Polish Journal of Veterinary Sciences Vol. 27, No. 3 (2024), 407–418

DOI 10.24425/pjvs.2024.151734

Original article

CpG islands: Features and distribution in the genomes of porcine parvovirus

M.-S. Xu[1], Z.-K. Zhou[3], R.-Y. Xiong[2], L.-B. Zhang[2], C.-Q. Yu[4], Q. Liu[2]

[1] Chongqing Three Gorges Vocational College, Wanzhou 404155, China
[2] Nanchong Key Laboratory of Disease Prevention, Control and Detection in Livestock and Poultry, Nanchong Vocational and Technical College, Nanchong 637131, China
[3] Nanchong Academy of Agricultural Sciences, Nanchong 637002, China
[4] Engineering Center of Agricultural Biosafety Assessment and Biotechnology, School of Advanced Agricultural Sciences, Yibin Vocational and Technical College, Yibin 644199, China

Correspondence to: Q. Liu, e-mail: liuqiang_yyy@163.com, tel.: +86-133-1461-1425;
M.-S. Xu, e-mail: maosenshuai@126.com, tel.: +86-185-2328-6494

Abstract

Porcine parvovirus disease is a reproductive disorder caused by the porcine parvovirus (PPV) in sows and is characterised by miscarriage, stillbirth and mummification in pregnant sows. Porcine parvovirus disease poses a significant threat to pork herds and seriously hinders healthy and sustainable development of the pig farming industry. Currently, there is no effective treatment for porcine parvovirus disease except for prevention and control measures. Based on genotype differences, PPV can be classified into at least eight subtypes, PPV1-PPV8. Epigenetic mechanisms, particularly cytosine methylation of cytosine-phosphate-guanine (CpG) dinucleotides, are proven to have a significant impact on the life cycle of various viruses. Therefore, we selected the PPV genome as the research object and analysed the number, distribution and length of CpG islands in the genome of strains PPV1-PPV8. PPV1-6 had AT rich genomes (GC content ≤50%), whereas PPV7 had a GC content >50%. PPV1, PPV4, PPV5 and PPV6 contained fewer CpG islands (1-5), PPV7 contained moderate CpG islands (6-11) and PPV2 and PPV3 contained more CpG islands (12-16). This study provides a foundation for exploring novel antiviral treatment strategies from an epigenetic perspective.

Keywords: CpG islands, DNA methylation, porcine parvovirus

Introduction

Parvoviruses infect a wide range of hosts, including vertebrates, invertebrates and humans (Streck et al. 2024, Vargas-Bermudez et al. 2023). According to the different host types, the family Parvoviridae can be classified into two subfamilies: Parvovirinae and Densovirinae, which infect vertebrates and invertebrates, respectively (Cotmore et al. 2019, Penzes et al. 2020). The subfamily of parvoviruses consists of eight virus genera, namely, the Aleutian mink disease virus (Amdoparvovirus), avian parvovirus (Aveparvovirus), Bocaparvovirus (Bocaparvovirus), replicativirus (Copiparvovirus), dependent virus (dependent virus), erythrovirus (erythroparvovirus), protoparvovirus (Protoparvovirus) and tetraparvovirus (Vargas-Bermudez et al. 2023). In 1965, Mayr and Mahnel first discovered the porcine parvovirus (PPV), which is one of the main pathogens that causes reproductive disorders in sows (Mayr et al. 1968). In 1967, Cartwright et al. (Cartwright et al. 1969). isolated porcine parvovirus from disease samples when conducting etiological research on infertility, abortion and stillbirth in pigs, thus proving the pathogenic effect of PPV for the first time. Following viral infection in the pre-pregnancy period, the virus invades the embryo or foetus through the placenta, causing abortion, embryo death, foetal malformation, foetal mummification and infertility in sows; however, the sows themselves have no apparent clinical symptoms (Meszaros et al. 2017). Porcine parvovirus may play a role in diarrhoea, dermatitis and respiratory diseases, in addition to coinfection with porcine reproductive and respiratory syndrome virus, porcine circovirus type 2 and other common viruses (Chen et al. 2023, Faustini et al. 2024). In recent years, rising cases of PPV infections have caused marked economic losses to the pig industry.

PPV virus particles are hexagonal or circular, without a capsule, with a diameter of 20-23 nm and an icosahedral equiaxed stereosymmetry (Streck and Truyen 2020, Vargas-Bermudez et al. 2023). The PPV genome is a single-stranded linear DNA, with a size of 4-6 kb and palindromic sequences at both ends (Streck and Truyen 2020, Vargas-Bermudez et al. 2023). Three structural proteins are encoded by the PPV genome: VP1, VP2 and VP3 (Ranz et al. 1989, Vargas-Bermudez et al. 2023). VP1 and VP2 overlap in many amino acid sequences (Streck and Truyen. 2020, Vargas-Bermudez et al. 2023). VP1 is a structural protein necessary for PPV replication and viral particle packaging, and functions to stabilise viral single-stranded DNA (Streck and Truyen 2020, Vargas-Bermudez et al. 2023). Additionally, the VP1 N-terminus contains a proline-rich region that plays an important role in extracellular to intracellular viral transfer, and VP1 has an N-terminus-rich alkaline residue that helps to bind with the single--stranded DNA, thereby initiating the replication and packaging of viral DNA (Streck and Truyen 2020, Vargas-Bermudez et al. 2023). VP2 is the main component of capsid proteins and exhibits a haemagglutinating activity. The peptide encoded by VP2 can self--assemble into virus-like particles, therefore functioning as an effective antigen transporter that induces strong cellular immunity (Streck and Truyen 2020, Vargas-Bermudez et al. 2023). VP3 is a product of the post-translational cleavage processing of VP2 and appears only after capsid assembly and viral genome packaging (Simpson et al. 2002). The PPV genome encodes three non-structural proteins: NS1, NS2 and NS3 (Streck and Truyen 2020, Vargas-Bermudez et al. 2023). The NS1 protein is the main non-structural protein involved in PPV infection, and its genes are highly conserved (Streck and Truyen 2020). However, the specific function of this encoded protein remains unclear. To date, there are eight known genotypes of PPV that can infect pigs: PPV1-8 (Vargas-Bermudez et al. 2023). PPV1 can cause reproductive disorders in sows that clinically manifest as miscarriages, mummified foetuses and stillbirths, causing substantial economic losses to the pig farming industry (Streck and Truyen 2020). PPV2-8 is a new type of porcine parvovirus that has been reported both domestically and internationally in recent years (Jager et al. 2021). Currently, research on these new types of porcine parvoviruses is focused on epidemiological studies (Li et al. 2021) thus, their pathogenicity and mechanisms remain unknown.

DNA methylation is a form of DNA chemical modification that can alter genetic expression without altering the DNA sequence (Moore et al. 2013, Mattei et al. 2022). Cytosine methylation, which occurs only in eukaryotes, refers to the covalent modification of the methyl group on the 5th carbon atom of cytosine in cytosine-phosphate-guanine (CpG) dinucleotides (under the catalytic action of DNA methyltransferases (DNMTs) to form 5-methylcytosine (5mC)) (Moore et al. 2013, Mattei et al. 2022). DNA methylation often occurs in CpG sequence density-enriched regions; namely, CpG islands (CGIs). CGIs exist in three regions: transcription initiation sites, gene bodies and transposable elements (Moore et al. 2013, Mattei et al. 2022).

