Iranian Journal of Virology 2022;16 (2):1-14.

Original Article

Genome Characterization and Phylogeny Analysis of Invertebrates Parvoviruses

Fariborzi N[1], Alipour H[*2], Azizi K[1], Eskandarzade N[3], Ghorbani A*[4]

1. Department of Biology and Control of Diseases Vector, School of Health, Shiraz University of Medical Sciences: Shiraz, Iran.
2. Research Center for Health Sciences, Institute of Health, Shiraz University of Medical Sciences, Shiraz, Iran.
3. Department of Basic Sciences, School of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman, Iran.
4. Nuclear Agriculture Research School, Nuclear Science and Technology Research Institute (NSTRI), Karaj, Iran.

Abstract

Non-conserved sequences in the parvoviruses' genome are involved in the virus-host interactions, so analysis of the complete genome sequence of the parvoviruses broadens our perspective about parvoviruses' host tropism and evolution. This study aimed to analyze the phylogeny of the 25 parvoviruses that infect invertebrates from Densovirinae and Hamaparvovirinae subfamilies based on the complete genome and NS1 gene sequences. According to the whole-genome phylogenetic tree, hamaparvoviruses are close to each other; however, they are in the middle of densoviruses clades. Applying non-NS1 sequences for constructing the phylogenetic tree causes Hamaparvovirinae to disperse among Densovirinae members, which shows the possible ancestor relationship of these two subfamilies. The divergence difference between the NS1 nucleotide sequence and the whole genome was higher in Densovirinae compared with Hamaparvovirinae. In Hamaparvovirinae, there is approximately no difference in divergence of NS1 gene sequence compared with the whole genome, which indicates that non-NS1 sequences in densoviruses are more diverse than these sequences in Hamaparvovirinae members. The evolutionary divergence between Hamaparvovirinae and Densovirinae was slightly higher in the whole genome compared with NS1 sequences. As both Hamaparvovirinae and Densovirinae used in this study infect invertebrates, the low difference in divergence value between them could be related to their host. Each parvovirus in this study has its unique 5´ and 3´ untranslated regions, which could be used as a genetic fingerprint indicator for parvovirus detection.

Keywords: Hamaparvovirinae, Densovirinae, Parvoviridae, Bioinformatics, Phylogeny

Introduction

Parvoviridae, a large diverse family of animal viruses, was established in 1975 (1). The parvoviruses wait for the rapidly dividing host genome to enter the S-phase replication stage, then uncoat and multiply in the nucleus, which results in cell necrosis (1). Some members of this family cause mild diseases, whereas others, such as canine parvovirus (CPV), are highly pathogenic (2).

Until the year 2020, the ability of parvoviruses to infect a broad-spectrum host, from insects to mammals, served as the basis for the phylogeny-supported division of this family into Densovirinae and Parvovirinae subfamilies that respectively infect invertebrate and vertebrate hosts (1). Nevertheless, new chaparvoviruses with a close phylogenetic relationship with densoviruses were unexpectedly isolated from vertebrates which caused heterogeneity in Densovirinae.

As a result, in the recent classification, a new subfamily Hamaparvovirinae was introduced as its members infect both vertebrates and invertebrates (3-7).

[*] **Corresponding authors:**
Abozar Ghorbani, emails: anbghorbany@aeoi.org.ir and Hamzeh Alipour, email: alipoorh@sums.ac.ir

Parvoviruses have single-stranded DNA (~ 3.9-6.3 kbp), which has two gene cassettes (1). The first gene on the N-terminal side is NS whose expression produces proteins that either initiate replication (NS1) or help egress from the nucleus (NS2). The second gene, VP, which produces capsid proteins (VPs) is located on the C-terminal side. NS1 protein or replicase is a multi-domain protein that harbors domains called superfamily endonuclease domain (HUH), DNA-binding domain, and a helicase superfamily 3 (SF3) domain; latter is a highly conserved tripartite motif (167 aa) serves as the taxon demarcation criterion for the phylogeny inference of Parvoviridae (7,8). Because the NS1 cannot act as a polymerase, the host polymerase is necessary for parvovirus replication (8). The coding frames are flanked by the two untranslated regions of 3'UTR and 5'UTR (1).

Parvoviruses' untranslated regions (UTRs) form various secondary structures such as hairpin-like and/or G-quadruplexes to participate in different steps of the cell cycle: genome replication, encapsidation, integration, as well as interactions with host factors. The terminal regions of the parvovirus genome are such diverse that even among closely related viruses or within the same genus, the length of the regions, the percentage of GC, and the shape of the regions are different (9). The influencing factors on the diversity of UTRs are unknown, but it seems that host factors involve in this evolution. As a result, analyzing UTRs is crucial for understanding the interactions between viruses and their hosts. For example, the terminal repetitive regions in UTRs have been discovered to be the starting points for homologous or non-homologous recombination by inducing the DNA repair response in the host, which can affect host tropism/ switching (10).

The emergence of whole-genome sequencing techniques has led us to examine the viral genome regions other than conserved sequences in old and newly discovered parvoviruses. This study aimed to analyze the whole-genome phylogeny of the parvoviruses that infect invertebrates and compare it with NS1-based phylogeny. We then continued the analysis by briefly examining the parvovirus genome organization and UTRs.

Methods

Data Retrieval

Whole-genome sequences (WGS) of 25 parvoviruses isolated from different genera in two subfamilies (Densovirinae and Hamaparvovirinae) that infect invertebrates were downloaded from the publicly available NCBI GenBank database. The viruses were grouped into two subfamilies according to the recent information available on the internal committee on the taxonomy of viruses (ICTV) website (https://ictv.global/ taxonomy).

All ambiguous positions were removed from raw sequences, and the sequences aligned using the ClustalW (version 2.1) in Geneious Prime 2019 (Biomatters, New Zealand).

