Novel lineage of anelloviruses with large genomes identified in dolphins

Matthew D de Koch, Mart Krupovic, Russell Fielding, Kendal Smith, Kelsie Schiavone, Katharine R Hall, Vincent S Reid, Diallo Boyea, Emma L Smith, Kara Schmidlin, et al.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Virology | Full-Length Text

Novel lineage of anelloviruses with large genomes identified in dolphins

Matthew D. De Koch,[1] Mart Krupovic,[2] Russell Fielding,[3] Kendal Smith,[1] Kelsie Schiavone,[4] Katharine R. Hall,[4] Vincent S. Reid,[5] Diallo Boyea,[6] Emma L. Smith,[7] Kara Schmidlin,[1] Rafaela S. Fontenele,[1] Eugene V. Koonin,[8] Darren P. Martin,[9] Simona Kraberger,[1] Arvind Varsani[1,10]

AUTHOR AFFILIATIONS See affiliation list on p. 15.

ABSTRACT Anellovirus infections are ubiquitous in mammals but lack any clear disease association, suggesting a commensal virus-host relationship. Although anelloviruses have been identified in numerous mammalian hosts, their presence in members of the family Delphinidae has yet to be reported. Here, using a metagenomic approach, we characterize complete anellovirus genomes (n = 69) from four Delphinidae host species: short-finned pilot whale (*Globicephala macrorhynchus*, n = 19), killer whale (*Orcinus orca*, n = 9), false killer whale (*Pseudorca crassidens*, n = 6), and pantropical spotted dolphin (*Steno attenuatus*, n = 1). Sequence comparison of the open reading frame 1 (ORF1) encoding the capsid protein, the only conserved gene shared by all anelloviruses, shows that the Delphinidae anelloviruses form a novel genus-level clade that encompasses 22 unique species-level groupings. We provide evidence that different Delphinidae species can be co-infected by multiple anelloviruses belonging to distinct species groupings. Notably, the ORF1 protein of the Delphinidae anelloviruses is considerably larger than those encoded by all previously described anelloviruses from other hosts (spanning 14 vertebrate orders and including 27 families). Comprehensive analysis of the ORF1 sequences and predicted protein structures showed that the increased size of these proteins results from divergent elaborations within the capsid-distal P2 subdomain and elongation of the C-terminal domain of ORF1. Comparative structural and phylogenetic analyses suggest that acquisition of the P2 subdomain and its diversification occurred convergently in the anelloviruses associated with primate and Delphinidae hosts. Collectively, our results further the appreciation of diversity and evolution of the ubiquitous and enigmatic viruses in the family *Anelloviridae*.

IMPORTANCE Anelloviruses are ubiquitous in mammals, but their infection has not yet been linked to any disease, suggesting a commensal virus-host relationship. Here, we describe the first anelloviruses associated with diverse species of dolphins. The dolphinid anelloviruses represent a new genus (tentatively named "Qoptorquevirus") and encode open reading frame 1 (ORF1) (capsid) proteins that are considerably larger than those encoded by previously described anelloviruses from other hosts. Comprehensive analysis of the ORF1 sequences and predicted protein structures revealed the underlying structural basis for such an extravagant ORF1 size and suggested that ORF1 size increased convergently in the anelloviruses associated with primate and Delphinidae hosts, respectively. Collectively, our results provide insights into the diversity and evolution of *Anelloviridae*. Further exploration of the anellovirus diversity, especially in the host species that have not yet been sampled, is expected to further clarify their evolutionary trajectory and explain the unusual virus-host commensal relationship.

KEYWORDS pantropical spotted dolphin, killer whale, short-finned pilot whale, false killer whale, Delphinidae, *Anelloviridae*, *Qoptorquevirus*, single jelly-roll fold

Editor Koenraad van Doorslaer, College of Agriculture & Life Sciences, University of Arizona, Tucson, Arizona, USA

Address correspondence to Mart Krupovic, mart.krupovic@pasteur.fr, or Arvind Varsani, arvind.varsani@asu.edu.

The authors declare no conflict of interest.

Received 5 August 2024
Accepted 27 November 2024
Published 12 December 2024

Anelloviruses have small icosahedral virions that encapsidate negative-sense, single-stranded circular DNA genomes. These viruses have been identified in a variety of terrestrial vertebrate hosts, yet their prevalence and diversity in marine mammals remain poorly understood (1). Anelloviruses are classified into the family *Anelloviridae*, which currently includes ~30 genera comprising ~150 species, although due to the extensive diversity of anelloviruses, their taxonomy is in constant flux. The vast majority of the known anellovirus hosts are mammals, with members of only one genus infecting birds (1, 2). Mammalian anellovirus genomes range in size from 1.6 to 3.9 kb and contain one large and two or three smaller protein-coding open reading frames (ORFs). The exact functions of these genes are not entirely known, especially those of the smaller ORFs. ORF1 is the signature ORF of anelloviruses and encodes a single jelly-roll (JR) capsid protein with an N-terminal arginine-rich domain, similar to the DNA-binding R-arm found in the capsid proteins of circoviruses and many other viruses with icosahedral capsids and ssRNA or ssDNA genomes (3). The smaller ORF2 contains a serine-rich region, which likely binds template DNA (4). However, the molecular mechanisms of anellovirus replication remain enigmatic because their small genomes encode no proteins homologous to any known replication enzymes (3), and neither helper viruses nor the involvement of the host replication machinery has been demonstrated.

Following the initial discovery of an anellovirus (named torque teno virus) in the blood of a post-transfusion patient from Japan in 1997 (5), related viruses have been found in various animals including non-human primates (6–8), pinnipeds (9–11), birds (12), felids (13, 14), bears (3, 15), and rodents (16, 17). Transmission of anelloviruses is believed to mainly occur via the fecal-oral and saliva routes (18).

The key characteristic of anelloviruses, which continues to puzzle researchers, is their ubiquity in vertebrates without a clear disease association. Anellovirus prevalence has been shown to range from 5% to 90% in human populations, with initial presence often detected in the first few months following birth (4). Anellovirus DNA has been detected in whole blood, saliva, tissues, and feces (19–23). Although a few studies have suggested possible associations with certain viral infections (24) and disease (25–27), detailed investigation of these potential relationships has been hindered by the omnipresence of anelloviruses and the lack of robust cell culture models (28). One popular hypothesis is that anelloviruses are commensals or even mutualists of their hosts, being a part of the so-called healthy human virome (4, 29).

Although anelloviruses have been observed in numerous terrestrial mammalian species, very few studies have investigated their prevalence and diversity in marine mammals. Currently, pinnipeds remain the only marine mammals that have been assessed for the presence of anelloviruses (3, 9–11, 30). To close this knowledge gap, we accessed archived samples collected for other studies of Delphinidae species in the Caribbean. In particular, we analyzed the presence of anelloviruses in tissue samples from four species: short-finned pilot whale (*Globicephala macrorhynchus*), killer whale (*Orcinus orca*), false killer whale (*Pseudorca crassidens*), and pantropical spotted dolphin (*Stenella attenuata*). Although Delphinidae population demographics are often difficult to accurately estimate, historic whaling data indicate that the four species studied here are among the most abundant cetacean species in St. Vincent and the Grenadines (31). We have previously identified a polyomavirus and cressdnaviricots in killer whale and short-finned pilot whale samples (32). In this study, we build on this previous work with a focus on anelloviruses. We recovered 69 anellovirus genomes that fall into 22 species groupings (TTDelV g1 through g22) within a putative new genus of Delphinidae-infecting anelloviruses. Comprehensive phylogenetic analyses coupled with comparative structural analysis of the ORF1 proteins provided insights into the diversity and evolution of these viruses.

MATERIALS AND METHODS

Sample collection

Tissue samples from four Delphinidae species (short-finned pilot whale, killer whale, false killer whale, and pantropical spotted dolphin) were obtained through collaboration with artisanal subsistence whalers from the Caribbean island of St. Vincent, where local communities have carried out this traditional practice for more than a century (33). The various tissue samples (kidney, liver, or muscle) were collected from the deceased animals with the permission of the whalers and associated partners.

The tissue samples were first imported to the University of the South (Sewanee, TN, USA), and a subsample of each sample was transported to Arizona State University (Tempe, AZ, USA). These subsamples were stored at −80°C until processing.

Viral nucleic acid isolation and high-throughput sequencing

The tissue samples (kidney, liver, or muscle) were collected from 55 marine mammals between 2015 and 2016 (Table 1). From each sample, ~5 g of tissue was individually homogenized in SM buffer (G-Biosciences, USA). Viral DNA was then extracted from 200 µL of each homogenate, using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA), following the manufacturer's protocol. To enrich circular DNAs, we carried out rolling-circle amplification (RCA), using the phi29 DNA Polymerase Kit (Watchmaker Genomics, USA).

The RCA products were pooled per animal species, and these were then used to generate libraries using the TruSeq DNA Nano kit (Illumina, Inc.) and sequenced (2 × 100 bp) on an Illumina HiSeq4000 sequencer at Macrogen Inc. (USA). The raw reads were quality trimmed using Trimmomatic version 0.39 (34) and then *de novo* assembled with MEGAHIT version 1.2.9 (35). *De novo*-assembled contigs > 750 nts were screened against a local NCBI viral protein RefSeq database (Release 220) using BLASTx (36).

Based on the *de novo*-assembled anellovirus contigs, we designed a pair of abutting primers to screen each sample and recover complete genomes by PCR. In each reaction, one forward primer (cetacean anellovirus F: 5′-TGA GTT TAC TGC GCS AGY GGT CAA T-3′) was paired with one of two reverse primers (cetacean anellovirus R: 5′-GCC ATT CGT CAC CCC ACT TAC TTA T-3′ or 5′-GCC ATT CGT CAG TGT AGT TAC TTA T-3′). PCR cycling conditions were applied according to primer annealing temperatures of 55°C and the manufacturer's recommendations. The amplicons were resolved on a 0.7% agarose gel, and the amplicons of ~3.5–3.8 kb were excised and purified using the MEGAquick-spin Plus Total Fragment DNA Purification Kit (iNtRON Biotechnology, South Korea). Then, the purified amplicons from each PCR were ligated into the pJET 1.2 vector (Thermo Fisher Scientific, USA) and transformed into XL blue *Escherichia coli*-competent cells. DNA from the recombinant plasmid was extracted using the DNA-spin Plasmid DNA Purification Kit (iNtRON Biotechnology, South Korea) and sequenced at Plasmidsaurus, USA, using Oxford Nanopore long-read sequencing technology.

