

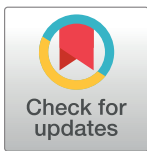
RESEARCH ARTICLE

# Homologous recombination is a force in the evolution of canine distemper virus

Chaowen Yuan<sup>1</sup>, Wenxin Liu<sup>2</sup>, Yingbo Wang<sup>1</sup>, Jinlong Hou<sup>1</sup>, Liguozhang<sup>3</sup>, Guoqing Wang<sup>1\*</sup>

**1** College of life and health sciences, Northeastern University, Shenyang, Liaoning, China, **2** Laboratory of Hematology, Affiliated hospital of Guangdong Medical College, Zhanjiang, Guangzhou, China, **3** Center for Animal Disease Emergency of Liaoning province, Shenyang, Liaoning, China

\* [wanguoqing@mail.neu.edu.cn](mailto:wanguoqing@mail.neu.edu.cn)



## Abstract

Canine distemper virus (CDV) is the causative agent of canine distemper (CD) that is a highly contagious, lethal, multisystemic viral disease of receptive carnivores. The prevalence of CDV is a major concern in susceptible animals. Presently, it is unclear whether intragenic recombination can contribute to gene mutations and segment reassortment in the virus. In this study, 25 full-length CDV genome sequences were subjected to phylogenetic and recombinational analyses. The results of phylogenetic analysis, intragenic recombination, and nucleotide selection pressure indicated that mutation and recombination occurred in the six individual genes segment (H, F, P, N, L, M) of the CDV genome. The analysis also revealed pronounced genetic diversity in the CDV genome according to the geographically distinct lineages (genotypes), namely Asia-1, Asia-2, Asia-3, Europe, America-1, and America-2. The six recombination events were detected using SimPlot and RDP programs. The analysis of selection pressure demonstrated that a majority of the nucleotides in the CDV individual gene were under negative selection. Collectively, these data suggested that homologous recombination acts as a key force driving the genetic diversity and evolution of canine distemper virus.

## OPEN ACCESS

**Citation:** Yuan C, Liu W, Wang Y, Hou J, Zhang L, Wang G (2017) Homologous recombination is a force in the evolution of canine distemper virus. PLoS ONE 12(4): e0175416. <https://doi.org/10.1371/journal.pone.0175416>

**Editor:** Dongsheng Zhou, Beijing Institute of Microbiology and Epidemiology, CHINA

**Received:** January 6, 2017

**Accepted:** March 24, 2017

**Published:** April 10, 2017

**Copyright:** © 2017 Yuan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was financed by the Natural Science Foundation of China (No.31272615). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Canine distemper (CD) is caused by the canine distemper virus (CDV) that belongs to the Morbillivirus genus of the Paramyxoviridae family. The length of the CDV genome is 15,690 nucleotides (nt), and it possesses a single-stranded negative RNA encoding six nonoverlapping transcriptional units producing eight proteins [1, 2]. The genomic RNA that is tightly encapsulated by the nucleocapsid (N) protein serves as a template for transcription and replication by the viral polymerase (L) protein and its cofactor phosphoprotein (P). The N, P, and L proteins together with the viral RNA constitute the ribonucleoprotein (RNP) [3], which directs the sequential synthesis of capped and polyadenylated mRNAs from six transcriptional units or the replication of full-length encapsulated antigenomes [4]. The viral envelope contains two integral membrane proteins, fusion (F) and hemagglutinin (H), as well as, a membrane-associated protein [matrix (M)] that altogether establish the contact with the RNP [5]. The H glycoprotein facilitates the binding of the virus to

the cell membrane; the F protein accomplishes the fusion of the two membranes, which enables the entry of the viral RNP into the cytoplasm [6].

Canine distemper is a highly contagious and lethal disease that affects several many carnivorous species. This virus infects a wide variety of animal families, such as Mustelidae (ferrets, minks, skunks, weasels, and badgers), *Procyonidae* (raccoons), Ursidae (bears and pandas), *Viverridae* (civets, genets, and linsangs), Hyaenidae (hyenas), and Felidae (lions and tigers). The genome of CDV encodes the genes for the following components: matrix (M = 1008 bp), fusion (F = 1989 bp), hemagglutinin (H = 1824 bp), nucleocapsid (N = 1572 bp), polymerase (L = 6555 bp), and phosphoprotein (P = 1524 bp).

Recombination could serve as a critical process for the evolution of RNA viruses [7]. To date, studies regarding the recombination in the CDV genome are not available. Thus, in the present study, twenty-five full-length CDV genome sequences were subjected to phylogeny, recombination, and selection pressure analyses. As a consequence, six such CDV phylogroups were identified among which the homologous recombination occurred among, and the individual genes mentioned above were under negative selection pressure. These data indicated that one recombination and reassortment might contribute to the molecular diversity of CDV.

## Materials and methods

### Sequence data and alignment

The complete sequences of the CDV genome and its six genes (N, P, L, M, H, and F) were retrieved from GenBank. All sequences were aligned using ClustalW [8] and BioEdit [9], followed by visual confirmation. The virus name, GenBank accession number, isolation time, and place are summarized in [Table 1](#).

### Phylogenetic analysis

The gene sequences mentioned above were trimmed by MEGA V5.0 [10]. Seven transversional models with a proportion of invariable sites and a substitution model comprised of gamma-shaped distribution of rates across these sites (CDV genome = GTR+G+I, H = GTR+G, N = GTR+G, P = TrN+G, F = TIM+G, M = TrN+G, L = GTR+G+I) were assessed by ModelTest v3.7 [11], PAUP\* v4b10 (Swofford, 2003), and MrBayes v3.1.1 that explored the distance (neighbor-joining) and characteristics [Bayesian, maximum likelihood (ML)-based phylogenetic methods] of the nucleotides sequences [12]. The results of the analysis were confirmed using Bayesian and ML approaches with the MEGA 5.0 software package [10]. The phylogenetic tree was tested by bootstrapping with 1000 replicates. The phocine distemper virus (PDV) (GenBank accession number KC802221) was indicated as an outgroup in the seven models.

