes. DVGs, by accumulating mutations—often in the absence of strong selective pressures—could harbor genetic variants that, under changing hosts or environmental conditions, enable the virus to rapidly adapt to new niches. As proposed by [101], recombination between full-length genomes and DVGs may allow viruses to escape catastrophic mutational loads and restore fitness under positive selection of the certain recombinants [101]. Through this same mechanism, recombination events could generate viral genomes that are genetically distant from their ancestors—effectively representing evolutionary jumps. While this last possibility is intriguing, it remains speculative and warrants further investigation. Future studies employing advanced modeling techniques and empirical analyses are essential to elucidate the role of DVGs in facilitating saltatory evolution and their broader impact on viral evolution and host-pathogen dynamics.

# CHAPTER 8

# Conclusions

Each Chapter in this Thesis has allowed us to approach the knowledge of host-virus interactions from a different perspective (in the time scale and in the protagonist). Despite the fact that we have followed a biologic systems approach, the human mind (and mine concretely) its limited and so is its ability to process and integrate all the information. This means that, for sure, we may not be fully grasping the entire complexity of these interactions.

Even so, I present here some of the conclusions I have reached.

Conclusions on the long-term virus-host interactions studied:

1. The epigenetic regulation profoundly affects *A. thaliana* antiviral responses and TuMV evolution.

   (a) TuMV evolving in mutants in DNA methylation and histone modification pathways increases in viral accumulation and virulence, with more permissive host genotypes selecting for less virulent viruses and more resistant ones selecting for more virulent viruses.

   (b) Viruses evolving in the epigenetic mutants show faster rates of evolution of the phenotypic traits than viruses evolving in the wild-type plant.

   (c) TuMV VPg is a major target of selection in viral adaptation, independent of the plant's genotype. The flexible loop, spanning residues 79 to 120, accumulates changes under positive selection.

(d) Viral adaptation induced changes in the host responses. We identified a core set of 1029 genes whose expression was perturbed by viral adaptation in all epigenetic mutants and in the wild-type plants. The overexpressed genes were associated with SA and SAR, ethylene-signaling pathways, and the response to chitin. Conversely, the genes repressed by viral adaptation were involved in JA, amine, and aromatic amino acid pathways, as well as the repression of phenylpropanoid metabolism. These responses reflect an exacerbated defense reaction at the cost of plant growth, ultimately leading to weaker plants.

(e) DNA modification (*ddm1* and *polv* mutats) but not histone methylation (*jmj14*) affects the expression of genes nearby TEs. TEs near gene promoters are more likely to correlate with changes in gene expression than those within genes.

(f) The integration of host transcriptomics and the virus-host PPIN allowed us to pinpoint 13 Arabidopsis genes critical for virus adaptation. Between them, the downregulation of *pal1* – that encodes for the first enzyme of the phenylpropanoid synthesis pathway and a physical interactor of the TuMV viroporin 6K1– by the evolved viruses could be responsible for the weaker plant phenotype.

2. The accumulation and diversity of DVGs during the evolution of HCoV-OC43 in BHK-21 and HCT-8 cells, and MHV in CCL-9.1 cells depend on both cellular and viral factors.

(a) DVGs accumulate and evolve in both HCoV-OC43 and MHV, with high MOI promoting greater variability and persistence of certain DVGs. This suggests a selective process where fitter DVGs are favored, potentially influencing viral evolution.

(b) Deletions, particularly in the *E* and *ns12.9* cistrons, were the most common and abundant DVGs. These deletions may modulate pathogenesis, with some of them appearing independently across lineages and cell types, suggesting an adaptive advantage linked to host cell survival trade-offs.

(c) The evolution of DVGs was influenced by both viral and host cell factors, with HCoV-OC43 showing more defined patterns of deletion compared to MHV. This highlights the importance of considering host factors and using diverse cell lines in future studies.

(d) Persistent deletions may be positively selected to modulate viral replication and pathogenesis, opening the possibility of developing DVGs as antiviral agents, either by inhibiting wild-type virus replication or as broad-spectrum antivirals.

(e) High MOI conditions led to more defined patterns of DVG accumulation, particularly in deletion hotspots within cistrons encoding structural and accessory proteins. In contrast, low MOI showed more variability in DVG size distribution, underscoring the impact of MOI on DVG evolution dynamics.

Conclusions on the accumulation and diversity dynamics of OrV during the course of a primary infection in *C. elegans* and the transcriptomic responses of the animals:

1. OrV accumulation follows an stamping machine replication mode.

2. OrV generates all types of DVGs, with deletions and one-nucleotide insertions being the most pervasive ones. Specific hotspots of deletions, 5'cb, and 3'cb suggests that the generation is not totally random but mediated by still unknown processes.

3. A persistent deletion in the RNA2 is compatible with the transcription of a subgenomic RNA that would potentially produce a shorter version of the free $\delta$ protein. This *in silico* result would require experimental validation.

4. The identification of a putative peptide encoded in the antisense RNA1. This *in silico* result would require experimental validation.

5. Micro- and macroevolutionary evidence of selective pressure in the RdRP carboxy terminal tail highlights this region as a potential key player in the OrV life cycle.

6. We have mapped the transcriptomic responses of the animals to a primary OrV infection with the highest temporal resolution up to date. A classification of the gene expression by their dynamics allowed us to distinguish a set of early responding genes, a group of genes highly correlated with viral accumulation, and a group of late responding genes that have importance in the recovery stage of the animal. Comparison of the responses against OrV with those against several virulent pathogens allowed us to identify a set of OrV-specific response genes.

7. Some of the highlighted genes have been experimentally validated, allowing the identification of new *C. elegans* pro and antiviral factors.

# References


[1] Acevedo, M. A., Dillemuth, F. P., Flick, A. J., Faldyn, M. J., and Elderd, B. D. (2019). Virulence-driven trade-offs in disease transmission: A meta-analysis. *Evolution*, 73(4):636–647.

[2] Alazem, M. and Lin, N. S. (2015). Roles of plant hormones in the regulation of host-virus interactions. *Molecular Plant Pathology*, 16(5):529–540.

[3] Alizon, S. and Michalakis, Y. (2015). Adaptive virulence evolution: The good old fitness-based approach. *Trends in Ecology and Evolution*, 30(5):248–254.

[4] Althouse, B. M. and Hanley, K. A. (2015). The tortoise or the hare? Impacts of within-host dynamics on transmission success of arthropod-borne viruses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1675).

[5] Ambrós, S., Olmo-Uceda, M. J., Corrêa, R. L., and Elena, S. F. (2024). Phenotypic and genomic changes during Turnip mosaic virus adaptation to *Arabidopsis thaliana* mutants lacking epigenetic regulatory factors. *Evolution*, 78(1):69–85.

[6] Anderson, R. M. and May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, 85 (Pt 2):411–426.

[7] Andronov, L., Han, M., Zhu, Y., Balaji, A., Roy, A., Barentine, A., Patel, P., Garhyan, J., Qi, L., and Moerner, W. (2024). Nanoscale cellular organization of viral RNA and proteins in SARS-CoV-2 replication organelles. *Nature Communications*, 15(1):1–17.

[8] Aratani, Y., Uemura, T., Hagihara, T., Matsui, K., and Toyota, M. (2023). Green leaf volatile sensory calcium transduction in Arabidopsis. *Nature Communications*, 14(1):1–16.

[9] Bangham, C. M. and Kirkwood, T. B. (1990). Defective interfering particles: Effects in modulating virus growth and persistence. *Virology*, 179(2):821–826.

[10] Beauclair, G., Mura, M., Combredet, C., Tangy, F., Jouvenet, N., and Komarova, A. (2018). Di-tector: Defective interfering viral genomes' detector for next-generation sequencing data. *Rna*, 24:1285–1296.

[11] Ben-Shachar, R. and Koelle, K. (2018). Transmission-clearance trade-offs indicate that dengue virulence evolution depends on epidemiological context. *Nature Communications*, 9(1).

[12] Blanquart, F., Grabowski, M. K., Herbeck, J., Nalugoda, F., Serwadda, D., Eller, M. A., Robb, M. L., Gray, R., Kigozi, G., Laeyendecker, O., Lythgoe, K. A., Nakigozi, G., Quinn, T. C., Reynolds, S. J., Wawer, M. J., and Fraser, C. (2016). A transmission-virulence evolutionary trade-off explains attenuation of hiv-1 in uganda. *eLIFE*.

[13] Bohan, D. and Maury, W. (2021). Enveloped rna virus utilization of phosphatidylserine receptors: Advantages of exploiting a conserved, widely available mechanism of entry. *PLoS Pathogens*, 17(9):10–12.

[14] Borges, F. and Martienssen, R. A. (2015). The expanding world of small RNAs in plants.

[15] Bosma, T., Karagiannis, K., Santana-Quintero, L., Ilyushina, N., Zagorodnyaya, T., Petrovskaya, S., Laassri, M., Donnelly, R., Rubin, S., Simonyan, V., and Sauder, C. (2019). Identification and quantification of defective virus genomes in high throughput sequencing data using dvg-profiler, a novel post-sequence alignment processing algorithm. *PLoS ONE*, 14:1–34.

[16] Bosque, G., Folch-Fortuny, A., Picó, J., Ferrer, A., and Elena, S. F. (2014). Topology analysis and visualization of Potyvirus protein-protein interaction network. *BMC systems biology*, 8:129.

[17] Buchmann, K. (2014). Evolution of innate immunity: Clues from invertebrates via fish to mammals. *Frontiers in Immunology*, 5(SEP):1–8.

[18] Butkovic, A., Ellis, T. J., Gonzalez, R., Jaegle, B., Nordborg, M., and Elena, S. F. (2024). Genetic basis of textitArabidopsis thaliana responses to infection by naïve and adapted isolates of turnip mosaic virus. *eLife*, 13:1–29.

[19] Butler, O. E. (1993). *Parable of the sower*. A Four Walls Eight Windows first edition. New York : Four Walls Eight Windows, [1993] ©1993.

[20] Carr, J. P., Tungadi, T., Donnelly, R., Bravo-Cazar, A., Rhee, S. J., Watt, L. G., Mutuku, J. M., Wamonje, F. O., Murphy, A. M., Arinaitwe, W., Pate, A. E., Cunniffe, N. J., and Gilligan, C. A. (2020). Modelling and manipulation of aphid-mediated spread of non-persistently transmitted viruses. *Virus Research*, 277.

[21] Casorla-Perez, L. A., Guennoun, R., Cubillas, C., Peng, B., Kornfeld, K., and Wang, D. (2022). Orsay virus infection of *Caenorhabditis elegans* is modulated by zinc and dependent on lipids. *Journal of Virology*, 96(22).

[22] Casteel, C. L., Yang, C., Nanduri, A. C., De Jong, H. N., Whitham, S. A., and Jander, G. (2014). The NIa-Pro protein of Turnip mosaic virus improves growth and reproduction of the aphid vector, *Myzus persicae* (green peach aphid). *Plant Journal*, 77(4):653–663.

[23] Castiglioni, V. G., Olmo-Uceda, M. J., Martín, S., Félix, M. A., González, R., and Elena, S. F. (2024a). Experimental evolution of an RNA virus in *Caenorhabditis elegans*. *Infection, Genetics and Evolution*, 123.

[24] Castiglioni, V. G., Olmo-Uceda, M. J., Villena-Giménez, A., Muñoz-Sánchez, J. C., Legarda, E. G., and Elena, S. F. (2024b). Story of an infection: Viral dynamics and host responses in the *Caenorhabditis elegans*-Orsay virus pathosystem. *Sci. Adv*, 10:5945.

[25] Castiglioni, V. G., Villena-Giménez, A., González-Sánchez, A., Toft, C., Gómez, G. G., and Elena, S. F. (2024c). Chronic infection of *Caenorhabditis elegans* by Orsay virus induces age-dependent immunity and superinfection exclusion. *Preprint*, pages 1–41.

[26] Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348(6300):454–455.

[27] Chao, L. and Elena, S. F. (2017). Nonlinear trade-offs allow the cooperation game to evolve from prisoner's dilemma to snowdrift. *Proceedings of the Royal Society B: Biological Sciences*, 284(1854).

[28] Cheng, X., Xiong, R., Li, Y., Li, F., Zhou, X., and Wang, A. (2017). Sumoylation of turnip mosaic virus RNA polymerase promotes viral infection by counteracting the host NPR1-mediated immune response. *Plant Cell*, 29(3):508–525.

[29] Cheon, Y. J., Hwang, H., and Kim, K. (2020). Plasticity of pheromone-mediated avoidance behavior in C. elegans. *Journal of Neurogenetics*, 34(3-4):420–426.

[30] Chkuaseli, T. and White, K. A. (2020). Activation of viral transcription by stepwise largescale folding of an RNA virus genome. *Nucleic Acids Research*, 48(16):9285–9300.

[31] Cubillas, C., Sandoval Del Prado, L. E., Goldacker, S., Fujii, C., Pinski, A. N., Zielke, J., and Wang, D. (2023). The *alg-1* gene is necessary for Orsay Virus replication in *Caenorhabditis elegans* . *Journal of Virology*, 97(4):1–14.

[32] Darwin, C. (1859). *On the origin of species by means of natural selection, or Preservation of favoured races in the struggle for life*. London: John Murray, New York, NY.

[33] DeDiego, M., Álvarez, E., Almazán, F., Rejas, M., Lamirande, E., Roberts, A., Shieh, W.-J., Zaki, S., Subbarao, K., and Enjuanes, L. (2007). A severe acute respiratory syndrome coronavirus that lacks the e gene is attenuated in vitro and in vivo. *Journal of Virology*, 81:1701–1713.

[34] den Boon, J. A., Nishikiori, M., Zhan, H., and Ahlquist, P. (2024). Positive-strand RNA virus genome replication organelles: structure, assembly, control. *Trends in Genetics*, 40(8):681–693.

[35] Díaz-Muñoz, S. L. (2019). Uncovering virus-virus interactions by unifying approaches and harnessing high-throughput tools. *mSystems*, 4(3):1–5.

[36] Díaz-Muñoz, S. L., Sanjuán, R., and West, S. (2017). Sociovirology: conflict, cooperation, and communication among viruses. *Cell Host and Microbe*, 22(4):437–441.

[37] Dolgin, E. (2019). The secret social lives of viruses. *Nature*, 570(7761):290–292.

[38] Domingo, E. (2019). *Virus as populations : composition, complexity, quasispecies, dynamics, and biological implications*. Academic Press, second edition.

[39] Domingo, E. (2020). Molecular basis of genetic variation of viruses. In *Virus as Populations*, pages 35–71. Elsevier.

[40] Domingo, E., Biebricher, C. K., Eigen, M., and Holland, J. J. (2001). *Quasispecies and RNA virus evolution: principles and consequences*. Molecular biology intelligence unit ; 14. Landes Bioscience.

[41] Domingo, E., Menéndez-Arias, L., and Holland, J. J. (1997). RNA virus fitness. *Reviews in Medical Virology*, 7(2):87–96.

[42] Domingo, E., Schuster, P., Elena, S. F., and Perales, C., editors (2023). *Viral Fitness and Evolution*, volume 439. Springer International Publishing.

[43] dos Santos, C. and Franco, O. L. (2023). Pathogenesis-Related Proteins (PRs) with enzyme activity activating plant defense responses. *Plants*, 12(11):1–13.

[44] Doumayrou, J., Avellan, A., Froissart, R., and Michalakis, Y. (2013). An experimental test of the transmission-virulence trade-off hypothesis in a plant virus. *Evolution*, 67(2):477–486.

[45] Duarte, E., Clarke, D., Moya, A., Domingo, E., and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13):6015–6019.

[46] Dufresne, P. J., Ubalijoro, E., Fortin, M. G., and Laliberte, J. F. (2008). Arabidopsis thaliana class II poly(A)-binding proteins are required for efficient multiplication of turnip mosaic virus. *Journal of General Virology*, 89(9):2339–2348.

[47] Dunoyer, P., Thomas, C., Harrison, S., Revers, F., and Maule, A. (2004). A Cysteine-Rich plant protein potentiates Potyvirus movement through an interaction with the Virus Genome-linked protein VPg . *Journal of Virology*, 78(5):2301–2309.

[48] Efstathiou, S., Ottens, F., Schütter, L. S., Ravanelli, S., Charmpilas, N., Gutschmidt, A., Le Pen, J., Gehring, N. H., Miska, E. A., Bouças, J., and Hoppe, T. (2022). ER-associated RNA silencing promotes ER quality control. *Nature Cell Biology*, 24(12):1714–1725.

[49] Elena, S. F. and Rodrigo, G. (2012). Towards an integrated molecular model of plant-virus interactions. *Current Opinion in Virology*, 2(6):719–724.

[50] Ewald, P. W. (1987). Transmission modes and evolution of the parasitism-mutualism continuum. *Annals of the New York Academy of Sciences*, 503(1):295–306.

[51] Félix, M. A. and Wang, D. (2019). Natural viruses of *Caenorhabditis* nematodes. *Annual Review of Genetics*, 53:313–326.

[52] Felsenstein, J. and Yokoyama, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics*, 83(4):845–859.

[53] Filipowicz, A. and Allard, P. (2025). *Caenorhabditis elegans* as a model for environmental epigenetics. *Current environmental health reports*, 12(1):6.

[54] Forrester, N. L., Guerbois, M., Seymour, R. L., Spratt, H., and Weaver, S. C. (2012). Vector-borne transmission imposes a severe bottleneck on an RNA virus population. *PLoS Pathogens*, 8(9).

[55] Frank, J. A. and Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Current Opinion in Virology*, 25:81–89.

[56] Fraser, C., Hollingsworth, T. D., Chapman, R., De Wolf, F., and Hanage, W. P. (2007). Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 104(44):17441–17446.

[57] Froissart, R., Doumayrou, J., Vuillaume, F., Alizon, S., and Michalakis, Y. (2010). The virulence-transmission trade-off in vector-borne plant viruses: A review of (non-)existing studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1548):1907–1918.

[58] Gainey, D., Shubin, A., and Hunter, C. (2024). Reported transgenerational responses to *Pseudomonas aeruginosa* in *C . elegans* are not robust. *eLife*, pages 1–35.

[59] Gao, M., Hao, Z., Ning, Y., and He, Z. (2024). Revisiting growth–defence trade-offs and breeding strategies in crops. *Plant Biotechnology Journal*, 22(5):1198–1205.

[60] García-Ordóñez, L. and Pagán, I. (2024). Vertical and horizontal transmission of plant viruses: two extremes of a continuum? *npj Viruses*, 2(1):1–10.

[61] Gerrish, P. J. and Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, 102-103(1-6):127–144.

[62] Girgis, S., Xu, Z., Oikonomopoulos, S., Fedorova, A. D., Tchesnokov, E. P., Gordon, C. J., Schmeing, T. M., Götte, M., Sonenberg, N., Baranov, P. V., Ragoussis, J., Hobman, T. C., and Pelletier, J. (2022). Evolution of naturally arising SARS-CoV-2 defective interfering particles. *Communications Biology*, 5(1).

[63] Goncalvez, A. P., Engle, R. E., St. Claire, M., Purcell, R. H., and Lai, C. J. (2007). Monoclonal antibody-mediated enhancement of dengue virus infection in vitro and in vivo and strategies for prevention. *Proceedings of the National Academy of Sciences of the United States of America*, 104(22):9422–9427.

[64] González, R., Butković, A., and Elena, S. F. (2020). From foes to friends: Viral infections expand the limits of host phenotypic plasticity. *Advances in Virus Research*, 106:85–121.

[65] González, R., Butkovic, A., Escaray, F. J., Martínez-Latorre, J., Melero, Í., Pérez-Parets, E., Gómez-Cadenas, A., Carrasco, P., and Elena, S. F. (2021). Plant virus evolution under strong drought conditions results in a transition from parasitism to mutualism. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6).

[66] González, R. and Félix, M. A. (2024a). *Caenorhabditis elegans* immune responses to microsporidia and viruses. *Developmental and Comparative Immunology*, 154(February 2023).

[67] González, R. and Félix, M. A. (2024b). Naturally-associated bacteria modulate Orsay virus infection of Caenorhabditis elegans. *PLoS Pathogens*, 20(1).

[68] Gravato-Nobre, M. J. and Hodgkin, J. (2005). *Caenorhabditis elegans* as a model for innate immunity to pathogens.

[69] Hajishengallis, G., Netea, M. G., and Chavakis, T. (2023). Innate immune memory, trained immunity and nomenclature clarification. *Nature Immunology*, 24(9):1393–1394.

[70] Hannan Parker, A., Wilkinson, S. W., and Ton, J. (2022). Epigenetics: a catalyst of plant immunity against pathogens.

[71] Hillung, J., Olmo-Uceda, M. J., Muñoz-Sánchez, J. C., and Elena, S. F. (2024). Accumulation dynamics of defective genomes during experimental evolution of two betacoronaviruses. *Viruses*, 16(4).

[72] Holmes, E. C. (2009). The evolution and emergence of rna viruses. *Oxford University Press*.

[73] Huang, A. S. and Baltimore, D. (1970). Defective viral particles and viral disease processes. *Nature Publishing Group*, 226:325–327.

[74] Huang, T. S., Wei, T., Laliberté, J. F., and Wang, A. (2010). A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. *Plant Physiology*, 152(1):255–266.

[75] Hunter, M. and Fusco, D. (2022). Superinfection exclusion: A viral strategy with short-term benefits and long-term drawbacks. *PLoS Computational Biology*, 18(5):1–18.

[76] Jaramillo-Mesa, H. and Rakotondrafara, A. M. (2023). All eggs in one basket: How potyvirus infection is controlled at a single cap-independent translation event. *Seminars in Cell and Developmental Biology*, 148-149(January):51–61.

[77] Ji, M., Zhao, J., Han, K., Cui, W., Wu, X., Chen, B., Lu, Y., Peng, J., Zheng, H., Rao, S., Wu, G., Chen, J., and Yan, F. (2021). Turnip mosaic virus P1 suppresses JA biosynthesis by degrading cpSRP54 that delivers AOCs onto the thylakoid membrane to facilitate viral infection. *PLoS Pathogens*, 17(12):1–24.

[78] Jiang, H., Chen, K., Sandoval, L. E., Leung, C., and Wang, D. (2017). An evolutionarily conserved pathway essential for orsay virus infection of Caenorhabditis elegans. *mBio*, 8(5).

[79] Jiang, H., Sandoval Del Prado, L. E., Leung, C., and Wang, D. (2020). Huntingtin-interacting protein family members have a conserved pro-viral function from *Caenorhabditis elegans* to humans. *Proceedings of the National Academy of Sciences of the United States of America*, 117(36):22462–22472.

[80] Jiang, J., Patarroyo, C., Garcia Cabanillas, D., Zheng, H., and Laliberté, J.-F. (2015). The vesicle-forming 6K 2 protein of turnip mosaic virus interacts with the COPII coatomer Sec24a for viral systemic infection . *Journal of Virology*, 89(13):6695–6710.

[81] Jones, J. D. and Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117):323–329.

[82] Kachroo, A. and Robin, G. P. (2013). Systemic signaling during plant defense. *Current Opinion in Plant Biology*, 16(4):527–533.

[83] Kim, D. H. and Ewbank, J. J. (2018). Signaling in the innate immune response. *WormBook : the online review of C. elegans biology*, 2018:1–35.

[84] King, C. R. and Mehle, A. (2022). Retasking of canonical antiviral factors into proviral effectors. *Current Opinion in Virology*, 56.

[85] Kirkwood, T. B. and Bangham, C. R. (1994). Cycles, chaos, and evolution in virus cultures: A model of defective interfering particles. *Proceedings of the National Academy of Sciences of the United States of America*, 91(18):8685–8689.

[86] Kolbert, Z., Barroso, J. B., Boscari, A., Corpas, F. J., Gupta, K. J., Hancock, J. T., Lindermayr, C., Palma, J. M., Petřivalský, M., Wendehenne, D., and Loake, G. J. (2024). Interorgan, intraorgan and interplant communication mediated by nitric oxide and related species. *New Phytologist*, pages 786–797.

[87] Kover, P. X. and Clay, K. (1998). Trade-off between virulence and vertical transmission and the maintenance of a virulent plant pathogen. *American Naturalist*, 152(2):165–175.

[88] Ku, D. D. (1982). A pheromone influences larval development in the nematod *Caenorhabditis elegans*. *Science*, 218(5):578–581.

[89] Kumar, R. and Dasgupta, I. (2024). Abscisic acid: An emerging player in plant-virus interactions. *Plant Physiology and Biochemistry*, 215(April):109046.

[90] Kuo, C. J., Hansen, M., and Troemel, E. (2018). Autophagy and innate immunity: Insights from invertebrate model organisms.

[91] Lai, M. M. (1992). RNA recombination in animal and plant viruses. *Microbiological Reviews*, 56(1):61–79.

[92] Lanz-Mendoza, H., Gálvez, D., and Contreras-Gardunõ, J. (2024). The plasticity of immune memory in invertebrates. *Journal of Experimental Biology*, 227:1–11.

[93] Lázaro, J. T., Albó, A., Alarcón, T., Elena, S. F., and Sardanyés, J. (2024). No two without three: Modeling dynamics of the trio RNA virus-defective interfering genomes-satellite RNAs. *Communications in Nonlinear Science and Numerical Simulation*, 133.

[94] Lažetić, V., Wu, F., Cohen, L. B., Reddy, K. C., Chang, Y. T., Gang, S. S., Bhabha, G., and Troemel, E. R. (2022). The transcription factor ZIP-1 promotes resistance to intracellular infection in Caenorhabditis elegans. *Nature Communications*, 13(1).

[95] Le Pen, J., Jiang, H., Di Domenico, T., Kneuss, E., Kosałka, J., Leung, C., Morgan, M., Much, C., Rudolph, K. L., Enright, A. J., O'Carroll, D., Wang, D., and Miska, E. A. (2018). Terminal uridylyltransferases target RNA viruses as part of the innate immune system. *Nature Structural and Molecular Biology*, 25(9):778–786.

[96] Ledvina, H. E. and Whiteley, A. T. (2024). Conservation and similarity of bacterial and eukaryotic innate immunity.

[97] Leeks, A. (2021). *Social evolution in viruses*. PhD thesis, University of Oxford.

[98] Leeks, A., Bono, L. M., Ampolini, E. A., Souza, L. S., Höfler, T., Mattson, C. L., Dye, A. E., and Díaz-Muñoz, S. L. (2023). Open questions in the social lives of viruses. *Journal of Evolutionary Biology*, 36(11):1551–1567.

[99] Lei, M., Tan, Y., Tu, H., and Tan, W. (2024). Neuronal basis and diverse mechanisms of pathogen avoidance in Caenorhabditis elegans.

[100] Li, Y., Xiong, R., Bernards, M., and Wang, A. (2016). Recruitment of Arabidopsis RNA helicase AtRH9 to the viral replication complex by viral replicase to promote Turnip mosaic virus replication. *Scientific Reports*, 6(May):1–12.

[101] Lin, C. H., Tam, H. M. H., Yang, C. Y., Hsieh, F. C., Wang, J. L., Yang, C. C., Hsu, H. W., Liu, H. P., and Wu, H. Y. (2023). Evolution of the coronavirus spike protein in the full-length genome and defective viral genome under diverse selection pressures. *Journal of General Virology*, 104(11).

[102] Liu, L. and Chen, X. (2018). Intercellular and systemic trafficking of RNAs in plants. *Nature Plants*, 4(11):869–878.

[103] Liu, Y. and Sun, J. (2021). Detection of pathogens and regulation of immunity by the *Caenorhabditis elegans* nervous system. *mBio*, 12(2).

[104] Liu, Y., Zhang, X., Liu, J., Xia, H., Zou, J., Muruato, A., Periasamy, S., Kurhade, C., Plante, J., Bopp, N., Kalveram, B., Bukreyev, A., Ren, P., Wang, T., Menachery, V., Plante, K., Xie, X., Weaver, S., and Shi, P. (2022). A live-attenuated sars-cov-2 vaccine candidate with accessory protein deletions. *Nature Communications*, 13.

[105] Llopis-Almela, O., Lázaro, J. T., Elena, S. F., and Sardanyés, J. (2025). Global bifurcation in a virus, defective genomes, satellite rnas tripartite system: Breakdown of a coexistence quasi-neutral curve. *Chaos, Solitons and Fractals*, 192.

[106] Lopez-Gomollon, S. and Baulcombe, D. C. (2022). Roles of RNA silencing in viral and non-viral plant immunity and in the crosstalk between disease resistance systems. *Nature Reviews Molecular Cell Biology*, 23(10):645–662.

[107] Lázaro, J. T., Albó, A., Alarcón, T., Elena, S. F., and Sardanyés, J. (2024). No two without three: Modeling dynamics of the trio rna virus-defective interfering genomes-satellite rnas. *Communications in Nonlinear Science and Numerical Simulation*, 133.

[108] Mahmoudabadi, G. and Phillips, R. (2018). A comprehensive and quantitative exploration of thousands of viral genomes. *eLife*, 7:1–26.

[109] Martineau, C. N., Kirienko, N. V., and Pujol, N. (2021). Innate immunity in *C. elegans*. In *Current Topics in Developmental Biology*, volume 144, pages 309–351. Academic Press Inc.

[110] Martínez, F., Carrasco, J. L., Toft, C., Hillung, J., Giménez-Santamarina, S., Yenush, L., Rodrigo, G., and Elena, S. F. (2023). A binary interaction map between turnip mosaic virus and Arabidopsis thaliana proteomes. *Communications Biology*, 6(1):1–18.

[111] Matkowski, H. and Daszkowska–Golec, A. (2025). Wisdom comes after facts – An update on plants priming using phytohormones. *Journal of Plant Physiology*, 305(December 2024):154414.

[112] Melero, I., González, R., and Elena, S. F. (2023). Host developmental stages shape the evolution of a plant RNA virus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1873).

[113] Messenger, S. L., Molineux, I. J., and Bull, J. J. (1999). Virulence evolution in a virus obeys a trade-off. *Proceedings of the Royal Society B: Biological Sciences*, 266(1417):397–404.

[114] Miralles, R., Gerrish, P. J., Moya, A., and Elena, S. F. (1999). Clonal interference and the evolution of RNA viruses. *Science*, 285(5434):1745–1747.

[115] Moore, R. S., Kaletsky, R., and Murphy, C. T. (2019). Piwi/PRG-1 argonaute and TGF-$\beta$ mediate transgenerational learned pathogenic avoidance. *Cell*, 177(7):1827–1841.e12.

[116] Movahed, N., Sun, J., Vali, H., Laliberté, J. F., and Zheng, H. (2019). A host ER fusogen is recruited by turnip mosaic virus for maturation of viral replication vesicles. *Plant Physiology*, 179(2):507–518.

[117] Moya, A., Holmes, E. C., and González-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nature reviews. Microbiology*, 2(4):279–288.

[118] Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9.

[119] Muñoz-Sánchez, J., Lázaro, J., Hillung, J., Olmo-Uceda, M. J., Sardanyés, J., and Elena, S. F. (2025). Quasineutral multistability in an epidemiological-like model for defective-helper betacoronavirus infection in cell cultures. *Applied Mathematical Modelling*, 137:115673.

[120] Muñoz-Sánchez, J. C., Olmo-Uceda, M. J., Oteo, J. A., and Elena, S. (2024). Quantifying defective and wild-type viruses from high-throughput rna sequencing. *Bioinformatics*, 40.

[121] Netea, M. G., Schlitzer, A., Placek, K., Joosten, L. A., and Schultze, J. L. (2019). Innate and adaptive immune memory: an evolutionary continuum in the host's response to pathogens. *Cell Host and Microbe*, 25(1):13–26.

[122] Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563.

[123] Olmo-Uceda, M. J., Muñoz-Sánchez, J. C., Lasso-Giraldo, W., Arnau, V., Díaz-Villanueva, W., and Elena, S. F. (2022). Dvgfinder: A metasearch tool for identifying fefective viral genomes in rna-seq data. *Viruses*, 14:1114.

[124] Owusu, I., Quaye, O., Passalacqua, K., and Wobus, C. (2021). Egress of non-enveloped enteric RNA viruses. *Journal of General Virology*, 102(3).

[125] Pagán, I., Montes, N., Milgroom, M. G., and García-Arenal, F. (2014). Vertical Transmission Selects for Reduced Virulence in a Plant Virus and for Increased Resistance in the Host. *PLoS Pathogens*, 10(7):23–25.

[126] Panek, J., Gang, S. S., Reddy, K. C., Luallen, R. J., Fulzele, A., Bennett, E. J., and Troemel, E. R. (2020). A cullin-RING ubiquitin ligase promotes thermotolerance as part of the intracellular pathogen response in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 117(14):7950–7960.

[127] Payne, S. (2023). 9 - virus evolution and genetics. In Payne, S., editor, *Viruses (Second Edition)*, pages 107–118. Academic Press, second edition edition.

[128] Peck, K. M. and Lauring, A. S. (2018). Complexities of viral mutation rates. *Journal of Virology*, 92.

[129] Pradel, E., Zhang, Y., Pujol, N., Matsuyama, T., Bargmann, C. I., and Ewbank, J. J. (2007). Detection and avoidance of a natural product from the pathogenic bacterium *Serratia marcescens* by *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(7):2295–2300.

[130] Presloid, J. B., Ebendick-Corpus, B. E., Zárate, S., and Novella, I. S. (2008). Antagonistic pleiotropy involving promoter sequences in a virus. *Journal of Molecular Biology*, 382(2):342–352.

[131] Ramirez-Prado, J. S., Abulfaraj, A. A., Rayapuram, N., Benhamed, M., and Hirt, H. (2018). Plant immunity: from signaling to epigenetic control of defense.

[132] Ramos-Cruz, D., Troyee, A. N., and Becker, C. (2021). Epigenetics in plant organismic interactions.

[133] Rechavi, O. and Lev, I. (2017). Principles of transgenerational small RNA inheritance in *Caenorhabditis elegans*. *Current Biology*, 27(14):R720–R730.

[134] Reddy, K. C., Dror, T., Sowa, J. N., Panek, J., Chen, K., Lim, E. S., Wang, D., and Troemel, E. R. (2017). An intracellular pathogen response pathway promotes proteostasis in *C. elegans*. *Current Biology*, 27(22):3544–3553.e5.

[135] Reddy, K. C., Dror, T., Underwood, R. S., Osman, G. A., Elder, C. R., Desjardins, C. A., Cuomo, C. A., Barkoulas, M., and Troemel, E. R. (2019). Antagonistic paralogs control a switch between growth and pathogen resistance in C. elegans. *PLoS Pathogens*, 15(1).

[136] Rezelj, V. V., Carrau, L., Merwaiss, F., Levi, L. I., Erazo, D., Tran, Q. D., Henrion-Lacritick, A., Gausson, V., Suzuki, Y., Shengjuler, D., Meyer, B., Vallet, T., Weger-Lucarelli, J., Bernhauerová, V., Titievsky, A., Sharov, V., Pietropaoli, S., Diaz-Salinas, M. A., Legros, V., Pardigon, N., Barba-Spaeth, G., Brodsky, L., Saleh, M. C., and Vignuzzi, M. (2021). Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. *Nature Communications*, 12:1–14.

[137] Routh, A. and Johnson, J. (2014). Discovery of functional genomic motifs in viruses with virema-a virus recombination mapper-for analysis of next-generation sequencing data. *Nucleic Acids Research*, 42:1–10.

[138] Ryu, L., Cheon, Y., Huh, Y. H., Pyo, S., Chinta, S., Choi, H., Butcher, R. A., and Kim, K. (2018). Feeding state regulates pheromone-mediated avoidance behavior via the insulin signaling pathway in *Caenorhabditis elegans* . *The EMBO Journal*, 37(15):1–14.

[139] Sandoval, L. E., Jiang, H., and Wang, D. (2019). The dietary restriction-like gene *drl-1*, which encodes a putative serine/threonine kinase, is essential for Orsay virus infection in *Caenorhabditis elegans*. *Journal of virology*, 93(3).

[140] Schaack, G. A. and Mehle, A. (2020). Experimental approaches to identify host factors important for influenza virus. *Cold Spring Harbor Perspectives in Medicine*, 10(12):1–19.

[141] Schneider, M., Johnson, J. R., Krogan, N. J., and Chanda, S. K. (2016). The virus-host interactome: Knowing the players to understand the game. In *Viral Pathogenesis: From Basics to Systems Biology*, number January, chapter 12, pages 157–167. Academic Press, 3 edition.

[142] Schulenburg, H., Kurz, C. L., and Ewbank, J. J. (2004). Evolution of the innate immune system: The worm perspective.

[143] Segredo-otero, E. and Sanjuán, R. (2022). Cooperative virus-virus interactions : An evolutionary perspective. *BioDesign Research*, 2022.

[144] Sethna, P. B., Hung, S. L., and Brian, D. A. (1989). Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proceedings of the National Academy of Sciences of the United States of America*, 86(14):5626–5630.

[145] Shah, J. and Zeier, J. (2013). Long-distance communication and signal amplification in systemic acquired resistance. *Frontiers in Plant Science*, 4(FEB):1–16.

[146] Shah, P. S., Beesabathuni, N. S., Fishburn, A. T., Kenaston, M. W., Minami, S. A., Pham, O. H., and Tucker, I. (2022). Systems biology of virus-host protein interactions: From hypothesis generation to mechanisms of replication and pathogenesis. *Annual Review of Virology*, 9(1):397–415.

[147] Shapiro, J. W. and Turner, P. E. (2018). Evolution of mutualism from parasitism in experimental virus populations. *Evolution*, 72(3):707–712.

[148] Sharov, V., Rezelj, V. V., Galatenko, V. V., Titievsky, A., Panov, J., Chumakov, K., Andino, R., Vignuzzi, M., and Brodsky, L. (2021). Intra- and inter-cellular modeling of dynamic interaction between zika virus and its naturally occurring defective viral genomes. *Journal of Virology*, 95.

[149] Shearer, H. L., Cheng, Y. T., Wang, L., Liu, J., Boyle, P., Després, C., Zhang, Y., Li, X., and Fobert, P. R. (2012). Arabidopsis clade I TGA Transcription factors regulate plant defenses in an NPR1-independent fashion. *Molecular Plant-Microbe Interactions*, 25(11):1459–1468.

[150] Shomar, H., Georjon, H., Feng, Y., Olympio, B., Guillaume, M., Tesson, F., Cury, J., Wu, F., and Bernheim, A. (2024). Viperin immunity evolved across the tree of life through serial innovations on a conserved scaffold. *Nature Ecology and Evolution*, 8(9):1667–1679.

[151] Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. *Journal of General Virology*, 96(6):1193–1206.

[152] Smith, T. (1904). Some problems in the life history of pathogenic microorganisms. *Science*, 20(520):817–832.

[153] Söderhäll, K. (2010). *Invertebrate immunity* . Advances in experimental medicine and biology ; v. 708. Springer Science+Business Media, New York, N.Y.

[154] Song, Y. Y., Zeng, R. S., Xu, J. F., Li, J., Shen, X., and Yihdego, W. G. (2010). Interplant communication of tomato plants through underground common mycorrhizal networks. *PLoS ONE*, 5(10).

[155] Sowa, J. N., Jiang, H., Somasundaram, L., Tecle, E., Xu, G., Wang, D., and Troemel, E. R. (2020). The *Caenorhabditis elegans* RIG-I homolog DRH-1 mediates the Intracellular Pathogen Response upon viral infection. *Journal of Virology*, 94(2).

[156] Sparacio, A. P., Trojanowski, N. F., Snetselaar, K., Nelson, M. D., and Raizen, D. M. (2020). Teething during sleep: Ultrastructural analysis of pharyngeal muscle and cuticular grinder during the molt in Caenorhabditis elegans. *PLoS ONE*, 15(5):1–24.

[157] Sterken, M. G., Snoek, L. B., Bosman, K. J., Daamen, J., Riksen, J. A., Bakker, J., Pijlman, G. P., and Kammenga, J. E. (2014). A heritable antiviral RNAi response limits orsay virus infection in Caenorhabditis elegans N2. *PLoS ONE*, 9(2).

[158] Sun, Y., Kim, E., Felt, S., Taylor, L., Agarwal, D., Grant, G., and Lopez, C. (2019). A specific sequence in the genome of respiratory syncytial virus regulates the generation of copy-back defective viral genomes. *PLoS Pathogens*, 15.

[159] Tang, Z. and Wang, A. (2023). *Western Molecular identification and characterization of Heat shock protein 70 family proteins essential for Turnip mosaic virus infection in Arabidopsis thaliana*. Thesis, Western University.

[160] Tanguy, M., Véron, L., Stempor, P., Ahringer, J., Sarkies, P., and Miska, E. A. (2017). An alternative STAT signaling pathway acts in viral immunity in *Caenorhabditis elegans*. *mBio*, 8(5).

[161] Tecle, E., Chhan, C. B., Franklin, L., Underwood, R. S., Hanna-Rose, W., and Troemel, E. R. (2021). The purine nucleoside phosphorylase pnp-1 regulates epithelial cell resistance to infection in *C. elegans*. *PLoS pathogens*, 17(4):e1009350.

[162] Tenthorey, J. L., Emerman, M., and Malik, H. S. (2022). Evolutionary landscapes of host-virus arms races. *Annual Review of Immunology*.

[163] The UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*, 51(November 2022):523–531.

[164] Thivierge, K., Cotton, S., Dufresne, P. J., Mathieu, I., Beauchemin, C., Ide, C., Fortin, M. G., and Laliberté, J. F. (2008). Eukaryotic elongation factor 1A interacts with Turnip mosaic virus RNA-dependent RNA polymerase and VPg-Pro in virus-induced vesicles. *Virology*, 377(1):216–225.

[165] Tran, T. D. and Luallen, R. J. (2023). An organismal understanding of *C. elegans* innate immune responses, from pathogen recognition to multigenerational resistance.

[166] Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E., and Higgs, S. (2007). A single mutation in Chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathogens*, 3(12):1895–1906.

[167] Turner, P. E. and Chao, L. (1999). Prisoner's dilemma in an RNA virus. *Nature*, 398(6726):441–443.

[168] Turner, P. E. and Chao, L. (2003). Escape from prisoner's dilemma in RNA phage $\phi$6. *American Naturalist*, 161(3):497–505.

[169] Van Gennip, R. G., Drolet, B. S., Rozo Lopez, P., Roost, A. J., Boonstra, J., and Van Rijn, P. A. (2019). Vector competence is strongly affected by a small deletion or point mutations in bluetongue virus. *Parasites and Vectors*, 12(1):1–16.

[170] van Sluijs, L., Liu, J., Schrama, M., van Hamond, S., Vromans, S. P., Scholten, M. H., Žibrat, N., Riksen, J. A., Pijlman, G. P., Sterken, M. G., and Kammenga, J. E. (2021). Virus infection modulates male sexual behaviour in Caenorhabditis elegans. *Molecular Ecology*, 30(24):6776–6790.

[171] Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, 1(July):1–30.

[172] van Wersch, S., Tian, L., Hoy, R., and Li, X. (2020). Plant NLRs: The whistleblowers of plant immunity. *Plant Communications*, 1(1):100016.

[173] Vignuzzi, M. and López, C. (2019). Defective viral genomes are key drivers of the virus–host interaction. *Nature Microbiology*, 4:1075–1087.

[174] Vignuzzi, M., López, C. B., Defilippi, P., Franco, D., and García-Sastre, A. (2019). Virus populations: Quantification, complexity dynamics, and predictability. *Annual Review of Virology*, 6:123–151.

[175] Vijayapalani, P., Maeshima, M., Nagasaki-Takekuchi, N., and Miller, W. A. (2012). Interaction of the trans-frame potyvirus protein P3N-PIPO with host protein PCaP1 facilitates potyvirus movement. *PLoS Pathogens*, 8(4).

[176] Vivier, E. and Malissen, B. (2005). Innate and adaptive immunity: Specificities and signaling hierarchies revisited. *Nature Immunology*, 6(1):17–21.

[177] Von Magnus, P. and Gard, S. (1947). Studies on interferenee in experimental influenza. ii. purification and centrifugation experiments. *Arkiv for Kemi, Mineralogi och Geologi*, page 4 pp.

[178] Wang, J., Long, X., Chern, M., and Chen, X. (2021). Understanding the molecular mechanisms of trade-offs between plant growth and immunity. *Science China Life Sciences*, 64(2):234–241.

[179] Wang, Y., Sun, X., Feng, L., Zhang, K., and Yang, W. (2024). Nervous system guides behavioral immunity in *Caenorhabditis elegans*.

[180] Wei, T., Zhang, C., Hou, X., Sanfaçon, H., and Wang, A. (2013). The SNARE protein Syp71 is essential for turnip mosaic virus infection by mediating fusion of virus-induced vesicles with chloroplasts. *PLoS Pathogens*, 9(5).

[181] Whitfield, Z. J., Dolan, P. T., Kunitomi, M., Tassetto, M., Seetin, M. G., Oh, S., Heiner, C., Paxinos, E., and Andino, R. (2017). The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Current Biology*, 27(22):3511–3519.e7.

[182] Whitney, K. D. and Garland, T. (2010). Did genetic drift drive increases in genome complexity? *PLoS Genetics*, 6(8).

[183] Wibisono, P. and Sun, J. (2021). Neuro-immune communication in *C. elegans* defense against pathogen infection.

[184] Willis, A. R., Sukhdeo, R., and Reinke, A. W. (2021). Remembering your enemies: mechanisms of within-generation and multigenerational immune priming in Caenorhabditis elegans.

[185] Wu, G., Fang, X., Yu, T., Chen, J., and Yan, F. (2024). Turnip mosaic virus pathogenesis and host resistance mechanisms in Brassica. *Horticultural Plant Journal*, 10(4):947–960.

[186] Wu, G., Jia, Z., Ding, K., Zheng, H., Lu, Y., Lin, L., Peng, J., Rao, S., Wang, A., Chen, J., and Yan, F. (2022). Turnip mosaic virus co-opts the vacuolar sorting receptor VSR4 to promote viral genome replication in plants by targeting viral replication vesicles to the endosome. *PLoS Pathogens*, 18(1):1–29.

[187] Wu, H. Y. and Brian, D. A. (2010). Subgenomic messenger RNA amplification in coronaviruses. *Proceedings of the National Academy of Sciences of the United States of America*, 107(27):12257–12262.

[188] Wu, J. H. and Rappaport, I. (1962). Competition between infectious tobacco mosaic virus nucleic acid and intact virus on Nicotiana glutinosa. *Nature*, 193:908.

[189] Wu, X., Liu, H., and Guo, Y. R. (2025). Insights into virus - host interactions : Lessons from *Caenorhabditis elegans* - Orsay virus model. *Current Medical Science*.

[190] Xiao, Y., Lidsky, P. V., Shirogane, Y., Aviner, R., Wu, C. T., Li, W., Zheng, W., Talbot, D., Catching, A., Doitsh, G., Su, W., Gekko, C. E., Nayak, A., Ernst, J. D., Brodsky, L., Brodsky, E., Rousseau, E., Capponi, S., Bianco, S., Nakamura, R., Jackson, P. K., Frydman, J., and Andino, R. (2021). A defective viral genome strategy elicits broad protective immunity against respiratory viruses. *Cell*, 184:6037–6051.e14.

[191] Xiong, R. and Wang, A. (2013). SCE1, the SUMO-conjugating enzyme in plants that interacts with NIb, the RNA-dependent RNA polymerase of turnip mosaic virus, is required for viral infection. *Journal of Virology*, 87(8):4704–4715.

[192] Yang, B., Wang, J., Zheng, X., and Wang, X. (2023). Nematode pheromones: structures and functions. *Molecules*, 28(5).

[193] Yeh, J. X., Fan, Y., Bartlett, M. L., Zhang, X., Sadowski, N., Hauer, D. A., Timp, W., and Griffin, D. E. (2022). Treatment of sindbis virus-infected neurons with antibody to e2 alters synthesis of complete and nsp1-expressing defective viral rnas. *mBio*, 13.

[194] Yuan, M., Ngou, B. P. M., Ding, P., and Xin, X. F. (2021). PTI-ETI crosstalk: an integrative view of plant immunity. *Current Opinion in Plant Biology*, 62:102030.

[195] Zafirov, D., Giovinazzo, N., Lecampion, C., Field, B., Ducassou, J. N., Couté, Y., Browning, K. S., Robaglia, C., and Gallois, J. L. (2023). Arabidopsis eIF4E1 protects the translational machinery during TuMV infection and restricts virus accumulation. *PLoS Pathogens*, 19(11):1–25.

[196] Zhang, Y., Lu, H., and Bargmann, C. I. (2005). Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans*. *Nature*, 438(7065):179–184.

[197] Zhiqi, H., Tingyi, W., Dongdong, C., Lan, S., Guangheng, Z., Qian, Q., and Li, Z. (2025). Leucine-Rich Repeat Protein family regulates stress tolerance and development in plants. *Rice Science*, 32(1):32–43.

[198] Zhou, Y., Chen, H., Zhong, W., and Tao, Y. J. (2024). Collagen and actin network mediate antiviral immunity against Orsay virus in C. elegans intestinal cells. *PLoS Pathogens*, 20(1):1–19.

# Resum Ampliat

**Capítol 1: Introducció**

Els virus són simbionts intracel · lulars obligats que depenen de la maquinària cel · lular de l'hoste per a la seua replicació. Malgrat la seua simplicitat, són genèticament diversos i evolutivament exitosos. Els virus d'ARN, en particular, posseïxen genomes compactes, altes taxes de mutació i cicles de replicació ràpids. Estes característiques els permeten adaptar-se ràpidament, evadir la immunitat i creuar barreres entre espècies, la qual cosa els convertix en models ideals per a estudiar l'evolució, la genètica de poblacions i les interacciones hoste-patogen.

Amb genomes que varien d'1.8 a 31 kb, els virus d'ARN solen utilitzar marcs de lectura superposats i proteïnes multifuncionals per a maximitzar l'eficiència codificadora. La replicació per polimerases d'ARN dependents d'ARN (RdRPs), que manquen d'activitat de correcció d'errors, dona com a resultat altes taxes de mutació i la formació de *quasispecies*, un núvol de variants genèticament relacionades. Esta diversitat permet l'exploració d'un vast espai de seqüències, proporcionant la matèria primera per a la selecció i l'adaptació ràpida.

La genètica de poblacions ajuda a explicar com la mutació, la recombinació, la selecció i la deriva donen forma a l'evolució viral. Els modes de replicació—*geomètric* o *de màquina d'estampat*—influïxen en com s'acumulen les mutacions. La replicació geomètrica amplifica ràpidament els genomes mutants, mentres que la replicació de màquina d'estampat limita la propagació de mutacions. Les respostes immunitàries de l'hoste, els colls de botella en la replicació i la dinàmica dins de l'hoste també influïxen en les trajectòries evolutives.

Una característica destacada de les poblacions virals és la presència de *genomes virals defectuosos* (DVGs), produïts durant la replicació. Encara que incapaços de replicar-se de manera autònoma, els DVGs poden replicar-se en *trans* amb l'ajuda dels

virus *wildtype* (WT). Alguns DVGs, coneguts com *partícules defectuoses interferents* (DIPs), competixen amb els virus WT pels recursos de replicació, interferint així amb la seua replicació. Primer identificats en el virus de la influença, els DVGs ara són reconeguts en molts virus d'ARN. Durant molt de temps considerats subproductes de la replicació, ara s'entén que els DVGs són funcionalment rellevants. Poden atenuar infeccions en reduir la replicació dels WT i activar fortes respostes immunitàries innates. Alguns DVGs poden codificar pèptids o ARN reguladors, la qual cosa suggerix un potencial funcional més ampli. També influïxen en l'evolució viral: en alta multiplicitat d'infecció (MOI), les interaccions entre WT i DVGs s'assemblen a sistemes socials, on els DVGs poden actuar com *tramposos* aprofitant-se de la replicació dels WT. No obstant això, els DVGs també poden donar suport a la persistència o modular la immunitat de l'hoste, revelant un espectre des de rols cooperatius a antagònics. Els colls de botella en la transmissió poden eliminar o fixar els DVGs depenent de la selecció i la deriva genètica. Així, els DVGs són fonamentals per a l'ecologia i evolució viral. Modulen la replicació, la immunitat i la diversitat genètica, desafiant les visions tradicionals de les poblacions de virus d'ARN i destacant la importància de la dinàmica de poblacions en l'adaptabilitat viral.

L'evolució viral també és impulsada per les defenses de l'hoste. Amb el temps, els hostes han desenrotllat mecanismes per a detectar i eliminar virus. Mentres que la immunitat innata proporciona una defensa ràpida i no específica, la immunitat adaptativa oferix especificitat i memòria. Malgrat les diferències taxonòmiques, moltes respostes antivirals estan conservades, especialment la interferència d'ARN (RNAi).

Els organismes model ajuden a disseccionar estes respostes. En plantes, *Arabidopsis thaliana* és un model clau a causa del seu coneixement genètic i el seu cicle de vida curt. Les seues estratègies antivirals inclouen el silenciament d'ARN, la repressió translacional i la senyalització hormonal que involucra àcid salicílic, àcid jasmónico o etilé. Les proteïnes tipus Dicer i les proteïnes Argonauta guien les respostes basades en RNAi contra els ARN virals.

La regulació epigenètica també juga un paper. La metilació de l'ADN, sovint guiada per xicotets ARN a través de la via de metilació dirigida per ARN (RdDM), pot silenciar seqüències virals i donar forma a la memòria immunitària. No obstant això, com estes modificacions retroalimenten l'evolució del virus continua sent poc explorat.