Anellovirus sequence analysis

Anellovirus genomes were annotated with Geneious Prime (Dotmatics, USA). A ORF1 data set was assembled from the genomes identified in this study as well as those of representatives from each species available in GenBank (downloaded on 1st December 2023).

The ORF1 sequences were translated and aligned with MAFFT version 7.113 (37). The alignment was trimmed using TrimAl with a 0.2 gap threshold (38), and the resulting trimmed alignment was used to infer a maximum likelihood phylogenetic tree using PhyML 3 (39) using the best fit substitution model VT + G + I determined using ProtTest 3 (40). Branches with less than 0.8 aLRT branch support were collapsed using TreeGraph2 version 2.14 (41).

The ORF1 amino acid sequence lengths characterized in this study were compared to those of representative species across 14 distinct host orders using Rstudio (42) with

TABLE 1 Summary of sample information for all anelloviruses recovered in this study including source/host, Delphinidae host family demographic information, sampling location, sample date, sample type, anellovirus species grouping, and anellovirus genome accession number

Sample ID	Host species	Sampling location	Sample date	Sample type	Virus species grouping	Accession no.
USW2	*Stenella* spp.	St. Vincent island	3-August-15	Kidney	TTDelV g18	PP790224
					TTDelV g14	PP790225
USW5	*Orcinus orca*	St. Vincent island	26-August-15	Muscle, fresh	TTDelV g20	PP790226
					TTDelV g22	PP790227
					TTDelV g22	PP790228
USW7	*Orcinus orca*	St. Vincent island	26-August-15	Liver	TTDelV g21	PP790229
					TTDelV g19	PP790230
					TTDelV g16	PP790231
USW8	*Orcinus orca*	St. Vincent island	26-August-15	Muscle, fresh	TTDelV g20	PP790232
					TTDelV g20	PP790233
USW9	*Globicephala macrorhynchus*	St. Vincent island	22-September-15	Muscle, fresh	TTDelV g8	PP790234
					TTDelV g5	PP790235
					TTDelV g8	PP790236
USW10	*Globicephala macrorhynchus*	St. Vincent island	22-September-15	Kidney	TTDelV g8	PP790237
					TTDelV g8	PP790238
					TTDelV g8	PP790239
USW11	*Globicephala macrorhynchus*	St. Vincent island	28-September-15	Muscle, fresh	TTDelV g8	PP790240
					TTDelV g8	PP790241
USW12	*Pseudorca crassidens*	St. Vincent island	7-October-15	Muscle, fresh	TTDelV g3	PP790242
					TTDelV g3	PP790243
					TTDelV g3	PP790244
USW13	*Pseudorca crassidens*	St. Vincent island	12-October-15	Muscle, fresh	TTDelV g4	PP790245
USW15	*Pseudorca crassidens*	St. Vincent island	29-October-15	Liver	TTDelV g3	PP790246
					TTDelV g3	PP790247
USW16	*Pseudorca crassidens*	St. Vincent island	29-October-15	Kidney	TTDelV g8	PP790248
					TTDelV g3	PP790249
USW17	*Pseudorca crassidens*	St. Vincent island	29-October-15	Muscle, fresh	TTDelV g3	PP790250
					TTDelV g3	PP790251
					TTDelV g3	PP790252
USW18	*Pseudorca crassidens*	St. Vincent island	4-November-15	Liver	TTDelV g6	PP790253
					TTDelV g4	PP790254
					TTDelV g2	PP790255
USW22	*Globicephala macrorhynchus*	St. Vincent island	13-March-16	Muscle, fresh	TTDelV g8	PP790256
USW23	*Globicephala macrorhynchus*	St. Vincent island	13-March-16	Liver	TTDelV g8	PP790257
					TTDelV g8	PP790258
USW24	*Globicephala macrorhynchus*	St. Vincent island	13-March-16	Liver	TTDelV g8	PP790259
					TTDelV g8	PP790260
USW25	*Globicephala macrorhynchus*	St. Vincent island	13-March-16	Kidney	TTDelV g1	PP790261
					TTDelV g1	PP790262
USW26	*Globicephala macrorhynchus*	St. Vincent island	13-March-16	Muscle, fresh	TTDelV g5	PP790263
					TTDelV g5	PP790264
USW28	*Globicephala macrorhynchus*	St. Vincent island	15-March-16	Muscle, fresh	TTDelV g8	PP790265
					TTDelV g13	PP790266
USW29	*Orcinus orca*	St. Vincent island	28-April-16	Muscle, fresh	TTDelV g8	PP790267
USW30	*Orcinus orca*	St. Vincent island	28-April-16	Liver	TTDelV g7	PP790268
					TTDelV g11	PP790269
					TTDelV g2	PP790270
USW35	*Orcinus orca*	St. Vincent island	28-April-16	Liver	TTDelV g17	PP790271
USW36	*Orcinus orca*	St. Vincent island	28-April-16	Muscle, fresh	TTDelV g15	PP790272
USW39	*Globicephala macrorhynchus*	St. Vincent island	27-May-16	Muscle, fresh	TTDelV g2	PP790273
					TTDelV g5	PP790274

(*Continued on next page*)

TABLE 1 Summary of sample information for all anelloviruses recovered in this study including source/host, Delphinidae host family demographic information, sampling location, sample date, sample type, anellovirus species grouping, and anellovirus genome accession number (*Continued*)

Sample ID	Host species	Sampling location	Sample date	Sample type	Virus species grouping	Accession no.
					TTDelV g8	PP790275
USW40	*Globicephala macrorhynchus*	St. Vincent island	27-May-16	Muscle, fresh	TTDelV g2	PP790276
USW42	*Orcinus orca*	St. Vincent island	9-June-16	Muscle, fresh	TTDelV g3	PP790277
					TTDelV g3	PP790278
USW44	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Liver	TTDelV g10	PP790279
					TTDelV g8	PP790280
USW45	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Liver	TTDelV g1	PP790281
USW48	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Liver	TTDelV g8	PP790282
					TTDelV g5	PP790283
					TTDelV g8	PP790284
USW49	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Kidney	TTDelV g1	PP790285
USW50	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Liver	TTDelV g3	PP790286
					TTDelV g3	PP790287
					TTDelV g12	PP790288
USW51	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Muscle, fresh	TTDelV g3	PP790289
USW52	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Kidney	TTDelV g8	PP790289
USW54	*Globicephala macrorhynchus*	St. Vincent island	23-June-16	Muscle, fresh	TTDelV g9	PP790291
USW55	*Orcinus orca*	St. Vincent island	9-June-16	Muscle, fresh	TTDelV g5	PP790292

ggplot2 (43) and tidyr (44) packages. To visualize relative ORF1 amino acid sequence lengths, these data were compiled to construct a jitter dotplot.

Pairwise identity analyses for all ORF1 nucleotide and amino acid sequences were determined using SDT version 1.2 (45). Clinker (46) was used to illustrate genome organization and show the percentage identity of the encoded protein sequences (ORF1, ORF2, and ORF3) of representative genomes from the 22 TTDelV species groupings.

Recombination events evident within the whole genome alignment were detected, and recombination breakpoint distributions were analyzed with RDP5.57 (47) using default parameters. For optimal recombination detection, sequences were auto-masked. Only events detected with >3 different recombination detection methods implemented in RDP5 with phylogenetic support for recombination and a *P*-value of <0.05 were considered credible.

The number of arginine (R) and lysine (K) were counted in the N terminus of the ORF1 protein of all the representative anelloviruses and used to determine a Pearson correlation coefficient with *orf1* gene and genome size using DATAtab (48).

Protein structure prediction and analysis

ORF1 structures of anelloviruses representing all 22 species described in this study were modeled using AlphaFold2 (49) via a local installation of ColabFold version 1.5.5 (50). Multiple sequence alignments of ORF1 from anelloviruses of delphinids and other hosts were used as an input for modeling, with six recycles. When appropriate, the models were further refined using DeepFold (51) with zero recycles. The quality of the generated structural models was assessed using the local distance difference test (52). The obtained ORF1 structural models of TTDelVs were compared using DALI (53) to the previously published ORF1 models representing all established genera of the *Anelloviridae* family (3). For a more thorough comparison, the structures were aligned using the MatchMaker algorithm (54). Protein structures and structural models were visualized using ChimeraX version 1.7.1 (55).

RESULTS AND DISCUSSION

Characterization of anelloviruses from four delphinid species

In this study, a total of 69 complete anellovirus genomes (size range of 3,480–3,780 nts) were sequenced from 35 individuals from four Delphinidae species. A total of 36 anellovirus genomes were recovered from the short-finned pilot whales (n = 19), 17 genomes from the killer whales (n = 9), 14 genomes from the false killer whales (n = 6), and 2 genomes from the pantropical spotted dolphin (n = 1). On average, about two distinct anelloviruses were recovered from each sample (Table 1).

Anelloviruses were identified from three tissue types. In total, 11 anellovirus genomes were recovered from kidney samples (n = 6); 25 genomes from liver samples (n = 11); and 33 genomes from muscle samples (n = 18) (Table 1). We did not find any clear correlation between organ types and anelloviruses identified. Although the majority of anelloviruses were recovered from the muscle samples, these samples represented ~50% of our sample types.

Open reading frames ORF1, ORF2, and ORF3 were identified and annotated for all 69 genomes (deposited in GenBank under accession numbers PP790224–PP790292). The killer whale samples had the largest anellovirus genome size range, spanning from 3,480 to 3,780 nts. Anelloviruses from the short-finned pilot whales had a similar range, from 3,505 to 3,766 nts. Those from the false killer whales had a slightly smaller range, from 3,619 to 3,754 nts. Finally, the pantropical spotted dolphin sequences had the smallest genomes, from 3,500 to 3,541 nts. However, this dolphin species (n = 1) is the smallest number of recovered anellovirus genomes among the four species. Thus, although in terms of the genome length distribution, TTDelVs fall within the range of previously described anellovirus genomes (1,600–3,900 nts), they clearly include some of the largest genomes described thus far. Notably, all other known anelloviruses that have genomes larger than 3,500 nts infect primates (3).