### Recombination analysis

The recombination events were assessed using RDP4.0 [13] and confirmed by Similarity Plot, Boot Scan, and Find Sites (using maximization of  $\chi^2$ ) implementing the sub-programs of SimPlot v3.5.1 [14] or identified by BLAST in GenBank. After the CDV genomes, viral H, N, P, M, F, and L genes were aligned using ClustalW [8]. The Phi test in SplitsTree 4.11 [15] provided statistically significant evidence for recombination. Subsequently, the ML trees were constructed by MEGA 5.0, and bootstrap (1000 replicates) values were shown. Shimodaira–Hasegawa (SH) test affirmed the statistical difference between the phylogenetic trees estimated from different regions. The statistical significance was set at  $p < 0.05$ . All the analyses were conducted in triplicate [16].

**Table 1. CDV genome sequences used in this study.**

GenBank accession no.	Strain	Host	Country	Years
AY466011	98–2654	raccoon	USA	1998
AY443350	00–2601	raccoon	USA	2000
JX681125	HLJ1-06	fox	China	2006
AY542312	98–2646	raccoon	USA	1998
AY445077	98–2645	raccoon	USA	1998
EU716337	164071	<i>Canis familiaris</i>	USA	2004
AB687721	CYN07-hV	<i>Macaca fascicularis</i>	Japan	2008
JN896987	VR-1587	NA	NA	NA
AY649446	01–2689	raccoon	USA	2001
AY386315	5804	NA	USA	NA
AY386316	5804P	NA	USA	NA
AB687720	CYN07-dV	<i>Macaca fascicularis</i>	Japan	2008
KC427278	Hebei	mink	China	2008
HM852904	MKY-KM08	<i>Macaca mulatta</i>	China	2008
HM063009	Shuskiy	mink	Kazakhstan	1989
HM046486	Phoca/Caspian/2007	seal	Kazakhstan:	2007
GU138403	recombinant Snyder Hill	ferret	USA	2010
AF164967	A75/17	NA	USA	1999
EU726268	CDV3	mink	China	NA
AB475097	M25CR	NA	Japan	NA
AB462810	007Lm-1vp	Vero cells	Japan	NA
AB474397	007Lm	NA	Japan	NA
AB476401	011C	<i>Canis lupus familiaris</i>	Japan	NA
AB476402	50Con	<i>Canis lupus familiaris</i>	Japan	NA
AB475099	55L	<i>Canis lupus familiaris</i>	Japan	NA
KC802221	PDV/Wadden_Sea.NLD/1988	<i>Phoca vitulina</i>	Netherlands	1988

<https://doi.org/10.1371/journal.pone.0175416.t001>

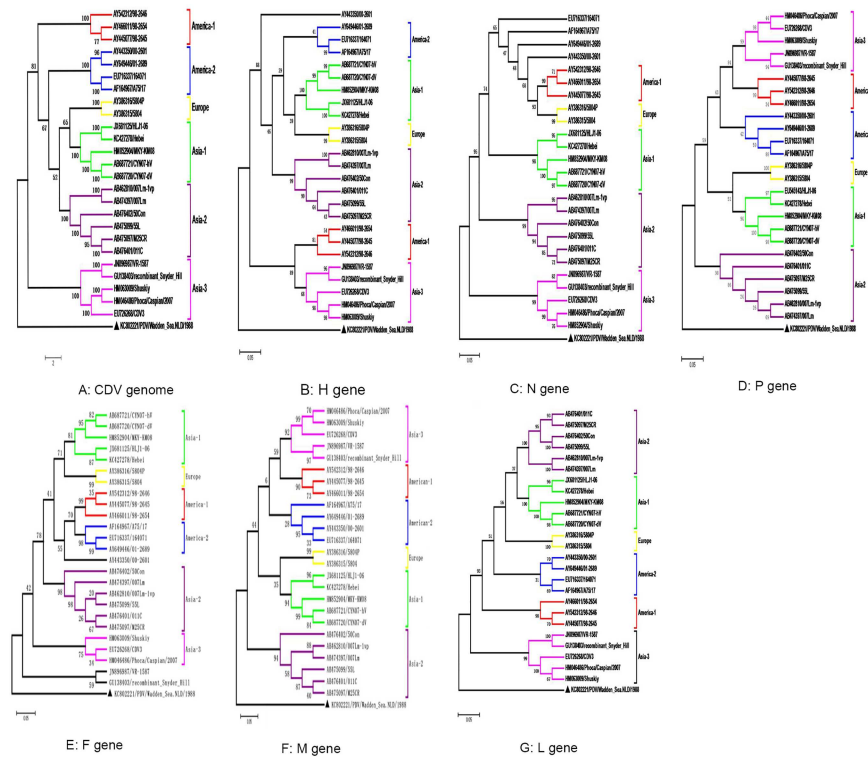
## Selection pressure analysis

The non-neutral selection was calculated by the ratio of nonsynonymous (dN) to synonymous (dS) substitutions using ML phylogenetic reconstruction and the general reversible nucleotide substitution model available through the Datamonkey web server [17]. To detect non-neutral selection, single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), internal fixed-effects likelihood (IFEL), and random effects likelihood (REL) within the HyPhy software package [18] were implemented in the Datamonkey. The significance levels were set at  $p = 0.1$ ,  $p = 0.1$ ,  $p = 0.1$ , respectively, and Bayes factor = 50 was used to estimate the rates of dN and dS within each codon. The values  $dN/dS > 1$ ,  $dN/dS = 1$  and  $dN/dS < 1$  were used to define the positive selection (adaptive molecular evolution), neutral mutations, and negative selection (purifying selection), respectively.