En animals, *Caenorhabditis elegans* és un altre sistema model. El virus d'Orsay (OrV), l'únic virus natural conegut que infecta a *C. elegans*, causa símptomes lleus, la qual cosa suggerix una relació coevolutiva de llarg termini. Igual que les plantes,

*C. elegans* ha d'equilibrar respostes immunitàries efectives amb el risc d'autodany, il · lustrant el compromís entre immunitat i autotolerància.

Estos models revelen com els hostes detecten amenaces virals, orquestren respostes immunitàries i evolucionen sota pressió, oferint perspectives sobre estratègies antivirals conservades al llarg de la vida.

Les interacciones virus-hoste són fonamentalment coevolutivas, modelades per adaptacions recíproques al llarg del temps. Els virus evolucionen mecanismes per a explotar la maquinària de l'hoste, mentres que els hostes desenrotllen contramesures, creant una carrera armamentista evolutiva. La hipòtesi de la *Reina Roja* il · lustra esta dinàmica: els dos han d'evolucionar contínuament per a mantindre la seua aptitud relativa. Mentres que els virus muten ràpidament, el seu espai de seqüències està restringit per les demandes funcionals. En contrast, les defenses de l'hoste sovint evolucionen amb menys restriccions, permetent una major diversificació [162].

Els intercanvis entre virulència i transmissió són fonamentals per a l'evolució viral. Una alta virulència pot reduir la supervivència de l'hoste i limitar la propagació del virus, mentres que una virulència baixa pot afavorir la transmissió. Per exemple, una virulència reduïda en el VIH-1 es correlaciona amb una major transmissibilitat [12]. Anderson i May [6] van demostrar que, depenent de la disponibilitat de l'hoste i la dinàmica de transmissió, la virulència òptima pot ser baixa, alta o intermèdia. A partir d'això, Ewald [50] va proposar que el mode de transmissió influïx en l'evolució de la virulència. La transmissió vertical (de pares a fills) tendix a afavorir una baixa virulència o mutualisme, mentres que la transmissió horitzontal (entre individus) sovint selecciona per a una major virulència. Existixen dades empíriques que analitzen esta teoria, especialment en sistemes de virus transmesos per vectors i en virus de plantes [57, 125, 60].

Els virus coopten factors de l'hoste per a facilitar la replicació. En mamífers, les proteïnes virals sovint segresten components de la immunitat innata per al seu benefici [84]. En *Arabidopsis*, el factor de transcripció TGA1—normalment involucrat en respostes antibacterianes—interactua amb la RdRP de TuMV per a ajudar a la replicació viral [110, 149]. Els hostes també poden cooptar seqüències virals. Amb el temps, els *elements virals endògens* (EVEs) poden convertir-se en part dels genomes dels hostes, contribuint a la immunitat o a funcions noves com el desenrotllament placentari. Estos gens virals domesticats il · lustren la interacció evolutiva entre virus i hostes [55].

**Capítol 2: Objectius, Estructura de la Tesi i Altres Contribucions**

En esta Tesi, amb un aproximació de la biologia de sistemes, investigue la interacció dinàmica entre l'evolució viral i les respostes transcriptòmiques de l'hoste. Basat en quatre articles publicats, este treball analitza dades de seqüenciació d'ARN en diversos sistemes hoste-virus i escales evolutives per a explorar com les poblacions virals s'adapten i com responen els hostes a nivell fenotípic i molecular.

La Tesi aborda dos aspectes complementaris de la biologia dels virus d'ARN: (1) les trajectòries evolutives a llarg termini de tres virus (+)ssRNA—TuMV en *A. thaliana*, i els betacoronavirus virus d'hepatitis murina (MHV) i coronavirus humà OC43 (HCoV-OC43) en cèl · lules de mamífer—i (2) la dinàmica d'infecció a curt termini de l'OrV en *C. elegans*.

En combinar evolució experimental, seqüenciació d'alt rendiment i modelatge computacional, el treball proporciona una perspectiva multidimensional sobre les interaccions virus-hoste, que va des d'infeccions individuals fins a evolució experimental.

Els estudis també s'organitzen en dos eixos principals: un centrat en la diversitat viral, la dinàmica de quasispecies i l'aparició de DVGs; i l'altre centrat en les respostes de l'hoste a la infecció. Este enfocament integrador busca una comprensió completa de com evolucionen les poblacions virals en diferents contextos biològics i com els hostes remodelen els seus transcriptomes en resposta a la infecció.

Els objectius generals i específics d'esta Tesi són:

- Aplicar un enfocament de biologia de sistemes a l'estudi d'infeccions virals causades per virus d'ARN monocatenari de sentit positiu ((+)ssRNA), enfocant-se tant en la dinàmica viral com en les respostes de l'hoste en dos contextos temporals distints: a curt i a llarg termini.
  - Estudiar el procés d'adaptació del TuMV a genotips de *A. thaliana* amb defectes en diferents vies epigenètiques.
  - Analitzar la resposta transcripcional de genotips de *A. thaliana* amb defectes en vies epigenètiques infectats amb aïllats de TuMV naïve i ben adaptats a *A. thaliana*.
  - Estudiar l'acumulació de DVG de dos betacoronavirus durant la seua evolució en diferents tipus de cèl · lules mamíferes susceptibles.
  - Estudiar l'acumulació d'OrV, el seu espectre mutacional, incloent DVGs, al llarg de la infecció en *C. elegans*.
  - Mapear les respostes transcriptòmiques de *C. elegans* durant el procés d'infecció per l'OrV.

El treball en esta Tesi no hauria sigut possible sense la col · laboració de les científiques que van realitzar tots els experiments. Per orde d'aparició dels articles publicats: la Dra. Silvia Ambrós, la Dra. Julia Hillung i la Dra. Victoria G. Castiglioni han format part del meu desenvolupament com a científica i dels resultats presentats ací.

Per la meua part, he realitzat totes les anàlisis de dades presentades en esta Tesi. Això inclou el preprocessament i control de qualitat de les dades d'RNA-seq, la identificació de variants i quantificació de la diversitat viral (variants de nucleòtid únic, mesures de diversitat nucleotídica i caracterització de DVGs), anàlisi transcriptòmic de l'hoste (expressió diferencial, agrupament d'expressió dinàmica en perfils, enriquiment funcional, ...), la majoria de les anàlisis estadístiques i el modelatge de les dinàmiques de replicació viral. Quan va ser necessari, es van desenrotllar o van ampliar progammes personalitzats per a complir amb les necessitats específiques de cada conjunt de dades. La majoria del codi desenrotllat està disponible de manera oberta en els repositoris referenciats en les seccions corresponents de Mètodes de cada capítol.

Algunes de les ferramentes computacionals que vaig desenrotllar i vaig aplicar en este treball, encara que no són part d'esta Tesi, han donat lloc a publicacions independents, com *DVGfinder* [123] o *SDgenerator* [120], que proporcionen recursos reutilitzables per a l'estudi dels DVGs en dades d'RNA-seq. La resta de les contribucions publicades durant aquest temps estan llistades al Capítol 2.

**Capítols 3 a 6: Resultats publicats**

Els Capítols 3, 4 i 5 corresponen a tres articles publicats i se centren en la dinàmica evolutiva a llarg termini dels virus. Els Capítols 3 i 4 s'enfoquen en l'evolució a llarg termini d'un virus d'ARN que infecta plantes amb alguns reguladors epigenètics alterats. Estos capítols es basen en un estudi d'evolució experimental de TuMV en mutants de *A. thaliana* amb defectes en vies epigenètiques clau.

**En el Capítol 3**, publicat en *Evolution*, investiguem com un virus emergent d'ARN evoluciona, interactua i s'adapta a poblacions d'una nova espècie hoste amb mecanismes de defensa controlats epigenèticament alterats. Els gens involucrats en la regulació epigenètica en *Arabidopsis thaliana* poden modular la susceptibilitat de la planta als virus, però si també influïxen en l'evolució viral continua sent desconegut. Ací, investiguem com els gens de l'hoste involucrats en processos epigenètics modelen l'evolució fenotípica de TuMV. Evolucionem llinatges de TuMV durant 12 passes en quatre genotips de *A. thaliana*, cada un afectat en diferents components de dos vies epigenètiques principals: metilació d'ADN (mutants en *dcl2 dcl3 dcl4*, *ddm1* i *polv*) i modificació d'histones (mutant en *jmj14*). En cada passada, registrem l'àrea sota la

corba de progressió de la ferida (AUDPS), la infectivitat (I) i la càrrega viral (VL), i analitzem les taxes d'evolució fenotípica. Els tres trets van evolucionar significativament durant l'experiment, amb AUDPS i I augmentant en tots els genotips, i VL exhibint dinàmiques més heterogènies. Estos trets van evolucionar més ràpid en els genotips amb major permissivitat basal. L'evolució fenotípica es va convergir a través dels llinatges, suggerint que les restriccions de selecció imposades per l'hoste limiten el paisatge adaptatiu del virus.

L'anàlisi de la diversitat viral (RNA-seq del primer (P1) i l'última passada (P12)) va mostrar un augment net en la diversitat genètica entre els llinatges virals. L'anàlisi de la diversitat genètica viral dins dels llinatges de TuMV (mesurat com APD) va mostrar un augment general significatiu després de 12 passes ($\sim 30.2\% \pm 2.9$), amb diferències estadísticament significatives entre els genotips vegetals. La magnitud de l'augment en diversitat va dependre tant del genotip com del número de passada, variant des d'un 20.0% en *ddm1* fins a un 44.7% en *dcl2 dcl3 dcl4*. Les línies evolutives del tipus silvestre (WT) van mostrar el segon menor increment. La distribució genòmica de la diversitat va revelar un augment en els llocs segregants de 518 (passe 1) a 1,206 (passe 12), sent la majoria variants úniques o singletons (82.0% i 86.1%, respectivament). La diversitat nucleotídica dins dels genotips hostes va augmentar 2.11 vegades, i la diversitat total de la mostra va augmentar 2.23 vegades. La diversitat entre genotips hostes va augmentar 48.3 vegades, la qual cosa indica una forta diferenciació genètica entre ambients hostes. El coeficient de diferenciació nucleotídica (NST) també va augmentar significativament, mostrant que les poblacions inicialment homogènies van divergir a causa d'evolució específica de l'hoste.

Els valors del test D de Tajima van ser significativament negatius després de les dos passades ($D_1 = -2.457$, $D_{12} = -2.567$), la qual cosa suggerix selecció purificadora, expansió ràpida o mutacions deletèries segregantes. L'alta proporció de variants úniques recolza una expansió poblacional durant la infecció sistèmica. L'anàlisi $d_N - d_S$ a nivell de cistró va indicar una selecció purificadora generalitzada en nou cistrons, intensificant-se per a la passada 12. NIb va ser particularment afectat, mostrant una variabilitat inicial que després va ser homogeneïtzada (excepte en *polv*). Només VPg va mostrar signes de selecció diversificant, més forta a l'inici i debilitant-se amb el temps.

L'anàlisi de codons FEL va confirmar selecció purificadora en els codons G3, Y25 i P78. Nou mutacions no sinònimes van estar sota selecció positiva. El codó 79 va mostrar inicialment les variants L79F i L79I, que van desaparéixer per a la passada 12. Altres huit mutacions seleccionades es van agrupar en VPg (aa 110-120), revertint sovint a al · lels ancestrals per a la passada 12. Per exemple, el codó 115 va presentar

set al · lels nous diferents després de la passada 1, però en la seua majoria va tornar a l'ancestral N al final. Esta diversificació transitòria seguida de reversió és consistent amb una dinàmica de "salt de granota" (leap-frog), la qual cosa suggerix la reaparició de genotips primerencs per damunt de variants recents. La reversió probablement no és mutacional; en el seu lloc, la persistència de variants ancestrals a baixa freqüència és una explicació més parsimoniosa.

No es va trobar una associació consistent entre mutacions fixades i genotips hostes, amb l'excepció de R118H, present sol en línies evolucionades en *ddm1* i prèviament relacionada amb contextos de resistència. VPg, un objectiu conegut de selecció als potivirus, complix funcions essencials en la replicació viral, traducció i interaccions proteiques. La seua participació en l'adaptació ha sigut observada en múltiples estudis, especialment en relació amb interaccions amb factors eIF de l'hoste i esdeveniments de ruptura de resistència.

L'absència de mutacions fixades específiques de l'hoste, malgrat la divergència fenotípica, pot semblar sorprenent. Això, combinat amb canvis paral · lels entre genotips, suggerix que l'adaptació podria reflectir pressions ambientals compartides en lloc de restriccions específiques de l'hoste. No obstant això, esta interpretació pot reconciliar-se invocant variants de baixa freqüència per davall del llindar de detecció que col · lectivament determinen el fenotip, com s'ha descrit en altres sistemes virals. Així, la complexitat de les quasispecies virals, més que les seqüencies consens, podria impulsar l'adaptació específica a l'hoste sota pressió epigenètica.

**En el Capítol 4**, publicat en *BMC Genomics*, investiguem les respostes transcriptòmiques dels genotips mutants i de tipus silvestre de *A. thaliana* a la infecció amb aïllaments ancestrals o evolucionats de TuMV del conjunt de dades estudiat en el Capítol 3. En l'anàlisi transcriptòmic es van seguir tres estratègies de contrast. Primer, es van investigar les diferències entre cada transcriptoma mutant i el WT en la mateixa condició (control, infectat amb virus ancestral i infectat amb virus evolucionat). En condicions de control, *dcl2 dcl3 dcl4* va presentar el major nombre de gens diferencialment expressats (DEGs, 2073), seguit de *ddm1* (809). Després de la infecció amb el virus ancestral, *ddm1* va mostrar el major nombre de DEGs (1267), i després de la infecció amb virus evolucionats, *ddm1* i *polv* van mostrar un augment de DEGs, mentres que *jmj14* i *dcl2 dcl3 dcl4* van mostrar disminucions. Un total de 74 DEGs van ser compartits entre els quatre genotips mutants després de la infecció amb virus evolucionats, sent la majoria subexpresados. La proporció de DEGs compartits va augmentar del 1% en condicions de control al 72% amb virus evolucionats. La pertorbació transcriptómica va variar entre els mutants, amb *jmj14* mostrant la major pressió selectiva. Els mutants epigenètics

mostren canvis en un conjunt central de gens comuns. Sis DEGs van ser sobreexpressats en tots els genotips, independentment de l'estat d'infecció, incloent-hi gens involucrats en la biosíntesi de metionina i defensa (MRD1), desenrotllament floral (PI4KC3), i ARN no codificants. Estos gens són regulats per RdDM. A més, cinc DEGs van ser compartits per plantes infectades, independentment de l'evolució viral, relacionats amb respostes a l'estrés, meiosi i la metiltransferasa MET2. Segon, es va avaluar l'impacte de l'adaptació viral en les respostes de les plantes comparant transcriptomes de plantes infectades amb les línies virals evolucionades (P12) *vs.* ancestrals (P1). Els mutants com *jmj14* i *polv* van exhibir canvis d'expressió gènica més significatius que les plantes WT, amb distribucions alterades de gens activats i reprimits. L'anàlisi d'enriquiment funcional va destacar vies comunes, com la resposta a l'àcid salicílic, activada per virus evolucionats, i categories específiques, com la biosíntesi de carotenoides, enriquides en mutants específics com *ddm1* i *jmj14*. Un conjunt central de 1029 gens es va expressar diferencialment en tots els genotips de plantes, vinculat a l'adaptació de TuMV. Estos gens van mostrar patrons consistents de sobre- i infra-expressió, amb excepcions específiques en alguns genotips. L'anàlisi funcional va revelar processos enriquits relacionats amb la resposta a quitina, senyalització de l'àcid salicílic i metabolisme de l'àcid jasmònic. Tercer, en incloure l'anàlisi d'expressió gènica entre les plantes infectades amb el virus ancestral i les condicions de control, classifiquem cada gen en un dels nou perfils bàsics d'expressió. Totes estes anàlisis van revelar tant objectius comuns de l'adaptació viral com respostes transcriptòmiques específiques de genotip. Com era d'esperar, els gens relacionats amb la immunitat van ser diferencialment expressats després de la infecció. No obstant això, també es van sobrerepresentar categories funcionals addicionals, la qual cosa suggerix que l'adaptació viral està impulsada per factors més enllà dels gens de resistència clàssics.

A més, un estudi previ de Corrêa *et al.* va mostrar que els TEs a les regions centromériques i pericentromériques van ser induïts d'hora en les plantes WT infectades amb TuMV evolucionat. Estos TEs van ser més prominentment induïts o reprimits en etapes posteriors de la infecció. En el nostre estudi, trobem una correlació significativa entre els DEGs i els TEs prop dels seus llocs d'inici de transcripció (TSS). Mentres que la proporció de DEGs associats amb TEs va ser lleugerament més alta en tots els mutants que en WT, les plantes *ddm1* van mostrar un augment significatiu de tals DEGs. Les dades suggerixen que els TEs prop dels promotors de gens tenen més probabilitats de correlacionar-se amb canvis en l'expressió gènica que aquells situats dins dels gens. Notablement, *rpp7*, associat amb la resistència a malalties, va tindre una alta densitat de TEs dins de la seua seqüència i va ser reprimit en

plantes *ddm1* i *jmj14* infectades amb virus evolucionats. Este gen està regulat per un mecanisme epigenètic que involucra el silenciament d'un ET "domesticat" en el seu intró. Una anàlisi d'enriquiment dels TEs associats amb DEGs que responen a l'adaptació de TuMV va revelar una sobrerepresentació significativa d'elements de classe II (transposons d'ADN) i una subrepresentació d'elements de classe I (retrotransposons). Els TEs de classe I, particularment Gypsy, van anar significativament menys freqüents en els genotips, excepte en les plantes *ddm1*, on uns certs tipus de SINE i LINE van ser enriquits. D'altra banda, els transposons de classe II, especialment Helitrons, van ser sobrerepresentats, amb un marcat enriquiment tant a les regions del promotor com en les seqüències codificants dels DEGs, la qual cosa suggerix el seu possible paper en el conflicte epigenètic i en les vies de resposta a l'estrés.

Finalment, explorem com l'adaptació de TuMV podria alterar la xarxa d'interacció proteïna-proteïna (PPI) entre el virus i l'hoste, integrant els nostres resultats transcriptòmics amb dades existents de PPI. Els resultats van mostrar que en les plantes *ddm1*, va haver-hi una sobrerepresentació significativa de interactors activats, especialment per a proteïnes virals com P3 i P3N-PIPO. VPg va interactuar tant amb proteïnes reprimides com activades, algunes involucrades en la iniciació de la traducció i remodelat de la cromatina, mentres que la replicasa viral NIb va mostrar interaccions alterades amb factors de transcripció i ligases d'ubiquitina. Notablement, 6K1, una viroporina descrita com a interactora de PAL1, una fenilalanina amoni-liasa, la qual cosa suggerix un vincle potencial amb l'adaptació viral i la via del fenilpropanoid, que juga un paper en la immunitat vegetal. Estes troballes apunten a importants direccions futures en la investigació de la interacció virus-hoste.

**Capítol 5**, publicat en *Viruses*, presenta un estudi experimental separat en evolució centrat en la generació i acumulació de DVGs en dos betacoronavirus diferents. Analitzem la dinàmica a llarg termini dels DVGs durant passades serials dels dos betacoronavirus esmentats, HCoV-OC43 i MHV, en cultiu cel · lular. Les combinacions estudiades van ser: HCoV-OC43 en cèl · lules BHK-21 i en cèl · lules HCT-8 i MHV en cèl · lules CCL-9.1. Els virus també van ser evolucionats sota multiplicitats altes i baixes d'infecció (MOI). Es va realitzar un RNA-seq en les passades 0, 11, 23, 35 i 47. Identifiquem *in silico* i caracteritzem DVGs i analitzem els seus patrons d'acumulació. Els DVGs van mostrar una major diversitat i abundància al llarg del temps, amb delecions i insercions com les característiques més comunes en riquesa i abundància. En HCoV-OC43, les distribucions i dinàmiques de grandària dels DVGs van diferir entre condicions de MOI. Els DVGs curts i llargs van dominar sota baix MOI, mentres que van sorgir i es van fer prevalents noves grandàries enlaire MOI, suggerint una

tendència de selecció. Per a MHV, la distribució de grandàries de DVGs va ser més uniforme, amb menys diferències pronunciades entre tractaments de MOI. També identifiquem regions hotspot per a delecions a les regions codificants de proteïnes estructurals i accessòries, particularment baixe alt MOI. E i Nsp12, les dos proteïnes virals amb potencial de viroporines, van ser identificades com les més afectades per les delecions. Finalment, estudiem la persistència d'alguns DVGs, especialment delecions, i identifiquem hotspots on els DVGs persistents s'acumulen. La identificació de les mateixes delecions persistents en línies independents va reforçar la idea que alguns DVGs han d'estar sota selecció positiva. Estos resultats destaquen la generació i evolució generalitzada de DVGs, modelades per factors tant virals com ambientals, i el seu paper potencial en la modulació de l'aptitud i persistència viral.

El Capítol 6 és centra en la dinàmica a curt termini de la infecció per OrV en *C. elegans*. A l'Apèndix A es mostren resultats addicionals al Capítol 6, centrats exclusivament en les dinàmiques virals.

**En el Capítol 6**, publicat en *Science Advances*, examinem la progressió de la infecció per OrV durant el desenrotllament larvari en *C. elegans*, amb l'objectiu d'entendre com la càrrega viral i les respostes de l'hoste evolucionen amb el temps. OrV és l'únic virus natural conegut de *C. elegans*, i generalment té un impacte mínim en l'aptitud a causa del robust sistema immune innat de l'hoste. Es va seguir la progressió de la infecció des de 0 fins a 44 hores post inoculació (hpi) amb diferents mètodes (RT-qPCR, RNA-seq i smiFISH). Identifiquem quatre etapes distintes d'infecció, marcades per canvis dinàmics en la càrrega viral, amb RNA2 (codificant de la càpside) aconseguint el seu pic a l'inici de l'egresió viral. *(i)* Una fase prereplicació (0 a 6 hpi), *(ii)* una replicació viral exponencial que porta a un pic de càrrega viral (8 a 12 hpi), *(iii)* una fase d'altiplà amb nivells moderats de càrrega viral (aproximadament la mitat del primer pic) (14 a 36 hpi), i *(iv)* una fase d'infecció residual persistent o resolució de la infecció, en la qual la càrrega viral disminuïx però es manté per damunt dels nivells inicials (38 a 44 hpi). L'anàlisi transcriptòmic de dotze punts temporals ens va permetre caracteritzar els canvis temporals en l'expressió gènica de l'animal. Primer, mitjançant la identificació dels gens diferencialment expressats en cada etapa, assenyalem un conjunt central de 29 gens ubics en la replicació viral exponencial i etapes moderades (com *B0507.6*, *col-133*, *ddn-1*, *droe-8*, *eol-1*, *fbxa-182* o *math-15*). Vint-i-un d'ells pertanyen a la resposta a patògens intracel · lulars i altres han sigut descrits prèviament com a dependents de ZIP-1. La resta dels gens—*col-133*, *math-15*, *trpl-5*, *Y43F88.12* i *Y6G8.5*—no havien sigut prèviament associats a la resposta de l'animal a OrV. Una anàlisi d'enriquiment funcional sobre els gens regulats positiva i negativament

per punt temporal va revelar dos grans ones d'activitat transcriptòmica, una entre 12 i 18 hpi i una altra a les 26 hpi, que van involucrar tant ontologies gèniques com vies metabòliques. Les respostes primerenques van incloure la regulació positiva de gens de vigilància d'RNA i guia axonal, mentres que els gens relacionats amb la traducció en mitocondri, formació de cutícula i metabolisme d'àcids grassos van ser reprimits en diferents etapes de la infecció. Segon, per a identificar gens l'expressió dels quals va ser significativament influenciada per la infecció amb el temps, apliquem una prova de raó de versemblança (LRT) per a la interacció condició i temps. Este enfocament va descobrir 2313 DEGs, que van ser agrupats en clústers basats en els seus perfils temporals, revelant diferents patrons de regulació primerenca i tardana. Notablement, 45 gens van mostrar fortes correlacions positives ($r \geq 0.8$) amb la càrrega viral, incloent-hi gens coneguts dependents de IPR i ZIP-1, mentres que 13 uns altres, com *elo-9* i *clec-24*, van mostrar patrons anticorrelacionats. Tercer, per a distingir les respostes específiques de OrV de les respostes generals a altres patògens, comparem les nostres dades amb 56 conjunts transcriptòmics, incloent-hi infeccions per bacteris Gram-positius i Gram-negatius, fongs i altres experiments amb OrV. Esta comparació va identificar 188 gens alterats exclusivament per l'OrV i 208 gens comunament afectats en totes les categories de patògens. Els gens específics de OrV van mostrar perfils d'expressió diversos i es van enriquir en components citoplasmàtics i mitocondrials, biosíntesis d'aminoàcids, processos relacionats amb pèptids i la via del fosfat de pentosa. Els gens compartits van ser enriquits a la regió extracelul · lar, respostes immunes innates, funcions de col · lagen i cutícula, i biosíntesi de lípids, i van incloure el 60% dels gens IPR i nombrosos efectors antimicrobians. Finalment, la validació experimental d'alguns dels gens identificats, ens va permetre suggerir nous factors antivirals (*lys-2*, *sago-2*) i provirals (*C53A5.9*, *clec-60*, *eol-1*, *ppw-1* i *vsra-1*).

En el *Apèndix A* es mostren **resultats addicionals** als publicats en el **Capítol 6**. Els resultats extra se centren en l'evolució a curt termini i les dinàmiques poblacionals de l'OrV durant la infecció primària en *C. elegans*. Es va trobar que l'OrV seguix un model de replicació tipus "màquina d'estampat", amb poblacions virals dominades per replicació de les plantilles produïdes a l'inici. L'anàlisi de diversitat viral va revelar diversitat de nucleòtids en la cadena negativa que no es va transmetre a la cadena positiva, la qual cosa suggerix restriccions de replicació dins de les cèl · lules hoste. L'anàlisi micro- a macroevolutiu va destacar la cua de la RdRP com una regió sota conflicte evolutiu, potencialment reflectint pressions d'interacció hoste-virus. També informem de la primera evidència de DVGs en OrV, identificant una deleció recurrent en RNA2 que podria funcionar com un RNA subgenòmic produint una proteïna delta

més curta. A més, la predicció de nous marcs de lectura oberts xicotets (smORFs) va revelar un nou pèptid candidat, oferint noves perspectives sobre el potencial codificant de OrV i les seues estratègies evolutives.

Junts, els estudis d'aquesta Tesi fan diverses contribucions originals:

- Demostrar que l'adaptació viral està fortament condicionada pel genotip de l'hoste, particularment en el context de les vies immunològiques regulades epigenèticament.
- Proporcionar una visió a nivell de sistemes de les respostes transcriptòmiques vegetals a l'evolució viral, revelant canvis no sols en els gens immunològics clàssics sinó també en els processos metabòlics i reguladors.
- Caracteritzar els patrons d'evolució i acumulació de DVGs en dos betacoronavirus en diferents contextos (tipus cel · lulars i MOI), i identificar com les diferents condicions modelen la seua diversitat mentres proporcionen noves maneres d'analitzar *in silico* la població de DVGs.
- Descriure les dinàmiques temporals de les respostes transcriptòmiques de l'hoste durant la infecció per l'OrV, incloent-hi la identificació de respostes sostingudes i específiques per etapa d'infecció.
- Caracteritzar per primera vegada l'acumulació de genoma i antigenoma en les dinàmiques a curt termini de l'OrV i proposar el model de màquina d'estampat com el mecanisme de replicació predominant.
- Presentar la primera evidència de DVGs en OrV i suggerir la presència d'un RNA subgenòmic i un nou pèptid candidat.

Estes troballes secunden col · lectivament un model en el qual l'evolució dels virus d'ARN està estretament entrellaçada amb els mecanismes reguladors de l'hoste, i on la diversitat genètica dins de les poblacions virals té el potencial d'influir en l'adaptació i persistència viral.

En **Capítol 7**, presente una **discussió general** explorant alguns dels resultats obtinguts. Com ara l'evolució de la virulència, l'existència de compromisos en l'evolució experimental, o el paper dels DVGs en l'adaptació viral.

**Capítol 8: Conclusions**

*Conclusions sobre les Interaccions Virus-Hoste a llarg termini:*

Les dinàmiques coevolutives a llarg termini entre virus i els seus hostes van revelar diversos aspectes crítics. Primer, la regulació epigenètica exercix una profunda influència

en les respostes antivirals de *A. thaliana* i l'evolució de TuMV. TuMV va evolucionar en plantes mutants defectuoses en les vies de metilació de l'ADN i modificació d'histones, mostrant major acumulació viral i major virulència. És notable que els genotips més permissius de l'hoste van tendir a seleccionar formes virals menys virulentes, mentres que els genotips més resistents van afavorir l'evolució de ceps més virulents.

En estos contextos epigenèticament compromesos, TuMV va exhibir una evolució accelerada de trets fenotípics en comparació amb els que evolucionaven en plantes tipus salvatge. Un punt focal d'adaptació va ser la proteïna viral VPg, que es va confirmar com un important objectiu de selecció, independentment del genotip de la planta. La regió flexible de VPg, que abasta els residus 79 a 120, va acumular mutacions sota pressió de selecció positiva.

És important destacar que l'adaptació viral va tindre efectes recíprocs en l'expressió gènica de l'hoste. Un conjunt central de 1029 gens va mostrar una expressió alterada en tots els mutants epigenètics, així com en plantes tipus salvatge després de l'evolució viral. Els gens amb expressió augmentada es van associar principalment amb SA i SAR, senyalització d'etilé i respostes a quitina, la qual cosa indica una resposta defensiva amplificada. En contrast, els gens amb expressió disminuïda van participar en JA, metabolisme d'amines i aminoàcids aromàtics, i la supressió de la biosíntesi de fenilpropanoids, reflectint un compromís costós entre l'activació de defenses i el creixement de la planta, la qual cosa va resultar en fenotips de plantes més febles.

A més, la naturalesa de la modificació epigenètica va influir en l'expressió gènica prop de ET. Els mutants en les vies de modificació de l'ADN (*ddm1* i *polv*)—però no en la metilació d'histones (*jmj14*)—van afectar l'expressió gènica a les regions pròximes a TEs, particularment quan els TEs estaven prop dels promotors gènics. Estes associacions suggerixen que els TEs proximals als promotors són més propensos a influir en els canvis d'expressió gènica que aquells dins dels cossos gènics.

La integració de la transcriptòmica de l'hoste i la xarxa d'interacció proteïna-proteïna virus-hoste va permetre la identificació de 13 gens de *Arabidopsis* crítics per a l'adaptació viral. Entre ells, la disminució de *pal1*, que codifica el primer enzim en la via de síntesi de fenilpropanoids i que interacciona físicament amb la viroporina 6K1 de TuMV, per ceps virals evolucionats pot contribuir al fenotip de major debilitat observat en les plantes.

*Conclusions sobre l'Acumulació i Diversitat dels DVGs en Coronavirus:*

L'acumulació i diversitat de genomes virals defectuosos (DVGs) durant l'evolució del HCoV-OC43 en cèl · lules BHK-21 i HCT-8, i del virus MHV en cèl · lules CCL-9.1, es va trobar que depenien de factors cel · lulars i virals. Els DVGs no sols es van

acumular en els dos virus, sinó que també van evolucionar, amb condicions d'alta MOI promovent major diversitat i la persistència de formes específiques de DVG. Este patró suggerix la presència de mecanismes de selecció que afavorixen DVGs més aptes, la qual cosa podria impactar l'evolució viral.

Els DVGs més prevalents van involucrar delecions, particularment dins dels cistrons *E* i *ns12.9*. Estes delecions podrien influir en la patogenicitat i es van observar que van sorgir independentment en diferents llinatges i tipus cel · lulars, la qual cosa indica un possible avantatge adaptatiu que podria reflectir un compromís que afavorisca la supervivència de les cèl · lules hospedadores.

L'evolució dels DVGs va estar modelada tant pel genotip viral com per la identitat de la cèl · lula hoste. HcoV-OC43 va mostrar patrons de deleció més definits i estructurats en comparació amb MHV, destacant la importància dels factors específics de l'hoste i donant suport a l'ús de múltiples línies cel · lulars en futures investigacions sobre DVGs.

Les delecions persistents probablement experimenten selecció positiva per a modular la replicació viral i la patogènesi, la qual cosa planteja la possibilitat d'explotar els DVGs com a agents terapèutics. Els DVGs podrien inhibir la replicació del virus de tipus salvatge o servir com a antivirals d'ampli espectre.

A més, les condicions d'alta MOI van portar a patrons ben definits d'acumulació de DVGs, particularment en punts calents de deleció que afecten els cistrons que codifiquen proteïnes estructurals i accessòries. En contrast, les condicions de baixa MOI van resultar en major variabilitat en la distribució de la grandària dels DVGs, subratllant que la MOI és un factor clau en la dinàmica de l'evolució dels DVGs.

*Conclusions sobre la dinàmica d'acumulació i diversitat de l'OrV durant el curs d'una infecció primària en C. elegans i les respostes transcriptòmiques dels animals:*

Durant una infecció primària, OrV exhibix un mode de replicació tipus "màquina d'estampat".

Produïx una àmplia varietat de DVGs, sent les delecions, i les insercions de un nucleótid les més abundants. Els punts calents específics de deleció, així com les estructures cb de 5' i 3', suggerixen que la generació de DVG no és completament aleatòria, sinó que pot estar impulsada per mecanismes moleculars encara desconeguts.

Una deleció persistent en RNA2 és compatible amb la transcripció d'un ARN subgenòmic que podria donar lloc a una versió més curta de la proteïna delta lliure. Este resultat, obtingut *in silico*, requerix confirmació experimental.

A més, identifiquem un pèptid putatiu codificat per la cadena negativa de l'RNA1. Este producte putatiu també està pendent de validació experimental.

Es va trobar evidència de pressió selectiva tant a nivell micro com macroevolutiu en la cua carboxil-terminal de la RdRP, destacant esta regió com un possible element funcional clau en el cicle de vida de l'OrV.

Este treball també proporciona la resolució temporal més alta fins a la data de la resposta transcriptòmica de *C. elegans* a una infecció primària per OrV. En classificar els gens segons la seua dinàmica d'expressió, identifiquem gens de resposta primerenca, gens l'expressió dels quals es va correlacionar estretament amb la càrrega viral, i gens de resposta tardana associats amb la recuperació. Les comparacions amb respostes de l'hoste a altres patògens virulents van revelar un conjunt únic de gens de resposta específics de l'OrV.

Finalment, la validació experimental de diversos d'estos gens va confirmar els seus rols funcionals, identificant nous factors provirals (*C53A5.9*, *clec-60*, *eol-1*, *ppw-1* i *vsra-1*) i factors antivirals (*lys-2* i *sago-2*) en la resposta de *C. elegans* a la infecció per l'OrV.

# A: Extra Results to Chapter 6

# Extra Results: Dynamics of within-host replication and generation of diversity during single infections of Orsay virus in *Caenorhabiditis elegans*

María José Olmo-Uceda

## Contents

**1 Introduction** **2**

**2 Material and Methods** **4**

2.1 RNA extraction and RNA-seq preprocessing . . . . . 4

2.2 Splitting fragments by sense strand . . . . . 4

2.3 Quantifying the viral genomic and antigenomic RNAs . . . . . 4

2.4 Diversity metrics . . . . . 5

2.5 Selection metrics . . . . . 5

2.6 Statistics . . . . . 6

2.7 Replication model definition and fitting . . . . . 6

2.8 OrV proteins study . . . . . 6

**3 Results and Discussion** **7**

3.1 Characterization of OrV accumulation dynamics . . . . . 7

3.1.1 Accumulation of viral RNA1 and RNA2 is different through the infection process and strand dependent . . . . . 7

3.1.2 A stamping machine replication model is coherent with the observed differential accumulation of strands . . . . . 10

3.2 Characterization of OrV diversity dynamics, short-term scale. . . . . 11

3.2.1 Viral diversity is significantly different between the RNA viral species . . . . . 11

3.2.2 Nucleotide diversity strand differences accumulate in specific genomic regions . . . . . 13

3.2.3 Negative strand of RNA1 presents the highest number of SNV with 12 and 18 hpi samples been specially heavy in SNV allele frequency . . . . . 15

3.2.4 Within-cell evolution processes or biological constrains in the short-term evolution . . . . . 17

3.2.5 Two examples of the within-cell constrain phenomena . . . . . 20

3.2.6 19C>U SNV is detected exclusively in the (+) strand and would reduce the minimum free energy of the 5'UTR RNA1 secondary structure . . . 22
3.2.7 A variety of low frequency variants appears in a convergent manner during the infection process . 25
3.2.8 A sliding window of $K_A/K_S$ values revealed silent and non-silent mutational hotspots . . . 27
3.2.9 Signatures of selection at different time-scales of evolution . . . 29
3.2.10 RdRP tail as a target of selection at different time-scales . . . 33
3.2.11 Deletions in RNA 2 are the predominant DVG type . . . 36
3.3 New insights into the transcription of OrV genome . . . 44
3.3.1 A sgRNA candidate (or a really persistent deletion) would code for a shorter free-$\delta$ version . . . 44
3.3.2 Small ORFs in OrV and a new peptide candidate . . . 47
**4 Conclusion** **48**

# 1 Introduction

Evolution refers to any change in the genetic composition of a population over time, no matter on the time-scale involved, or the transience of the change [16]. In the context of viral infections, genetic diversity is the raw material upon which deterministic and stochastic forces operate. RNA viruses generate significant variability due to high mutation rates, high effective population sizes, and the lack of proofreading error-prone replication mechanisms in most of them [52]. The newly generated variability is shaped by selective pressures, including host immune responses, and stochastic processes, such as genetic drift and viral bottlenecks [15].

Orsay virus (OrV) was first identified in 2011 infecting a natural population of *Caenorhabditis elegans* [22]. It is a non-eveloped, positive-sense single stranded RNA virus ($(+)ssRNA$) with a bi-segmented genome that encodes for three known proteins and a fusion version of two of them. RNA1 (3421 bases) encodes for a RNA dependent RNA polymerase (RdRP) while RNA2 (2574 bases) encodes for the capsid protein (CP, also called $\alpha$ protein), the $\delta$ protein and, through a ribosome frameshift, a fusion of CP and $\delta$ protein [33]. CP and the CP-$\delta$ fusion are the structural proteins. The OrV particle, as the other nodaviruses, has an icosahedral structure with T = 3 symmetry. The CP-$\delta$ fusion is necessary for cell attachment and entry [19] while the free $\delta$ protein is necessary for the non-lytic egression of the new particles [64]. OrV, together with Santeuil virus, Le Blanc virus and Melnik virus form a new clade of nematode-infecting noda-like viruses [21].

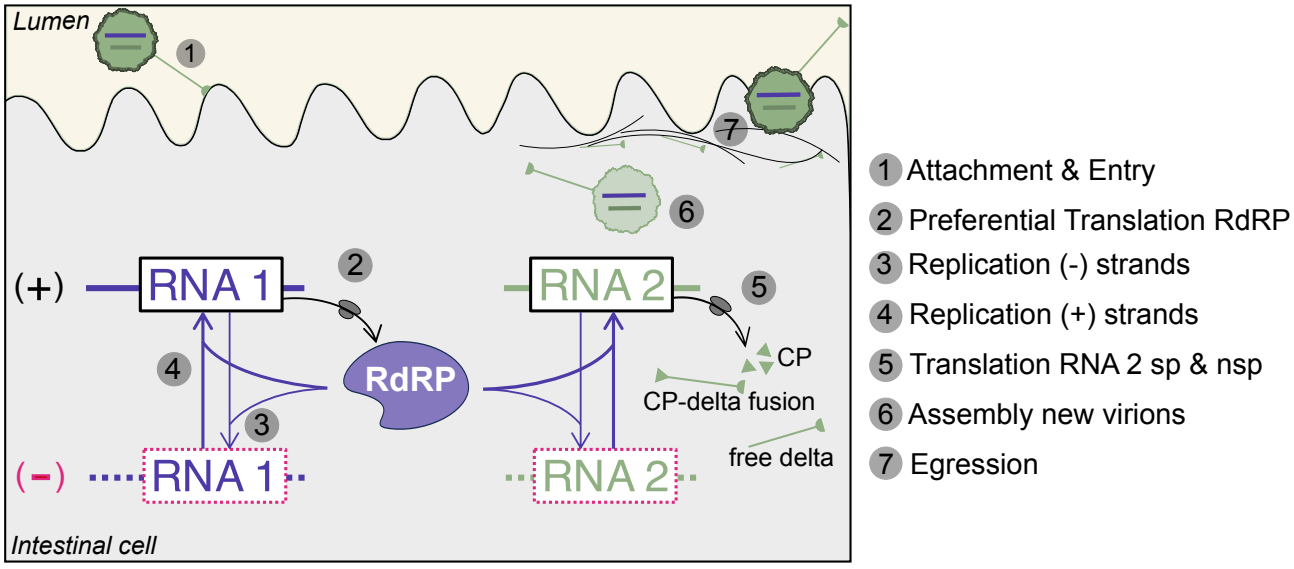



Figure 1: **Schematic representation of the OrV infection cycle.** OrV follows an orofecal route. Once the nematode ingests the virus, the virion can entry into the intestinal cell (anterior tract). The host ribosomes translate the RdRP that start making copies of the genome through a negative strand intermediate. At some point of the process, the structural (sp) and non-structural (nsp) proteins coded in the RNA2 are translated and the new virions assembly and egress in a non-litic manner.

We can hypothesize that once a particle enters the cell, the released genomic (+)RNA can be translated by the ribosomal host machinery. Then, the viral RdRP will use the (+)RNA to make intermediate antigenomic (-)RNAs that serve as templates for generating the progeny of (+)RNA copies (Figure 1). As in other (+)RNA viruses, the positive strand has multiple functions: *(i)* it is the genomic material of the virus (vRNA), *(ii)* it is the template from which the ribosomes will translate the viral proteins and, *(iii)* it is the template from which the RdRP synthesize the negative strand. By its part, the negative strand has a principal function: to serve as template to the RdRP to produce the progeny of genomic RNAs. Because the positive sense strand will form part of the viral progeny while the negative sense strand is required as intermediate of the process, several (+)RNA viruses have adapted their replication modes in order to accumulate less antigenomes. A stamping machine replication mode, where the antigenome derived from the parental virus is the predominant template for making new genomic copies, has been reported in several (+)RNA virus [40], although there is no evidence of the mode of replication in OrV. We will address the question in this work.

As mentioned above, RNA viruses have the highest mutation rates, but previous studies have shown that OrV (JUv1580 and JUv2572 strains) remains genetically stable after being passed several times in susceptible *C. elegans* [8]. No proofreading activity has been reported for OrV, raising the question of the primary nature of the stabilizing forces at play. The intra-host evolution of RNA viruses responsible for short-term acute infections is typically characterized by strong purifying selection [24], and we hypothesize that this could be the case for OrV as well. During acute infections with rapid viral replication, changes in the ancestral sequence are less likely to occur compared to chronic infections. However, physiological barriers may still exist, and both within-host and between-host bottlenecks can significantly impact the composition of the viral quasispecies [47]. Current evidence suggests that OrV can maintain persistent infections [10], providing a unique opportunity to study these processes in the pathosystem. For the time being, the process analyzed here pertains to the primary infection.

Characterizing the quasispecies that emerge in the short term is essential for understanding the virus's long-term evolution, as both stochastic and deterministic processes (drift and selection) work

on the existing genetic pool. This work aims to characterize the viral populations that emerge during a single OrV infection. In a previous experiment [9], we followed the course of an OrV infection in synchronized L1 *C. elegans*, taking periodic samples of the process from 0 to 44 hours post-inoculation (hpi). The results focused on virus accumulation, the virus tissular localization and, the host's responses at each infection stage. Here, we will delve deeper into the virus diversity and accumulation dynamics of positives and negative-sense strand of both RNAs, addressing some of the questions proposed above. Additionally, we present the first evidence of defective viral genomes (DVGs) generated in OrV and we propose the existence of a sgRNA that codes for a shorter version of the free $\delta$ protein. We will also propose a new peptide candidate for further investigation.

# 2 Material and Methods

Complete details of the experimental design, the generation of the time course experiment and the preparation of the samples are given in [9].

## 2.1 RNA extraction and RNA-seq preprocessing

Library preparation and Illumina sequencing were done by Novogene Europe (www.novogene.com) using a NovaSeq 6000 platform and a Lnc- stranded mRNA-seq library method, ribosomal RNA depletion, directional library preparation, 150 paired end, and 6 Gb raw data per sample. Novogene checked the quality of the libraries using a Qubit 4 Fluorometer (Thermo Fisher Scientific), qPCR for quantification, and Bioanalyzer for size distribution detection.

The quality of the Fastq files was assessed with FASTQC [2] and MultiQC [18] and preprocessed as paired reads with BBDuk (https://sourceforge.net/projects/bbmap/). Briefly, adapters were eliminated, the first 10 5' nucleotides of each read were cut, and the 3' end sequences with average quality below 10 were removed. Processed reads shorter than 60 nucleotides were discarded. (The parameter values were set to: ktrim = r, k = 31, mink = 11, qtrim = rl, trimq = 10, maq = 5, and forcetrimleft = 10, and minlength = 60).

## 2.2 Splitting fragments by sense strand

The processed fastq were aligned to the OrV reference with bwa mem [37] with default parameters. The resulting alignments were binarized and sorted with samtools [12]. Then, each alignment file was splitted in two having into account the strand orientation. Briefly, reads were identified as mapping the (+) or the (-)strand with the `bedtools bamtobed`. Then, "+" of the second in pair and "-" of the first in pair were classify as plus strand reads and "+" of the first in pair and "-" of the second in pair were saved as the minus strand reads. All homemade scripts are available at the github repository.

## 2.3 Quantifying the viral genomic and antigenomic RNAs

Total fragments of each viral specie ((+)RNA1, (+)RNA2, (+)RNA1, and (+)RNA2) were extracted with a house custom scripts from results of `samtools coverage` on the splitted alignments. Then, each viral molecular species was normalized by the length of their genomic segment (RNA 1: 3421

and RNA 2: 2574 nts) and the size of the library (number of fragments in the clean fastq files) with Eq. 1.

$$VPMB_i = \frac{viral\ genomes_i}{bases\ sequenced}10^6 = \frac{fragments_i \cancel{length(read)}/length_i}{library\ size \cancel{length(read)}}10^6 = \frac{fragments_i}{library\ size length_i}10^6 \tag{1}$$

For the profile analysis, VPMB values were centered and scaled as $z$-score values Eq. 2.

$$z_i = \frac{VPMV_i - \mu_i}{\sigma_i} \tag{2}$$

### 2.4 Diversity metrics

Nucleotide counts per position were obtained from BAM files with pysamstats version 1.1.2 (https://hpc.nih.gov/a
Then, the counts files were used to calculate the nucleotide frequencies use in the diversity metrics. Normalized Shannon entropy ($S_n$) (Eq. 3) was calculated for each genomic position as a measure of the diversity per locus with `NormShannon` of `QSutils` implementation, v1.16.1 (Guerrero-Murillo and Gregori Font, 2023). Higher ($S_n$) values indicate more heterogeneous viral populations whereas lower values indicate genetically homogeneous populations (Shannon, 1948; Gregori et al., 2014).

$$S_n = \frac{1}{log(N)} \sum_{i=1}^{N} p_i log(p_i) \tag{3}$$

For calculating the mean normalized entropy per sample, $\bar{S_n}$, only positions with at least 20 depth coverage were considered.

Second, single nucleotide variants were called with LoFreq [62] from the splitted-by-strand alignments (sec. 2.2) without optical duplicates. Only variants with an allele frequency up to 0.01, a minimum total coverage of 20 and a minimum SNV quality (Phred) of 20 were considered in the posterior analysis.

Last, DVGs were identify and characterized with `DVGfinder` metasearch tool [46]. The metasearch mode was used but only DVGs detected in at least 5 reads with ViReMa were included in posterior analysis. The richness was expressed as the number of different DVGs and the abundance was expressed in two ways: relative to the total sample as the sum of counts per total million reads ($\sum CPTM$) and relative to the viral population as the sum of counts per each hundred thousand of viral reads ($\sum CPVHT$).

### 2.5 Selection metrics

The direction of selection ($DoS$) statistic was calculated as in [58]:

$$DoS = \frac{D_n}{D_n + D_s} - \frac{P_n}{P_n + P_s} \tag{4}$$

, where $Dn$ and $Ds$ are the number of non-synonymous and synonymous substitutions and $P_n$ and $P_s$, the number of the non-synonymous and synonymous polymorphisms. $DoS > 1$ provides evidence of positive selection, while $DoS < 1$ is taken as an evidence of negative selection. $DoS \approx 1$ represents neutral evolution. The advantage of $DoS$ compared to the well-known and widely used $\omega$ is two fold: it does not assume neutrality of nonsynonymous changes and it does not considers only fixed substitutions, meaning it is more sentitive to detect weak purifying selection. A Tajima's neutrality test was performed with MEGA for the three coding sequences on the $1^{st}$ codon position.

### 2.6 Statistics

All statistics were performed in R version 4.4.2 (2024-10-31) in RStudio, v.2024.9.1.394 "Cranberry Hibiscus" [48]. Generalized linear mixed models (GLMM) were fitted to the data using the `glmer` function from the `lme4` package (Bates et al., 2015). The response variables were modeled using a Gamma distribution with a log-link function to account for the positive and skewed nature of the data. The models were fitted using maximum likelihood estimation (Laplace approximation). The statistical significance of the fixed effects was assessed using Wald $z$-tests with $P$ values calculated via asymptotic approximations.

The similarity between time series was evaluated using the Dynamic Time Warping (DTW) algorithm (`dtw` R implementation [25]). To assess variablility in the distance calculations, we performed a subsampling procedure by generating 1000 bootstrap time series. For each time series, we randomly selected one replicate per time point from a total of three available replicates, ensuring that the variation between replicates at each time point was captured.

### 2.7 Replication model definition and fitting

The `ode` function with `rk4` method from `deSolve` v1.40, was used to solve the system of ordinary differential equations [57]. The optimization of the parameters was performed with `optim` function from `stats` with the L-BFGS-B method, from the family of quasi-Newton methods that approximate the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) [5,38]. The fitting was applied to the data from 0 to 12 hpi.

### 2.8 OrV proteins study

The palmprint motifs of the OrV RdRP were identified using the palmID tool from the Serratus toolkit [3]. The protein structures were predicted with Alphafold 2 in the Colabfold implementation [34,42], and Alphafold 3 Server [1]. For the structural alignments, the DALI web server against all the PDB database was used [32].

# 3 Results and Discussion

## 3.1 Characterization of OrV accumulation dynamics

### 3.1.1 Accumulation of viral RNA1 and RNA2 is different through the infection process and strand dependent

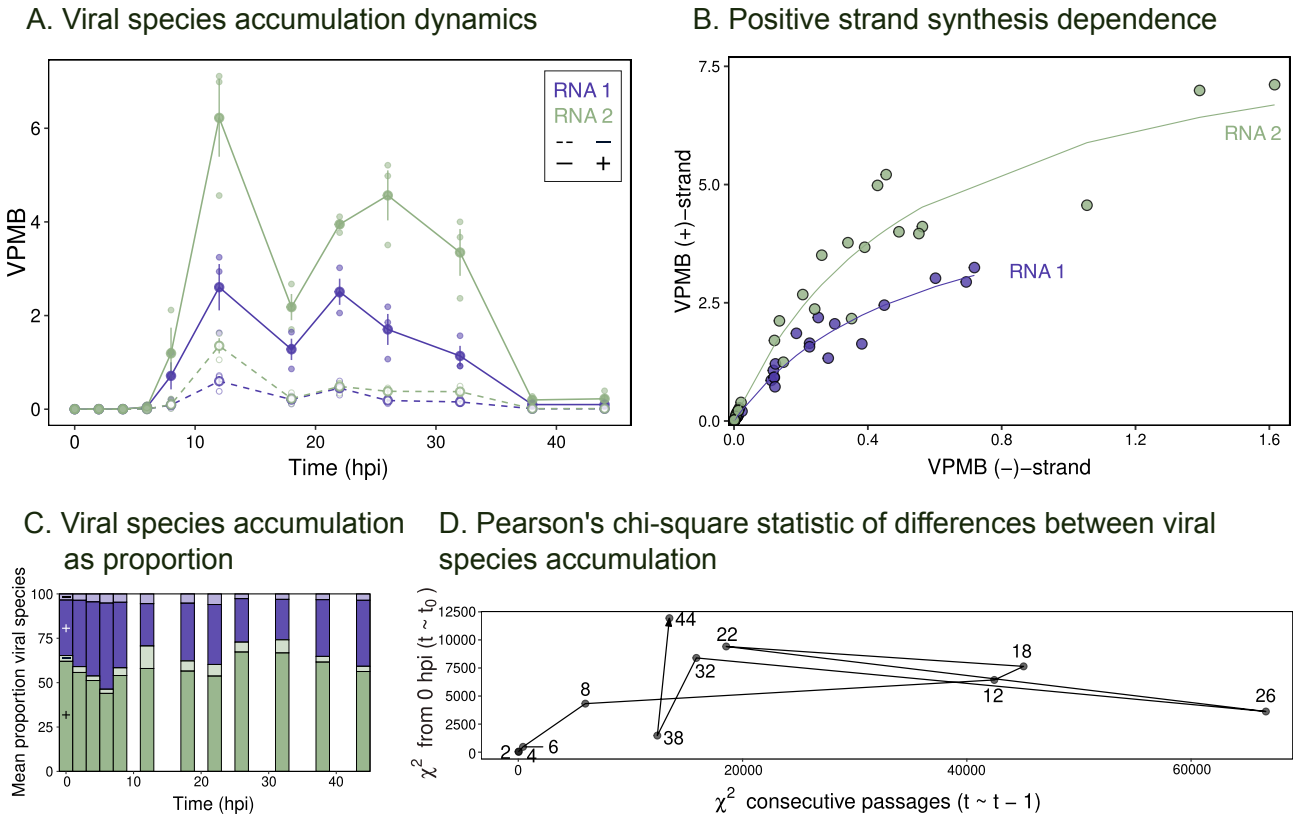



Figure 2: **OrV quantification**. (A) Viral species accumulation through the infection process. (B) Positive strand synthesis dependence on negative strand accumulation. (C) Viral species accumulation as proportion. As in the other plots, purple for RNA 1 and green for RNA 2. Lighter and darker colors for (-) and (+)RNAs respectively. (D) Pearson's $\chi^2$ statistic of the differences between the viral species accumulation between consecutive time points ($t \sim t-1$) and between each time point and the composition at the origen ($t \sim t_0$).

