Chapter

The Diversity of Parvovirus Telomeres

Marianne Laugel, Emilie Lecomte, Eduard Ayuso, Oumeya Adjali, Mathieu Mével and Magalie Penaud-Budloo

Abstract

Parvoviridae are small viruses composed of a 4–6 kb linear single-stranded DNA protected by an icosahedral capsid. The viral genes coding non-structural (NS), capsid, and accessory proteins are flanked by intriguing sequences, namely the telomeres. Telomeres are essential for parvovirus genome replication, encapsidation, and integration. Similar (homotelomeric) or different (heterotelomeric) at the two ends, they all contain imperfect palindromes that fold into hairpin structures. Up to 550 nucleotides in length, they harbor a wide variety of motifs and structures known to be recognized by host cell factors. Our study aims to comprehensively analyze parvovirus ends to better understand the role of these particular sequences in the virus life cycle. Forty *Parvoviridae* terminal repeats (TR) were publicly available in databases. The folding and specific DNA secondary structures, such as G4 and triplex, were systematically analyzed. A principal component analysis was carried out from the prediction data to determine variables signing parvovirus groups. A special focus will be put on adeno-associated virus (AAV) inverted terminal repeats (ITR), a member of the genus *Dependoparvovirus* used as vectors for gene therapy. This chapter highlights the diversity of the *Parvoviridae* telomeres regarding shape and secondary structures, providing information that could be relevant for virus-host interactions studies.

Keywords: parvovirus, telomeres, DNA folding, DNA secondary structure, adeno-associated virus

1. Introduction

Many linear DNA viruses possess terminal repeats (TRs) known to be critical for viral genome stability and propagation [1]. A parallel can be drawn with human chromosome telomeres that are composed of GC-rich repeat sequences of 5–10 nucleotides. In cells, telomeres are critical to maintain the linear structure of the chromosomes. They can adopt specific secondary structures, such as G-quadruplexes (G4), providing structural characteristics for protein binding and genomic stability. In addition, Cellular telomeres play a role in transcription regulation, chromatin compaction, subcellular localization, and chromosome segregation.

Similarly, *Parvoviridae* TRs have been demonstrated to be essential to several steps of the virus cycle. They vary in shape and size from approximatively 100 to 550 nucleotides [2]. Due to the presence of palindromic repeats, TRs can fold into T-, I-, J-, Y-, U- shape or simple hairpin-like structures. Up to now, the global shape

has been named without any consensus, for example, Y-shape also being called "rabbit ears."

Viral genomes in some genera (*Ambi-* and *Itera-densovirus*; *Ave-*, *Dependo-*and *Erythro-parvovirus*) are homotelomeric meaning that both termini are similar but inverted, whereas, in other genera (*Brevi-* and *Hepan-densovirus*; *Amdo-*, *Boca-* and *Proto-parvovirus*), the 5′ and 3′ ends of the linear genome differ and therefore are called heterotelomeric. The strand polarity packed in viral capsids may be related to the left and right TRs dissimilarity. Indeed, most of the heterotelomeric parvoviruses encapsidate only one strand polarity, mainly negative. This preference may be due to inefficient nicking during replication or incomplete packaging signal at one TR [2]; for example, the minute virus of mice (MVM), a virus of the subfamily *Parvovirinae* and genus *Protoparvovirus*, harbors a Y-shape left end and a longer U-shape structure on its right end. After replication and ori resolution, the single-stranded DNA of minus polarity is preferentially displaced from the left TR and encapsidated. For parvovirus with both polarities, the proportion can range from 1 to 50% and may be influenced by the host cell in which the virus is produced [3].

Parvoviral TRs are involved in many steps of the virus life cycle. They contain most of the *cis*-acting information required for genome replication and encapsidation, including tetranucleotide repeats that serve as binding sites for NS1 (Rep) oligomer, a resolution site necessary for the completion of the DNA strand copy, and a packaging signal. Recognized as DNA double-strand breaks (DSB) in the host cell, TR can trigger a DNA damage response (DDR), leading to the circularization and concatemerization of the viral genomes either by non-homologous end-joining (NHEJ) or homologous recombination (HR) [4]. Finally, transcription regulation elements are contained in the genome ends. For example, the MVM TRs contain both symmetric and asymmetric binding sites for transcription factors that modulate expression from the adjacent P4 promoter [5] and the Acheta Domestica Densovirus TRs contain a TATA box used for transcription initiation of NS gene on one side and VP on the other side [6].

The TRs secondary structures, motifs composition, and their role in the virus-cell cycle have been under-examined. In this study, DNA secondary structures of the *Parvoviridae* TRs, including non-canonical secondary structures, have been predicted. We have shown a high diversity of parvovirus telomeres characteristics even within a genus. This chapter may provide significant knowledge for *Parvoviridae* classification and interaction with host cells.

2. The intriguing shape diversity of parvovirus telomeres

2.1 Size, GC content, and shape of parvovirus ends

Although *Parvoviridae* genomes have been extensively studied, in particular for phylogenetic and evolutionary analyses, the sequence and characteristics of their telomeres are not clearly described. Therefore, we analyzed *Parvoviridae* TRs sequences publicly available in the NCBI GenBank database. *Parvoviridae* complete genomes were downloaded. At least one representative virus per genus was selected based on their notoriety and information available on the internal committee on the taxonomy of viruses (ICTV) website. TRs were annotated following GenBank annotation or information available in the literature (**Table S1**). TR sequences of homotelomeric genomes were then verified by aligning the 5′ and 3′ ends. Sequences differing in length and showing no homology (with no common palindromic regions between 5′ and 3′ ends) were discarded from the data set. Finally, the presence of palindromic regions was verified by RNAfold (method described

later) [7]. A total of 40 *Parvoviridae* 5′ and the 3′ TRs sequences were extracted for further analysis. Among those are 17 *Densovirinae*, 22 *Parvovirinae,* and 1 unclassified *Parvoviridae* telomeres.

First, the length of each TR was determined and listed in **Table S1**. Interestingly, TR length varies within a single genus (**Table 1**), for example, going from 122 nucleotides for the PcDV to 550 nucleotides for the GmDV in the same genus *Ambidensovirus*. Second, the percentage of GC was calculated for each parvovirus TR. **Figure 1** highlights the GC content diversity of TRs between parvoviruses. The minimum and maximum GC content was observed for the left TR of AalDV2 and AAV2 with 32.4% and 69.7%, respectively. Within the genus *Ambidensovirus*, the percentage of GC ranges from 35% to 58.5% with the lowest GC content being attributed to the 5′ TR of the PcDV (**Figure 1b**). Comparatively, telomeres of the human chromosomes contain 50–55% of G and C bases whereas the whole human genome contains 40.9% GC on average [8].