## Results

### Phylogenetic relationships of CDVs

All of the 25 full-length CDV strains originated from China, USA, or Japan, except 2 strains that were isolated from Kazakhstan. (Table 1). To determine the genetic relationships among the CDVs, seven phylogenetic trees were constructed (Fig 1) based on the available N, H, P, M, F, L segments and CDV genome sequences. The results showed that the CDVs could be



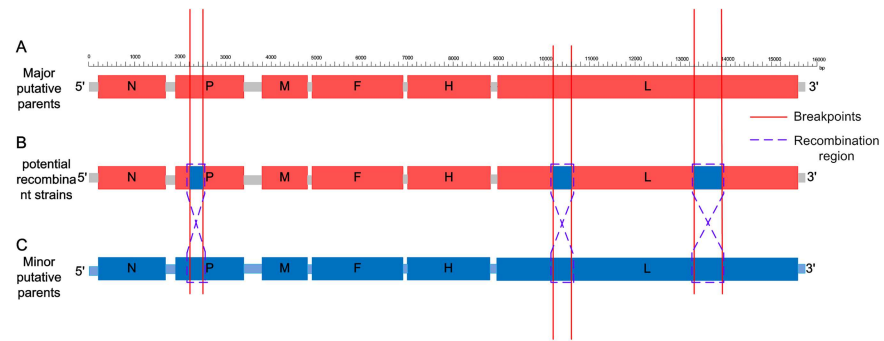
**Fig 1. Phylogenetic tree analysis of 25 full-length CDV genomes.** The distances were calculated by the ClustalW program with the MEGA 5.0 software package. The ML algorithm was used to generate the tree. Bootstrap values were calculated based on 1000 replicates. The result contains six groups: Asia-1, Asia-2, Asia-3, Europe, America-1, and America-2. Triangle (▲) indicates the out-group strains analyzed in this study.

<https://doi.org/10.1371/journal.pone.0175416.g001>

divided into six lineages: Asia-1, Asia-2, Asia-3, Europe, America-1, and America-2. However, in the phylogenetic tree of the N gene, we can only find five lineages including America-1, Asia-1, Asia-2, and Asia-3. We also observed that one strain (AY443350/00-2601) in the H gene and two strains (GU138403/recombinant-Snyder-Hill, JN896987/VR-1587) in the F gene were independent.

### Recombination analysis of CDV using RDP4 and SimPlot programs

In order to seek evidence for the natural recombination, the CDV complete genome sequence alignment dataset was analyzed by the RDP4 software package, which scanned the recombinant sequence. The six CDV isolates exhibited robust recombination signals. The p-values demonstrating the statistical significance of these programs were shown in [S1 Table](#). The breakpoints were also confirmed by the RDP4 program [13]. Each of the recombination isolate encompassed two strains, which were suggested as the representatives of its putative parental lineages (Fig 2). To further validate the authenticity of the recombination events, SimPlot was employed that exhibited a result similar to that of the RDP4 program except for the breakpoint. The crossover sites depend on the size of the sliding window in SimPlot, and hence, the recombination breakpoints in RDP4 and SimPlot are different. The gold-standard bioinformatic approach for demonstrating the presence of recombination is a set of statistically incongruent phylogenetic trees [19]; thus, phylogenetic trees were also constructed using MEGA 5.0 to further determine the recombination events. Eventually, six groups of potential recombination events have been determined (Table 2).



**Fig 2. A schematic map of the genome for the recombinant strains of CDV.** (A) Major putative parents for recombinant strains. (B) Potential recombinant strains of CDV. (C) Minor putative parents for recombinant strains. The same breakpoints in different genome schematics were labeled with a red line, and the recombinant fragments were labeled with a blue dotted line. The x-axis indicated the nucleotide position along the alignment (gap 100 bp).

<https://doi.org/10.1371/journal.pone.0175416.g002>

**One intragenic recombinant was identified in CDV strain HLJ1-06 and its L gene.** The analyses using RDP and SimPlot programs showed that the first potential recombinant event occurred in the strain (JX681125/HLJ1-06). The two strains (KC427278/Hebei and HM046486/Phoca/Caspian/2007) isolated from China and Kazakhstan might be the major and minor putative parents, respectively, and the breakpoint regions of the potential recombinant strain, JX681125/HLJ1-06, were localized from 12771–14312 (in SimPlot) and 12922–14259 (in RDP4) bp. To further analyze the identified recombination event, the strain JX681125/HLJ1-06 was utilized as the query in the SimPlot program (Fig 3A). One potential breakpoint was localized at the parsimonious region, 12771–14312 bp, as identified by the maximization of  $X^2$ . The sequence of HLJ1-06 showed greater affinity with one putative parent lineage of Hebei position 1–12770 and 14313–15690 bp than the other putative parent Phoca/Caspian/2007. However, the sequence of Phoca/Caspian/2007 from 12771–14312 bp shared a greater similarity with HLJ1-06 than Hebei. An identical evidence was displayed by the Boot Scan result (Fig 3B). Owing to the difference in the breakpoint regions in SimPlot and RDP, we selected the total partial 12922–14259 bp region for analysis. The recombination events were further substantiated by constructing the phylogenetic tree using MEGA 5.0. In the putative non-recombinant regions (positions 1–12921 + 14260–15690), HLJ1-06 and Hebei were clustered into the same sub-lineage (Fig 3C), while Phoca/Caspian/2007 was grouped into a distinct sub-lineage. However, in the putative recombinant regions, 12922–14259 bp, the arrangement of the phylogenetic tree reflecting the relationship of the three isolates contradicted the previous tree (Fig 3D). The topology of the two phylogenetic trees around the breakpoint showed a significant statistical discrepancy when the PDV was included in the analyzed data (Shimodaira–Hasegawa test,  $p < 0.001$ ) constituting a robust evidence for recombination.

In order to identify the gene and the portion of the segment that leads to the potential recombinant event, all the six individual genes of CDV were analyzed using RDP4, SimPlot, and MEGA 5.0 programs. The putative recombinant events occurring in the L gene were similar to that described in the CDV genome. The two potential recombinant events exhibited a high similarity among the recombinant strains, major and minor putative parents, except the breakpoints. The breakpoints in the L gene were located from 3743–5297 (in SimPlot) and 3749–5299 (in RDP4). We also analyzed this potential recombinant event with the method used for the CDV genome strain (JX681125/HLJ1-06) (Fig 3E–3H). Collectively, these analyses provided significant evidence for the recombination event described above.