Distribution of pairwise genetic distances between delphinid anelloviruses

According to the current ICTV taxonomy guidelines, a 69% ORF1 pairwise nucleotide identity is used as a species demarcation threshold for the classification of viruses in the family *Anelloviridae* (2). Based on this threshold, we can group the 69 anelloviruses from this study into 22 unique species-level assemblages—torque teno Delphinidae virus groups 1 through 22 (TTDelV g1 through g22; Fig. 1; Data S1). Genome organization of the representative genomes from these 22 unique species level assemblages is provided in Fig. 1. The ORF1 nucleotide pairwise identity values range from 55% to 99% similarity among all the 69 TTDelVs (Data S1). At a full genome level, the 69 Delphinidae anelloviruses share 59%–99% identity (Data S1).

The BLASTx analysis revealed that cetacean-infecting anelloviruses are most closely related to porcine-infecting viruses in the genus *Kappatorquevirus*. Given that cetaceans (family Delphinidae) and porcines (family Suidae) both belong to the same mammalian order, Artiodactyla, this observation is compatible with the coevolution of Artiodactyla-infecting anelloviruses with their hosts, following the divergence of the corresponding host lineages from their last common ancestor. Molecular analyses among groups in the Artiodactyla order indicate that the Delphinidae and Suidae families diverged ~50 million years ago (56).

Anellovirus ORF1 protein phylogeny and species groupings

In the phylogenetic tree of the ORF1 amino acid sequences of the anelloviruses characterized in this study and those available from GenBank, the Delphinidae-infecting anelloviruses clustered together in a distinct clade (Fig. 2), representing a putative new genus (proposed genus name "Qoptorquevirus"). The Delphinidae hosts that were considered here all belong to the infraorder Cetacea and the order Artiodactyla. Many of the host tanglegram lines span across several of the 22 TTDelV species groupings. In particular, of the 35 individual cetacean hosts, 23 harbored more than one variant of the

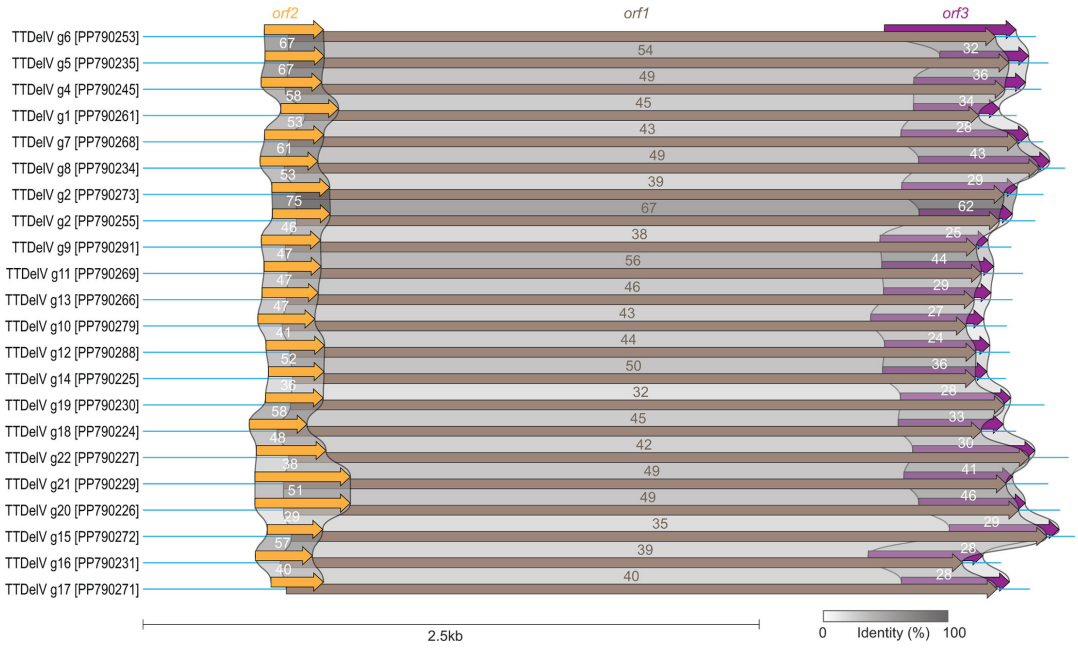

FIG 1 Genome organization of TTDelVs. Genome organization and percentage identity of the encoded protein sequences of ORF1, ORF2, and ORF3 of representative genomes from the 22 TTDelV species groupings.

anellovirus. Moreover, 11 individual hosts harbored multiple anelloviruses across more than one TTDelV species group. Thus, co-infections with diverse anelloviruses occur across these cetacean hosts, but notably, not beyond the infraorder Cetacea. Individuals co-infected with multiple anelloviruses have been documented broadly in several host groups, including humans (4, 23, 57), pinnipeds (11, 30), ursids (3, 58), rodents (16, 59), and felids (13, 60).

In a few cases, TTDelV species groupings coincide with cetacean host species. For example, 18 of the 20 anellovirus genomes from TTDelV g8 and five of the six anellovirus genomes from TTDelV g5 were identified from short-finned pilot whales. In addition, 9 of the 14 anellovirus genomes from TTDelV g3 were identified in false killer whales. In the majority of the other anellovirus species groupings, however, greater host diversity is observed, suggesting the occasional cross-species transmissions of TTDelVs. At the higher host taxonomic level, anelloviruses appear to have coevolved with their hosts, with no observed spillover to other animals thus far.

Delphinid anelloviruses display recombination patterns similar to those in other anellovirus groups

The 69 genomes displayed evidence of 30 independent recombination events with detectable recombination breakpoints clustered in the non-coding genome region. This pattern, with recombination breakpoint hotspots in the non-coding region and very few breakpoints detectable in ORF1 (Fig. 3), mirrors that observed in anelloviruses from multiple different genera (3, 11, 13, 23, 61).

In all but 2 of the 30 detected recombination events, one or both of the parental sequences shared <90% sequence identity with the respective genome regions that contributed to recombinants. Furthermore, for 13 of the 30 recombination events, only one parental sequence was identified (Data S2). Together, these observations suggest that the 69 sequences we analyzed comprise a severely under-representative sample of Delphinidae anelloviruses.

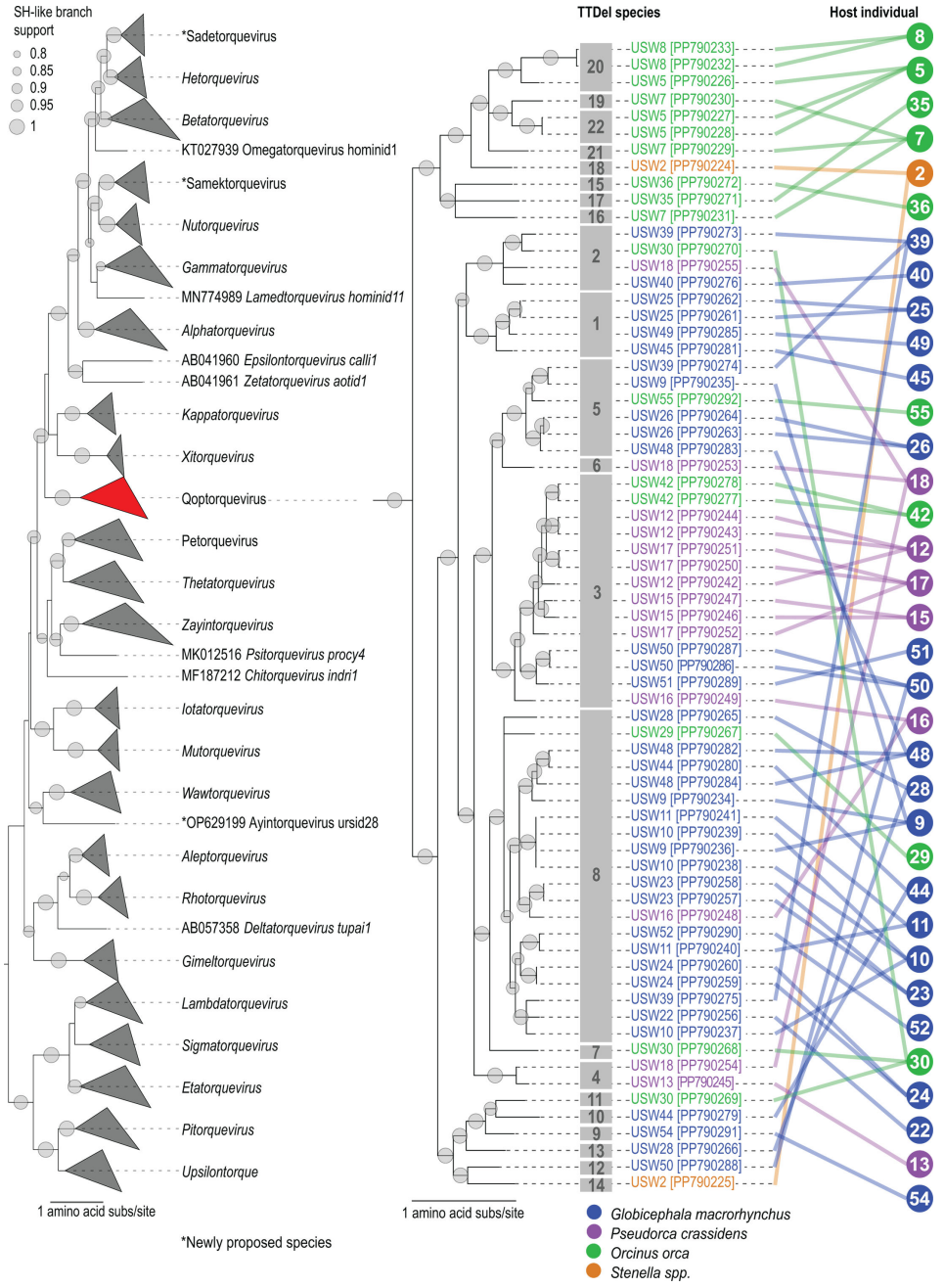

FIG 2 Anelloviruses sampled from Delphinidae species cluster together and reveal diverse co-infections. The maximum-likelihood phylogenetic tree on the left depicts the evolutionary relationships among representative anellovirus species ORF1 proteins together with those recovered from Delphinidae family hosts in this study. All the anellovirus genomes recovered here cluster together, across 22 TTDelV species groupings, in one large clade (shown in red). The clade of the Delphinidae is shown on the right, with a tanglegram for each to show the individual and host species each genome was recovered from. These virus-host relationships reveal co-infection dynamics among the Delphinidae family anelloviruses.