Genetic Diversity Analysis

The number of base substitutions per site over both whole-genome and NS1 genes between and within each subfamily was conducted using the Maximum Likelihood algorithm using the MEGA11 package (11).

Phylogenetic Analysis

Initially, a distance-based phylogenetic tree based on whole-genome sequences and NS1 genes of 25 parvoviruses (Supplementary Table 1) was constructed to infer the phylogenetic relationship between Densovirinae and Hamaparvovirinae subfamilies members. As some densoviruses have ambisense genomes, all ORFs were set to the forward direction in the whole-genome analysis. After sequences were aligned via the ClustalW approach implemented in Geneious Prime 2019 (Biomatters, New Zealand), the data were exported into MEGA11 to construct the optimal trees using Maximum Likelihood and Neighbor-Joining methods with bootstrap values of 1000 replicates and a 50% threshold score.

The evolutionary distances were computed by the Maximum Likelihood model. We used Adeno-associated virus – 2 (NC_001401) as an outgroup for all phylogenetic analyses.

Divergence clock rates of whole genome sequences were estimated using the Timetree tool with the UPGMA method in the MEGA11

(11). Transition/transversion (r) and dN/dS were calculated using MEGA11 with default parameters. SplitsTree V5 software was used for probable DNA recombination in all sequences.

Genome organization and UTRs

Visualization of genome organization was performed using Open Reading Frame (ORF) Finder (available at https://www.ncbi.nlm.nih.gov/orffinder/). The most thermodynamically stable secondary structures of 5´UTRs and 3´UTRs were predicted using the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The tandem repeats were determined in 5´UTRs and 3´UTRs using Tandem Repeat Finder tools (https://tandem.bu.edu/trf/trf.html).

Results and Discussion

The conserved areas in NS1 are the global criteria of phylogenetic analysis of the Parvoviridae. However, regions other than NS1 are susceptible to high mutation and recombination and it is believed that their genetic changes are responsible for host switching in parvoviruses. In this study, we attempted to construct a phylogenic tree based on both the whole-genome and NS1 sequences of parvoviruses that infect invertebrates and discuss the divergence of parvoviruses. We then examine genome organization and non-coding regions of these viruses.

Genetic Diversity of Parvoviruses

The family Parvoviridae comprises large, diverse, single-stranded DNA viruses that share conserved SF3 helicase domain (167 aa) in NS1 protein at the family level. When two parvoviruses share more than 30% similarity in the amino acid sequence of NS1, they are assigned to the same genus, and when they share more than 85% identity, they are considered the same species.

Discovering new densoviruses caused genus-level paraphyly in Densovirinae, which forced scientists to divide Densovirinae into two subfamilies: Densovirinae and Hamaparvovirinae. The members of this new subfamily Hamaparvovirinae show less than 20% of NS1 sequence similarity with other parvoviruses; however, they share 30% of NS1 sequence identity (7).

Table 1 shows the average evolutionary divergence over sequence pairs within each subfamily member that infects invertebrates.

Densovirinae showed less intragroup divergence compared with Hamaparvovirinae. The divergence values based on the both NS1 nucleotide sequence and the whole genome are higher than the data obtained based on the NS1 amino acid sequence by Morais et al. (12).

They found that the average genetic distances in Hamaparvovirinae, Densovirinae and Parvovirinae members were close to each other: 0.49, 0.50 and 0.50, respectively (12).

According to Table 1, the intra-group divergence difference between NS1 and the whole genome was higher in Densovirinae (0.82 vs. 0.92) compared with Hamaparvovirinae (1.08 vs. 1.07).

In Hamaparvovirinae, there is approximately no divergence difference in NS1 compared with the whole genome. This phenomenon demonstrates that non-NS1 sequences in densoviruses are more diverse than these sequences in Hamaparvovirinae members (Table 1) .

In Table 2, the evolutionary divergence is compared between two subfamilies. The evolutionary divergence between Hamaparvovirinae and Densovirinae was slightly higher according to the whole genome compared with NS1 sequences.

As both Hamaparvovirinae and Densovirinae used in this study infect invertebrates, the lowest difference in whole-genome divergence value between them could be related to their host; however, this hypothesis cannot be approved by the findings of this study.

The software couldn't find any recombination and positive selection (dn/ds>1) that were statistically significant between our sequences.

Whole-Genome Phylogeny Analysis

Except for the NS1 region, which is the prime criterion for phylogeny analysis, other parts of the parvovirus's genome such as UTRs are diverse even among closely related viruses. The association of increased host range with genetic dynamic in regions other than NS1 confirms that whole-genome screening is a

valuable strategy to evaluate parvovirus-host interactions.

Table 1. Estimates of an average intra-group evolutionary divergence over sequence pairs within subfamilies members. The number of base substitutions per site from averaging over all sequence pairs within each group is shown. Analyses were conducted using the Maximum Composite Likelihood model (13).
This analysis involved 25 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 1245 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 (11).

Subfamily	Average evolutionary divergence in the whole genome	Average evolutionary divergence in the NS1 gene
Densovirinae	0.92	0.82
Hamaparvovirinae	1.07	1.08

Table 2. Estimates of evolutionary divergence over sequence pairs between Hamaparvovirinae and Densovirinae subfamilies.
The number of base substitutions per site from averaging over all sequence pairs between groups is shown. Analyses were conducted using the Maximum Composite Likelihood model (13). This analysis involved 25 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 1245 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 (11).