**Table 2. CDV strains with evidence for potential recombination.**

Gene	Event	GenBank accession no.	Strain	Putative parent lineage (s) <sup>a</sup>	SimPlot identified breakpoint regions		RDP identified breakpoint regions	Final identified breakpoint regions	Country
Genome	I	JX681125	HLJ1-06	KC427278/HeBei; HM046486/Phoca/Caspian/2007	12771–14312		12922–14259	12922–14259	China
	II	AY466011	98–2654	EU726268/CDV3; EU716337/164071	2015–2833	7337–9131	2134–2504	2134–2504	USA
	III	AY542312	98–2646	EU726268/CDV3; EU716337/164071	2015–2833	7337–9131	2134–2504	2134–2504	USA
	IV	AY445077	98–2645	EU726268/CDV3; EU716337/164071	2015–2833	7337–9131	2134–2504	2134–2504	USA
	V	AB462810	007Lm-1vp	AY649446/01-2689; AB476401/011C	9846–10558		9789–10558	9846–10558	Japan
	VI	AB474397	007Lm	AY649446/01-2689; AB476401/011C	9846–10558		9789–10558	9846–10558	Japan
L gene	I	JX681125	HLJ1-06	KC427278/HeBei; HM046486/Phoca/Caspian/2007	3743–5297		3749–5255	3749–5255	China
	V	AB462810	007Lm-1vp	AY649446/01-2689; AB476401/011C	683–1562		620–1547	683–1547	Japan
	VI	AB474397	007Lm	AY649446/01-2689; AB476401/011C	683–1562		620–1547	683–1547	Japan
P gene	II	AY466011	98–2654	EU726268/CDV3; EU716337/164071	345–977		334–1075	345–977	USA
	III	AY542312	98–2646	EU726268/CDV3; EU716337/164071	345–975		333–1158	345–977	USA
	IV	AY445077	98–2645	EU726268/CDV3; EU716337/164071	345–977		334–1075	345–977	USA

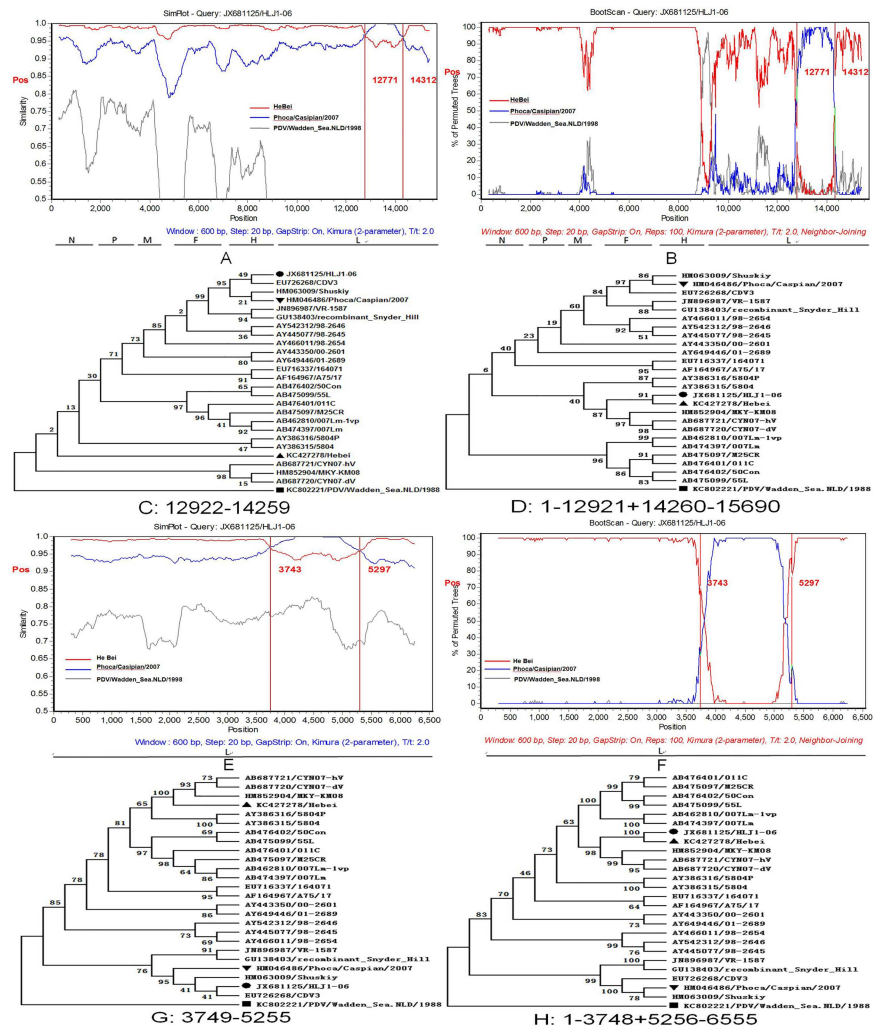
<sup>a</sup> The major and minor putative parental lineages of recombination isolates.

<https://doi.org/10.1371/journal.pone.0175416.t002>

**Three similar intragenic recombinants were identified in the America-1 group of the CDV genome and P gene.** The America-1 group contains three strains (AY445077/98-2645, AY542312/98-2646, and AY466011/98-2654). According to the analysis of the three genome sequences by RDP4 and SimPlot, three potential recombination events harbored identical breakpoint regions in the CDV genome (Table 2). The three sequences aligned using ClustalW and compared by the DNASTar program showed a high sequence similarity (> 99%) with each other (date not shown). Here, AY445077/98-2645 served as a representative strain for further analyses. Based on the results of RDP4 and SimPlot, the breakpoint regions of the potential recombinant strain (AY445077/98-2645) were located from 2015–2833 (SimPlot) and 2134–2504 (RDP4). The American and the Chinese strains, EU716337/164071 and EU726268/CDV3, respectively, act as the major and minor putative parents.

To further analyze the identified recombination events, the AY445077/98-2645 strain was used as the query in the SimPlot program (Fig 4A). The result demonstrated that the sequence of AY445077/98-2645 showed a greater relatedness with one putative parent lineage of EU716337/164071 in the region from position 1–2014 and 2834–15690 bp than the other putative parent (EU726268/CDV3). However, the sequence, EU726268/CDV3, from 2015–2833 bp shared a greater similarity with AY445077/98-2645 than EU716337/164071. The identical evidence was provided by the Boot Scanning result (Fig 4B).