To visualize the general shape and the secondary structures of the viral TRs, the folding of each parvovirus TR was predicted by RNAfold program using parameters of the Turner model for single-stranded RNA and DNA and the Matthews model for double-stranded DNA [7]. Additionally, mFold program was used in the DNA mode to corroborate the predictions [9]. The most thermodynamically stable structures, or minimum free energy (MFE) structures, obtained on the RNAfold web server were used to propose a classification of the TR (**Figure 2**). Four groups were constituted according to the number of hairpin loops at their extremity and named H1 (previously named U- and I-shapes in the literature), H2 (corresponding to J-, Y-, T-shapes), H3, and H4 (**Figure 2**, **Table S1**). This classification based on the number of terminal hairpin loops after folding and on additional structural characteristics may be more informative and precise than the global shape. Moreover, this nomenclature is applicable to all parvovirus TR. Interestingly, TR sequences and shapes differ within a genus. For example, among *Ambidensovirus*, CpDV and DicDV 5′ TRs are both classified in the H1 group although they only share 43% of sequence homology. In the genus *Bocaparvovirus*, HBoV1 and BpV1 left ends are 62% homologous in sequence but form a terminal H1 and H2 shape, respectively. Phylogenetic and evolution analyses of *Parvoviridae* have been constructed on the basis of NS1 proteins homology. Telomeres have never been considered as a classification criterion.

2.2 Comprehensive analysis of DNA secondary structural elements

The global analysis of the parvovirus TR has highlighted their broad diversity, even within the same genus. To study the TR divergence, an in-depth prediction of the secondary structures followed by a principal component analysis (PCA) have

Subfamily	Genus	5′ TR length	3′ TR length
Parvovirinae	*Bocaparvovirus*	[140–161]	[161–200]
	Dependoparvovirus	[141–455]	[141–455]
	Erythroparvovirus	[94–383]	[94–383]
Densovirinae	*Ambidensovirus*	[122–550]	[122–550]
	Brevidensovirus	[98–182]	[134–165]
	Iteradensovirus	[101–271]	[101–271]

Table 1.
Minimal and maximal length of five-prime and three-prime terminal repeats within genera of the Parvoviridae *family.*

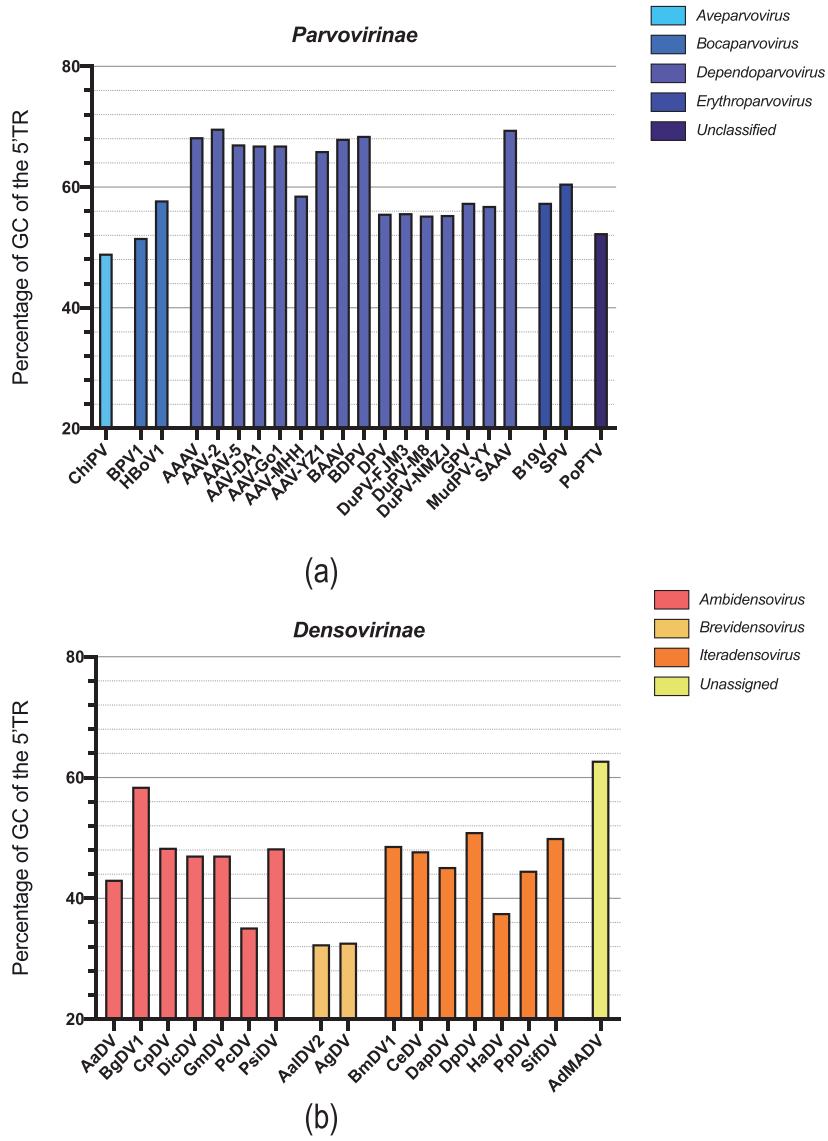


Figure 1.
GC content of 5-prime terminal repeats of the Parvoviridae *family: Differences inside the* Parvovirinae *(a) and the* Densovirinae *(b) subfamilies. The telomere sequences were downloaded from the NCBI GenBank database (see* ***Table S1*** *for accession numbers). GC content was calculated using the APE program.*

been realized. Secondary structure elements (**Figure 2**) and non-B form DNA structures were included as variables in the PCA.

Non-canonical specific structures are susceptible to be recognized by cellular proteins and thus to be essential in the virus-host interactions. For example, a recent study reported that special structures in DNA, such as quadruplex structures, can preferentially bind to IFI16 and trigger more potent type I IFN responses than those produced by the same sequence in dsDNA [10]. Such structures are intrinsic in many viral genomes, such as those of EBV and HPV [11]. Rich in GC, viral telomeres may also contain non-B DNA structures, such as G-quadruplexes (G4) or triplexes.

Therefore, putative G4 and triplexes were determined in all the parvoviral TR. G4 have been non-canonical DNA secondary structures formed by G-rich sequences (**Figure 3a**). Present in human telomeres, they are suggested to participate in chromosome stability maintenance [12]. G4 have also been shown to be present and play major roles in almost all virus families [13]. G4 have also been described in some parvovirus

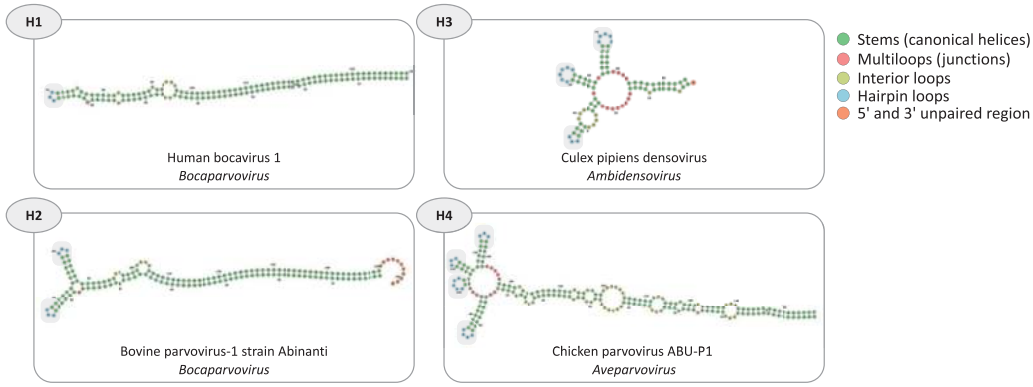


Figure 2.
Examples of five-prime terminal repeat folding among the Parvoviridae *family. The most thermodynamically stable structures or minimum free energy (MFE) structures were obtained using the RNAfold program (RNA mode: http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). Four groups (H1 to H4) were generated according to the number of hairpin loops (in blue) found at the five-prime TR extremity. The following DNA secondary structures were identified and counted: Stems (green), multiloops (red), interior loops (yellow), and hairpin loops (blue).*