	Densovirinae	
	Whole genome	NS1 gene
Hamaparvo-virinae	1.250	1.16

Figure 1 shows the phylogeny analysis of the complete genome sequences of 25 parvoviruses from Hamaparvovirinae and Densovirinae subfamilies that infect invertebrates. All members of Densovirinae are not close to each other because applying non-NS1 sequences of the genome in the whole-genome phylogeny causes members of Hamaparvovirinae to disperse among Densovirinae members (Fig 1).
The proximity of hamparvoviruses to some densoviruses in the whole-genome tree strengthens the possibility that densoviruses are the ancestors of hamaparvoviruses or vice versa. Analysis of insect-specific viruses has shown that they could overcome the host range limitation and expand their host diversity (14,15). For example, chaphamaparvoviruses in Hamaparvovirinae, which typically infected invertebrates, were isolated from vertebrate hosts such as duck and large crayfish (12,16); however, genus Brevihamaparvovirus, have only been isolated from insects (insect-specific viruses) .
Figure 2 shows the phylogeny trees of the NS1 gene sequence of viruses that infect invertebrates. Although members of the Hamaparvovirinae are close to each other in a separate cluster from the densoviruses, Decapod hepanhamaparvovirus 1 is more distant from its subfamily members (Fig 2).
This result indicates that Decapod hepanhamaparvovirus 1 is closer to its subfamily members when constructing a phylogeny tree of the whole genome compared with the NS1-based tree.
In Figure 1, three densoviruses with marine hosts: Ostreid aquambidensovirus 1, Asteroid aquambidensovirus 1 and Decapod aquambidensovirus 1 are close to the outgroup and show paraphyly in the tree.
Ostreid aquambidensovirus 1 and Decapod aquambidensovirus 1 are close to each other; however, Asteroid aquambidensovirus 1 share a common ancestor with other denso and hamaparvoviruses. In the NS1 gene evolutionary clock, these three viruses are the last descendants and show no paraphyly (Fig 3). Whole genome sequences of viruses were the target for determining relative divergence time analysis.
The result showed Hemipteran scindoambidensovirus 1 with 0.64 relative times is the oldest virus. Hamaparovirinae and Densovirinae were divergent from 0.59 relative time from together. This analysis also confirms Hamaparovirinaes were divergent from Densovirinaes and a Densovirinae is close to its ancestor (Fig 3).
According to the study conducted by Morris et al., both VP and NS1 had the same phylogenetic value for drawing a phylogenetic tree; however, the genetic diversity among Parvoviruses was greater for structural than non-structural proteins (12,17).

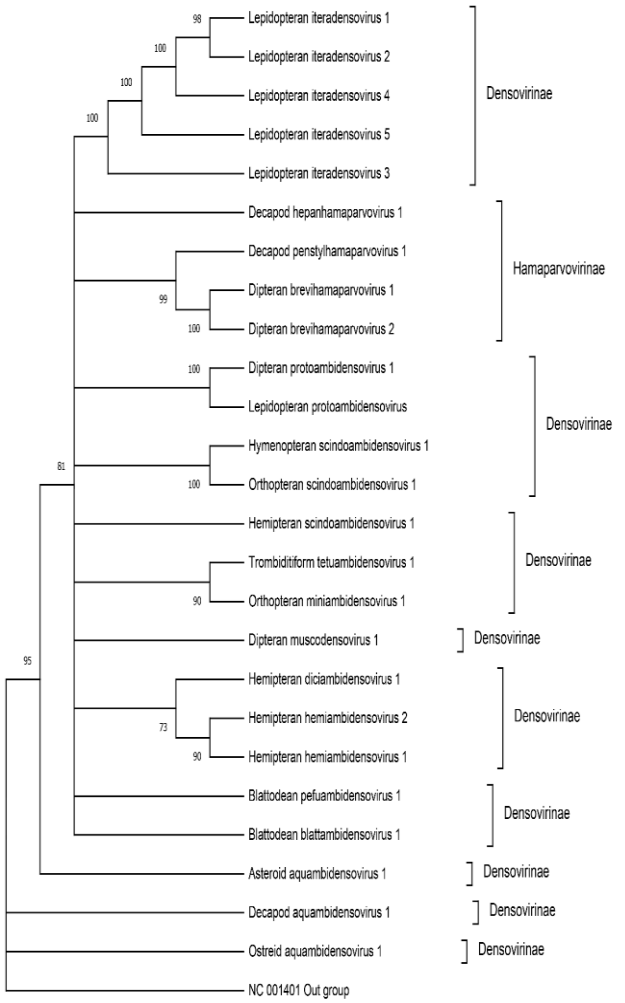


Fig 1. Phylogenetic analysis of full-length genomes of Densovirinae and Hamaparvovirinae subfamilies. This analysis involved 26 nucleotide sequences. All ambiguous positions were removed from each sequence pair. The evolutionary distances (units of the number of base substitutions per site) were computed using the Maximum Likelihood method. The evolutionary history was inferred using the Neighbor-Joining method in MEGA11 (11).

In the same research, the entropy of the NS1 sequence was very low and the same in three subfamilies, which indicated that NS1 conserved sequences undergo purifying selection and could be properly used for phylogenetic analyses (12). However, the VP sequence had a high entropy and positive natural selection with no sign of recombination which suggested that the events such as the recombination or mutation of the VP gene were related to determining the new host (12). For example, canine parvovirus is believed to be descended from a single ancestor as a result of a shift in host selection due to mutations in the feline pan-leukopenia virus (18,19).

In canine parvoviruses, most of the non-synonymous substitutions occurred in the main antigenic protein, capsid protein 2 (VP2), which determined viral tissue tropism and host range (20). The surface of the viruses is uneven due to the protrusion of the VP loops.

In previous research, it has been determined that similar parvoviruses were more related in terms of folding and topology of VP protein than nucleotide sequence or amino acid sequence, which could be applied in phylogenetic analysis (21).