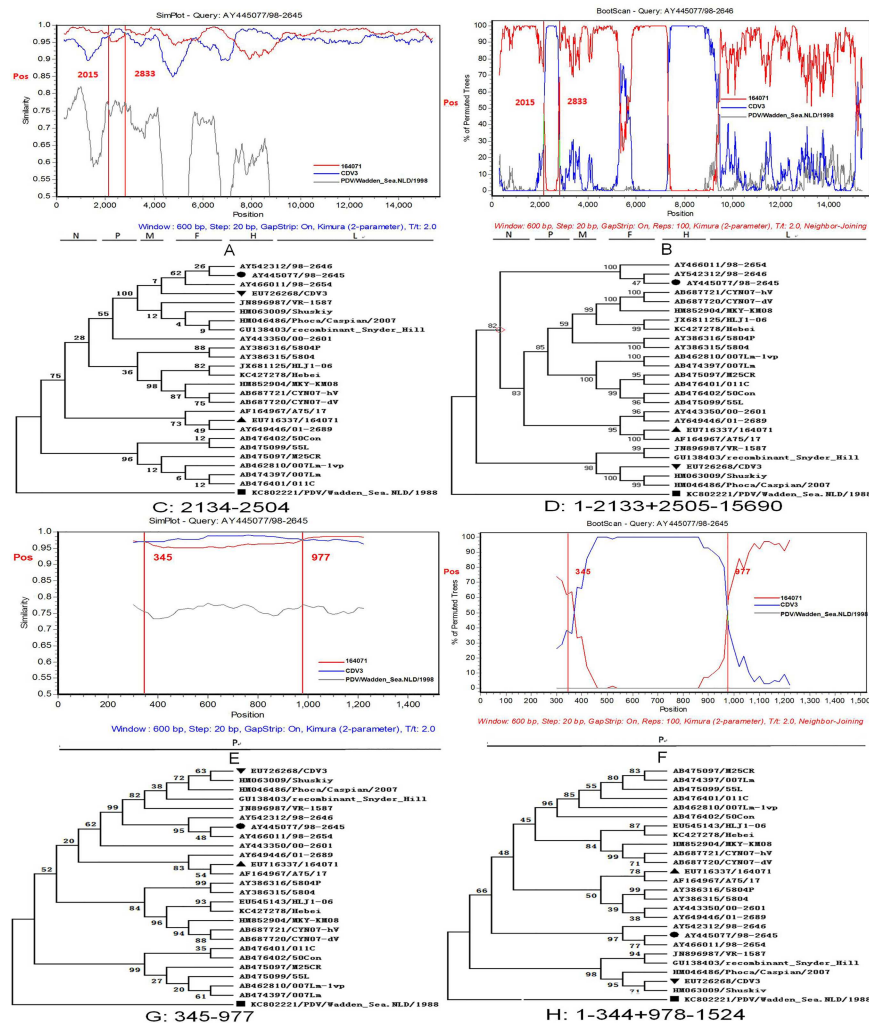
Subsequently, according to the potential recombinant events and cause of the differences in the breakpoint regions in SimPlot and RDP, we finally selected the conservative interval



**Fig 3. Analysis of recombination breakpoints in the CDV genome strain HLJ1-06 using the SimPlot program and MEGA 5.0 software package.** (A) A similarity Plot analysis of the genome sequences of JX681125/HLJ1-06 and its putative parents (KC427278/HeBei and HM046486/Phoca/Caspian/2007). The JX681125/HLJ1-06 was used as the query. The y-axis presented the percentage of identity with a window size of 600 bp and a step size of 20 bp. The vertical line indicated the breakpoint regions. (B) Boot Scan analysis of JX681125/HLJ1-06 and its parent sequences. The y-axis presented the percentage of permuted trees using a sliding window. (E) Similarity Plot analysis of the L gene sequences of JX681125/HLJ1-06 and its putative parents (KC427278/HeBei and HM046486/Phoca/Caspian/2007). (F) Boot Scan analysis of JX681125/HLJ1-06 and its parent sequences in L gene. The remaining are identical as in panels A and B. A PDV isolate, KC80222/PDV/Wadden\_Sea.NLD/1998 was used as an outgroup. (C and D) The phylogenetic tree of the segment in different regions in the CDV genome from 12922–14259 and 1–12921 + 14260–15690 bp. (G and H) The phylogenetic tree of the segment in different regions in the L gene from 3749–5255 and 1–3748 + 5256–6555 bp. The recombinant strain and their major and minor putative parents are indicated with “●”, “▲”, and “▼”, respectively. The PDV isolates are marked with a “■”.

<https://doi.org/10.1371/journal.pone.0175416.g003>

(2134–2504) for analysis. The phylogenetic trees were also constructed using MEGA 5.0 to determine the actual occurrences in the potential recombinant event. In the putative non-recombinant regions (positions 1–2133b + 2505–15690 bp), AY445077/98-2645 and EU716337/164071 were clustered into the same sub-lineage (Fig 4C), while EU726268/CDV3 was grouped into a distinct sub-lineage. However, in the putative recombinant regions, 2134–2504 bp, the



**Fig 4. Evidence for homologous recombination in the CDV genome and L gene.** (A) Similarity Plot analysis of the genome sequences of AY445077/98-2645 and its putative parents (EU716337/164071 and EU726268/CDV3). The AY445077/98-2645 was used as the query. The y-axis indicated the percentage of identity with a window size of 600 bp and a step size of 20 bp. The vertical line indicated the breakpoint regions. (B) Boot Scan analysis of AY445077/98-2645 and its parent sequences. The y-axis presented the percentage of permuted trees using a sliding window. (E) The similarity plot analysis of the L gene sequences of AY445077/98-2645 and its putative parents (EU716337/164071 and EU726268/CDV3). (F) The Boot Scan analysis of AY445077/98-2645 and its parent sequences in the P gene. The remaining results are the similar as in panels A and B. A PDV isolate, KC80222/PDV/Wadden\_Sea. NLD/1988 was used as an outgroup. (C and D) The phylogenetic tree of the segment in different regions in the CDV genome from positions 2134–2504 and positions 1–2133 + 2505–15690 bp. (G and H) The phylogenetic tree of the segment in different regions in the P gene from 345–977 and 1–344 + 978–1524 bp. The recombinant strain and their major and minor putative parents are indicated with “●”, “▲”, and “▼”, respectively. The PDV isolates are marked with a “■”.

<https://doi.org/10.1371/journal.pone.0175416.g004>

arrangement of the phylogenetic tree reflecting the relationship of the three isolates was converse to the previous tree (Fig 4D).