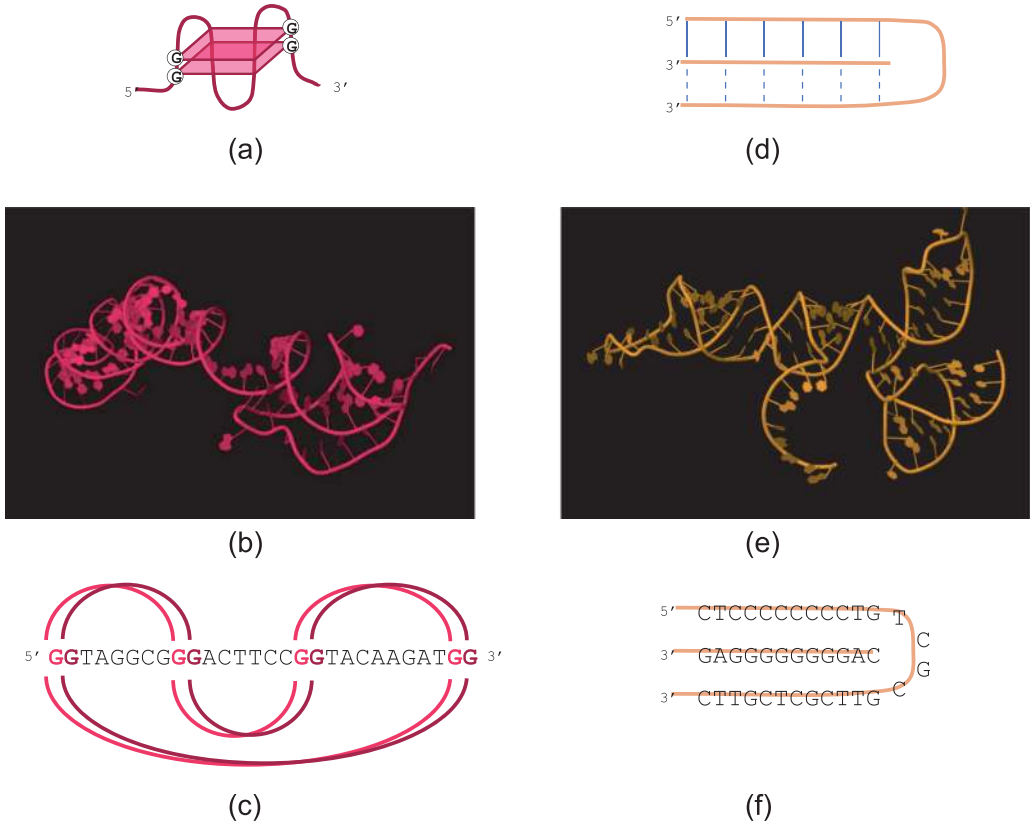


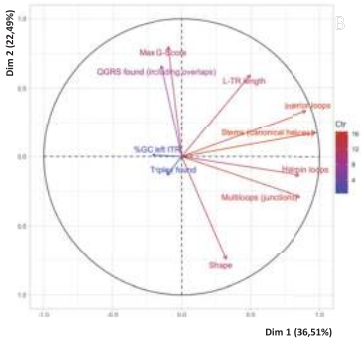
Figure 3.
Two non-B secondary structures: G-quadruplex (G4) and triplex. (a) 3D representation of a theoretical G4. (b) 3D representation of a G4 in the parvovirus B19 five-prime telomere and correspondence to the G tetrads sequence (c). (d) Triplex theoretical 2D representation. (e) 3D representation of a triplex of the five-prime bovine AAV terminal repeat and (f) correspondence to the sequence. (b) and (e) figures were obtained using the Jmol software (Jmol: An open-source Java viewer for chemical structures in 3D. http://www.jmol.org/).

telomeres [14] but has not been systematically predicted in all parvovirus ends. G4 were predicted using the online tool QGRS-mapper [15] using the search parameters—QGRS max length 45, min G-group size 2, loop size from 0 to 12. These criteria are deliberately drastic to increase the stringency and relevance of the G4 prediction. Three values were

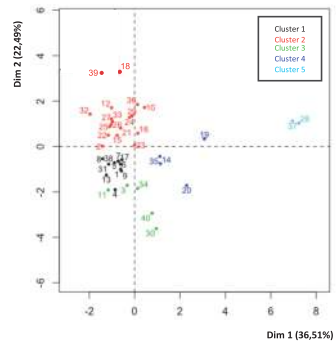
collected—the raw number of predictive G4, with and without overlaps, and the QGRS max-score rewarding the G4 that are more likely to form. The *erythroparvovirus* B19V contains four G4 without overlaps which represent the maximum number of these non-B motifs for parvovirus TR. Including overlaps, CeDV and SifDV TRs harbor the highest number of putative tetraplex DNA structures with 296 G4. Of note, the two *brevidensoviruses* lack any predictive G4 in their ends. No correlation exist between the length of TR and the number of predictive G4 (data not shown).

In parallel, triplexes are important non-B form DNA structures for protein recognition, such as for the binding of p53 factor [16]. Triplex can form at homopurine:homopyrimidine sequences with mirror symmetry (**Figure 3b**). The triplex package of the R program was used to predict the existence of intramolecular triplex DNA structures in parvovirus TR [17]. Only two triplexes were found, both in the *bocaparvovirus* BAAV ends.

Finally, a PCA was performed for the forty left TRs and with the following variables—length, GC content, shape, max G-score, raw number of G4 with overlaps, and secondary structures elements (hairpins loops, interior loops, junction loops, and stems) collected from RNAfold analysis. The R package FactoMineR was used [18]. The main PCA variables are the stems and loops (**Figure 4a**). The "hairpin loops" criteria is one of the most important element allowing division of parvoviruses into groups, hence the relevance of our proposed classification in shapes H1 to H4. Clustering was subsequently realized on the three most informative dimensions corresponding to more than 70% of the cumulative variance. Five clusters were obtained (**Figure 4b**).



