

# Comprehensive hallmark gene sequence, genomic and structural analysis clarifies new and established taxa within the *Picornavirales*

Richard Mayne<sup>1</sup>, Donald B. Smith<sup>1</sup>, Katherine Brown<sup>2</sup>, Yan Ping Chen<sup>3</sup>, Andrew E. Firth<sup>2</sup>, Kazuhiko Katayama<sup>4</sup>, Nick J. Knowles<sup>5</sup>, Peter Simmonds<sup>6</sup>

<sup>1</sup>Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, 3 South Parks Road, Oxford OX1 3SY, United Kingdom

<sup>2</sup>Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, United Kingdom

<sup>3</sup>US Department of Agriculture (USDA) – Agricultural Research Center (ARS), Bee Research Laboratory, Building 306, 10300 Baltimore Avenue, BARC-EAST, Beltsville, MD 20705, United States

<sup>4</sup>Laboratory of Virology, Graduate School of Infection Control Sciences, Kitasato University 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

<sup>5</sup>The Pirbright Institute, Ash Road, Pirbright, Woking, Surrey GU24 0NF, United Kingdom

<sup>6</sup>Institute of Biomedicine, University of Turku, Kiinamyllynkatu 10, Turku 20014, Finland

\*Corresponding author. Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, 3 South Parks Road, Oxford OX1 3SY, United Kingdom.  
E-mail: [Richard.Mayne@ndm.ox.ac.uk](mailto:Richard.Mayne@ndm.ox.ac.uk)

## Abstract

The order *Picornavirales* is a group of highly diverse RNA viruses that includes many pathogens of significance to human and veterinary health, agriculture, and the wider environment. However, the wide range of viruses assigned to the order, together with their genomic variability, and the recent description of numerous ‘picorna-like’ viruses derived from metagenomic analyses of environmental samples, challenge the established taxonomic classification of members of the order and the criteria for their classification. Here, we combine the existing gold standard, hallmark RNA-directed RNA-polymerase (RdRP) gene sequence-based analysis with helicase sequence-based phylogeny, RdRP structural prediction through the use of ColabFold and Fold Tree, and analysis of coding-complete genomes using GRAViTy-V2, to genetically classify 525 picornaviral genomes and recently described ‘picorna-like’ viruses. All analyses were conducted with a bespoke, fully automated pipeline for retrieval of genome sequences, domain prediction and extraction, phylogenetic analysis, and output conditioning, which is available as open-source software. Our results reveal broad support for established families as well as for 6 novel families, and 32 new genera. In instances where inconsistencies were found between classification methods, we demonstrate how examination of the pipeline’s output may be used to reconcile differences with respect to the genomic features quantified by the analysis. Automated multimodal taxonomic analysis may save significant resources over manual methods and better define demarcation criteria for families and genera.

**Keywords** Picornavirus, taxonomy, bioinformatics, phylogenetics, ICTV

## 1. Introduction

As of early 2026, the International Committee on Taxonomy of Viruses (ICTV) has assigned 775 species in eight families in the order *Picornavirales*. However, the community recognizes at least 1000 additional unclassified picorna- and picorna-like viruses in publicly available databases, many of which have been assembled from data generated during high-throughput metagenomic sequencing studies (Zell et al. 2017, Thongsripong et al. 2021, do Ramos et al. 2024). Members of the *Picornavirales* are eukaryote-infecting, non-enveloped, positive-strand RNA viruses whose distinguishing genomic characteristic is a three-component ‘replication block’ encoding a helicase (Hel), a proteinase (Pro), and an RNA-directed RNA-polymerase (RdRP), which are cleaved by an auto-proteolytic mechanism (Le Gall et al. 2008).

Additionally, picornavirals may have a 5’ covalently genome-linked protein (VPg) and a 3’ poly(A) tail, though these features are not present in all constituent families. Possession of an icosahedral virion with characteristic jelly-roll fold capsid proteins and T = 3 symmetry was previously another of the order’s defining characteristics, prior to the discovery of exceptions to this rule (Koonin et al. 2008). Indeed, apart from the commonality in replication-associated proteins, the order is notable for its remarkably high genome organizational diversity, even among other RNA viruses. For example, picorna- and picorna-like viruses may be monopartite or bipartite; mono-, di-, or polycistronic; possess a variable number of structural proteins and internal ribosomal entry sites (IRES) (Le Gall et al. 2008), and are prone to both recombination events (Lukashev 2010) and genome

**Received:** January 15, 2026. **Revised:** March 13, 2026. **Accepted:** March 24, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

**Table 1** Current family members of the *Picornavirales*.

Family	Host range	Genome	Ref
<i>Caliciviridae</i>	Mammalia	7–8 kb, monopartite polycistronic	(Vinjé et al. 2019)
<i>Dicistroviridae</i>	Arthropoda	9–10 kb, monopartite bicistronic	(Valles et al. 2017a)
<i>Iflaviridae</i>	Arthropoda	9–11 kb, monopartite monocistronic	(Valles et al. 2017b)
<i>Marnaviridae</i>	Var. marine protists	9–10 kb, monopartite monocistronic	(Lang et al. 2021)
<i>Picornaviridae</i>	Animalia	7–10 kb, monopartite monocistronic	(Zell et al. 2017)
<i>Polycipiviridae</i>	Arthropoda	10–12 kb, monopartite polycistronic	(Olendraite et al. 2019)
<i>Secoviridae</i>	Plantae	9–14 kb, mono-/bipartite monocistronic	(Fuchs et al. 2022)
<i>Soliniviridae</i>	Arthropoda	10–11 kb, monopartite polycistronic	(Brown et al. 2019)
<i>Noraviridae</i>	<i>Drosophila</i>	12 kb, monopartite polycistronic	(Habayeb et al. 2006)

organization reversal (Zhang et al. 2023). This variability makes viruses in the order challenging to classify at lower taxonomic ranks.

There are currently nine accepted families within *Picornavirales* (Table 1) which include a number of pathogens with significant relevance to human, veterinary, agricultural, and environmental health. However, several previously described and characterized viruses are not currently classified to family rank, including ‘posaviruses’ [isolated in porcine faeces, suggested in Ref. (Han et al. 2021) to be a new family ‘Posaliviridae’] and Mayfield virus, a pathogen of bees (*Apidae*) (Pascall et al. 2021).

Taxonomic classification of viruses may only be ratified by the ICTV in cases where the nucleotide sequence for a whole coding region(s) of the genome is known (Simmonds et al. 2017, 2023). The gold standard for RNA virus classification at family rank and higher ranks remains, however, based on the evolutionary relationships of the hallmark RdRP gene, as this genomic component is the defining common feature within the realm *Riboviria* and therefore uniquely allows virus evolutionary histories to be mapped across the realm. Despite recent international efforts to standardize and adapt RdRP analysis for scale (Charon et al. 2024), in practice, there is little consensus in the community on what tools and algorithms might be used to achieve this, and manual curation of alignments is commonplace. Such approaches are rational and effective for demonstrating taxonomic relationships between more divergent viruses with common ancestry, but are, however, not scalable and ignore additional genomic, structural, and functional (e.g. antigenic) features which are otherwise useful for determining the characteristics and groupings of viruses at the rank of genus and species. For example, the widely accepted *Picornaviridae* subfamilies take into consideration not just homologies in the 3CD (Pro-RdRP) domain, but also structural proteins and genome organization (Zell et al. 2021); meanwhile, *Enterovirus* species are demarcated by a 25% sequence divergence threshold in the major capsid protein VP1 gene.

Generating taxonomies in this manner tends to be extremely time consuming and requires significant amounts of input from subject matter experts. It is likely that, with the power of metagenomic approaches, we will continue to identify more novel viruses that should be assigned to *Picornavirales* and beyond at a rate that greatly exceeds our abilities to classify them with existing methods. This presents an urgent and currently unmet requirement for new, scalable taxonomic methods that can provide more information than the simple linear relationships conveyed by hallmark gene phylogenies.