DNA methylation plays a crucial role in embryonic and germ cell development (Isagawa et al. 2011). Research has found that early embryonic and germ cell development in mammals involves overall demethylation and re-establishment of methylation maps (Isagawa et al. 2011). The target tissues for PPV infec-

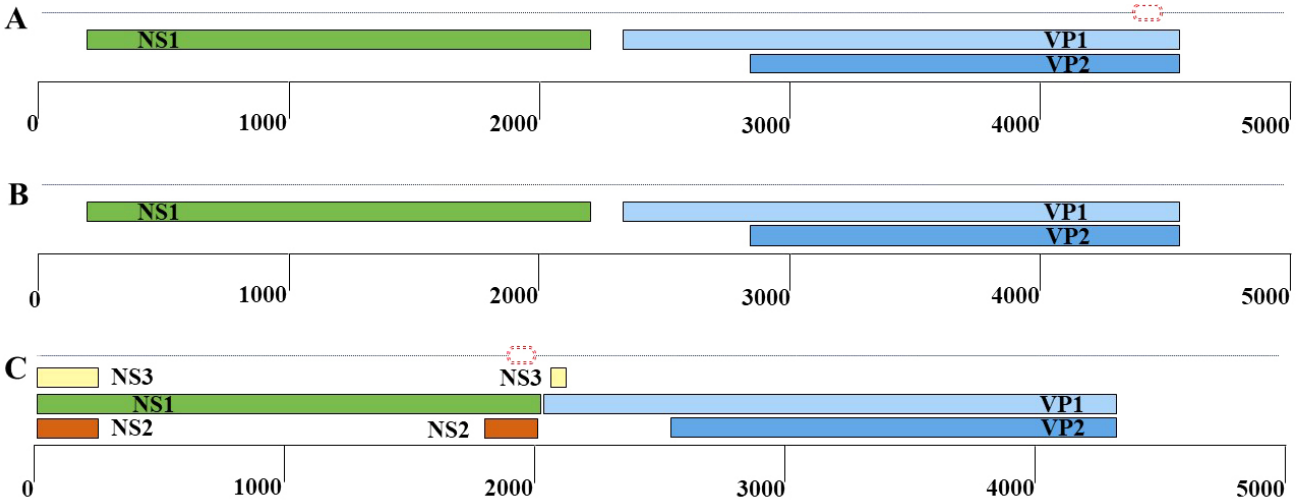

Fig. 1. CpG islands in porcine parvovirus 1 (PPV1) genomes. A. Schematic representation of CpG isolation from the Kresse strain virus. B. Schematic representation of CpG isolation from NADL-2. C. Schematic representation of the CpG isolate in the MZ706996 strain. The red circle represents the CpG island in the genome.

tion are the reproductive organs, causing subsequent reproductive disorders (Vargas-Bermudez et al. 2023). PPV infection in the early stages of pregnancy (30-50 days) causes foetal abortion and sows exhibit recurrent oestrus and infertility, whereas infection during mid-pregnancy (50-60 days) can cause stillbirth, mummified foetuses and malformed foetuses (Joo et al. 1976). Infection at 70 days of pregnancy can lead to miscarriage and stillbirth (Joo et al. 1976). In the late stages of pregnancy (after 70 days), most infected pigs give birth normally; however, the resulting piglets are small and have a high risk of mortality (Joo et al. 1976). Piglets that survive can still carry the virus. Therefore, different stages of infection may result in varying clinical symptoms. In this preliminary study, we explored the biological characteristics of viruses from the perspective of DNA methylation, laying the foundation for correlation research between epigenetics and viral pathogenicity.

Materials and Methods

Sequencing information

PPV1-PPV8 sequences were collected from the GenBank of National Biotechnology Information Center, using the search terms 'Porcine Parvovirus' and 'complete genome'.

CpG island analysis

We conducted a CpG island analysis of the genome sequences of PPV1-PPV8. Two online software packages were used: Meth Primer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) and CpG Plot (www.ebi.ac.uk/Tools/seqstats/embass _cpg) to calculate the CpG islands of the PPV genome sequences for each genotype. CpG islands were calculated by detecting the GC content and observing the expected CpG dinucleotide ratio of a sequence window. The standard settings for CpG islands were as follows: sequence window length should not be <100 bases, GC content should not be <50%, and observation/expected value should not be <60%.

Results

CGI distribution in PPV1

The search results indicated that PPV1 (with a known full-length genome sequence) includes three strains: MZ577027.1, MZ577026.1 and MZ706996.1. Using the MZ706996.1 strain as an example, the total genome length was 4494 bp. The predicted results indicated that the sequence contained a CpG island located between positions 1898 and 1999 with a length of 102 bp (Fig. 1C). In addition, strains with known sequences (including the NADL-8, NADL-2, and Kresse strains) were analysed. The results indicated that the Kresse strain contained a CpG island located at positions 4384-4490 bp, with a length of 106 bp (Fig.1A). The NADL-8 and NADL-2 strains did not contain CGI (Fig.1B).

CGI distribution in PPV2

The analysisshowed that there are a total of 19 strains of PPV2 with known full-length genome sequences (Fig. 2A and Table 1) with a GC content of 44-45% containing 12-16 CGI. Using the MW051675 strain as an example, the total genome length was 5324 bp. The prediction results indicated that the sequence con-

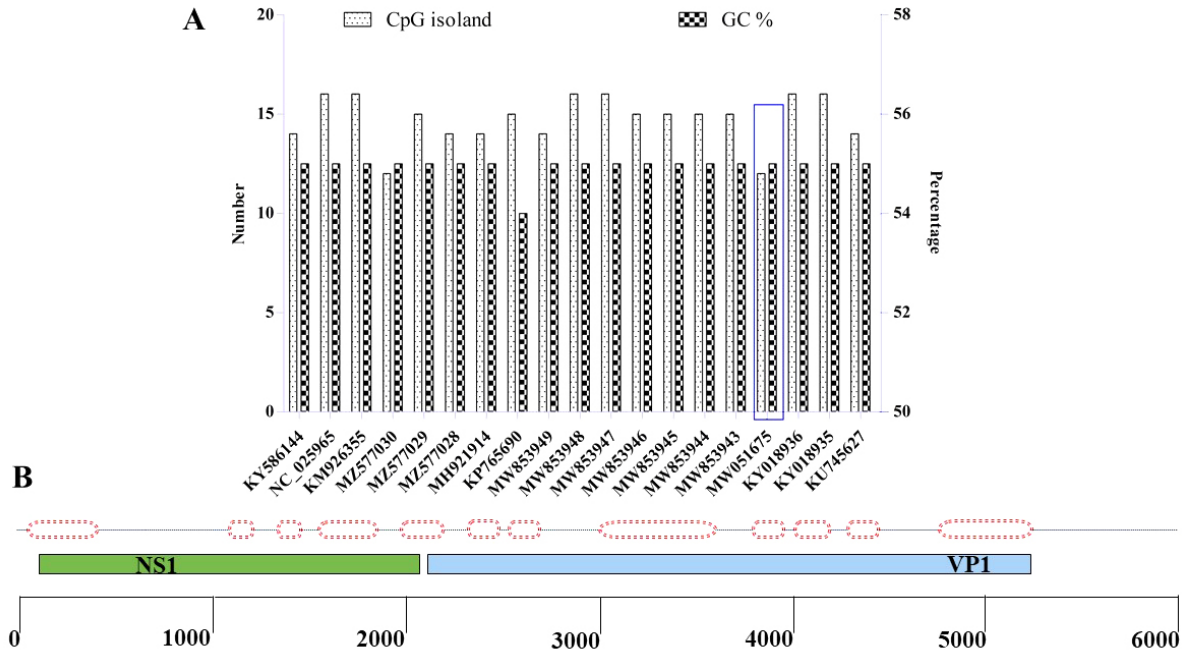

Fig. 2. GC and CpG content of porcine parvovirus 2 (PPV2) genomes. A. The value of GC percentage and CpG island in PPV2 genomes. Left vertical bars represent the value of the CpG island in the genomes. Right vertical bars represent value of GC content. Light blue box indicates the representative strains analysed. B. Schematic representation of the CpG island in the MW051675 strain virus. Red circles symbolize the CpG island in the genomes.