(a)



(b)

	Abbrevation	Genus	Cluster
1	AAAV	*Dependoparvovirus*	1
4	AAV-2	*Dependoparvovirus*	1
5	AAV-5	*Dependoparvovirus*	1
6	AAV-DA1	*Dependoparvovirus*	1
7	AAV-Go1	*Dependoparvovirus*	1
8	AAV-MHH	*Dependoparvovirus*	1
9	AAV-YZ1	*Dependoparvovirus*	1
13	BAAV	*Dependoparvovirus*	1
17	BPV1	*Bocaparvovirus*	1
31	HBoV1	*Bocaparvovirus*	1
38	SAAV	*Dependoparvovirus*	1

	Abbrevation	Genus	Cluster
2	AaDV	*Ambidensovirus*	2
10	AdMADV	*Unassigned*	2
12	B19V	*Erythroparvovirus*	2
15	BgDV1	*Ambidensovirus*	2
16	BmDV1	*Iteradensovirus*	2
18	CeDV	*Iteradensovirus*	2
21	DapDV	*Iteradensovirus*	2
22	DicDV	*Ambidensovirus*	2
23	DpDV	*Iteradensovirus*	2
24	DPV	*Dependoparvovirus*	2
25	DuPV-FJM3	*Dependoparvovirus*	2
26	DuPV-M8	*Dependoparvovirus*	2
27	DuPV-NMZJ	*Dependoparvovirus*	2
29	GPV	*Dependoparvovirus*	2
32	MokPV	*Unclassified Parvoviridae*	2
33	MudPV-YY	*Dependoparvovirus*	2
36	PpDV	*Iteradensovirus*	2
39	SifDV	*Iteradensovirus*	2

	Abbrevation	Genus	Cluster
3	AalDV2	*Brevidensovirus*	3
11	AgDV	*Brevidensovirus*	3
30	HaDV	*Iteradensovirus*	3
34	PcDV	*Ambidensovirus*	3
40	SPV	*Erythroparvovirus*	3
14	BDPV	*Dependoparvovirus*	4
19	CpDV	*Ambidensovirus*	4
20	ChiPV	*Aveparvovirus*	4
35	PoPTV	*Unclassified Parvovirinae*	4
28	GmDV	*Ambidensovirus*	5
37	PsiDV	*Ambidensovirus*	5

(c)

Figure 4.
Principal component analysis (PCA) conducted on the five-prime telomere repeats data set. The PCA analysis was conducted using the R software and the Factoshiny package. (a) Contribution graph of each variable. (b) Clusterisation of the 40 parvovirus TR. (c) Correspondence of each parvovirus in the 5 clusters.

Cluster 1, composed of individuals such as the HBoV1 and AAV2, one of the most famous parvoviruses in the gene therapy field, is characterized by a high value for the variable "GC content." Parvovirus B19 belongs to cluster 2, a group characterized by high values for the G4 scores and TR length. Cluster 3 mainly depends on the shape class. Individuals in cluster 4 hold a similar number of multiloops and hairpin loops. Finally, viruses in cluster 5 share many DNA structure common features (stems, interior loops, hairpin loops, multiloops, and length). Clusters do not perfectly correlate with phylogenetic classification (**Figure 4c**), however, we observed that cluster 1 is only composed of *Dependoparvovirus* and *Bocaparvovirus* and cluster 5 contains two *Ambidensovirus*, GmDV and PsiDV. Interestingly, the latter highly differs from other groups.

3. Focus on adeno-associated virus (AAV) inverted terminal repeats (ITR)

The use of vectors derived from the adeno-associated virus (AAV) for gene delivery encounters a growing success for the treatment of a variety of human

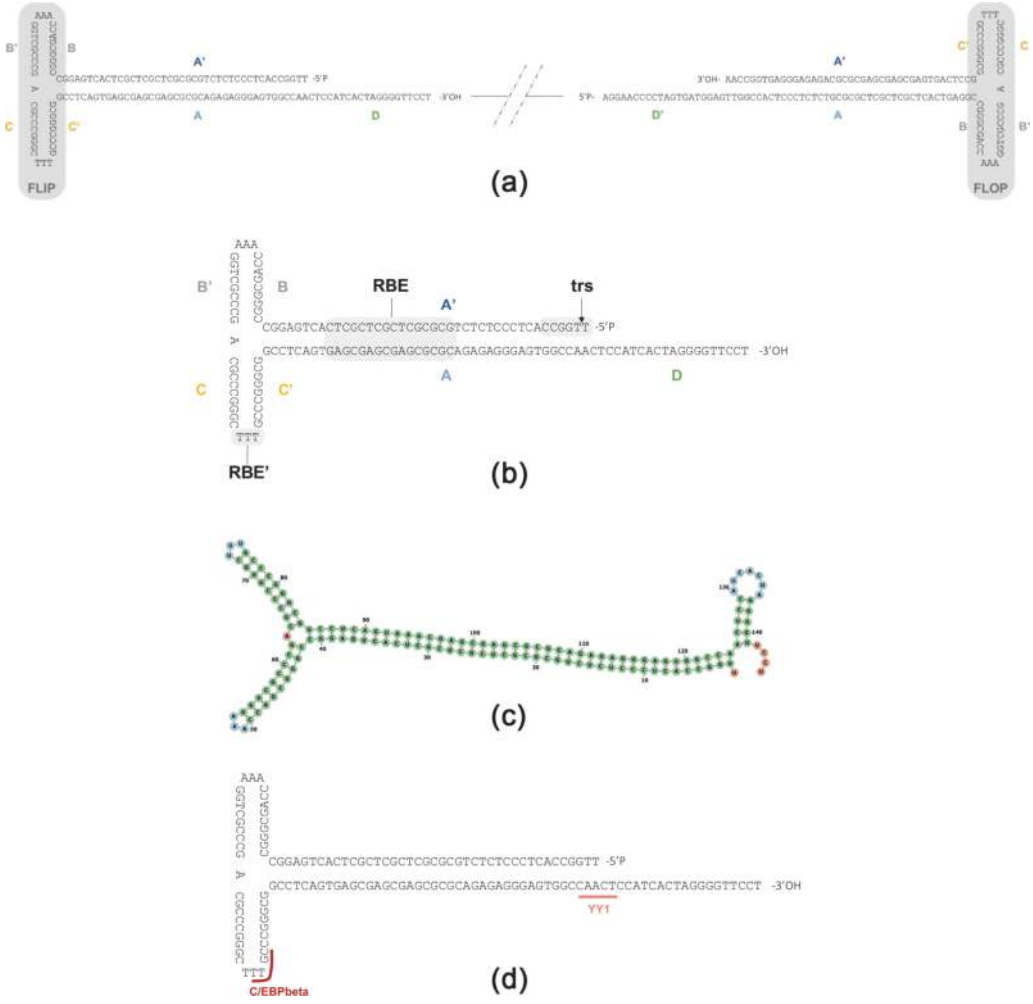


Figure 5.
Inverted terminal repeats of the adeno-associated virus (AAV) serotype 2. (a) Scheme of five prime and three-prime ITRs of the wild-type AAV serotype 2. (b) Two-dimensional drawing of the five-prime ITR in flip configuration. (c) Predictive folding of the five-prime AAV2 ITR using RNAfold. The color code is the same that for ***Figure 2*** *(http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). (d) Putative human transcription factor binding sites in AAV2 ITR. ITR regions are named based on Lusby and Berns' publication [22].*

diseases [19]. Nevertheless, the scientific community has recently faced tragic toxicity of AAV vectors administered intravenously at high doses in several clinical trials [20]. AAV vectors are generated by inserting a recombinant genome usually flanked by AAV-2 inverted terminal repeats (ITR) in an AAV capsid. The recent side-effects observed in human trials have raised the question of DNA sensing, in particular, ITR detection and subsequent cellular responses [21]. Considering the importance of providing new knowledge in this field, a special focus on AAV-2 ITR was included in our study.