The accumulation of OrV during the infection process was describe as highly dynamic [9] and an isotonic regression model found the process could be significantly divided in four phases: ($i$) a prereplication phase (0 to 6 hpi), ($ii$) an exponential viral replication that leads to a viral load peak (8 to 12 hpi), (textitiii) a plateau phase with moderate viral load levels (approximate half of the first pick)(14 to 36 hpi), and ($iv$) a phase of persistent residual infection or resolution of the infection, in which the viral load falls but remains above initial levels (38 to 44 hpi) . Here we wanted to better characterize the dynamics of the same infection process accounting also for the strand accumulation. Hereafter and to simplify the nomenclature, we will refer to the positive strand of RNA1 and RNA2 as (+)RNA1 and (+)RNA2 and to the negative strand as (-)RNA1 and (-)RNA2 respectively.

In order to have a comparable measure of the four viral species, we used viral genomes per million of sequenced base ($VPMB$) units. The number of the viral species counts were normalized by the size of

the RNA (RNA1 is approximately 1/3 larger than RNA2) and the total library size per million of bases (Eq. 1). This measure assumes that all the genome is transcribed regularly and that the sequencing process access equally to all the transcripts. The exploration of the coverage plots indicates that these is not true. If the irregularities are due to biological and/or technical noise or are reflecting a more complex transcription process (as the presence of DVGs or intermediary transcripts as subgenomic RNAs) will be one of the questions addressed in this work.

A GLMM was used to assess the effects of the RNA segment, time post-inoculation, and the strand on the **viral accumulation** (VPMB), with sample as random effect. Within the entire process, RNA2 accumulates significantly more than RNA1 ($estimate = 0.3454,\ SE = 0.0638,\ P < 0.0001$), and the (+)RNA species accumulates over the (-)RNA ($estimate = 2.3140,\ SE = 0.0627, P < 0.0001$). The time variable also show a significant positive effect ($estimate = 0.1011,\ SE = 0.0322,\ P = 0.0017$). As expected, the accumulation of (+)RNA and (-)RNA strands was found highly correlated (Spearman's rank correlation for RNA1: $r_s = 0.9848, P < 0.0001$; RNA2: $r_s = 0.9799, P < 0.0001$). We found the relationship between the (+)RNA and (-)RNA strands of $RNA_i$ accumulation follows a saturation curve that can be express as ($i = 1, 2$):

$$(+)RNA_i = \frac{a_i \cdot (-)RNA_i}{b_i + (-)RNA_i} \tag{5}$$

In this context, $a_i$ represents the maximum asymptotic value to which the concentration of $(+)RNA_i$ converges and $b_i$ the concentration of $(-)RNA_i$ at which half of the maximum asymptotic value occurs. The model was fitted to the data for both RNA1 and RNA2 using nonlinear least squares regression with the following results: $a_{RNA1} = 5.3852 \pm 0.6035$, $a_{RNA2} = 8.9801 \pm 0.7970$, $b_{RNA1} = 0.5380 \pm 0.1033$, and $b_{RNA2} = 0.5546 \pm 0.1014$. These estimates were all four highly significant ($P < 0.0001$). The residual standard errors were 0.1957 for RNA1 and 0.492 for RNA2 (34 df), indicating a good fit (Fig. 2B).

We then investigate the **proportion** of the four molecular viral species along the infection process (Fig. 2D). A Pearson's $\chi^2$ test for count data was performed between the consecutive passages to evaluate the progressive changes and between each passage and the origin (0 hpi). All comparisons performed showed highly significant differences (Table 1).

| hpi | statistic consecutive | $P$-adjusted | hpi | statistic origin | $P$-adjusted |
|---|---|---|---|---|---|
| 0∼2 | 28.94 | 0.0164 | 0∼2 | 28.94 | 0.0164 |
| 2∼4 | 31.08 | 0.0094 | 0∼4 | 52.12 | $< 0.0001$ |
| 6∼8 | 5980.57 | $< 0.0001$ | 0∼8 | 4323.07 | $< 0.0001$ |
| 8∼12 | 42451.91 | $< 0.0001$ | 0∼12 | 6432.98 | $< 0.0001$ |
| 12∼18 | 45062.99 | $< 0.0001$ | 0∼18 | 7642.31 | $< 0.0001$ |
| 18∼22 | 18539.37 | $< 0.0001$ | 0∼22 | 9415.49 | $< 0.0001$ |
| 22∼26 | 66668.69 | $< 0.0001$ | 0∼26 | 3622.79 | $< 0.0001$ |
| 26∼32 | 15879.98 | $< 0.0001$ | 0∼32 | 8395.40 | $< 0.0001$ |
| 32∼38 | 12399.04 | $< 0.0001$ | 0∼38 | 1477.74 | $< 0.0001$ |
| 38∼44 | 13477.51 | $< 0.0001$ | 0∼44 | 11936.46 | $< 0.0001$ |

Table 1: Results of the Pearson's $\chi^2$ test for count data performed between the consecutive passages and between each passage and the proportions at the origen (0 hpi). $P$-adjusted values were obtain with the Benjamini and Hochberg (1995) method.

The largest changes in the viral species composition where found between 22 and 26 hpi ($\chi^2 = 66668.69, P < 0.0001$), followed by the changes between 12 to 18 hpi ($\chi^2 = 45062.99, P < 0.0001$) and 8 to 12 hpi ($\chi^2 = 42451.91, P < 0.0001$). The changes with respect the origen were not so

drastic and 44 hpi was found the most different timepoint ($\chi^2 = 11936.46, P < 0.0001$) followed by 22 hpi ($\chi^2 = 9415.49, P < 0.0001$) and 32 hpi ($\chi^2 = 8395.40, P < 0.0001$). Interestingly, a part from their nearest 2 and 4 hpi, time 32 hpi was the one that, although significantly different from 0 hpi, accumulated the smallest changes between them.

The differences between the viral species were not just on magnitude, but also in their **temporal dynamics**. In order to compare their profiles, we re-scaled the values using the $z$-scores transformation (Fig. 3A). These allowed us to qualitatively detect: $(i)$ A delay in the accumulation of RNA2 in the second wave. The maximum accumulation peak, specially for (+)RNA2 is reached later than (+)RNA1, with the posterior decay was also delayed. $ii)$ Comparing both peaks, 12 and 22 hpi for RNA1 and 12 and 22-26 hpi for RNA2, we can distinguish that (+)RNA1 values are similar between them, while second peak reaches less abundance for RNA2 (more than 2 SD in (-)RNA2 and almost 1 SD for (+)RNA2). We then apply a Dynamic Time Warping (DTW) algorithm in order to quantify the differences between viral species profiles. Fig. 3B shows the normalized Euclidian distances computed, mean and SD in the lower and upper triangle respectively. Profiles of the (-)RNA show the smallest differences ($0.0851 \pm 0.0123$) between them while in the (+)RNA, RNA1 and RNA2 showed $0.6312 \pm 0.0880$. Furthermore, distances within the same RNA but between strands were higher for RNA2 ($0.6620 \pm 0.0575$) than for RNA1 ($0.3204 \pm 0.0414$).

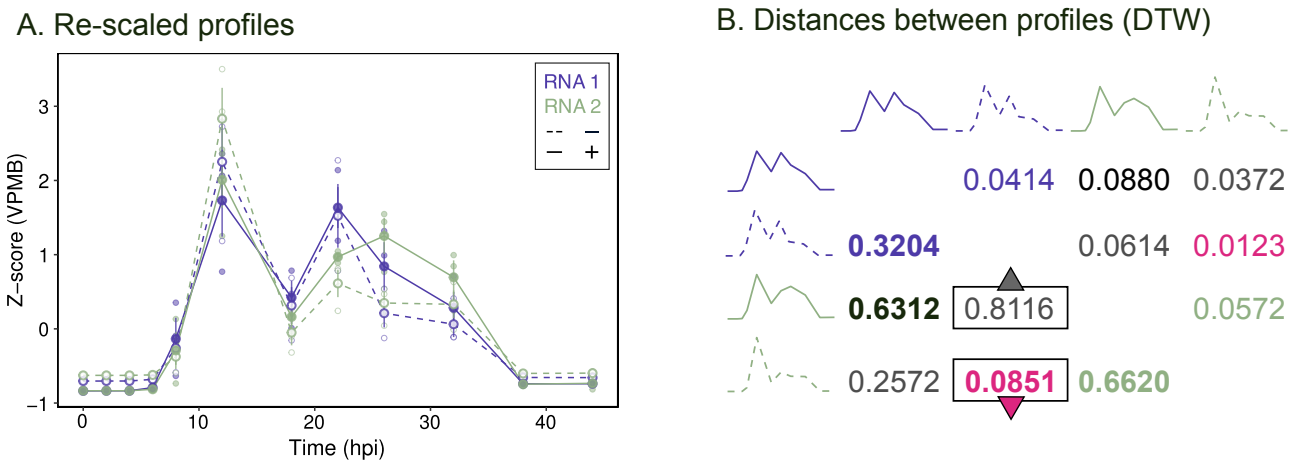



Figure 3: **Accumulation dynamics**. (A) Re-scaled accumulation profile. A $z$-score transformation of each viral specie time series was performed to facilitate the comparisons. (B) Distances between each pair of profiles. DTW distances calculated for each pair of profiles, the mean of 1000 bootstrap resampling in the lower triangle with 1 sd in the upper triangle. Viral species are represented with their $z$-score profiles. In order up to down and left to right: (+)RNA1, (-)RNA1, (+)RNA2, (-)RNA2.

Recapitulating, the accumulation of viral species was found highly dynamic with the (+)RNAs been more expressed than the (-)RNAs of both segments. RNA2 accumulated more than RNA1 in both polarities. The proportion of the viral species changed through time, with the 38 hpi been the most similar to the origin proportions. The biggest shifts in molecular viral species between consecutive passages were found in 22 - 26 and 12 - 18 comparisons, the timepoints with the highest viral load. Finally, the dynamics of the viral species was found significantly different between all the species except for the (-)RNA species.

| **State variables** (units) | $t_0$ | Description |
|---|---|---|
| $G_1$ (VPMB) | 0.00010 | (+) RNA1. |
| $G_2$ (VPMB) | 0.00019 | (+) RNA2. |
| $A_1$ (VPMB) | 0 | (-) RNA1. |
| $A_2$ (VPMB) | 0 | (-) RNA2. |
| **Parameters** (units) | Estimates | Description |
| $\rho_1$ (VPMB/h) | 10.0148 | RNA1 production rate. |
| $\rho_2$ (VPMB/h) | 9.6582 | RNA2 production rate. |
| $\sigma_1$ (VPMB/h) | 0.9863 | Factor of proportionality for SMR — GR continuum in RNA1 |
| $\sigma_2$ (VPMB/h) | 0.9846 | Factor of proportionality for SMR — GR continuum in RNA2 |
| $\Omega$ | - | Logistic growth restriction to viral amplification due to finite cellular resources. |
| $K$ (VPMB) | 8.8682 | Carrying capacity. |

Table 2: **State variables and parameters estimates with L-BFGS-B** VPMB = number of genomes per million of sequenced bases.

### 3.1.2 A stamping machine replication model is coherent with the observed differential accumulation of strands

From the above results, we hypothesize that the observed asymmetry in the strands accumulation was caused because a stamping machine mode (SMR) was predominantly governing the dynamics of replication. In a pure SMR, all the progeny is synthetised from the parental virus. In the case of a (+)ssRNA, the preferred template would be the (-)RNA genome (antigenome) derived from the infecting particle (+)RNA. The opposite strategy is the geometric replication model (GR), where the new progeny (no matter the strand sense) has the same probability of been used as template than the progenitor virus [17, 51, 59]. Intermediate modes of replication, where the SMR predominates but a small portion of the progeny can also be amplified (GR), have been described for bacteriophage $\phi 6$ [11] and for TuMV [40, 53].

We propose here a basic model focused on the first two stages of the viral replication cycle (Table 2, Fig. 4A). With this model, we wanted to address a simple question, can a smaller replication rate of the (-)RNA with respect the (+)RNA explain the lower accumulation of antigenome? For that, we considered a replication rate ($\rho_i$) and the introduction of a proportionality factor ($\sigma_i \in [0, 1)$) for the $\rho_i$ in the production of the (-)RNAs. If both genome and antigenome are produced at the same rate, $\sigma \to 0$, meaning a GR is governing the replication mode. On the other hand, the lower the production of antigenome compared to genome, the more $\sigma_i \to 1$ and more predomination of the SMR (Fig. 4A-B).

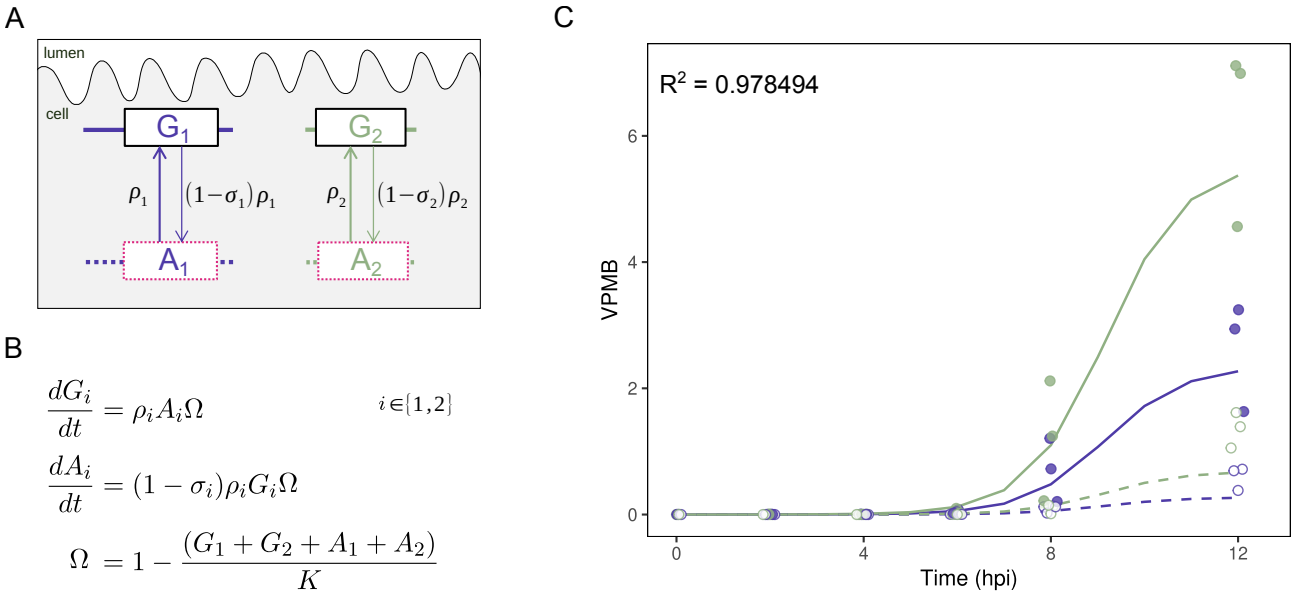



Figure 4: **Stamping machine replication model for OrV**.

The ODE model (Fig. 4B) was fitted to the experimental results (0 to 12 hpi) as described in the corresponding Methods section. The initial values of the state variables were set to the mean values at 0 hpi. Results of the fit ($\sigma_1 = 0.9862; \sigma_2 = 0.9846$, Table 2) showed that the asymmetry in the production rate with $\rho_i > (1 - \sigma_i)\rho_i$ effectively predicts the differential accumulation of the strands ($R^2 = 0.9785$, Fig. 4C). Therefore, we conclude that most of OrV (+)RNA progeny results from an SRM mechanism.

RNA viruses, and specifically (+)ssRNA viruses, have the highest mutation rates [52] but OrV has been described as genetically stable [8]. Its molecular evolution stasis could be explained by a strong negative selection. Additionally, the mode of replication highly impacts the spread of mutations and a stamping machine replication strategy would be favoring the maintenance of the original sequence [54]. We have test that a SMR mode is compatible with the differential strands accumulation found in our short-time experiment. Now on, we will further investigate the impacts in the short-term viral diversity.

## 3.2 Characterization of OrV diversity dynamics, short-term scale.

### 3.2.1 Viral diversity is significantly different between the RNA viral species

We hypothesize that the asymmetric accumulation of the viral species could have an impact on their diversity. To test this, we studied three different aspects of viral diversity: nucleotide diversity, the presence of strand-specific SNV, and the presence of DVGs).

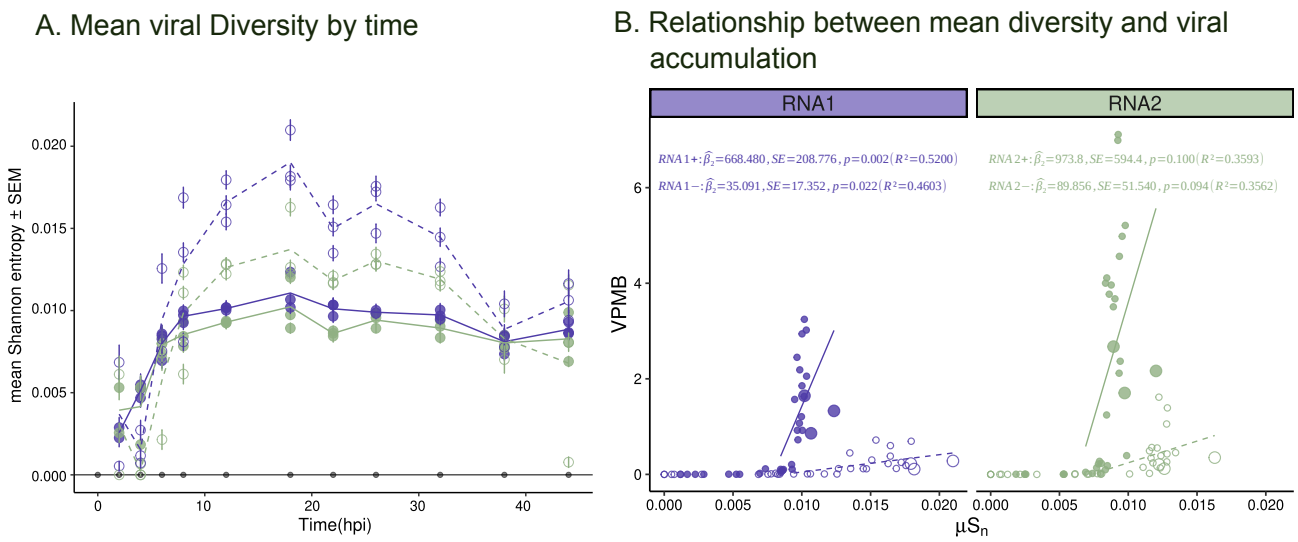



Figure 5: **Mean nucleotide diversity**. (A) $\bar{S_n}$ dynamics. Each empty dot represents the $\bar{S_n} \pm 1sd$ of the samples. The solid points represent the $\bar{S_n} \pm 1sd$ of the three replicates per time. Green is RNA1 and purple is RNA2. Solid lines for the positive strands and dashed lines for the negative strand. (B) Relationship between viral accumulation and $\bar{S_n}$. The lines represent the linear regression models post-breakpoint. Slopes and statistics showed in the corresponding equations. Green for RNA1 and purple for RNA2. Solid lines and solid points for the positive strands and dashed lines and empty points for the negative strands. The 18 hpi samples are represented with a bigger point size.

We used the normalized Shannon entropy index ($S_n$, hereafter) as a measure of nucleotide variability by position, where 0 represents the case of least variability (*i.e.*, a position is always occupied by the same nucleotide) and 1 represents the case of maximum variability, where each of the four nucleotides may occupy that position with equal probability.

The mean $S_n$ ($\bar{S}_n$) per sample is shown in Fig. 5A. Results show that $\bar{S}_n$ increased and stabilized at a certain value from 6 hpi in the positive strand while the negative strand reaches their maximum values later at 18 hpi. A linear mixed model was used to assess the effects of the RNA segment, time post-inoculation, and strand on $S_n$, with a random intercept for each sample. The model was fitted using restricted maximum likelihood (REML).

The results show that RNA2 had a significantly reduced $\bar{S}_n$ compared to RNA1 (estimate = -0.002, $t$ = -2.251, $P$ = 0.0267). Time post-inoculation had a significant positive effect, indicating that as time increases, the entropy also increases (estimate = 0.00016, $t$ = 3.336, $P$ = 0.0016). And the negative strand showed a marginally significant effect on $S_n$, with higher values of $S_n$ than the positive strand (estimate = -0.0018, $t$ = -1.983, $P$ = 0.0503). Interactions between RNA segment, hours post-inoculation (hpi), and strand were considered in the model but were generally not significant, with only a small effect of strand orientation interacting with hpi (estimate = -0.00006494, $t$ = -1.680, $P$ = 0.09612).

Very interestingly, 18 hpi some viral molecular species reached their maximum diversity values while we have described that the virus accumulation suffered a huge decay at that time (Fig. 5A). This observation prompted us to dig into the relation between the virus accumulation and its diversity.

A first visualization of the existing relationship between viral accumulation (VPMB) and diversity $\bar{S}_n$ showed a non linear relation (Fig. 5B). The linear relation appears after certain threshold in $\bar{S}_n$, so we analyzed that relationship using segmented regression models. These models revealed a distinct

change in the relationship between VPMB and $\bar{S}_n$, indicating a breakpoint after which the slope of the relationship significantly altered. Segmented regression models identified breakpoints ($\bar{S}_n$ values) for each viral species, which are summarized below in 3:

After the breakpoint, the relationship between VPMB and $\bar{S}_n$ was characterized by steeper slopes, indicating a stronger positive association between viral quantity and diversity. The slopes of the linear terms in the segmented models are in 3. The models showed significant shifts in the slope for the post-breakpoint phases for all species except `(-)RNA2` and `(+)RNA2`, which showed a less statistically significant change in slope after the breakpoint.

| **RNA** | **Breakpoint,** $\bar{S}_n \pm SE$ | $Slope \pm SE$**,** $P$ | | **Model fit** | |
|---|---|---|---|---|---|
| | | **Pre-breakpoint** | **Post-breakpoint** | **RSS** | $R^2$ |
| (-)RNA1 | $0.009 \pm 0.003$ | $1.173 \pm 12.823$ $P = 0.928$ | $35.091 \pm 17.352$, $P = 0.022$ | 0.1581 | 0.4603 |
| (+)RNA1 | $0.008 \pm 0.001$ | $10.196 \pm 115.452$, $P = 0.930$ | $668.480 \pm 208.776$, $P = 0.002$ | 0.7430 | 0.5200 |
| (-)RNA2 | $0.007 \pm 0.003$ | $0.878 \pm 34.586$, $P = 0.980$ | $89.856 \pm 51.540$, $P = 0.094$ | 0.3352 | 0.3562 |
| (+)RNA2 | $0.006 \pm 0.002$ | $0.191 \pm 470.800$, $P = 1.000$ | $973.800 \pm 594.400$, $P = 0.10$ | 1.8500 | 0.3593 |

Table 3: Summary of segmented model results for different RNA species. The table includes breakpoints, slopes before and after the breakpoint, as well as model fit statistics (RSS and $R^2$).

These findings suggest that the relationship between viral quantity (VPMB) and diversity ($\bar{S}_n$) is not linear. Instead, there is a clear threshold at which the relationship changes significantly, as indicated by the breakpoints and the changes in the slopes post-breakpoint. The steeper post-breakpoint slopes suggest that viral replication and diversity exhibit a stronger association beyond these threshold values. Additionally, we can infer that the positive strand reaches its plateau of $S_n$ (with only slight differences between the RNAs), while the negative strand, with much lower accumulation, has not yet reached its plateau (if it exists).

### 3.2.2 Nucleotide diversity strand differences accumulate in specific genomic regions

$\bar{S}_n$ was then evaluated. Only positions with at least 10 reads were included. A pattern of diversity hotspots start emerging at 6 hpi and was mainly mantained until the last time point (Fig. 6).

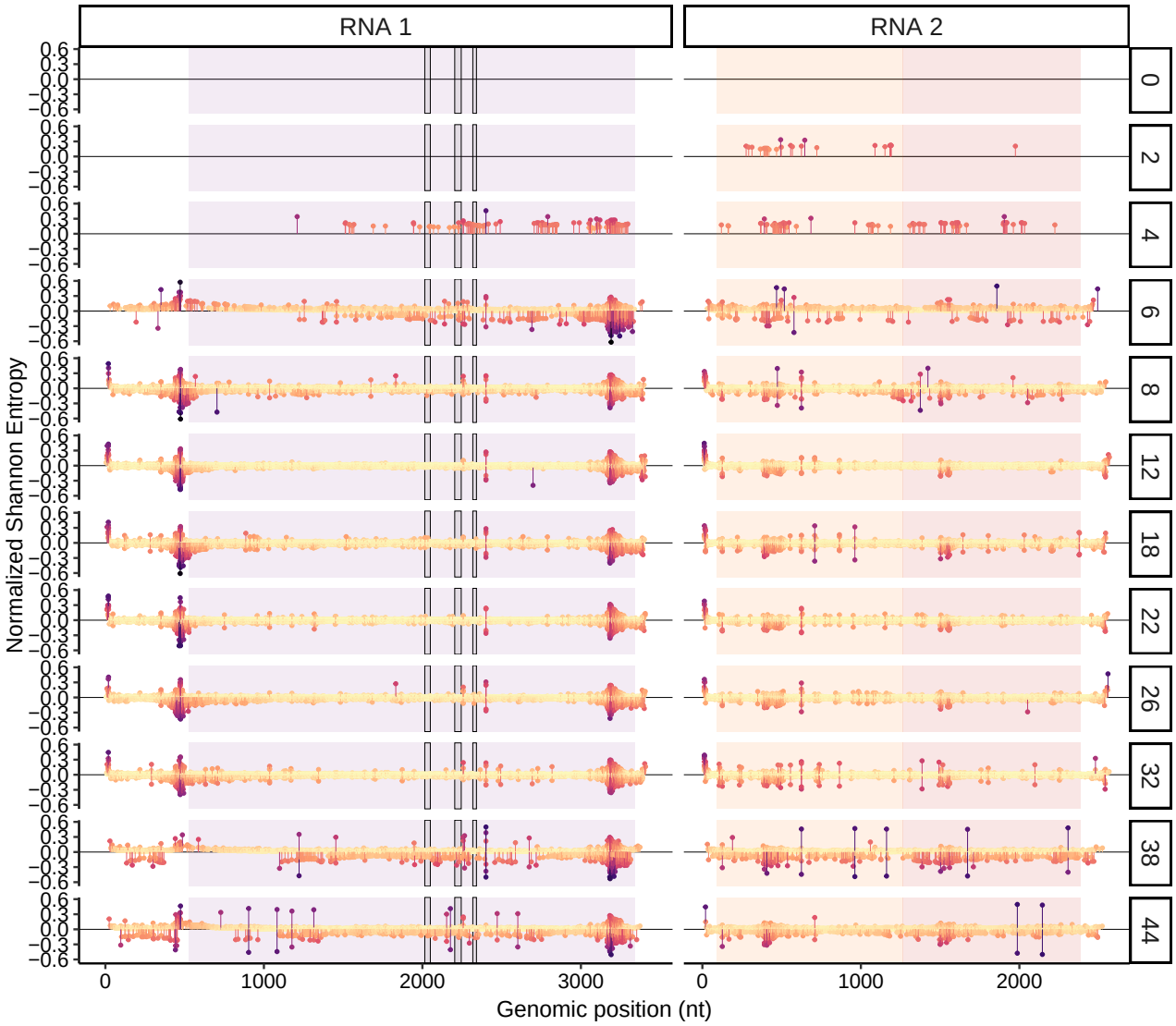



Figure 6: **Normalized Shannon Entropy as a measure of the diversity per genomic position**. $S_{np}$ was calculated independently for each strand, only positions with at least 10 reads were included. For simplicity $S_{np} = 0$ are not colored. Colors of dots and segments range from less entropy (yellow) to more entropy (purple). Background colors indicate coding regions. Gray rectangles on the coordinates of A, B and C motifs from the RdRP Palmprint domain

A linear mixed model (LMM) with strand and timepoint as fixed effects, and position and sample as random effects was run per each RNA segment. The results indicate that positive strand had significantly less $S_np$ than the negative strand in both RNAs (RNA1: estimate $= -0.0037 \pm 0.0002, P < 0.0001$; RNA2: estimate $= -0.0013 \pm 0.0002, P < 0.0001$). But differences due to the time and the interaction strand and time where no significant.

The LMM founded differences of $S_{np}$ in the strands, with the (-)RNA of position $p$ been predominantly greater than in (+)RNA. We then asked which positions showed the largest diversity differences depending on the strand. Because time had not a significant effect in the LMM, we use the average of the $S_{np}$ residuals (residual $\bar{S}_{np} = \bar{S}^{+}_{np} - \bar{S}^{-}_{np}$) to measure the differences in diversity. Then, a 50-window with 10 nt of step was used to quantify the number of $p$ with residual $S_{np}$ above the 99 percentile (P99) or under the 5 percentile (P5). Positions above the thresholds are represented as red dots in Fig. 7,

top. In Fig. 7 bottom, the results of the sliding-window strategy indicate these positions accumulate mainly in the start of the 5'UTR of both segments. Last region of the RdRP, and the final of the first third of CP and $\delta$ genes show also these type of events.

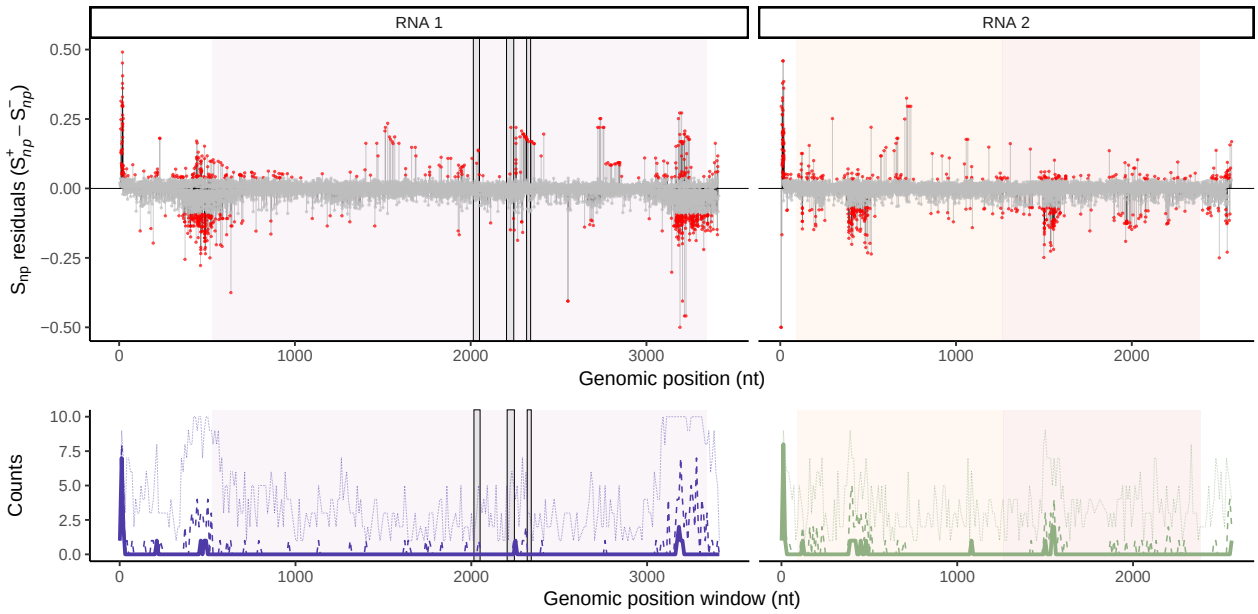



Figure 7: **Diversity strand bias accumulates in specific genomic regions**. *Top. residual* $S_{np} = \mu S^{+}_{np} - \mu S^{-}_{np}$ results above P99 and below P1 are colored in red. Only *residual* $S_{np}$ under Q1 or above Q3 are represented. *Bottom* Sliding window count (window = 50 nt, step = 10 nt) of *residual* $S_{np}$. Solid lines: values under P1 and above P99; dashed lines: values under P5 or above P95 and dotted line: values under Q1 and over Q3 were considered. Background colors indicate coding regions. Gray rectangles on the coordinates of A, B and C motifs from the RdRP Palmprint domain.

### 3.2.3 Negative strand of RNA1 presents the highest number of SNV with 12 and 18 hpi samples been specially heavy in SNV allele frequency

Following the diversity study, we performed a single nucleotide variant (SNV) calling at the viral species level. In addition to quality filters, only SNV with an allele frequency ($AF$) $> 0.01$ were considered.

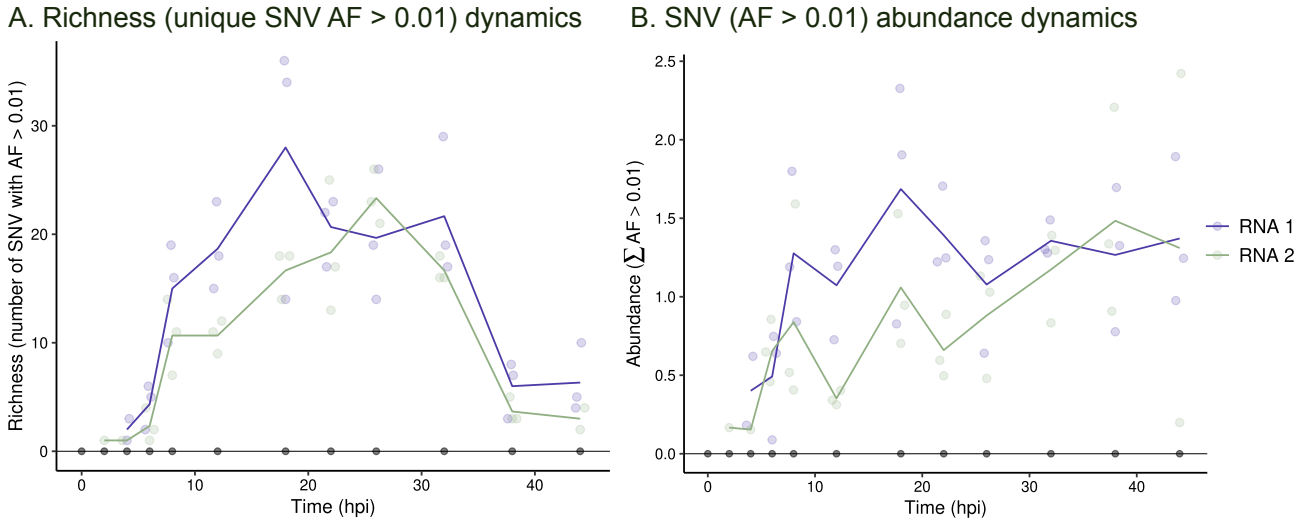



Figure 8: **SNV richness and accumulated AF**. (A) SNV richness dynamics. The richness was calculated as the number of SNV with an $AF > 0.01$ per sample. Line represent $\mu_{Richness}$ profile of RNA1 (purple) and RNA2 (green). (B) SNV abundance dynamics. Abundance was calculated as the $\sum AF$ of all SNV with an $AF > 0.01$. Line represent the $\mu SNV abundance$ of RNA1 (purple) and RNA2 (green).

A preliminary summary of the number of SNVs identified in each RNA segment is shown in Fig.8A. Richness was calculated as the number of SNVs above the $AF$ threshold in each sample, regardless of the strand orientation. A LMM with sample as a random effect indicated that RNA1 contained more unique SNVs than RNA2 (estimate $= 3.756, P = 0.0058$). When richness was rescaled to the genome size (RNA1 is 1/3 longer than RNA2), the differences were no longer significant. Fig.8B shows the sum of the $AF$ over time. $AF$ accumulation was used as a proxy for the total abundance of SNVs in each sample. A similar analysis indicated that RNA1 had a slightly higher weight in SNVs compared to RNA2 (estimate $= -0.2977, P = 0.0435$), but when values were rescaled to RNA size, the differences disappeared.

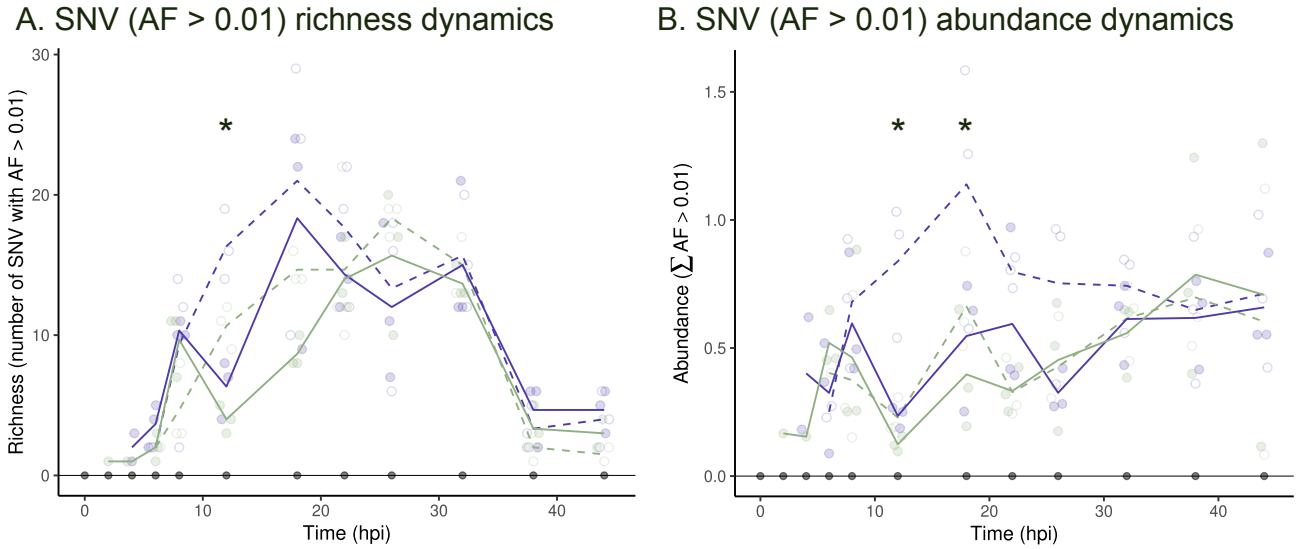



Figure 9: **SNV richness and accumulated AF**. (A) SNV richness dynamics. The richness was calculated as the number of SNV with an $AF > 0.01$ per sample. Line represent the $\mu_{Richness}$ profile of positive (solid) and negative (dashed) strains of RNA1 (purple) and RNA2 (green). (B) SNV abundance dynamics. Abundance was calculated as the $\sum AF$ of all SNV with an $AF > 0.01$. Line represent the $\mu_{SNVabundance}$. The asterisks indicate the time points where a *post hoc* Tukey test found a specie with a significant higher value ($P < 0.05$).

The most interesting results were observed when the analysis was performed separately by strand orientation. A LMM was used to evaluate differences in SNV richness between viral species, with sample as a random effect. *Post hoc* comparisons using the Tukey method revealed that **(-)RNA1** had significantly higher SNV richness compared to (+)RNA 2 ($P = 0.0043$). When analysing richness by time point (hpi) with a Friedman rank sum test and sample as a random effect, 12 hpi presented the most significant differences ($P = 0.0293$). A *post hoc* Tukey test identify (-)RNA1 as the viral specie with higher richness ($P < 0.0001$ from the both (+)RNA, and $P = 0.0002$ when comparing with (-)RNA), and (-)RNA2 present a significant higher number of SNV than (+)RNA2 ($P < 0.0001$) (Fig. 9).

The $\bar{AF}$ of the SNVs per time was in general below 0.1. To evaluate the weight of the SNV in the samples we use the $\sum AF$ of the SNV detected, as a proxy of the SNV abundance. We conducted a similar analysis and found that overall, (-)RNA1 present the highest proportion of SNV ($P < 0.0001$ comparing with both strands of RNA 2 and $P = 0.0004$ comparing with (+)RNA1). When performed the analysis by time point (Friedman rank sum test and sample as a random effect), we observed significant differences at 12 hpi ($P = 0.0421$) and 18 hpi ($P = 0.0421$), with the (-)RNA1 been the viral specie with more SNV weight (all comparisons $P \leq 0.001$).

That (-)RNA1, the less accumulated specie, presented the higher SNVs richness and weight was a result. In fact, intuitively we could expect that all these results could be influence by the number of transcripts, with more accumulated species been more diverse. But, when tested, after rescaling by the number of viral transcripts, the number of points with differences between species increase (Friedman rank sum test with sample as random effect). Significant differences were observed across different time pointsshocking for both richness and abundance. Specifically, for rescaled richness, significant differences were found at 8 hpi ($P = 0.0421$), 12 hpi ($P = 0.0293$), 18 hpi ($P = 0.0421$), 26 hpi ($P = 0.0421$), and 32 hpi ($P = 0.0293$). For rescaled abundance, significant differences were observed at 12 hpi ($P = 0.0293$), 18 hpi ($P = 0.0421$), 22 hpi ($P = 0.0421$), 26 hpi ($P = 0.0421$), and 32 hpi ($P = 0.0293$). In all cases, the pattern we obtained previously is maintained or even accentuated, the negative strands present a higher richness and abundance of SNVs than the positive strands.

These results suggest two ideas. (i) Could the peak of viral diversity at 18 hpi explain some sort of immune scape that allowed a second wave of viral accumulation? However, given that signs of egression were detected as early as 14 hpi, it is also possible that this peak in diversity represents variants with slower replication cycles. As faster-replicating variants complete their cycles, the slower ones may have had an opportunity to complete their replication. And (ii) is the presence of SNVs in the (-) but not in the (+) strand pointing to a within-cell constrain process? We will address this question bellow.

### 3.2.4 Within-cell evolution processes or biological constrains in the short-term evolution

The results obtained open multiple questions. If (+) strands are the most accumulated strand but the (-) strand presented a more richer composition, does it means that there is a different mutation rate per strand or selection and genetic bottlenecks are determining which genomes would be replicated from (-) to (+)? We do not pretend to fully solve this question here but shedding some light on it encourages us to continue exploring the data.

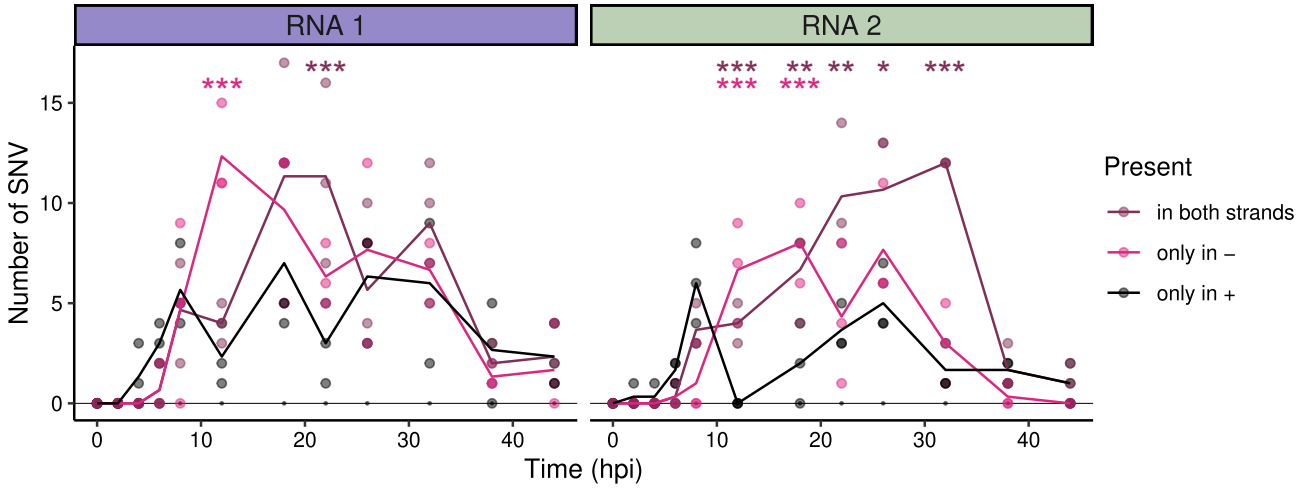



Figure 10: **Presesence of SNV strand dependent.** Caracterization of the composition of SNV in the samples depending on if they are founded in both strands (in bourgogne), or only in the megative (pink) or the positive strand (black). Asterisks indicate significance level of the pairwise Tukey test

We first analyze the number of SNV that, within a sample, were found only in the (+) strand, only in the (-) or in both strands. Fig. 10 resume the results. Just the SNV with and $AF$ ¿ 0.01 were considered. From the accumulation results 3.1.1 we suspected that OrV replicates following (at least predominantly) a stamping machine model. In order to simplify, we will assume that if a SNV appears at the (-)RNA but it is not propagated to the (+)RNA, selection or a bottleneck process has occurred. The opposite case (a SNV appearing only in the (+) strand) will be interpreted as a mutation during the (-) to (+) copying.

A LMM with sample as random effect found that in the RNA2, the SNVs were found in both strands more than in the (+) strand (estimate $= 2.2500, P < 0.001$) or in the (-) strand only (estimate $= 1.5833, P = 0203$). When analyzing by time we found that at 12 hpi, despite (+)RNA2 was the more accumulated viral specie(3.1.1), the number of SNV found only in the (+) strand where bellow founded in both strands (estimate $= -4.000, P = 0.0026$) or in the (-)RNA (estimate $= -6.667, P < 0.001$). At 18 the pattern is repeated and at 22 and 32 hpi the SNV founded at both strains is higher than at each strain by itself.

In RNA1 these differences were found only when assesing by time point. Then, 12 hpi have significantly more SNV only in the (-) than only in the (+) strand (estimate $= 10.000, P < 0.001$) and at 22 hpi there are found in both strains more than only in the (+)RNA (estimate $= 8.333, P < 0.001$). To sum up, SNVs unique to the (+) strand are no considerable. In RNA1 there are more unique SNV in the (-) strand than in the (+) at 12 hpi and in RNA 2 at 12 and 18 hpi. The SNVs that are propagated between strands (presumably while copying the (-) to the (+) strand) are more important than the specific to (+) and (-) strands from 22 hpi (in both RNAs) and on (in the RNA 2).

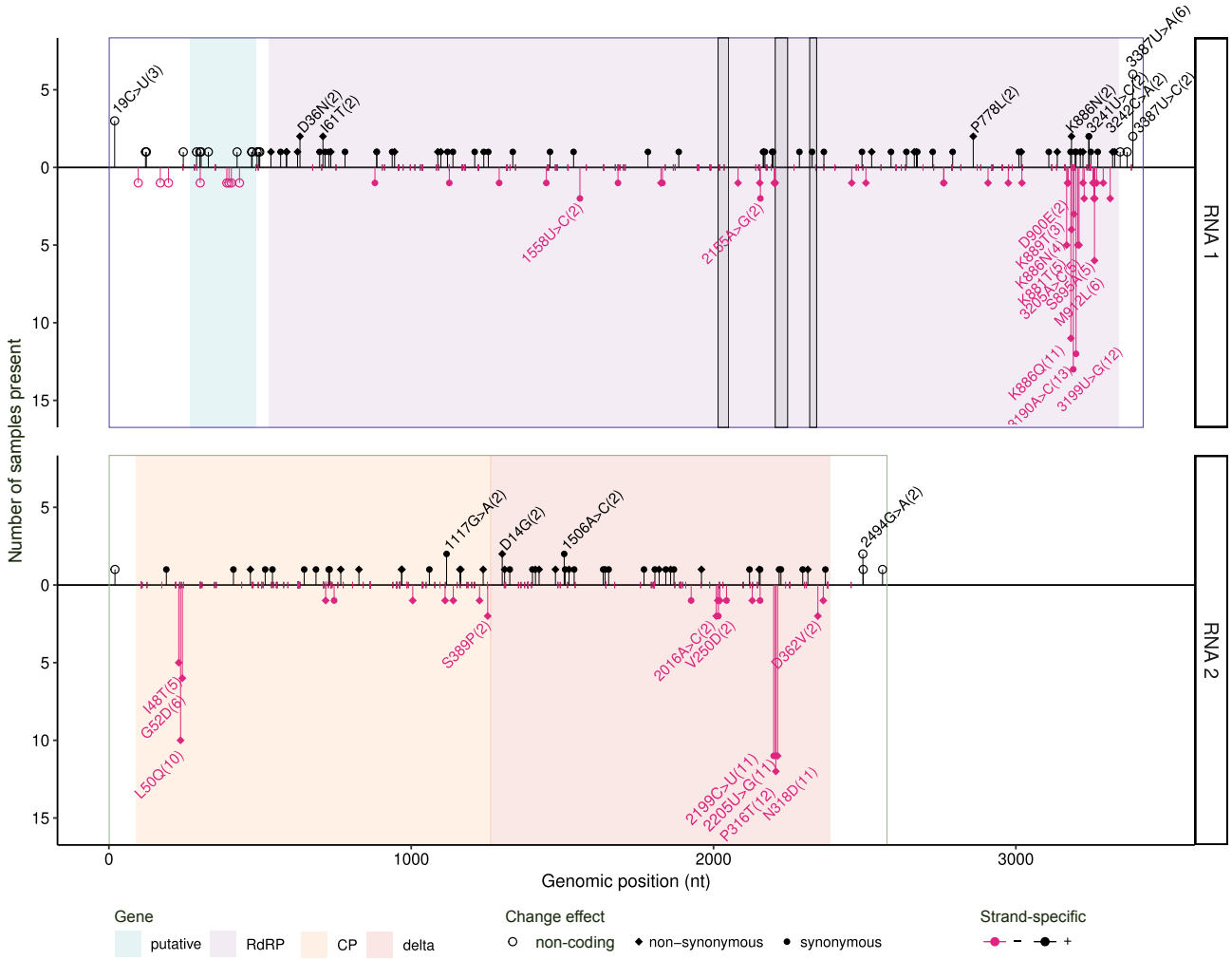



Figure 11: **Genomic regions with strand-specific SNVs**. The SNV found only in one of the strands is represented in its genomic position, with the segment height representing the number of samples where is independently found. (+)specific in black and (-)specific in pink. Background colors mark the coding regions as in other plots, the SNVs are labeled when founded in more than one sample unless the region is specially dense. The label is with the nucleotide coding if the change is in a non-coding region or is synonymous and with the aminoacid coding if is non-synonymous, even for the (-)RNA specific. Between parenthesis is indicated the number of samples where the SNV appear with an $AF > 0.01$. Shape of the points represent the change effect as specified in the legend.

This is experiment is not an strict time course, because once a time point sample is analyzed, it is destroyed by the process. The samples are completely independent of each other, and we cannot infer whether there is selection of variants by following their frequency history across the continuum of a lineage. Furthermore, we have samples less constrained to the conditions of the previous analysed point (because they are independent). We then asked if there exist SNV found exclusively in one of the strands, not necessarily in the same sample but in the whole experiment. And if they exist, were placed in specific genomic regions. The results obtained suggest that the C-terminal domain of RdRP is specially dense in this phenomenon (Fig. 11). Also, I48T, L50Q and G52D variants corresponding to the CP protein appear in five of more samples but only in the (-) strand. SNVs at positions 2199, 2205, 2206 (P316T) and 2212 (N318D), corresponding to the $\delta$ protein, were identified in at least eleven samples, but again, only in the antigenome. A difference detected using Fisher's exact test showed that the proportion of non-synonymous variants is significantly greater in the (-) strand ($P < 0.0001$, odds ratio = 3.53, 95% CI: 1.96, $\infty$). The number of samples where these SNVs appear is also greater

in the (-) strand (Wilcoxon rank sum test with continuity correction: $W = 6155.5,\ P < 0.0001$). No differences were found between the $\mu AF$ of both strand-restricted SNVs (only SNVs with $AF > 0.01$ were included in the analysis).