In this article, we present an analysis of both classified and unclassified picorna- and picorna-like viruses, which combines analysis of two evolutionarily conserved genes (RdRP and Hel), genomic component and organization similarity, and predicted RdRP protein secondary

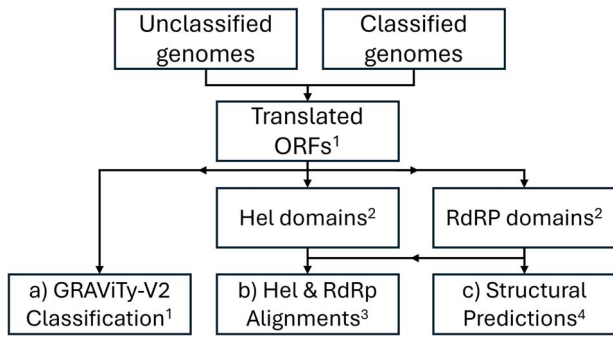
structure relationships. This multimodal approach builds on our previous work in a recent reclassification and promotion of genera within the *Flaviviridae* to three families within the order *Amarillovirales* (Simmonds et al. 2025). The current study expands this analytical approach through the introduction of an open-source, all-in-one software pipeline which is optimized for use on personal computers and incorporates features such as automated domain extraction and bespoke elements for collating and contrasting discrepancies between methods. We proceed to describe how our results clarify established and proposed taxonomic groupings for *Picornavirales* and conclude by discussing both limitations and future work in broadening our method to the study of other taxa and at greater scales.

## 2. Materials and methods

The three analytical methodologies were used (Fig. 1) to generate four sets of taxonomic assignments:

- Alignment and phylogenetic tree generation of both RdRP and Hel domains, using Mafft v7.520 (Katoh and Standley 2013) and IQ-TREE 3.0.1 (Wong et al. 2025).
- GRAViTy-V2 2.2 (Mayne et al. 2024), which is a framework for discovering and classifying viruses using protein profile hidden Markov models (PPHMMs). A composite Jaccard distance (CJD) metric of shared PPHMM similarity and relative genomic position is used to generate taxonomic assignments from coding-complete genomes.
- Comparison of structural characteristics of extracted RdRP domains, through ColabFold (Mirdita et al. 2022) using a lightweight implementation, ColabFold Local 1.5.5 (<https://github.com/YoshitakaMo/localcolabfold>) and Fold-Tree (build 02 06 25) (Moi et al. 2025).

All analytical methods were executed as part of an automated pipeline, created by the authors, called ‘UniViT’ (<https://github.com/Mayne941/UniViT>). Output from these tools was in the form of phylogenetic trees rendered with IToL v7 (Letunic and Bork 2024), alignments from the hallmark gene methods, and various statistics and PPHMM graphs from GRAViTy-V2. Initial classifications at family and genus ranks were generated through a bespoke script as a component of the UniViT pipeline that automatically tabulated leaves from dendrograms for both sequence-based phylogenies and the structural tree and set demarcation boundaries at a fixed depth. These were joined to tabulated GRAViTy-V2 clusters, which were automatically generated by the tool’s classifier module. All classifications were manually checked and disagreements between the phylogenies were recorded.



**Figure 1** Process flow for three analytical pipelines used (a–c); tools: <sup>1</sup>GRAViTy-V2; <sup>2</sup>InterPro scan; <sup>3</sup>Mafft, IQ-Tree3; <sup>4</sup>ColabFold local, fold tree.

Trees were compared through the use of Robinson-Foulds metric computed on tree pairs, and through generating tanglegrams. Tanglegrams were rendered in Python 3.10, using a bespoke implementation of the popular algorithm by Nguyen *et al.* (Nguyen *et al.* 2022), over three untangling iterations. Genome diagrams were created using Vectorgraphic\_genome v0.1 ([https://github.com/donaldsmithictv-cell/Vectorgraphic\\_genome](https://github.com/donaldsmithictv-cell/Vectorgraphic_genome)).

## 2.1 Data acquisition and preprocessing

Previously classified picornaviral sequences were sourced from the ICTV Virus Metadata Resource (VMR, MSL40.v1), which was filtered to one representative example per species ( $n = 244$ ). Unclassified *Picornavirales* sequences were sourced from National Center for Biotechnology Information (NCBI) Virus, using the filters ‘Unclassified Picornavirales’ and ‘complete genome’ ( $n = 281$ ). As outgroups, additional single representative virus sequences from picorna-like virus families *Potyviridae* ( $n = 13$ ) and *Hypoviridae* ( $n = 2$ ) were included, as were representative members of neighbouring *Amarillovirales* ( $n = 4$ ), all of which were sourced from the VMR. The full dataset is included in [Supplementary Information 1](#).

Open reading frames (ORFs) were extracted from whole genomes using GRAViTy-V2 and were then input to InterPro Scan (Jones *et al.* 2014) to generate protein domain predictions based on the Pfam, PROSITE, PANTHER, SMART, and CDD databases. XML outputs from InterPro Scan were parsed with a bespoke script to extract RdRP and Hel amino acid sequences. Sequences were omitted downstream if either or both of the RdRP and Hel domains were not identified.

## 2.2 Hallmark gene alignments

Amino acid multiple sequence alignments were generated for RdRP and Hel sequences output by InterPro Scan, using Mafft with settings:

```
--localpair --iterate 1000 --leavegappyregion
```

Alignments were not manually curated but were examined for errors using SSE v1.4 (Simmonds 2012). Maximum-likelihood trees were generated using IQ-TREE using the Q.pfam+F+I+R10 substitution model for both genes and 1000 bootstrap iterations.

## 2.3 GRAViTy-V2

GRAViTy-V2 was run on the dataset described in section 2.1, in single-pass mode using the ‘new classification full’ pipeline. Run parameters and input sequences are included in [Supplementary Information 1](#).

A separate GRAViTy-V2 analysis was performed to include representative single members of each family within *Riboviria*, using two-pass mode but otherwise identical settings to the single-pass run, to reveal relative depths and comparative demarcation boundaries for both orders and families. As an alignment-free taxonomy framework, GRAViTy-V2 represents groups of viruses with no evolutionary relatedness with a CJD distance of 1.0; distances are not empirical measurements of evolutionary distance but do offer a relative scale to compare relatedness between members of homologous groups.

## 2.4 Structural predictions

RdRP sequences were used as input to a structural prediction pipeline, using ColabFold Local. The pipeline then determined the highest confidence models from the output for each structural prediction and predicted Local Distance Difference Test (pLDDT), which were used to generate phylogenetic trees using FoldTree. Structures with amino acid length >20% shorter than median length, or  $\leq 80\%$  mean pLDDT confidence scores were omitted from the analysis. A subset of predicted protein structures was manually checked by making family-to-family comparisons ( $n = 25$ ) using FoldSeek to align and search for homologues in public database. Both pLDDT score for the structure and root mean square deviation (RMSD) of atomic positions for the comparison with the experimentally derived structure were reported. Molecular views were rendered using UCSF ChimeraX (Meng *et al.* 2023).