Delphinid anelloviruses encode by far the largest ORF1 proteins within the family *Anelloviridae*

Comparative analysis of the ORF1 amino acid sequence lengths from representative species associated with 14 distinct host orders showed that those from Delphinidae anelloviruses have by far the largest ORF1 proteins detected to date. In TTDelVs, ORF1 proteins range in length from 910 to 1,020 amino acids (*orf1* genes 2,733–3,063 nts),

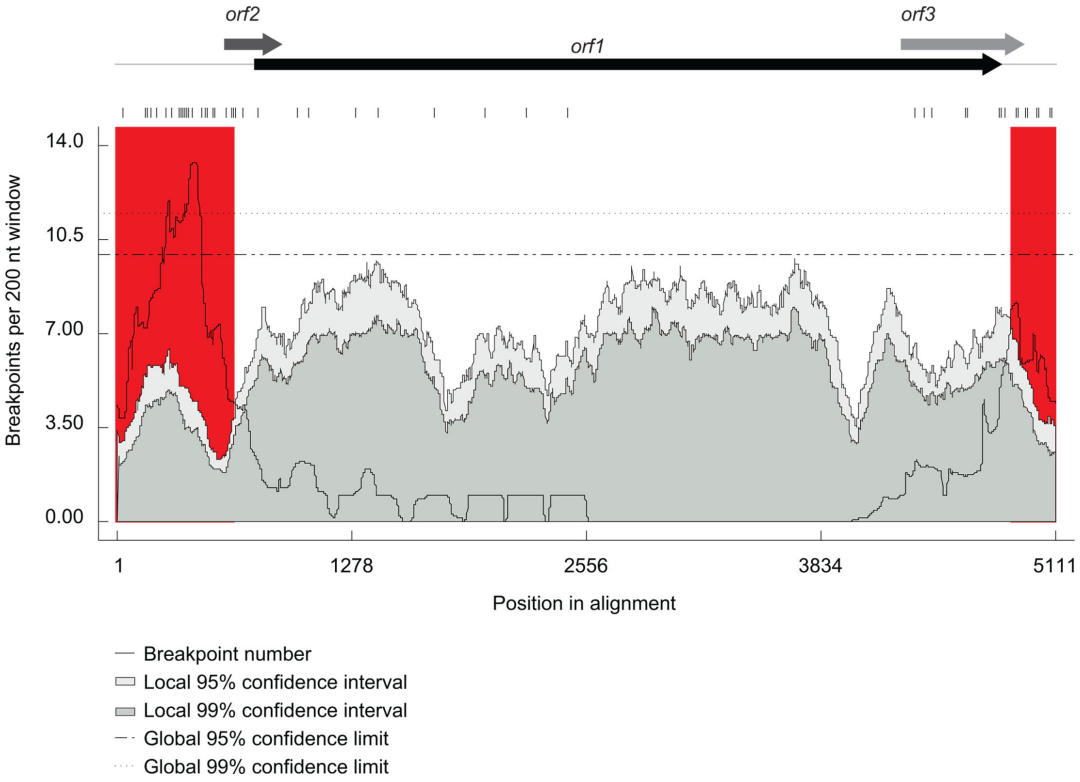

FIG 3 Distribution of recombination breakpoints along delphinid anellovirus genomes. Breakpoint hotspots are indicated in red. The light and dark sage color areas of the plots indicate 95% and 99% confidence intervals, respectively.

whereas ORF1 from anelloviruses infecting all other hosts are considerably shorter (342–773 amino acids; *orf1* genes 1,028–2,022 nts; Fig. 4A). ORF1s of anelloviruses from Suidae family hosts (GenBank accession numbers AB076001, AY823991, and JQ406846), the closest mammalian relatives of cetaceans, contain 624–635 amino acids. Thus, ORF1s of Delphinidae-derived anelloviruses are 247 amino acids longer than the previously known largest ORF1 proteins (773 amino acids; *Alphatorquevirus cerco3*; KP296853) encoded by members of the *Alphatorquevirus* genus (3). Notably, Delphinidae-derived anelloviruses break the trend observed previously, namely, that the genome length of anelloviruses is correlated with the ORF1 length (3).

Analysis of the number of arginine (R) and lysine (K) residues in the N terminus of ORF1 relative to the genome and *orf1* gene lengths showed that there is a positive correlation (r = 0.83 for genome length and r = 0.74 for *orf1* gene length; Fig. 4B). This result is consistent with that obtained by Requião et al. (62) showing that the positively charged arginine domains located in the N-terminal region of capsid proteins of diverse RNA and DNA viruses are positively correlated with the genome size.

Comparative analysis of the genomes and ORF1 lengths for representative species from each genus illustrates that although there is an overall positive relationship between genome and ORF1 lengths, there are a number of exceptions. For instance, the genomes of certain members of the *Alphatorquevirus*, *Epsilontorquevirus,* and *Zetatorquevirus* are larger than those of delphinid anelloviruses, yet their ORF1 proteins are substantially shorter. Overall, among the representative anellovirus sequences, *orf1* of TTDelV species is disproportionately larger than ORF1 of all previously described anelloviruses (turquoise bars; Fig. 5). In particular, the *orf1* in *Alphatorquevirus* members with similar sized or larger genomes are about 500 nts shorter than *orf1* sequences in TTDelV species (burgundy bars; Fig. 5). It does appear, however, that among Delphinid anelloviruses themselves, there is a positive correlation between *orf1* length and the genome length.

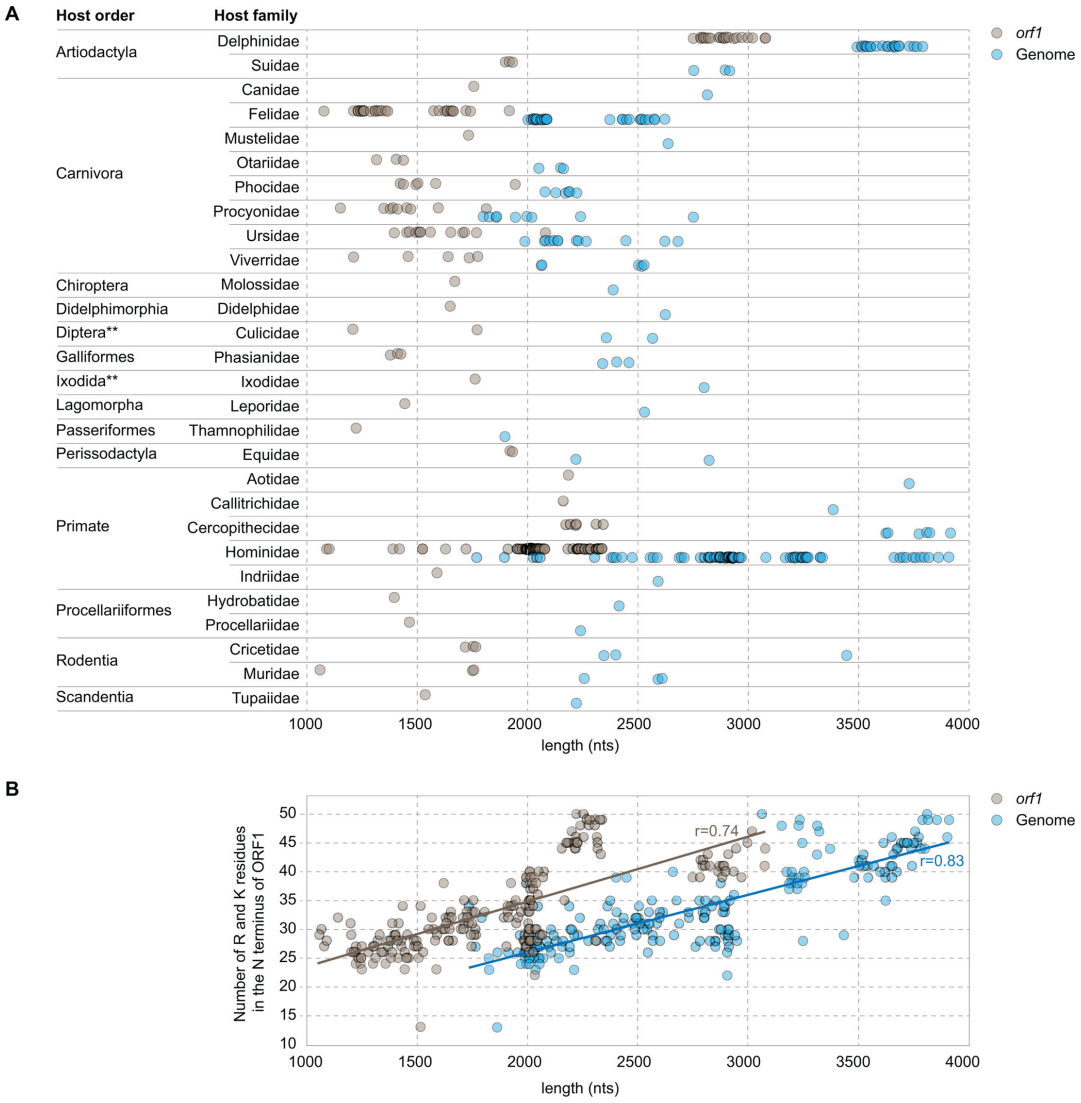

FIG 4 Anellovirus genome and *orf1* length distributions. (A) Distribution of anellovirus genome and *orf1* lengths is plotted for each host order and family. Circles represent individual anelloviruses that are representative of indicated host species. *The order Artiodactyla includes the Delphinidae family, which is associated with the 22 TTDelV species groupings of anelloviruses sequenced from oceanic dolphins in this study, and the Suidae family with three anelloviruses (GenBank accession numbers AB076001, AY823991, and JQ406846) identified in swine hosts, in separate studies. **The orders Diptera and Ixodida include invertebrate species, which are unlikely hosts for anelloviruses. Thus, such anelloviruses are likely derived from the blood meal of their vertebrate hosts. In this analysis, all Delphinidae family anelloviruses contain the largest ORF1-encoding genes. (B) Distribution of the number of arginine (R) and lysine (K) residues in the N-terminus of representative ORF1 proteins in relation to the *orf1* gene ($r = 0.74$ and *P*-value of 0.001 for the Pearson correlation) and genome ($r = 0.83$ and *P*-value of 0.001 for the Pearson correlation) sizes of anelloviruses (one from each species).