Table 1. Locations and lengths of CpG islands within genome of PPV2.

Accession no.	Length (bp)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)
KY586144	5316	50-329	485-624	1077-1179	1317-1416	1524-1828	1895-2178	2338-2651	2657-2761	2980-3603	3756-3960	4039-4208	4314-4437	4760-4864	4871-5258		
NC_025965	5426	48-531	550-662	918-1023	1357-1506	1609-1911	1971-2274	2422-2564	2628-2846	3064-3348	3401-3564	3571-3720	3766-3921	3924-4029	4126-4326	4400-4530	4821-5370
KM926355	5426	48-531	550-662	918-1023	1357-1506	1609-1911	1971-2274	2422-2564	2628-2846	3064-3348	3401-3564	3571-3720	3766-3921	3924-4029	4126-4326	4400-4530	4820-5369
MZ577030	5426	48-405	581-717	1161-1264	1367-1500	1610-1912	1970-2270	2422-2556	2644-2844	3062-3778					4123-4316	4370-4544	4820-5370
MZ577029	5426	48-413	573-691	915-1024	1161-1263	1334-1500	1608-1915	1970-2262	2417-2559	2576-2847	3064-3566	3570-3720	3795-3936	4127-4301	4403-4517		4954-5370
MZ577028	5426	48-411	570-714	1161-1263	1400-1503	1595-1916	1966-2262	2417-2693	2762-2866	2886-2997	3064-3762	3837-3997	4119-4316	4369-4530			4954-5370
MH921914	5350	49-550	640-778	1220-1322	1430-1559	1654-1969	2031-2329	2481-2621	2800-2907	3123-3407	3460-3624	3854-4054	4182-4375	4425-4603			4879-5293
KP765690	5491	49-459	612-760	1207-1309	1447-1548	1654-1960	2014-2308	2463-2775	2791-2890	3113-3221	3243-3603	3656-3758	3885-4126	4169-4340	4444-4567		4889-5295
MW853949	5388	50-307	1027-1129	1237-1370	1473-1778	1835-2136	2288-2426	2607-2732	2752-2863	2930-3431	3434-3644	3790-3912	3989-4172	4266-4410			4673-5332
MW853948	5454	49-469	637-777	976-1080	1217-1319	1382-1556	1662-1771	1778-1966	2029-2326	2478-2618	2797-2900	3120-3621	3626-3834	3853-4051	4175-4362	4456-4600	4876-5396
MW853947	5754	50-363	373-585	662-799	1004-1103	1245-1348	1487-1588	1692-1999	2051-2347	2505-2776	2848-2951	2971-3082	3149-3863	3922-4080	4175-4401	4454-4629	4911 -5698
MW853946	5488	50-499	688-826	1268-1370	1441-1605	1716-2022	2080-2377	2529-2669	2848-2951	3171-3504	3508-3827	3944-4102	4226-4413	4507-4651	4549-4652	4927-5430	
MW853945	5381	50-335	495-613	837-946	1083-1185	1256-1420	1531-1837	1894-2184	2339-2481	2498-2769	2986-3488	3761-3966	4049-4220	4327-4439	4761-4861	4876-5325	
MW853944	5464	50-510	686-826	1267-1368	1473-1605	1714-2019	2102-2375	2517-2661	2749-2949	3169-3502	3506-3883	3902-4100	4232-4411	4504-4649	4550-4650	4939-5359	
MW853943	5573	49-466	626-744	968-1077	1214-1316	1385-1551	1662-1967	2025-2315	2470-2612	2629-2900	3117-3619	3623-3765	3892-4097	4180-4354	4457-4570	5007-5517	
MW051675	5324	50-435	1096-1198	1273-1444	1544-1848	1910-2208	2355-2499	2563-2781	2999-3657	3859-3981	4058-4249	4341-4479	4757-5266				
KY018936	5350	49-559	977-1083	1214-1321	1397-1557	1666-1970	2055-2329	2482-2623	2687-2905	3124-3456	3460-3624	3627-3777	3824-3980	3983-4105	4182-4371	4464-4603	4887-5293
KY018935	5350	49-559	977-1083	1214-1321	1397-1557	1666-1970	2055-2329	2482-2623	2687-2905	3124-3456	3460-3624	3627-3777	3824-3980	3983-4105	4182-4371	4464-4603	4886-5293
KU745627	5350	49-464	640-774	1220-1323	1450-1559	1668-1970	2034-2329	2476-2612	2800-2903	3121-3840	3856-4054	4178-4373	4459-4603	4879-5115	5123-5293		

tained 12 CGI located at positions 50-435, 1096-1198, 1273-1444, 1544-1848, 1910-2208, 2235-2499, 2563-2781, 2999-3657, 3859-3981, 4058-4249, 4341-4479, and 4757-5266 bp, with lengths ranging from 103 to 659 bp (Fig. 2B).

CGI distribution in PPV3

The results indicated that a total of eight PPV3 had known full-length genome sequence (Fig. 3A and Table 2), with a GC content of 50% containing 12-14 CGI. Taking the KU167029 strain as an example, its genome length was 5081 bp. The prediction results showed that the sequence contained 12 CGI located at positions 49-215, 763-1164, 1358-1481, 1562-1895, 2142-2276, 2526-2641, 2790-3024, 3121-3224, 3276-3378, 3398-4450, and 4760-4935 bp, with lengths ranging from 103 to 523 bp (Fig. 3B).

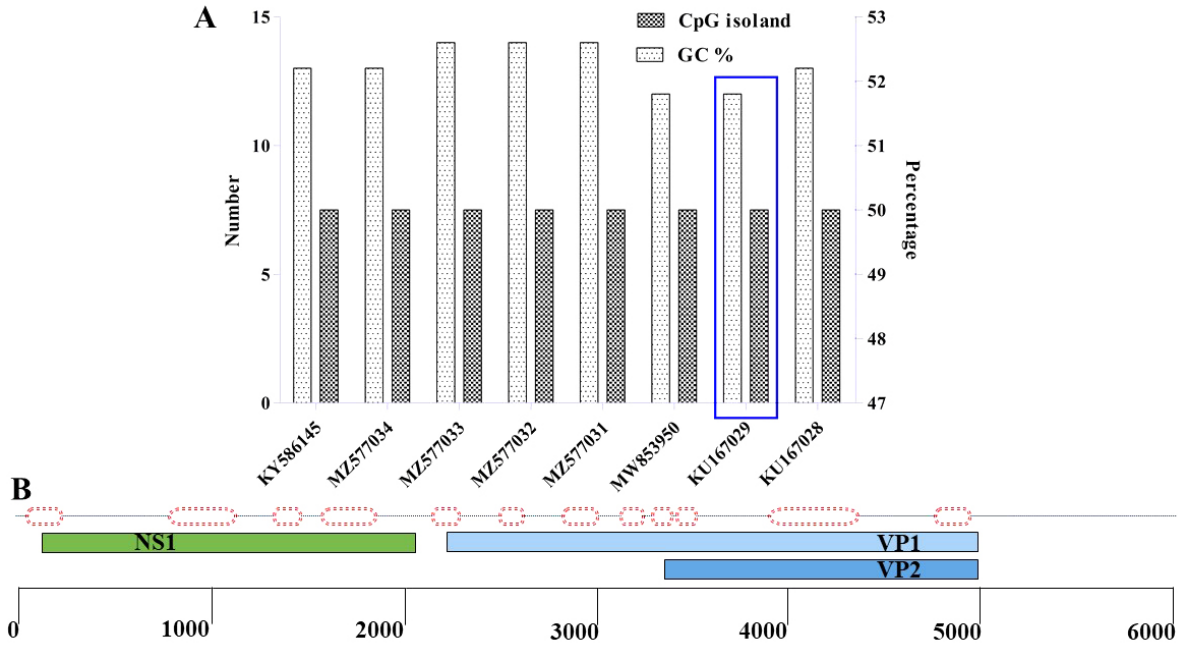

Fig. 3. GC and CpG content of porcine parvovirus 3 (PPV3) genomes. A. The value of GC percentage and CpG islands in PPV3 genomes. Left vertical bars represent the value of CpG islands in the genomes. Right vertical bars represent value of GC content. Light blue box indicates the representative strains analysed. B. Schematic representation of the CpG island in the KU167029 strain virus. Red circles symbolize the CpG island in the genomes.