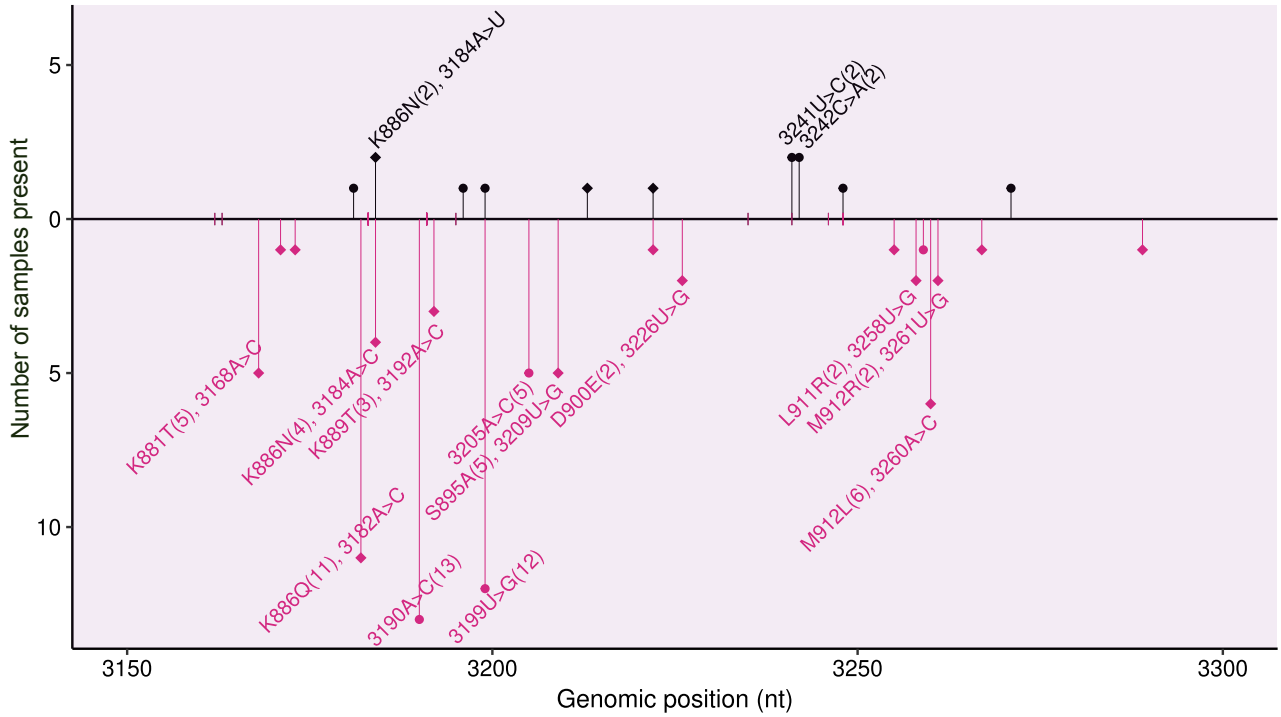
These results suggest that the presence of these SNVs in the (-) strands that did not propagate into progeny of (+) strand, is likely not due to randomness but the result of a biological constraint. If the nature of these constrains operates at the RNA level or at the protein level is far from speculation here. But what we have learn from OrV evolution experiments is that, in laboratory conditions, this RNA virus remains genetically stable in an apparent molecular evolution stasis (Castiglioni, Olmo-Uceda, Martín et al. 2024).

### 3.2.5 Two examples of the within-cell constrain phenomena

Positions 3168 to 3261 of RNA1 were identified as particularly rich in SNVs that were found in the (-) strand but not in the (+) strand (Fig. 12A). As we pointed before, this phenomenon could suggest that negative selection is operating at the within-cell scale. These SNVs accumulate in the C-terminal region, previously predicted as disordered ( [21]), and the AlphaFold 3 model of the OrV RdRP situate them in a terminal tail (Fig. 12B). Of the SNVs that appear in more than one sample, nine would produce a non-synonymous change and three a synonymous mutation (3190A>C (codon 888), 3199U>G (codon 891) and 3205A>C (codon 893)). Four of the codons affected coded for a Lysine, a positive charged residue. Codon 886 is of particular interest, as positions 3182, 3183, and 3184 all undergo mutations. However, only the 3183A>C mutation (K886T) is found in both strands. The 3182A>C (K886Q) and 3184A>C (K886N) mutations appear in eleven and four independent samples, respectively, but only on the (-) strand. In contrast, the 3184A>U mutation, which would also result in K886N, is found in two samples on the (+) strand. This finding suggests that some constraints may be acting more strongly at the RNA level.

RdRP shares a 48 % of aminoacid identity with Santeuil virus, a 47 % with Melnik virus and a 38 % with Le Blanc virus. Despite the C-terminal is not specially conserved between the nodaviruses ( [21]), the prediction of a similar tail in Santeuil virus (879-982 residues) and Le Blanc virus RdRPs (873-942 residues) and the presence of tandems of basic residues in them (K and R in OrV and Santeuil and R in Le Blanc virus) could be associated with protein-RNA interaction functions.

A. Zoom in the RdRP region rich in strand specific SNVs




B. RdRP with the positions of the SNV found only in the (-) strand

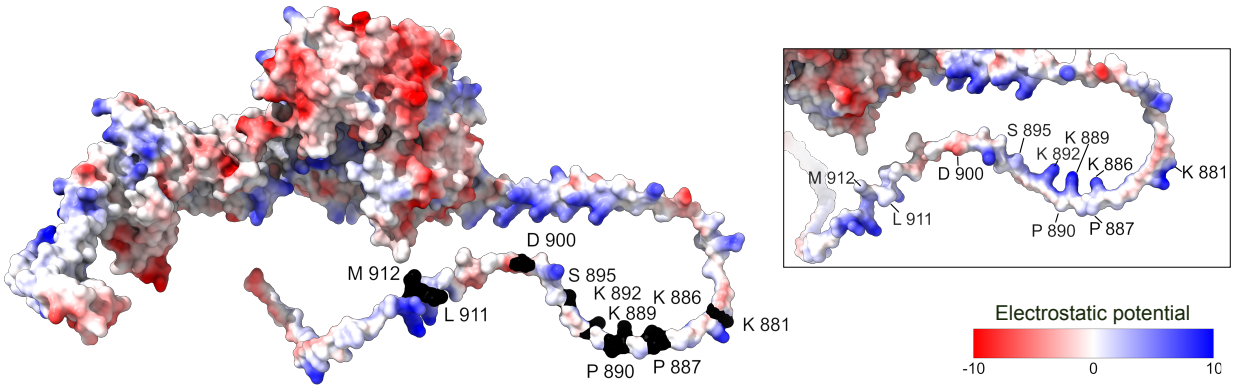



Figure 12: **RdRP**.

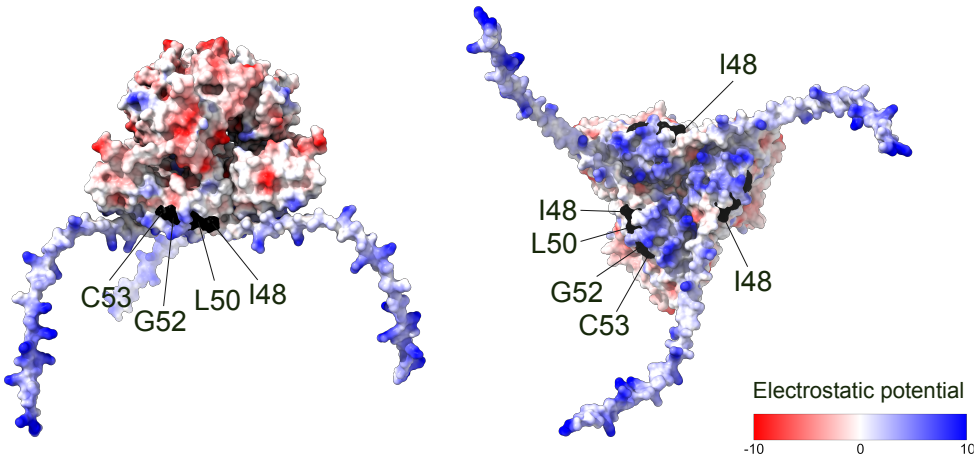



Figure 13: **Capsid protein**.

The CP, but specifically its S domain (44-214) is the protein more conserved between OrV, Le Blanc (79%) and Santeuil (62%) viruses [29]. Residues I48, L50 and G52 are all non-polar but they are near from R41 and K51, two basic residues that could be involved in RNA packaging [29]. The change to a polar (T and Q) and to an acidic (D) residue in a part of the structure (exposed at the capsid interior) where basic residues predominate, could disrupte the RNA-CP interactions(Fig.13). Other posible hypothesis of the nature of these constrain is that these residues could be implicated in the assembly of the capsomers into virions (they are on the inner-lateral part of trimers) (Fig.13). Furthermore, a Fixed Effect Likelihood (FEL) analysis estimates that the $53^{rd}$ codon is under purifying selection ($\alpha = 45.727, \beta = 0, P = 0.0740$) when analyzing the two known strains of OrV (results below).

We cannot dismiss the possibility that these frequent mutations, restricted to the (-) strand, may be operative and playing a role at the RNA level. Apart of that, these results suggest some examples where purifying selection could be operating at the within-cell level. If these constrain operates at the RNA level, the protein level or both should require further investigation.

#### 3.2.6 19C>U SNV is detected exclusively in the (+) strand and would reduce the minimum free energy of the 5'UTR RNA1 secondary structure

Following up with the above analyses, there were no significant differences in the $\bar{AF}$ of the strand-specific SNVs, but an outlier from the (+) strand-only group captured our attemption. RNA 1 19C>U, appear in at least three samples with a $\bar{AF} > 0.5$.

A. Wild type conformation

B. With the SNV RNA1: 19C>U

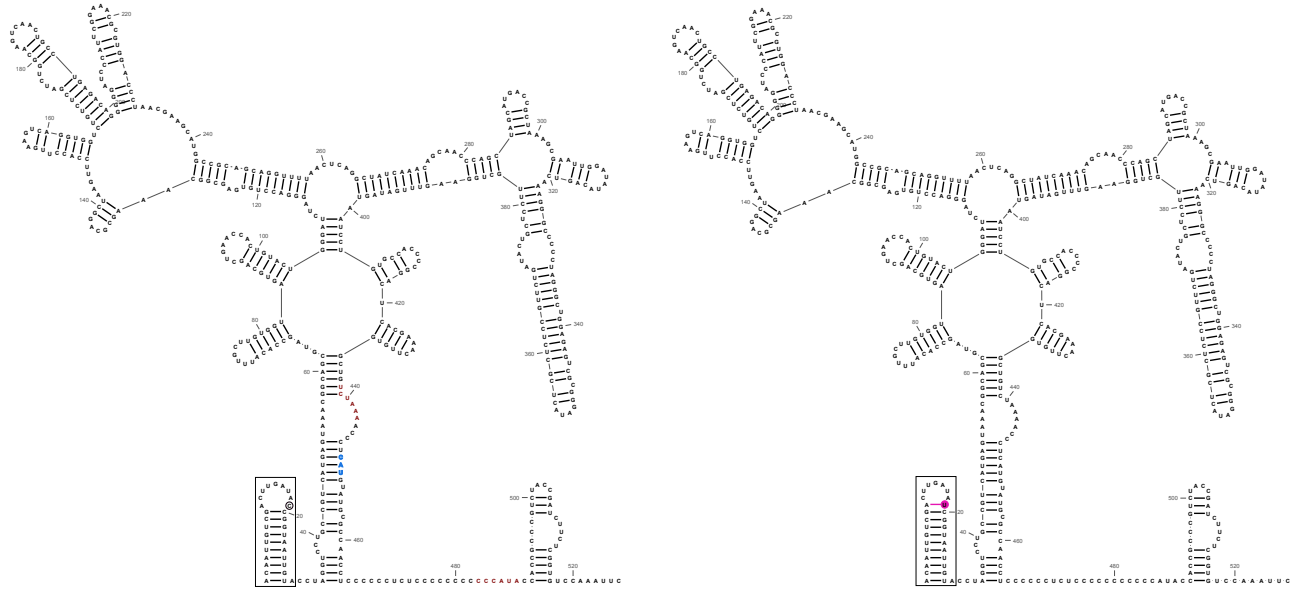


Figure 14: **5' UTR of RNA 1 (1:526) secondary structure prediction**. Secondary structure of the (A) wild-type and (B) 19C>U SNV predicted conformations. The change of a C (in black in A) for a U (in pink in B) at the 19 position would facilitate a longer stem-loop structure.

This variant is located in the 5' UTR region of RNA1. The RNA secondary structure prediction [39], both with and without the variant, showed a reduction in the minimum free energy across all window sizes investigated: 1:50 (-13.29 kcal/mol → -15.19 kcal/mol), 1:100 (-39.87 kcal/mol → -41.77 kcal/mol), 1:150 (-73.97 kcal/mol → -75.87 kcal/mol), 1:270 (-139.82 kcal/mol → -141.72 kcal/mol), complete 5' UTR (-230.66 kcal/mol → -232.56 kcal/mol). The reduction was associated with an increase in one more base-pairing and the reduction of the loop size of the first stem-loop structure (Fig. 14). The implications of this change in the stem-loop configuration, should be further investigated.

### 3.2.7 A variety of low frequency variants appears in a convergent manner during the infection process

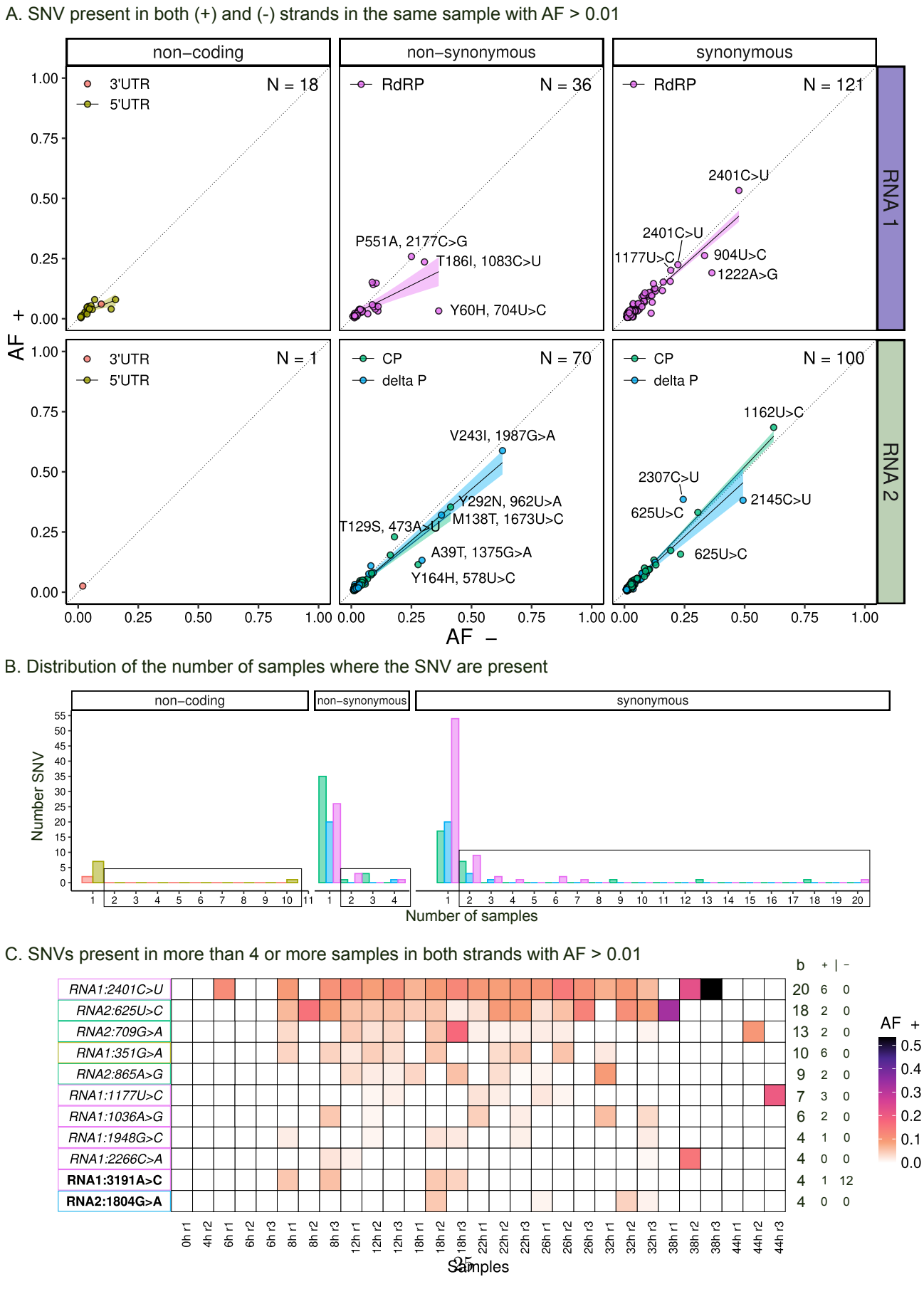



Figure 15: **SNVs found in both strands with** $AF > 0.01$. (A) $AF$ correlation in both strands. $N$ indicates de total number of SNV per RNA and amino acid effect. (B) Distribution of the number of samples where the SNV is present. Colors indicate proteins or UTR regions as in A). (C) Heatmap of the SNV founded in both strands in four or more samples.

In order to continue investigating the viral variants, we focus now on the SNVs found in both strands of the same sample. These SNV show a strong linear correlation between the $AF$ of both strands (Pearson's moment correlation: $r = 0.9292$, 95% CI $[0.9133, 0.9423]$, $t = 46.645$, $df = 344$, $P < 0.0001$) (Fig. 15A).

A Fisher's exact test found a statistically significant difference in the frequency of non-synonymous variants between the two RNA types, with RNA1 having a significantly lower proportion of non-synonymous variants compared to RNA2 ($P = 0.0003$, odds ratio = 0.43) (Fig. 15A).

Next, we asked if the SNV were spurious or they appear in a more persistent manner. The 21.1 % of all the SNVs found in both strands appear in more than one sample. When separating by protein and amino acid effect, we obtain the following percentages (non-synonymous — synonymous): RdRP (13.3 %— 22.9%), CP (10.3% — 37%) and $\delta$ (4.76% — 16.7%) (Fig. 15B).

Fig. 15C shows the $AF$ of all the SNVs founded in more than three samples in both strands. Only two of the selected are non-synonymous mutations(bolded in 15C). RNA1:3191A>C correspond to the RdRP K889Q and RNA2:1804G>A to $\delta$ V182I. K889Q (RNA1:3191A>C) is the unique present in both strands or exclusively in the (-) strand, where it appears in 12 different samples. K889 was previously identified as a residue with a potential biological constraint (Fig. 11). Here, the mutation occurs at position 3191 instead of 3192, but the pattern seems to confirm the previous results.

RNA1:2401C>U is a synonymous SNV that appear in twenty samples at both strands (and in six more but only in the (+) strand). Its maximum $AF$ (0.53 in the (+) and 0.48 in the (-) strand) is reached at a latter 38 hpi sample (38 hpi, rep. 3). In other 38 hpi sample represents the 22% of their population at that locus. Although this is not a lineages experiment where we could track its dynamics, the persistent presence of the SNV from the early stages of the infection and the significant $AF$ observed in some of the samples (Fig. 16A) prompted us to take a closer look at it. OrV has only four annotated proteins but its genome presents smaller ORFs that could be generating small peptides or playing regulatory roles.

Small Open Reading Frames (smORFs) have been documented in viruses, but traditionally, the primary focus has been on the major ORFs [30]. With the development of ribosomal profiling (RiboSeq), the translation of many of these smORFs has now been confirmed, and they are receiving growing attention [4,14]. We have identified a smORF in RNA1 coded between 2191 and 2544 but the transition C $\rightarrow$ U in 2401 position would introduce an OPAL codon (**U**GA), truncating the putative native peptide (118 amino acids) into a 70 amino acids one. A structural comparison with DALI server of the predicted 3D conformations of the complete and the truncated smORF did not found any similarities in the full PDB (Protein Data Base). Furthermore, C $\rightarrow$ U is the third most frequent mutation of the SNVs with an $AF > 0.01$ (17 % of changes are C $\rightarrow$ U, only bellow U $\rightarrow$ C, 23% and G $\rightarrow$ A, 20%) (Fig. 16B). Only in the RNA 1 exist 218 possible codons were a transition C $\rightarrow$ U would convert a sense codon into an stop codon so the significance of this association should be better characterized.

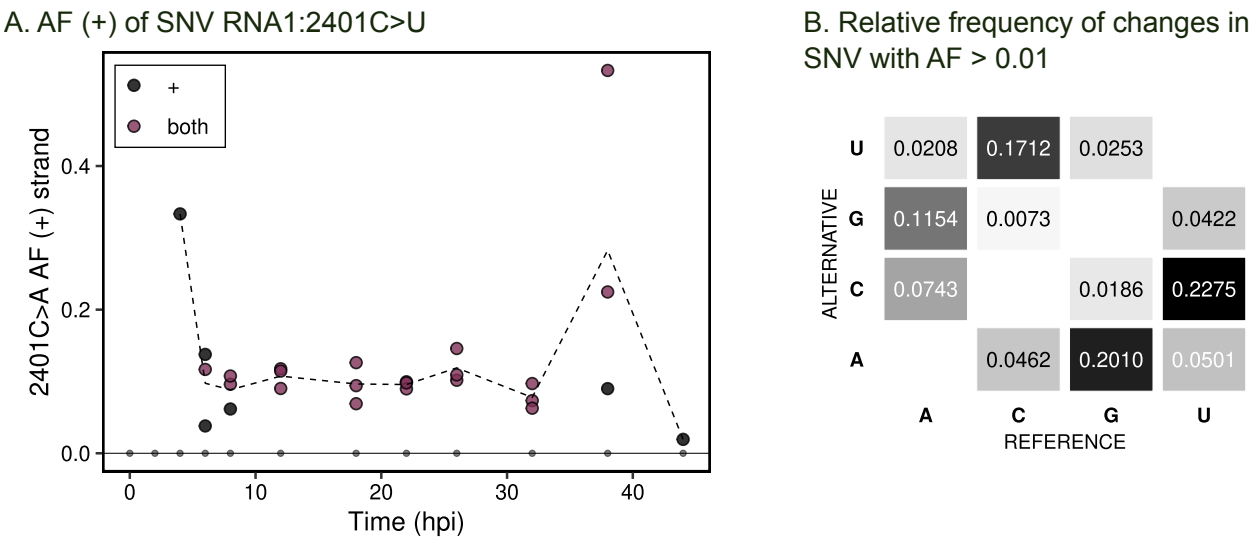



Figure 16: **AF of SNV RNA1:2401C>U and relative frequencies of nucleotide changes in the SNV with** $AF > 0.01$. (A) SNV RNA1:2401C¿U was found independently in nine of the twelve time points. From 12 hpi to 32 hpi in both strands (bourgogne points) or in the (+)RNA (black points). The dotted line represents the $\bar{AF}$. (B) Relative frequencies of the nucleotide changes (REFERENCE → ALTERNATIVE) in the SNVs with $AF > 0.01$. White to black background from less to more frequent.

### 3.2.8 A sliding window of $K_A/K_S$ values revealed silent and non-silent mutational hotspots

Next, we follow a more conservative approach by studying the effect of the mutations at the codon level. There were no fixed variants in any of the samples. This result was concordant with the apparent molecular stasis previously described in a 10-passages evolution experiment [8]. We then calculate the number of non-synonymous ($K_A$) and synonymous mutations ($K_S$) in sliding windows of four amino acids with one or two amino acids step. Because this analyses was focus on the codon effect, we used the SNV found at least in the (+)RNA with an $AF > 0.01$.

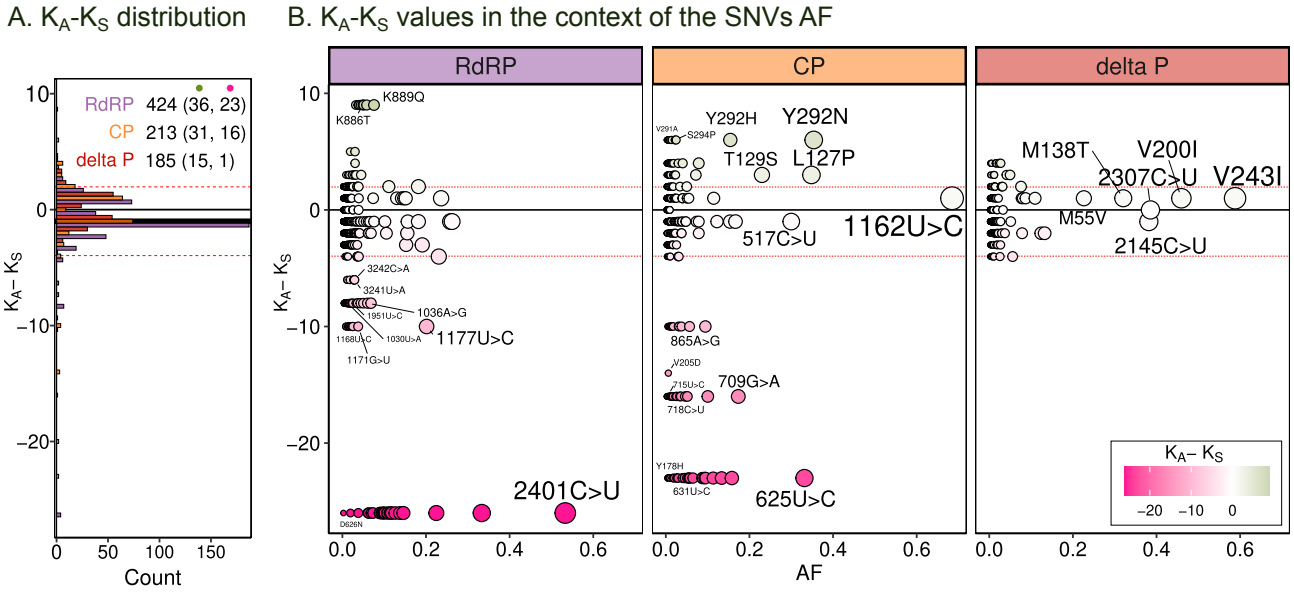



Figure 17

Results show significant differences between the $K_A - K_S$ estimated for RdRP, CP and $\delta$ (Fig. 18; Kruskal-Wallis rank sum test: $H = 15.743$, $df = 2$, $P = 0.0004$). A pairwise Dunn test with Benjamini-Hochberg (BH) correction indicate CP and $\delta$ had less negative $K_A - K_S$ values than RdRP (CP *vs.* RdRP: $z = 3.6059$, $P = 0.0005$; $\delta$ *vs.* RdRP: $z = 2.7184$, $P = 0.0049$). In Fig. 17B, is represented the relation of each SNV considered with their $K_A - K_S$ and AF.

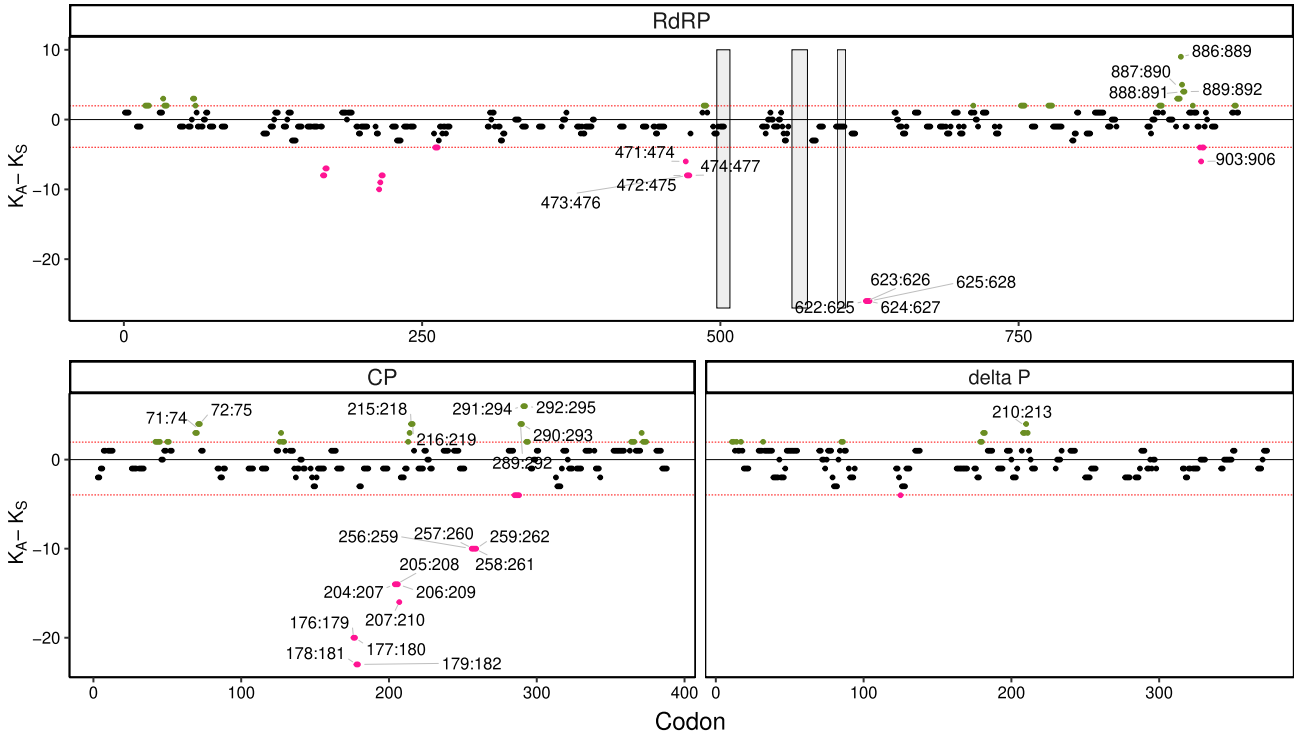



Figure 18: $K_A - K_S$ **values computed using four amino acid-sliding window and one amino acid-step**. All the SNVs present at least in the (+)RNA with an $AF > 0.01$ were used for the calculation. $K_A$ and $K_S$ correspond to the number of non-synonymous and synonymous mutations respectively in the 4-amino acids window. Red dotted lines indicate ±2 median absolute deviations (MAD) of values for all three proteins. Residues with a difference ±3 MAD were labeled. As in other plots, the localization of the conserved subdomains A, B and C of the palmprint is represented as grey rectangles.

In the RdRP, the sites with the lowest $K_A - K_S$, *i.e.*, where the synonymous SNVs overcome the non-synonymous, were (residues in genomic order): 167-173, 214-220, 261-266, 471-477, 622-628 and 902-908. While regions with the highest positive $K_A - K_S$ values where in: 18-24, 33-39, 58-63, 486-491, 712-715, 752-758, 775-781,868-873, 883-892, 896-899, and 931-935. Because the values were not normally distributed, the threshold for selecting residues that significantly deviated from the median value was set to the median $K_A - K_S \pm 2 median absolute deviations (MAD)$.

In the CP, the most negative values of $K_A - K_S$ accumulated in regions (Fig. 18): 176-182, 204-210, 256-262 and 285-291. While the most positive ones were in regions: 42-48, 50-54, 69-75, 126-132, 213-219, 289-297, 364-369 and 371-377. Of special interest is the window that includes the Y292N variant, the non-synonymous mutation with the highest $AF$ (Fig. 17B).

$\delta$ had the less number of residues with a $K_A - K_S \pm 2MAD$. Additionally, the $AF$ of the SNVs in the selected positions had an $AF < 0.1$ (Fig. 17). Only 125-128 residues window where below the $K_A - K_S$ threshold while five regions were found with a $K_A - K_S > 2MAD$: 11-20, 32-35, 85-89, 179-185 and 208-214.

$K_A - K_S$-window counting is not a robust metric for identifying selection, but in the context of this experiment, we had to discard other methods based on tracking SNV $AF$'s. Analyzing per-sample consensus sequences was also discarded because any of the samples presented a fixed mutation. Only three samples had a coding variant above the frequency of the ancestral. Synonymous mutation 2401C>U (RNA1) appear in a 38 hpi sample at 53% but was found in multiple samples at lower frequency 3.2.7. In this analysis, the RdRP $625^{th}$ codon (where position 2401 is part) obtained the lowest $K_A - K_S$. In addition, synonymous mutation 1162U>C (RNA2) was found at the 62% of the $AF$ (being the highest $AF$ reached by a SNV in the entire experiment) but only in one sample. No more variants were found at the same position at any sample and the $K_A - K_S$ value for its amino acid window was 1. In this case, probably drift and no selection was the responsible of the SNV overcomming the wild-type frequency.

The results of the $K_A - K_S$ window strategy employed here are biased by how frequently a mutation is detected in a given position. If the mutation is the same or different is not differentiated so highly recurrent mutations may disproportionately affect the estimation of $K_A - K_S$, potentially leading to an over- or underestimation of selective pressures at specific sites. But this can be also its strength, allowing us to tease apart spurious from more recurrent events. Nevertheless, this approach can still be a valuable method for identifying mutational hotspots.

### 3.2.9 Signatures of selection at different time-scales of evolution

As a (+)ssRNA virus, the mutation rate of OrV should be within the ballpark of typical values estimated for most viruses ($10^{-6} - 10^{-3}$ substitutions per site and replication round, [52]). Considering the known mutational rates, for a genome of 3421 + 2574 bases we should expect from 0.03 to 30 mutations per copying event of RNA1 and from 0.02 to 22 mutations per copy of RNA2. We have not estimated here the rate of mutation for OrV but the mutation frequency found in this work suggest OrV mutation rate should be within the low range of estimated for a ssRNA viruses. More investigation should be done to clarify this aspect.

What we know is that OrV shows a remarkable genetic stability within the short-term, with no fixed variants at the end of the infectious cycle. Only four variants accumulate above the 50% in the population to considerate that they are competing with the ancestral but these events are punctually found in a few samples and drift and non natural selection could be the responsible for their increase. This scenario is coherent with other describe acute infections [47].

Moreover, we found mutational hotspots were synonymous or non-synonymous variants were recurrently emerging. Although these variants were not dominating in frequency the viral cloud, the fact that they exist in the genomic pool, make them available for deterministic and stochastic processes at longer time scales.

Here, we asked whether signatures of selection can be found at the different time scales (Fig.19).

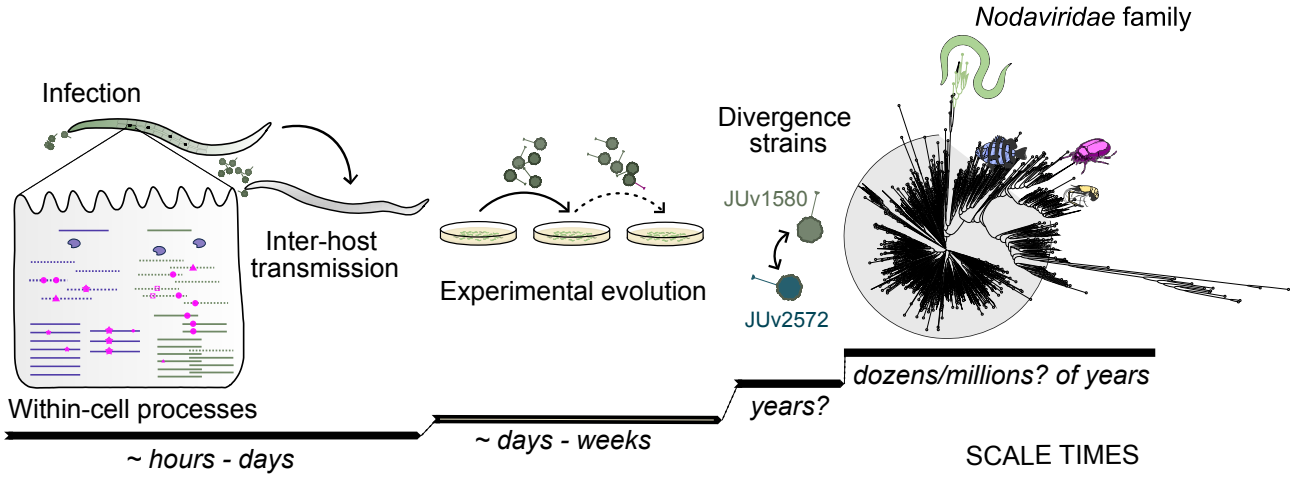



Figure 19: **Different time scales.**.

Results obtain in this short-term infection process (the complete process analyzed here only took 44 hpi) revealed an asymmetry in the SNV presence in both viral strands. Under the SMR model we considered for OrV, if a SNV is found in the (-)RNA but not in the (+)RNA, the simplest hypothesis is that the change in the SNV has not allowed the RdRP engage or copied, not been selected in the very short term. Other possibilities included that the SNV in the (-)RNAs could be playing a role, not as intermediate template of the next generation of (+)RNAs, but at the RNA level. We have highlighted the positions that under the above hypothesis presented a constrain 3.2.4.

We have analyzed the variants that have been identified in both strands, *i.e.*, the variants that have complete the replication cycle and could be potentially producing proteins for the viral pool. From these set, we identified hotspots where synonymous or non-synonymous mutations predominate. But always interpreted under the general result that the viral quasispecies is dominated by a master sequence that is exactly the OrV reference.

One hypothesis is that the predominant pressures that shape the viral populations analyzed, happened only at the intra-host level. Other hypothesis is that inter-host pressures can additionally be acting. The egression of viral particles to the lumen was detected from 14 hpi [9] so samples after that time point could be infected with subsequent viral generations. The 70 % of animals was infected at 8 hpi [9], so the hypothesis that the second generation of viruses have infected the 30% of animals that remained uninfected at the first peak, cannot be fully ruled out. Additionally, the detection of a second lower peak of the (-)RNAs accumulation at 22 hpi is also concordant with two infection cycles, where the second wave was produced for the descendants of the initial viral stock. But independently of at which level can be selection acting, there exist a predominance of sites were the changes accumulated are synonymous over non-synonymous (in 3.2.8).

By expanding the scale time, we used the experimental evolution of the virus in a 10-passages design (with 6 days before extraction of viral supernatant). The evolution finished with only two or three mutations fixed per lineage, and only one per lineage was non-synonymous: RdRP Q41R in one lineage and CP Y292H in the other [8]. The DoS statistic [58], was estimated for OrV proteins. $D_n$ and $D_s$ were extracted from the substitutions in the evolved JUv1580_vlc lineages while $P_n$ and $P_s$ refers to the polymorphisms founded in the short-term experiment at least in (+)RNAs with an $AF > 0.01$ (Table 4). Adding the substitutions of the evolved JUv2572_vlc lineages to the computation changes slightly the DoS values but the same pattern is observed ($DoS_{RdRP} = 0.2265$; $DoS_{CP} = 0.4044$ and $DoS_{delta} = -0.4304$).

| Protein | $D_n$ | $D_s$ | $P_n$ | $P_s$ | $DoS$ |
|---|---|---|---|---|---|
| RdRP | 1 | 1 | 61 | 162 | 0.2265 |
| CP | 1 | 0 | 47 | 89 | 0.6544 |
| $\delta$ | 0 | 1 | 34 | 45 | -0.4304 |

Table 4: **Direction of selection in OrV JUv1580 proteins**. Columns indicate: $D_n$ = number of non silent substitution, $D_s$ = number of silent substitution, $P_n$ = Number of non-synonymous polymorphisms ($AF$ in (+) $> 0.01$), $P_s$ = Number of synonymous polymorphisms ($AF$ in (+) $> 0.01$).

Additionally, an analysis of the SNVs detected at the JUv1580_vlc evolved lineages ($AF > 0.01$) with the same $K_A - K_S$ window strategy was performed (Fig. 20, midterm evolution, MT).

In the evolution experiment, inter-host selection pressures added to the intra-host selection pressures present in the short-term experiment. Factors such as competition for encapsidation, egression efficiency, transmissibility, and resistance to environmental stressors in the host growth medium are shaping the evolved viral population. As a result, some viral variants that reached high frequencies in the cellular cytoplasm (and could be present in the short-term experiments) may not be able to overcome subsequent barriers. By comparing mutational hotspots of both short-term and the evolution experiment (for interest we call it medium-term) we were asking for the genomic spots that were more permissible to maintain the mutations at both scale times. And by focusing on the shared hotspots the aim was to identify the regions with more potential to change.

In RdRP, the amino acid windows between 883 - 892, 896 - 899 and 902 - 908 (disordered tail), CP 289 - 295 (protrusions from P domain) and $\delta$ 14 - 19 residues were mutation hotspots in both scale times. In the RdRP disorder tail, we also found an accumulation of SNVs that appearing in the (-)RNAs were not found in the (+)RNAs (see 3.2.4) what make this region interesting for more exploration. But before that, we wanted to go further into the scale time by considering the other viral strain.

OrV has two strains with public available references. JUv1580_vlc, the strain used in this work and JUv2572_vlc. The differences in their infection phenotypes were studied [8, 21, 27], as well as their experimental evolution in laboratory conditions and the characterization of their genomic differences [8]. In the evolution experiment, two lineages of JUv2572_vlc and other two of JUv2572_vlc strain were evolved. We use here the four evolved sequences (JUv1580_vlc RL1_evolved, JUv1580_vlc RL2_evolved, JUv2572_vlc 10_5_evolved and JUv2572_vlc RL3_evolved), together with their ancestral sequences (JUv1580_vlc and JUv2572_vlc) as well as the two other JUv1580 assemblies publicly available (NC028097.1 and HM030970.2). A Tajima's neutrality test was performed for the three coding sequences on the first codon position, suggesting that any of the viral genes have evolved neutrally (RdRP $D = 1.970$, CP $D = 0.662$ and $\delta$ $D = 1.449$)(Table 5) suggest that any of the viral genes have evolved neutrally. A codon-based test of neutrality between all the sequences serve us to confirm that both strains have evolved under purifying selection. With values of $d_N - d_S$ being compatible with negative selection. The values of $d_N - d_S$ for the ancestral strains (JUv1580_vlc and JUv2572_vlc) were for the RdRP: -7.4751 ($P < 0.0001$), for the CP: -3.7960 ($P = 0.0002$) and for $\delta$: -3.6534 ($P = 0.0004$).

A recursive idea has been present in this work, OrV shows small changes in the short and in the medium-term evolution. If the mutation rate is within the expected range for a (+)ssRNA virus, this genetic stability could be explain, at least in part, by a strong purifying selection. Results reported here confirm this observation.

| Protein | $m$ | $S$ | $p_s$ | $\Theta$ | $\Pi$ | $D$ |
|---|---|---|---|---|---|---|
| RdRP | 8 | 13 | 0.0139 | 0.0054 | 0.0074 | 1.9699 |
| CP | 8 | 6 | 0.0153 | 0.0059 | 0.0067 | 0.6624 |
| $\delta$ | 8 | 2 | 0.0058 | 0.0022 | 0.0031 | 1.4488 |

Table 5: **Tajima's $D$ neutrality test for OrV proteins of Juv1580 and JUv2572**. Columns indicate: $m$ =number of sequences, $n$ =total number of sites, $S$ =Number of segregating sites, $p_s = S/n$, $\Theta = p_s/a_1$, $\pi$ =nucleotide diversity, and $D$ is the Tajima test statistic.

Next, using the same set of sequences we performed a fixed effects likelihood (FEL) test to identify the codons under selection (Supplementary Table 1). In Fig.20 are represented together the residues affected with constrains or with selection signatures at the different scale times.

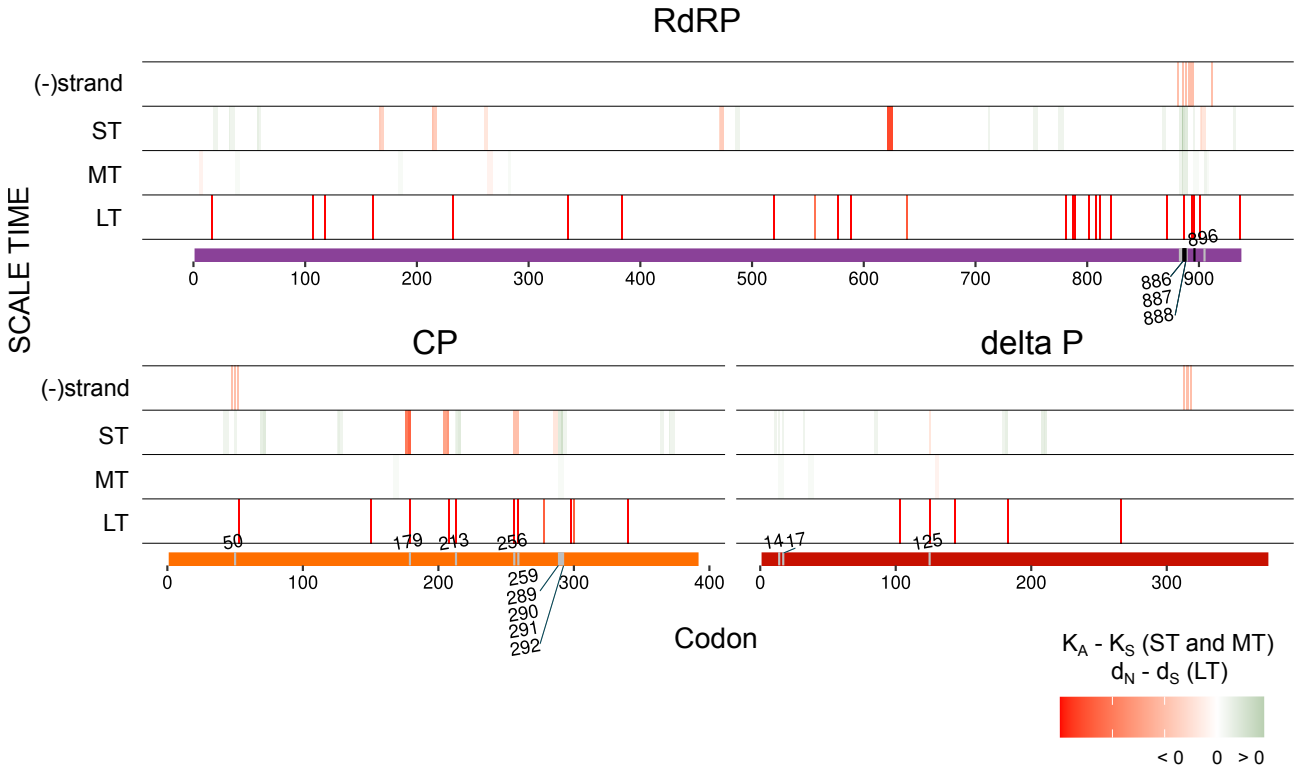
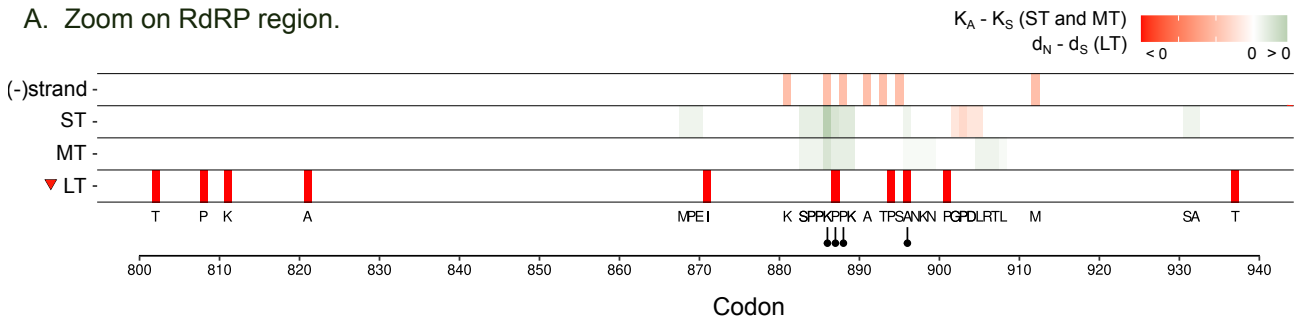



Figure 20: **Signatures of selection at different time scales**. Heatmap with the selected values of $K_A - K_S$ or $d_N - d_S$ at each scale time. Time scales are: (-) strand (the shorter time), for the residues mutated only in the (-) strand but not propagated to the (+) strand; short-term (ST) for this experiment and medium-term (MT) for the experimental evolution in [8]. In both ST and MT , colors represent the values of the $K_A - K_S$ window. In the long-term (LT) the $d_N - d_S$ estimations ($P < 0.1$) from FEL test for the set of JUv1580 and JUv2572 ancestrals and evolved sequences. On the corresponding gene segment are marked the residues with two (in gray) or three (in black) scale times are affected.

The C-terminal of the RdRP was revealed as specially dense in these events, been the unique region highlighted in the four time scales. The direction of the constrain was also interesting, while a predominance of non-synonymous mutations seems to rule the short-term and the midterm of some codons, in the long-term all the residues are under negative selection. Fig. 21A shows a zoom of the region with the residues involved. In the case of the short and medium time scales (the infection and the experimental evolution experiments) there is colored the first position of the 4-amino acids window used for the computation. In Fig. 21B, the codons under negative selection are shown on the

OrV RdRP structure prediction.




B. OrV RdRP prediction colored by their electrostatic potential.

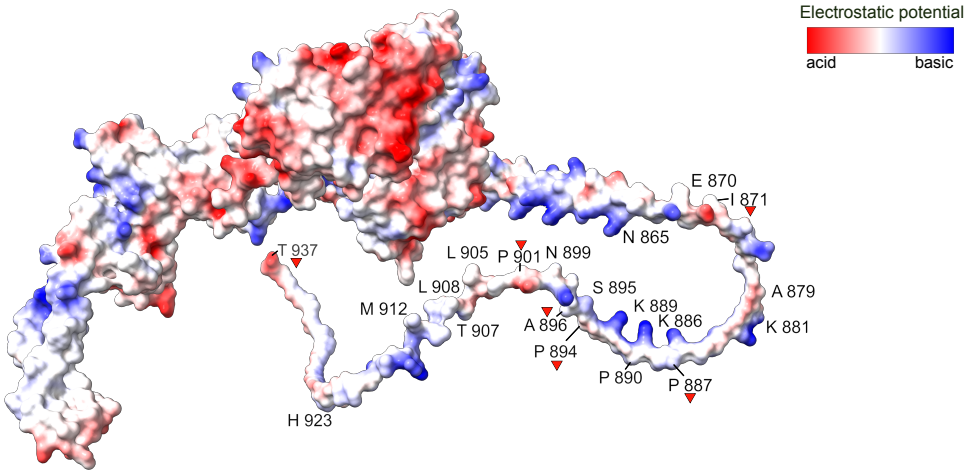



Figure 21: **OrV RdRP.**. (A) Zoom on the RdRP were selected residues are accumulated. (B) OrV RdRP with some of the selected residues. Residues under negative selection (by FEL) are indicated as red triangles.

We decided to further investigate this region.

### 3.2.10 RdRP tail as a target of selection at different time-scales

Results in this work pointed the N-terminal region of the OrV RdRP several times. In this part of the replicase, we found an accumulation of mutations that appear recurrently only in the (-)RNAs. The same region is a hotspot of non-synonymous (886-892) and synonymous mutations (residues 903-906) in the short-term and in the evolution experiment. Finally, when analyzing the long-term evolutionary patterns of the ancestral and evolved sequences of the two OrV strains, the final part of the RdRP was particularly enriched with codons under negative selection.

The N-terminal domain of OrV forms a disordered tail with a predomination of positive charged residues ( 21A). At the sequence level, this region is one of the less conserved between the nematode noda-like viruses [21], but results here suggest that some structural characteristics are conserved.

This disordered tail is not only found in OrV but also in the other know nematode-infecting noda-like viruses. In order to compare with a more distant virus we have chosen the flock house virus (FHV). FHV belongs to the *Alphanodavirus* genus and present some differences with the nematode

virus clade. It is able to infect a wide range of hosts, including different genera of insects. Structural and disordered degree prediction allow us to identify a similar disordered tail in FHV RdRP (Fig. 22A-B). A 2-sample test for equality of proportions with continuity correction, found the proportion of positive residues significantly higher in the last part of the RdRP compared with the complete protein in OrV ($\chi^2 = 6.2467,\ P = 0.0062$) and in FHV ($\chi^2 = 3.2731,\ P = 0.0352$) but not in the rest of the viruses (Fig. 22B). Despite that, a visualization of the surface electrostatic charges showed the region was dominated by basic residues at least in small windows (Fig. 22A). In the FHV RdRP tail, a significant reduction of acidic residues was also found ($\chi^2 = 4.1393,\ P = 0.04190$)(Fig. 22B). The compositional bias of the disordered region was then systematically analyzed using the same strategy. The equality of each residues proportion in the disordered and in the complete protein was assessed using a directional alternative test. All the viruses showed a significant enrichment in P in the disordered tail (Fig. 22C). Except Santeuil virus, they were also enriched in one basic residue, OrV in K and Le Blanc virus and FHV in R (Fig.C).

As far as we know, the structure neither the function of any nodaviruses RdRP disordered tail has been previously described. FHV RdRP has been deeply study as a model of non-enveloped (+)ssRNA virus. The virus induces replication organelles in the mitochondrial outer membrane via formation of a 12-mer ring crown [**?**, 13, 44, 60]. The cryo-EM reconstruction of the replication crown did not include the residues before M861 but following the orientation of the existing model, the tail would be facing the cytosolic side. Because the predominance of basic residues it could be implicated in the RNA-protein interactions needed for example, for recruiting the RNA templates, a function proposed in [60]. If OrV uses replication organelles as other (+)ssRNA viruses, and more specifically, through the formation of replication crowns as FHV, CHIKV or SARS-CoV-2 is still unknown. What it has been described is the transcriptionally impact that the OrV infection has in genes related with the mitochondria organelle [9]. This and the structural conservation should be took into account for further investigation.