## 2.5 Hardware

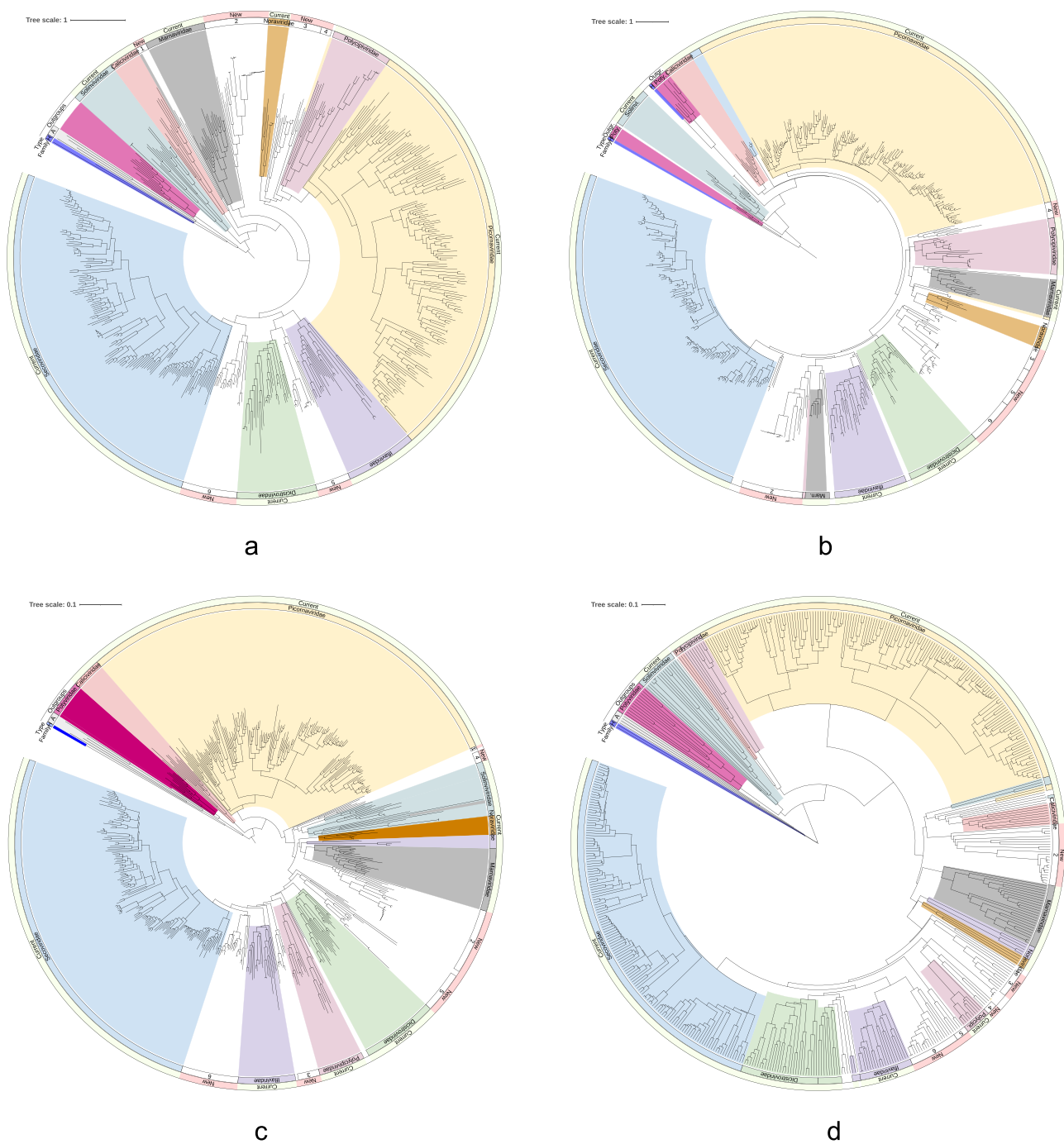
All analytical work was conducted on a personal computer running Ubuntu 22.04 LTS, using an AMD Ryzen 7 7600X 12-core CPU with 32 Gb DDR5 RAM at 6000 MHz. Data were read from and written to an M.2 NVMe SSD at a maximum speed of 3000 Mbps. ColabFold was run using an NVidia GeForce RTX 2080 SUPER with 8 Gb GDDR5 RAM and 3072 Compute Unified Device Architecture (CUDA) cores.

## 3. Results

### 3.1 Comparison of analysis methods

All four analysis methods produced phylogenies that broadly reflected established family rank classifications, though several rearrangements in genera were observed among methods (Figs 2 and 3, [Supplementary Information 2–4](#)). Among the approaches, GRAViTy-V2 classification was most divergent from the RdRP phylogeny (10.5% of classifications disagreed at family rank), followed by the Hel phylogeny (7.8%) and the ColabFold structural tree (4.5%) (Table 2). Robinson-Foulds distances between trees ranged from 0.39 to 0.54, with the Hel phylogeny being the least similar to the RdRP tree. Although genera were generally conserved between RdRP and Hel sequence-based analyses in the larger, more stable families, there were comparatively frequent discrepancies in family ranks. Although RdRP and Hel domains are anticipated to have co-evolved as a result of their proximity within the replication block (Koonin *et al.* 2008), the shorter sequence length of the conserved Hel domain confers less resolution for making sequence-based comparisons.

Computing time for the complete pipeline was  $\sim 94$  h. Run times were comparatively short for both hallmark gene alignment and tree generation steps, at 3 min each. The GRAViTy-V2 analysis took approximately six times longer, and the structural prediction pipeline