Table 2. Locations and lengths of CpG islands within genome of PPV3.

Accession no.	Length (bp)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)
KY586145	5113	50-186	783-983	1014-1149	1332-1454	1534-1867	2115-2251	2505-2608	2763-2986	3095-3197	3372-3535	3894-4061	4236-4462	4733-4922	
MZ577034	5070	50-203	795-1017	1022-1161	1346-1470	1551-1884	2113-2267	2779-3012	3111-3213	3265-3367	3387-3548	3917-4133	4254-4439	4754-4927	
MZ577033	5179	49-253	802-1211	1396-1523	1600-1942	1962-2063	2181-2315	2541-2674	2839-3061	3161-3263	3315-3417	3437-3586	3966-4182	4293-4491	4780-4976
MZ577032	5179	49-253	802-1211	1396-1523	1600-1942	1962-2063	2181-2315	2574-2674	2829-3062	3161-3263	3315-3416	3438-3586	3967-4183	4248-4488	4805-4977
MZ577031	5179	49-252	694-794	802-1213	1396-1520	1600-1934	1962-2063	2166-2315	2565-2675	2829-3052	3315-3417	3437-3598	3961-4126	4216-4480	4804-4983
MW853950	4966	650-1060	1244-1369	1447-1782	1810-1911	2014-2163	2421-2523	2621-2900	3163-3265	3285-3446	3809-3974	4064-4328	4652-4831		
KU167029	5081	49-215	763-1164	1358-1481	1562-1895	2142-2276	2526-2641	2790-3024	3121-3224	3276-3378	3398-3547	3928-4450	4760-4935		
KU167028	5081	49-213	763-1172	1371-1481	1562-1895	2129-2276	2526-2635	2791-3024	3121-3224	3276-3377	3399-3547	3928-4144	4165-4450	4761-4937	

CGI distribution in PPV4

A total of 14 PPV4 exhibited known full-length genome sequences (Fig. 4A and Table 3), of which 12 sequences contain 1-3 CGI s with an average GC content of 41%. Taking the MH921911 strain as an example, the total genome length was 5479 bp. The prediction results indicated that there is a CpG island located at positions 593-694 bp, with a length of 102 bp (Fig. 4B).

CGI distribution in PPV5

The retrieval results showed that PPV5 with a known full-length genome sequence included 16 strains (Fig. 5A and Table 4), with a GC content of 39% and 1-3 CGI s. The JX896318 strain was used as an example, the total genome length was 5516 bp, which contained two CGI s, located at positions 49-159 and 3165-3292 bp, respectively, with lengths of 111 and 128 bp, respectively (Fig. 5B).

CGI distribution in PPV6

We identified a total of 38 known full-length genome sequences of PPV6 strains (Fig. 6A and Table 5), with an average GC content of 45-46% and containing 3-5 CGI s. For example, MH447537has a total genome length of 5976 bp. and four CGI s located at positions 826-941, 1721-1821, 2046-2168 and 5036-5161 bp, with lengths ranging from 101 to 126 bp (Fig. 6B).

CGI distribution in PPV7

The search results indicated that PPV7, which has a known full-length genome sequence, included 44 strains (Fig. 7A and Table 6) with a GC content of 53-55% containing 6-11 CGI s. Using the MZ577044

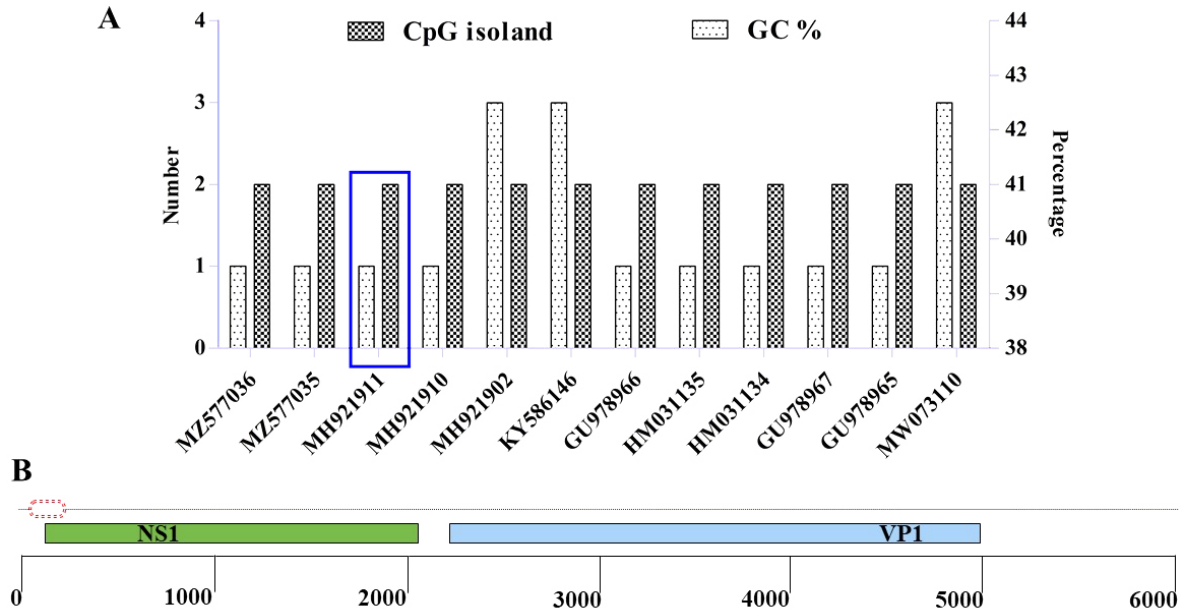

Fig. 4. GC and CpG content of porcine parvovirus 4 (PPV4) genomes. A. The value of GC percentage and CpG islands in PPV4 genomes. Left vertical bars represent the value of CpG islands in the genomes. Right vertical bars represent value of GC content. Light blue box indicates the representative strains analysed. B. Schematic representation of the CpG islands in the MH921911 strain virus. Red circles symbolize CpG islands in the genomes.

Table 3. Locations and lengths of CpG islands within genome of PPV4.

Accession no.	Length (bp)	Location/Length (bps)	Location/Length (bps)	Location/Length (bps)
MZ577036	5414	484-585		
MZ577035	5414	485-585		
MH921911	5479	593-694		
MH921910	5484	596-697		
MH921902	5509	49-393	975-1079	2001-2125
KY586146	5851	50-160	734-834	5669-5794
GU978966	5597	669-769		
HM031135	5552	624-724		
HM031134	5400	485-585		
GU978967	5644	716-816		
GU978965	5644	715-818		
MW073110	5907	49-216	233-392	978-1078

strain as an example, the total genome length was 3999 bp. The prediction results indicated that the sequence contained seven CGI s located at positions 48-178, 389-897, 1158-1899, 1932-2463, 2537-2767, 2819-3393 and 3665-3832 bp, respectively, with lengths ranging from 168 to 742 bp (Fig. 7B).