Moreover, all viruses showed an enrichment in P at their disordered tails. It has been described that viruses with short P-rich motifs are able to hijack the host's endosomal sorting complexes required for transport (ESCRT) pathway and using for its within-cell transportation and egression [61]. This specially applies for enveloped viruses but non-enveloped viruses as hepatitis A virus, are also able to recruit the ESCRT machinery [50]. OrV infection has a huge impact in the lipid metabolism of the animals [6, 9, 26] and the intracellular localization of the viral RNAs and the CP proteins during the last part of the cycle, suggest that particles are concentrated inside vesicle-like structures [20] so the ability to manipulate lipid trafficking would be highly convenient. In other example, a P-rich regions in the N-terminal unstructured region of NS3 of Dengue virus (same phylum than nodaviruses) was identified as crucial for the formation of infective particles. In the same work, authors speculate that NS3 could be, in addition to its known functions as protease and helicase, a platform required for the interactions with viral and/or host factors needed for the production of the viral particles [23]. Exploring more this interaction idea, P is the unique planar amino acid and their presence in an exposed structure would be an easy target for the host proline-recognition machinery. P-rich binding domains are an important group of modular recognition domains in animals. Most interesting, P-rich regions are the most common motifs in *Droshophila* and the second more common in *C. elegans*. This recognition domains are critical for the assembly of several intracellular signaling pathways [65].

Finally, it is important to note that in FHV, the disordered tail overlaps with the sgmRNA3. FHV sgmRNA codes for B1, a 102 peptide of unknown function and for B2, a 106 peptide that is a RNA silencing suppressor. Other examples of overlapping or nested genes where one of them is a host silencing suppressor, are found in plant virus [36,41]. Furthermore, genes that overlap are significantly enriched in disordered regions [36, 41]. But the presence of a sgRNA has not been described in the

nematode virus clade. A wider investigation of the *Nodaviridae* family and the appearing of the RNA silencing suppressor could help us to understand the nature of some of the evolutionary constrains the RdRP tail is facing.

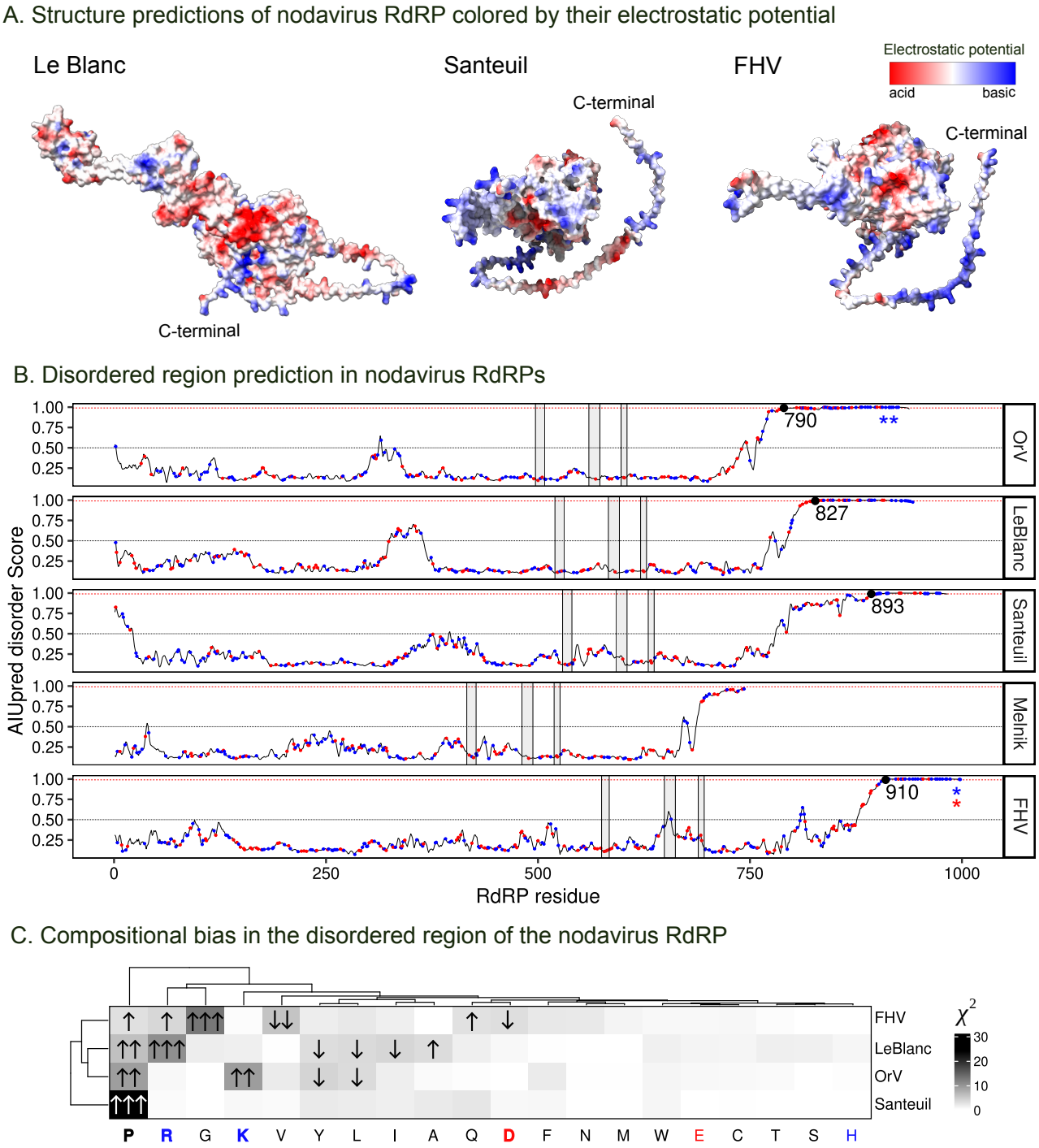



Figure 22: **Nodavirus RdRP disordered coil.**. (A) Structural predictions of the nodaviruses RdRP colored by their electrostatic potential. (B) Disordered region prediction in nodaviruses RdRP. (C) Viral species accumulation as proportion. (D) Compositional bias in the disordered region of the nodavirus RdRP.

### 3.2.11 Deletions in RNA 2 are the predominant DVG type

As part as the viral diversity study, we look for the DVGs present in the samples. For that purpose, our metasearch tool DVGfinder was used [46]. For this analysis only the defectives with at least five counts were took into account. Here, the information about the strand was ignored so instead of using the breakpoint and re-initiation site terminology we will use start and end when referring to the joining point closer to the 5' and to the 3' respectively.

#### DVG Richness

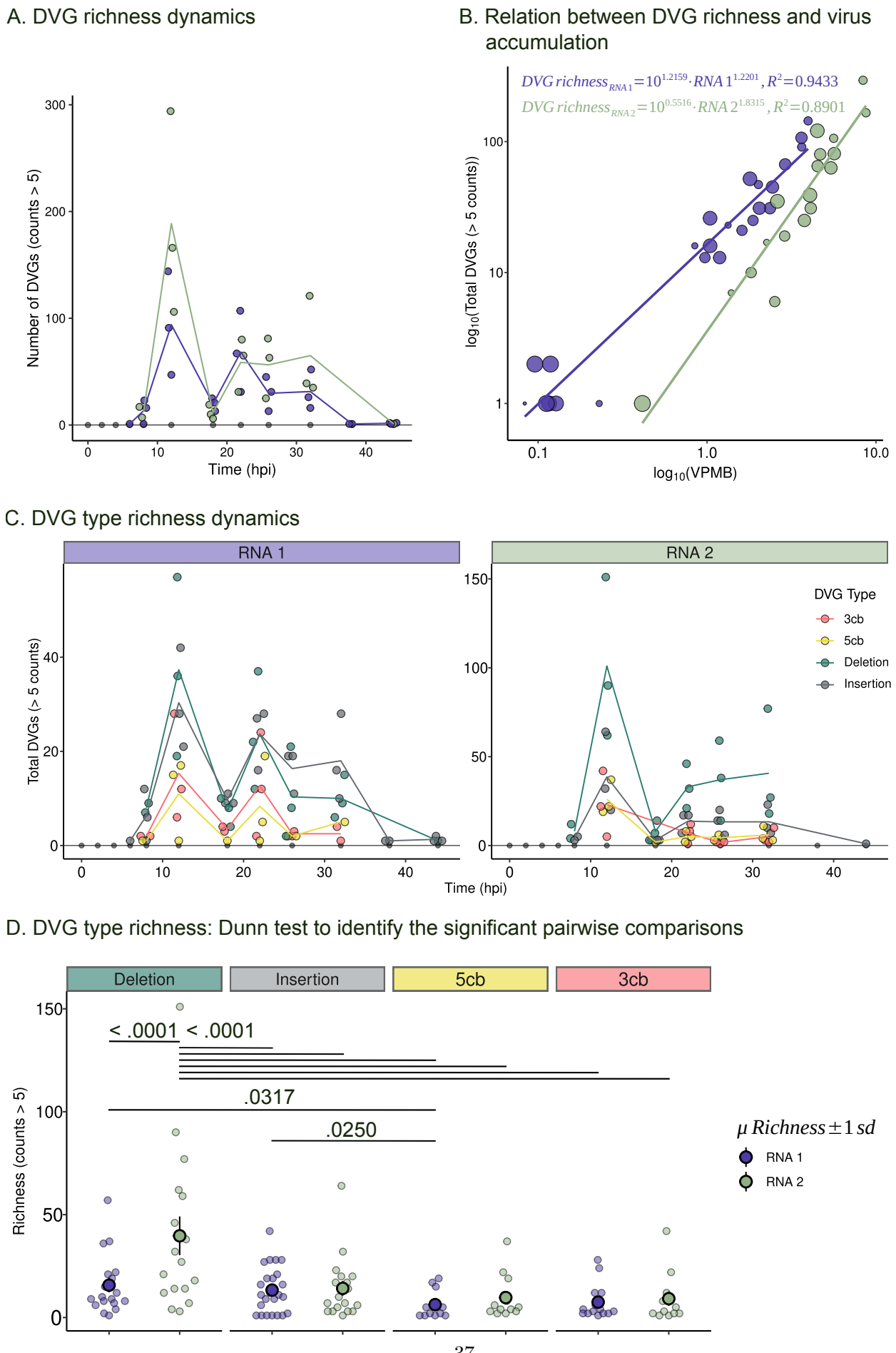





Figure 23: **DVG Richness**. A) DVG richness dynamics. B) Relation between DVG richness and virus accumulation. The relationship between the C) Viral species accumulation as proportion. D) Positive strand synthesis dependence reaseamble a saturation curve.

We first evaluate the DVG richness as the number of different types of defective genomes present in each sample. Fig. 23A shows the total richness per time and viral RNA. To evaluate the apparent positive relationship between the number of DVGs and the number of viral genomes (see Fig. 23B), three models were compared: a linear model, a log-linear model, and a log-log model. Among these, the log-log model showed the best overall statistical fit. It explained the highest proportion of variance ($R^2 = 0.9302$), exhibited the lowest residual standard error (0.426), and displayed a significantly higher $F$-statistic (183, $P \ll 0.0001$), indicating strong performance and more accurate predictions compared to the other models.

The log-log regression models for both RNA1 and RNA2 demonstrate a strong power-law relationship between the count of defective genomes and the total number of viral genomes of each segment(sum of VPMB). For RNA1, the model yields the following power-law equation:

$$\text{DVG richness}_{\text{RNA1}} = 10^{1.2159} \cdot \text{RNA1}^{1.2201}$$

This relationship is highly significant, with a $P$-value for the slope $\log_{10}(\text{VPMB})$ of $2.06 \times 10^{-15}$, $R^2 = 0.9458$, residual standard error = 0.1735, and $F = 383.7$ (22 $d.f.$).

For RNA2, the corresponding power-law equation was:

$$\text{DVG richness}_{\text{RNA2}} = 10^{0.5516} \cdot \text{RNA2}^{1.8315}$$

This model also demonstrates a highly significant relationship, with a $P$-value for the slope of $2.69 \times 10^{-9}$, adjusted $R^2 = 0.8901$, residual standard error = 0.1997, and $F = 138.7$ (16 $d.f.$). The differences in the model statistics show that the power-law relationship for RNA2 is slightly weaker than for RNA1.

The power-law exponents for RNA1 and RNA2 indicate that RNA2 accumulates DVGs at a 50.2% faster rate than RNA1. The 78.3% smaller intercept estimated for RNA2 suggests that the DVGs start to accumulate later, when higher levels of viruses are had been already produced. This could suggest that, despite RNA2 is approximately one-third shorter, its replication dynamics lead to a higher number of DVGs relative to the total genomes, possibly due to higher replication rates or greater error-prone replication.

When exploring the type of defectives that have been generated, we saw a predominance of deletions in RNA1, followed by insertions in both segments.

A LMM was fitted to assess the effects of RNA segment, DVG type, and hours post inoculation on the DVG richness, with sample as a random effect. The interaction between RNA2 and deletion was significant (`estimate` = 23.63, $P = 0.0003$), indicating a significant increase in richness under this condition. Time barely contribute ($P = 0.183$) to the richness in the complete model.

The *post hoc* pairwise comparisons revealed a significant highest richness of deletions in RNA2 than the rest of the DVG types in both viral segments ($P$ ł0.0001). Deletion in RNA1 and insertion in RNA1 shown a higher richness than 5' cb in the same viral segment ($P = 0.0317$ and $P = 0.0250$, respectively) (Fig.23D).

**DVG Abundance**

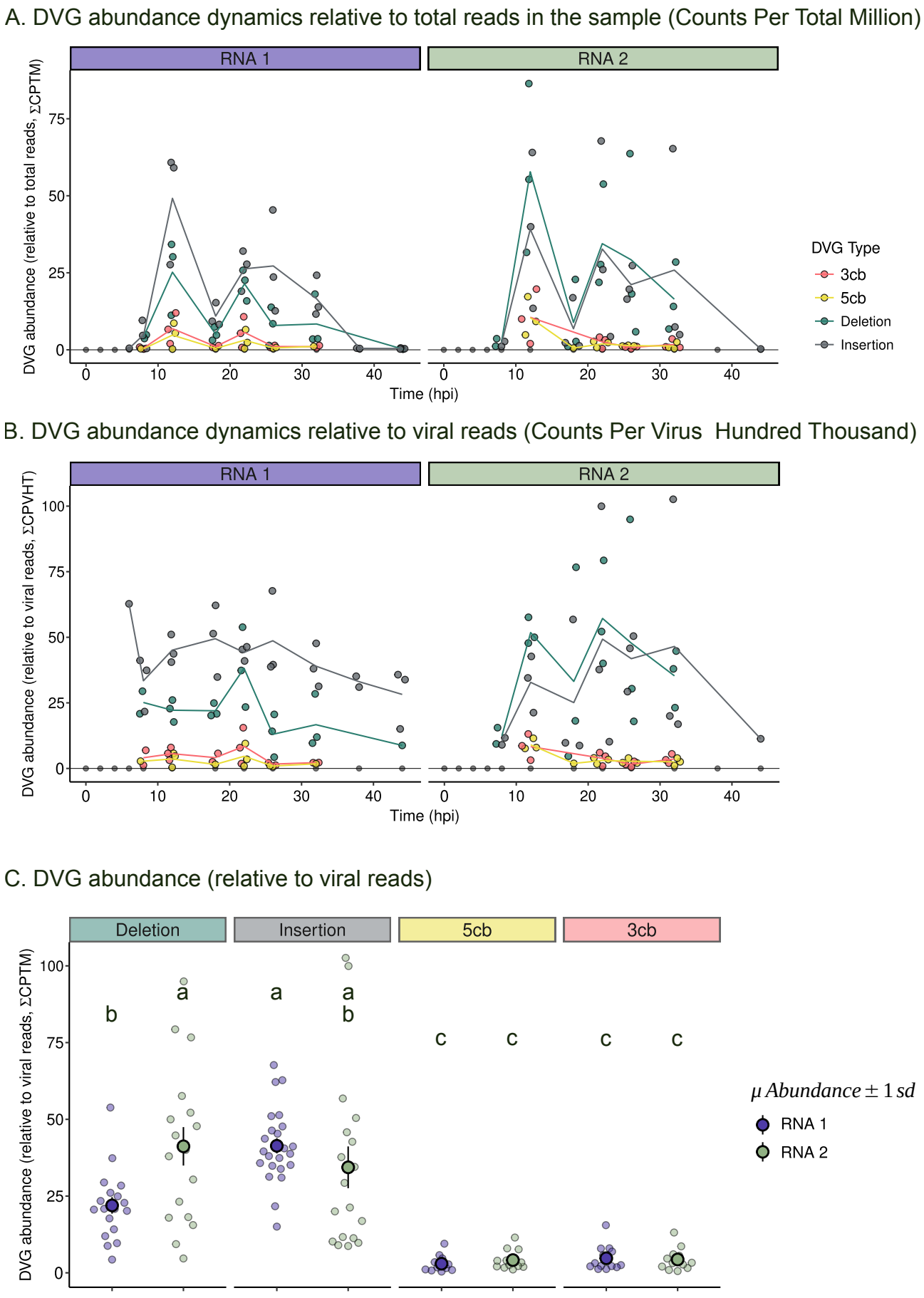





Figure 24: **DVG Abundance**. (A) DVG abundance dynamics relative to total reads in sample (CPTM). (B) DVG abundance dynamics relative to virus accumulation (CPVHT). (C) DVG abundance (CPVHT): Dunn test to identify significant pairwise comparisons.

We then focused in the abundance of the DVGs with two different normalization approaches. First, we use a normalization to the total number of reads sequenced (CPTM, counts per total million). This revealed the same pattern of positive dependence to the viral accumulation but with a general predominance of insertions in RNA1 (Fig. 24A), despite a similar richness in deletions was previously found for the same viral segment (see Fig. 23A).

But we were particularly interested in temporal dynamics of the proportion of DVGs in the population along the infection process. For that, we relativize the DVG counts to the viral genomes (CPVHT, counts per virus hundred thousand, Fig. 24B). This proportional approach revealed that ones the DVGs appear, their proportion remains quasi-stable from 8 to 32 hpi. A LMM testing for the effects of RNA, DVG type and time, while incorporating sample as random effect, confirmed that the time was less important when evaluate the abundance as proportion (textitP = 0.3653) than when is evaluated relative to the total library (textitP = 0.1615). *Post hoc* pairwise comparisons indicate deletions in RNA2 and insertions in RNA1 as the most abundant DVGs ($P$ ł0.0001 for all comparisons except when comparing with deletion DVG in RNA1 were the difference was less significant: $P = 0.0161$ with deletion DVG RNA2 and $P = 0.0011$ with insertion DVG RNA1). Insertions in RNA2 are also part of the of the most abundant DVGs group ("a" in Fig. 24C, $P$ ł0.0001 with all the copy-backs). Deletions in RNA1 were found more abundant than copy-backs (5'cb-RNA1: $P = 0.0043$, 5'cb-RNA2: $P = 0.0088$, 3'cb-RNA1: $P = 0.0140$, 3'cb-RNA2: $P = 0.0161$).

Lastly, we studied how richness and abundance relate to each other. *A priori*, we should expect a positive relationship between them, but the results suggest insertions are replicated at a higher rate than the rest of the defective types (Fig. 25A). The linear regression model confirmed the rate for insertions is notably higher (RNA1 slope 1.44, RNA2 slope 1.14) compared to deletions (RNA1 slope 0.653, RNA2 slope 0.572), 5'cb (RNA1 slope 0.410, RNA2 slope 0.446) and 3'cb (RNA1 slope 0.452, RNA2 slope 0.468) types. The statistical significance for these models are all highly significant ($P$ ¡ 0.001), indicating a strong positive relationship between DVG richness and their relative abundance in the samples. Notably, the goodness of fit values for most models were high, with 3' and 5' cb types showing the strongest fits (RNA1 $R^2 = 0.984$, RNA2 $R^2 = 0.992$), suggesting a well-defined replication pattern for these DVGs. The dynamics of the ratio abundance / richness (or the mean DVG abundance) appears more stable in RNA2 than in RNA1, where insertions have two picks of higher mean abundance at 12 and 26 hpi while deletions has one at 22 hpi. Copy-backs remain more stable across the time series. In RNA2, the mean density was lower for all DVG types and insertions only has a first pick of higher density at 18 hpi while deletions have it at the same time point (Fig. 25B).

To summarize, we computationally found all types of DVGs in the course of the OrV infection. From 6 hpi on, their richness strongly correlates with the virus accumulation following a power-law relationship with a later appearance but at a higher rate of generation for RNA2. Deletions and insertions are the DVGs showing the highest richness and the most abundante in both RNAs. The proportion of DVGs seems to be quite stable from 8 to 32 hpi but if the population of DVGs is also stable in its composition will be discused in the next section.

**DVG composition**

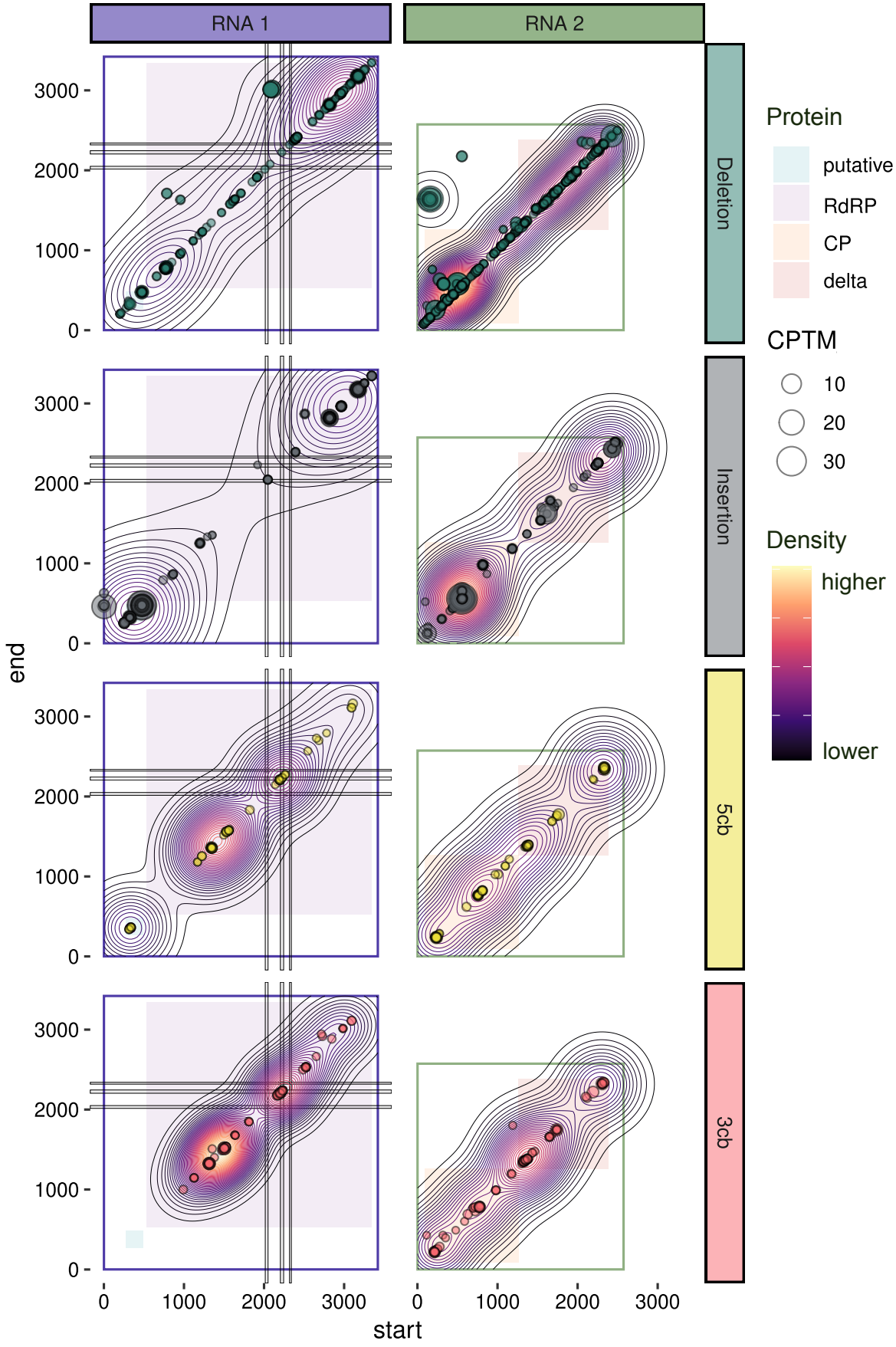



Figure 26: **DVG hotspots**. The start and end coordinates of all the DVGs detected at least in 5 reads. The size represents the abundance relative to the total library size per million reads (CPTM). Background colors represent the protein products coded in each region and an outer box marks the limit of each genome segment. In RNA1, the conserved motifs of the palmprint (A, B and C) are represented as gray rectangles.

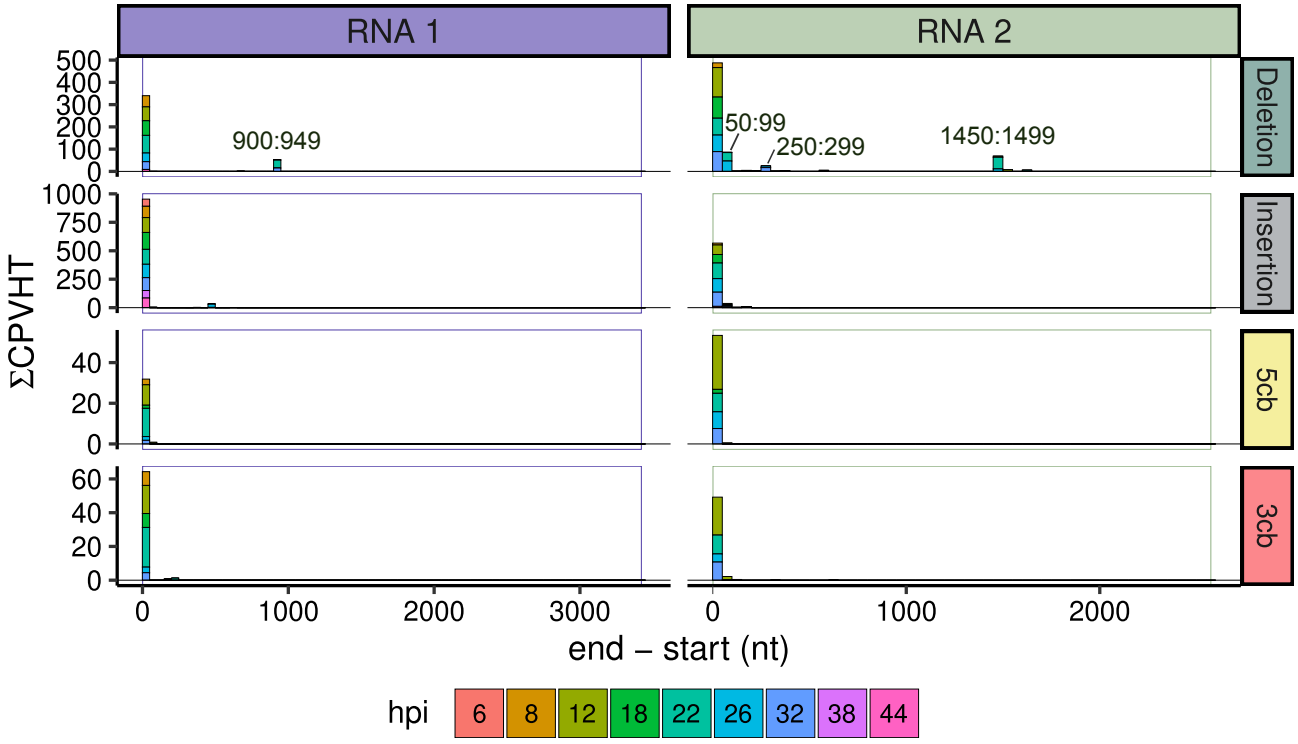



Figure 27: **Delta distribution**. Distribution of the start-end distances of all the DVGs detected at least in 5 reads. The colors filling the bars indicate the time points were a DVG with that delta was found.

After analyzing the richness and abundance with which DVGs appear, we wanted to know whether DVGs appear randomly spread in the genome or, as in other RNA viruses, they accumulated in specific regions [31].

A first look at the genome space revealed the presence of hotspots for all four types of DVGs (Fig. 26). Several small deletions accumulate across both segments (in the diagonal in Fig. 26 and the first bar (0 - 49 nucleotides) in Fig. 27) but some bigger and interesting deletions appear. In RNA1, there are three hotspots of convergent deletions far away from the diagonal, one of them with the start after the palmprint motif A and the end approximately in 3000, that is a 900-946 nucleotides deletion (see also Fig. 27). But even most interesting, the RNA segment with the structural proteins, present a heavely hotspot with a start 150 and the end 1460 (more than 1500 nucleotides deleted, Fig. 27) that we will further study as a subgenomic candidate.

Insertions appeared much less spread than deletions. In RNA1, they concentrate in the 5'UTR region and in the last quart of the RdRP while in RNA2 are three hotspots, the more dense of them, in the middle of the capsid gene (Fig. 26). The majority of these insertions are only of one nucleotide and are in homopolymeric regions. At RNA1, the most ubiquitous and abundant insertion was found at position 475, a {C}11 sequence, the region with the longest homopolymer. In the RNA2, 557 and 560 were the two positions in which more insertions were identified. Both coordinates are the start and the end of an {A}4 sequence that is also flanked by other homopolymeric sequences. If this insertion events are real or artifacts introduced during the sequencing should be addressed.

Finally, 5' and 3' cb show a similar accumulation pattern. This similarity is specially evident in RNA2.

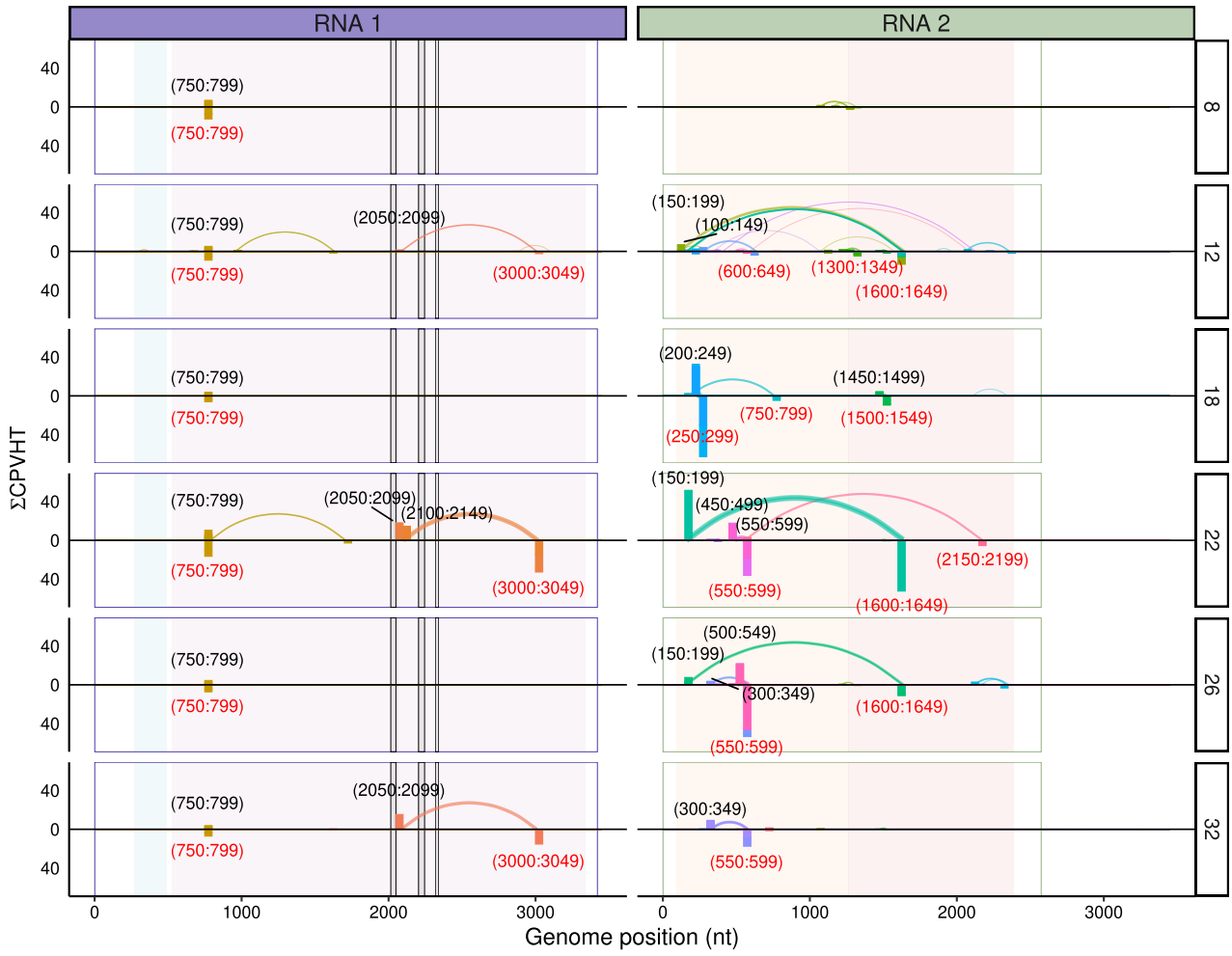



Figure 28: **Deletions arc plot**. Fuzzy positions (50 nucleotide window) of the start and end coordinates of the deletions detected at least in 5 reads. The colors filling the bars and the arcs joining an up and down oriented bar are references of the same deletion. Up oriented bars and black labels and down-bars with red labels for start and end positions respectively. As in other plots, background colors indicate the position of the genes, purple and green boxes represent the limits of RNA1 and RNA2 respectively. In the RNA1, gray rectangles mark the position of the A, B and C conserved subdomains of the RdRP palmprint.

Then, we took a closer look to the positions that with more frequency were start and end points of the abnormal joining, weighted by their relative abundance (read counts per hundred thousand of viral counts, VPVHT). A window size of 50 nucleotides was chosen to visualize the frequency plots. In the case of the deletions, an arc plot was also included to facilitate the visualization of the big deletions (Fig. 28). There are some results that should have more attention and that will require experimental validation. We have dig computationally in a few of them.

## 3.3 New insights into the transcription of OrV genome

### 3.3.1 A sgRNA candidate (or a really persistent deletion) would code for a shorter free-$\delta$ version

At 12 hpi appear the two big deletions we introduced previously. The one starting after motif A of the RdRP palmprint was also detected at 22 and 32 hpi at a considerable relative abundance. Other big deletions appear in RNA2 but the one deleting the major part of the CP (150-199 to 1600-1649 at 12, 22 and 26 hpi, Fig. 28) specially caught our. attention. Some (+)RNA viruses (such as coronaviruses) translate their structural proteins using as template subgenomic RNA (sgRNA) [55,63]. The existence of sgRNA has not been described in OrV. Our results suggest that there is a deletion with non strict coordinates in the RNA2 that is detected from 12 hpi in eight independent samples. The same fuzzy deletion was found (as the most abundant DVG) when analyzing the two evolved JUv1580 samples from [8], at similar or higher concentration. But also appear in the other two evolved JUv2572 samples from the same study.

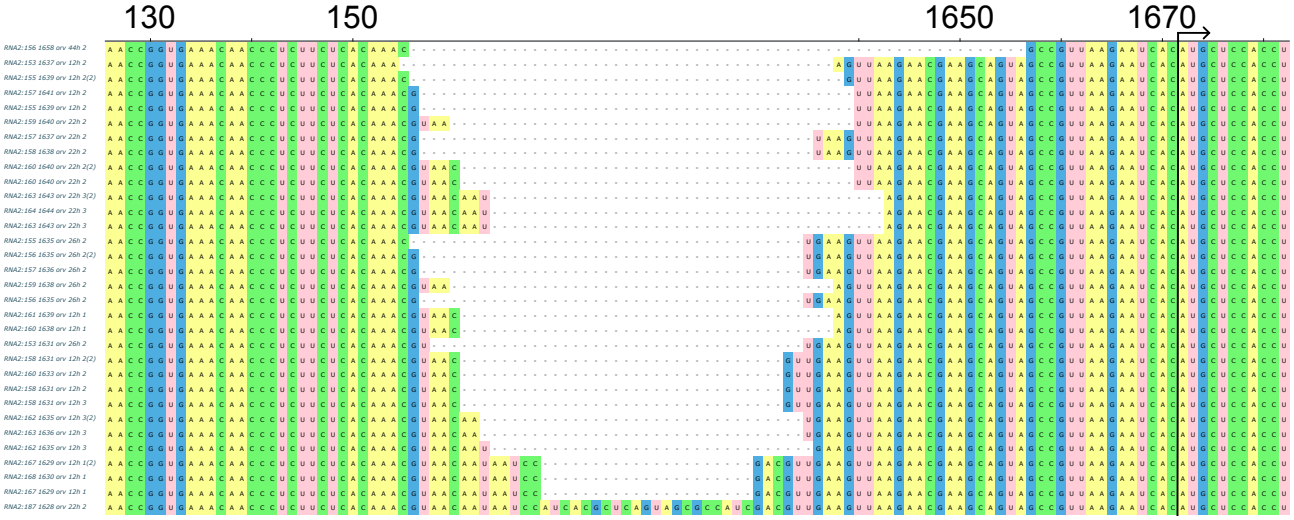



Figure 29: **Most abundant deletion in RNA2**. Alignment of all the deletions in RNA2 hotspot left intact a posible start codon in 1672 position.

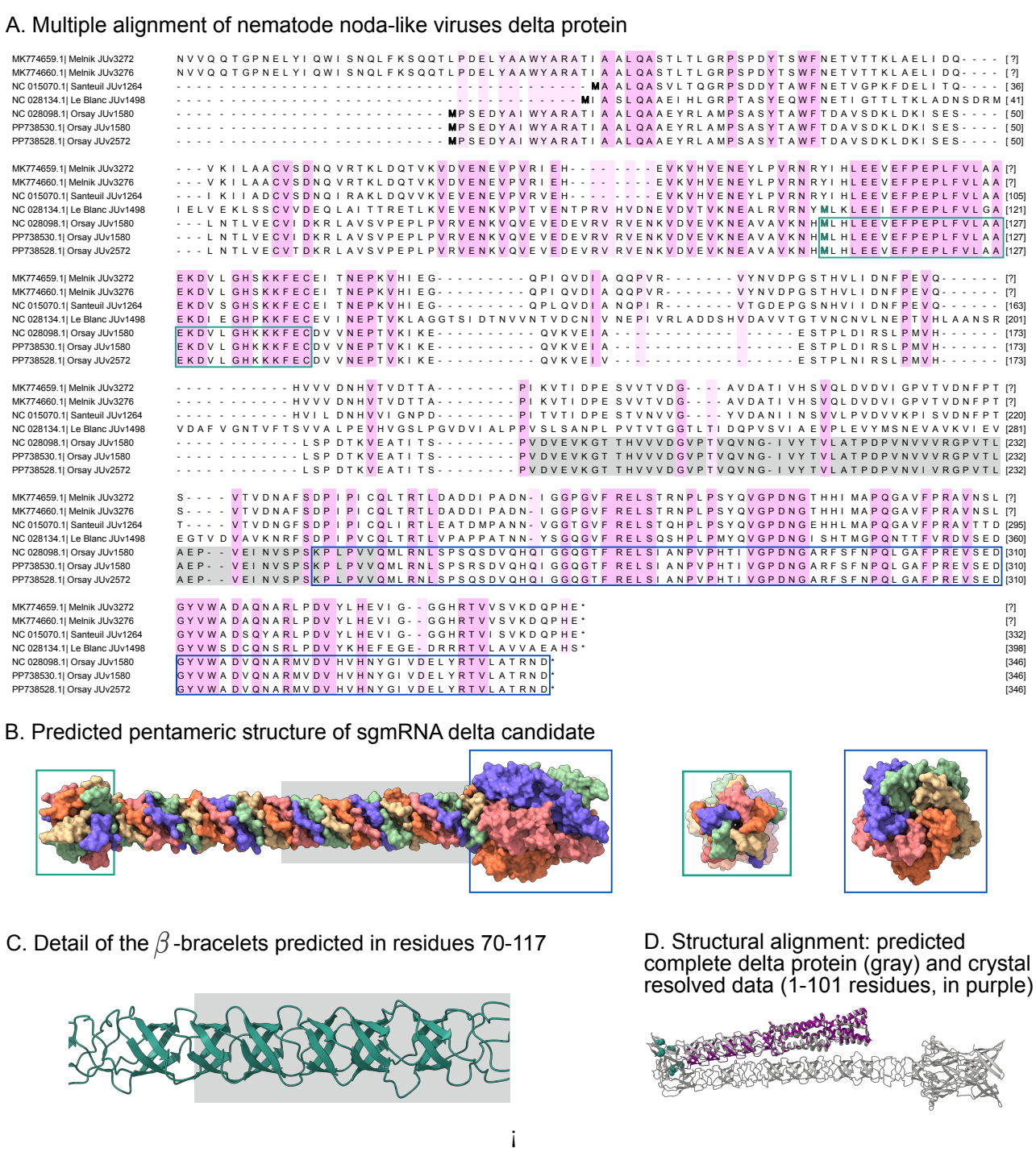



Figure 30: **sgRNA candidate of $\delta$ protein**. (A) Multiple alignment of the $\delta$ protein genes of the four nematode-infecting nodaviruses. The assembly sequences available at NCBI with differences in its nucleotide sequences have been use. Melnik virus is the last nematode infecting virus described and its sequence is incomplete. We have use it here but the presence of $\delta$ free protein has not been addressed yet, in fact, Melnik lacks the ATG codon that would allow to the other noda-like viruses express free $\delta$. Black bolded M indicates the start of the complete $\delta$ protein sequence while green bolded M marks the candidate start of the delta sgRNA version. In dark pink the conserved residues and in light pink the residues conserved in at least two species or strains from the same specie. Boxes indicate the globular domains, in green the N-terminal and in blue the C-terminal. The gray background shows the V-rich region identified by PROSITE motif search tool. (B) Alphafold 3.0 prediction of the protein structure coded by the sgRNA $\delta$ candidate identified in this work. Swab view from N-terminal globular domain, the fiber structure and the C-terminal domain. Frontal view of the N-terminal globular domain. Frontal view of the C-terminal globular domain. (C) Detail of the $\beta$-bracelets fold formed in the fiber (residues 70-117) with gray background indicating the V-rich region. (D) Structural alignment of the crystal structure of $\delta$ (1-101) (PDB ID: 5W82) in purple and the predicted structure of the complete version of free $\delta$ in gray. In green are the M residues that would correspond to the start of subgenomic $\delta$ version.

Despite the deletion has not exact coordinates, the alignment of all the reconstructed sequences allowed us to check that all the deletion versions left intact a start codon at position 1672 (Fig 29). That AUG is in frame with the $\delta$ protein gene and its putative product could code for a free $\delta$ version lacking the first 109 residues (M210 in Fig. 30A, predicted structure with Alphafold 3.0 Fig. 30B). We then check if that internal start was conserved in other related virus. Despite Melnik virus was recently discover [21] and there was no complete assembly of its genome, we also included it in order to have an actualized version of the multiple sequence alignment provided by [28]. We found that the same internal start is conserved in the other OrV strain JUv2572 and in Le Blanc virus (Fig. 30A). We already discovered that a similar deletion was conserved in JUv2572 evolved samples, if it is also generated during Le Blanc viral cycle should be clarify.

Then a prediction of a pentameric version of the sgRNA $\delta$ was performed with Alphafold 3.0 (Fig.30B). The structure coincides with the one predicted previously. Starts with a smaller globular domain that would correspond to the internal globular domain predicted by [28] and ends with a larger globular domain (Fig. 30B). Between both extremes, a fiber-like structure with a V-rich region (residues 79-140, gray background in Fig. 30A,B and C) was identified by the PROSITE motif search tool. A closer look to the fiber (Fig. 30C) allow as to identify more $\beta$-bracelets fold motifs, the ones discover in the crystal structure of $\delta$(1-101) [28]. [Fig. 30D, in purple $\delta$(1-101), PDB ID: 5W82]. To assess how much the sgRNA $\delta$ variant differs from the full $\delta$ version, the prediction for the latter was also performed using AlphaFold 3.0. Structural alignment between the complete $\delta$ and the crystal structure of the first 101 residues shows almost perfect overlap (Fig. 30D). When compared to the shorter sgRNA $\delta$ version, it is evident that the residues forming the articulation in the full version correspond to those forming the N-terminal globular domain. In other words, the product of the subgenomic candidate would result in a $\delta$ version lacking the articulation, with the shorter fiber ending in $\alpha$-helices, but adopting a more globular at the ends, swab-like shape.

It has been described that free $\delta$ facilitates OrV exit [64] although it has not been detected yet by Western blot with the current version of the antibody. Through the recombinant expression of the complete $\delta$ protein, [64] found that $\delta$ co-localize with the apical membrane of the host intestine. The authors also show that a truncated version of $\delta$ (C-terminal $\delta$ fragment, 67 to 346 residues) had a more disperse localization while $\delta$(1-66) was enough to mantain the apical localization. With that results, the first 66 residues are considered needed for the $\delta$-apical membrane interaction. The sgmRNA $\delta$ version would lack the $\alpha$ helices module but also is different from the truncated version that show the more disperse localization. Where would be their localization should be addressed but one hypothesis is that by loosing the $\alpha$ helices module will loose its hability to interact with ACT-5. By other part, recent investigations (unpublished work from Yusong R. Guo *et al.*) showed that the $\delta$-CTD directly interactes with FSHR-1. FSHR-1 encodes a putative neuropeptide receptor [56], is located in plasma membrane and expressed in several structures including intestine. Interestingly, *fshr-1* mutants has several disease phenotypes, including a pathogen susceptibility increased [49]. How OrV $\delta$ interacts with FSHR-1 and modulates their signaling function would be an interesting pathway of exploration. But the last results from Guo *et al.* suggest that the shorter free-$\delta$ version produced by the sgmRNA we proposed here could mantain the interaction property.

The results here suggest that the deletion identified in the RNA2 between the fuzzy positions 150-1650 is a consistent candidate to further investigate as a possible sgRNA. But if it was not a sgRNA, its presence as a persistent and ubiquitous DVG makes it interesting *per se*.

### 3.3.2 Small ORFs in OrV and a new peptide candidate

OrV has only four annotated proteins: RdRP encoded by RNA1 and CP, free $\delta$ and CP-$\delta$ fusion by RNA2. The atypically large size of the 5'UTR region of the RNA1, together with some peculiarities in the most 3' end of the 5'UTR region impulse us to take a closer look. A rapid prediction with the `FindORFs` tool from UGENE toolkit [45] considering only AUG as initiation codon, revealed 21 new ORFs in RNA1 and 17 new ORFs in RNA2 (Fig. 31, A and B respectively).

Of all the new ORFs, the 72 residues peptide encoded in the (-)RNA1 [nucleotides 272 to 487] was predicted by Prodigal v2.6.3 (confidence score = 81.91 %) and by PROKKA (in green in Fig. 31A). If a peptide was produced from this putative ORF, their genomic position at the end of the (-)RNA1 strand would probably require for an internal ribosomal entry site (IRES) for its translation. The IRESite web tool contains a database of viral and cellular IRES [43]. A search of the sequence upstream the putative peptide (from nucleotides site 491 to 654) had positive matches with several IRES database, being the best score with the IRES from the taura syndrome virus (TSV) (IRESite_Id:77, score = 26.0 bits (14), *E*-value = 8.5). TSV is a small picornavirus virus that infects invertebrates (a wide range of penaeid prawns).

The structure of the peptide and the $n$-mers [$n \in \{2, 20\}$] was then predicted with AlphaFold 2.0 and 3.0. Alphafold 2.0 achieved the better average pLDDT for the monomer (Fig. 32B) while AlphaFold 3.0 was better for the dimeric predictions. The dimeric form of the putative peptide (Fig. 32C) had the best score ($\mu pLDDT = 58.4778 \pm 0.2542$) (Table 6).

A structural alignment of the predicted structures with the DALI server [32] found 832 results for the monomeric structure and 829 for the dimeric prediction with a $Z > 2$. About the monomer structure, the matches with highest $Z$ values correspond to the BACTERIOPHYTOCHROME from *Stigmatella aurantiaca* (PDBs: 8UPK, 8UPH and 8UQI, Z = 4.1, 4.1 and 3.8 respectively) and the SENSOR HISTIDINE KINASE CPXA from *Escherichia coli* (PDB: 8UK7, Z = 3.7). All of them involved in signaling. Second, a frequency word count revealed that heat-shock, followed by uridylate-specific were the most frequent description of the monomer matches (Fig. 32D-E). *C. elegans*uses terminal uridylyltransferases as part of the antiviral response [35]. This activities would target the viral RNA for degradation and some viruses have evolved mechanisms to conteract. This is the case of the nsp15 of SARS-CoV-2. Different available structures of this protein have been positive matched to the putative monomer of OrV. The 9BIH PDB model, was the one with the highest score ($Z = 3.2$) (Fig. 32E). By other part, the importance of the heat shock response in viral infections has been widely addressed in several organisms but more specifically in *C. elegans* [7]. The fact that a OrV protein could be mimicking a HSP motif or even working as that, would require further investigation. The structural alignments of the dimeric prediction obtained a less $\mu Z$ than the monomeric matches ($2.39 \pm 0.271$ *vs.* $2.46 \pm 0.299$)(Fig. 32) but still, some interesting coincidences were found. The best matches were with the virulent factor PHIST from *Plasmodium falciparum* and with AriA, involved in the antiviral defense of *Escherichia coli* (PBD: 4JLE and 8V46, $Z = 3.4$ and 3.3). Another interesting coincidence was with the deneddylase from the human alphaherpesvirus (PBD: 4TT0, $Z = 3.1$). Additionally, the dimeric structure was most frequently aligned with proteins involved in RNA processing. MEDIATOR OF RNA POLYMERASE II TRANSCRIPTION SUBUNIT, DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB1 or PRE-MRNA-SPLICING FACTOR 8 were the most redundant related terms within the descriptions.

We believe the results summarized here are interesting enough to consider this putative peptide as a promising candidate for further investigation.

| n | Alphafold version | $\mu pLDDT \pm SEM$ |
|---|---|---|
| **1** | **2** | **53.4354** 1.2032 |
| 1 | 3 | 47.7608 ± 0.3660 |
| 2 | 2 | 34.3601 ± 0.4473 |
| **2** | **3** | **58.4778** ± 0.2542 |
| 3 | 3 | 44.5788 ± 0.1600 |
| 4 | 3 | 35.9023 ± 0.1109 |
| 5 | 3 | 33.5537 ± 0.1116 |
| 6 | 3 | 31.8185 ± 0.0884 |
| 7 | 3 | 30.3391 ± 0.0918 |
| 8 | 3 | 30.3646 ± 0.0845 |
| 9 | 3 | 28.1610 ± 0.0762 |
| 10 | 3 | 25.6378 ± 0.0701 |
| 11 | 3 | 27.6473 ± 0.0654 |
| 12 | 3 | 27.7242 ± 0.0675 |
| 13 | 3 | 27.1384 ± 0.0611 |
| 14 | 3 | 27.2985 ± 0.0520 |
| 15 | 3 | 26.8289 ± 0.0541 |
| 16 | 3 | 26.5972 ± 0.0489 |
| 17 | 3 | 27.7678 ± 0.0541 |
| 18 | 3 | 25.3379 ± 0.0348 |
| 19 | 3 | 27.1220 ± 0.0447 |
| 20 | 3 | 28.6583 ± 0.0439 |

Table 6: Putative peptide encoded in the 3' end of (-)RNA1. (A) Quality of the predictions of Alphafold 2.0 (gray) and 3.0 (black) for $n$-mers. (B) Predicted monomer structure. (C) Predicted dimer structure, (D) Biological functions associated to proteins that fold like the predicted monomer. (E) $Z$ scores distribution of the associated proteins that fold like the predicted monomer. The similar proteins with the two more redundant biological terms are indicated with the corresponding colors. (F)(G) Similar to (D)(E) for the predicted dimer.

# 4 Conclusion

OrV, the only virus identified infecting natural populations of the nematode *C. elegans*, was previously described as being in a state of molecular evolutionary stasis [8]. Here, we present a characterization of the viral accumulation and diversity dynamics that unfold during a primary infection. As in other (+)RNA viruses, the accumulation of the viral genome and antigenome follows a stamping machine replication mode. By other part, this short term study confirm that most of the diversity produced is purified but some of the minor variants are convergent through the entire process. Some genomic regions seemed to show higher levels of constrains operating at the shorter scale. This results, together with a more macroevolution approach highlighted the RdRP tail as an interesting target of evolution. For the first time, we have *in silico* identified the presence of DVGs and proposed a persistent deletion as a candidate of sgRNA that would code for a shorter version of the free $\delta$ protein. Finally, a new putative peptide has been proposed for validation.

# References

[1] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A.J. Ballard, J. Bambrick, S.W. Bodenstein, D.A. Evans, C.C. Hung, M. O'Neill, D. Reiman,

K. Tunyasuvunakool, A. Wu, Z. a nd Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A.I. Cowen-Rivers, A. Cowie, M. Figurnov, F.B. Fuchs, H. Gladman, R. Jain, Y.A. Khan, C.M.R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E.D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, and J.M. Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

[2] S. Andrews. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, jun 2010.

[3] A. Babaian and R. Edgar. Ribovirus classification by a polymerase barcode sequence. *PeerJ*, 10:1–17, 2022.

[4] C. Baena-Angulo, A.I. Platero, and J.P. Couso. Cis to trans: small ORF functions emerging through evolution. *Trends in Genetics*, nov 2024.

[5] R.H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[6] L.A. Casorla-Perez, R. Guennoun, C. Cubillas, B. Peng, K. Kornfeld, and D. Wang. Orsay Virus Infection of Caenorhabditis elegans Is Modulated by Zinc and Dependent on Lipids. *Journal of Virology*, 96(22), nov 2022.

[7] V.G. Castiglioni and S.F. Elena. Orsay virus infection increases Caenorhabditis elegans resistance to heat-shock. *Biology Letters*, 20(8), 2024.