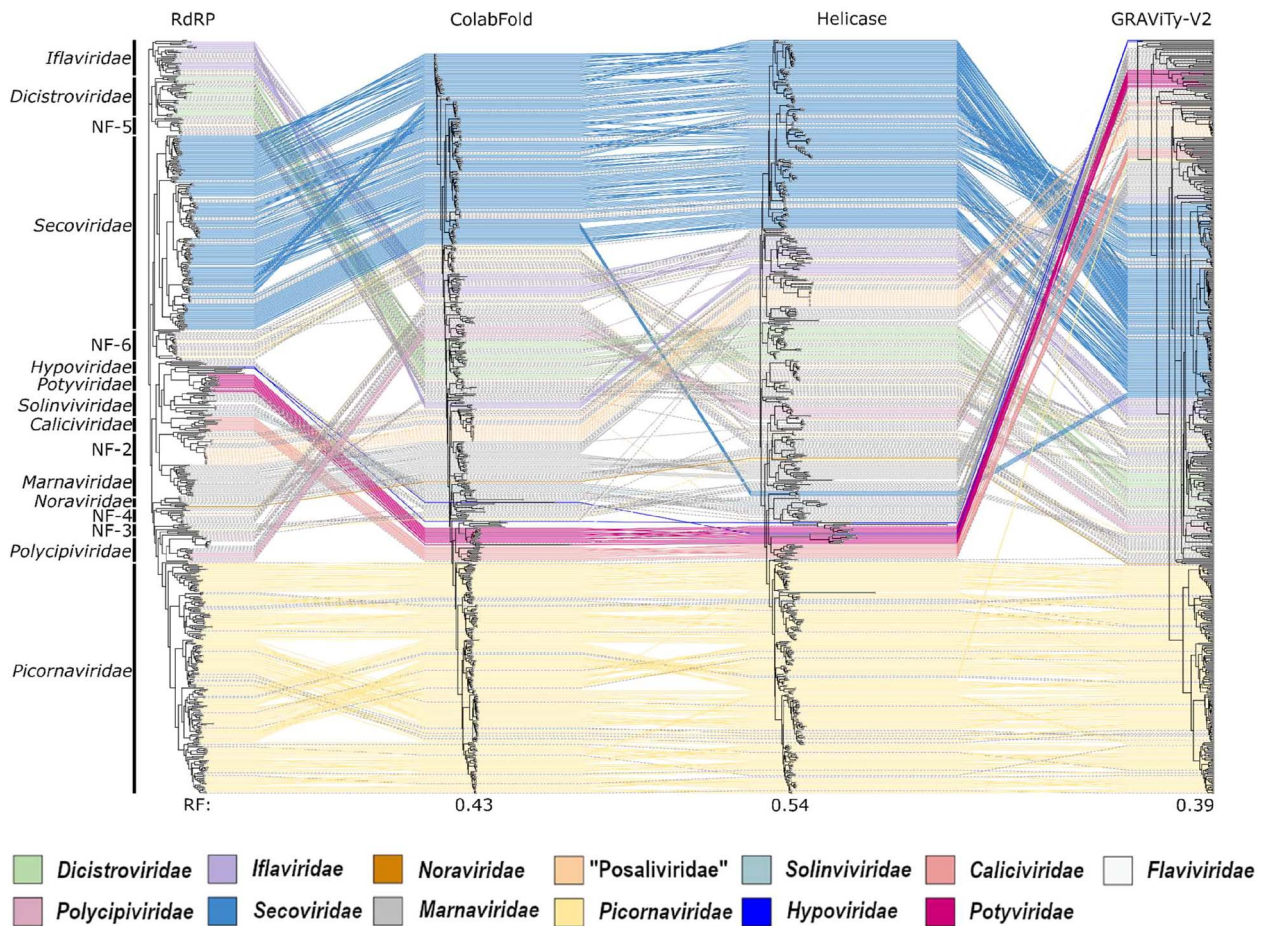


**Figure 2** Phylogenetic trees showing conservation of both new and established family groups between all four analysis methods; clades are coloured by family where the majority of sequences within a clade correspond to a previously defined family, while numerical annotations indicate proposed new families; (a) RdRP phylogeny; (b) Hel phylogeny; branches for flaviviruses KC410872, AF160193 trimmed; (c) RdRP structural predictions; branches trimmed for oat mosaic virus AJ306718 and hypovirus MH766501; (d) GRAViTy-V2 analysis. H.: *Hypoviridae*, A.: *Amarillovirales*.

duration was nearly 2000 times longer. The long run times and comparatively high computing requirements (i.e. a discrete CUDA-enabled graphics card) for deep learning models such as ColabFold represent a key obstacle to their wider adoption and therefore necessitate a strong justification for their use in these contexts.

Phylogenetic placements for *Caliciviridae*, *Marnaviridae*, *Picornaviridae*, *Secoviridae*, as well as prospective family ‘*Posaliviridae*’, were generally stable between analysis methods, with few or no misclassifications or discrepancies (Supplementary Information 2 and 3). Non-picornina-like virus families used as outgroup controls

(*Hypoviridae*, *Potyviridae*, *Flaviviridae*) clustered appropriately in all methods except for in the Hel and GRAViTy-V2 phylogenies, where *Potyviridae* sequences were observed to partially cluster within *Picornavirales*, though inclusion of other ribovirion families in the second GRAViTy-V2 analysis caused them to group appropriately (section 3.5). Previously suggested subfamilies and genera within *Picornaviridae* (Zell et al. 2021) were well supported in all analyses, as were the established genera within *Secoviridae* (*Comovirus*, *Fabavirus*, *Mersevivirus*, *Nepovirus*, *Cheravirus*, *Sadwavirus*, *Sequivirus*, *Stralarivirus*, *Torradorvirus*, and *Waikavirus*).



**Figure 3** Comparison of phylogenetic trees highlights conservation of both established and new family groupings across all analysis methods, but numerous genus rank rearrangements and, in the GRAViTyV2 phylogeny, inversions in relative depths of some families; grey hashed lines indicate previously unclassified sequences; RF: Robinson-Foulds metric, calculated against the RdRP phylogeny.

**Table 2** Summary statistics for each analysis method, comparing run time, number of sequences requiring trimming (N trimmed), number of family discrepancies versus RdRP sequence-based phylogeny (N fam. disc.), and Robinson-Foulds distance versus sequence-based RdRP tree.

	RdRP	Hel	ColabFold	GRAViTy-V2
Run time (min)	3	2	5580	18
N trimmed	0	2	2	0
N fam. disc.	–	43	25	58
Norm. Robinson-Foulds	–	0.54	0.43	0.39

–, not available.

Classifications within *Solinviviridae*, *Polycipiviridae*, *Iflaviridae*, and *Dicistroviridae* were less stable between analysis methods. These families tend to exhibit greater variation in genome organization, number of cistrons, and host ranges in comparison with other families within *Picornavirales* (Valles et al. 2017a, Olendraite et al. 2019, Zell et al. 2021, Mayne et al. 2024).

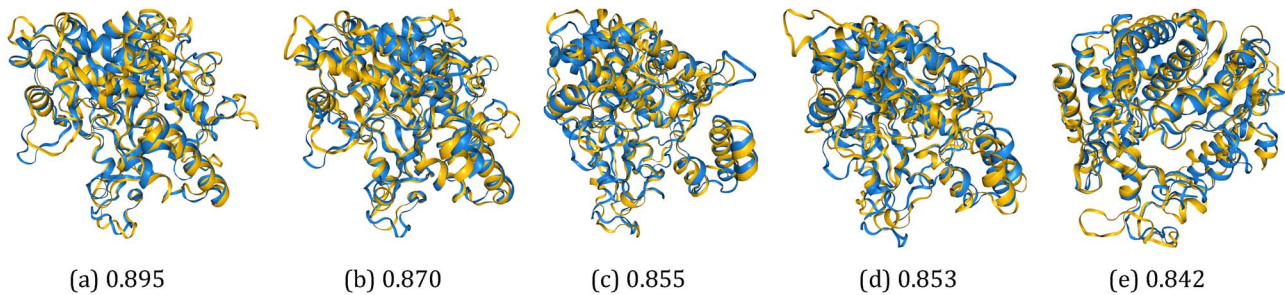
### 3.2 RNA-directed RNA-polymerase sequence-based analysis supports established classifications and indicates new taxa

All previously classified genomes were placed within their appropriate family clusters in the sequence-based RdRP phylogeny, and each family rank grouping was supported by bootstrapping ( $\geq 0.7$ ). Bootstrap support was similarly high for the majority of established

genus groupings. These observations highlight the value of RdRP phylogeny in ribovirion taxonomy, as well as the utility of automated RdRP domain extraction that was used in the UniViT pipeline. In addition to established families in the *Picornavirales*, the analysis also revealed seven additional clusters whose branch depths were indicative of family rank groups (provisionally named ‘NF-1–7’, section 3.6), all of which were supported by bootstrap values of 0.76 or higher.

### 3.3 Helicase phylogeny supports most classifications inferred from RNA-directed RNA-polymerase sequence-based analyses

The Hel phylogeny was concordant with the RdRP analysis regarding family placements for *Caliciviridae*, *Noroviridae*, NF-4 (*Macrobrachium rosenbergii* virus 10–13), *Picornaviridae* and *Iflaviridae*, and was highly



**Figure 4** RdRP domain structural prediction comparisons between infectious flacherie virus (*Iflavirus flacherie*, blue) and random selections of structures across the dataset reveal biologically feasible tertiary structures with anticipated degrees of homology (by pLDDT score); (a) Deformed wing virus (*Iflavirus aladeformis*), RMSD: 3.14; (b) *Macrobrachium rosenbergii* virus 12 (NF-4), RMSD: 2.95; (c) Mobovirus A (*Picornaviridae*), RMSD: 3.27; (d) Mayfield virus 1 (*Noraviridae*), RMSD: 3.43; (e) *Biomphalaria virus 1* ('Nora-2'), RMSD: 3.26.

similar (barring one or two outliers) for *Secoviridae*, *Dicistroviridae*, *Marnaviridae*, and NF-2 ('Posaliviridae'). Major differences were, however, observed for *Solinviviridae*, in which a presumptive new genus ('Solinvi-1', *Riptortus pedestris* virus 1) formed a new clade distinct from any family. Furthermore, a presumptive new family, NF-1, apparent in the RdRP phylogeny, was not reproduced in the Hel analysis, with members clustering with *Marnaviridae*. Similarly, both a presumptive new genus within *Polycipiviridae* ('Polycipi-1') and the entire NF-3 clade either outgrouped or clustered with established families upon analysis of the Hel tree. Hel domains within picornaviruses are considered highly conserved (Yin et al. 2024) and are known to make coherent phylogenies (Koonin et al. 2008) but are however less conserved than RdRP domains and are additionally shorter. Picornaviral Hel domains are additionally prone to recombination, especially in pathogenic enteroviruses (Wang and Wen 2024), which may further limit its usefulness in taxonomic investigations. These factors can make Hel phylogenies more challenging to interpret, and for this reason less reliable than counterpart RdRP sequence-based analyses.