Discussion

PPV has only one serotype that can be classified into four categories based on pathogenicity and tissue preference: The first type is represented by the NADL-8 strain, which is highly virulent (Meszaros et al. 2017). Oral administration can cross the placental barrier, leading to foetal infection and the formation of viremia. The second type (represented by NADL-2) is an attenuated vaccine strain that cannot cross the placental barrier after oral administration, is non-pathogenic to pregnant sows and foetuses and does not cause viraemia (Jozwik et al. 2009). Therefore, the second type can be used as an attenuated vaccine strain. When the NADL-2 strain is inoculated into the uterus, the virus has the ability to replicate and infect, which can lead to foetal death. The third type (represented by the Kresse strain) is a highly virulent strain of dermatitis that can kill piglets with insufficient immunity (Meszaros et al. 2017). The fourth type is enteritis type strains, which mainly cause intestinal lesions (Csagola et al. 2016).

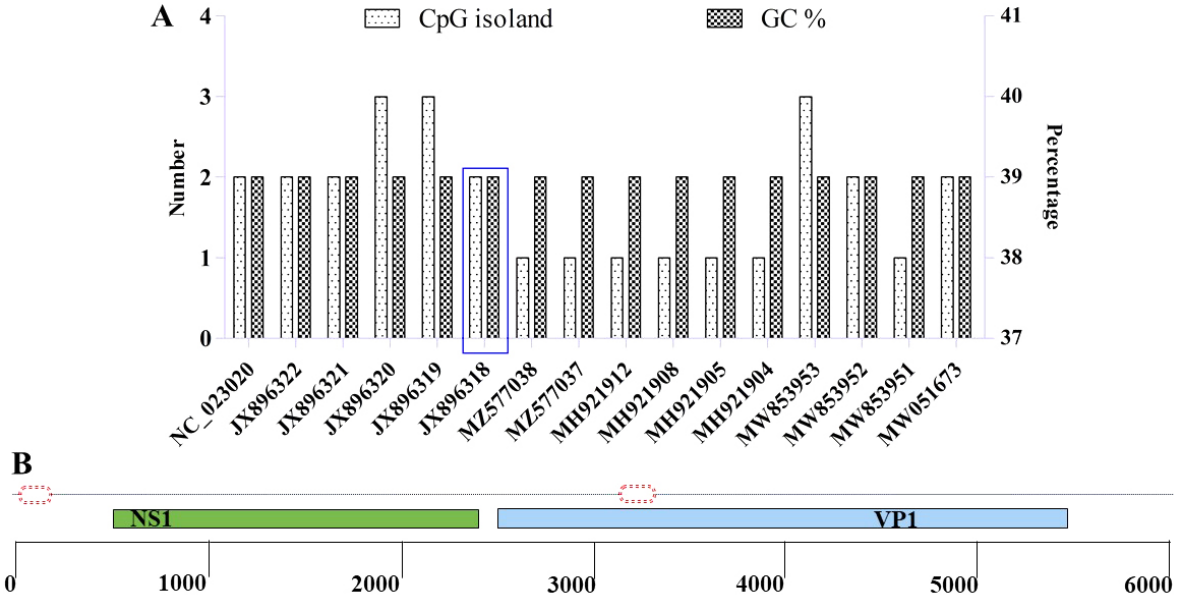

Fig. 5. GC and CpG content of porcine parvovirus 5 (PPV5) genomes. A. The value of GC percentage and CpG islands in PPV5 genomes. Left vertical bars represent the value of CpG islands in the genomes. Right vertical bars represent value of GC content. Light blue box indicates the representative strains analysed. B. Schematic representation of the CpG islands in the JX896318 strain virus. Red circles symbolize CpG islands in the genomes.

Table 4. Locations and lengths of CpG islands within genome of PPV5.

Accession no.	Length (bp)	Location/Length (bps)	Location/Length (bps)	Location/Length (bps)
NC_023020	5805	49-406	3460-3587	
JX896322	5553	49-156	3205-3332	
JX896321	5805	49-406	3460-3587	
JX896320	5756	49-188	194-357	3411-3538
JX896319	5788	49-209	228-386	3440-3567
JX896318	5516	49-159	3165-3292	
MZ577038	5369		3056-3194	
MZ577037	5369		3067-3194	
MH921912	5386		3067-3221	
MH921908	5366		3064-3191	
MH921905	5354	3052-3179		
MH921904	5351	3049-3176		
MW853953	5760	50-157	3211-3338	5576-5703
MW853952	5662	50-157	3211-3338	
MW853951	5408		2916-3043	
MW051673	5584	50-182	3237-3364	

Reports have shown that PPV DNA was found to be methylated independently from its origin (Toth et al. 2013). Like the encapsidated negative strand, the positive strand has been shown to be hypomethylated, thus indicating that PPV DNA maintains hypomethylation throughout the entire lifecycle of the virus, including replication and packaging (Toth et al. 2013). This study showed that the PPV1 strain Kresse contains a CpG isolate, whereas NADL-8 and NADL-2 do not. NADL-8 may be the reason for the short analysis sequence, which was only 3000 bp. The PPV NADL-2 strain contained 60 CpGs. According to previous reports, single nucleotide polymorphisms (SNPs) in the PPV genome were analysed based on available sequences from DNA databases, and the CpG site in the PPV genome was found to be more variable than the GC or C and G sites (Toth et al. 2013). This finding supports the high mutation rate of CpG sites in the PPV genome, which may explain the low number of CGIs.

Previous studies have shown that methylation strat-

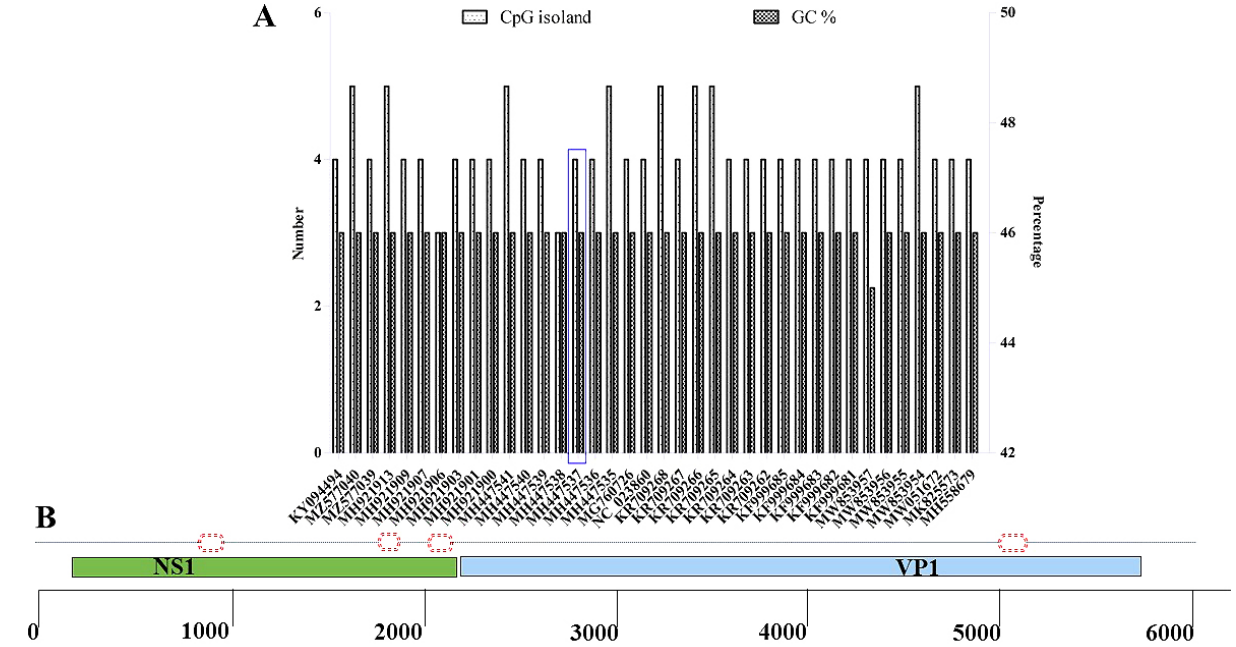

Fig. 6. GC and CpG content of porcine parvovirus 6 (PPV6) genomes. A. The value of GC percentage and CpG islands in PPV6 genomes. Left vertical bars represent the value of CpG islands in the genomes. Right vertical bars represent the GC content values. Light blue box indicates the representative strains analysed. B. Schematic representation of CpG islands in the MH447537 strain virus. Red circles symbolize CpG islands in the genomes.