ORF1 proteins of TTDelVs contain a conserved jelly-roll core and a hypervariable projection domain

To analyze the structural differences between the extravagantly large TTDelV ORF1 proteins and those encoded by other anelloviruses, we modeled the ORF1 structures of TTDelVs representing all 22 putative species described in this study using AlphaFold2 (see Materials and Methods). The predicted local distance difference test scores for the obtained models ranged between 94.9 and 96.6, indicating high confidence (Table S1). As observed in the previously analyzed anellovirus ORF1s (3), the ORF1 proteins of TTDelVs can be structurally dissected into the central core, consisting of the jelly-roll (JR) and projection (P) domains and extended N-terminal and C-terminal regions (Fig. 6A; Fig.

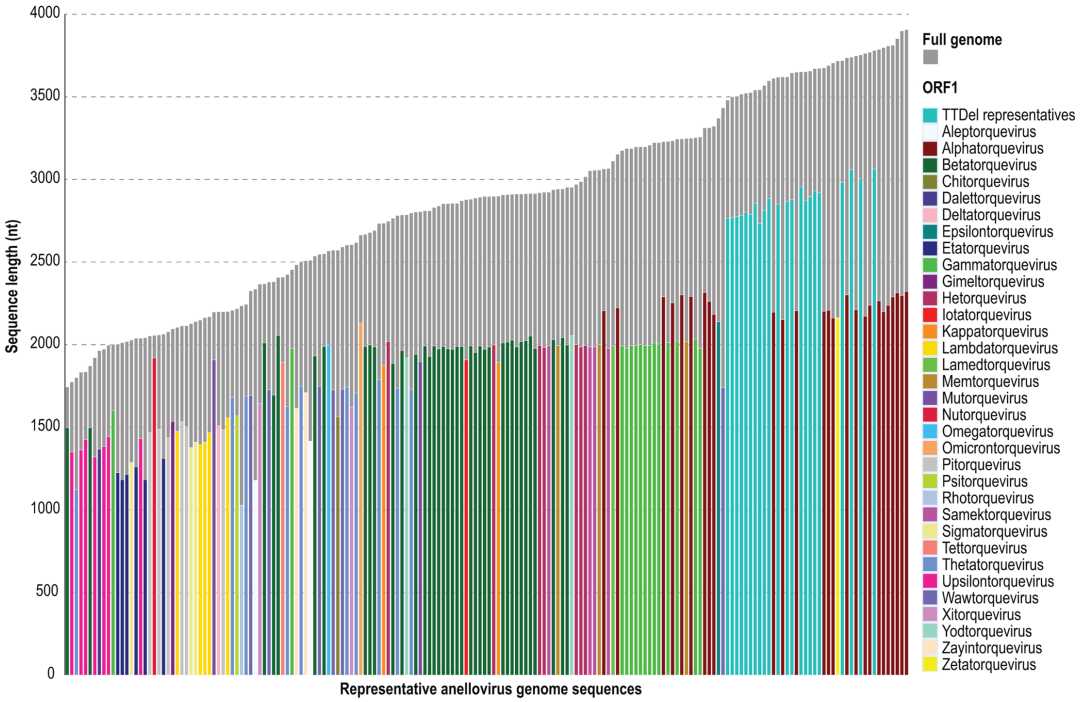

FIG 5 *orf1* to full genome length (nucleotide) ratio of representative anellovirus species. The 22 novel TTDel sequences (turquoise bars) were analyzed alongside 162 representative sequences from all 33 established anellovirus genera. *orf1* lengths and full genome lengths are organized by size.

S1). The canonical JR domain comprises eight antiparallel β-strands (B through I), which form two juxtaposed β-sheets, BIDG and CHEF (63, 64). In anelloviruses, the BIDG β-sheet is extended by an additional β-strand (C′) inserted between β-strands C and D, forming a five-stranded β-sheet C′BIDG. The P-domain is an insertion within the loop connecting the JR domain β-strands H and I (HI-loop). In the modeled monomer structures, the N- and C-terminal regions vary in terms of length, secondary structure elements, and their relative position with respect to the central core, suggesting flexibility in these regions, although within the virion, they likely occupy fixed positions. The same domain organization was found in all TTDelV ORF1 proteins, including the largest ones (>1,000 amino acids) encoded by TTDelV g8, TTDelV g15, and TTDelV g22 (Fig. 6B).

The JR domain plays a key role in the formation of the icosahedral T = 1 capsid, whereas the P domain points outward and is predicted to be important for host cell surface receptor recognition and/or evasion of host immune responses (3, 65). By contrast, the terminal regions are likely implicated in capsid assembly and stabilization through interactions with the neighboring subunits and genomic DNA. In particular, the N-terminal region is enriched in arginine residues and is likely to be implicated in genome binding and nuclear localization, as suggested for circoviruses (66). Below, we restricted our analyses to the central core, i.e., JR and P domains, of ORF1 (Fig. 6A).

Comparison of the TTDelV ORF1 homologs from different species revealed the conservation of the JR domain contrasted by extensive variability of the P domain, both in terms of topology and sequence conservation (Fig. 6; Data S2). The P domain can be further divided into two subdomains, P1 and P2. The P1 subdomain is proximal to the JR domain and serves as a platform for P2, which occupies the most distal position with respect to the virion shell (3, 65). While the P1 subdomain was largely conserved among TTDelV ORF1s, P2 displayed considerable variation, with distinct secondary structure elements predicted for different homologs, including additional α-helices, β-strands, and extended loops (Fig. S2). The P2 subdomain has been hypothesized to represent a point of contact with the host immune system (65) and is thus expected to be under constant

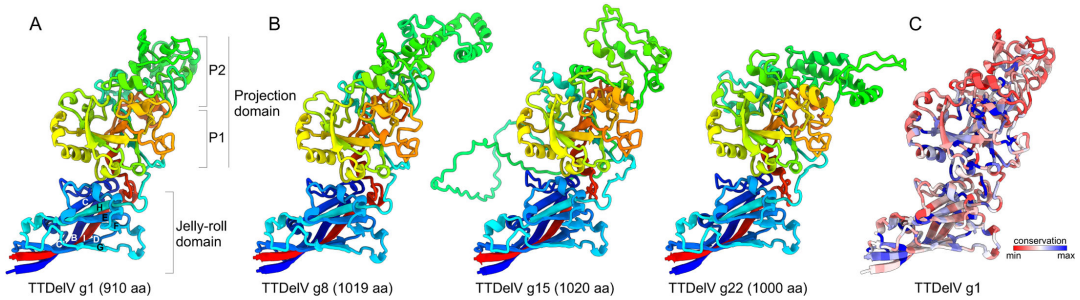

FIG 6 Structural models of selected TTDelV ORF1. (A) The central core of TTDelV ORF1 exemplified using the protein of TTDelV1, the smallest of the TTDelV ORF1s. The β-strands (B through I) constituting the jelly-roll domain are labeled. The jelly-roll and projection (P) domains as well as subdomains P1 and P2 are also indicated. (B) Structural models of the three largest ORF1 proteins encoded by TTDelVs. The sizes of the corresponding ORF1 proteins are indicated in parentheses. The structures in panels A and B are colored using the rainbow scheme from N-terminus (blue) to C-terminus (red). (C) Structural model of TTDelV1 ORF1 colored according to sequence conservation among TTDelVs representing the 22 species, with the more conserved and variable positions shown in blue and red, respectively.

pressure to change in order to avoid clearance. The observed hypervariance of P2 is consistent with this possibility.

The larger sizes of TTDelV ORF1 are attributed to the expanded central core and C-terminal domains

We next used DALI to compare the ORF1 structural models of TTDelVs to each other and the previously generated models of anellovirus capsid proteins representing all established genera (3). In this analysis, all TTDelV ORF1 homologs clustered together (Fig. 7), consistent with the sequence-based phylogenetic analysis (Fig. 2), further supporting the grouping of the 22 TTDelV species into a genus. Notably, the three largest TTDelV ORF1 homologs did not form a group, suggesting that their sizes increased independently of each other. Although the N-terminal regions were overall of similar lengths across all TTDelV ORF1 homologs, the central core and the C-terminal region varied in length. Comparison of the structural neighbors in the dendrogram showed that an increase in ORF1 size can occur by the expansion of either the central core (e.g., compare TTDelV g15 versus TTDelV g17 [679 versus 627 amino acids, respectively]) or the C-terminal region (e.g., compare TTDelV g13 and TTDelV g8 [223 versus 257 amino acids, respectively]) (Fig. 7). A similar pattern is also characteristic of anelloviruses from other genera. For instance, in the structure-based dendrogram and phylogeny, gammatorqueviruses consistently form a clade with alphatorqueviruses, betatorqueviruses, epsilontorqueviruses, hetotorqueviruses, zetatorqueviruses, and omegatorqueviruses (Fig. 2; Fig. 8). However, gammatorqueviruses encode ORF1 proteins that are >100 amino acids shorter than homologs from the other genera in this cluster, and the difference can be attributed to both smaller central core and C-terminal regions (Fig. 7).

The central core of TTDelV ORF1 is nearly twice as large as that of their homologs from other anelloviruses (on average 630 versus 361 aa, median 627 versus 348 amino acids) and is >100 amino acids larger than for the previous record holder, alphatorquevirus, with 500 amino acids. The C-terminal domains of TTDelV ORF1 are also considerably larger compared to anelloviruses from other genera, namely, on average 240 versus 124 amino acids (median 236 versus 134 amino acids). Although the N-terminal region of ORF1 in TTDelVs was slightly larger compared to other anelloviruses (on average 87 versus 65 amino acids), its contribution to the overall size of ORF1 is modest. Thus, the augmentation of the central core and the C-terminal domain is the key factor underlying the dramatic expansion in the size of TTDelVs ORF1.