The homotelomeric AAV2 possesses two identical ITR of 145 nucleotide-long. The first 125 bases contain three palindromic sequences allowing the ITR to form a T-shape structure composed of two small inverted repeat sequences (BB′ and CC′) and a larger repeated sequence (AA′) (**Figure 5b**). According to our analysis, AAV2 ITR belongs to group H2 (**Figure 5c**). A fourth proximal region called D remains single-stranded if not annealed to the opposite polarity strand or not in an intramolecular manner to the D′ region in 3′. Each ITR can be found in two alternative configurations termed "flip" and "flop" distinguishable by the BB'-CC' orientation (**Figure 5a**), as a direct result of the replication mechanism. There are nomenclature inconsistencies in the literature. Here, ITR regions are named based on Lusby and Berns' publication [22] and ordered as followed ABB′CC′A′D from 5′ to 3′.

For historical reasons and the sake of convenience, most of the AAV vectors contain the ITR of AAV serotype 2, the sole viral sequences required for the replication and packaging of the recombinant genome in AAV capsids. Additional functions of AAV2 ITR have been described, such as a promoter activity [23], a role in the virus persistence either through genome integration [24] or recombination to form monomeric or concatemeric episomes [25].

Strikingly, the GC content of AAV2 ITR corresponds to the highest score of all studied parvoviral telomeres (69%). No predictive G4 was found using stringent parameters, unlike Satkunanathan *et al.* who described 18 QGRS inside the AAV2 ITR sequences [14]. In addition, no potential triplex structure was found. According to the PCA (**Figure 4**), AAV2 belongs to cluster 1 with several other *Dependoparvoviruses*.

Putative binding sites for human transcription factors (TF) have already been described in AAV2 ITR [26, 27]. We completed this work by analyzing human TF

Serotype	Accession number	Transcription factors				
		C/EBPbeta [T00581]	Pax-5 [T00070]	YY1 [T00915]	AP-2alphaA [T00035]	GR-alpha [T00337]
1	NC_002077	2	2	1	1	0
2	NC_001401	1	0	1	0	0
3A	JB292182.1	1	2	2	0	0
3B	AF028705.1	1	2	1	0	0
4	NC_001829	1	0	1	0	0
5	NC_006152.1	2	0	0	0	2
6	AF028704	1	0	1	0	0
7	NC_006260	0	2	1	1	0
AAV_CHC1017	MK139265.1	1	0	1	0	0

Table 2.
Putative recognition sites of human transcription factors in inverted terminal repeats of AdenoAssociated virus serotypes. Predictions was realized using Alggen-promo tool with 0% sequence dissimilarity (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3).

for AAV serotypes 1 to 7 ITRs using the Alggen-Promo tool with 0% sequence dissimilarity (**Table 2**) [28, 29]. Five human TF sites were found in AAV ITR: C/EBPbeta, Pax-5, YY1, AP-2alphaA, and GR-alpha. The C/EBPbeta was found in most of the AAV serotypes and unlike the other found TF, is mainly involved in immune responses. In our study, the GR-alpha was only found in AAV5 ITR contains two predictive TF binding sites, one for C/EBPbeta and other for YY1 (**Figure 5d**). YY1 participates in the initial steps of replication by binding to the p5 promoter region of AAV [30, 31].

4. Conclusion

The genomic and structural diversity of parvovirus is today classified by phylogeny analysis showing an expected separation between parvoviruses and densoviruses, but its robustness is relative, suggesting that the introduction of new sequences could change our perception of their evolutionary history [32]. The diversity of sequences, structures, and genomic organizations of parvoviruses suggest evolutionary histories that are probably more complex than those illustrated by current phylogenies. These observations led us to analyze and characterize the intriguing terminal sequences present in all parvoviruses, namely the telomeres.

This chapter highlights the diversity of *Parvoviridae* telomeres through a complete analysis of the terminal ends folding and secondary structures. Their length, GC content, and global shape vary even within a genus between phylogenetically closely related viruses. Evolution also led to heterotelomeric viruses with completely different left and right extremities. The diversity suggests high importance of these particular structures. Yet, factors involved in the TR selective pressure are unknown. Cotmore and Tattersall suggested a link between the resolution mechanism, the strand polarity, and the TR conformation [4, 33], while Tijssen et al. suggested that the significant differences in size and secondary structure of genome end between genera might reflect a dependence on specific cellular factors necessary for replication and encapsidation [34]. Consistently, we hypothesized that TR may have evolved according to the interactions with their replicase, helper virus co-factors, and/or cell host proteins. Based on data integration of predictive DNA secondary structures in a PCA, new groups were made that were distinct from the ICTV phylogenetic classification conducted from the NS replicase sequences.

Additionally, the significance of specific secondary structures in the parvovirus life cycle and the relation with strand polarity of the packaged linear genome are interesting topics deserving further investigations. The MVM, canine parvovirus (CPV), BPV1 (*Parvovirinae*), and the AalDV2 (*Densovirinae*) encapsidate only or predominantly negative-strand polarity genomes and possess heterotelomeric TR [35, 36]. On the contrary, the homotelomeric AAV2 encapsidates both strands polarities at the same level. By having different shapes and different secondary structure elements, the TR directly impacts the polarity of the encapsidated strand.

Finally, a special emphasis was put on the ITRs of the adeno-associated virus serotype 2, taking into consideration its importance in the world of gene transfer using viral vectors. Particular motifs and secondary structures within AAV ITR may have a significant impact on gene transfer efficiency. Indeed, it has already been demonstrated that AAV2 ITRs are detected by cellular factors belonging to the NHEJ and HR-DNA damage pathways [37]. The viral telomeres may also be recognized by DNA sensors which subsequently could restrict AAV vectors transduction or activate innate immune responses [21]. Consistent with this hypothesis, a variety of cellular proteins have been shown to interact with AAV2 ITR, such as nucleophosmin (NPM1), a protein involved in ribosome biogenesis and nucleolus

transport of basic proteins. Notably, NPM1 binds preferentially G4. The restriction factor FKBP52 in its phosphorylated form also binds to the ITR in the D region, inhibits the second strand synthesis, and consequently decreases transgene expression [38]. Thus, the involvement of ITR recognition by cellular factors is central to understand the extent of subsequent responses to the rAAV DNA that can negatively impact the therapeutic gene expression and cause potential safety concerns for the patients. Using drastic parameters, no putative G4 or triplex were found in AAV2 ITR contrary to a previous study [14]. The formation of these non-conventional DNA motifs highly depends on the adjacent sequences as well as pH and ion concentration conditions and thus requires to be confirmed experimentally.