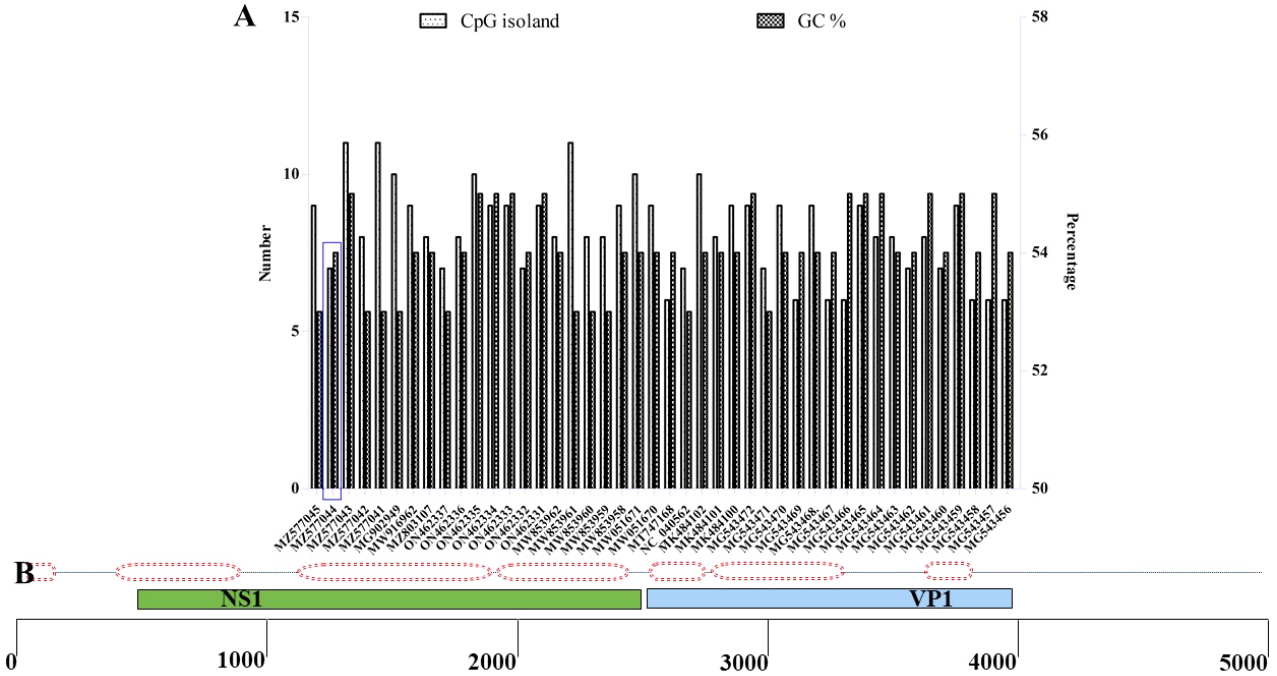

Fig. 7. GC and CpG content of porcine parvovirus 7 (PPV7) genomes. A. The value of GC percentage and CpG islands in PPV7 genomes. Left vertical bars represent the value of CpG islands in the genomes. Right vertical bars represent GC content values. Light blue box indicates the representative strains analysed. B. Schematic representation of the CpG islands in the MZ577044 strain virus. Red circles symbolize CpG islands in the genomes.

egies for combating viruses include methylation of the viral genome and demethylation of the host genome (Gao et al. 2021, Verdikt et al 2022). This study analyses the potential methylation sites of various types of PPV genomes and explores the methylation characteristics of PPV from an epigenetic perspective. The function of methylation is to shut down gene activity, and the function of demethylation is to activate gene expression (Einkauf et al. 2022, Rehman et al. 2023). Once the DNA and RNA of a virus are methylated, their replication, transcription, and translation activities will inevitably decrease (Einkauf et al. 2022, Rehman et al. 2023). The implementation of DNA methylation requires the involvement of DNA methyltransferase

Table 5. Locations and lengths of CpG islands within genome of PPV6.

Accession no.	Length (bp)	Location/Length (bps)	Location/Length (bps)	Location/Length (bps)	Location/Length (bps)	Location/Length (bps)
KY094494	6199	933-1046	1826-1926	2151-2273	5142-5266	
MZ577040	6400	84-251	1096-1209	1989-2089	2314-2436	5305-5429
MZ577039	6199	933-1046	1826-1926	2151-2273	5142-5266	
MH921913	5490	818-933	1713-1813	2038-2160	3254-3374	5029-5153
MH921909	5967	821-939	1719-1819	2044-2166	5035-5159	
MH921907	5974	824-939	1719-1819	2044-2166	5034-5159	
MH921906	5965		1719-1819	2044-2166	5035-5159	
MH921903	5965		1712-1813	2038-2160	5029-5153	5681-5784
MH921901	5994	836-951	1731-1831	2056-2178	5046-5171	
MH921900	5959	818-933	1713-1813	2038-2177	5029-5153	
MH447541	5940	818-933	1713-1813	2038-2160	3254-3374	5029-5153
MH447540	5959	818-933	1713-1813	2038-2177	5029-5153	
MH447539	5996	838-953	1733-1833	2058-2180	5048-5173	
MH447538	5967		1721-1821	2046-2168	5037-5161	
MH447537	5976	826-941	1721-1821	2046-2168	5036-5161	
MH447536	5969	823-941	1721-1821	2046-2168	5037-5161	
MH447535	5979	839-952	1732-1832	2057-2179	5048-5172	5700-5803
MG760726	6123	856-971	1751-1851	2076-2198	5067-5191	
NC_023860	6148	856-971	1751-1851	2076-2198	5066-5191	
KR709268	6148	858-971	1751-1851	2076-2198	3292-3412	5067-5191
KR709267	6117	856-971	1751-1851	2076-2198	5067-5191	
KR709266	6144	858-971	1751-1851	2076-2198	3292-3412	5067-5191
KR709265	6144	858-971	1751-1851	2076-2198	3292-3412	5067-5191
KR709264	6143	858-971	1751-1851	2076-2198	5067-5191	
KR709263	6128	856-971	1751-1851	2076-2198	5066 -5191	
KR709262	6147	858-971	1751-1851	2076-2198	5067-5191	
KF999685	6148	856-971	1751-1851	2076-2198	5066-5191	
KF999684	6163	858-971	2076-2187	4532-4643	5072-5191	
KF999683	6148	858-971	2076-2197	4532-4643	5073-5191	
KF999682	6136	858-972	2076-2187	4532-4643	5072 -5191	
KF999681	6136	858-972	2076-2187	4532-4643	5072-5191	
MW853957	6571	943-1056	1836-1936	2161-2283	5152-5276	
MW853956	6287	65-216	2279-2396	4737-4847	5275-5394	
MW853955	6309	86-237	2300-2417	4758-4868	5296-5415	
MW853954	6328	51-222	1059-1174	1954-2054	2279-2401	5270-5394
MW051672	6382	977-1090	2195-2317	3411-3531	5164-5310	
MK825573	6180	2145-2251	4601-4712	4602-4713	5141-5260	
MH558679	6144	858-971	2076-2198	3292-3412	5068-5191	

(DMT), a methyl donor, and methylation of specific bases (Rehman et al. 2023). On the 35th day of pregnancy, PPV infection leads to embryonic death (Joo et al. 1976). At approximately 70 days of pregnancy, the clinical symptoms of foetal infection are obscure (Joo et al. 1976). Embryonic development undergoes cell reprogramming, allowing an individual cell to transition to a pluripotent embryo or a pluripotent stem cell (Isagawa et al. 2011, Han et al. 2024). During this process, the genomic DNA sequence remains unchanged (Isagawa et al. 2011, Han et al. 2024). Although this process is short, the changes in cell fate and gene expression regulation are very drastic. Early embryos can observe a wide range of highly methylated states

Table 6. Locations and lengths of CpG islands within genome of PPV7.