### 3.4 Structural predictions follow RNA-directed RNA-polymerase

RdRP structural predictions for eight RdRP domain sequences failed (listed as 'conversion failed' in Supplementary Information 2) due to low pLDDT scores and were omitted from the analysis. Visual sense checks of molecular structure alignments confirmed that automated domain extraction produced biologically feasible structures, and that this was repeatable across the dataset (Fig. 4).

Classifications for almost all family and genus groupings in the structural phylogeny were congruent with the RdRP sequence-based analysis, with noteworthy differences arising in only NF-1, *Polycipiviridae*, and *Dicistroviridae*. As with the Hel phylogeny, the structure-based tree did not support the creation of NF-1, instead grouping all three viruses within *Solinviviridae*. The problematic sequences from *Polycipiviridae* and *Dicistroviridae* all clustered with other families (*Solinviviridae* and NF-5, respectively), in contravention of the RdRP phylogeny but consistent with the Hel and GRAViTy-V2 analyses.

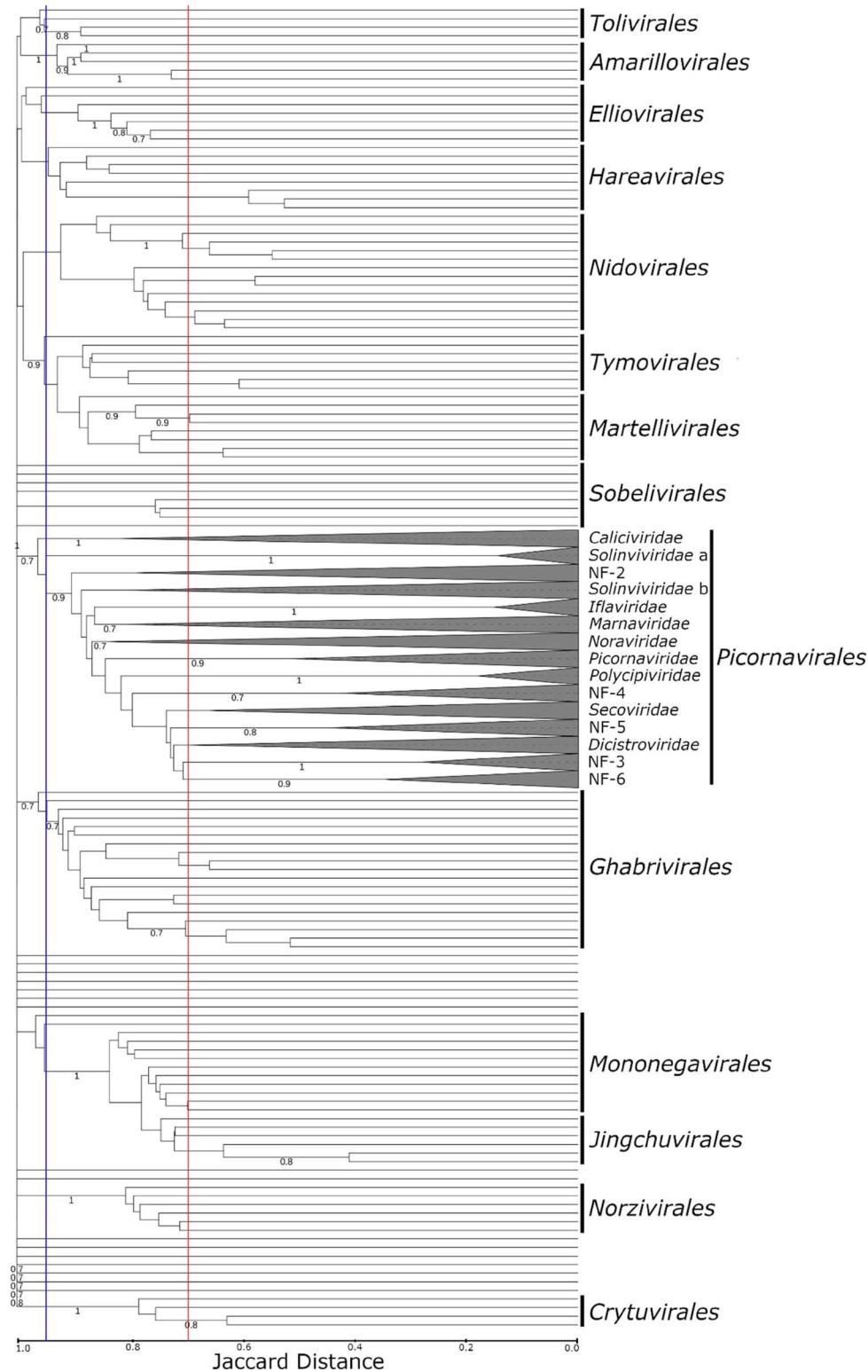
RdRP sequence-based phylogenies are generally expected to follow RdRP structural phylogenies, as protein structural conservation should be strongly correlated with sequence similarity. This correlation is, however, weakened in more distantly related protein structures (Balaji and Srinivasan 2007). Accordingly, the disagreements that were observed with sequence-based phylogeny arose in viruses that were also hard to place by other methods, indicating comparatively greater diversity in these clusters.

### 3.5 GRAViTy-V2 heatmaps reveal extent of picornaviral diversity

The GRAViTy-V2 phylogeny was the most divergent from the RdRP sequence-based phylogeny, though families *Marnaviridae*, *Picornaviridae*, *Secoviridae*, NF-2, NF-4, and NF-5 were largely concordant between the two approaches. In the majority of instances where there were violations, largely in *Solinviviridae*, *Polycipiviridae*, and *Iflaviridae*, violations tended to present as creating outgroup clusters, rather than placing discrepant genomes within other families. This behaviour is typical of how changes in genome composition and/or organization, which are not reflected in conventional hallmark gene analyses, can influence CJDs calculated by GRAViTy-V2, and highlights how an analysis method that is based on complete genome sequences can generate additional information over hallmark gene alignments.

Viruses within *Solinviviridae*, for example, may be either mono- or dicistronic, likely resulting from a ribosomal frameshift site between capsid and capsid extension proteins (Valles et al. 2014, 2016, Brown et al. 2019). Accordingly, the dicistronic *Invictavirus solenopsis* cluster of solinviviruses formed an outgroup in the GRAViTyV2 analysis. Similarly, the GRAViTy-V2 analysis split *Polycipiviridae* into two clusters, as polycipiviruses within the *Sopolycivirus* cluster possess an additional, sixth ORF (Le Gall et al. 2008, Olendraite et al. 2019). While measuring cistronicity of a genome cannot in isolation be used as a method of producing taxonomic groupings, these data contribute to the depiction of genome organizational relationships through shared normalized PPHMM ratio pairwise comparison matrices and PPHMM location 'barcode' graphs (Supplementary Information 5).

GRAViTy-V2 output also revealed differences in genome organization through creating outgroups. *Iflaviridae* and *Dicistroviridae* were split into two clusters each. *Iflaviridae* are monopartite and functionally monocistronic with a single polyprotein strategy, whereas *Dicistroviridae* are monopartite but dicistronic, using dual IRES-driven translation to separately express functional and structural proteins. Members of both families may assume one of two genome organizations: 'Type I', wherein the genome is arranged (from 5' → 3') as {structural block → replication block}, and 'Type II', an inverted arrangement of {replication block → structural block}. In Type II genomes, the orientation of genes within the structural block may additionally be inverted. These examples further emphasize how analysis of complete genomes as part of a taxonomy workflow can uncover both structural and functional characteristics, which may be used to resolve discrepancies between other complementary methods and hallmark gene phylogenies in instances where no consensus is found.