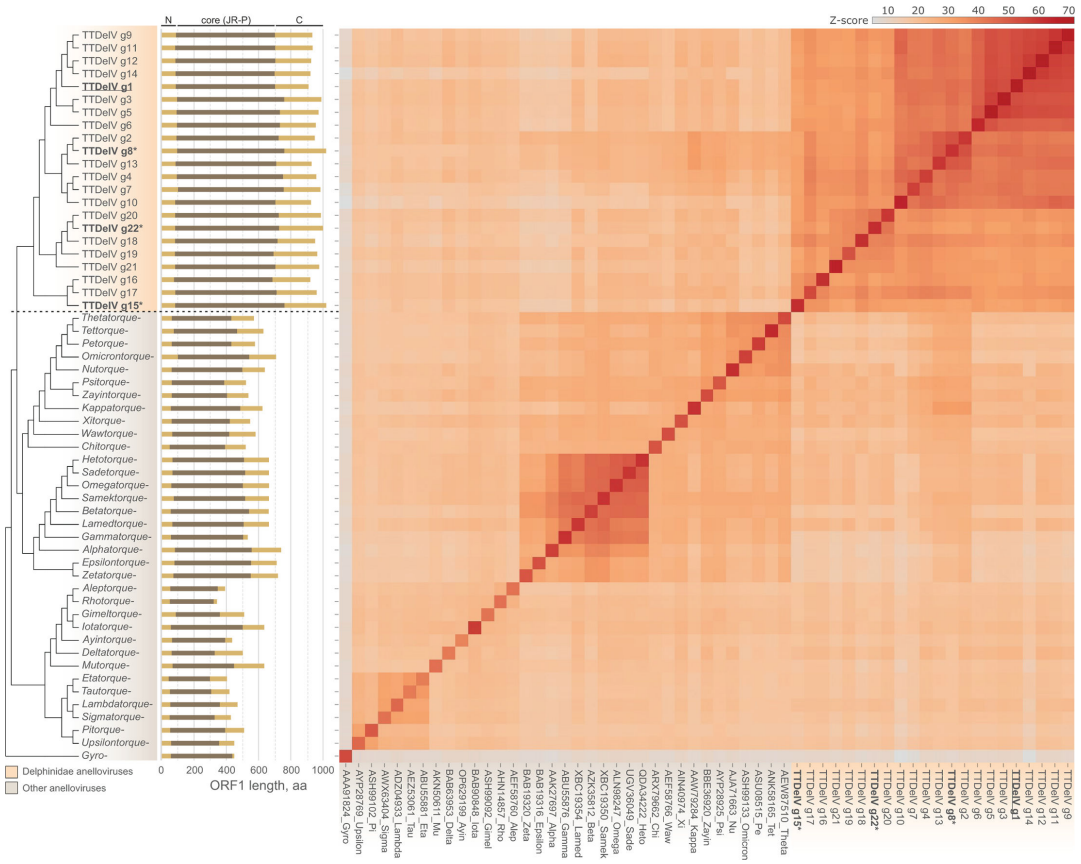

FIG 7 Structural relationships between ORF1 proteins of anelloviruses from different genera. Cluster dendrogram and heatmap were constructed on the basis of the comparison of pairwise *Z*-scores calculated using DALI. The previously classified anelloviruses and TTDeIVs are indicated with different background colors on the dendrogram. The former are labeled with an abbreviated genus name, and the corresponding ORF1 GenBank accession numbers are provided at the bottom of the heatmap. The bar plot depicts the lengths of ORF1 proteins compared with the relative positions of the terminal (N and C) domains and the core region including the jelly-roll and projection (JR-P) domains indicated in light and dark brown, respectively. The three largest ORF1 homologs (>1,000 amino acids) are indicated with asterisks, whereas the smallest one (910 amino acids, TTDelV g1) is underlined.

Analysis of the P-subdomains could guide higher rank classification of anelloviruses

We have previously shown that anelloviruses display dramatic variation in the complexity of the insertions within the HI-loop of the JR domains of ORF1 proteins and hypothesized that anelloviruses evolved from a circovirus-like ancestor with a simple JR capsid protein through acquisition and gradual augmentation of the P-domain (3). Comprehensive comparison of the predicted ORF1 structures suggests that anelloviruses can be broadly divided into four groups: (i) viruses with an insertion into the EF-loop that lack insertion into the HI-loop, represented by members of the *Gyrovirus* genus; (ii) viruses with a simple α-helical insertion into the HI-loop of the JR domain; (iii) viruses with a simple P domain consisting of the P1-subdomain alone; and (iv) viruses with an elaborate P-domain containing both P1 and P2 subdomains (Fig. 8). The latter group can be further subdivided into three subgroups based on the complexity of the P2 subdomain, and the evolution of the more complex P-domain from the simpler state through gradual augmentation can be plausibly reconstructed.

The P1 subdomain consists of a conserved core of three- or four-stranded β-sheet and a β-hairpin (stem hairpin), with the two elements being roughly perpendicular to each other. The P2 subdomain is an insertion into the apex of the loop in the P1 stem hairpin, with the β-sheet serving as a platform onto which the P2 subdomain is mounted (Fig. 8). Viruses from the genera *Wawtorqueviruses*, *Zayintorquevirus*, and

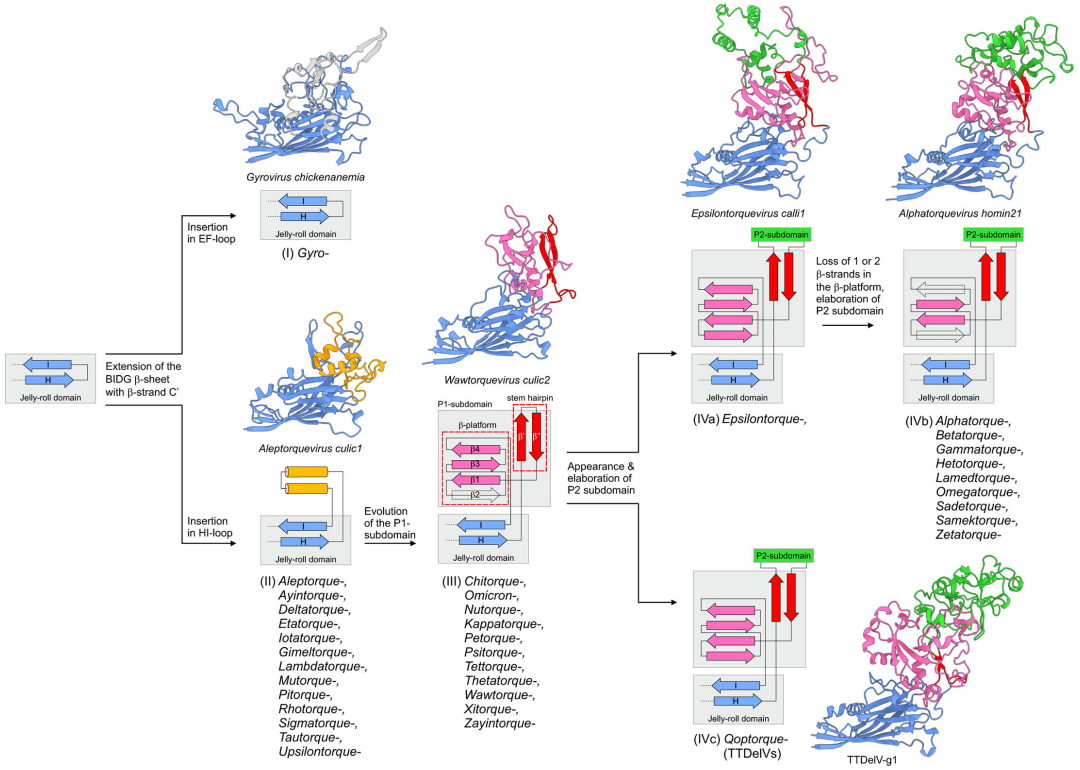

FIG 8 Putative evolutionary transitions in the capsid protein structure throughout anellovirus evolution. Each structural group is depicted by a structure of a representative ORF1 and a topological diagram of the insertion within the HI-loop of the jelly-roll domain. Only H and I β-strands of the jelly-roll domain are shown. Genera of anelloviruses that encode ORF1 of each structural group are listed below the corresponding topological diagram. The four major structural groups are indicated with Roman numbers, with group IV being further divided into three subgroups, IVa, IVb, and IVc. Major structural changes are explained next to the corresponding arrows. The structural models are colored according to the domain organization: jelly-roll domain, blue; β-platform and stem hairpin of the P1 subdomain, magenta and red, respectively; P2 subdomain, green; gyrovirus-specific insertion within the EF-loop, gray; α-helical insertion within the HI-loop, orange.

Psitorquevirus might resemble the ancestral state of the P1 subdomain, with the stem hairpin lacking any pronounced insertions in the loop. By contrast, in a number of anellovirus genera, the P1 subdomain consists of a four-stranded β-sheet and the loop connecting the two β-strands of the stem hairpin is more extended, although still lacking extensive secondary structure other than random coil. By contrast, epsilontorqueviruses already contain a more elaborate P2 subdomain with several α-helices. In phylogenetic analyses (Fig. 2), epsilontorqueviruses together with zetatorqueviruses are basal to a clade including alphatorqueviruses, betatorqueviruses, gammatorqueviruses, hetotorqueviruses, omegatorqueviruses, sadetorqueviruses, samektorqueviruses, and lamedtorqueviruses. Notably, with the exception of epsilontorqueviruses, ORF1s of all these anelloviruses, including zetatorqueviruses, have apparently lost one or two β-strands of the four-stranded β-sheet of the P1 subdomain. The TTDelV ORF1 proteins display an organization of the P1 subdomain most similar to that of epsilontorqueviruses, with a four-stranded β-sheet in the P1 subdomain. However, both in phylogenetic analyses and in the structural clustering dendrogram, TTDelVs consistently cluster outside of the group IV anelloviruses with the large ORF1s (Fig. 2 and 7). Thus, it appears that TTDelVs did not evolve directly from epsilontorqueviruses or other viruses in this group but rather convergently augmented their P2 subdomains.

Due to the high sequence divergence, including extensive size variation, of ORF1 proteins, reconstruction of deeper evolutionary relationships between viruses from different genera of *Anelloviridae* proves challenging, with phylogenetic analyses systematically failing to resolve the deep branching points. Detailed structural

comparisons can aid higher rank classification of anelloviruses. For instance, the four structural groups of anelloviruses presented above potentially could be unified into separate subfamilies.

Conclusion

In this study, we characterized a novel group of Delphinidae-infecting anelloviruses. According to ORF1 pairwise nucleotide sequence comparisons, the amino acid sequence phylogeny, and comparisons of the ORF1 structural models, the 69 recovered Delphinid viral genomes were categorized into 22 species-level assemblages (TTDelV g1–22). In the phylogenetic trees, the TTDelV species groupings cluster together, forming a putative new genus, which we propose naming "*Qoptorquevirus*." The TTDelV species groupings correspond to the cetacean host species, suggesting virus-host coevolution. Furthermore, we identified diverse co-infections within the infraorder Cetacea, which aligns with anellovirus infection patterns found in other vertebrates.