In order to identify the gene and the position of the segment that leads to the potential recombinant events, all the six individual genes were analyzed in RDP4, SimPlot, and MEGA 5.0 programs. Interestingly, we found a potential recombinant event occurring in the P gene that was similar to that described in the potential recombinant event (AY445077/98-2645,



breakpoint regions from 2134–2504 bp). Compared to the two potential recombinant events, these breakpoint regions exhibit a high similarity to the recombinant strain and the major and minor putative parents except the breakpoints. The breakpoint regions in the P gene were located from 345–977 (SimPlot) and 334–1075 (RDP4) bp. We also analyzed this potential recombination event with the method used in the CDV genome strain (AY445077/98-2645) and observed identical results (Fig 4E–4H).

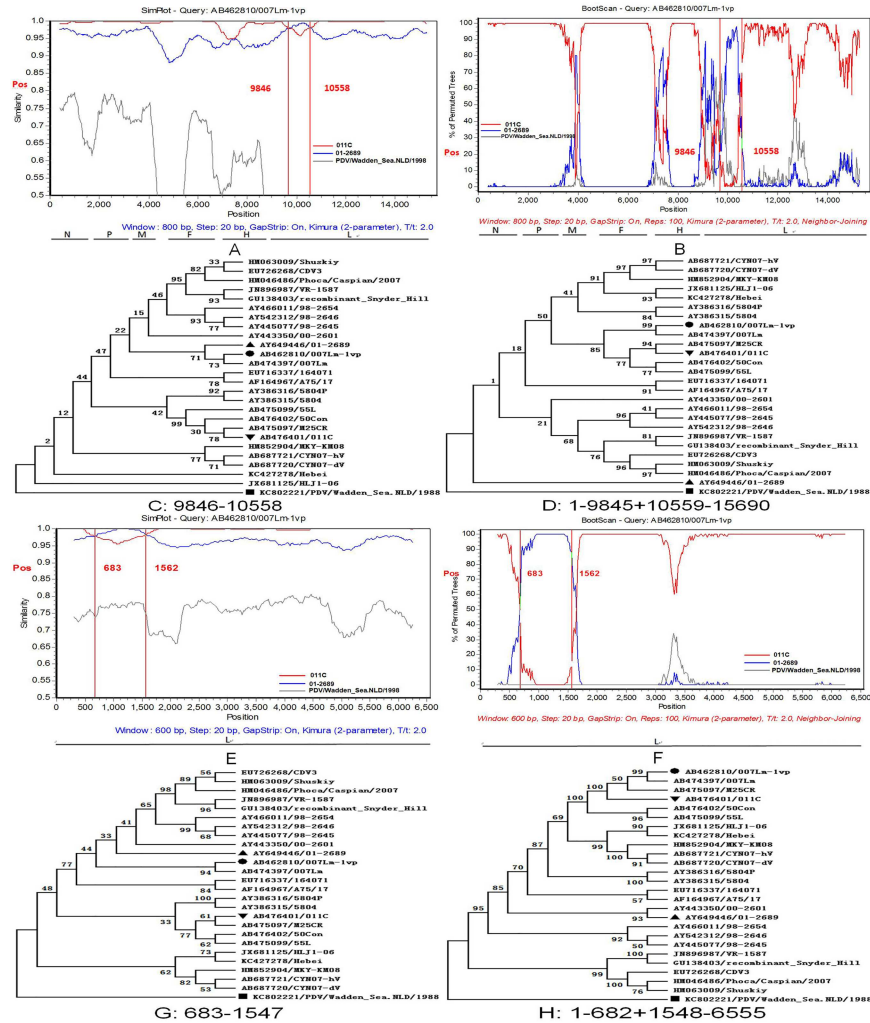
**Last two similar intragenic recombinants were identified in strains 007Lm-1vp and 007Lm of the CDV genome.** In the case of the last two potential recombinant strains are AB462810/007Lm-1vp and AB474397/007Lm, the Japan strain AB476401/011C, and the American strain AY649446/01-2689 act as their major and minor putative parents. The two sequences of the recombinant strains were aligned using ClustalW and compared by DNASTar; we found > 99% homology (date not shown). Thus, we just select AB462810/007Lm-1vp as a representative strain for further analysis. Based on the results obtained by RDP4 and SimPlot, the breakpoint regions of the potential recombinant strain (AB462810/007Lm-1vp) were located from 9846–10558 (SimPlot) and 9789–10558 (RDP4).

To further analyze the identified recombination event, the AB462810/007Lm-1vp strain was used as the query in the SimPlot program (Fig 5A). This revealed that the sequence of AB462810/007Lm-1vp showed a greater affinity with one putative parent lineage of AB476401/011C than the other putative parent AY649446/01-2689 in the region 1–9719 + 10635–15690 bp. However, the sequence from 9720–10634 bp in AY649446/01-2689 shared a greater similarity with AB462810/007Lm-1vp than AB476401/011C. An identical evidence was provided by Boot Scanning (Fig 5B). Due to the difference of the breakpoint regions in SimPlot and RDP, we finally selected the partial 9846–10558 bp region for analysis. Then, the recombination events were further confirmed by constructing the phylogenetic tree using MEGA 5.0. In the putative non-recombinant regions (positions 1–9845 and 10559–15690 bp), AB462810/007Lm-1vp and AB476401/011C were clustered into the same sub-lineage (Fig 5C), while Phoca/Caspian/2007 was grouped into a distinct sub-lineage. However, in the putative recombinant regions 9846–10558 bp, the arrangement of the phylogenetic tree reflecting the relationship of the three isolates was in contrast to the previous tree (Fig 5D). The topology of the two phylogenetic trees around the breakpoint showed a significant statistical discrepancy when the PDV was included in the analyzed data (Shimodaira–Hasegawa test,  $p < 0.001$ ).

In order to identify the gene and position of the segment leading to the potential recombinant events, all the six individual genes were analyzed in RDP4, SimPlot and MEGA 5.0 programs. Interestingly, we found a potential recombinant event occurring in the L gene which was the similar to that described in the potential recombinant event (AB462810/007Lm-1vp, breakpoint regions from 9846–10558 bp). Compared to the two potential recombinant events, these regions have a high similarity than that containing the recombinant strain, major and minor putative parents except the breakpoints. The breakpoint regions in the L gene were located from 683–1562 (in SimPlot) and 620–1547 (RDP4). We also analyzed this potential recombinant event with the method employed for the analysis of the CDV genome strain, AB462810/007Lm-1vp, and observed similar results (Fig 5E–5G).