Open Reading Frame (ORF) Organization

The genomic organization of the densoviruses and hamaparvoviruses that infect invertebrates is depicted in Figure 4. Hamaparvoviruses' genomes are typically smaller than densoviruses' genomes, ~4 kbp vs. ~6 kbp respectively.

In densoviruses, the genome organization does not have a specific pattern and some have an ambisense genome; however, hamaparvoviruses have a unisense genome. Most densoviruses have multiple VPs open reading frames (Fig 4).

The number of ORFs in densoviruses is larger than in hamaparvoviruses; for example, densoviruses have NS3 in their genome, which is not present in hamaparvoviruses (Fig 4).

There are different architectures in the ORF organization of Densovirinae related to different genera; however, densoviruses with the same ORF architect are close to each other in the whole-genome tree (Fig 1).

ORF organization in Hamaprvovirinae is almost the same (Fig 4). The largest genomes among the hamaparvoviruses are related to Decapod hepanhamaparvovirus 1 in Hepanhamaparvovirus genus with 6085 nt long which is very close to the average length of the genome in densoviruses and it is placed close to densoviruses in the whole-genome tree (Fig 1).

Regardless of the difference in genomic transcription and expression mechanism, the members of the *Parvoviridae* have almost similar genome organization, so the NS1 gene is on the left side and the VP gene is on the

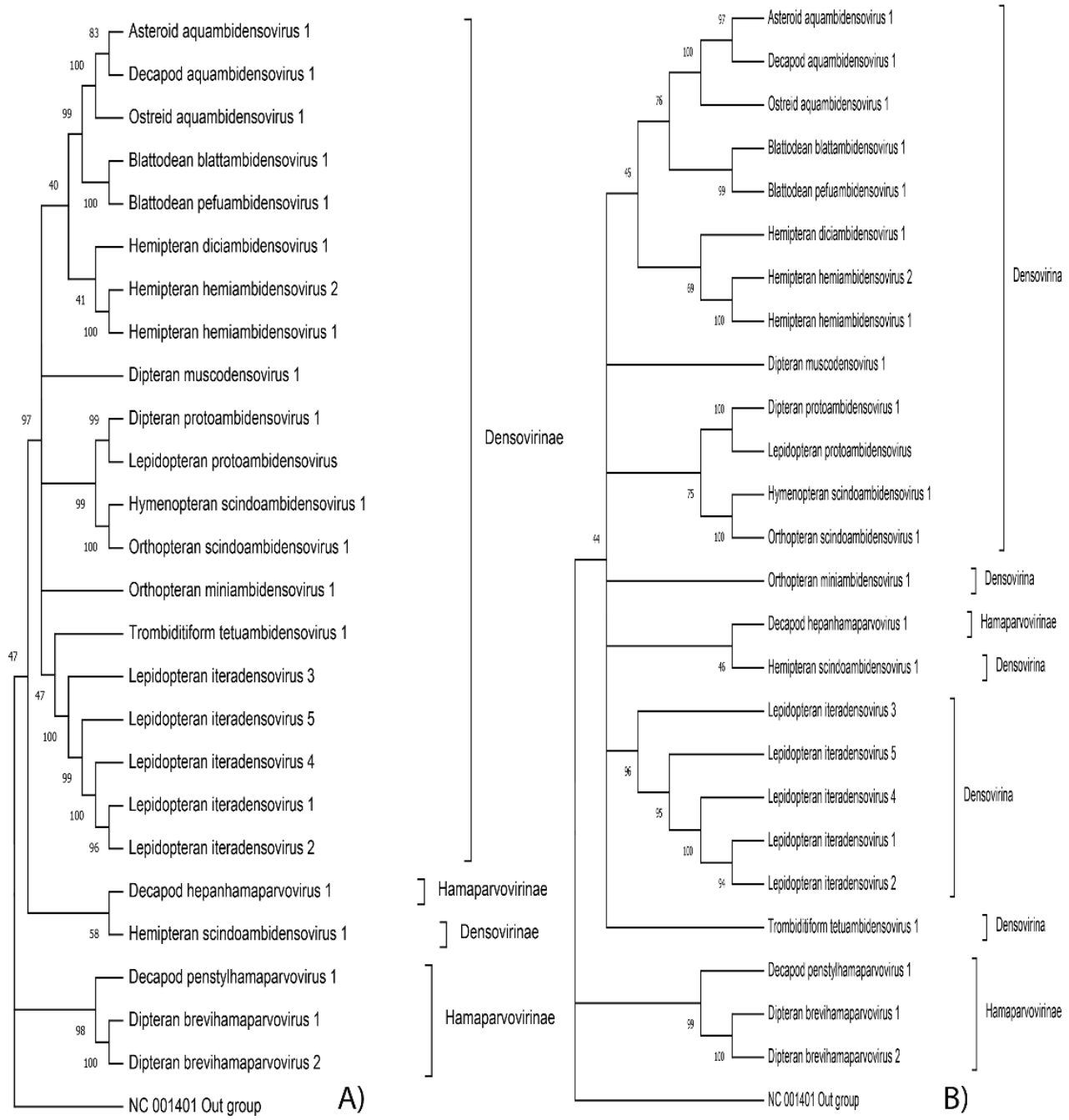


Fig 2. Phylogenetic analysis of NS1 gene of viruses from Densovirinae and Hamaparvovirinae subfamilies. This analysis involved 26 nucleotide sequences. All ambiguous positions were removed from each sequence pair. The evolutionary distances (units of the number of base substitutions per site) were computed using the Maximum Likelihood method. The evolutionary history was inferred using the Neighbor-Joining method (A) Maximum Likelihood (B) in MEGA 11 (11). This analysis involved 26 nucleotide sequences.

right side. If sequence similarity is not established for a genus (30-40% similarity in NS1 amino acid sequences), it is possible to assign a parvovirus to a specific genus based on its genomic organization. NS gene encodes up to 1, 5, and 3 proteins in Parvovirinae, Densovirinae, and Hamaprvovirinae, respectively (21). The capsid protein of members of this family is assembled by viral proteins encoded by the cap gene.