Remarkably, the Delphinid anelloviruses encode ORF1 proteins that are much larger than those of other known anelloviruses, including primate anelloviruses with similar genome sizes. Analysis of the structural models of the Delphinid anelloviruses showed that they have the same characteristic ORF1 domains as other anelloviruses, namely, the central core, the extended N-terminal positively charged region, and the extended C-terminal region. The central core consists of the conserved JR domain and the variable P domain. The JR-domain plays a key role in capsid assembly and is therefore subject to strong purifying selection. In contrast, the P-domain likely interacts with the host immune system and/or cell surface receptors and likely evolves under diversifying selective pressure to evade the immune response. Consistent with this notion, we observed hypervariability in the P-domain of the Delphinid anelloviruses. Our observations suggest that Delphinid anelloviruses evolved the large, elaborate P-domains convergently with the primate anelloviruses. Finally, we argue that structural comparisons of the ORF1 proteins can guide higher-level classification of anelloviruses, which proved to be challenging based on the nucleotide or amino acid sequence comparisons due to their extreme divergence. To conclude, here we expand the identified host range of anelloviruses to Delphinidae family members and provide insight into their virus-host interactions, evolution, and classification.

ACKNOWLEDGMENTS

The molecular work was supported by funds from the Biodesign Institute and School of Life Sciences (Arizona State University) awarded to A.V. and Barrett Honors College (Arizona State University) awarded to K.S. The sample collection was supported with funds from the Faculty Development Grants program at the University of the South awarded to R.F. and internal departmental support for K.S. and K.R.H. E.V.K. is funded by the Intramural Research Program of the National Institutes of Health (National Library of Medicine).

AUTHOR AFFILIATIONS

[1]The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[2]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, Paris, France
[3]HTC Honors College, Coastal Carolina University, Conway, South Carolina, USA
[4]Department of Earth and Environmental Systems, The University of the South, Sewanee, Tennessee, USA
[5]Barrouallie Whaler's Project, Barrouallie, Saint Vincent and the Grenadines
[6]Independent Researcher, Barrouallie, Saint Vincent and the Grenadines
[7]Department of Chemical and Biological Sciences, The University of the West Indies at Cave Hill, Bridgetown, Saint Michael, Barbados

[8]National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, USA
[9]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory, Western Cape, South Africa
[10]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Rondebosch, Cape Town, South Africa

AUTHOR ORCIDs

Mart Krupovic http://orcid.org/0000-0001-5486-0098
Eugene V. Koonin http://orcid.org/0000-0003-3943-8299
Arvind Varsani http://orcid.org/0000-0003-4111-2415

AUTHOR CONTRIBUTIONS

Matthew D. De Koch, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Mart Krupovic, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Russell Fielding, Conceptualization, Formal analysis, Investigation, Methodology, Resources, Writing – review and editing | Kendal Smith, Formal analysis, Investigation, Methodology, Writing – review and editing | Kelsie Schiavone, Investigation, Methodology, Writing – review and editing | Katharine R. Hall, Investigation, Methodology, Writing – review and editing | Vincent S. Reid, Investigation, Methodology, Writing – review and editing | Diallo Boyea, Investigation, Methodology, Writing – review and editing | Emma L. Smith, Investigation, Methodology, Writing – review and editing | Kara Schmidlin, Formal analysis, Investigation, Methodology, Writing – review and editing | Rafaela S. Fontenele, Investigation, Methodology, Writing – review and editing | Eugene V. Koonin, Investigation, Methodology, Writing – review and editing | Darren P. Martin, Formal analysis, Investigation, Methodology, Visualization, Writing – review and editing | Simona Kraberger, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review and editing | Arvind Varsani, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Anellovirus genomes have been deposited in GenBank under accession numbers PP790224-PP790292. The raw reads have been deposited under BioProject PRJNA1189461, BioSample numbers SAMN44970975, SAMN44970976, SAMN4497077, and SAMN44970978, and SRA numbers SRR31500901, SRR31500900, SRR31500899, and SRR31500898.

ETHICS APPROVAL

Samples collected in this study are from deceased animals. The samples were exported under CITES Permit 16US774223/9 to the USA and imported into the USA under NMFS Permit 19091. Protocols for processing these samples in the lab are covered under the Arizona State University Institutional Biosafety Committee (SPROTO202300000004).

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Data Set S1 (JVI01370-24-s0001.xlsx). ORF1 pairwise identity matrix.
Data Set S2 (JVI01370-24-s0002.xlsx). Recombination breakpoints.
Figure S1 (JVI01370-24-s0003.pdf). Structural model of the TTDelV1 ORF1.
Figure S2 (JVI01370-24-s0004.pdf). Structural diversity of the P-domains from TTDelVs representing all 22 species.
Table S1 (JVI01370-24-s0005.pdf). pLDDT scores.

REFERENCES

1. Varsani A, Kraberger S, Opriessnig T, Maggi F, Celer V, Okamoto H, Biagini P. 2023. Anelloviridae taxonomy update 2023. Arch Virol 168:277. https://doi.org/10.1007/s00705-023-05903-6
2. Varsani A, Opriessnig T, Celer V, Maggi F, Okamoto H, Blomström A-L, Cadar D, Harrach B, Biagini P, Kraberger S. 2021. Taxonomic update for mammalian anelloviruses (family *Anelloviridae*). Arch Virol 166:2943–2953. https://doi.org/10.1007/s00705-021-05192-x
3. Butkovic A, Kraberger S, Smeele Z, Martin DP, Schmidlin K, Fontenele RS, Shero MR, Beltran RS, Kirkham AL, Aleamotu'a M, Burns JM, Koonin EV, Varsani A, Krupovic M. 2023. Evolution of anelloviruses from a circovirus-like ancestor through gradual augmentation of the jelly-roll capsid protein. Virus Evol 9:vead035. https://doi.org/10.1093/ve/vead035
4. Kaczorowska J, van der Hoek L. 2020. Human anelloviruses: diverse, omnipresent and commensal members of the virome. FEMS Microbiol Rev 44:305–313. https://doi.org/10.1093/femsre/fuaa007
5. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochem Biophys Res Commun 241:92–97. https://doi.org/10.1006/bbrc.1997.7765
6. Spandole S, Cimponeriu D, Berca LM, Mihăescu G. 2015. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. Arch Virol 160:893–908. https://doi.org/10.1007/s00705-015-2363-9
7. Laubscher F, Kaiser L, Cordey S. 2023. SCANellome: analysis of the genomic diversity of human and non-human primate anelloviruses from metagenomics data. Viruses 15:1575. https://doi.org/10.3390/v15071575
8. Paietta EN, Kraberger S, Lund MC, Vargas KL, Custer JM, Ehmke E, Yoder AD, Varsani A. 2024. Diverse Circular DNA viral communities in blood, oral, and fecal samples of captive lemurs. Viruses 16:1099. https://doi.org/10.3390/v16071099
9. Crane A, Goebel ME, Kraberger S, Stone AC, Varsani A. 2018. Novel anelloviruses identified in buccal swabs of Antarctic fur seals. Virus Genes 54:719–723. https://doi.org/10.1007/s11262-018-1585-9
10. Canova R, Budaszewski RF, Weber MN, da Silva MS, Puhl DE, Battisti LO, Soares JF, Wagner PG, Varela APM, Mayer FQ, Canal CW. 2021. Spleen and lung virome analysis of South American fur seals (*Arctocephalus australis*) collected on the southern Brazilian coast. Infect Genet Evol 92:104862. https://doi.org/10.1016/j.meegid.2021.104862
11. Fahsbender E, Burns JM, Kim S, Kraberger S, Frankfurter G, Eilers AA, Shero MR, Beltran R, Kirkham A, McCorkell R, Berngartt RK, Male MF, Ballard G, Ainley DG, Breitbart M, Varsani A. 2017. Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica. Virus Evol 3:vex017. https://doi.org/10.1093/ve/vex017
12. Rijsewijk FAM, Dos Santos HF, Teixeira TF, Cibulski SP, Varela APM, Dezen D, Franco AC, Roehe PM. 2011. Discovery of a genome of a distant relative of chicken anemia virus reveals a new member of the genus Gyrovirus. Arch Virol 156:1097–1100. https://doi.org/10.1007/s00705-011-0971-6
13. Kraberger S, Serieys LEK, Richet C, Fountain-Jones NM, Baele G, Bishop JM, Nehring M, Ivan JS, Newkirk ES, Squires JR, Lund MC, Riley SPD, Wilmers CC, van Helden PD, Van Doorslaer K, Culver M, VandeWoude S, Martin DP, Varsani A. 2021. Complex evolutionary history of felid anelloviruses. Virology 562:176–189. https://doi.org/10.1016/j.virol.2021.07.013
14. Cavalcante LTF, Cosentino MAC, D'arc M, Moreira FRR, Mouta R, Augusto AM, Troccoli F, Soares MA, Santos AF. 2023. Characterization of a new anellovirus species infecting an ocelot (*Leopardus pardalis*) in Brazil. Genet Mol Biol 46:e20230015. https://doi.org/10.1590/1678-4685-GMB-2023-0015
15. Fry TL, Owens LA, Ketz AC, Atwood TC, Dunay E, Goldberg TL, Hunt K. 2023. Serum virome of southern beaufort sea polar bears (*Ursus maritimus*) during a period of rapid climate change. Conserv Physiol 11:coad054. https://doi.org/10.1093/conphys/coad054
16. de Souza WM, Fumagalli MJ, de Araujo J, Sabino-Santos G Jr, Maia FGM, Romeiro MF, Modha S, Nardi MS, Queiroz LH, Durigon EL, Nunes MRT, Murcia PR, Figueiredo LTM. 2018. Discovery of novel anelloviruses in small mammals expands the host range and diversity of the Anelloviridae. Virology 514:9–17. https://doi.org/10.1016/j.virol.2017.11.001
17. Lund MC, Larsen BB, Rowsey DM, Otto HW, Gryseels S, Kraberger S, Custer JM, Steger L, Yule KM, Harris RE, Worobey M, Van Doorslaer K, Upham NS, Varsani A. 2023. Using archived and biocollection samples towards deciphering the DNA virus diversity associated with rodent species in the families cricetidae and heteromyidae. Virology 585:42–60. https://doi.org/10.1016/j.virol.2023.05.006
18. Maggi F, Pifferi M, Fornai C, Andreoli E, Tempestini E, Vatteroni M, Presciuttini S, Marchi S, Pietrobelli A, Boner A, Pistello M, Bendinelli M. 2003. TT virus in the nasal secretions of children with acute respiratory diseases: relations to viremia and disease severity. J Virol 77:2418–2425. https://doi.org/10.1128/jvi.77.4.2418-2425.2003
19. Deng X, Terunuma H, Handema R, Sakamoto M, Kitamura T, Ito M, Akahane Y. 2000. Higher prevalence and viral load of TT virus in saliva than in the corresponding serum: another possible transmission route and replication site of TT virus. J Med Virol 62:531–537.
20. Okamoto H, Takahashi M, Kato N, Fukuda M, Tawara A, Fukuda S, Tanaka T, Miyakawa Y, Mayumi M. 2000. Sequestration of TT virus of restricted genotypes in peripheral blood mononuclear cells. J Virol 74:10236–10239. https://doi.org/10.1128/jvi.74.21.10236-10239.2000
21. Matsubara H, Michitaka K, Horiike N, Yano M, Akbar SM, Torisu M, Onji M. 2000. Existence of TT virus DNA in extracellular body fluids from normal healthy Japanese subjects. Intervirology 43:16–19. https://doi.org/10.1159/000025018
22. Kaczorowska J, Deijs M, Klein M, Bakker M, Jebbink MF, Sparreboom M, Kinsella CM, Timmerman AL, van der Hoek L. 2022. Diversity and long-term dynamics of human blood anelloviruses. J Virol 96:e0010922. https://doi.org/10.1128/jvi.00109-22
23. Arze CA, Springer S, Dudas G, Patel S, Bhattacharyya A, Swaminathan H, Brugnara C, Delagrave S, Ong T, Kahvejian A, Echelard Y, Weinstein EG, Hajjar RJ, Andersen KG, Yozwiak NL. 2021. Global genome analysis reveals a vast and dynamic anellovirus landscape within the human virome. Cell Host Microbe 29:1305–1315. https://doi.org/10.1016/j.chom.2021.07.001
24. Biagini P, Charrel RN, de Micco P, de Lamballerie X. 2003. Association of TT virus primary infection with rhinitis in a newborn. Clin Infect Dis 36:128–129. https://doi.org/10.1086/345552
25. Pan S, Yu T, Wang Y, Lu R, Wang H, Xie Y, Feng X. 2018. Identification of a torque teno mini virus (TTMV) in hodgkin's lymphoma patients. Front Microbiol 9:1680. https://doi.org/10.3389/fmicb.2018.01680
26. Liang G, Conrad MA, Kelsen JR, Kessler LR, Breton J, Albenberg LG, Marakos S, Galgano A, Devas N, Erlichman J, Zhang H, Mattei L, Bittinger K, Baldassano RN, Bushman FD. 2020. Dynamics of the stool virome in very early-onset inflammatory bowel disease. J Crohns Colitis 14:1600–1610. https://doi.org/10.1093/ecco-jcc/jjaa094
27. Al-Qahtani AA, Alabsi ES, AbuOdeh R, Thalib L, El Zowalaty ME, Nasrallah GK. 2016. Prevalence of anelloviruses (TTV, TTMDV, and TTMV) in healthy blood donors and in patients infected with HBV or HCV in Qatar. Virol J 13:208. https://doi.org/10.1186/s12985-016-0664-6