[8] V.G. Castiglioni, M.J. Olmo-Uceda, S. Martín, M.A. Félix, R. González, and S.F. Elena. Experimental evolution of an RNA virus in Caenorhabditis elegans. *Infection, Genetics and Evolution*, 123, sep 2024.

[9] V.G. Castiglioni, M.J. Olmo-Uceda, A. Villena-Giménez, J.C. Muñoz-Sánchez, E.G. Legarda, and S.F. Elena. Story of an infection: Viral dynamics and host responses in the caenorhabditis elegans-orsay virus pathosystem. *Science Advances*, 10:5945, 2024.

[10] V.G. Castiglioni, A. Villena-Giménez, A. González-Sánchez, C. Toft, G.G. Gómez, and S.F. Elena. Chronic infection of Caenorhabditis elegans by Orsay virus induces age- dependent immunity and superinfection exclusion. *Preprint*, pages 1–41, 2024.

[11] L. Chao, C.U. Rang, and L.E. Wong. Distribution of Spontaneous Mutants and Inferences about the Replication Mode of the RNA Bacteriophage $\phi$6. *Journal of Virology*, 76(7):3276–3281, 2002.

[12] P. Danecek, J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies, and H. Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2):1–4, 2021.

[13] J.A. den Boon, M. Nishikiori, H. Zhan, and P. Ahlquist. Positive-strand rna virus genome replication organelles: structure, assembly, control. *Trends in Genetics*, 40:681–693, 2024.

[14] D. DiMaio. Viral miniproteins. *Annu Rev Microbiol.*, 68(1):21–43, 2014.

[15] P.T Dolan, Z.J. Whitfield, and R. Andino. Mechanisms and concepts in rna virus population dynamics and evolution. *Annual review of virology*, 5:69–92, 9 2018.

[16] E. Domingo. *Virus as populations : composition, complexity, quasispecies, dynamics, and biological implications*. Second edition edition, 2019.

[17] S. Duffy, L.A. Shackelton, and E.C. Holmes. Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, 9(4):267–276, 2008.

[18] P. Ewels, M. Magnusson, S. Lundin, and M. Käller. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, 2016.

[19] Y. Fan, Y.R. Guo, W. Yuan, Y. Zhou, M.V. Holt, T. Wang, B. Demeler, N.L. Young, W. Zhong, and Y.J. Tao. Structure of a pentameric virion-associated fiber with a potential role in orsay virus entry to host cells. *PLoS Pathogens*, 13:1–24, 2017.

[20] C.J. Franz, H. Renshaw, L. Frezal, Y. Jiang, M.A. Félix, and D. Wang. Orsay, santeuil and le blanc viruses primarily infect intestinal cells in caenorhabditis nematodes. *Virology*, 448:255–264, 1 2014.

[21] L. Frézal, H. Jung, S. Tahan, D. Wang, and M.A. Félix. Noda-Like RNA viruses infecting Caenorhabditis nematodes: sympatry, diversity, and reassortment . *Journal of Virology*, 93(21), nov 2019.

[22] M.A. Félix, A. Ashe, J. Piffaretti, G. Wu, I. Nuez, T. Bélicard, Y. Jiang, G. Zhao, C.J. Franz, L.D. Goldstein, M. Sanroman, E.A. Miska, and D. Wang. Natural and experimental infection of caenorhabditis nematodes by novel viruses related to nodaviruses. *PLoS Biology*, 9, 2011.

[23] L.G. Gebhard, N.G. Iglesias, L.A. Byk, C.V. Filomatori, F.A. De Maio, and A.V. Gamarnik. A Proline-Rich N-Terminal Region of the Dengue Virus NS3 Is Crucial for Infectious Particle Production. *Journal of Virology*, 90(11):5451–5461, 2016.

[24] J.L. Geoghegan and E.C. Holmes. Evolutionary virology at 40. *Genetics*, 210:1151–1162, 12 2018.

[25] Toni Giorgino. Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software*, 31(7):1–24, 2009.

[26] R. González and M.A. Félix. Caenorhabditis elegans immune responses to microsporidia and viruses. *Developmental and Comparative Immunology*, 154(February 2023), 2024.

[27] R. González and M.A. Félix. Naturally-associated bacteria modulate Orsay virus infection of Caenorhabditis elegans. *PLoS Pathogens*, 20(1), jan 2024.

[28] Y.R. Guo, Y. Fan, Y. Zhou, M. Jin, J.L. Zhang, H. Jiang, M.V. Holt, T. Wang, N.L. Young, D. Wang, W. Zhong, and Y.J. Tao. Orsay Virus CP-$\delta$ Adopts a Novel $\beta$-Bracelet Structural Fold and Incorporates into Virions as a Head Fiber. *Journal of Virology*, 94(21), 2020.

[29] Y.R. Guo, C.F. Hryc, J. Jakana, H. Jiang, D. Wang, W. Chiu, W. Zhong, and Y.J. Tao. Crystal structure of a nematode-infecting virus. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35):12781–12786, sep 2014.

[30] R.P. Hellens, C.M. Brown, M.A.W. Chisnall, P.M. Waterhouse, and R.C. Macknight. The emerging world of small orfs. *Trends in Plant Science*, 21(4):317–328, apr 2016.

[31] J. Hillung, M.J. Olmo-Uceda, J.C. Muñoz-Sánchez, and S.F. Elena. Accumulation dynamics of defective genomes during experimental evolution of two betacoronaviruses. *bioRxiv*, 2024.

[32] L. Holm, A. Laiho, P. Törönen, and M. Salgado. DALI shines a light on remote homologs: One hundred discoveries. *Protein Science*, 32(1):1–18, 2023.

[33] H. Jiang, C.J. Franz, G. Wu, H. Renshaw, G. Zhao, A.E. Firth, and D. Wang. Orsay virus utilizes ribosomal frameshifting to express a novel protein that is incorporated into virions. *Virology*, 450-451:213–221, feb 2014.

[34] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[35] J. Le Pen, H. Jiang, T. Di Domenico, E. Kneuss, J. Kosałka, C. Leung, M. Morgan, C. Much, K.L.M. Rudolph, A.J. Enright, D. O'Carroll, D. Wang, and E.A. Miska. Terminal uridylyltransferases target RNA viruses as part of the innate immune system. *Nature Structural and Molecular Biology*, 25(9):778–786, sep 2018.

[36] E.G. Legarda, S.F. Elena, and A.R. Mushegian. Emergence of two distinct spatial folds in a pair of plant virus proteins encoded by nested genes. *Journal of Biological Chemistry*, 300(5):107218, 2024.

[37] H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Oxford University Press*, 00(00):1–3, 2013.

[38] T. Liu and D. Li. Convergence of the BFGS-SQP method for degenerate problems. *Numerical Functional Analysis and Optimization*, 28(7-8):927–944, 2007.

[39] R. Lorenz, S.H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, and I.L. Hofacker. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1):26, 2011.

[40] F. Martínez, J. Sardanyés, S.F. Elena, and J.A. Daròs. Dynamics of a plant RNA virus intracellular accumulation: Stamping machine vs. geometric replication. *Genetics*, 188(3):637–646, jul 2011.

[41] Z. Mérai, Z. Kerényi, A. Molnár, E. Barta, A. Válóczi, G. Bisztray, Z. Havelda, J. Burgyán, and D. Silhavy. Aureusvirus P14 Is an Efficient RNA Silencing Suppressor That Binds Double-Stranded RNAs without Size Specificity. *Journal of Virology*, 79(11):7217–7226, jun 2005.

[42] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6):679–682, 2022.

[43] M. Mokrejš, T. Mašek, V. Vopálenský, P. Hlubuček, P. Delbos, and M. Pospíšek. IRESite A tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Research*, 38(SUPPL.1):131–136, 2009.

[44] M. Nishikiori, J.A. Den Boon, N. Unchwaniwala, and P. Ahlquist. Crowning touches in positive-strand rna virus genome replication complex structure and function. *Annual Review of Virology*, 9:193–212, 2022.

[45] K. Okonechnikov, O. Golosova, M. Fursov, A. Varlamov, Y. Vaskin, I. Efremov, O.G. German Grehov, D. Kandrov, K. Rasputin, M. Syabro, and T. Tleukenov. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, 28(8):1166–1167, 2012.

[46] M.J. Olmo-Uceda, J.C. Muñoz-Sánchez, W. Lasso-Giraldo, V. Arnau, W. Díaz-Villanueva, and S.F. Elena. DVGfinder: A metasearch tool for identifying defective viral genomes in rna-seq data. *Viruses*, 14(5):1114–10134, 2022.

[47] E.Z. Poirier and M. Vignuzzi. Virus population dynamics during infection, apr 2017.

[48] Posit team. *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA, 2024.

[49] J.R. Powell, D.H. Kim, and F.M. Ausubel. The G protein-coupled receptor FSHR-1 is required for the Caenorhabditis elegans innate immune response. Technical report, 2009.

[50] X. Ren and J.H. Hurley. Proline-rich regions and motifs in trafficking: From ESCRT interaction to viral exploitation. *Traffic*, 12(10):1282–1290, 2011.

[51] M. Safari and M.J. Roossinck. How does the genome structure and lifestyle of a virus affect its population variation?, dec 2014.

[52] R. Sanjuán, M.R. Nebot, N. Chirico, L.M. Mansky, and R. Belshaw. Viral Mutation Rates. *Journal of Virology*, 84(19):9733–9748, 2010.

[53] J. Sardanyés, J.A. Martínez, F.and Daròs, and S.F. Elena. Dynamics of alternative modes of RNA replication for positive-sense RNA viruses. *Journal of the Royal Society Interface*, 9(69):768–776, 2012.

[54] J. Sardanyés, R.V. Solé, and S.F. Elena. Replication Mode and Landscape Topology Differentially Affect RNA Virus Mutational Load and Robustness. *Journal of Virology*, 83(23):12579–12589, dec 2009.

[55] P.B. Sethna, S.L. Hung, and D.A. Brian. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proceedings of the National Academy of Sciences of the United States of America*, 86(14):5626–5630, 1989.

[56] D. Sieburth, Q. Ch'ng, M. Dybbs, M. Tavazoie, S. Kennedy, D. Wang, D. Dupuy, J.F. Rual, D.E. Hill, M. Vidal, G. Ruvkun, and J.M. Kaplan. Systematic analysis of genes required for synapse structure and function. *Nature*, 436(7050):510–516, jul 2005.

[57] K.E.R. Soetaert, T. Petzoldt, and R.W. Setzer. Solving differential equations in r: package desolve. *Journal of statistical software*, 33, 2010.

[58] N. Stoletzki and A. Eyre-Walker. Estimation of the neutrality index. 2010.

[59] G. Thébaud, J. Chadœuf, M.J. Morelli, J.W. McCauley, and D.T. Haydon. The relationship between mutation frequency and replication strategy in positive-sense single-stranded RNA viruses. *Proceedings of the Royal Society B: Biological Sciences*, 277(1682):809–817, 2010.

[60] N. Unchwaniwala, H. Zhan, J.A. den Boon, and P. Ahlquist. Cryo-electron microscopy of nodavirus rna replication organelles illuminates positive-strand rna virus genome replication. *Current Opinion in Virology*, 51:74–79, 12 2021.

[61] J. Votteler and W.I. Sundquist. Virus budding and the ESCRT pathway. *Cell Host Microbe*, 14(3), 2014.

[62] A. Wilm, P.P.K. Aw, D. Bertrand, G.H.T. Yeo, S.H. Ong, C.H. Wong, C.C. Khor, R. Petric, M.L. Hibberd, and N. Nagarajan. LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, 40(22):11189–11201, dec 2012.

[63] H.Y. Wu and D.A. Brian. Subgenomic messenger RNA amplification in coronaviruses. *Proceedings of the National Academy of Sciences of the United States of America*, 107(27):12257–12262, 2010.

[64] W. Yuan, Y. Zhou, Y. Fan, Y.J. Tao, and W. Zhong. Orsay $\delta$ Protein Is Required for Nonlytic Viral Egress. *Journal of Virology*, 92(14), jul 2018.

[65] A. Zarrinpar, R.P. Bhattacharyya, and W.A. Lim. The structure and function of proline recognition domains. *Science's STKE : signal transduction knowledge environment*, 2003(179):1–11, 2003.

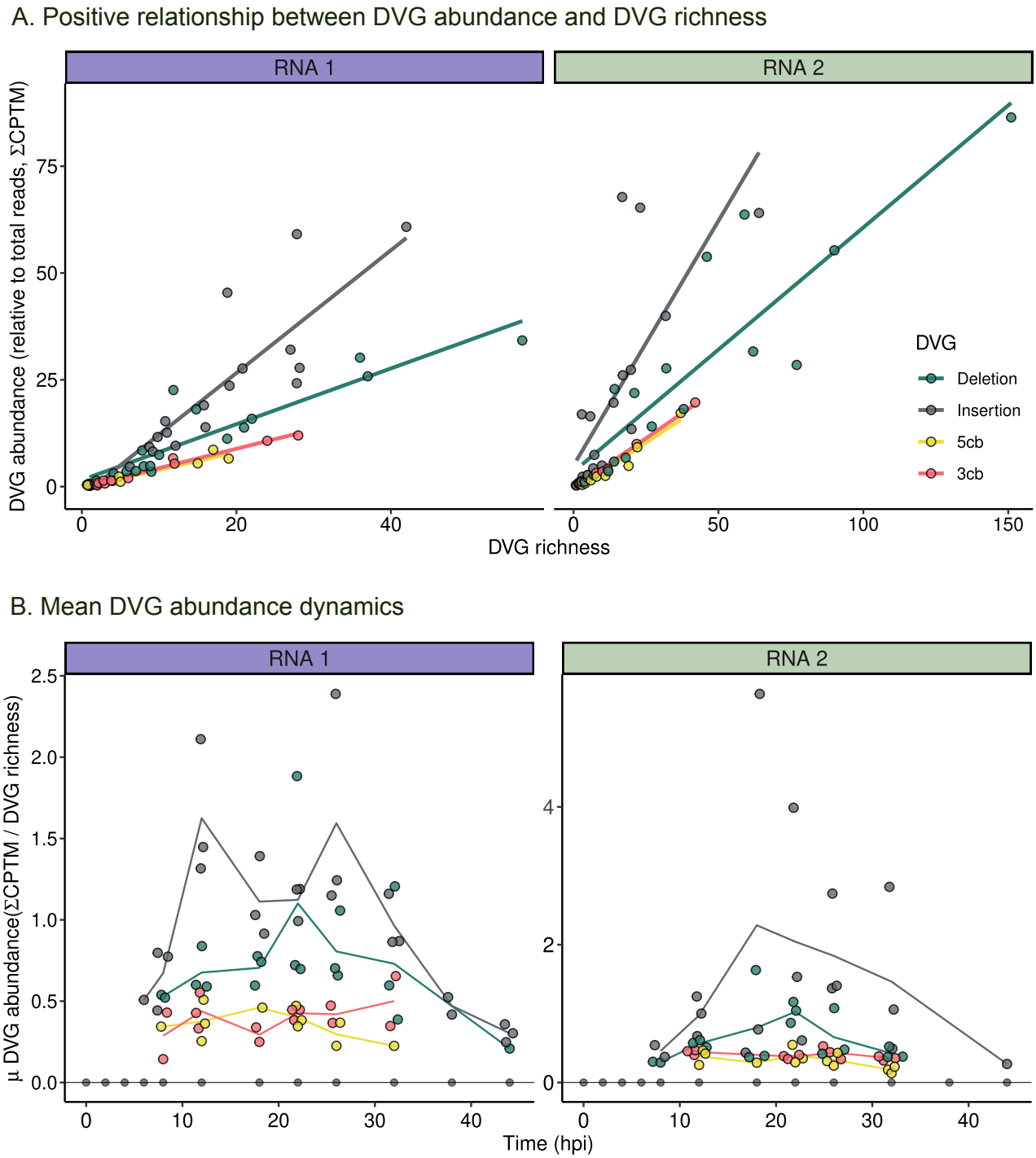
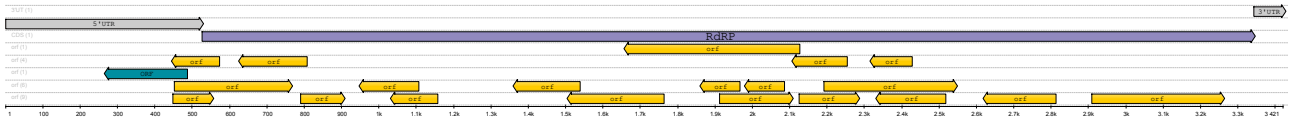



Figure 25: **Relation DVG abundance and richness**. (A) Positive relationship between DVG abundance and DVG richness. (B) Mean DVG abundance dynamics.

A. Annotated proteins and predicted ORFs in OrV JUv1580_vlc RNA 1




B. Annotated proteins and predicted ORFs in OrV JUv1580_vlc RNA 2

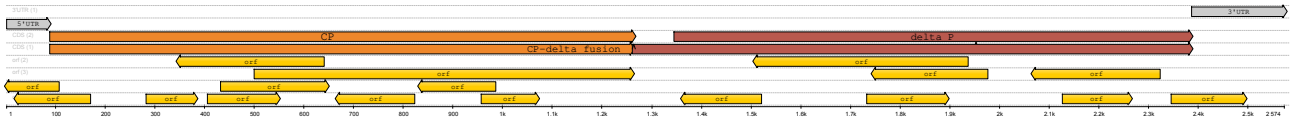



Figure 31: **Annotated (RdRP: purple, CP: orange, $\delta$: red) and predicted proteins (blue and yellow) in JUv1580_vlc genome.**. (A) RNA1, in green the new peptide candidate. (B) RNA2.

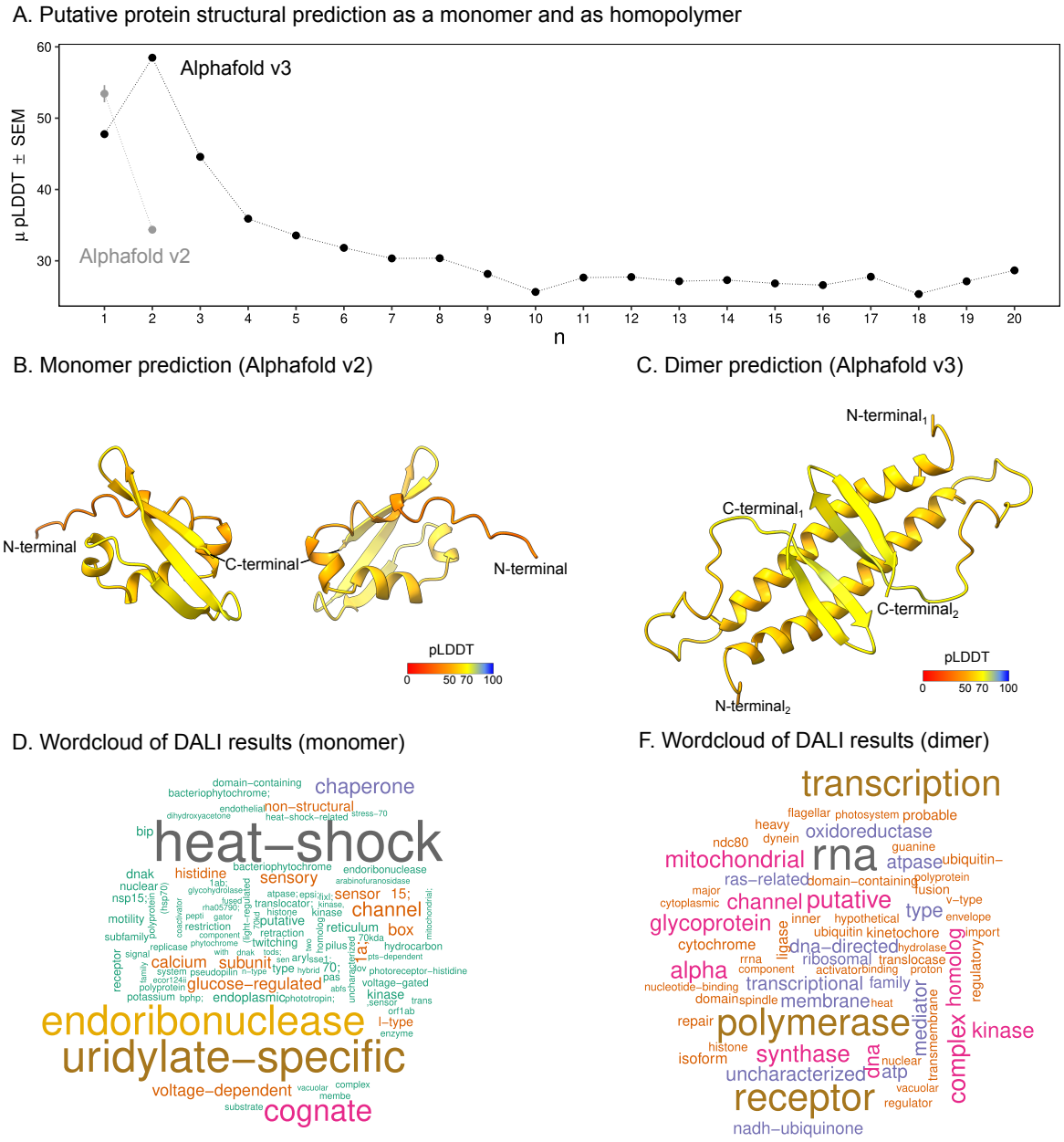


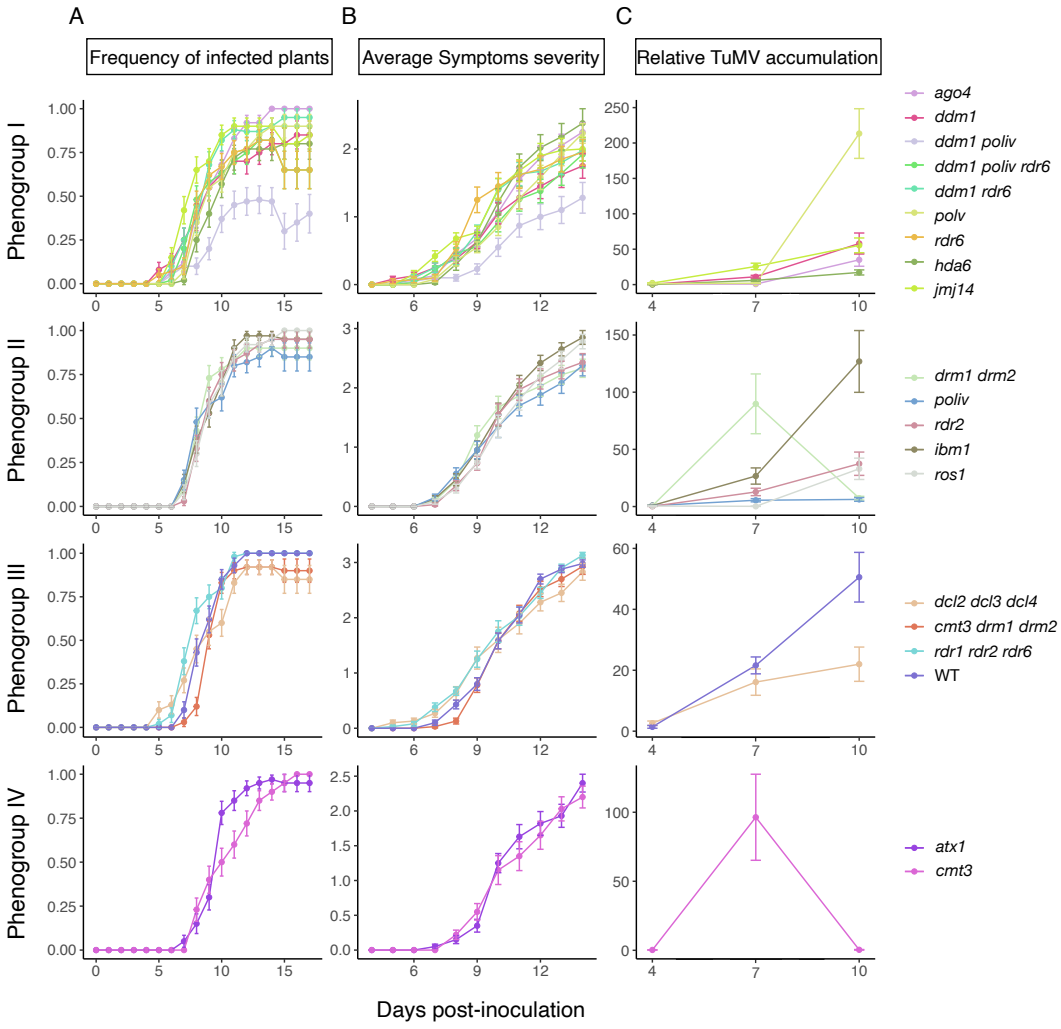

Figure 32: **A putative new peptide on the reverse strand of RNA 1**. (A) Mean prediction scores (pLDDT) of AlphaFold v2 (gray) and Alphafold v3 (black) as monomer (n = 1) and homopolymers ($n \in \{2, 20\}$) sh ows the better prediction scores are for the dimeric estructure with Alphafold v3 followed by the monomer prediction by AlphaFold v2. (B) Monomer prediction (AlphaFold v2) and (C) dimeric prediction (AlphaFold v3) colored by pLDDT values. (D) Results of the structural alignment with against the complete PDB database represented as a wordcloud of the descriptions and as a distribution of the Z values of the matches for the monomeric (D and E) and the dimeric structures (F,G). Colored dots in (E) and (G) represent the Z values of the two most common matches.

B: Supplementary material from published papers

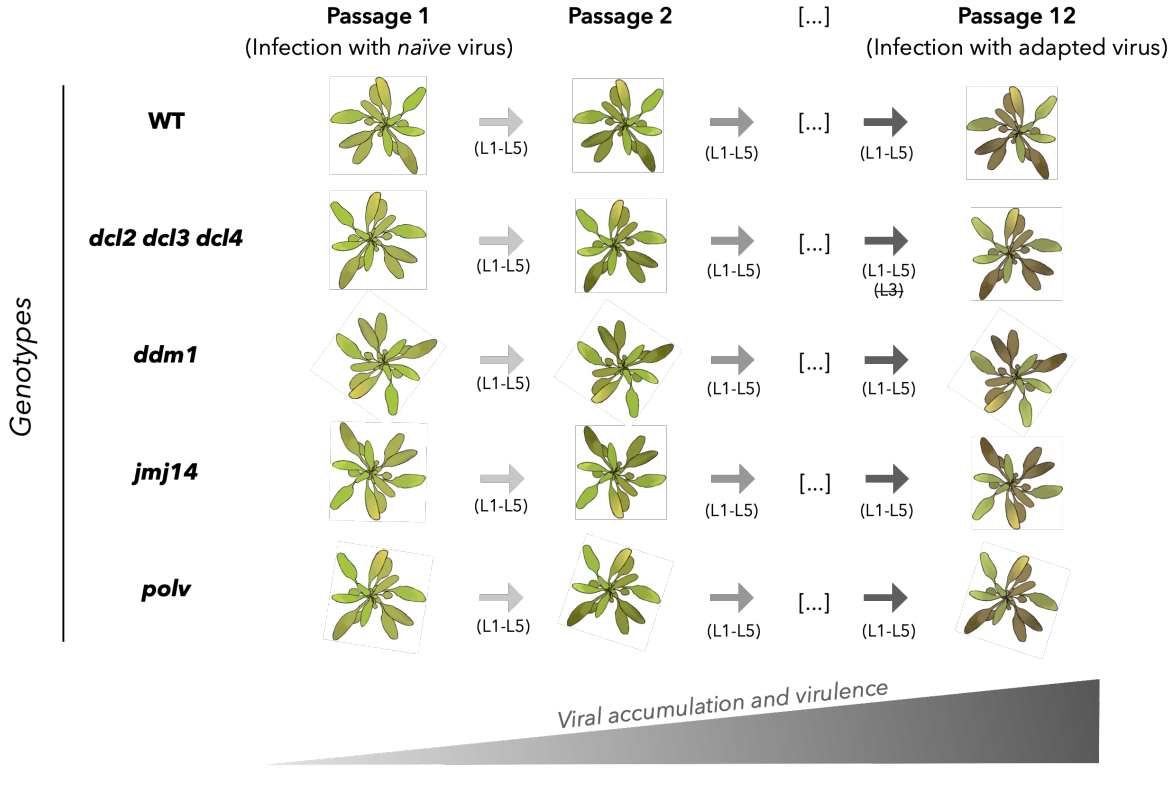
## .1 Supplementary to Chapter 3

## Supplementary Material

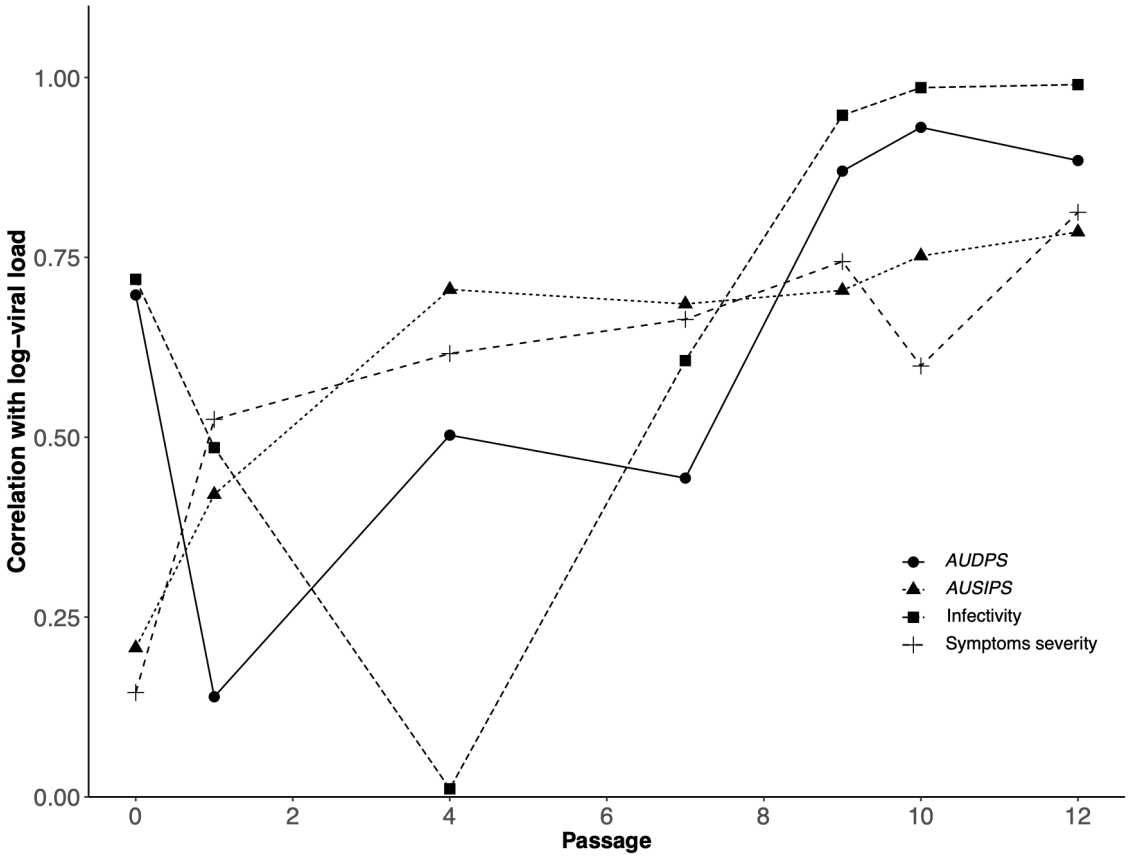
**Supplementary Figure 1**. Disease-related phenotypes measured for all 20 *A. thaliana* genotypes infected with TuMV, ordered according the four phenogroups defined in Table S1. (A) Incidence curves showing the frequency of infected plants along the 17 dpi. Infectivity values were estimated as the frequency of infection at 17 dpi; *AUDPS* values were estimated as the area under the corresponding incidence curves. (B) Symptoms development curves showing the evolution of disease severity during 14 dpi. Symptom severity was estimated as the mean values observed at 14 dpi; *AUSIPS* values are estimated as the area under the corresponding symptoms progression curves. (C) Increase in virus accumulation during the course of infections. Virus accumulation was evaluated by absolute RT-qPCR and expressed as fold-changes with respect to the corresponding values at 4 dpi. In all panels, dots represent the mean across plants and error bars represent ±1 SEM.



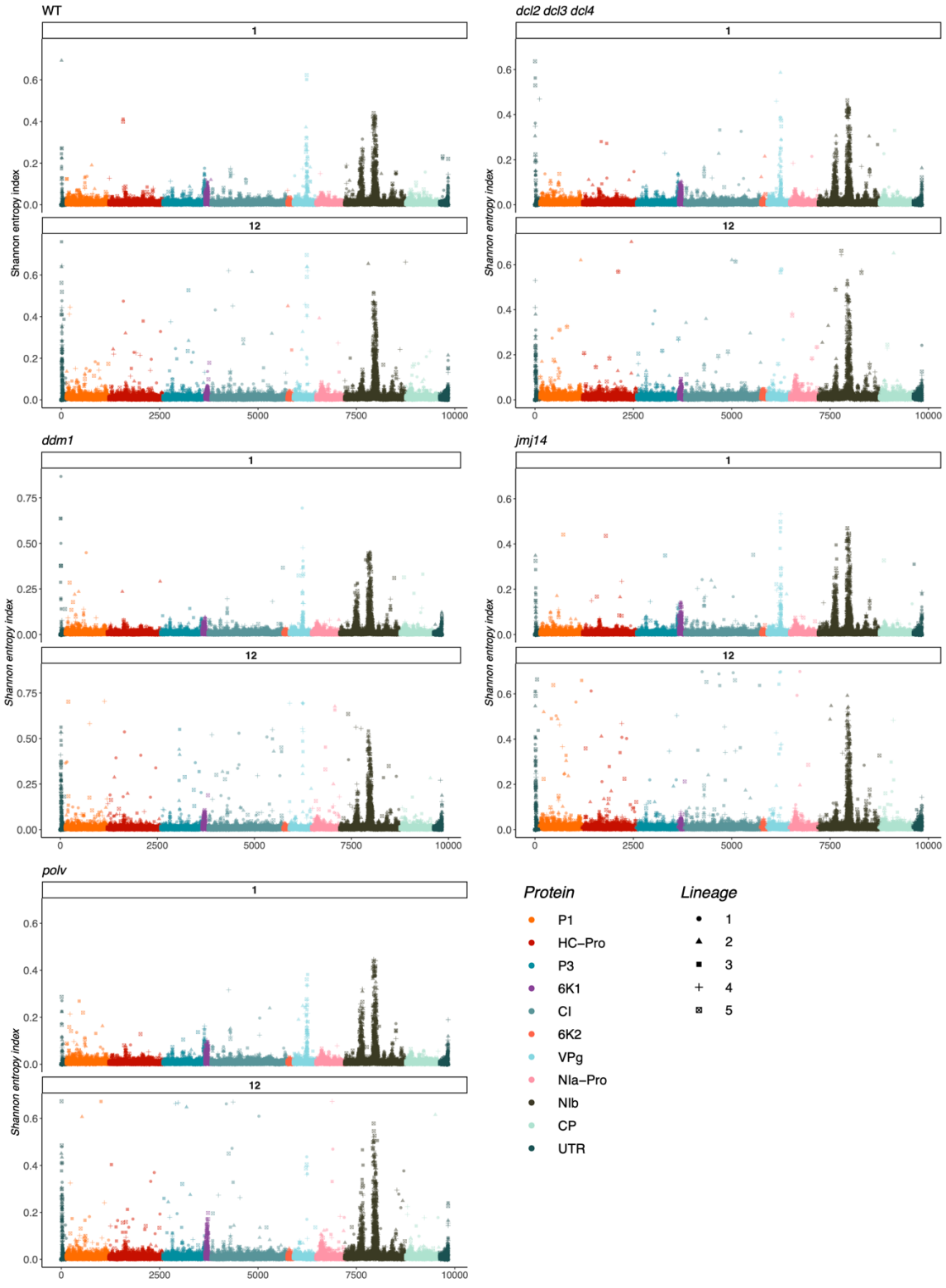
**Supplementary Figure 2.** Schematic representation of the evolution experiment.



**Supplementary Figure 3.** Evolution of the association between disease-related traits and log-viral load, measured as partial correlations controlling for plant genotype.



**Supplementary Figure 4.** Distribution of variability, measured as Shannon's entropy, along TuMV genome at passages 1 and 12. Viral cistrons are indicated with different colors and replicates with different symbols (see legend).



**Supplementary Table 1**. Evaluation of the goodness-of-fit of models with an increasing number of clusters (phenogroups).

| Number of clusters | *RSS*[1] | –Log-likelihood | *BIC*[2] | Akaike's weight |
|---|---|---|---|---|
| 2 | 0.127 | 18.555 | 42.891 | 0.007 |
| 3 | 0.093 | 14.224 | 37.119 | 0.124 |
| 4 | 0.083 | 11.194 | 33.950 | 0.602 |
| 5 | 0.037 | 11.852 | 38.155 | 0.074 |
| 6 | 0.033 | 10.219 | 37.780 | 0.089 |
| 7 | 0.032 | 8.881 | 37.994 | 0.077 |
| 8 | 0.017 | 9.225 | 41.572 | 0.013 |
| 9 | 0.016 | 8.304 | 42.621 | 0.008 |
| 10 | 0.016 | 7.474 | 43.851 | 0.004 |

[1]Residual sum of squares

[2]In all cases, $n = 20$ genotypes.

**Supplementary Table 2**. Results of the MANCOVA analysis fitting multivariate data to equation (1).

| Effect | $\Lambda$[1] | d.f.[2] | $F$ | $P$ | $\eta_P^2$ | $1-\beta$ |
|---|---|---|---|---|---|---|
| Intersection ($A$) | 0.985 | 4, 272 | 4520.645 | < 0.001 | 0.985 | 1 |
| Plant genotype ($G$) | 0.883 | 16, 831.6 | 2.168 | 0.005 | 0.031 | 0.925 |
| Passage ($t$) | 0.485 | 4, 272 | 72.212 | < 0.001 | 0.515 | 1 |
| Passage by Plant genotype by ($t \times G$) | 0.779 | 16, 831.6 | 4.426 | < 0.001 | 0.061 | 1 |
| Lineage within Plant genotype ($L(G)$) | 0.882 | 80, 1075.4 | 0.434 | 1 | 0.031 | 0.727 |
| Passage by Lineage within Plant genotype ($t \times L(G)$) | 0.431 | 80, 1075.4 | 3.193 | < 0.001 | 0.190 | 1 |

[1]Wilk's $\Lambda$.

[2]numerator, denominator

**Supplementary Table 3**. Results of the ANCOVA analysis fitting log-viral load data to equation (1).

| Effect | $SS$[1] | d.f. | $F$ | $P$ | $\eta_P^2$ | $1-\beta$ |
|---|---|---|---|---|---|---|
| Intersection ($A$) | 2256.775 | 1 | 229435.017 | < 0.001 | 1 | 1 |
| Plant genotype ($G$) | 0.684 | 4 | 17.390 | < 0.01 | 0.777 | 1 |
| Passage ($t$) | 9.988 | 1 | 48.286 | < 0.0001 | 0.279 | 1 |
| Passage by Plant genotype by ($t \times G$) | 4.526 | 4 | 5.470 | 0.004 | 0.149 | 0.972 |
| Lineage within Plant genotype ($L(G)$) | 0.197 | 20 | 0.048 | 1 | 0.008 | 0.068 |
| Passage by Lineage within Plant genotype ($t \times L(G)$) | 31.642 | 20 | 7.649 | < 0.001 | 0.550 | 1 |
| Error | 25.856 | 125 | | | | |

[1]Type III sum of squares.

**Supplementary Table 4**. Results of the MANOVA analysis fitting log-viral load data to equation (3).

| Effect | $\Lambda$[1] | d.f.[2] | $F$ | $P$ | $\eta_P^2$ | $1-\beta$ |
|---|---|---|---|---|---|---|
| Intersection ($v$) | 0.011 | 5, 16 | 2798.767 | < 0.001 | 0.989 | 1 |
| Epigenetic pathway ($P$) | 0.189 | 10, 32 | 4.155 | 0.010 | 0.565 | 0.988 |
| Plant genotype within pathway ($G(P)$) | 0.387 | 10, 32 | 1.944 | 0.075 | 0.378 | 0.757 |

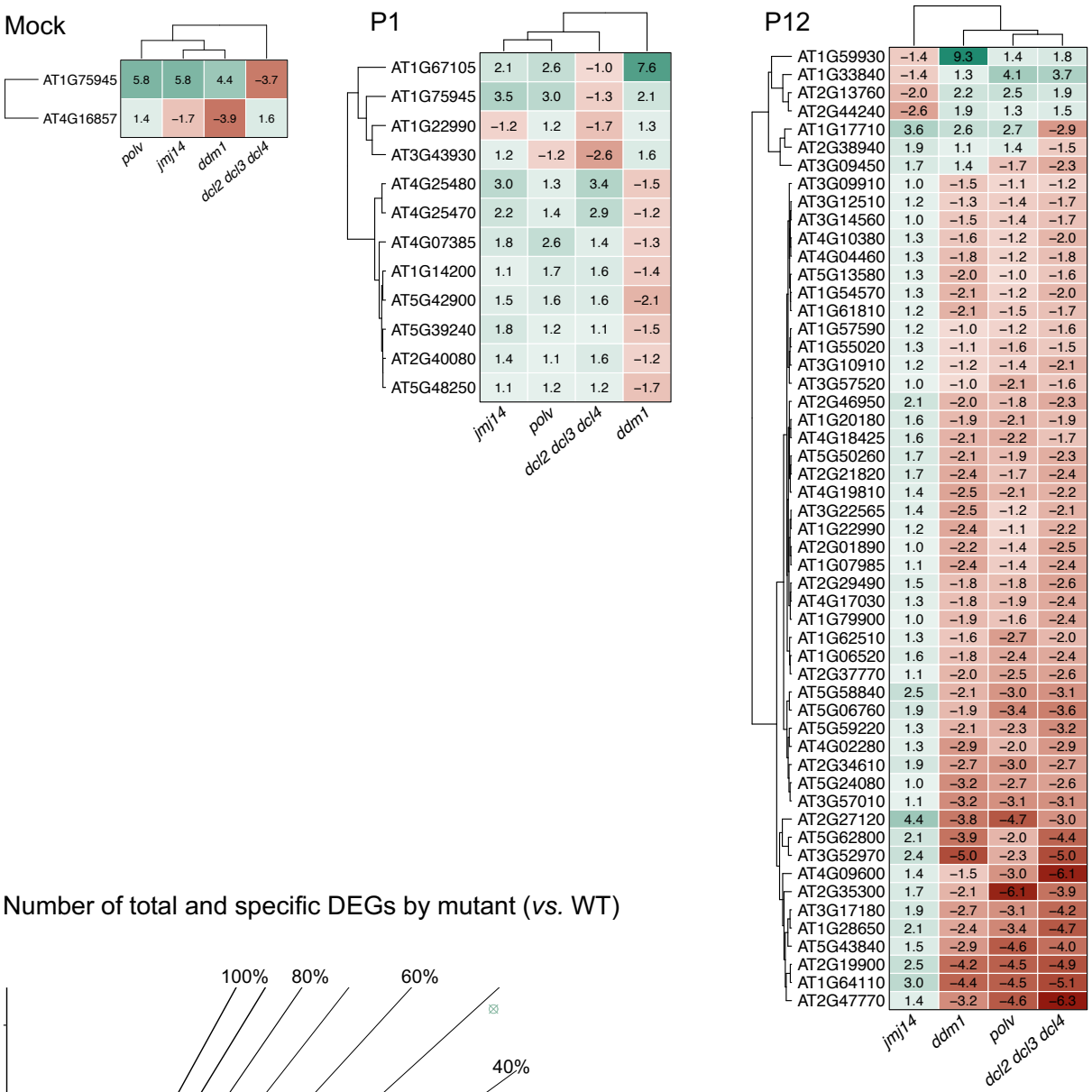
[1]Wilk's $\Lambda$.

[2]numerator, denominator.

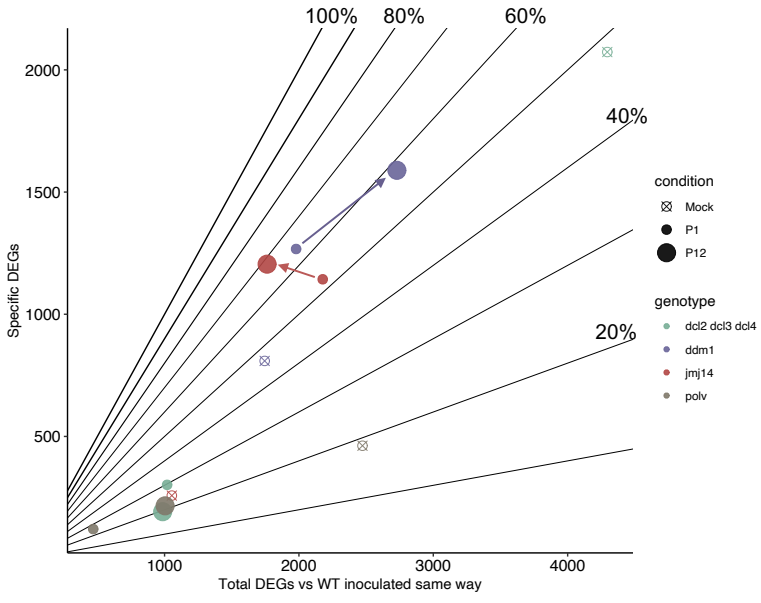
35

## .2 Supplementary to Chapter 4

A Common DEGs (mutant *vs* WT) with opposite regulation across mutants

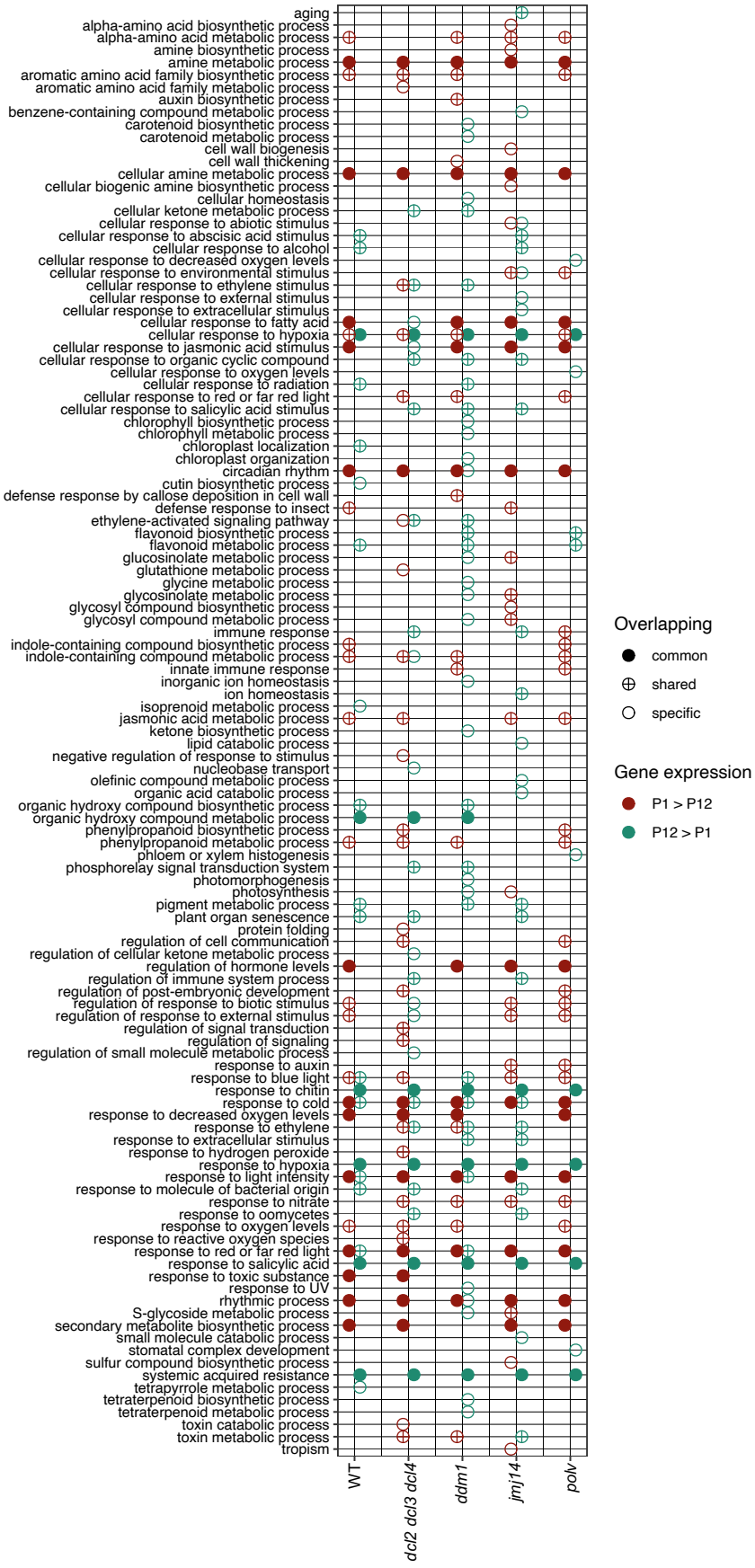



B Number of total and specific DEGs by mutant (*vs.* WT)




Supplementary Fig. S1

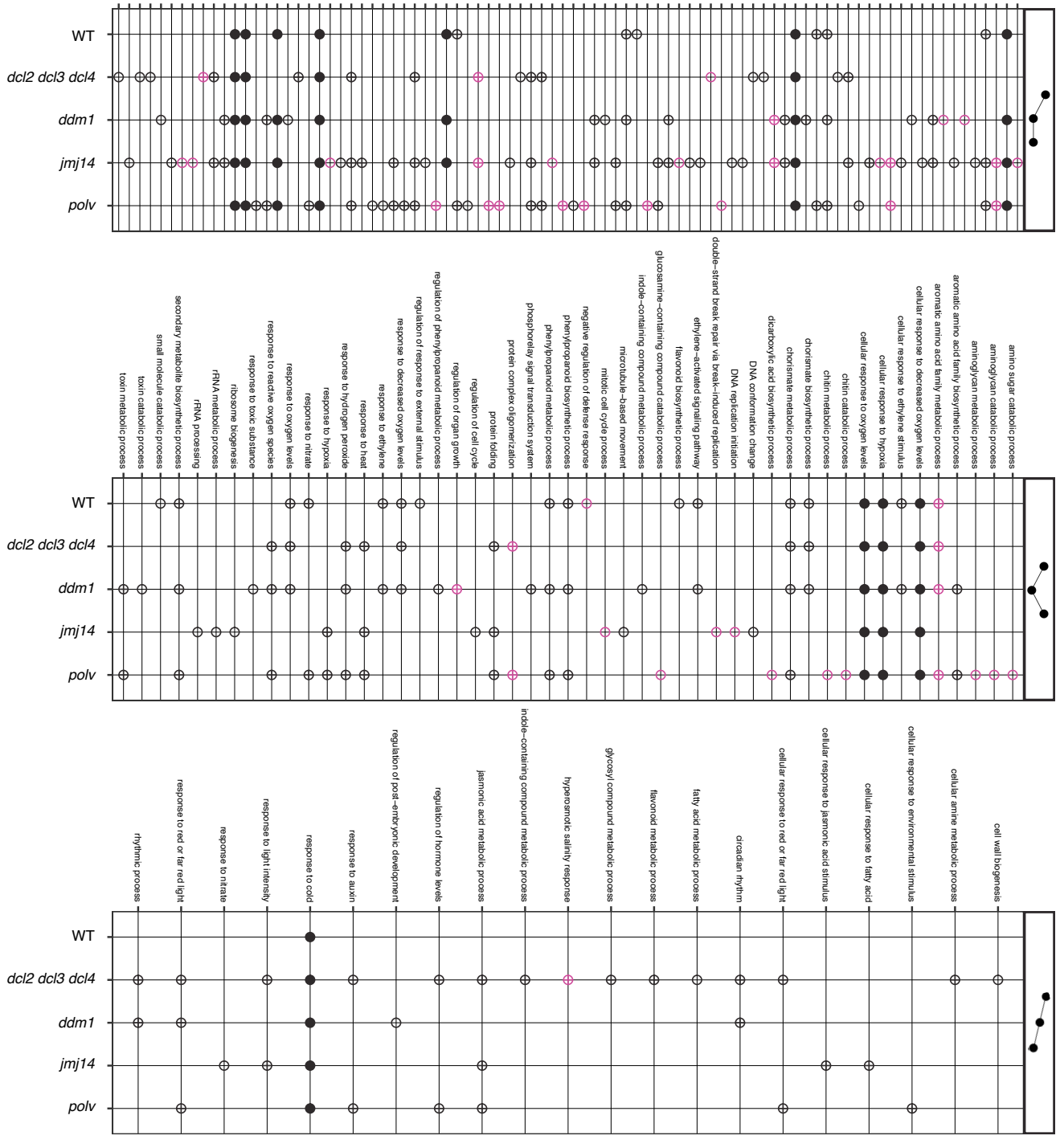
Functional enrichment analyses. GO: Biological processes separated by the direction of the change and its genotype specificity (Bonferroni-Hochberg adjusted $P < 0.001$).