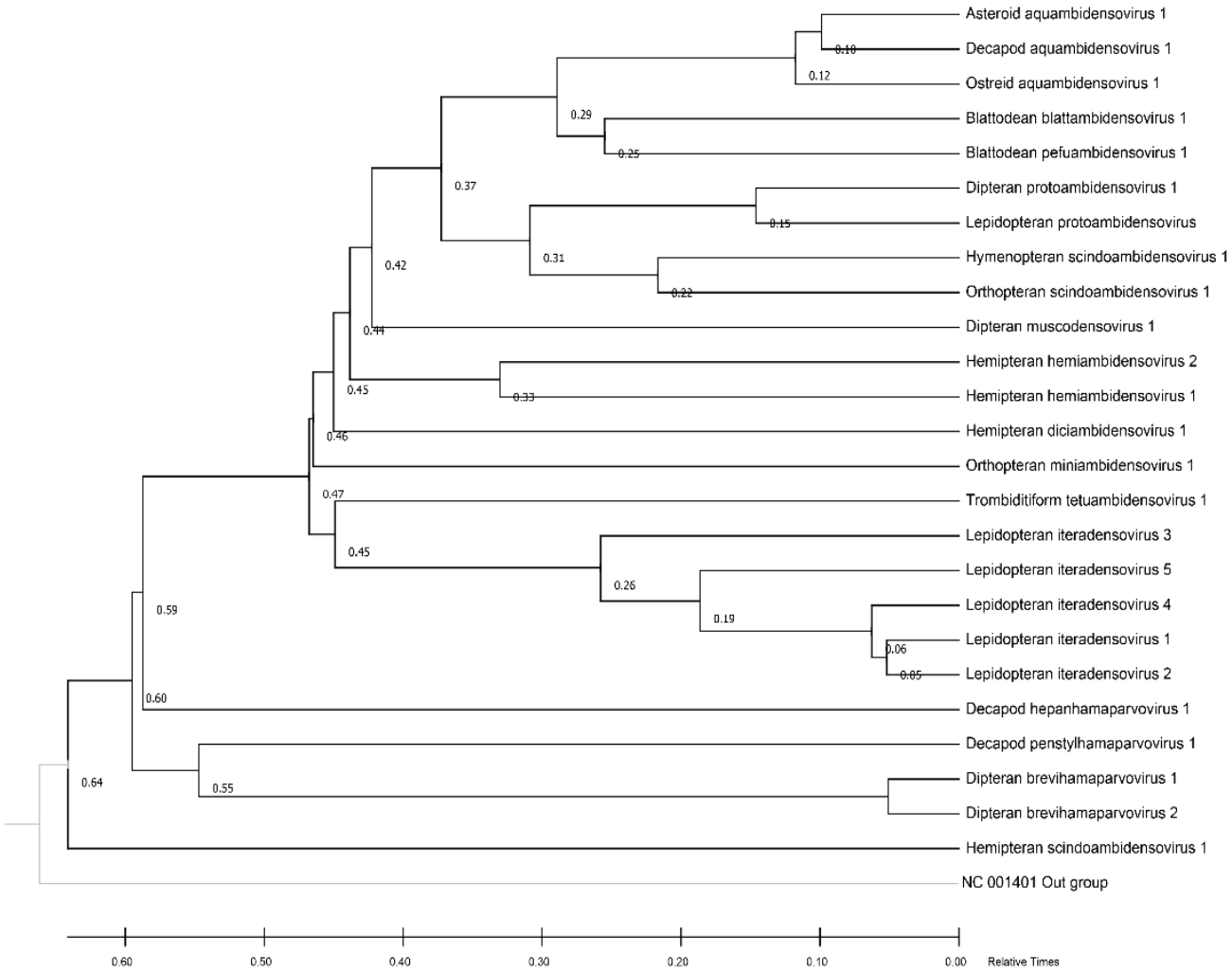


Fig 3. Timtree and divergence clock of *Densovirinae and Hamaparvovirinae* subfamilies. Relative divergence time was calculated using the UPGMA method.

Cap gene in *Parvovirinae*, *Densovirinae*, and *Hamaprvovirinae* encodes a maximum of 3, 4, and 1 VP, respectively (21). NP, a small regulatory protein, is commonly found in *Parvovirinae* and *Hamaprvovirinae* (21).

The reason for the difference in the number of transcribed proteins in the members of different subfamilies is unknown.

In densoviruses, leaky scanning is responsible for protein expression; however, alternative splicing has a regulatory role in hamaparvoviruses (21–23). Densoviruses also use overlapping frames to create a series of genus-specific supporting proteins (21). VPs with different sizes are produced in different parvoviruses through truncation from the N-terminal region. VP 2-5 is smaller and VP1 is the largest and the expression rate of smaller VPs is higher (23).

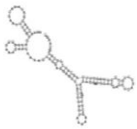
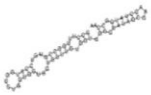
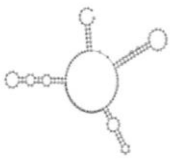
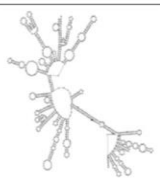
Untranslated Regions

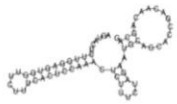
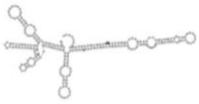
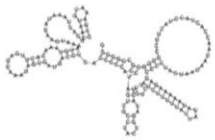
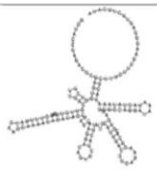
Parvoviruses' UTRs can adapt non-B DNA structures for recognition and binding of host proteins to affect communication with the host (10,24,25).