## Selection pressure on the CDVs

The ratio of dN:dS substitutions/nucleotide site was estimated from an alignment of the six individual genes in CDV. The predominant selection pressure on the six individual genes was negative as reflected in a dN/dS value of < 1 (S2 Table). Using FEL ( $p = 0.1$ ), 99% of the codons were under the negative selection pressure. However, we also found nine representative positive selection sites in the six individual genes of CDV; four positive selection sites were



**Fig 5. Analysis of recombination breakpoints in CDV genome strain 007Lm-1vp by SimPlot program and MEGA 5.0 software package.** (A) Similarity Plot analysis of the genome sequences of AY462810/007Lm-1vp and its putative parents (AB476401/011C and AY649446/01-2689). The AY462810/007Lm-1vp was used as the query. The y-axis exhibited the percentage of identity with a window size of 600 bp (800 bp) and a step size of 20 bp. The vertical line indicated the breakpoint regions. (B) Boot Scan analysis of AY462810/007Lm-1vp and its parent sequences. The y-axis represented the percentage of permuted trees using a sliding window. (E) The similarity Plot analysis of the L gene sequences of AY462810/007Lm-1vp and its putative parents (AB476401/011C and AY649446/01-2689). (F) Boot Scan analysis of AY462810/007Lm-1vp and its parent sequences in the L gene. The remaining are similar as in panels A and B. A PDV isolate, KC80222/PDV/Wadden\_Sea.NLD/1998 was used as an outgroup. (C and D) Phylogenetic tree of the segment in different regions in the CDV genome from 9846–10558 and 1–9845 + 10559–15690 bp. (G and H) Phylogenetic tree of the segment in different regions in the L gene from positions 683–1547 and positions 1–682 + 1548–6555 bp. The recombinant strain and their major and minor putative parents are indicated with “●”, “▲”, and “▼”, respectively; the PDV isolates are marked with a “■”.

<https://doi.org/10.1371/journal.pone.0175416.g005>

located in the H (positions 178), N gene (positions 456), M (positions 312), and L genes (positions 2137), respectively. Two positive selection sites (at positions 79 and 110) were detected in the F gene; whereas, three positive selection sites (at positions 148, 256 and 278) were detected in the P gene (FEL p-value < 1.0; SLAC p-value < 0.3) (S3 Table).

## Discussion

During the past two decades, despite the vaccination procedures adopted in the world, CDV continues to pose a severe threat to breeding foxes, raccoon dogs, minks, and domestic dogs [20, 21]. The prevalence of CDV may result in significant economic losses to the fur industry. Genetic diversity has been observed between wild-type CDVs and the vaccine strains [20, 22–26]. A majority of the live-modified CDV vaccine strains isolated could be dated to 1930–1950s [22, 27–29]. The diversity between the vaccine strains and wild-type CDVs may be attributed to several mechanisms, such as adapted to new host species [30, 31], antigenic escape [27, 32–34] and/or genetic recombination between wild-type strains [35], variedly driving the evolution of the virus. However, which intragenic recombination occurs and whether it contributes to the evolution of the virus is yet unclear.

In this study, 25 full-length CDV strains isolated from breeding foxes, raccoon, minks, seal, ferret, *Macaca fascicularis*, *Macaca mulatta*, and *Canis lupus familiaris* were sequenced and analyzed. Based on the phylogenetic analysis of the CDV genomes, all the CDV isolates were divided into six lineages: Asia-1, Asia-2, Asia-3, Europe, America-1, and America-2. According to the phylogenetic analysis, all the CDV isolates detected in naturally occurring cases clustered according to the geographical distribution. Comparing the phylogenetic trees between CDV genome sequences and six individual gene sequences, the topology of the phylogenetic trees in every individual gene were found to have a high identity with the phylogenetic tree of the genome sequence. Furthermore, the sequence homology of the CDV genome exhibited a high similarity than the six individual genes.

Homologous recombination drives genotype diversity in some RNA viruses [36–38]. Recombination, especially between different serotypes, may accelerate the evolution of the virus effectuating a rapid change in its epidemiology [39]. In this study, nine groups of potential recombination events were detected by RDP4 and SimPlot programs, and the six potential recombination events included six recombination strains (JX681125/HLJ1-06, AY445077/98-2645, AY542312/98-2646, AY466011/98-2654, AB462810/007Lm-1vp, and AB474397/007Lm). According to the phylogenetic analysis, the JX681125/HLJ1-06 strain belongs to the group of Asia-1. The other three strains, AY445077/98-2645, AY542312/98-2646, and AY466011/98-2654, belonged to the group of America-1. The remaining two strains, AB462810/007Lm-1vp and AB474397/007Lm, belonged to the group of Asia-2. Similar to that reported previously [40], we confirmed that JX681125/HLJ1-06, AB462810/007Lm-1vp, and AB474397/007Lm were wild-type strains. On the contrary, the other three strains, AY445077/98-2645, AY542312/98-2646, and AY466011/98-2654 were vaccine strains but not artificial [41]. Moreover, it could be speculated that the strains of the same group may have a high homology, thereby leading to similar recombination events. Thus, the six sequences were aligned using ClustalW and compared by DNASTar. Consequently, we found that the three strains, AY445077/98-2645, AY542312/98-2646, and AY466011/98-2654, showed a high sequence similarity (> 99%) with each other (date not shown), as well as, the other two strains (AB462810/007Lm-1vp and AB474397/007Lm). Finally, the three strains, X681125/HLJ1-06, AY445077/98-2645, and AB462810/007Lm-1vp, were selected as representatives for further analyses.

The RDP4 and SimPlot analyses detected the three recombination regions. The first recombination region was located from 12922–14259 bp in the JX681125/HLJ1-06 strain and the two strains, KC427278/Hebei and HM046486/ Phoca/Caspian/2007, might be the major and minor putative parents. The second recombination region was localized in the position 2135–2504 bp in the AY445077/98-2645 strain and the two strains, EU716337/164071 and EU726268/CDV3, might be the major and minor putative parents. The last recombination

region was found at 9846–10558 bp in the AB462810/007Lm-1vp strain and two strains, AB476401/011C and AY649446/01-2689, might act as the major and minor putative parents.