Acknowledgements

The authors would like to thank Judith Penzes for the initial discussion about phylogeny.

This work was supported by Nantes University Hospital.

Conflict of interest

The authors declare no conflict of interest.

Appendix

Subfamily	Genus	Virus name	Abbreviation	Accession number	5′ TR length	5′ TR shape	3′ TR length	3′ TR shape	Reference used for TR annotation
Parvovirinae	*Aveparvovirus*	Chicken parvovirus ABU-P1	ChiPV	GU214704.1	206	H4	206	H4	[39]
	Bocaparvovirus	Bovine parvovirus-1 strain Abinanti	BPV1 HBoV1	NC038895	161	H2	161	H1	[40]
		Human bocavirus 1		JQ923422	140	H1	200	H1	[41]
	Dependoparvovirus	adeno-associated virus 2	AAV-2	NC_001401	145	H2	145	H2	[42]
		adeno-associated virus 5	AAV-5	NC_006152.1	167	H2	167	H2	[43]
		Adeno-associated virus isolate MHH-05-2015	AAV-MHH	NC040671	174	H2	174	H2	[44]
		Adeno-associated virus-Go.1 (caprine)	AAV-Go1	DQ335246	167	H2	167	H2	[45]
		Avian adeno-associated virus ATCC VR-865	AAAV	NC004828	142	H2	142	H2	[46]
		Avian adeno-associated virus strain DA-1	AAV-DA1	NC006263	142	H2	143	H2	[47]
		Avian adeno-associated virus strain YZ-1	AAV-YZ1	GQ368252	141	H2	141	H2	[48]
		Bearded dragon parvovirus	BDPV	NC027429	257	H2	257	H2	[49]
		Bovine AAV	BAAV	NC005889	172	H2	172	H2	[50]
		Duck parvovirus strain FJM3	DuPV-FJM3	KR075690	359	H1	359	H1	[51]

Subfamily	Genus	Virus name	Abbreviation	Accession number	5′ TR length	5′ TR shape	3′ TR length	3′ TR shape	Reference used for TR annotation
		Duck parvovirus strain M8	DuPV-M8	KR029614	387	H1	387	H1	[52]
		Duck parvovirus strain NMZJD110	DuPV-NMZJ	KR075691.1	415	H1	415	H1	[52]
		Goose parvovirus	GPV	U25749	444	H1	444	H1	[52]
		Muscovy duck parvovirus FM	DPV	NC_006147.2	457	H1	455	H1	[51]
		Muscovy duck parvovirus YY	MudPV-YY	KU844281	452	H1	452	H1	[51]
		Serpentine adeno-associated virus	SAAV	NC006148	154	H2	154	H2	[53]
	Erythroparvovirus	B19 virus isolate J35 Simian adeno-associated virus	B19V	AY386330.1	383	H1	383	H1	[54]
			SPV	KT984498	94	H3	95	H3	[55]
	Unclassified Parvovirinae	Porcine partetravirus strain FMV10-1437266	PoPTV	NC022104.1	210	H2	210	H2	[56]
Densovirinae	*Ambidensovirus*	Acheta domestica densovirus	AaDV	HQ827781	144	H1	144	H1	[57]
		Blattella germanica densovirus 1	BgDV1	NC005041	217	H1	216	H1	[58]
		Culex pipiens densovirus	CpDV	NC012685	285	H1	285	Unclassified	[59]
		Diaphorina citri densovirus	DicDV	NC030296.1	210	H1	210	H1	[60]
		Galleria mellonella densovirus	GmDV	NC_004286	550	H2	550	H2	[61]
		Planococcus citri densovirus	PcDV	NC004289.1	122	H2	122	Unclassified	[62]
		Pseudoplusia includens densovirus	PsiDV	NC019492.1	540	H2	540	H2	[63]
	Brevidensovirus	Aedes albopictus densovirus 2	AalDV2	NC004285.	182	H2	134	H2	[64]
		Anopheles gambiae densonucleosis virus	AgDV	NC_011317.1	98	H2	165	H2	[65]

Subfamily	Genus	Virus name	Abbreviation	Accession number	5′ TR length	5′ TR shape	3′ TR length	3′ TR shape	Reference used for TR annotation
	Iteradensovirus	Bombyx mori densovirus 1	BmDV1	NC003346.1	230	H1	230	H1	[66]
		Casphalia extranea densovirus	CeDV	NC004288.1	230	H1	230	H1	[67]
		Danaus plexippus plexippus iteravirus isolate Granby	DapDV	NC023842	239	H1	239	H1	[68]
		Dendrolimus punctatus densovirus	DpDV	NC006555.1	200	H2	200	H2	[69]
		Helicoverpa armigera densovirus	HaDV	NC015718	101	H4	101	H1	[70]
		Papilio polyxenes densovirus	PpDV	NC018450.1	271	H1	271	H1	[71]
		Sibine fusca densovirus	SifDV	NC018399.1	230	H1	230	H1	[72]
	Unassigned	Acheta domesticus mini ambidensovirus isolate Kalamazoo	AdMADV	NC022564.1	199	H2	199	H2	[73]
Unclassified Parvoviridae		Mouse kidney parvovirus strain Centenary Institute	MokPV	NC040843.1	145	H1	118	H1	[74]

The phylogenetic classification used here refers to the most up-to-date from the International Comitee on Taxonomy of Virus (ICTV) published in 2020. Abbreviations were taken from the literature; when not existing, they were created taking the first letters of the virus name. The TR shape were annotated according to our classification proposed in this chapter. The reference used for ITR annotations does not always match with the first citation of the virus.

Table S1.
List of the forty Parvoviridae *terminal repeats analyzed in the study.*

Author details

Marianne Laugel, Emilie Lecomte, Eduard Ayuso, Oumeya Adjali, Mathieu Mével and Magalie Penaud-Budloo*
INSERM UMR 1089, University of Nantes, Centre Hospitalier Universitaire, Nantes, France

*Address all correspondence to: magalie.penaud-budloo@univ-nantes.fr

IntechOpen

References

[1] Deng Z, Wang Z, Lieberman PM. Telomeres and viruses: Common themes of genome maintenance. Frontiers in Oncology. 2012;**2**:201

[2] 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. Archives of Virology. 2020;**165**(9):2133-2146

[3] King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. Family - Parvoviridae. In: Virus Taxonomy. San Diego: Elsevier; 2012. pp. 405-425

[4] Cotmore SF, Tattersall P. Parvovirus diversity and DNA damage responses. Cold Spring Harbor Perspectives in Biology. 2013;**5**(2)

[5] 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):466-470

[6] Pham HT, Yu Q, Bergoin M, Tijssen P. A novel Ambisense Densovirus, Acheta domesticus Mini Ambidensovirus, from crickets. Genome Announcements. 2013;**1**(6)

[7] Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms for Molecular Biology. 2011;**6**(1):26

[8] Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. BMC Research Notes. 2019;**12**(1):106

[9] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research. 2003;**31**(13):3406-3415