The length of telomeres varies even among members of the same species, and there is still no correlation between telomere complexity and the host (9).

By analyzing the whole-genome sequence of parvoviruses, we might be able to better understand how genetic dynamics or host characteristics have led parvoviruses to become dual hosts (infecting both vertebrates and invertebrates) as well as classify newly discovered parvoviruses confidently.

Fig 4. Genome organization of *Densovirinae* and *Hamaparvovirinae* parvoviruses.

5′UTR	5′UTR size (bp) and %GC	Virus	3′UTR size (bp) and %GC	3′UTR
	36, 47.2	Asteroid aquambidensovirus 1 *Densovirinae; Aquambidensovirus*	505, 35.2	
	384, 43	Decapod aquambidensovirus 1 *Densovirinae; Aquambidensovirus*	346, 43.4	
	127, 45.7	Ostreid aquambidensovirus 1 *Densovirinae; Aquambidensovirus*	57, 38.6	
	281, 56.2	Blattodean blattambidensovirus 1 *Densovirinae; Blattambidensovirus*	242, 58.7	
	201, 59.7	Blattodean pefuambidensovirus 1 *Densovirinae; Pefuambidensovirus*	65, 58.5	
	80, 33.8	Dipteran muscodensovirus 1 *Densovirinae; Muscodensovirus*	181, 21.5	
	436, 45.9	Dipteran protoambidensovirus 1 *Densovirinae; Protoambidensovirus*	802, 37.9	
	656, 45.6	Lepidopteran protoambidensovirus *Densovirinae; Protoambidensovirus*	575, 46.4	

	63, 44.4	Hemipteran diciambidensovirus 1 *Densovirinae; Diciambidensovirus*	305, 45.9	
	173, 40.5	Hemipteran hemiambidensovirus 1 *Densovirinae; Hemiambidensovirus*	111, 43.2	
	237, 41.4	Hemipteran hemiambidensovirus 2 *Densovirinae; Hemiambidensovirus*	162, 38.9	
	158, 32.3	Hemipteran scindoambidensovirus 1 *Densovirinae; Scindoambidensovirus*	158, 35.4	
	368, 47.6	Hymenopteran scindoambidensovirus 1 *Densovirinae; Scindoambidensovirus*	290, 53.4	
	855, 37	Orthopteran scindoambidensovirus 1 *Densovirinae; Scindoambidensovirus*	195, 42.1	
	343, 40.8	Lepidopteran iteradensovirus 1 *Densovirinae; Iteradensovirus*	346, 38.2	
	353, 40.8	Lepidopteran iteradensovirus 2 *Densovirinae; Iteradensovirus*	292, 41.8	
	369, 33.1	Lepidopteran iteradensovirus 3 *Densovirinae; Iteradensovirus*	426, 34	

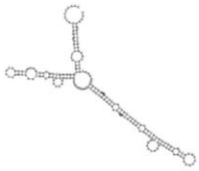
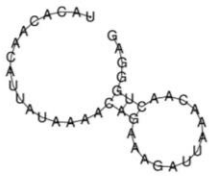
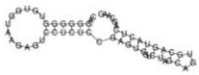
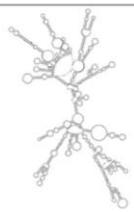
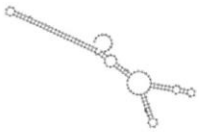
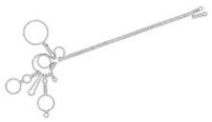
	348, 39.7	Lepidopteran iteradensovirus 4 *Densovirinae; Iteradensovirus*	346, 37.9	
	327, 43.4	Lepidopteran iteradensovirus 5 *Densovirinae; Iteradensovirus*	288, 39.2	
	282, 56.4	Orthopteran miniambidensovirus 1 *Densovirinae; Miniambidensovirus*	233, 60.5	
	208, 39.9	Trombiditiform tetuambidensovirus_1 *Densovirinae; Tetuambidensovirus*	40, 30	
	57, 57.9	Decapod hepanhamaparvovirus 1 *Hamaparvovirinae; Hepanhamaparvovirus*	1285, 40.1	
	157, 46.5	Decapod penstylhamaparvovirus 1 *Hamaparvovirinae; Penstylhamaparvovirus*	343, 42	
	383, 35.8	Dipteran brevihamaparvovirus 1 *Hamaparvovirinae; Brevihamaparvovirus*	421, 38.5	
	289, 34.3	Dipteran brevihamaparvovirus 2 *Hamaparvovirinae; Brevihamaparvovirus*	510, 40.2	

Table 3. Secondary structures of 5´UTR and 3´UTR of *Densovirinae* and *Hamaparvovirinae* subfamilies. Secondary structures of 5´UTR and 3´UTR were analyzed using the RNAfold tool (http://rna. tbi.univie.ac.at/cgi-bin/ RNAWebSuite/RNAfold.cgi). GC content was calculated using the Geneious Prime 2019 (Biomatters, New Zealand).

Table 4. Tandem repeats of 5´UTR and 3´UTR of *Densovirinae* and *Hamaparvovirinae* members. The tandem repeats were determined using Tandem Repeat Finder tools (https://tandem.bu.edu/trf/trf.html).