Accession no.	Length (bp)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)	Location/ Length (bps)
MZ577045	3999	48-202	392-899	1158-1664	1670-1918	1932-2442	2579-2793	2822-3042	3081-3312	3734-3833		
MZ577044	3999	48-178	389-897	1158-1899	1932-2463	2537-2767	2819-3393	3665-3832				
MZ577043	3999	48-246	382-627	640-904	1015-1118	1158-1892	1922-2443	2545-2659	2668-2794	2817-3197	3234-3393	3665-3832
MZ577042	3999	48-311	389-778	786-900	1166-2463	2537-2802	2808-3036	3051-3194	3726-3840			
MZ577041	3996	48-311	392-627	640-778	788-896	1175-1894	1965-2245	2253-2464	2537-2647	2692-2795	2820-3191	3716-3840
MG902949	4013	48-312	393-631	640-899	1162-1663	1676-1910	1942-2443	2538-2792	2815-3008	3048-3190	3730-3846	
MW916962	4041	49-352	421-666	680-937	1198-1931	1968-2451	2576-2834	2862-3074	3092-3226	3758-3907		
MZ803107	3990	48-203	471-612	640-901	1157-2443	2525-2793	2818-2944	3023-3334	3715-3823			
ON462337	3999	48-237	393-897	1175-1924	1929-2463	2524-2797	2816-3193	3723-3844				
ON462336	3995	48-251	389-898	975-1118	1158-2408	2523-2796	2817-2976	3047-3301	3722-3865			
ON462335	3984	48-176	477-898	1014-1117	1158-1911	1971-2463	2539-2777	2829-3047	3056-3168	3184-3368	3705-3805	
ON462334	3999	48-246	382-627	641-902	1015-1116	1158-1266	1268-2463	2537-2795	2822-3194	3723-3839		
ON462333	3999	48-311	382-628	640-902	1015-1116	1158-2442	2524-2792	2811-3195	3236-3393	3640-3843		
ON462332	3999	48-246	382-897	1185-1899	1932-2447	2531-2802	2806-3241	3716-3833				
ON462331	3999	48-246	382-613	649-902	994-1116	1159-1909	1941-2408	2523-2778	2823-3399	3665-3833		
MW853962	4143	51-321	409-511	537-1134	1300-2588	2669-2942	2961-3181	3229-3337	3871-3991			
MW853961	4161	51-316	530-1054	1059-1188	1300-2037	2082-2601	2675-2782	2790-2932	2961-3098	3209-3433	3438-3626	3856-3961
MW853960	4033	48-198	390-1003	1208-1910	1964-2466	2538-2792	2815-2974	2997-3190	3730-3843			
MW853959	4331	183-497	689-1302	1507-2217	2265-2765	2837-3091	3114-3273	3296-3488	4029-4142			
MW853958	4003	48-240	381-621	641-1117	1158-1909	1942-2463	2586-2795	2819-2979	3058-3196	3726-3840		
MW051671	4003	48-311	382-621	642-897	1156-1909	1941-2443	2537-2661	2823-2977	3053-3242	3351-3494	3723-3829	
MW051670	3955	48-175	392-902	1207-2445	2523-2622	2652-2784	2828-3014	3071-3293	3336-3472	3718-3823		
MT747168	3682	51-549	857-2113	2175-2444	2460-2665	2716-3138	3354-3465					
NC_040562	3999	48-242	389-894	1161-2444	2682-2795	2810-2979	3087-3188	3729-3829				
MK484102	3999	48-226	392-778	786-897	966-1072	1160-1918	1930-2463	2537-2792	2823-2954	3049-3194	3726-3832	
MK484101	3999	48-226	388-621	642-901	1162-1918	1932-2443	2539-2798	2826-3188	3665-3842			
MK484100	3999	48-226	392-778	786-897	966-1072	1160-1918	1932-2443	2537-3040	3049-3194	3726-3832		
MG543472	3999	48-226	389-894	1014-1118	1161-2465	2537-2762	2810-2973	3109-3209	3290-3393	3665-3832		
MG543471	3999	48-242	389-894	1161-2444	2682-2795	2810-2979	3087-3188	3729-3829				
MG543470	3984	48-241	389-897	1014-1118	1161-2443	2537-2644	2664-3014	3069-3194	3275-3378	3650-3817		
MG543469	3999	48-176	393-897	1014-1117	1162-2443	2537-3193	3207-3508					
MG543468	3999	48-226	389-778	792-897	1014-1118	1161-2443	2537-2796	2814-3075	3078-3193	3207-3337		
MG543467	3987	48-226	389-898	1157-2443	2537-2978	3079-3181	3195-3325					
MG543466	3999	48-226	389-898	1157-2465	2537-2793	2823-3036	3719-3829					
MG543465	3999	48-308	393-898	1157-2463	2537-2674	2682-2795	2810-2979	3083-3209	3290-3393	3665-3832		
MG543464	3999	48-226	389-629	639-779	784-899	1175-2443	2537-3209	3290-3393	3665-3832			
MG543463	3999	48-226	389-629	639-779	784-899	1175-2443	2537-3209	3290-3393	3665-3832			
MG543462	3999	48-176	389-898	1157-2443	2537-2796	2814-3075	3078-3209	3686-3831				
MG543461	3999	48-308	393-898	1157-2463	2537-2674	2682-2795	2810-2979	3083-3196	3725-3832			
MG543460	3987	48-226	389-899	1183-2463	2537-2978	3095-3197	3278-3381	3653-3820				
MG543459	3999	48-226	389-627	640-776	788-899	1206-2463	2537-2662	2667-2795	2816-3196	3725-3829		
MG543458	3999	48-226	389-897	1158-2443	2537-3209	3290-3393	3665-3832					
MG543457	3999	48-226	389-898	1159-2443	2537-3209	3290-3393	3665-3832					
MG543456	3999	48-226	389-898	1158-2443	2537-3209	3290-3393	3665-3832					

in both transcriptional and non-transcriptional regions (Han et al. 2024). And there may not be excess, readily available methyl groups to methylate the invading viral DNA. This may be the molecular biological basis for the rapid development of early embryos into death, which is not closely related to the type of virus. Can future corresponding viral therapies utilize this mechanism, especially viral DNA methylation, to suppress viral pathogenicity? At least this is a direction worth further research.

This study extended the analysis of the number and distribution of CpG islands in the PPV2-PPV7 genome (no data available for PPV8). The results indicated differences in the number and distribution of CpG islands. Research has shown that methylation of CpG islands can impair transcription factor binding, recruit inhibitory methyl-binding proteins and stably silence gene expression (Modutlwa et al. 2009). Although the PPV1 genome maintains low methylation independent of the tissue of origin throughout the virus lifecycle, further research is needed to investigate the methylation status of the PPV2-PPV7 genome and determine the extent to which DNA methylation of CpG islands regulates gene expression. Other studies have demonstrated that CpG islands, particularly those associated with gene promoters, are rarely methylated (Zhao et al. 2009). In addition, previous research has found that. The NADL-2 strain of PPV contains 60 CpGs but does not have a CpG island (Toth et al. 2013). Toll-like receptor 9 (TLR9) is a member of the Toll-like receptor family, which can be activated by non-methylated cytidine monophosphate guanosine DNA (CpG DNA) derived from pathogens or artificially synthesised oligonucleotides containing non-methylated CpG (CpG ODN) and directly or indirectly initiates innate immune responses through downstream signalling, thereby resisting pathogen invasion (Kim et al. 2013). Recent studies have shown that parvoviruses do not induce immune responses activated by TLR9 (Mattei et al. 2013). Therefore, further clarification is needed to determine whether methylation occurs at the CpG site outside the CpG island in the PPV genome and its role.