[10] Hároníková L, Coufal J, Kejnovská I, Jagelská EB, Fojta M, Dvořáková P, et al. IFI16 preferentially binds to DNA with Quadruplex structure and enhances DNA Quadruplex formation. PLoS One. 2016;**11**(6):e0157156

[11] Ma Z, Ni G, Damania B. Innate sensing of DNA virus genomes. Annual Review of Virology. 2018;**5**(1): 341-362

[12] Huppert JL. Structure, location and interactions of G-quadruplexes. The FEBS Journal. 2010;**277**(17):3452-3458

[13] Ruggiero E, Richter SN. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. Annual Reports in Medicinal Chemistry. 2020;**54**:101-131

[14] Satkunanathan S, Thorpe R, Zhao Y. The function of DNA binding protein nucleophosmin in AAV replication. Virology. 2017;**510**:46-54

[15] Kikin O, D'Antonio L, Bagga PS. QGRS mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Research. 2006;**34**(suppl_2):W676-W682

[16] 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

[17] Lexa M, Martínek T, Burgetová I, Kopeček D, Brázdová M. A dynamic programming algorithm for identification of triplex-forming sequences. Bioinformatics. 2011;**27**(18):2510-2517

[18] Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. Journal of Statistical Software. 2008; **25**(1):1-18

[19] Samulski RJ, Muzyczka N. AAV-mediated gene therapy for research and therapeutic purposes. Annual Review of Virology. 2014;**1**(1):427-451

[20] Wilson JM, Flotte TR. Moving forward after two deaths in a gene therapy trial of Myotubular myopathy. Human Gene Therapy. 2020;**31**(13–14): 695-696

[21] Muhuri M, Maeda Y, Ma H, Ram S, Fitzgerald KA, Tai PWL, et al. Overcoming innate immune barriers that impede AAV gene therapy vectors. The Journal of Clinical Investigation. 2021;**131**(1)

[22] Lusby E, Fife KH. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. Journal of Virology. 1980;**34**:8

[23] Earley LF, Conatser LM, Lue VM, Dobbins AL, Li C, Hirsch ML, et al. Adeno-associated virus serotype-specific inverted terminal repeat sequence role in vector transgene expression. Human Gene Therapy. 2020;**31**(3–4):151-162

[24] Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N, et al. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. The EMBO Journal. 1991;**10**(12):3941-3950

[25] Duan D, Sharma P, Yang J, Yue Y, Dudus L, Zhang Y, et al. Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term Episomal persistence in muscle tissue. Journal of Virology. 1998;**72**:10

[26] Ling C, Wang Y, Lu Y, Wang L, Jayandharan GR, Aslanidi GV, et al. Enhanced transgene expression from recombinant single-stranded D-sequence-substituted adeno-associated virus vectors in human cell lines *In vitro* and in murine hepatocytes *In vivo*. Journal of Virology. 2015;**89**(2):952-961

[27] Julien L, Chassagne J, Peccate C, Lorain S, Piétri-Rouxel F, Danos O, et al. RFX1 and RFX3 transcription factors interact with the D sequence of adeno-associated virus inverted terminal repeat and regulate AAV transduction. Scientific Reports. 2018;**8**(1):210

[28] Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, Albà MM. PROMO: Detection of known transcription regulatory elements using species-tailored searches. Bioinformatics. 2002;**18**(2):333-334

[29] Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Research. 2003;**31**(13):3651-3653

[30] Houbaviy HB, Burley SK. Thermodynamic analysis of the interaction between YY1 and the AAV P5 promoter initiator element. Chemistry & Biology. 2001;**8**(2):179-187

[31] Murphy M, Gomos-Klein J, Stankic M, Falck-Pedersen E. Adeno-associated virus type 2 p5 promoter: A rep-regulated DNA switch element functioning in transcription, replication, and site-specific integration. Journal of Virology. 2007;**81**(8):3721-3730

[32] Grenet A-SG, Salasc F, Francois S, Mutuel D, Dupressoir T, Multeau C, et al. Les densovirus : une « massive attaque » chez les arthropodes. Virologie. 2015;**19**(1):19-31

[33] Cotmore SF, Tattersall P. Parvoviruses: Small does not mean simple. Annual Review of Virology. 2014;**1**(1):517-537

[34] Tijssen P, Pénzes JJ, Yu Q, Pham HT, Bergoin M. Diversity of small, single-stranded DNA viruses of

invertebrates and their chaotic evolutionary past. Journal of Invertebrate Pathology. 2016;**140**:83-96

[35] Cotmore SF, Tattersall P. Genome packaging sense is controlled by the efficiency of the Nick site in the right-end replication origin of parvoviruses minute virus of mice and LuIII. Journal of Virology. 2005;**79**(4):2287-2300

[36] Yu Y, Zhang J, Wang J, Xi J, Zhang X, Li P, et al. Naturally-occurring right terminal hairpin mutations in three genotypes of canine parvovirus (CPV-2a, CPV-2b and CPV-2c) have no effect on their growth characteristics. Virus Research. 2019;**261**:31-36

[37] Adachi K, Nakai H. The Role of DNA Repair Pathways in Adeno-Associated Virus Infection and Viral Genome Replication / Recombination / Integration. DNA Repair and Human Health. 2011 Available from: https://www.intechopen.com/books/dna-repair-and-human-health/the-role-of-dna-repair-pathways-in-adeno-associated-virus-infection-and-viral-genome-replication-rec

[38] Qing K, Hansen J, Weigel-Kelley KA, Tan M, Zhou S, Srivastava A. Adeno-associated virus type 2-mediated gene transfer: Role of cellular FKBP52 protein in transgene expression. Journal of Virology. 2001;**75**(19):8968-8976

[39] Day JM, Zsak L. Determination and Analysis of the Full-Length Chicken Parvovirus Genome. Virology. 2010;**399**(1):59-64. DOI: 10.1016/j.virol.2009.12.027

[40] Qiu J, Cheng F, Pintel D. Molecular Characterization of Caprine Adeno-Associated Virus (AAV-Go.1) Reveals Striking Similarity to Human AAV5. Virology. 2006;**356**(1):208-216. DOI: 10.1016/j.virol.2006.07.024

[41] Shen W et al. Identification and Functional Analysis of Novel Nonstructural Proteins of Human Bocavirus 1, éd. par M. J. Imperiale. Journal of Virology. 2015;**89**(19):10097-10109. DOI: 10.1128/JVI.01374-15

[42] Lusby E, Fife KH. Nucleotide Sequence of the Inverted Terminal Repetition in Adeno-Associated Virus DNA. Journal of Virology. 1980;**34**:8

[43] Chiorini JA et al. Cloning and Characterization of Adeno-Associated Virus Type 5. Journal of Virology. 1999;**73**(2):1309-1319. DOI: 10.1128/JVI.73.2.1309-1319.1999

[44] Su XN et al. Isolation and Genetic Characterization of a Novel Adeno-Associated Virus from Muscovy Ducks in China. Poultry Science. 2017;**96**(11):3867-3871. DOI: 10.3382/ps/pex235