28. Desai M, Pal R, Deshmukh R, Banker D. 2005. Replication of TT virus in hepatocyte and leucocyte cell lines. J Med Virol 77:136–143. https://doi.org/10.1002/jmv.20426
29. Koonin EV, Dolja VV, Krupovic M. 2021. The healthy human virome: from virus–host symbiosis to disease. Curr Opin Virol 47:86–94. https://doi.org/10.1016/j.coviro.2021.02.002
30. Ng TFF, Wheeler E, Greig D, Waltzek TB, Gulland F, Breitbart M. 2011. Metagenomic identification of a novel anellovirus in Pacific harbor seal *(Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years. J Gen Virol 92:1318–1323. https://doi.org/10.1099/vir.0.029678-0
31. Fielding R, Kiszka JJ. 2021. Artisanal and aboriginal subsistence whaling in saint vincent and the grenadines (Eastern caribbean): history, catch characteristics, and needs for research and management. Front Mar Sci 8:668597. https://doi.org/10.3389/fmars.2021.668597
32. Smith K, Fielding R, Schiavone K, Hall KR, Reid VS, Boyea D, Smith EL, Schmidlin K, Fontenele RS, Kraberger S, Varsani A. 2021. Circular DNA viruses identified in short-finned pilot whale and orca tissue samples. Virology 559:156–164. https://doi.org/10.1016/j.virol.2021.04.004
33. Fielding R. 2018. The wake of the whale: hunter societies in the Caribbean and North Atlantic. Harvard University Press.
34. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
35. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033
36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
37. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
38. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348
39. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010
40. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165. https://doi.org/10.1093/bioinformatics/btr088
41. Stöver BC, Müller KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7. https://doi.org/10.1186/1471-2105-11-7
42. RStudio RT. 2020. Integrated development environment for R. RStudio. PBC, Boston, MA, USA.
43. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer Science & Business Media.
44. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. JOSS 4:1686. https://doi.org/10.21105/joss.01686
45. Muhire BM, Varsani A, Martin DP. 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9:e108277. https://doi.org/10.1371/journal.pone.0108277
46. Gilchrist CLM, Chooi YH. 2021. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics 37:2473–2475. https://doi.org/10.1093/bioinformatics/btab007
47. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B. 2021. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol 7:veaa087. https://doi.org/10.1093/ve/veaa087
48. DATAtab_Team. 2024. Online statistics calculator. Available from: https://datatab.net
49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2
50. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. Nat Methods 19:679–682. https://doi.org/10.1038/s41592-022-01488-1
51. Lee JW, Won JH, Jeon S, Choo Y, Yeon Y, Oh JS, Kim M, Kim S, Joung I, Jang C, Lee SJ, Kim TH, Jin KH, Song G, Kim ES, Yoo J, Paek E, Noh YK, Joo K. 2023. DeepFold: enhancing protein structure prediction through optimized loss functions, improved template features, and re-optimized energy function. Bioinformatics 39:btad712. https://doi.org/10.1093/bioinformatics/btad712
52. Mariani V, Biasini M, Barbato A, Schwede T. 2013. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics 29:2722–2728. https://doi.org/10.1093/bioinformatics/btt473
53. Holm L. 2022. Dali server: structural unification of protein families. Nucleic Acids Res 50:W210–W215. https://doi.org/10.1093/nar/gkac387
54. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. 2006. Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics 7:339. https://doi.org/10.1186/1471-2105-7-339
55. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023. UCSF ChimeraX: tools for structure building and analysis. Protein Sci 32:e4792. https://doi.org/10.1002/pro.4792
56. Soares P, Abrantes D, Rito T, Thomson N, Radivojac P, Li B, Macaulay V, Samuels DC, Pereira L. 2013. Evaluating purifying selection in the mitochondrial DNA of various mammalian species. PLoS One 8:e58993. https://doi.org/10.1371/journal.pone.0058993
57. Spandole-Dinu S, Cimponeriu DG, Crăciun A-M, Radu I, Nica S, Toma M, Alexiu OA, Iorga CS, Berca L-M, Nica R. 2018. Prevalence of human anelloviruses in Romanian healthy subjects and patients with common pathologies. BMC Infect Dis 18:334. https://doi.org/10.1186/s12879-018-3248-9
58. Wang X, Chen X, Song X, Cao L, Yang S, Shen Q, Ji L, Lu X, Zhang W. 2023. Identification of novel anelloviruses in the blood of giant panda (*Ailuropoda melanoleuca*). Comp Immunol Microbiol Infect Dis 100:102038. https://doi.org/10.1016/j.cimid.2023.102038
59. Nishiyama S, Dutia BM, Stewart JP, Meredith AL, Shaw DJ, Simmonds P, Sharp CP. 2014. Identification of novel anelloviruses with broad diversity in UK rodents. J Gen Virol 95:1544–1553. https://doi.org/10.1099/vir.0.065219-0
60. Zhang W, Wang H, Wang Y, Liu Z, Li J, Guo L, Yang S, Shen Q, Zhao X, Cui L, Hua X. 2016. Identification and genomic characterization of a novel species of feline anellovirus. Virol J 13:146. https://doi.org/10.1186/s12985-016-0601-8
61. Lefeuvre P, Lett JM, Varsani A, Martin DP. 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J Virol 83:2697–2707. https://doi.org/10.1128/JVI.02152-08
62. Requião RD, Carneiro RL, Moreira MH, Ribeiro-Alves M, Rossetto S, Palhano FL, Domitrovic T. 2020. Viruses with different genome types adopt a similar strategy to pack nucleic acids based on positively charged protein domains. Sci Rep 10:5470. https://doi.org/10.1038/s41598-020-62328-w
63. Krupovic M. 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr Opin Virol 3:578–586. https://doi.org/10.1016/j.coviro.2013.06.010
64. Rossmann MG, Johnson JE. 1989. Icosahedral RNA virus structure. Annu Rev Biochem 58:533–573. https://doi.org/10.1146/annurev.bi.58.070189.002533
65. Liou S-H, Boggavarapu R, Cohen N, Zhang Y, Sharma I, Zyheb L, Acharekar NM, Rodgers H, Islam S, Zeheb L, Pitts J, Arze C, Swaminathan H, Yozwiak N, Ong T, Hajjar RJ, Chang Y, Swanson KA, Delagrave S. 2024. Structure of anellovirus-like particles reveal a mechanism for immune evasion. Nat Commun 15:7219. https://doi.org/10.1101/2022.07.01.498313:2022.07.01.498313
66. Sarker S, Terrón MC, Khandokar Y, Aragão D, Hardy JM, Radjainia M, Jiménez-Zaragoza M, de Pablo PJ, Coulibaly F, Luque D, Raidal SR, Forwood JK. 2016. Structural insights into the assembly and regulation of distinct viral capsid complexes. Nat Commun 7:13014. https://doi.org/10.1038/ncomms13014