**Figure 5** GRAViTy-V2 analysis reveals relative depth of both new and proposed *Picornavirales* clusters ( $n = 525$ ) is consistent with the description of families within other representative ribovirion families ( $n = 135$ ); order and family lower demarcation boundaries are indicated at  $>0.95$  (blue) and  $>0.70$  (red), respectively; all sub-family branches collapsed and outgroups (described in section 3.5) trimmed.

When compared with other ribovirion families using GRAViTy-V2, the order *Picornavirales* was revealed to be monophyletic and showed a CJD demarcation boundary that was comparable to those of other orders in the *Riboviria* ( $>0.95$ ; Fig. 5). Similarly, families within

the *Picornavirales* were demarcated at comparable ranks to those of families in other orders of RNA viruses (Jaccard distances of 0.70–0.95). The Jaccard distance metric therefore supported the classification of both new and established clusters as families within *Picornavirales*,

**Table 3** Summary of proposed new *Picornavirales* genera, grouped by family. *n* = number of genomes per genus.

Family	New Genus	<i>n</i>	Exemplar
<i>Solinviviridae</i>	'Solinvi-1'	3	KU958718 <i>Riptus pedestris</i> virus-1
	'Solinvi-2'	10	KP714071 Kilifi virus
NF-1/ <i>Solinviviridae</i>	NF-1-1/'Solinvi-3'	3	ON125130 Freshwater macrophyte-associated picorna-like virus 5 isolate P 5
NF-2 ' <i>Posaliviridae</i> '	NF-2-1	1	KR019688 Posavirus 3
	NF-2-2	1	MN627463 <i>Erysiphe necator</i> -associated picorna-like virus 2
	NF-2-3	1	MH703051 Yongsan picorna-like virus 4
	NF-2-4	3	MK606535 Posavirus 12
	NF-2-5	4	LC131490 <i>Picornavirales</i> Bu-1
	NF-2-6	13	JF713720 Posavirus 1
<i>Noraviridae</i>	'Nora-1'	1	MT416538 Bumble bee Mayfield virus 1-like
	'Nora-2'	3	KY024317 <i>Biomphalaria</i> virus 1
NF-3	NF-3-1	2	MK533158 <i>Tetranychus urticae</i> -associated picorna-like virus 2
	NF-3-2	1	MW314611 <i>Linepithema humile</i> picorna-like virus 4
	NF-3-3	9	MG459012 Cherry virus Trakiya
NF-4	NF-4-1	4	ON382567 <i>Macrobrachium rosenbergii</i> virus 12
<i>Polycipiviridae</i>	'Polycipi-1'	1	KY024323 <i>Biomphalaria</i> virus 3
<i>Picornaviridae</i>	'Picorna-1'	1	MF370941 Mobovirus A
	'Picorna-2'	1	MN312220 <i>Macaca fascicularis</i> picornavirus
<i>Iflaviridae</i>	Ifla-1	2	MW281483 <i>Iflavirus achedomestici</i>
	Ifla-2	2	AB000906 <i>Iflavirus flacherie</i>
	Ifla-3	3	AF323747 <i>Iflavirus pernuda</i>
	Ifla-4	6	AF092924 <i>Iflavirus sacbroodi</i>
NF-5	NF-5-1	2	ON382575 <i>Macrobrachium rosenbergii</i> virus 14
	NF-5-2	11	MG662717 Blackbird arilivirus
NF-5/ <i>Dicistroviridae</i>	'Dicistro-1'/NF-5-3	2	OQ599888 Asiatic toad picorna-like virus 1
NF-6	NF-6-1	3	KY024322 <i>Biomphalaria</i> virus 2
	NF-6-2	17	Bat picornavirus isolate BtPV/13585-58/M.dau/DK/2014
			ON382570 <i>Macrobrachium rosenbergii</i> virus 15
<i>Secoviridae</i>	'Seco-1'	2	AB009958 Satsuma dwarf virus
	'Seco-2'	12	OK040225 Cassava Torrado-like virus
	'Seco-3'	26	MH323435 Green Sichuan pepper nepovirus
	'Seco-4'	3	AY303786 <i>Arabis</i> mosaic virus
	'Seco-5'	6	

and differentiated them from groupings at lower or higher taxonomic ranks.

### 3.6 Description of suggested new taxa

A consistent finding between all analysis methods was the existence of six statistically supported clusters, labelled as presumptive new families (Table 3, Supplementary Information 4). Divergence between these clusters was within a range of other families within other ribovirion orders (Fig. 5). Genome maps for exemplar members of each family (Fig. 6) revealed extensive diversity in genome length, content, and organization between each grouping, consistent with their assignments to separate families. An expanded version of Table 1 to include new families is included in Supplementary Information 6.

**NF-2** Corresponds to a previously proposed new family of 'posaviruses' isolated from porcine and various other mammalian faeces, '*Posaliviridae*'. All members of the clade possessed monocistronic genomes, and their grouping was strongly supported by all analysis methods.

**NF-3** Twelve dicistronic or polycistronic metagenomically derived sequences from various plant and insect hosts. NF-3 was not supported by the Hel analysis, where all members either formed distinct outgroups or clustered with *Noraviridae*.

**NF-4** A new grouping of four monocistronic *Macrobrachium rosenbergii* viruses, strongly supported by all analysis methods.

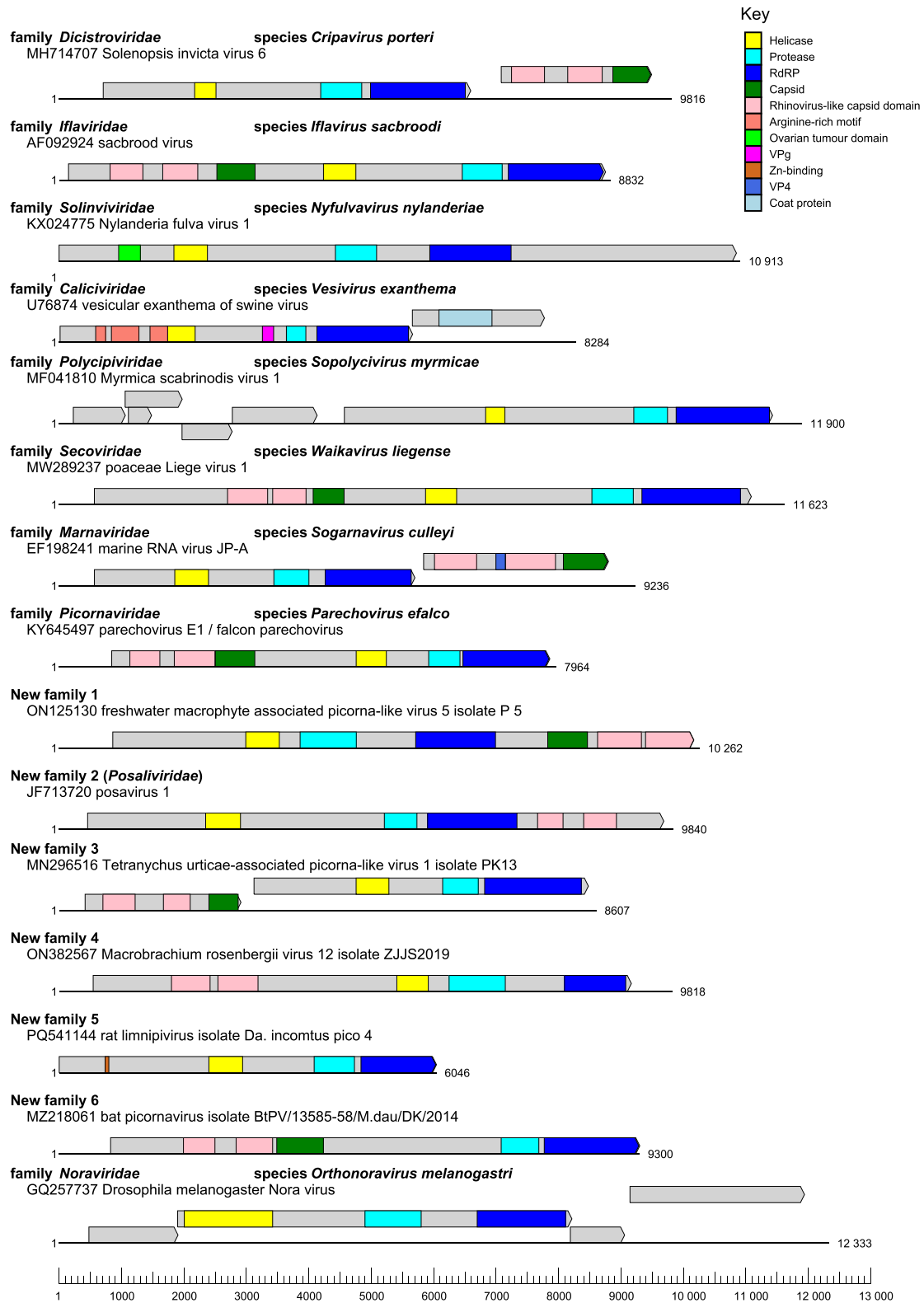
**NF-5** A cluster closely related to *Dicistroviridae*, containing metagenomically derived viruses from insect-, bird-, bat-, and rat-derived