[45] Qiu J, Cheng F, Pintel DJ. Expression Profiles of Bovine Adeno-Associated Virus and Avian Adeno-Associated Virus Display Significant Similarity to That of Adeno-Associated Virus Type 5. Journal of Virology. 2006;**80**(11):5482-5493. DOI: 10.1128/JVI.02735-05

[46] Bossis I, Chiorini JA. Cloning of an Avian Adeno-Associated Virus (AAAV) and Generation of Recombinant AAAV Particles. Journal of Virology. 2003;**77**(12):6799-6810. DOI: 10.1128/JVI.77.12.6799-6810.2003

[47] Estevez C, Villegas P. Sequence Analysis, Viral Rescue from Infectious Clones and Generation of Recombinant Virions of the Avian Adeno-Associated Virus. Virus Research. 2004;**105**(2):195-208. DOI: 10.1016/j.virusres.2004.05.010

[48] Wang J et al. Molecular Characterization and Phylogenetic Analysis of an Avian Adeno-Associated Virus Originating from a Chicken in China. Archives of Virology. 2011;**156**(1):71-77. DOI: 10.1007/s00705-010-0822-x

[49] Pénzes JJ et al. Novel parvoviruses in reptiles and genome sequence of a lizard parvovirus shed light on Dependoparvovirus genus evolution. Journal of General Virology. 2015;**96**(9): 2769-2779. DOI: 10.1099/vir.0.000215

[50] Schmidt M et al. Cloning and Characterization of a Bovine Adeno-Associated Virus. Journal of Virology. 2004;**78**(12):6509-6516. DOI: 10.1128/JVI.78.12.6509-6516.2004

[51] Zádori Z et al. Analysis of the complete nucleotide sequences of Goose and Muscovy Duck Pervoviruses indicates common ancestral origin with adeno-associated virus 2. Virology. 1995; **212**(2):562-573. DOI: 10.1006/viro.1995.1514

[52] Wang J et al. Molecular characterization of a novel Muscovy duck parvovirus isolate: evidence of recombination between classical MDPV and goose parvovirus strains. BMC Veterinary Research. 2017;**13**. DOI: 10.1186/s12917-017-1238-6

[53] Farkas SL. A parvovirus isolated from royal python (Python Regius) is a member of the genus Dependovirus. Journal of General Virology. 2004;**85**(3): 555-561. DOI: 10.1099/vir.0.19616-0

[54] Zhi N et al. Construction and sequencing of an infectious clone of the human parvovirus B19. Virology. 2004; **318**(1):142-152. DOI: 10.1016/j.virol.2003.09.011

[55] Kapusinszky B et al. Case-control comparison of enteric viromes in captive rhesus macaques with acute or idiopathic chronic diarrhea. Journal of Virology. 2017;**91**(18). DOI: 10.1128/JVI.00952-17

[56] Bellehumeur C et al. High-throughput sequencing revealed the presence of an unforeseen parvovirus species in Canadian swine: The porcine partetravirus. The Canadian Veterinary Journal. 2013;**54**(8):787-789

[57] Szelei J et al. Susceptibility of North-American and European crickets to acheta domesticus densovirus (AdDNV) and associated epizootics. Journal of Invertebrate Pathology. 2011;**106**(3): 394-399. DOI: 10.1016/j.jip.2010.12.009

[58] Mukha DV et al. Characterization of a new densovirus infecting the German cockroach, Blattella germanica. Journal of General Virology. 2006;**87**(6):1567-1575. DOI: 10.1099/vir.0.81638-0

[59] Baquerizo-Audiot E et al. Structure and expression strategy of the genome of Culex pipiens Densovirus, a Mosquito Densovirus with an Ambisense Organization. Journal of Virology. 2009; **83**(13):6863-6873. DOI: 10.1128/JVI.00524-09

[60] Nigg JC, Nouri S, Falk BW. Complete genome sequence of a putative Densovirus of the Asian Citrus Psyllid, Diaphorina citri. Genome Announcements. 2016;**4**(4). DOI: 10.1128/genomeA.00589-16

[61] Tijssen P et al. Organization and expression strategy of the Ambisense genome of densonucleosis virus of Galleria mellonella. Journal of Virology. 2003;**77**(19):10357-10365. DOI: 10.1128/JVI.77.19.10357-10365.2003

[62] Thao ML et al. Genetic characterization of a Putative Densovirus from the Mealybug Planococcus Citri. Current Microbiology. 2001;**43**(6):457-458. DOI: 10.1007/s002840010339

[63] Huynh OTH et al. Pseudoplusia includens Densovirus genome organization and expression strategy. Journal of Virology. 2012;**86**(23):13127-13128. DOI: 10.1128/JVI.02462-12

[64] Boublik Y, Jousset F-X, Bergoin M. Complete nucleotide sequence and

genomic organization of the Aedes Albopictus Parvovirus (AaPV) pathogenic for Aedes Aegypti Larvae. Virology. 1994;**200**(2):752-763. DOI: 10.1006/viro.1994.1239

[65] Ren X, Hoiczyk E, Rasgon JL. Viral Paratransgenesis in the malaria vector Anopheles gambiae. PLoS Pathogens. 2008;**4**(8). DOI: 10.1371/journal.ppat.1000135

[66] Bando H et al. Genome organization of the Densovirus from Bombyx Mori (BmDNV-1) and enzyme activity of its capsid. Journal of General Virology. 2001;**82**(11):2821-2825. DOI: 10.1099/0022-1317-82-11-2821

[67] Fédière G et al. Genome organization of Casphalia Extranea Densovirus, a New Iteravirus. Virology. 2002;**292**(2):299-308. DOI: 10.1006/viro.2001.1257

[68] Yu Q, Tijssen P. Iteradensovirus from the Monarch Butterfly, Danaus plexippus plexippus. Genome Announcements. 2014;**2**(2). DOI: 10.1128/genomeA.00321-14

[69] Wang J et al. Nucleotide sequence and genomic organization of a newly isolated densovirus infecting Dendrolimus punctatus. Journal of General Virology. 2005;**86**(8):2169-2173. DOI: 10.1099/vir.0.80898-0

[70] Pengjun X et al. Complete genome sequence of a monosense Densovirus infecting the cotton bollworm, Helicoverpa Armigera. Journal of Virology. 2012;**86**(19):10909. DOI: 10.1128/JVI.01912-12

[71] Qian Y et al. Papilio polyxenes Densovirus has an iteravirus-like genome organization. Journal of Virology. 2012;**86**(17):9534-9535. DOI: 10.1128/JVI.01368-12

[72] Qian Y et al. Iteravirus-like genome organization of a densovirus from Sibine Fusca Stoll. Journal of Virology. 2012;**86**(16):8897-8898. DOI: 10.1128/JVI.01267-12

[73] Pham HT et al. A novel Ambisense Densovirus, Acheta domesticus Mini Ambidensovirus, from crickets. Genome Announcements. 2013;**1**(6). DOI: 10.1128/genomeA.00914-13

[74] Roediger B et al. An atypical parvovirus driving chronic tubulointerstitial nephropathy and kidney fibrosis. Cell. 2018;**175**(2):530-543.e24. DOI: 10.1016/j.cell.2018.08.013