Acknowledgements

This study was supported by grants from the Nanchong Vocational and Technical College for basic scientific research (no. ZRA1904 and no. NZYBZ2002), the Nanchong High-Level Talents Special Support Program (Scientific Elite Of Ten-Thousands Talents Program), the Science and Technology Planning Project of Nanchong (Grant No. 22XCZX0003) and the Chongqing Three Gorges Vocational College for basic scientific research (no. SXZY-2023TD-005).

References

Cartwright SF, Lucas M, Huck RA (**1969**) A small haemagglutinating porcine DNA virus. I. Isolation and properties. J Comp Pathol 79: 371-377.

Chen Y, Luo S, Tan J, Zhang L, Qiu S, Hao Z, Wang N, Deng Z, Wang A, Yang Q, Yang Y, Wang C, Zhan Y (**2023**) Establishment and application of multiplex real-time PCR for simultaneous detection of four viruses associated with porcine reproductive failure. Front Microbiol 14: 1092273.

Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger AM, Hughes J, Mietzsch M, Modha S, Ogliastro M, Penzes JJ, Pintel DJ, Qiu J, Soderlund-Venermo M, Tattersall P, Tijssen P, Consortium IR (**2019**) ICTV virus taxonomy profile: Parvoviridae. J Gen Virol 100: 367-368.

Csagola A, Zadori Z, Meszaros I, Tuboly T (**2016**) Detection of porcine parvovirus 2 (ungulate tetraparvovirus 3) specific antibodies and examination of the serological profile of an infected swine herd. PloS One 11: 0151036.

Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, Parsons EM, Gladkov GT, Seiger KW, Blackmer JE, Jiang C, Yukl SA, Rosenberg ES, Yu XG, Lichterfeld M (**2022**) Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell 185: 266-282.

Faustini G, Tucciarone CM, Franzo G, Donneschi A, Boniotti MB, Alborali GL, Drigo M (**2024**) Molecular survey on porcine parvoviruses (PPV1-7) and their association with major pathogens in reproductive failure outbreaks in Northern Italy. Viruses 16: 157.

Gao ZJ, Li WP, Mao XT, Huang T, Wang HL, Li YN, Liu BQ, Zhong JY, Renjie C, Jin J, Li YY (**2021**) Single-nucleotide methylation specifically represses type I interferon in antiviral innate immunity. J Exp Med. 218: 20201798.

Han Q, Ma R, Liu N (**2024**) Epigenetic reprogramming in the transition from pluripotency to totipotency. J Cell Physiol 239: 31222.

Isagawa T, Nagae G, Shiraki N, Fujita T, Sato N, Ishikawa S, Kume S, Aburatani H (**2011**) DNA methylation profiling of embryonic stem cell differentiation into the three germ layers. PloS One 6: 26052.

Jager MC, Tomlinson JE, Lopez-Astacio RA, Parrish CR, Van de Walle GR (**2021**) Small but mighty: old and new parvoviruses of veterinary significance. Virol J 18: 210.

Joo HS, Donaldson-Wood CR, Johnson RH (**1976**) Observations on the pathogenesis of porcine parvovirus infection. Arch Virol 51: 123-129.

Jozwik A, Manteufel J, Selbitz HJ, Truyen U (**2009**) Vaccination against porcine parvovirus protects against disease, but does not prevent infection and virus shedding after challenge infection with a heterologous virus strain. J Gen Virol 90: 2437-2441.

Kim IY, Yan X, Tohme S, Ahmed A, Cordon-Cardo C, Shantha Kumara HM, Kim SK, Whelan RL (**2012**) CpG ODN, Toll like receptor (TLR)-9 agonist, inhibits metastatic colon adenocarcinoma in a murine hepatic tumor model. J Surg Res 174: 284-290.

Li J, Xiao Y, Qiu M, Li X, Li S, Lin H, Li X, Zhu J, Chen N (**2021**) A systematic investigation unveils high coinfection status of porcine parvovirus types 1 through 7 in China from 2016 to 2020. Microbiol Spectr 9: 0129421.

Mattei AL, Bailly N, Meissner A (**2022**) DNA methylation: a historical perspective. Trends Genet 38: 676-707.

Mattei LM, Cotmore SF, Li L, Tattersall P, Iwasaki A (**2013**) Toll-like receptor 9 in plasmacytoid dendritic cells fails to detect parvoviruses. J Virol 87: 3605-3608.

Mayr A, Bachmann PA, Siegl G, Mahnel H, Sheffy BE (**1968**) Characterization of a small porcine DNA virus. Arch Gesamte Virusforsch 25: 38-51.

Meszaros I, Olasz F, Csagola A, Tijssen P, Zadori Z (**2017**) Biology of porcine parvovirus (ungulate parvovirus 1). Viruses 9: 393.

Modutlwa N, Tada H, Sugahara Y, Shiraki K, Hara N, Deyashiki Y, Ando T, Maegawa T, Monguchi Y, Sajiki H (**2009**) Synthesis of deuterated benzyladenine and its application as a surrogate. Nucleic Acids Symp Ser (Oxf): 105-106.

Moore LD, Le T, Fan G (**2013**) DNA methylation and its basic function. Neuropsychopharmacology 38: 23-38.

Penzes JJ, Soderlund-Venermo M, Canuti M, Eis-Hubinger AM, Hughes J, Cotmore SF, Harrach B (**2020**) Reorganizing the family Parvoviridae: a revised taxonomy independent of the canonical approach based on host association. Arch Virol 165: 2133-2146.

Ranz AI, Manclus JJ, Diaz-Aroca E, Casal JI (**1989**) Porcine parvovirus: DNA sequence and genome organization. J Gen Virol 70: 2541-2553.

Rehman UU, Ghafoor D, Ullah A, Ahmad R, Hanif S (**2023**) Epigenetics regulation during virus-host interaction and their effects on the virus and host cell. Microb Pathog 182: 106271.

Simpson AA, Hebert B, Sullivan GM, Parrish CR, Zadori Z, Tijssen P, Rossmann MG (**2002**) The structure of porcine parvovirus: comparison with related viruses. J Mol Biol 315: 1189-1198.

Streck AF, Lopes TS, Lunge VR (**2024**) Antiviral alternatives against important members of the subfamily Parvovirinae: a review. Arch Virol 169: 52.

Streck AF, Truyen U (**2020**) Porcine Parvovirus. Curr Issues Mol Biol 37: 33-46.

Toth R, Meszaros I, Stefancsik R, Bartha D, Balint A, Zadori Z (**2013**) CpG distribution and methylation pattern in porcine parvovirus. PloS One 8: 85986.

Vargas-Bermudez DS, Mogollon JD, Franco-Rodriguez C, Jaime J (**2023**) The novel porcine parvoviruses: Current state of knowledge and their possible implications in clinical syndromes in pigs. Viruses 15: 2398.

Verdikt R, Bendoumou M, Bouchat S, Nestola L, Pasternak AO, Darcis G, Avettand-Fenoel V, Vanhulle C, Aït-Ammar A, Santangelo M, Plant E, Douce VL, Delacourt N, Cicilionytė A, Necsoi C, Corazza F, Passaes CPB, Schwartz C, Bizet M, Fuks F, Sáez-Cirión A, Rouzioux C, De Wit S, Berkhout B, Gautier V, Rohr O, Van Lint C (**2022**) Novel role of UHRF1 in the epigenetic repression of the latent HIV-1. EBioMedicine 79: 103985.

Zhao Z, Han L (**2009**) CpG islands: algorithms and applications in methylation studies. Biochem Biophys Res Commun 382: 643-645.