Virus	5′UTR		3′UTR	
	Indices	Consensus Size	Indices	Consensus Size
Dipteran brevihamaparvovirus 1	134-350	107	120-213	38
Dipteran brevihamaparvovirus 2	154-238	40	86-112 154-337	14 92
Lepidopteran protoambidensovirus	177-223	18	350-396	18
Lepidopteran iteradensovirus 1	1-230	230	117-347	230
Lepidopteran iteradensovirus 3	205-251	21	-	-
Lepidopteran iteradensovirus 4	1-271 1-175	271 175	75-346 195-346	271 175
Orthopteran miniambidensovirus 1	1-199	199	34-233	199
Hymenopteran scindoambidensovirus 1	104-238	134	13-143	134
Orthopteran scindoambidensovirus 1	1-144	144	51-195	144

The general shape and second structures of the 25 UTRs are visualized in Table 3. Among *Hamaprvovirinae*, Decapod hepanhamaparvovirus 1 with 1285 bp has the longest UTR. Among the densoviruses, the largest multi-loop terminal region is related to the 5´UTR of Orthopteran scindoambidensovirus 1 belonged to *Scindoambidensovirus* with 855 nt with a low GC level (37%) (Table 3). The minimum and maximum GC content are observed for the 3´UTR of Dipteran muscodensovirus 1 with 21.5% and 3´UTR Orthopteran miniambidensovirus 1 with 60.5%, respectively (Table 3). Although parvoviruses' UTRs can adapt a wide range of secondary motifs, only 9 sequences have telomeres, which are listed in Table 4.

Viruses with telomeres are from both *Hamaparvovirinae* and *Densovirinae* subfamilies. As shown in Table 4, most parvoviruses that infect lepidopterans have tandem repeats. Lepidopteran iteradensovirus 3 has a telomeric sequence only on the 5′UTR; however, Lepidopteran iteradensovirus 4 has the longest telomeric sequence with 271 nt long on both 3′ and 5′UTRs.

As previously approved, parvoviruses' TR length and shape vary within a single genus (9). For example, *Brevihamaparvovirus* members have the same genome organization but very different UTRs in terms of size and secondary structures (Fig 4 and Table 3).

In the linear DNA genome of parvoviruses, the coding region is located between the terminal regions of 3'UTR and 5'UTR that have 100-550 nucleotides of terminal repeats (TRs) in most parvoviruses (9). Due to the presence of palindromic repeats, terminal regions can fold into a wide variety of dynamic secondary structures in hairpin-like shapes or arbitrary T-, I-, J-, Y-, and U shapes to play a role in rolling hairpin replication as hinges to change the direction of replication (7,26). Moreover, TRs harbor cis-acting information elements such as the TATA box, which trigger interaction with host proteins such as transcription factors, or DNA damage response elements. These Interactions lead to the circularization and concatemerization of the viral genomes either by non-homologous end-joining (NHEJ) or homologous recombination (HR) (10).

The high percentage of GC, in the end, areas can adapt non-B DNA structures such as G-quadruplexes and triplex structures (9). G-quadruplexes are involved in genome stability while triplexes can form at homopurine: homopyrimidine sequences with mirror symmetry are involved in the recognition and binding of host proteins such as P53 to the virus genome (24). TRs in some genera are homotelomeric meaning that both termini are similar but inverted, whereas, in other genera, the linear genome differs at the two ends and therefore they are called heterotelomeric (7).

It seems that the type of TR sequences affects strand polarity in the virion in such a way that in viruses with heterotelomeres, encapsidation is usually completed with usually negative polarity. Negative strand polarity during packaging may be due to insufficient signal in one telomere. The determining factors of the evolutionary pressure on the selection of the type of TRs are not known and the great diversity of these structures shows the importance of these regions which are very little studied (9). Even viruses within the same genus have various telomeres, so Tijssen et al. proposed that the marked differences in genome end size and secondary structure between genera may indicate a reliance on particular cellular

components required for replication and encapsidation (27). As a result, TRs may have evolved in the direction of their need for interactions with their replicas, helper virus cofactors, and/or cell host proteins (27). For example, poisoning caused by the adeno-associated virus (AAV) vector in Parvovirinae, which is used for gene therapy, was probably related to the response of the host cell to the terminal areas of the virus (28,29).

Telomeres are not conserved sequences; therefore, phylogenetic evaluations cannot be assessed based on them, but their classification or examination can affect our understanding of parvoviruses host tropism/switching.

Telomeres have never been considered as a classification criterion till in a recent study 40 parvoviruses TR secondary structures were analyzed and used to divide parvoviruses into four groups (H1-H4) according to the number of hairpin loops (9). They concluded that the percentage of sequence homology within two viruses was not relevant neither to the number of hairpin loops nor the conformation of TR sequences even within a genus (9).

Conclusion

Whole-genome sequences of parvoviruses harbor non-conserved sequences whose genetic changes are relevant to increased host range; therefore, whole-genome screening can be informative for studying parvoviruses' host-virus interactions. Increasing our knowledge in this field will help us to find out the evolutionary history of Parvoviridae members regarding host switching.

Our result showed Hamaparovirinaes were divergent from Densovirinaes and a Densovirinae is close to its ancestor.

We showed 5´ and 3´UTRs diversity in all members of Densovirinae and Hamaparvovirinae as possible candidates for genetic fingerprints in parvovirus detection. We can also benefit from parvoviruses 5´ and 3´UTRs for plasmid vector designation because of their role in replication. Moreover, we provide the whole-genome phylogenetic analysis of parvoviruses found in invertebrates for the first time which helps other scientists for future studies.

Acknowledgment

None.

Conflict of Interest

No conflict of interest is declared.

Funding

The authors appreciate the support from the Vice-Chancellery for Research and Technology Affairs of Shiraz University of Medical Sciences with grant No. 24325.

Ethics Approval and Consent to Participate

The research reported here did not involve experimentation with human participants or animals.

Author Contributions

AG and NF designed the study and analyzed the data; NE and NF drafted the manuscript. AG, HA and KA edited and approved the final version of the manuscript.

References

1. Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger A-M, Hughes J, et al. ICTV virus taxonomy profile: Parvoviridae. J Gen Virol. 2019; 100(3):367.