samples. NF-5 formed an outgroup in the GRAViTy-V2 analysis. Constituent genomes were fairly short and monocistronic; several possessed zinc finger proteins detected via InterPro domain prediction.

**NF-6** Cluster closely related to *Secoviridae* but derived from non-plant hosts. Four genomes (ON162339, PP818844, NC 033456, MG552116) that were placed within NF-6 in the RdRP phylogeny grouped with *Noraviridae* in the Hel phylogeny and with NF-5 in the GRAViTy-V2 analysis. Genomes were monocistronic and possessed a complete {rhv-rhv-capsid protein} structural block.

An additional family-like grouping of three monocistronic viruses of freshwater hosts (NF-1; yellow catfish picornavirus (YCP) and freshwater macrophyte-associated picorna-like viruses -4 and -5), most closely related to *Marnaviridae*, was observed in the RdRP phylogeny. Although a single NF-1 member, YCP, did form a singleton family-like group in the structural tree, the other two sequences clustered with the soliniviruses in the ColabFold analysis. Conversely, the GRAViTy-V2 analysis placed YCP with three of the soliniviruses and the two freshwater macrophytes within a separate family-like grouping. The Hel phylogeny classified all three as marnaviruses. This highlights a rare instance of methodological disagreements with the gold standard method and represents an instance where expert supervision may be required to generate a definitive classification.

Thirty-two new bootstrap-supported sub-family rank clusters were identified (Table 3), 29 of which were composed entirely of previously unclassified sequences. The remaining four clusters were identified by splitting *Iflaviridae*, which currently only includes a single genus, at branch depths comparable to genera in other families within the



**Figure 6** Representative single genomes from each classified and novel family within *Picornavirales*, drawn to scale, show extensive variation in the presence and organization of protein-coding domains, as predicted by InterPro Scan; additional protein-coding domains may be present, but are too divergent from publicly available datasets to be detected with this tool.

order. Presumptive new genera were proposed in instances where at least one virus formed a distinct group in at least two of the four analysis methods; where the cluster was split between methods, primacy was given to the RdRP phylogeny.

A potential genus apparent by RdRP phylogeny (Supplementary Information 2), NF-5-3 in the family NF-5, contained two previously unclassified Asiatic toad picorna-like viruses. However, these sequences were grouped within *Dicistroviridae* (labelled

alternatively as the new genus ‘Dicistro-1’) by the other analysis methods. As with family assignments, this further instance is another example where human input may be required to resolve discrepancies.

#### 4. Discussion

The multimodal approach introduced in this study builds on our previous work in creating automated taxonomy workflows that are better suited to the current scale and pace of virus discovery. The majority of virus classifications in this analysis were comparable across the four methods used; in instances where there was disagreement, these corresponding sequences were automatically highlighted for further scrutiny. Genome organizational data from the full genomic analysis by GRAViTy-V2 were often able to reconcile discrepancies. While automated methods cannot replace manual classification by experts, we propose that the introduction of automated tools such as the UniViT pipeline may be used to perform an initial analysis of virus genome data, identify and align conserved gene blocks, and perform a range of comparison methods to guide taxonomic assignments.

While groupings were typically concordant between methods, there was a limited number of instances where all complementary methods disagreed with the RdRP hallmark gene phylogeny. This was observed once at the family rank (NF-1) and several times at the genus rank. In one instance, two sequences (Asiatic toad picorna-like viruses OQ599888, OQ599889) were placed in *Dicistroviridae* by RdRP sequence-based phylogeny, but within NF-5 in all other methods: these sequences were metagenomically assembled from lung and gut samples of Asiatic toads and were originally labelled as ‘unclassified *Picornaviridae*’ (Chen et al. 2023) by maximum likelihood tree of RdRP sequences. Such samples are inherently challenging to assemble and classify, as it is frequently unclear whether the host organism was infected by, or had consumed a host for (dicistroviruses are insect-borne), the assembled virus. This further highlights how careful supervision by experts is required, and how automated taxonomy workflows can help to realise greater efficiencies by expediting identification of problematic sequences.

While classical taxonomy workflows may not benefit substantially from incorporation of additional gene alignments or GRAViTy-V2 analysis, our results demonstrate that structural comparisons with ColabFold may be used to refine and support deeper clusters. Conversely, it should be recognized that a taxonomic classification is not a full evolutionary history of a virus. Beyond the RdRP hallmark gene, other genome components such as capsid genes may have entirely distinct evolutionary origins acquired through horizontal gene transfer (i.e. recombination and reassortment) that cannot be represented in the strictly hierarchical classification developed by the ICTV. Broader whole genome analysis, such as GRAViTy-V2, can be effective in identifying mosaicism genome modularity, and provide a potentially more complete account of a virus’ evolutionary history. Future work with UniViT will focus on the incorporation of further whole genome analyses methods and refining the pipeline to support its wider community adoption and its application for other RNA virus groups.

#### Acknowledgements

The authors extend their thanks and gratitude to Dr Kentaro Tohma for their assistance in reviewing the technical content of this article.

#### Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

#### Conflict of interest

None declared.

#### Funding

The Pirbright Institute receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) of the United Kingdom (projects BBS/E/1/00007037, BBS/E/PI/230001 A, BBS/E/PI/230002 C, and BBS/E/PI/23NB0004).

#### Data availability

The data underlying this article are available in the article and in its online supplementary material.