Supplementary Fig. S2

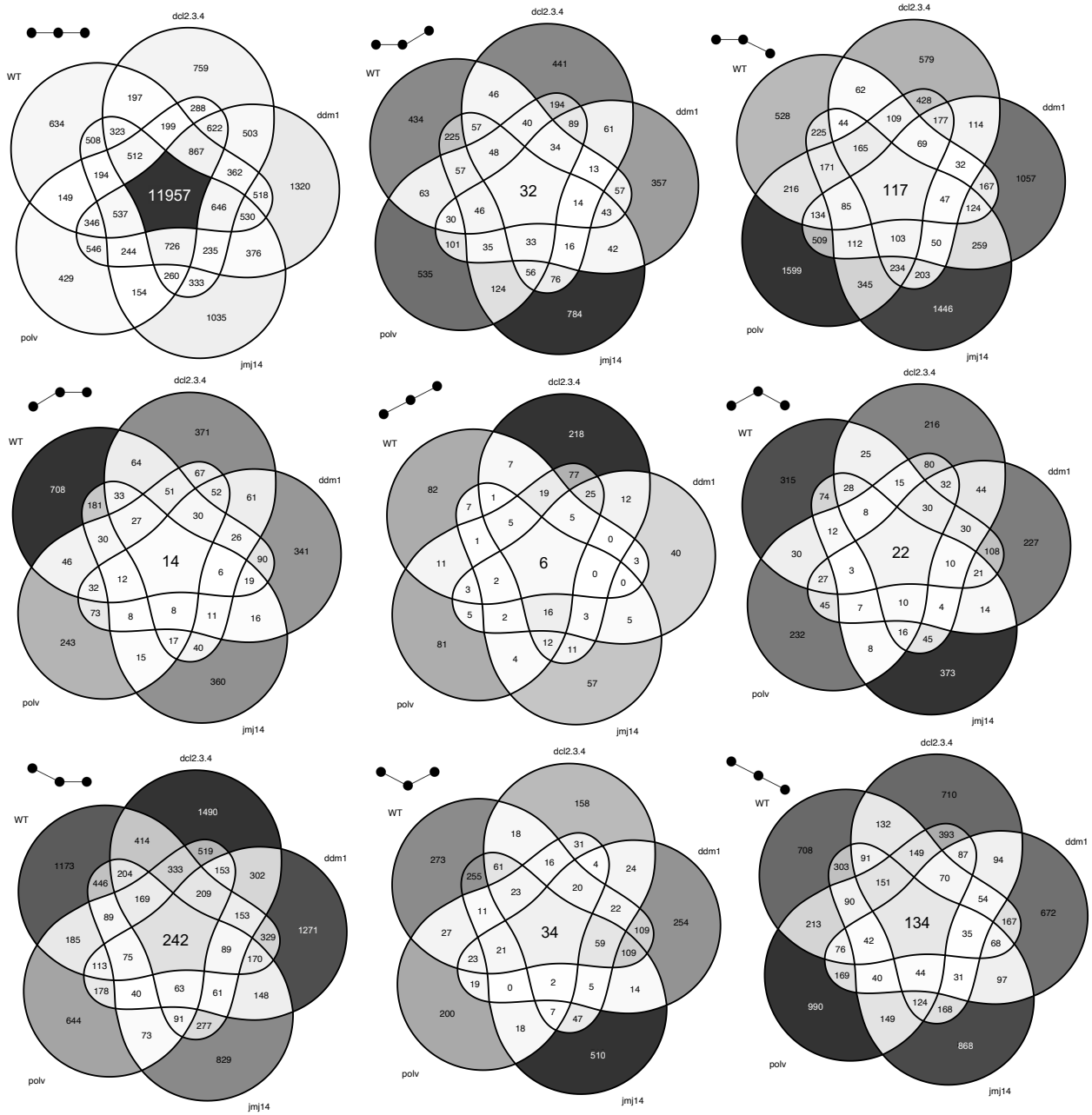
Supplementary Fig. S3

Supplementary Fig. S3 (cont.)

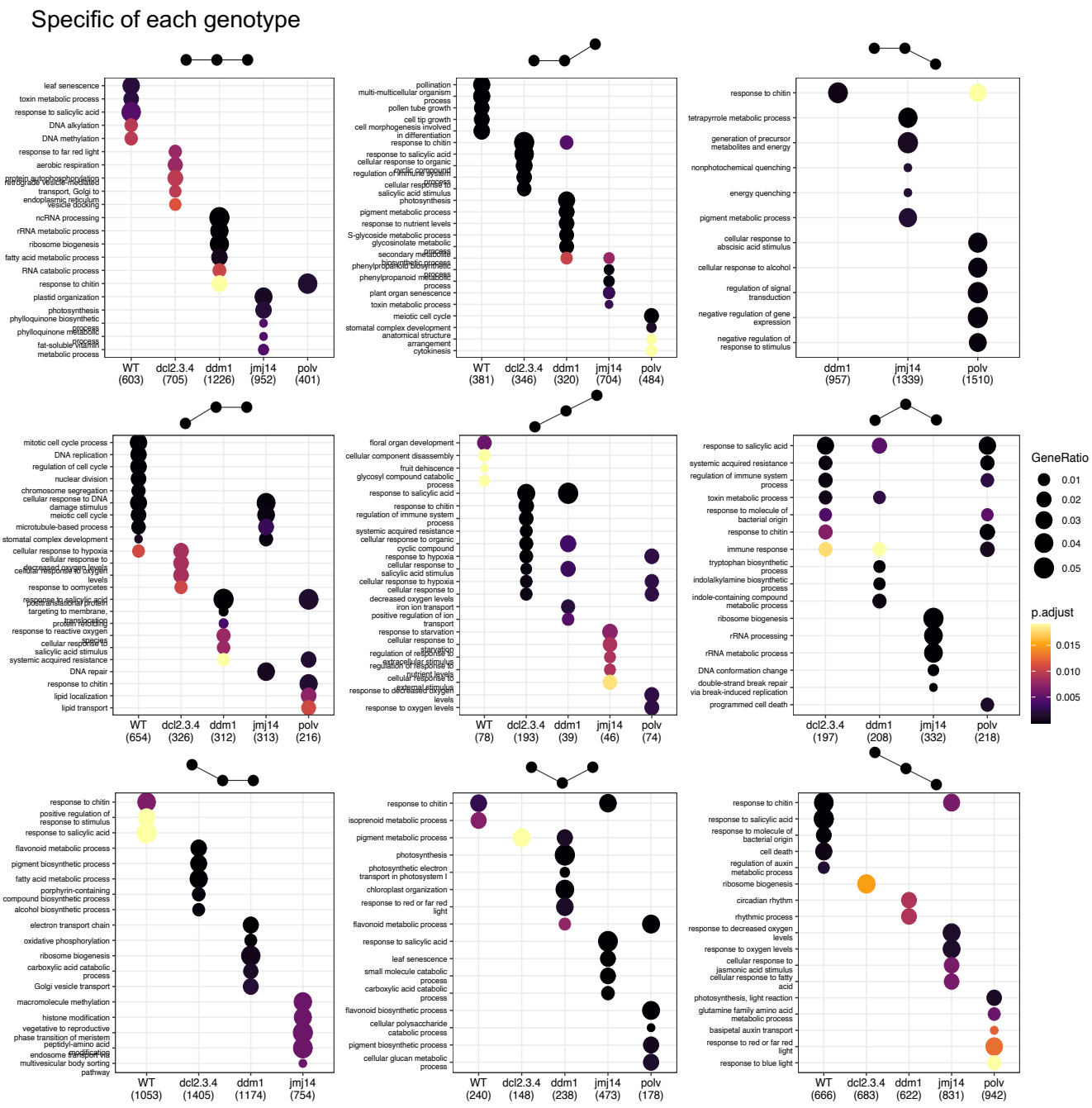



Supplementary Fig. S3 (cont.)

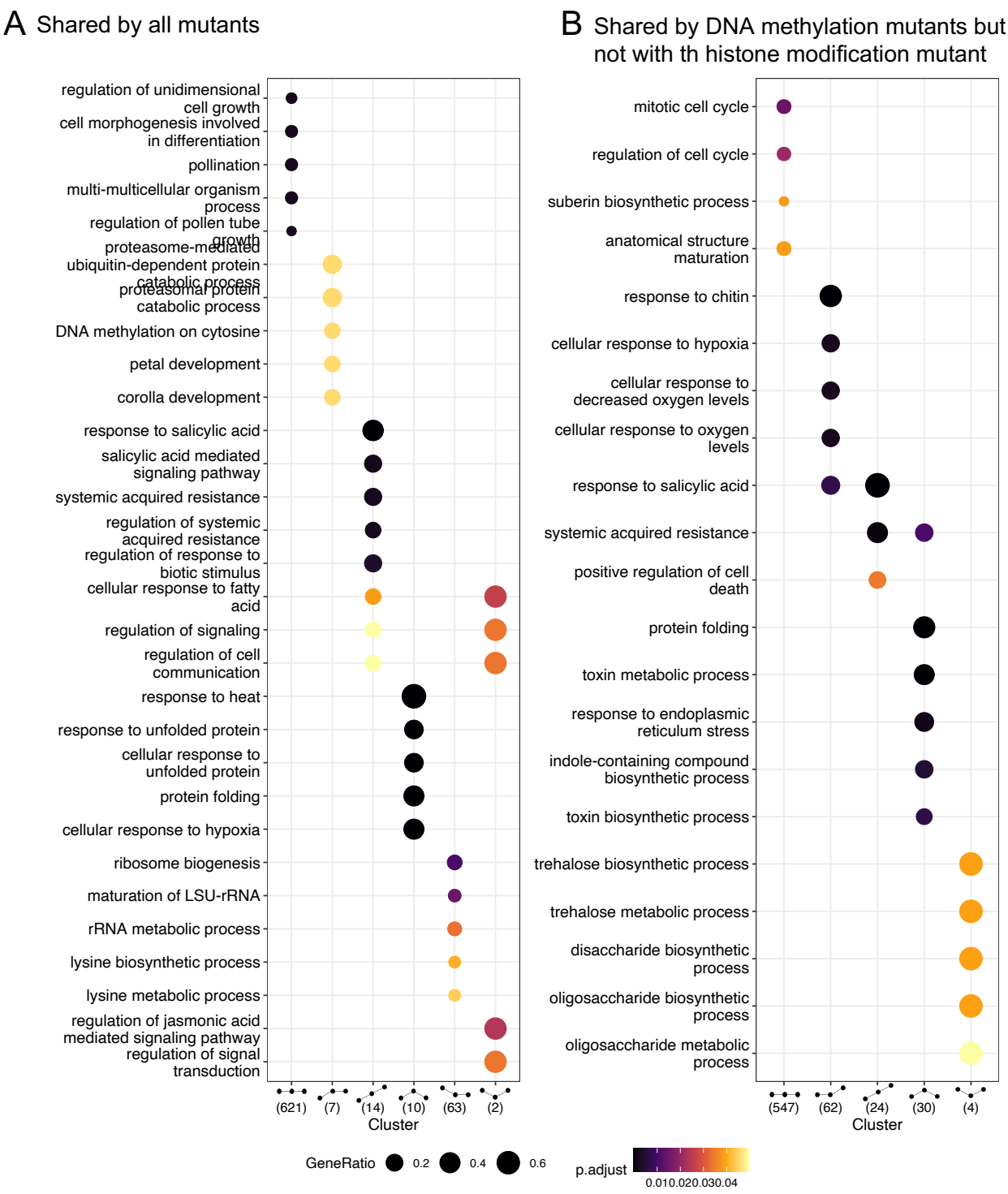
Shared genes per profiles and genotype



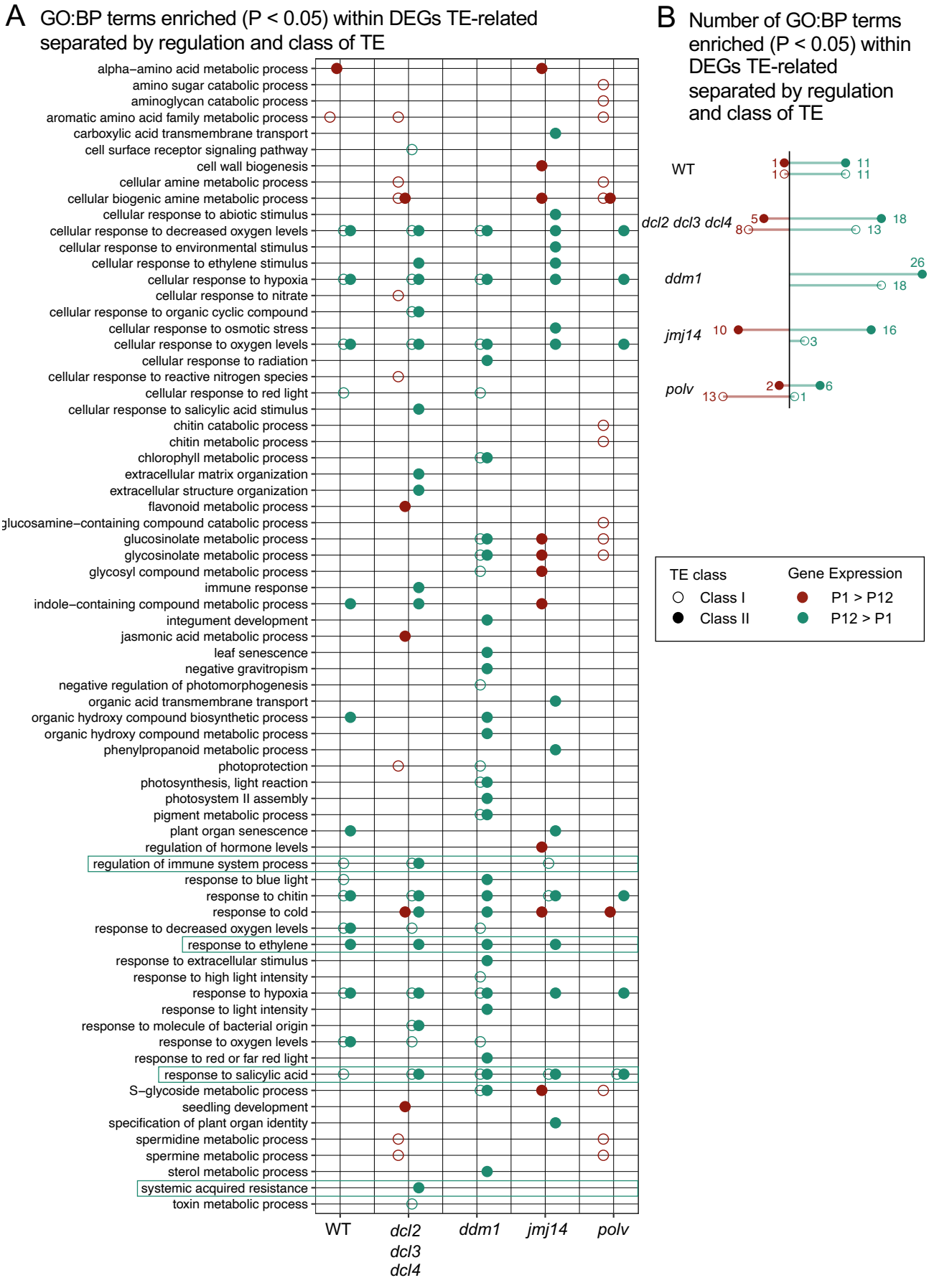

Supplementary Fig. S4



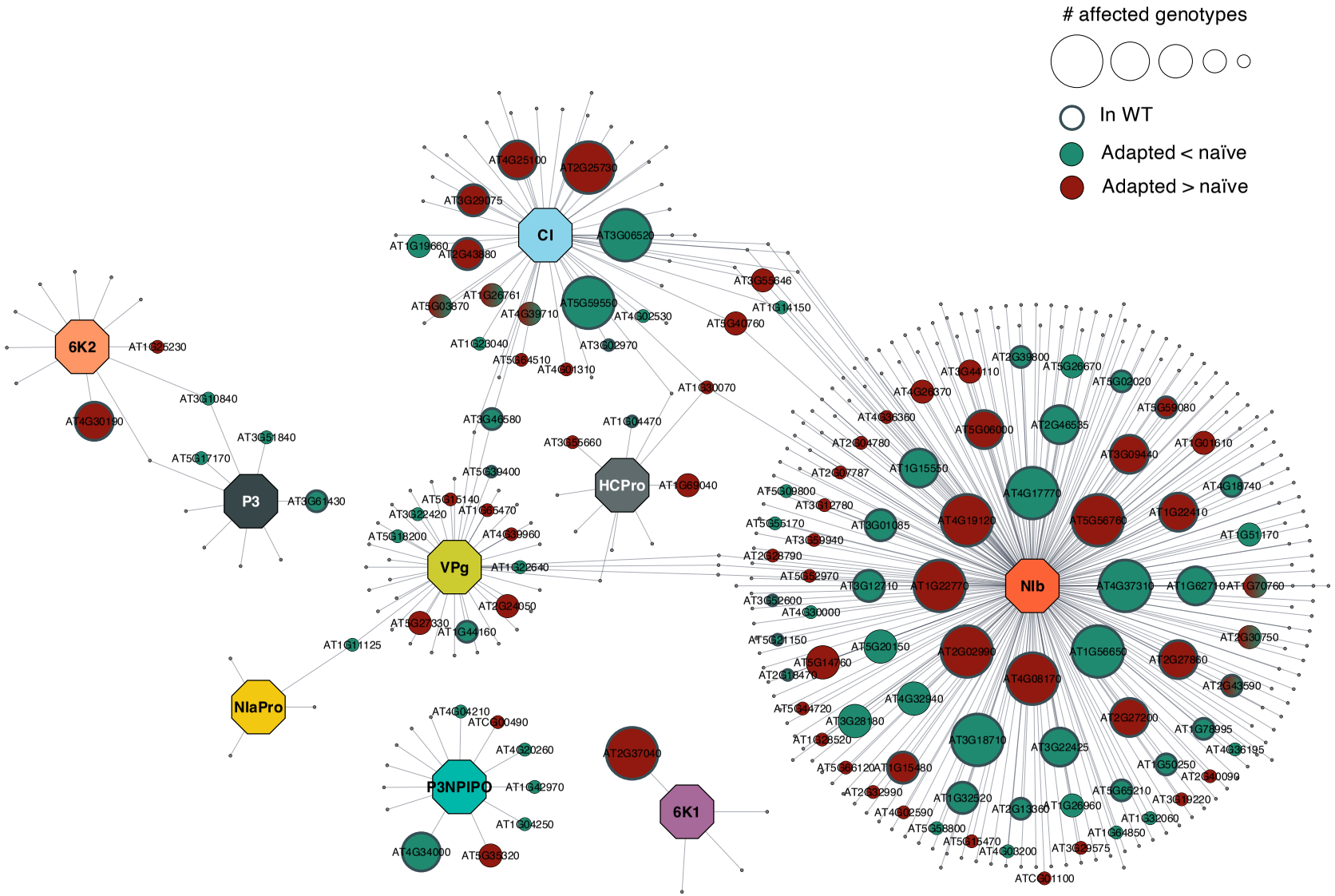

Supplemenary Fig. S5




Supplementary Fig. S6




Supplementary Fig. S7




Supplementary Fig. S8

**Supplementary Table S1.** Genes with a differential gene expression in all the mutants respect WT under the same inoculation conditions.

| Conditions intersection | Gene (common name) | Description/Comments |
|---|---|---|
| $\mathbf{P12 \cap P1 \cap Mock}$ | AT1G53480 (*T3F20.20*) | *METHIONINE OVERACCUMULATION 1 (MTO1) RESPONDING DOWN 1* (MRD1) gene, and encodes for a protein involved in down-regulation of methionine biosynthesis, and is essential for salicylic acid-mediated defense (Blevins et al. 2014 ) |
| | AT2G01422 | Noncoding RNA unknown function. found in experiments seeking for candidate genes directly downregulated by RdDM-related siRNAs (Kurihara et al. 2008) |
| | AT3G28899 | Encodes for a member of the ribosomal protein L34e superfamily protein that has was already found in experiments seeking for candidate genes directly downregulated by RdDM-related siRNAs (Kurihara et al. 2008) |
| | AT4G04221 | Antisense long-noncoding RNA. negatively regulates the expression of *LEUCINE RICH-REPEART (LRR) RECEPTOR-LIKE PROTEIN 46* (*RLP46*), a plasma-membrane located receptor involved in tissue differentiation, plant development and senescence that has been shown to be upregulated after infection (Steidele and Stam 2021) |
| | AT4G04223 | Noncoding primary transcript of unknown function |
| | AT5G24240 (*PI4KC3*) | Ubiquitin family protein involved in regulation of flower development and responses to abscisic acid (Akhter et al. 2015) |
| $\mathbf{P12 \cap P1 \cap Mock}$ | AT1G22990 (*HIPP22*) | *HEAVY METAL ASSOCIATED ISOPRENYLATED PLANT PROTEIN 2*. HIPP metallochaperone proteins are involved in metal homeostasis and detoxification but also in transcriptional responses to temperature and drought and plant-fungi interactions (De Abreu-Neto et al. 2013) |
| | AT1G53490 (*HEI10*) | *ENHANCER OF CELL INVASION 10*. RING finger-containing protein of the ZMM class required for class I crossover formation during meiosis (Serra et al. 2018) |
| | AT3G09450 | Encodes for a member of the fusaric acid resistance family protein. This family of proteins is involved in detoxification of fusaric acid, a secondary metabolite produced during fungal infections (Bouizgarne et al. 2005). |
| | AT4G14140 (*MET2*) | *DNA METHYLTRANSFERASE 2*. Essential component of the DNA methylation that promotes gene silencing in response to genome defense against transposons and DNA viruses. Geminiviruses' Rep protein, for example, actively repress the expression of DNA methyltransferases *MET1* and *CHROMOMETHYLASE 3* (*CMT3*) in locally and systemically infected tissues (Rodríguez-Negrete et al. 2013). |
| | AT4G15390 (*DLI3740c*) | Encodes for a HXXXD-type BAHD acyltransferase family protein. BAHD acyltransferases are often involved in the production of phenolic secondary metabolites, and particularly in brassinosteroids homeostasis (Zhang and Xu 2018). Brassinosteroids play key roles in plant adaptation to infections (Xiong et al. 2022). |
| $\mathbf{Mock \cap P1 - P1}$ | AT1G75945 | Mitochondrial membrane-associated hypothetical proteins |
| | AT2G33175 | Mitochondrial membrane-associated hypothetical proteins |


Akhter, S., Uddin, M.N., Jeong, I.S., Kim, D.W., Liu, X.M., & Bahk, J.D. (2015) Role of AtPI4Kgamma3, a type II phosphoinositide 4-kinase, in abiotic stress responses and floral transition. *Plant Biotechnol. J.*, **14**, 215-230.

Blevins, T., Pontvianne, F., Cocklin, R., Podicheti, R., Chandrasekhara, C., Yerneni, S., et al. (2014) A two-step process for epigenetic inheritance in Arabidopsis. *Mol. Cell.*, **54**, 30-42.

Bouizgarne, B., El-Maarouf-Bouteau, H., Frankart, C., Reboutier, D., Madiona, K., Pennarun, A.M., et al. (2005) Early physiological response of *Arabidopsis thaliana* cells to fusaric acid: toxic and signalling effects. *New Phytol.*, **169**, 209-218.

De Abreu-Neto, J.B., Turchetto-Zolet, A.C., Valter de Oliveira, L.F., Bodanese Zanettini, M.H., & Margis-Pinheiro, M. (2013) Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *FEBS J.*, **280**, 1604-1616.

Kurihara, Y., Matsui, A., Kawashima, M., Kaminuma, E., Ishida, J., Morosawa, T., et al. (2008) Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. *Biochem. Biophys. Res. Commun.*, **376**, 553-557.

Rodríguez-Negrete, E., Lozano-Durán, R., Piedra-Aguilera, A., Cruzado, L., Bejarano, E.R., & Castillo, A.G. (2013) Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. *New Phytol.*, **199**, 464-475.

Serra, H., Lambing, C., Griffin, C.H., Topp, S.D., Nageswaran D.C., Underwood, C.J., et al. (2018) Massive crossover elevation via recombination of *HEI10* and *recq4a recq4b* during Arabidopsis meiosis. *Proc. Natl. Acad. Sci. USA*, **115**, 2437-2442.

Steidele, C.E., & Stam, R. (2021) Multi-omics approach highlights differences between RLP classes in *Arabidopsis thaliana. BMC Genomics*, **22**, 557.

Xiong, J., Wan, X., Ran, M., Xu, X., Chen, L., & Yang, F. (2022) Brassinosteroids positively regulate plant immunity via BRI1-EMS-SUPPRESSOR 1-mediated *GLUCAN SYNTHASE-LIKE 8* transcription. *Front. Plant Sci.*, **13**, 854899.

Zhang, Z., & Xu, L. (2018) Arabidopsis BRASSINOSTEROID INACTIVATOR 2 is a typical BAHD acyltransferase involved in brassinosteroid homeostasis. *J. Exp. Bot.*, **69**, 1925-1941.

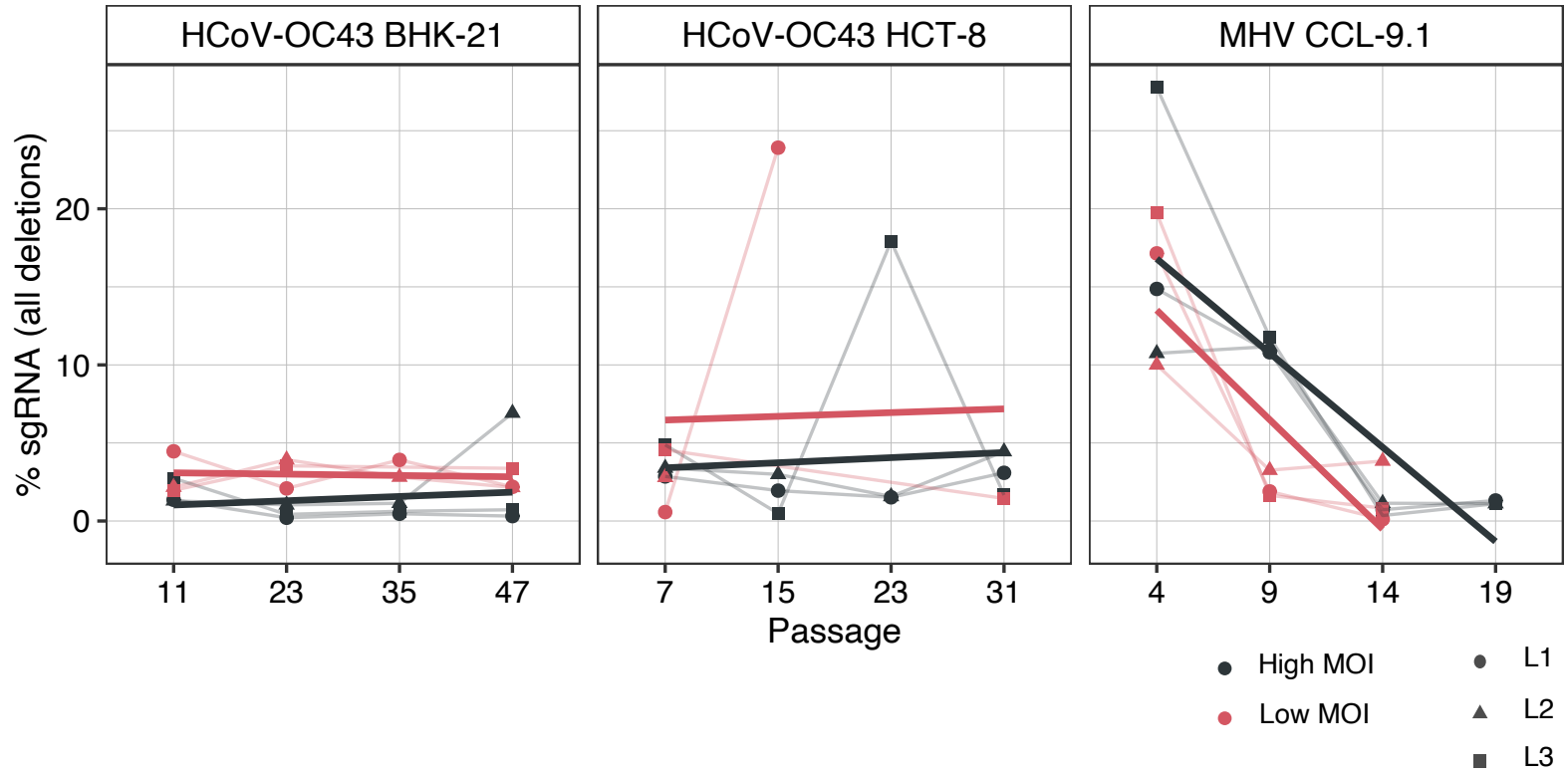
**Supplementary Table S2.** Genes with opposed regulation within the core of DEGs involved in TuMV adaptation to *A. thaliana*

| Locus | Gene name | Regulation adapted virus (activated, repressed) | Description |
|---|---|---|---|
| AT1G69480 | *PHO1-H10* | WT<br>*dcl2 dcl3 dcl4*<br>*ddm1*<br>*jmj14*<br>*polv* | Phosphate transporter PHO1 homolog 10; May transport inorganic phosphate (Pi); Belongs to the SYG1 (TC 2.A.94) family. |
| AT2G40610 | *EXPA8* | WT<br>*dcl2 dcl3 dcl4*<br>*ddm1*<br>*jmj14*<br>*polv* | Expansin-A8; Causes loosening and extension of plant cell walls by disrupting non-covalent bonding between cellulose microfibrils and matrix glucans. No enzymatic activity has been found (By similarity). Belongs to the expansin family. Expansin A subfamily. |
| AT4G09820 | *TT8* | WT<br>*dcl2 dcl3 dcl4*<br>*ddm1*<br>*jmj14*<br>*polv* | Transcription factor TT8; Transcription activator, when associated with MYB75/PAP1 or MYB90/PAP2. Involved in the control of flavonoid pigmentation. Plays a key role in regulating leucoanthocyanidin reductase (BANYULS) and dihydroflavonol-4-reductase (DFR). Not required for leucoanthocyanidin dioxygenase (LDOX) expression. |
| AT4G25480 | *DREB1A* | WT<br>*dcl2 dcl3 dcl4*<br>*ddm1*<br>*jmj14*<br>*polv* | Dehydration-responsive element-binding protein 1A; Transcriptional activator that binds specifically to the DNA sequence 5'-[AG]CCGAC-3'. Binding to the C-repeat/DRE element mediates cold-inducible transcription. CBF/DREB1 factors play a key role in freezing tolerance and cold acclimation; Belongs to the AP2/ERF transcription factor family. ERF subfamily. |
| AT5G65730 | *XTH6* | WT<br>*dcl2 dcl3 dcl4*<br>*ddm1*<br>*jmj14*<br>*polv* | Probable xyloglucan endotransglucosylase/hydrolase protein 6; Catalyzes xyloglucan endohydrolysis (XEH) and/or endotransglycosylation (XET). Cleaves and religates xyloglucan polymers, an essential constituent of the primary cell wall, and thereby participates in cell wall construction of growing tissues (By similarity); Belongs to the glycosyl hydrolase 16 family. XTH group 1 subfamily. |

**Supplementary Table S3.** Genes of the PPI network affected by viral adaptation in all plant genotypes.

| Viral protein | Plant protein | Common name | Description |
|---|---|---|---|
| **6K1** | AT2G37040 | PAL1 | Phenylalanine ammonia-lyase 1; This is a key enzyme of plant metabolism catalyzing the first reaction in the biosynthesis from L-phenylalanine of a wide variety of natural products based on the phenylpropane skeleton; Belongs to the PAL/histidase family. |
| **CI** | AT2G25730 | F3N11.1 | Zinc finger FYVE domain protein. |
| | AT3G06520 | F5E6.15 | Agenet domain-containing protein. |
| | AT5G59550 | DURF2 | E3 ubiquitin-protein ligase RDUF2; E3 ubiquitin-protein ligase involved in the positive regulation of abscisic acid-dependent drought stress responses. Possesses E3 ubiquitin ligase activity in vitro. |
| **NIb** | AT1G22770 | GI | Protein GIGANTEA; Involved in regulation of circadian rhythm and photoperiodic flowering. May play a role in maintenance of circadian amplitude and period length. Is involved in phytochrome B signaling. Stabilizes ADO3 and the circadian photoreceptor ADO1/ZTL. Regulates 'CONSTANS' (CO) in the long-day flowering pathway by modulating the ADO3-dependent protein stability of CDF1 and CDF2, but is not essential to activate CO transcription. Regulates, via the microRNA miR172, a CO-independent pathway that promotes photoperiodic flowering by inducing 'FLOWERING LOCUS T'. |
| | AT1G56650 | MYB75 | Transcription factor MYB75; Transcription activator, when associated with BHLH12/MYC1, EGL3, or GL3. Promotes the synthesis of phenylpropanoid-derived compounds such as anthocyanins and proanthocyanidin, probably together with GL3 and BHLH2. Regulates the expression of CHS, DFRA, LDOX, and BAN. |
| | AT2G02990 | RNS1 | Ribonuclease 1; May remobilize phosphate, particularly when cells senesce or when phosphate becomes limiting. |
| | AT3G18710 | PUB29 | U-box domain-containing protein 29; Functions as an E3 ubiquitin ligase. |
| | AT4G08170 | ITPK3 | Inositol-tetrakisphosphate 1-kinase 3; Kinase that can phosphorylate various inositol polyphosphate such as Ins(3,4,5,6)P4 or Ins(1,3,4)P3. Phosphorylates Ins(3,4,5,6)P4 to form InsP5. This reaction is thought to have regulatory importance, since Ins(3,4,5,6)P4 is an inhibitor of plasma membrane $Ca^{(2+)}$-activated $Cl^{(-)}$ channels, while Ins(1,3,4,5,6)P5 is not (By similarity). Also phosphorylates Ins(1,3,4)P3 or a racemic mixture of Ins(1,4,6)P3 and Ins(3,4,6)P3 to form InsP4. Ins(1,3,4,6)P4 is an essential molecule in the hexakisphosphate (InsP6) pathway (By similarity). |
| | AT4G17770 | TPS5 | Alpha,alpha-trehalose-phosphate synthase [UDP-forming] 5; In the N-terminal section; belongs to the glycosyltransferase 20 family. |
| | AT4G19120 | ERD3 | Probable methyltransferase PMT21; Belongs to the methyltransferase superfamily. |
| | AT4G37310 | C7A10.50 | Cytochrome P450, family 81, subfamily H, polypeptide 1; Belongs to the cytochrome P450 family. |
| | AT5G56760 | SAT5 | Serine acetyltransferase 5; Belongs to the transferase hexapeptide repeat family. |

## .3 Supplementary to Chapter 5


HCoV-OC43 BHK-21
HCoV-OC43 HCT-8
MHV CCL-9.1
% sgRNA (all deletions)
0
10
20
11
23
35
47
7
15
23
31
4
9
14
19
Passage
High MOI
Low MOI
L1
L2
L3

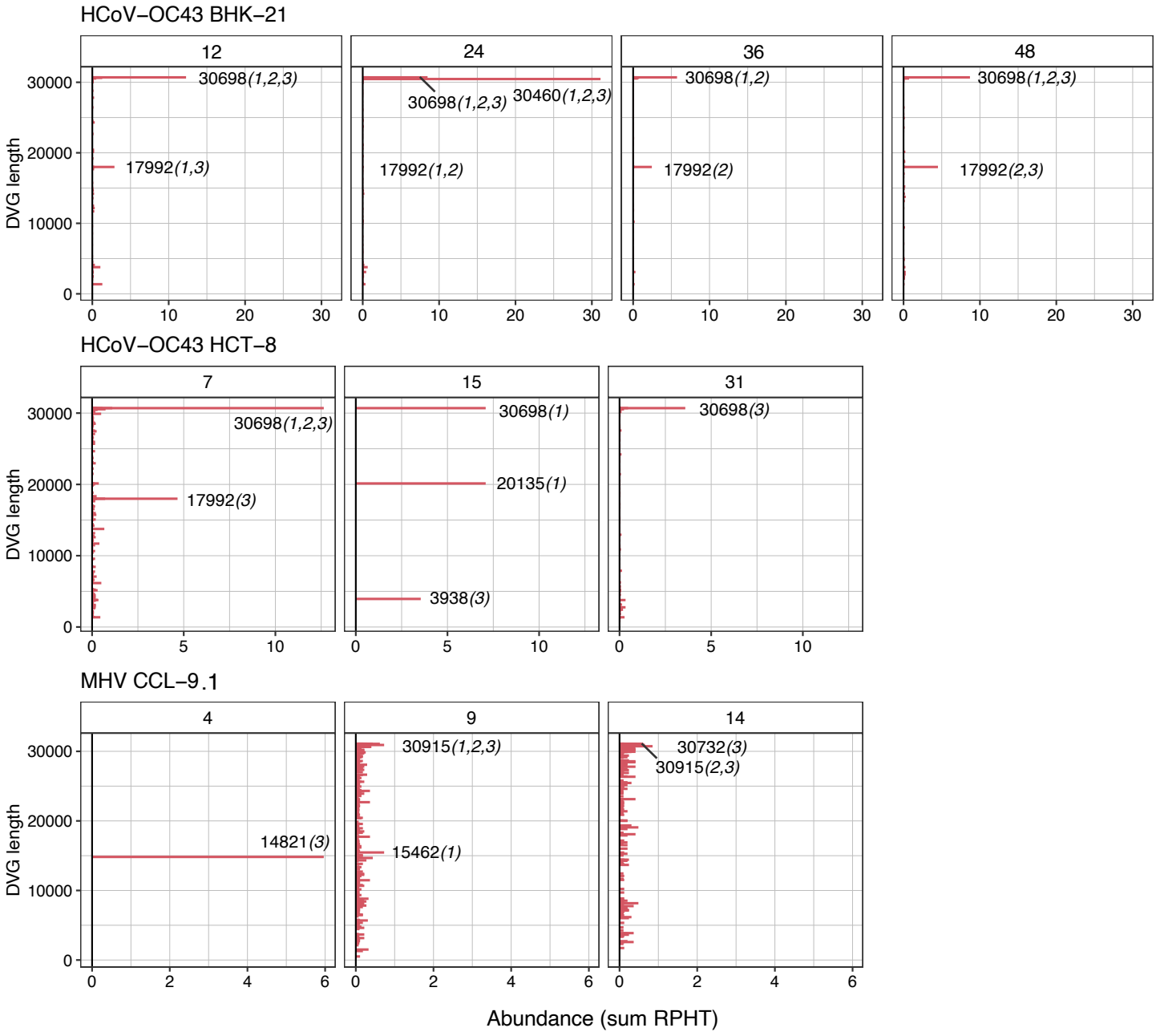

HCoV−OC43 BHK−21
12
24
36
48
30698(1,2,3)
17992(1,3)
30698(1,2,3)
30460(1,2,3)
17992(1,2)
30698(1,2)
17992(2)
30698(1,2,3)
17992(2,3)
HCoV−OC43 HCT−8
7
15
31
30698(1,2,3)
17992(3)
30698(1)
20135(1)
3938(3)
30698(3)
MHV CCL−9.1
4
9
14
14821(3)
30915(1,2,3)
15462(1)
30732(3)
30915(2,3)
DVG length
Abundance (sum RPHT)

**Table S1.** Statistics describing the variation in MOI for each experimental evolution lineage. Median values ±IQR are shown. Last column shows the fold difference between high and low MOIs (averaging lineages).

| Virus | Cells | Lineage | Low MOI | High MOI | $Log_2$-fold |
|---|---|---|---|---|---|
| HCoV-OC43 | BHK-21 | 1 | $3.74\times10^{-4}$ ±$2.50\times10^{-2}$ | 21.25 ±82.13 | 16.0 |
| | | 2 | $7.13\times10^{-4}$ ±$2.50\times10^{-2}$ | 31.25 ±208.50 | |
| | | 3 | $1.50\times10^{-4}$ ±$1.56\times10^{-2}$ | 25.00 ±145.75 | |
| | HTC-8 | 1 | $5.81\times10^{-5}$ ±$2.50\times10^{-3}$ | 0.88 ±4.76 | 12.6 |
| | | 2 | $1.21\times10^{-4}$ ±$3.37\times10^{-3}$ | 0.63 ±2.22 | |
| | | 3 | $3.00\times10^{-3}$ ±$2.96\times10^{-3}$ | 0.75 ±5.07 | |
| MHV | CCL-9.1 | 1 | $2.65\times10^{-6}$ ±$8.26\times10^{-6}$ | $6.96\times10^{-3}$ ±0.28 | 12.0 |
| | | 2 | $2.59\times10^{-6}$ ±$1.79\times10^{-5}$ | $1.06\times10^{-2}$ ±$9.02\times10^{-2}$ | |
| | | 3 | $8.69\times10^{-6}$ ±$1.10\times10^{-4}$ | $1.09\times10^{-2}$ ±$6.62\times10^{-2}$ | |

**Table S2.** Fitting of viral load data to the best-fitting ARIMA(*p, d, q*)[1] model based on the minimum BIC criterium. Errors represent ±1 SE.

| Virus | Cells | MOI | Lineage | ARIMA | BIC | *Intercept* | Slope | Moving average | Autocorrelation 1 | Autocorrelation 2 |
|---|---|---|---|---|---|---|---|---|---|---|
| HCoV-OC43 | BHK-21 | High | 1 | 0, 0, 1 | 64.54 | 8.717 ±0.143 | 0.015 ±0.005 | 0.319 ±0.118 | | |
| | | | 2 | 1, 0, 0 | 142.25 | 11.779 ±0.386 | 0.025 ±0.013 | | 0.394 ±0.138 | |
| | | | 3 | 0, 0, 0 | 232.22 | 18.415 ±0.612 | 0.041 ±0.022 | | | |
| | | Low | 1 | 1, 0, 0 | 546.11 | | −0.820 ±1.228 | | 0.821 ±0.076 | |
| | | | 2 | 2, 0, 1 | 143.65 | 13.336 ±0.465 | 0.017 ±0.016 | 0.565 ±0.161 | −0.302 ±0.176 | 0.565 ±0.136 |
| | | | 3 | 2, 0, 0 | 243.91 | 18.850 ±1.352 | 0.043 ±41.346 | | 0.268 ±0.268 | 0.284 ±0.146 |
| | HCT-8 | High | 1 | 0, 0, 0 | 631.98 | | −48.459 ±41.346 | | | |
| | | | 2 | 0, 0, 0 | 104.77 | 14.737 ±0.365 | −0.097 ±0.020 | | | |
| | | | 3 | 0, 0, 0 | −22.13 | 6.895 ±0.050 | −0.018 ±0.003 | | | |
| | | Low | 1 | 0, 0, 1 | 640.54 | | −635.858 ±74.643 | 0.720 ±0.137 | | |
| | | | 2 | 1, 0, 0 | 346.36 | | −6.764 ±0.807 | | 0.462 ±0.462 | |
| | | | 3 | 2, 0, 0 | 374.66 | | −4.469 ±2.255 | | 0.425 ±0.425 | 0.335 ±0.183 |
| MHV | CCL-9.1 | High | 1 | 0, 0, 0 | 66.58 | 7.8304 ±0.367 | −0.036 ±0.030 | | | |
| | | | 2 | 0, 0, 0 | 33.85 | 5.584 ±0.174 | 0.007 ±0.014 | | | |
| | | | 3 | 0, 0, 0 | −15.11 | 3.793 ±0.002 | 0.057 ±0.005 | | | |
| | | Low | 1 | 1, 0, 0 | 446.34 | | −789.109 ±330.142 | 0.781 ±0.130 | | |
| | | | 2 | 1, 0, 0 | 446.46 | | −697.742 ±338.711 | 0.774 ±0.142 | | |
| | | | 3 | 1, 0, 0 | 446.15 | | −866.648 ±317.089 | 0.779 ±0.125 | | |

[1]ARIMA parameters: $p \geq 0$: order of the autoregressive model (number of time lags); $d \geq 0$: is the differencing order; $q \geq 0$: order of the moving-average model.

# .4 Supplementary to Chapter 6

Science Advances
AAAS

# Supplementary Materials for

## Story of an infection: Viral dynamics and host responses in the *Caenorhabditis elegans*–Orsay virus pathosystem

Victoria G. Castiglioni *et al.*

Corresponding author: Victoria G. Castiglioni, v.g.castiglioni@csic.es; Santiago F. Elena, santiago.elena@csic.es

*Sci. Adv.* **10**, eadn5945 (2024)
DOI: 10.1126/sciadv.adn5945

**The PDF file includes:**

Figs. S1 to S11
Table S8
Legends for tables S1 to S7, and S9
References

**Other Supplementary Material for this manuscript includes the following:**

Tables S1 to S7, and S9

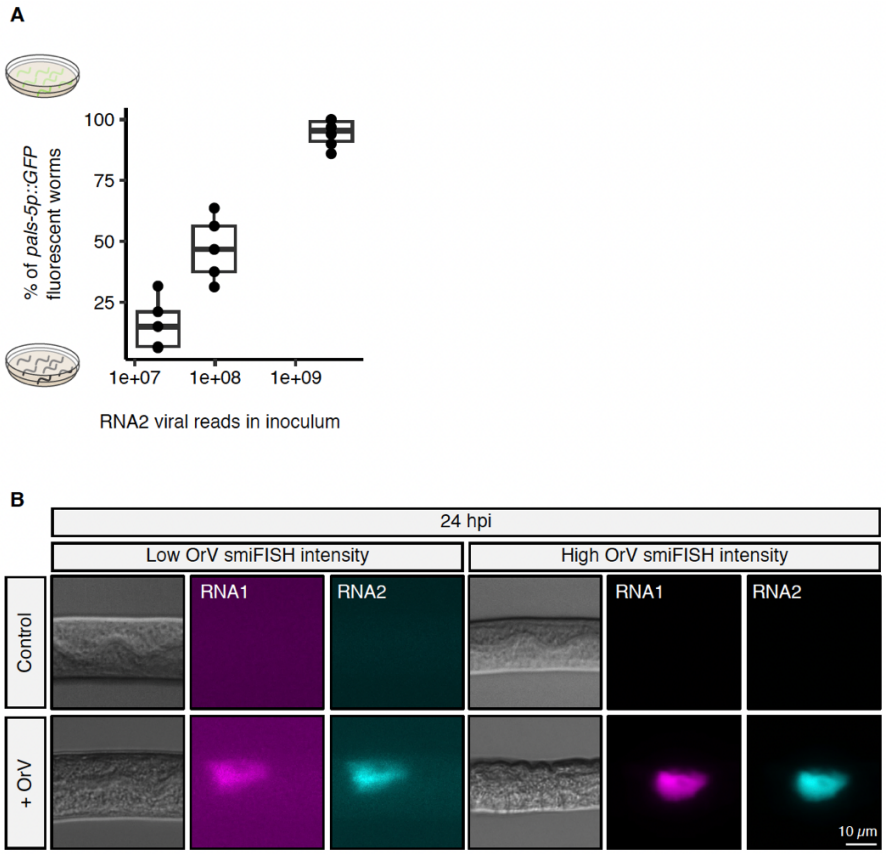



**Fig. S1. OrV positive animals in wild-type and sensitive background JU1580 strains**
(A) Percentage of *pals-5p*::GFP positive animals inoculated with 1.95 $\times 10^7$ to 2.76 $\times 10^9$ copies of OrV and analyzed at 24 hpi. $n = 89$ animals for 1.95 $\times 10^7$, 85 for 9.76$\times 10^7$, and 265 for 2.76 $\times 10^9$, within 5 or 6 replicates. (B) Representative pictures of mock-inoculated smiFISH at 24 hpi adjusted to low and high OrV smiFISH signal.

**A** Description of the 29 ubiquitous genes.

| Symbol | WB | IPR | ZIP-1 Dependent (30 min \| 4h) | Annotation |
|---|---|---|---|---|
| *B0507.6* | WBGene00015223 | ● | ○ ○ | Uncharacterized protein. |
| *C53A5.9* | WBGene00008267 | ● | ● ○ | Uncharacterized protein. |
| ***col-133*** | WBGene00000707 | ○ | ○ ○ | Col_cuticle_N domain-containing protein. |
| *ddn-1* | WBGene00015227 | ● | ● ○ | Downstream of Daf-Nineteen. |
| *droe-8* | WBGene00022550 | ● | ● ○ | Dietary restriction over expressed. |
| *eol-1* | WBGene00045401 | ● | ● ○ | RAI1 domain-containing protein. |
| *F26F2.1* | WBGene00009166 | ● | ○ ○ | Uncharacterized protein. |
| *F26F2.3* | WBGene00009168 | ○* | ○ ○ | Uncharacterized protein. |
| *F42C5.3* | WBGene00018345 | ● | ● ○ | Uncharacterized protein. |
| *F49H6.5* | WBGene00009908 | ● | ○ ● | Elp3 domain-containing protein. |
| *F57G4.11* | WBGene00195165 | ● | ○ ○ | FTH domain-containing protein. |
| *fbxa-182* | WBGene00018353 | ● | ● ○ | FTH domain-containing protein. |
| ***math-15*** | WBGene00015829 | ○ | ○ ○ | MATH (Meprin-associated Traf homology) domain containing. |
| *math-39* | WBGene00020358 | ○ | ● ● | MATH (Meprin-associated Traf homology) domain containing. |
| *pals-14* | WBGene00009061 | ● | ○ ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-27* | WBGene00007132 | ● | ● ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-30* | WBGene00013294 | ● | ● ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-33* | WBGene00021081 | ● | ● ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-37* | WBGene00138721 | ● | ● ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-38* | WBGene00008302 | ● | ○ ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-39* | WBGene00008301 | ● | ● ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-4* | WBGene00007658 | ● | ○ ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *pals-5* | WBGene00007659 | ● | ● ○ | Protein containing ALS2cr12 (ALS2CR12) signature. |
| *T08E11.1* | WBGene00020357 | ● | ○ ○ | Uncharacterized protein. |
| ***trpl-5*** | WBGene00004149 | ○ | ○ ○ | TRP-channel-Like. |
| *Y105C5A.9* | WBGene00013634 | ○ | ○ ● | Uncharacterized protein. |
| ***Y43F8B.12*** | WBGene00012822 | ○ | ○ ○ | Uncharacterized protein. |
| *Y46G5A.20* | WBGene00012910 | ● | ● ○ | Uncharacterized protein. |
| ***Y6G8.5*** | WBGene00044207 | ○ | ○ ○ | Uncharacterized protein. |

*Not part of the list in Table S8 Reddy et al (2019) but affected by *N. parisii* in drh-1(ok3495) (Reddy et al., 2017)

**B** Expression ubiquitous genes

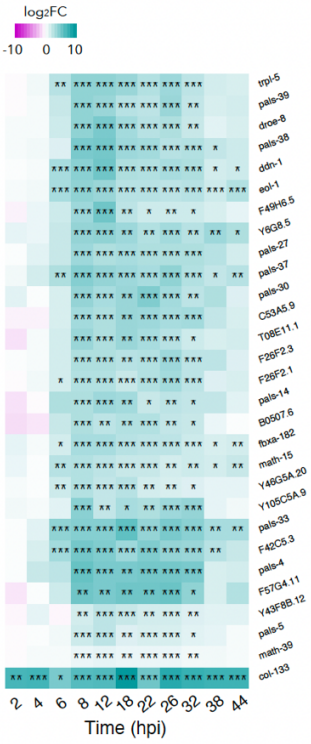



**Fig. S2. Ubiquitous genes**

(A) Description of the 29 genes ubiquitous to the exponential viral replication and moderate stages. (B) Expression profile of the 29 genes common to the exponential viral replication and moderate stages. All the genes present an up-regulated profile. The asterisks indicate the significance level at each time point: $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$.