2. Salarpour R, Zamani Ahmadmahmudi M, Rezaei M, Eskandarzade N. Paraoxonase activity assessment in dogs suffering from Parvovirus infection. J Hellenic Vet Med Soc. 2020;71(4):2499–2504.

3. Pénzes JJ, de Souza WM, Agbandje-McKenna M, Gifford RJ. An ancient lineage of highly divergent parvoviruses infects both vertebrate and invertebrate hosts. Viruses. 2019;11(6):525.

4. de Souza WM, Romeiro MF, Fumagalli MJ, Modha S, de Araujo J, Queiroz LH, et al. Chapparvoviruses occur in at least three vertebrate classes and have a broad biogeographic distribution. J Gen Virol. 2017; 98(2):225.

5. Williams SH, Che X, Garcia JA, Klena JD, Lee B, Muller D, et al. Viral diversity of house mice in New York City. mBio. 2018;9(2):e01354-17.

6. Cotmore SF, Agbandje-McKenna M, Chiorini JA, Mukha D V, Pintel DJ, Qiu J, et al. The family parvoviridae. Arch Virol. 2014;159(5):1239–47.
7. Pénzes JJ, Söderlund-Venermo M, Canuti M, Eis-Hübinger AM, Hughes J, Cotmore SF, et al. Reorganizing the family Parvoviridae: a revised taxonomy independent of the canonical approach based on host association. Arch Virol. 2020;165(9):2133–46.
8. Cotmore SF, Tattersall P. A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses. Parvoviruses London: Hoddler Arond. 2005; 171–81.
9. Laugel M, Lecomte E, Ayuso E, Adjali O, Mével M, Penaud-Budloo M. The Diversity of Parvovirus Telomeres. In: Veterinary Medicine and Science. Intech-Open; 2022.
10. Cotmore SF, Tattersall P. Parvovirus diversity and DNA damage responses. Cold Spring Harb Perspect Biol. 2013;5(2):a012989.
11. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.
12. Morais P, Trovão NS, Abecasis AB, Parreira R. Insect-specific viruses in the Parvoviridae family: Genetic lineage characterization and spatiotemporal dynamics of the recently established Brevihamapar-vovirus genus. Virus Res. 2022;313:198728.
13. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512–26.
14. Junglen S, Korries M, Grasse W, Wieseler J, Kopp A, Hermanns K, et al. Host range restriction of insect-specific flaviviruses occurs at several levels of the viral life cycle. mSphere. 2017;2(1):e00375-16.
15. Morais P, Trovão NS, Abecasis AB, Parreira R. Genetic lineage characterization and spatiotemporal dynamics of classical insect-specific flaviviruses: outcomes and limitations. Virus Res. 2021;303:198507.
16. Canuti M, Verhoeven JTP, Munro HJ, Roul S, Ojkic D, Robertson GJ, et al. Investigating the diversity and host range of novel parvoviruses from North American ducks using epidemiology, phylogenetics, genome structure, and codon usage analysis. Viruses. 2021;13(2): 193.
17. Lu G, Wu L, Ou J, Li S. Equine parvovirus-hepatitis in China: Characterization of its genetic diversity and evidence for natural recombination events between the Chinese and American strains. Front Vet Sci. 2020; 7:121.
18. Allison AB, Kohler DJ, Ortega A, Hoover EA, Grove DM, Holmes EC, et al. Host-specific parvovirus evolution in nature is recapitulated by in vitro adaptation to different carnivore species. PLoS Pathog. 2014; 10(11):e1004475.
19. Horiuchi M, Yamaguchi Y, Gojobori T, Mochizuki M, Nagasawa H, Toyoda Y, et al. Differences in the evolutionary pattern of feline panleukopenia virus and canine parvovirus. Virology. 1998;249(2):440–52.
20. Hueffer K, Parker JSL, Weichert WS, Geisel RE, Sgro J-Y, Parrish CR. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J Virol. 2003;77(3):1718–26.
21. Mietzsch M, Pénzes JJ, Agbandje-McKenna M. Twenty-five years of structural parvovirology. Viruses. 2019;11(4):362.
22. Bergoin M, Tijssen P. Densoviruses: a highly diverse group of arthropod parvoviruses. Insect Virol. 2010;59:82.
23. Qiu J, Yoto Y, Tullis GE, Pintel D. Parvovirus RNA processing strategies. Hodder Arnold, London, UK; 2006.
24. Brázdová M, Tichý V, Helma R, Bažantová P, Polášková A, Krejčí A, et al. p53 specifically binds triplex DNA in vitro and in cells. PLoS One. 2016; 11(12):e0167439.
25. Faisst S, Perros M, Deleu L, Spruyt N, Rommelaere J. Mapping of upstream regulatory elements in the P4 promoter of parvovirus minute virus of mice. Virology. 1994;202(1):466–70.
26. Deng Z, Wang Z, Lieberman PM. Telomeres and viruses: common themes of genome maintenance. Front Oncol. 2012;2:201.
27. Tijssen P, Pénzes JJ, Yu Q, Pham HT, Bergoin M. Diversity of small, single-stranded DNA viruses of invertebrates and their chaotic evolutionary past. J Invertebr Pathol. 2016;140:83–96.
28. Wilson JM, Flotte TR. Moving Forward after Two Deaths in a Gene Therapy Trial of Myotubular Myopathy: To learn from these tragedies, the scientific community must commit to full transparency and cooperation. Genet Eng Biotechnol News. 2020;40(8):14–6.
29. Muhuri M, Maeda Y, Ma H, Ram S, Fitzgerald KA, Tai PWL, et al. Overcoming innate immune barriers that impede AAV gene therapy vectors. J Clin Invest. 2021; 131(1).