#### References

- Balaji S, Srinivasan N. Comparison of sequence-based and structure-based phylogenetic trees of homologous proteins: inferences on protein evolution. *J Biosci* 2007;**32**:83–96. <https://doi.org/10.1007/s12038-007-0008-1>
- Brown K, Olendraitte I, Valles SM *et al.* ICTV virus taxonomy profile: *Soliniviridae*. *J Gen Virol* 2019;**100**:736–7. <https://doi.org/10.1099/jgv.0.001242>
- Charon J, Olendraitte I, Forgia M *et al.* Consensus statement from the first RdRp summit: advancing RNA virus discovery at scale across communities. *Front Virol* 2024;**4**:1371958. <https://doi.org/10.3389/fviro.2024.1371958>
- Chen Z, Zhao H, Li Z *et al.* First discovery of phenuiviruses within diverse RNA viromes of Asiatic toad (*Bufo gargarizans*) by metagenomics sequencing. *Viruses* 2023;**15**:750. <https://doi.org/10.3390/v15030750>
- Fuchs M, Hily J-M, Petrzik K *et al.* ICTV virus taxonomy profile: *Secoviridae* 2022. *J Gen Virol* 2022;**103**:001807. <https://doi.org/10.1099/jgv.0.001807>
- Habayeb MS, Ekengren SK, Hultmark D. Nora virus, a persistent virus in *Drosophila*, defines a new picorna-like virus family. *J Gen Virol* 2006;**87**:3045–51. <https://doi.org/10.1099/vir.0.81997-0>
- Han Z, Song Y, Xiao J *et al.* Monsavirus in monkey rectal swab and throat swab specimens in China: proposal for *Posaliviridae* as a new family in *Picornavirales*. *Virus Res* 2021;**303**:198501. <https://doi.org/10.1016/j.virusres.2021.198501>
- Jones P, Binns D, Chang H-Y *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**:1236–40. <https://doi.org/10.1093/bioinformatics/btu031>
- Katoh K, Standley D. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80. <https://doi.org/10.1093/molbev/mst010>
- Koonin EV, Wolf Y, Nagasaki K *et al.* The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat Rev Microbiol* 2008;**6**:925–39. <https://doi.org/10.1038/nrmicro2030>
- Lang AS, Vlok M, Culley AI *et al.* ICTV virus taxonomy profile: *Marnaviridae* 2021. *J Gen Virol* 2021;**102**:001633. <https://doi.org/10.1099/jgv.0.001633>
- Le Gall O, Christian P, Fauquet CM *et al.* *Picornavirales*, a proposed order of positive-sense single-stranded RNA viruses with a

- pseudo-T = 3 virion architecture. *Arch Virol* 2008;**153**:715–27. <https://doi.org/10.1007/s00705-008-0041-x>
- Leticia I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 2024;**52**:W78–82. <https://doi.org/10.1093/nar/gkae268>
- Lukashev AN. Recombination among picornaviruses. *Rev Med Virol* 2010;**20**:327–37. <https://doi.org/10.1002/rmv.660>
- Mayne R, Aiewsakun P, Turner D *et al*. GRAViTy-V2: a grounded viral taxonomy application. *NAR Genom Bioinform* 2024;**6**:4. <https://doi.org/10.1093/nargab/lqae183>
- Meng EC, Goddard TD, Pettersen EF *et al*. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci* 2023;**32**:e4792. <https://doi.org/10.1002/pro.4792>
- Mirdita M, Schütze K, Moriawaki Y *et al*. Colabfold: making protein folding accessible to all. *Nat Methods* 2022;**19**:679–82. <https://doi.org/10.1038/s41592-022-01488-1>
- Moi D, Bernard C, Steinegger M *et al*. Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses. *Nat Struct Mol Biol* 2025;**32**:2492–502. <https://doi.org/10.1038/s41594-025-01649-8>
- Nguyen N, Chawshin K, Berg CF *et al*. Shuffle & untangle: novel untangle methods for solving the tanglegram layout problem. *Bioinform Adv* 2022;**2**:vbac014. <https://doi.org/10.1093/bioadv/vbac014>
- Olendraitte I, Brown K, Valles SM *et al*. ICTV virus taxonomy profile: *Polycipiviridae*. *J Gen Virol* 2019;**100**:554–5. <https://doi.org/10.1099/jgv.0.001241>
- Pascal D, Tinsley M, Clark B *et al*. Virus prevalence and genetic diversity across a wild bumblebee community. *Front Microbiol* 2021;**12**:650747. <https://doi.org/10.3389/fmicb.2021.650747>
- do Ramos E, Barbosa MRF, Villanova F *et al*. Novel members of the order *Picornavirales* identified in freshwater from Guarapiranga reservoir in São Paulo. *Infect Genet Evol* 2024;**124**:105668. <https://doi.org/10.1016/j.meegid.2024.105668>
- Simmonds P. SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 2012;**5**:50. <https://doi.org/10.1186/1756-0500-5-50>
- Simmonds P, Adams MJ, Ö, M. *et al*. Consensus statement: virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 2017;**15**:161–8. <https://doi.org/10.1038/nrmicro.2016.177>
- Simmonds P, Adriaenssens EM, Zerbini FM *et al*. Four principles to establish a universal virus taxonomy. *PLoS Biol* 2023;**21**:1–23. <https://doi.org/10.1371/journal.pbio.3001922>
- Simmonds P, Butković A, Grove J *et al*. Taxonomic expansion and reorganization of *Flaviviridae*. *Nat Microbiol* 2025;**10**:3026–37. <https://doi.org/10.1038/s41564-025-02134-0>
- Thongsripong R, Chandler JA, Kittayapong P *et al*. Metagenomic shotgun sequencing reveals host species as an important driver of virome composition in mosquitoes. *Sci Rep* 2021;**11**:8448. <https://doi.org/10.1038/s41598-021-87122-0>
- Valles SM, Chen Y, Firth AE *et al*. ICTV virus taxonomy profile: *Dicistroviridae*. *J Gen Virol* 2017a;**98**:355–6. <https://doi.org/10.1099/jgv.0.000756>
- Valles SM, Chen Y, Firth AE *et al*. ICTV virus taxonomy profile: *Iflaviridae*. *J Gen Virol* 2017b;**98**:527–8. <https://doi.org/10.1099/jgv.0.000757>
- Valles SM, Bell S, Firth AE. *Solenopsis invicta* virus 3: mapping of structural proteins, ribosomal frameshifting, and similarities to *Acyrtosiphon pisum* virus and Kelp fly virus. *PLoS One* 2014;**9**:1–12. <https://doi.org/10.1371/journal.pone.0093497>
- Valles SM, Oi DH, Becnel JJ *et al*. Isolation and characterization of *Nylanderia fulva* virus 1, a positive-sense, single-stranded RNA virus infecting the tawny crazy ant, *Nylanderia fulva*. *Virology* 2016;**496**:244–54. <https://doi.org/10.1016/j.virol.2016.06.014>
- Vinje J, Estes MK, Esteves P *et al*. ICTV virus taxonomy profile: *Caliciviridae*. *J Gen Virol* 2019;**100**:1469–70. <https://doi.org/10.1099/jgv.0.001332>
- Wang Z, Wen H. A review of the recombination events, mechanisms and consequences of Coxsackievirus A6. *Infect Med* 2024;**3**:100115. <https://doi.org/10.1016/j.imj.2024.100115>
- Wong T, Ly-Trong N, Ren H *et al*. IQ-TREE 3: Phylogenomic inference software using complex evolutionary models. *EcoEvoRxiv*. 2025. [doi.org/10.32942/X2P62N](https://doi.org/10.32942/X2P62N)
- Yin C, Zhao H, Xia X *et al*. Picornavirus 2C proteins: structure-function relationships and interactions with host factors. *Front Cell Infect Microbiol* 2024;**14**:1347615. <https://doi.org/10.3389/fcimb.2024.1347615>
- Zell R, Delwart E, Gorbalenya AE *et al*. ICTV virus taxonomy profile: *Picornaviridae*. *J Gen Virol* 2017;**98**:2421–2. <https://doi.org/10.1099/jgv.0.000911>
- Zell R, Knowles NJ, Simmonds P. A proposed division of the family *Picornaviridae* into subfamilies based on phylogenetic relationships and functional genomic organization. *Arch Virol* 2021;**166**:2927–35. <https://doi.org/10.1007/s00705-021-05178-9>
- Zhang Y, Ye Z-X, Feng X-X *et al*. Prevalence of reversed genome organizations for viruses in the family *Iflaviridae*, order *Picornavirales*. *Microbiol Spectrum* 2023;**11**:e04738–22. <https://doi.org/10.1128/spectrum.04738-22>