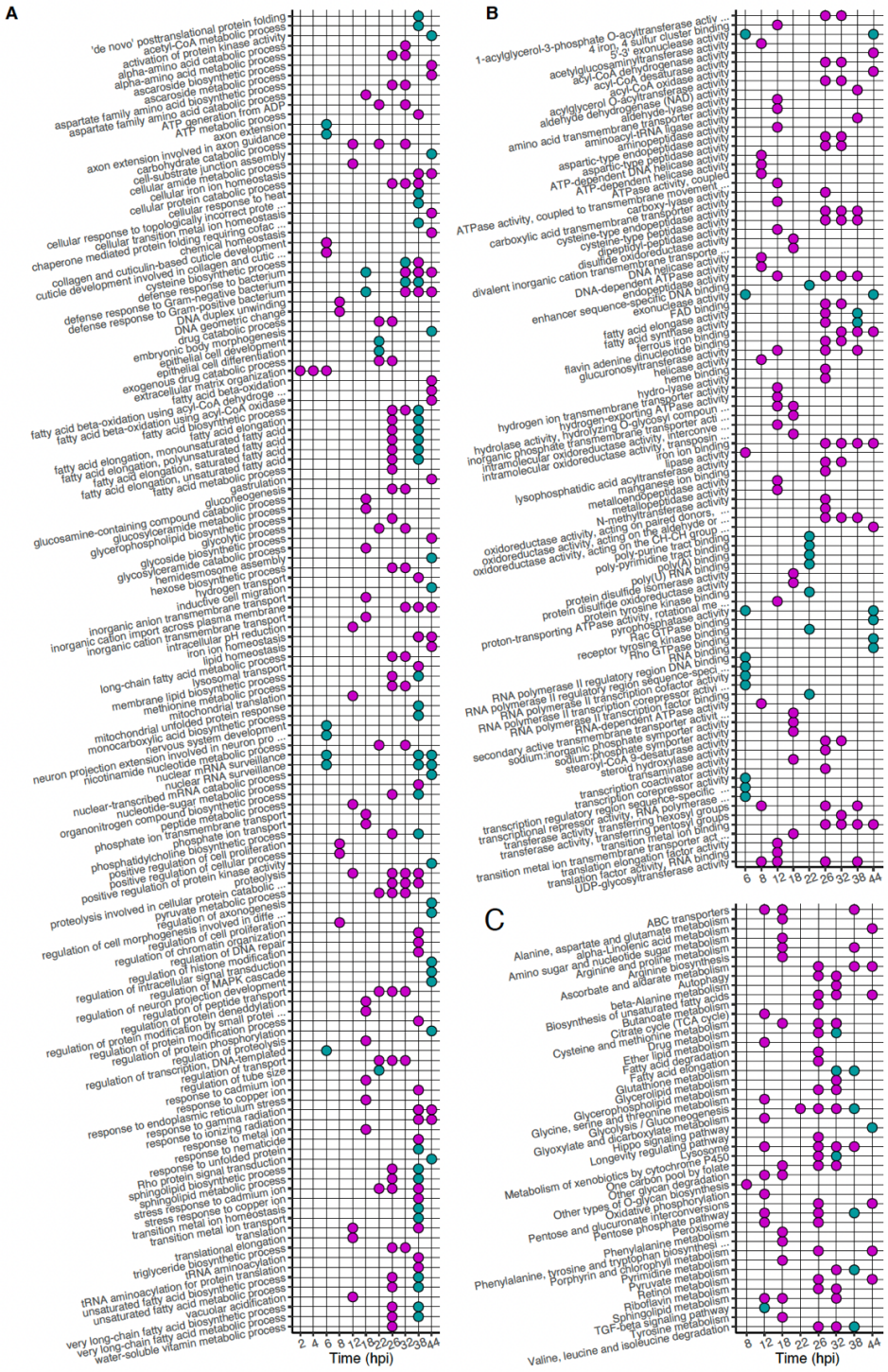



**Fig. S3. Functional enrichment by time point**

Up- (cyan) and down- (magenta) regulated DEGs where analyzed separately. Only significant terms are shown (adjusted $P \leq 0.05$). (A) GO:BP. (B) GO:MF. (C) KEGG pathways.

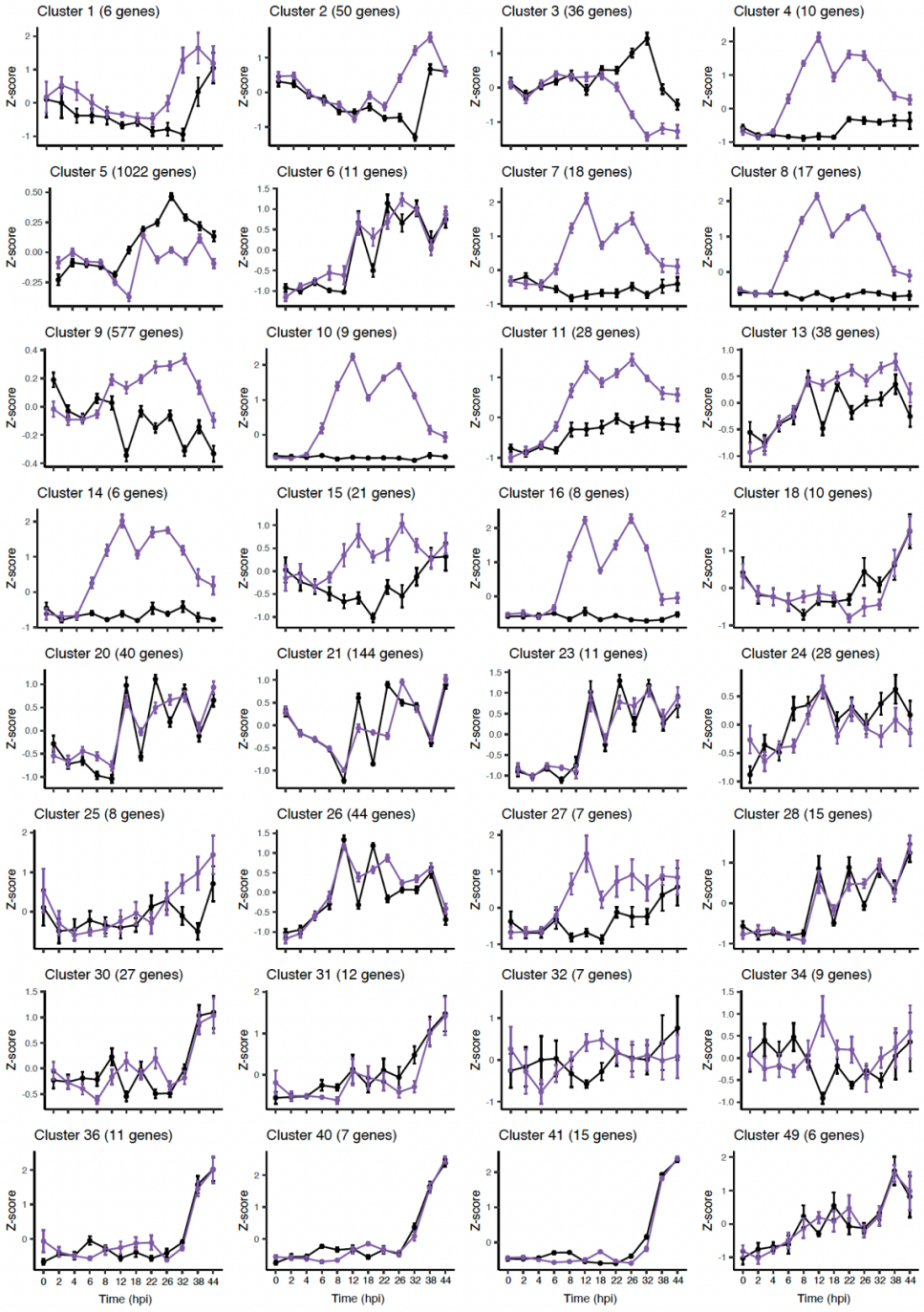



**Fig. S4. Expression profiles of the gene clusters**

All the significant DEGs (LRT analysis) where clustered as indicated in Material and Methods. The mean *z*-score (±1 SD) of the gene expression values of the control (black) and infected (purple) groups are represented by time point (from 0 to 44 hpi). Only clusters with at least 5 genes are shown.

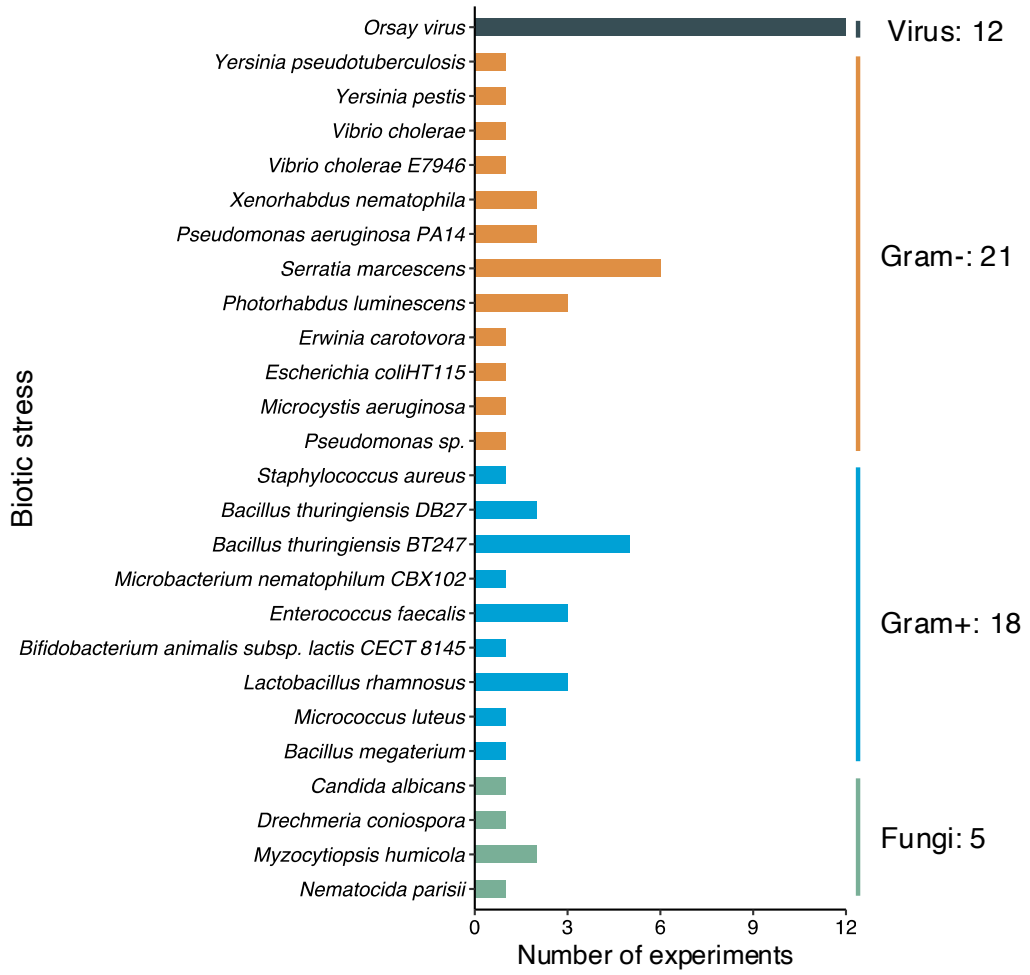
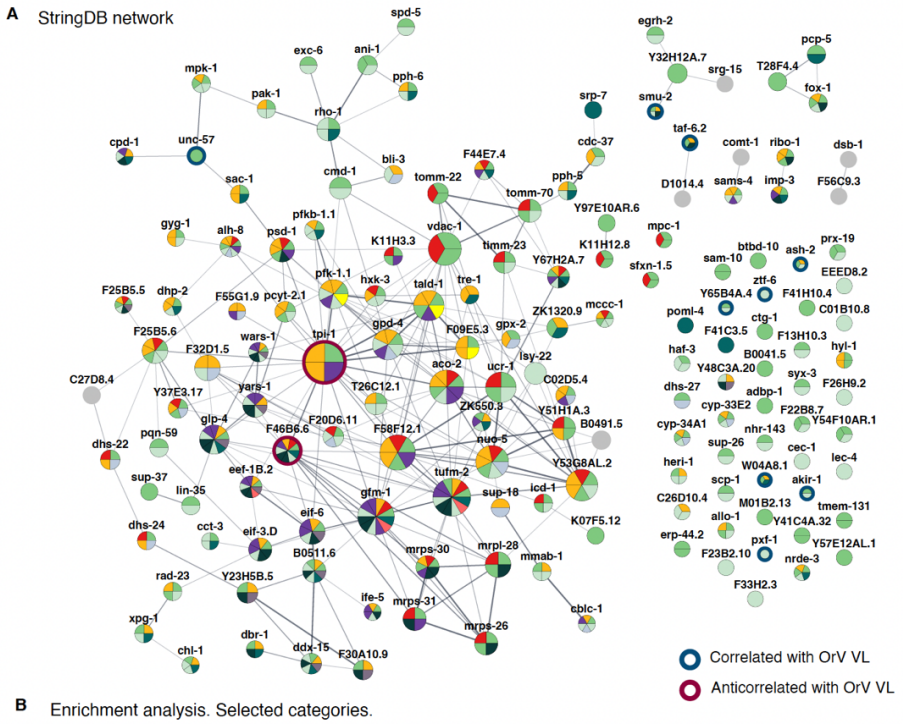



**Fig. S5. Biotic stress comparison analysis**

Number of experiments of each pathogen included in the comparison. Black: OrV, orange: Gram – bacteria, blue: Gram + bacteria and green: Fungi.

A StringDB network

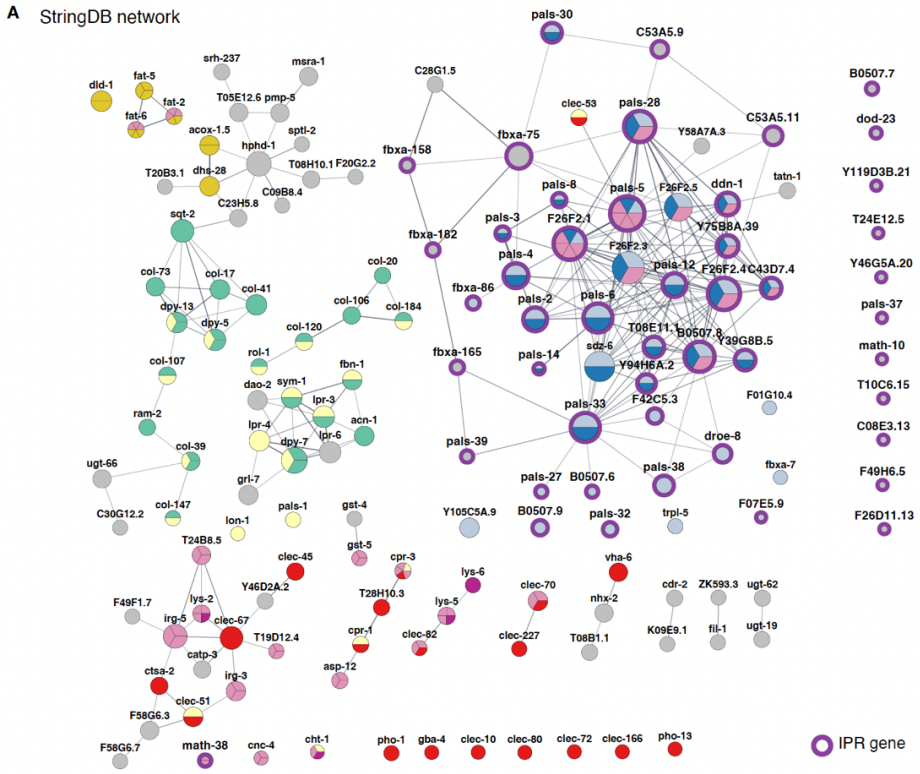



B Enrichment analysis. Selected categories.

| Category | Description | FDR |
|---|---|---|
| GO Molecular Function | Hydrolase activity | 3.90e-02 |
| Reactome Pathways | Translation | 2.40e-03 |
| GO Biological Process | Gene expression | 2.90e-03 |
| GO Biological Process | Peptide metabolic process | 2.60e-05 |
| UniProt Keywords | Protein biosynthesis | 3.73e-05 |
| KEGG Pathways | Biosynthesis of amino acids | 4.60e-04 |
| UniProt Keywords | Transit peptide | 1.75e-02 |
| GO Biological Process | ncRNA metabolic process | 4.68e-02 |
| GO Cellular Component | Cytoplasm | 0.00e+00 |
| COMPARTMENTS | Intracellular membrane-bounded organelle | 1.79e-05 |
| GO Molecular Function | Oxidoreductase activity | 1.13e-02 |
| GO Molecular Function | Binding | 1.30e-04 |
| GO Molecular Function | Nucleotide binding | 1.21e-02 |
| GO Cellular Component | Mitochondrion | 0.00e+00 |
| UniProt Keywords | Elongation factor | 1.99e-02 |
| GO Biological Process | Cellular metabolic process | 0.00e+00 |
| KEGG Pathways | Metabolic pathways | 5.20e-06 |
| KEGG Pathways | Pentose phosphate pathway | 3.74e-02 |

**Fig. S6. Specific response to OrV**

(A) StringDB interaction network of the genes described only in the OrV infection. Node color indicates enriched category (adjusted $P < 0.0468$) and size indicates degree. The genes with an expression profile correlated with OrV viral load (VL) ($r \geq 0.8$) are represented with blue border while the anticorrelated ones ($r \leq -0.8$) appear with burgundy border. (B) Terms of the functional enrichment analysis represented in (A).




B Enrichment analysis. Selected categories.

| | Category | Description | FDR |
|---|---|---|---|
| ● | STRING Clusters | Mostly uncharacterized, incl. Regulation of olfactory learning, and Negative regulators of DDX58/IFIH1 signaling | 0.00e+00 |
| | GO Molecular Function | Structural constituent of cuticle | 1.90e-06 |
| ● | STRING Clusters | Mixed, incl. Cuticle development involved in collagen and cuticulin-based cuticle molting cycle, and Ecdysis, collagen and cuticulin-based cuticle | 6.30e-03 |
| | UniProt Keywords | Collagen | 3.88e-02 |
| ● | GO Biological Process | Aminoglycan catabolic process | 9.80e-03 |
| | GO Biological Process | Peptidoglycan catabolic process | 1.82e-02 |
| ● | STRING Clusters | Mostly uncharacterized, incl. Protein of unknown function DUF713, and Meprin and TRAF homology | 0.00e+00 |
| | STRING Clusters | Mixed, incl. Regulation of olfactory learning, and Innate immune response | 0.00e+00 |
| | GO Biological Process | Response to other organism | 1.40e-03 |
| ● | GO Biological Process | Defense response to other organism | 2.30e-03 |
| | GO Biological Process | Innate immune response | 2.60e-03 |
| | KEGG Pathways | Biosynthesis of unsaturated fatty acids | 1.75e-02 |
| ● | UniProt Keywords | Fatty acid metabolism | 2.54e-02 |
| | UniProt Keywords | Fatty acid biosynthesis | 3.88e-02 |
| | KEGG Pathways | Lysosome | 2.30e-03 |
| ● | UniProt Keywords | Lectin | 3.26e-02 |
| ● | COMPARTMENTS | Extracellular region | 8.50e-04 |

**Fig. S7. Common response to biotic stresses**

(A) StringDB interaction network of the genes described in all pathogen categories. Node color indicates enriched category (adjusted $P < 0.0388$) and size indicates degree. Purple border indicates the genes involved in the IPR. (B) Terms of the functional enrichment analysis represented in (A).

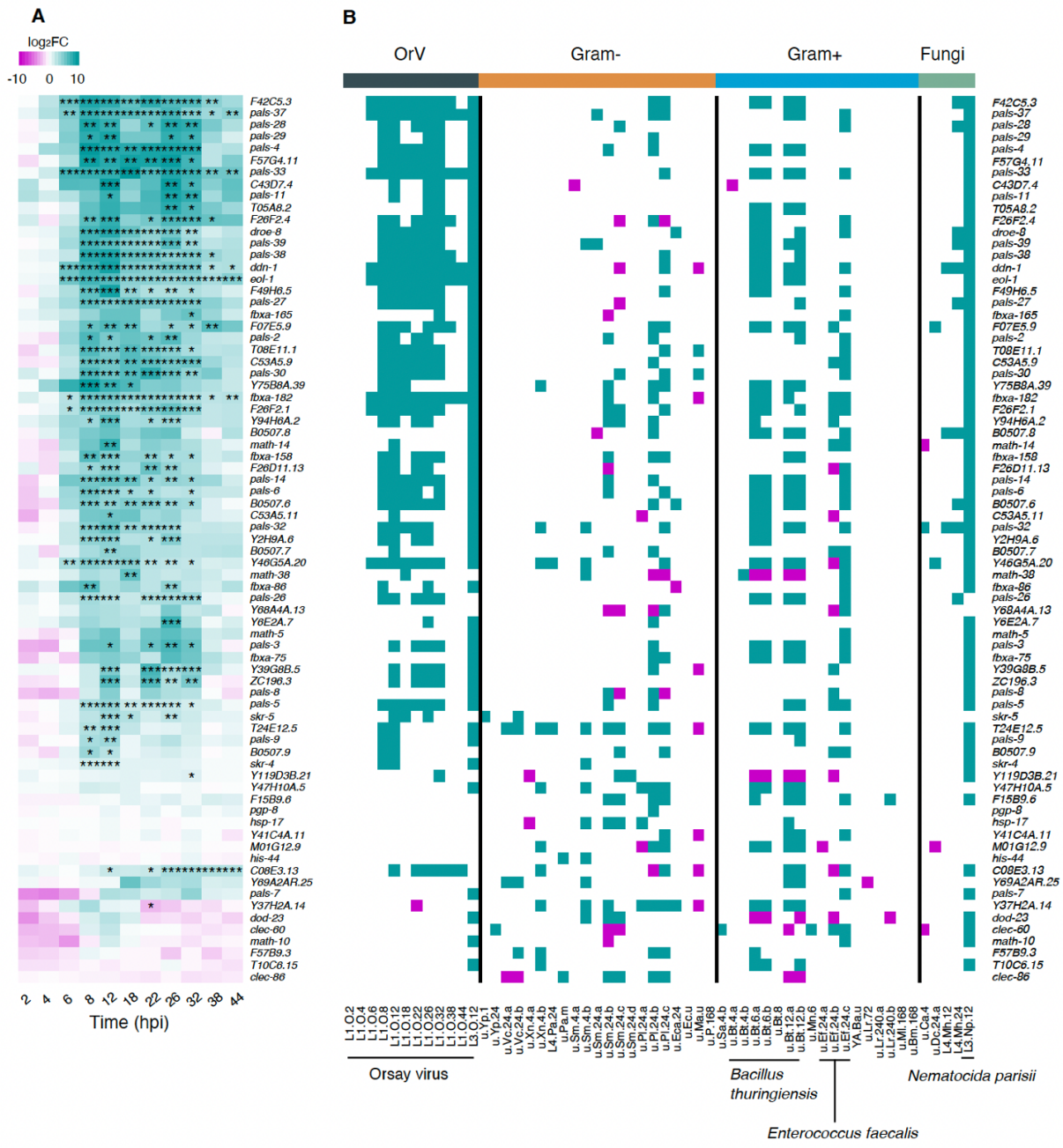



**Fig. S8. Differential expression of the DEGs related with the IPR in our time series experiments and comparison with other biotic stresses**

(A) Differential expression of IPR genes affected in our time series data. Asterisks indicate significance by timepoint (adjusted significance levels: $^{*}P < 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$). (B) Differential expression of the genes in A upon various biotic stresses. Cyan indicates upregulation and magenta indicates downregulation (adjusted $P < 0.05$).

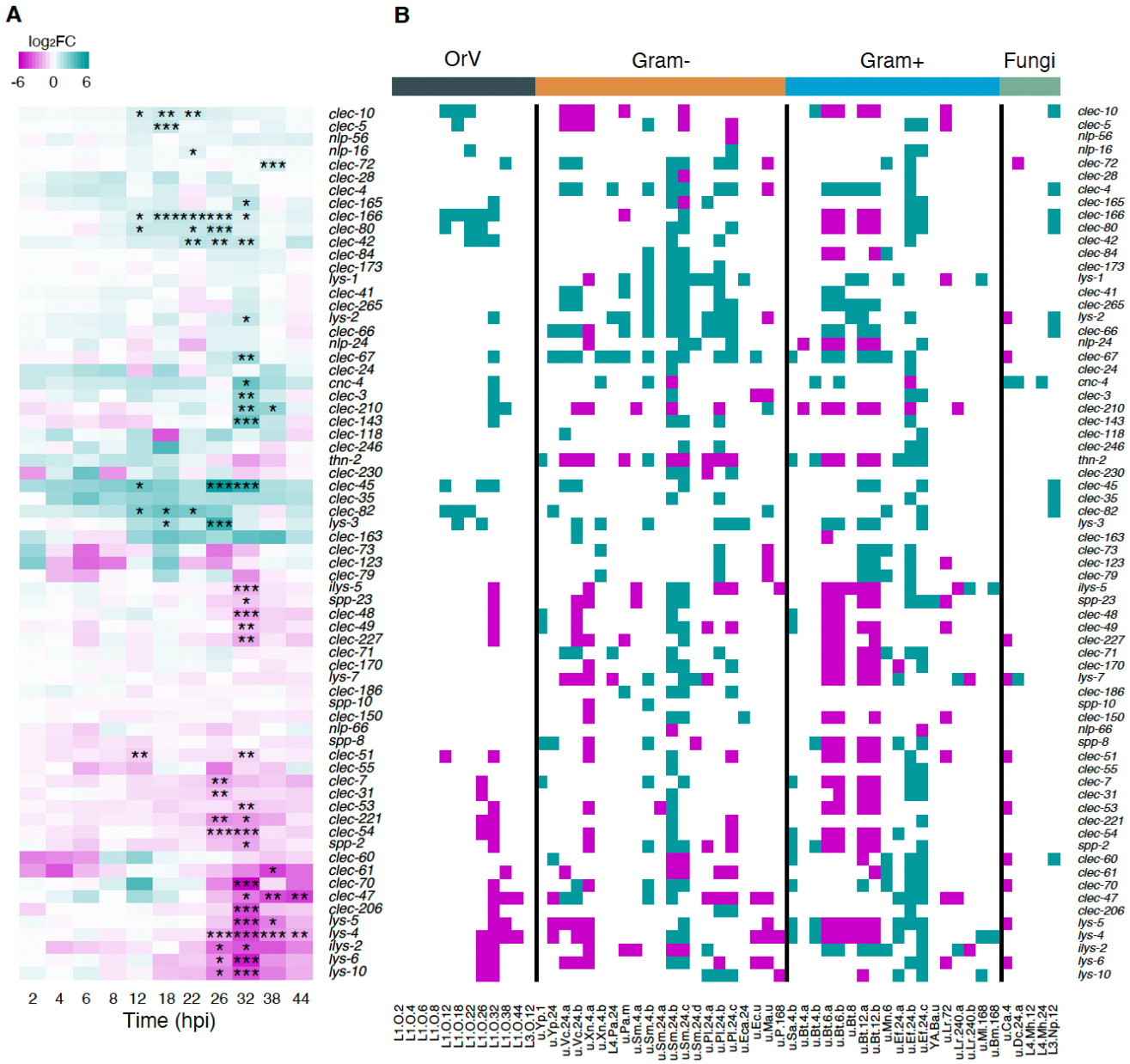



**Fig. S9. Differential expression of antimicrobial effectors identified as DEGs in our time series experiments and comparison with other biotic stresses**

(A) Differential expression of antimicrobial effector genes affected in our time series data. Asterisks indicate significance by timepoint (adjusted significance levels: $^{*}P < 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$). (B) Differential expression of genes in A upon various biotic stresses. Cyan indicates upregulation and magenta indicates downregulation (adjusted $P < 0.05$).

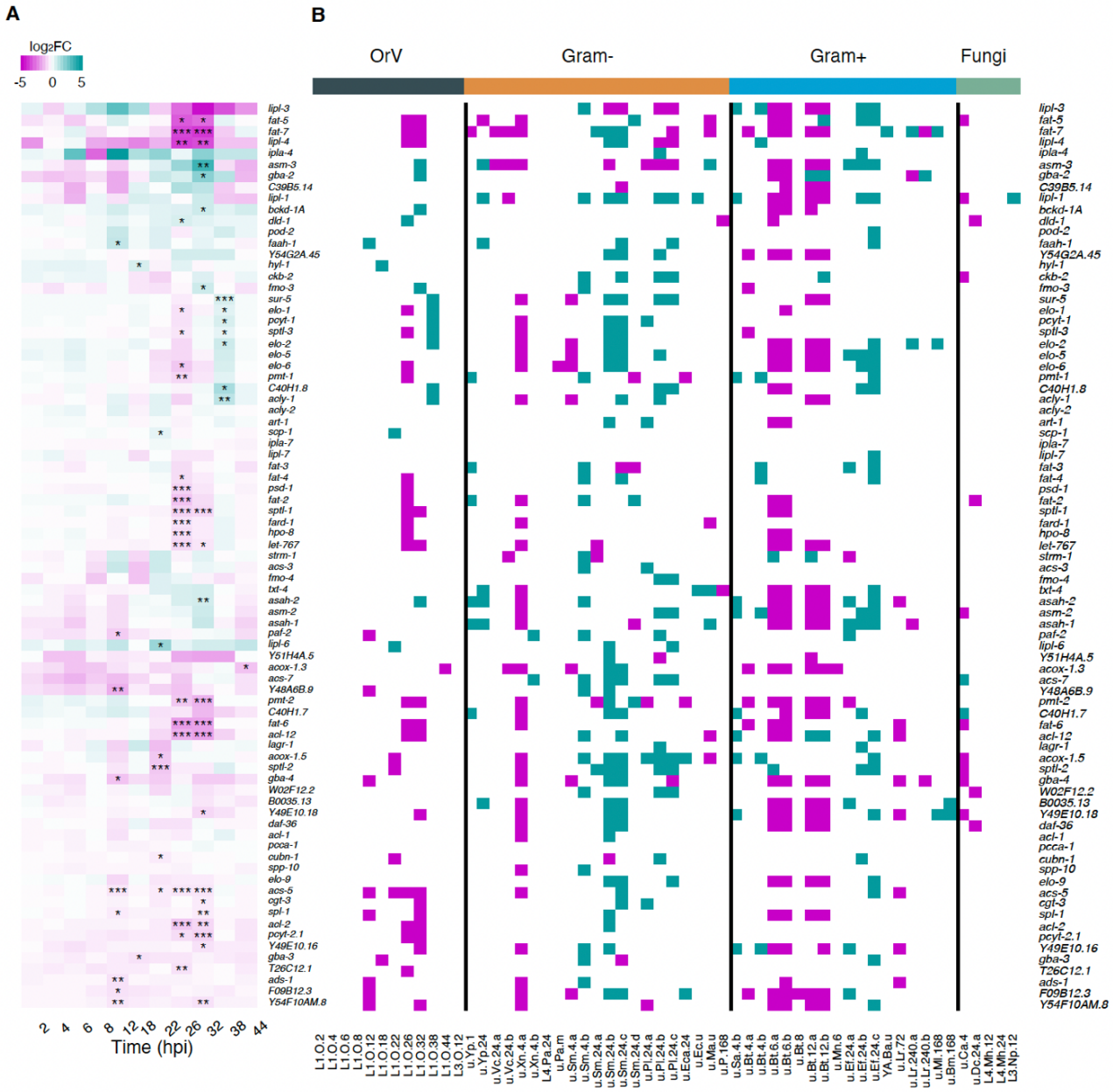



**Fig. S10. Differential expression of DEGs related with the lipid metabolism in our time series experiments and comparison with other biotic stresses**

(A) Differential expression of lipid metabolism related genes affected in our time series data. Asterisks indicate significance by timepoint (adjusted significance levels: $^{*}P < 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$). (B) Differential expression of the genes in A upon various biotic stresses (all performed on the wild-type). Cyan indicates upregulation and magenta indicates downregulation (adjusted $P < 0.05$).

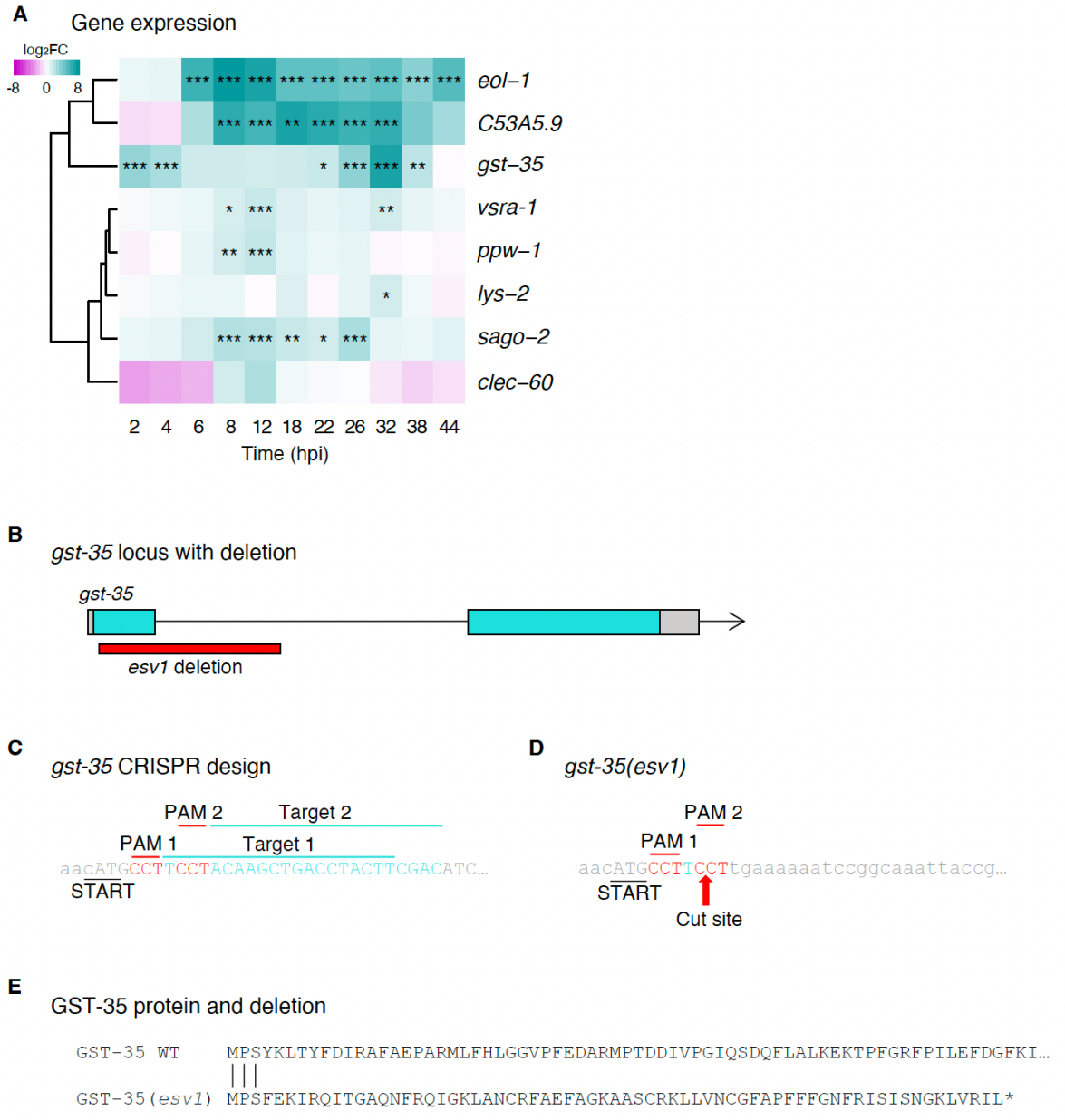



**Fig. S11. Expression of interesting DEGs after OrV exposure**

(A) Expression profiles of selected DEGs. (B) Schematic representation of *gst-35(esv1)* deletion. (C) *gst-35* CRISPR design. (D) *gst-35(esv1)* allele. (E) GST-35 protein and *esv1* deletion with horizontal bars indicating the conserved amino acids.

**Table S8.** Key resources.

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Deposited data** | | |
| Raw fastq files | This study | NCBI SRA: PRJNA1037048 |
| **Bacterial and virus strains** | | |
| *Escherichia coli* OP50 | Caenorhabditis Genetics Center (CGC) | RRID: WB-STRAIN:OP50 |
| Orsay virus (OrV) | Marie-Anne Félix's lab | NCBI:txid977912 |
| **Experimental models: Organisms/strains** | | |
| *C. elegans*: Strain: JU2624: *[myo-2::mcherry::unc54; lys-3p::eGFP::tbb-2] IV* in JU1580 background | Marie-Anne Félix's lab | N/A |
| *C. elegans*: Strain: SFE2: *drh-1(ok3495)IV; mjIs228[myo-2::mcherry::unc54; lys-3p::eGFP::tbb-2]?* | This study | N/A |
| *C. elegans*: Strain: ERT54: *jyIs8[pals-5p::GFP; myo-2p::mCherry] X* | Emily Troemel's lab | WB Strain: ERT54 |
| *C. elegans*: Strain: RB2519: *drh-1(ok3495)IV* | CGC | WB Strain: RB2519 |
| *C. elegans*: Strain: ERT288: *pals-22(jy3) III; jyIs8[pals-5p::GFP, myo-2::mCherry] X* | Emily Troemel's lab | WB Strain: ERT288 |
| *C. elegans*: Strain: NL2550: *ppw-1(pk2505) I* | CGC | WB Strain: NL2250 |
| *C. elegans*: Strain: WM154: *sago-2(tm894) I* | CGC | WB Strain: WM154 |
| *C. elegans*: Strain: WM153: *C04F12.1(tm1637) I* | CGC | WB Strain: WM153 |
| *C. elegans*: Strain: SFE9: *lys-2(tm2398); jyIs8[pals-5p::GFP + myo-2p::mCherry]X* | This study | N/A |
| *C. elegans*: Strain: CB6734: *clec-60(tm2319) II* | CGC | WB Strain: CB6734 |
| *C. elegans*: Strain: SFE10: *eol-1(tm6609); jyIs8[pals-5p::GFP + myo-2p::mCherry]X* | This study | N/A |
| *C. elegans*: Strain: SFE7: *C53A5.9(tm10539); jyIs8[pals-5p::GFP + myo-2p::mCherry]X* | This study | N/A |
| *C. elegans*: Strain: SFE11: *gst-35(esv1) II* | This study | N/A |
| **Oligonucleotides** | | |

| | | |
|---|---|---|
| See Table S9 for oligonucleotides | | |
| **Software and algorithms** | | |
| FastQC v0.12.1 | Andrews (*54*) | https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ |
| MultiQC v1.12 | Ewels et al. (*55*) | https://multiqc.info |
| bbduk.sh v39.01 | | https://sourceforge.net/projects/bbmap/<br>https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbduk-guide/ |
| STAR v.2.7.9a | Dobin et al. (*56*) | https://github.com/alexdobin/STAR |
| R version 4.2.1 (2022-06-23) | R Core Team (*66*) | https://www.r-project.org |
| RStudio 2022.02.3+492 "Prairie Trillium" | | https://posit.co/products/open-source/rstudio/ |
| R packages and versions listed in sessionInfo | This study | https://github.com/MJmaolu/OrV_progression/blob/main/sessionInfo.txt |
| **Other** | | |
| Scripts created to perform the bioinformatic analysis: from raw fastq files to the plots | This study | https://github.com/MJmaolu/OrV_progression/tree/main<br>https://git.csic.es/sfelenalab/ |

**Supplementary tables (separate Excel files):**

**Table S1:**

DEA significant results (by timepoint and the by the interaction condition by time).

**Table S2:**

Functional enrichment complete results.

**Table S3:**

Clusters LRT DEGs.

**Table S4:**

Biotic stresses experiments comparison.

**Table S5:**

Functional enrichment analysis of OrV-specific set of DEGs.

**Table S6:**

Overlap between OrV-specific responding genes and DEGs of *C. briggsae* upon Santeuli virus infection.

**Table S7:**

Functional enrichment analysis of DEGs common to all pathogen categories.

**Table S9:**

Pairs of primers used along the study for different purposes.

## REFERENCES AND NOTES


1. C. A. Suttle, Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).

2. R. Solé, S. F. Elena, *Viruses as Complex Adaptive Systems* (Primers in Complex Systems, Princeton Univ. Press, 2019).

3. A. X. Han, S. P. J. de Jong, C. A. Russell, Co-evolution of immunity and seasonal influence viruses. *Nat. Rev. Microbiol.* **21**, 805–817 (2023).

4. K. B. G. Scholthof, The past is present: Coevolution of viruses and host resistance within geographic centers of plant diversity. *Annu. Rev. Phythopahtol.* **61**, 119–136 (2023).

5. I. Engelmann, N. Pujol, Innate immunity in *C. elegans*, in *Invertebrate Immunity*, *Advances in Experimental Medicine and Biology*, K. Söderhäll, Ed. (Springer, 2010), pp. 105–121.

6. M. A. Félix, A. Ashe, J. Piffaretti, G. Wu, I. Nuez, T. Bélicard, Y. Jiang, G. Zhao, C. J. Franz, L. D. Goldstein, M. Sanroman, E. A. Miska, D. Wang, Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLoS Biol.* **9**, e1000586 (2011).

7. H. Jiang, C. J. Franz, G. Wu, H. Renshaw, G. Zhao, A. E. Firth, D. Wang, Orsay virus utilizes ribosomal frameshifting to express a novel protein that is incorporated into virions. *Virology* **450-451**, 213–221 (2014).

8. W. Yuan, Y. Zhou, Y. Fan, Y. J. Tao, W. Zhong, Orsay δ protein is required for nonlytic viral egress. *J. Virol.* **92**, e00745–e00718 (2018).

9. Y. Fan, Y. R. Guo, W. Yuan, Y. Zhou, M. V. Holt, T. Wang, B. Demeler, N. L. Young, W. Zhong, Y. J. Tao, Structure of a pentameric virion-associated fiber with a potential role in Orsay virus entry to host cells. *PLOS Pathog.* **13**, e1006231 (2017).

10. Y. R. Guo, Y. Fan, Y. Zhou, M. Jin, J. L. Zhang, H. Jiang, M. V. Holt, T. Wang, N. L. Young, D. Wang, W. Zhong, Y. J. Tao, Orsay virus CP-δ adopts a novel β-bracelet structural fold and incorporates into virions as a head fiber. *J. Virol.* **94**, e01560–e01520 (2020).

11. C. J. Franz, H. Renshaw, L. Frezal, Y. Jiang, M. A. Félix, D. Wang, Orsay, Santeuil and Le Blanc viruses primarily infect intestinal cells in *Caenorhabditis* nematodes. *Virology* **448**, 255–264 (2014).

12. A. Ashe, T. Bélicard, J. Le Pen, P. Jarkies, L. Frézal, N. J. Lehrbach, M. A. Félix, E. A. Miska, A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *eLife* **2**, e00994 (2013).

13. K. Chen, C. J. Franz, H. Jiang, Y. Jiang, D. Wang, An evolutionarily conserved transcriptional response to viral infection in *Caenorhabditis* nematodes. *BMC Genomics* **18**, 303 (2017).

14. J. N. Sowa, H. Jiang, L. Somasundaram, E. Tecle, G. Xu, D. Wang, E. R. Troemel, The *Caenorhabditis elegans* RIG-I homolog DRH-1 mediates the intracellular pathogen response upon viral infection. *J. Virol.* **94**, e01173–e01119 (2020).

15. J. Le Pen, H. Jiang, T. Di Domenico, E. Kneuss, J. Kosałka, C. Leung, M. Morgan, C. Much, K. L. M. Rudolph, A. J. Enright, D. O'Carroll, D. Wang, E. A. Miska, Terminal uridylyltransferases target RNA viruses as part of the innate immune system. *Nat. Struct. Mol. Biol.* **25**, 778–786 (2018).

16. M. A. Bakowski, C. A. Desjardins, M. G. Smelkinson, T. A. Dunbar, I. F. Lopez-Moyado, S. A. Rifkin, C. A. Cuomo, E. R. Troemel, Ubiquitin-mediated response to microsporidia and virus infection in *C. elegans*. *PLoS Pathog.* **10**, e1004200 (2014).

17. K. C. Reddy, T. Dror, J. N. Sowa, J. Panek, K. Chen, E. S. Lim, D. Wang, E. R. Troemel, An intracellular pathogen response pathway promotes proteostasis in *C. elegans*. *Curr. Biol.* **27**, 3544–3553.e5 (2017).

18. K. C. Reddy, T. Dror, R. S. Underwood, G. A. Osman, C. R. Elder, C. A. Desjardins, C. A. Cuomo, M. Barkoulas, E. R. Troemel, Antagonistic paralogs control a switch between growth and pathogen resistance in *C. elegans*. *PLoS Pathog.* **15**, e1007528 (2019).

19. L. A. Casorla-Perez, R. Guennoun, C. Cubillas, B. Peng, K. Kornfeld, D. Wang, Orsay virus infection of *Caenorhabditis elegans* is modulated by zinc and dependent on lipids. *J. Virol.* **96**, e01211–e01222 (2022).

20. H. Jiang, K. Chen, L. E. Sandoval, C. Leung, D. Wang, An evolutionarily conserved pathway essential for Orsay virus infection of *Caenorhabditis elegans*. *mBio* **8**, e00940–e00917 (2017).

21. Y. Zhou, W. Zhong, Y. J. Tao, Collagen and actin network mediate antiviral immunity against Orsay in *C. elegans* intestinal cells. *PLoS Pathog.* **20**, e1011366 (2024).

22. M. G. Sterken, L. B. Snoek, K. J. Bosman, J. Daamen, J. A. G. Riksen, J. Bakker, G. P. Pijlman, J. E. Kammenga, A heritable antiviral RNAi response limits Orsay virus infection in *Caenorhabditis elegans* N2. *PLOS ONE* **9**, e89760 (2014).

23. D. M. Parker, L. P. Winkenbach, A. Parker, S. Boyson, E. O. Nishimura, Improved methods for single-molecule fluorescence *in situ* hybridization and immunofluorescence in *Caenorhabditis elegans* embryos. *Curr. Protoc.* **1**, e299 (2021).

24. V. Lažetić, F. Wu, L. B. Cohen, K. C. Reddy, Y. T. Chang, S. S. Gang, G. Bhabha, E. R. Troemel, The transcription factor ZIP-1 promotes resistance to intracellular infection in *Caenorhabditis elegans*. *Nat. Commun.* **13**, 17 (2022).

25. E. J. E. Kim, H. G. Son, H. E. H. Park, Y. Jung, S. Kwon, S. J. V. Lee, *Caenorhabditis elegans algn-2* is critical for longevity conferred by enhanced nonsense-mediated mRNA decay. *iScience* **23**, 101713 (2020).

26. W. Yang, K. Dierking, H. Schulenburg, AnimalExp: A web-based application for a *Caenorhabditis elegans*-specific gene expression enrichment analysis. *Bioinformatics* **32**, 943–945 (2016).

27. G. A. Osman, M. K. Fasseas, S. L. Koneru, C. L. Essmann, K. Kyrou, M. A. Srinivasan, G. Zhang, P. Sarkies, M. A. Félix, M. Barkoulas, Natural infection of *C. elegans* by an oomycete reveals a new pathogen-specific immune response. *Curr. Biol.* **28**, 640–648.e5 (2018).

28. R. González, M. A. Félix, *Caenorhabditis elegans* immune responses to microsporidia and viruses. *Dev. Comp. Immunol.* **154**, 105148 (2024).

29. K. Dierking, W. Yang, H. Schulenburg, Antimicrobial effectors in the nematode *Caenorhabditis elegans*: An outgroup to the Arthropoda. *Philos. Trans. R. Soc. B* **371**, 20150299 (2016).

30. B. Pees, W. Yang, A. Zárate-Potes, H. Schulenburg, K. Dierking, High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates. *J. Innate Immun.* **8**, 129–142 (2015).

31. H. Yoon, J. L. Shaw, M. C. Haigis, A. Greka, Lipid metabolism in sickness and in health: Emerging regulators of lipotoxicity. *Mol. Cell* **81**, 3708–3730 (2021).

32. B. Song, X. Sheng, J. L. Justice, K. K. Lum, P. J. Metzger, K. C. Cook, J. C. Kostas, I. M. Cristea, Intercellular communication within the virus microenvironment affects the susceptibility of cells to secondary viral infections. *Sci. Adv.* **9**, eadg3433 (2023).

33. Y. Guo, Z. Xun, S. R. Coffman, F. Chen, The Shift of the intestinal microbiome in the innate immunity-deficient mutant *rde-1* strain of *C. elegans* upon Orsay virus infection. *Front. Microbiol.* **8**, 933 (2017).

34. I. Melero, R. González, S. F. Elena, Host developmental stages shape the evolution of a plant RNA virus. *Philos. Trans. R. Soc. B* **378**, 20220005 (2023).

35. J. E. Irazoqui, E. R. Troemel, R. L. Feinbaum, L. G. Luhachack, B. O. Cezairliyan, F. M. Ausubel, Distinct pathogenesis and host responses during infection of *C. elegans* by *P. aeruginosa* and *S. aureus*. *PLoS Pathog.* **6**, e1000982 (2010).

36. D. O'Rourke, D. Baban, M. Demidova, R. Mott, J. Hodgkin, Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. *Genome Res.* **16**, 1005–1016 (2006).

37. T. B. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, J. Middel, I. L. Cornelissen, H. S. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor, Y. van Kooyk, DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587–597 (2000).

38. Z. S. Yang, C. Y. Lin, M. B. Khan, M. C. Hsu, W. Assavalapsakul, A. Thitithanyanont, S. F. Wang, Understanding the role of galectins toward influenza A virus infection. *Expert Opin. Ther. Targets* **27**, 927–937 (2023).

39. C. Cubillas, L. E. Sandoval Del Prado, S. Goldacker, C. Fujii, A. N. Pinski, J. Zielke, D. Wang, The *alg-1* gene is necessary for Orsay virus replication in *Caenorhabditis elegans*. *J. Virol.* **97**, e0006523 (2023).

40. M. Pollari, S. De, A. Wang, K. Mäkinen, The potyviral silencing suppressor HCPro recruits and employs host ARGONAUTE1 in pro-viral functions. *PLOS Pathog.* **16**, e1008965 (2020).

41. U. Seroussi, A. Lugowski, L. Wadi, R. X. Lao, A. R. Willis, W. Zhao, A. E. Sundby, A. G. Charlesworth, A. W. Reinke, J. M. Claycomb, A comprehensive survey of *C. elegans* argonaute proteins reveals organism-wide gene regulatory networks and functions. *eLife* **12**, e83853 (2023).

42. J. Li, Y. Wang, H. Deng, S. Li, H. J. Qiu, Cellular metabolism hijacked by viruses for immunoevasion: Potential antiviral targets. *Front. Immunol.* **14**, 1228811 (2023).

43. S. M. Anderson, R. Pukkila-Worley, Immunometabolism in *Caenorhabditis elegans*. *PLOS Pathog.* **16**, e1008897 (2020).

44. Y. Qu, W. Wang, M. Z. X. Xiao, Y. Zheng, Q. Liang, The interplay between lipid droplets and virus infection. *J. Med. Virol.* **95**, e28967 (2023).

45. N. S. Heaton, G. Randall, Multifaceted roles for lipids in viral infection. *Trends Microbiol.* **19**, 368–375 (2011).

46. E. Ketter, G. Randall, Virus impact on lipids and membranes. *Annu. Rev. Virol.* **6**, 319–340 (2019).

47. L. Herrera-Moro Huitron, L. A. De Jesús-González, M. Martínez-Castillo, J. M. Ulloa-Aguilar, C. Cabello-Gutierrez, C. Helguera-Repetto, J. Garcia-Cordero, M. León Juárez, Multifaceted nature of lipid droplets in viral interactions and pathogenesis. *Microorganisms* **11**, 1851 (2023).

48. S. M. Anderson, H. K. Cheesman, N. D. Peterson, J. E. Salisbury, A. A. Soukas, R. Pukkila-Worley, The fatty acid oleate is required for innate immune activation and pathogen defense in *Caenorhabditis elegans*. *PLOS Pathog.* **15**, e1007893 (2019).

49. P. Deng, N. U. Naresh, Y. Du, L. T. Lamech, J. Yu, L. J. Zhu, R. Pukkila-Worley, C. M. Haynes, Mitochondrial UPR repression during *Pseudomonas aeruginosa* infection requires the bZIP protein ZIP-3. *Proc. Natl. Acad. Sci. USA* **116**, 6146–6151 (2019).

50. M. Kumar, S. Sharma, S. Mazumder, Role of UPR$^{mt}$ and mitochondrial dynamics in host immunity: it takes two to tango. *Front. Cell. Infect. Microbiol.* **13**, 1135203 (2023).

51. C. N. Martineau, N. V. Kirienko, N. Pujol, Innate immunity in *C. elegans*. *Curr. Top. Develop. Biol.* **144**, 309–351 (2021).

52. S. Brenner, The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).

53. T. Stiernagle, Maintenance of C. elegans (WormBook: The Online Review of C. Elegans Biology, 2006); http://www.wormbook.org/.

54. S. Andrews, A quality control tool for high throughput sequence data (2010); http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

55. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

56. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

57. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

58. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

59. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

60. N. Tsanov, A. Samacoits, R. Chouaib, A. M. Traboulsi, T. Gostan, C. Weber, C. Zimmer, K. Zibara, T. Walter, M. Peter, E. Bertrand, F. Mueller, smiFISH and FISH-quant—A flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).

61. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

62. G. Csardi, T. Nepusz, The igraph software package for complex network research. *Complex Syst.* **1695**, 1–9 (2006).

63. P. Davis, M. Zarowiecki, V. Arnaboldi, A. Becerra, S. Cain, J. Chan, W. J. Chen, J. Cho, E. da Veiga Beltrame, S. Diamantakis, S. Gao, D. Grigoriadis, C. A. Grove, T. W. Harris, R. Kishore, T. Le, R. Y. N. Lee, M. Luypaert, H. M. Müller, C. Nakamura, P. Nuin, M. Paulini, M. Quinton-Tulloch, D. Raciti, F. H. Rodgers, M. Russell, G. Schindelman, A. Singh, T. Stickland, K. van Auken, Q. Wang, G. Williams, A. J. Wright, K. Yook, M. Berriman, K. L.

Howe, T. Schedl, L. Stein, P. W. Sternberg, WormBase in 2022-data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics* **220**, iyac003 (2022).

64. J. A. Arribere, R. T. Bell, B. X. H. Fu, K. L. Artiles, P. S. Hartman, A. Z. Fire, Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**, 837–846 (2014).

65. S. Waaijers, J. Muñoz, C. Berends, J. J. Ramalho, S. S. Goerdayal, T. Y. Low, A. D. Zoumaro-Djayoon, M. Hoffmann, T. Koorman, R. P. Tas, M. Harterink, S. Seelk, J. Kerver, C. C. Hoogenraad, O. Bossinger, B. Tursun, S. van den Heuvel, A. J. R. Heck, M. Boxem, A tissue-specific protein purification approach in *Caenorhabditis elegans* identifies novel interaction partners of DLG-1/Discs large. *BMC Biol.* **14**, 66 (2016).

66. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).