In order to explore which gene can lead to the intragenic recombination, all the strains of six individual genes were also analyzed by RDP4 and SimPlot programs. Three parental recombination events were detected in P and L gene, they were the same as what we found in the strains (JX681125/HLJ1-06, AY445077/98-2645, and AB462810/007Lm-1vp) and no parental recombination event was found in other genes (H, N, F, and M gene). Previously report said the L gene exhibits a highly conserved property [40]. However, we found that P and L genes are the major recombination regions and we speculate that the L gene of Canine distemper virus may not have a high conservativeness.

In order to further illustrate the possibility of the six potential recombination events, a set of statistically incongruent phylogenetic trees (two for every event) were analyzed for their presence. [19]. The first phylogenetic tree was constructed in the recombination region and the second in the non-recombination region. The results indicated that the recombinant interval had a specific group of parents (major and minor putative parents), and the non-recombination interval had another group of parents, which might statistically support the phylogenetic recombination signal. These results suggested that all the phylogenetic trees supported the recombination result obtained from RDP4 and SimPlot analyses. In 2003, Holmes et al. indicated that the rates of homologous recombination in negative-sense RNA viruses were extremely low; however, the possibility of recombination in CDV should be investigated further [42]. The current study confirms the existence of homologous recombination in CDVs.

The data regarding the host of the virus are essential to understand the genetic diversity and evolution of CDV. In this study, we found that the hosts of recombination strains and their major and minor putative parents were different. The host of the first recombination strain JX681125/HLJ1-06 was a fox, and the hosts of its major (KC427278/Hebei) and minor putative (HM046486/ Phoca/Caspian/2007) parents were mink and seal, respectively. The host of the second recombination strain AY445077/98-2645 was a raccoon, but the hosts of its major (EU716337/164071) and minor putative (EU726268/CDV3) parents were mink and *Canis familiaris*, respectively. The host of the third recombination strain AB462810/007Lm-1vp was Vero cell, and the hosts of its major (AB476401/011C) and minor putative (AY649446/01-2689) parents were raccoon and *Canis lupus familiaris*. Therefore, we may speculate that CDV isolated from different species may occur in intragenic recombination and contribute towards the evolutionary process in the virus.

RNA viruses are genetically flexible. During the replication of their genomes, the nucleotide substitutions occur at high frequencies presumably to allow rapid adaptations to various selection pressures [43]. In the current study, considering that the six individual genes were vital to CDV, we assessed the nucleotide selection pressure in the six individual genes (H, F, P, N, M, L). No previous studies suggested whether these six individual genes were under negative or positive selection. In this study, the mean ratio of dN and dS rates (dN/dS) for the six individual genes were both < 1, and the rate for the six individual genes was M (0.074) < L (0.082) < N (0.095) < H (0.263) < F (0.310) < P (0.352) (S2 Table). These result indicated that the six individual genes were under purifying selection; P was under stronger negative selection than other genes, while M exhibited a highly conserved property. This finding strongly correlated with the higher evolutionary rate of P compared to other genes and postulated the correlation of the selection pressure and the rate of evolution [44, 45]. We also found nine positive selection sites (H-178, N-456, M-312, L-2137, F-79/110, and P-148/256/278), especially, positive natural selection involved in immune escape could cause similar sequences to evolve in unrelated strains.

In conclusion, six phylogroups and homologous recombination events, occurring in the CDVs, were identified based on the CDV genome and their gene segments. The majority of CDVs underwent a negative selection pressure. This information provides a valuable reference for the study of molecular epidemiology of CDVs.

## Supporting information

**S1 Table. CDV strains with evidence for potential recombination in RDP analysis.**  
(DOCX)

**S2 Table. The positive-selection sites represented by codons in six individual genes of CDV.**  
(DOCX)

**S3 Table. Evidence for positive and negative selection using four detection methods.**  
(DOCX)

## Author Contributions

**Conceptualization:** CWY GQW.

**Data curation:** CWY WXL YBW.

**Formal analysis:** CWY JLH.

**Funding acquisition:** GQW.

**Software:** CWY LGZ.

**Writing – original draft:** CWY YBW WXL.

**Writing – review & editing:** CWY.

## References

1. Radecke F, Spielhofer P, Schneider H, Kaelin K, Huber M, Dotsch C, et al. Rescue of measles viruses from cloned DNA. *The EMBO journal*. 1995; 14(23):5773–84. Epub 1995/12/01. PubMed Central PMCID: PMC394696. PMID: [8846771](https://pubmed.ncbi.nlm.nih.gov/8846771/)
2. Sidhu MS, Husar W, Cook SD, Dowling PC, Udem SA. Canine distemper terminal and intergenic non-protein coding nucleotide sequences: completion of the entire CDV genome sequence. *Virology*. 1993; 193(1):66–72. Epub 1993/03/01. <https://doi.org/10.1006/viro.1993.1103> PMID: [8438593](https://pubmed.ncbi.nlm.nih.gov/8438593/)
3. Sidhu MS, Chan J, Kaelin K, Spielhofer P, Radecke F, Schneider H, et al. Rescue of synthetic measles virus minireplicons: measles genomic termini direct efficient expression and propagation of a reporter gene. *Virology*. 1995; 208(2):800–7. Epub 1995/04/20. <https://doi.org/10.1006/viro.1995.1215> PMID: [7747454](https://pubmed.ncbi.nlm.nih.gov/7747454/)
4. Kolakofsky D, Pelet T, Garcin D, Hausmann S, Curran J, Roux L. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *Journal of virology*. 1998; 72(2):891–9. Epub 1998/01/28. PubMed Central PMCID: PMC124558. PMID: [9444980](https://pubmed.ncbi.nlm.nih.gov/9444980/)
5. Cathomen T, Mrkic B, Spehner D, Drillien R, Naef R, Pavlovic J, et al. A matrix-less measles virus is infectious and elicits extensive cell fusion: consequences for propagation in the brain. *The EMBO journal*. 1998; 17(14):3899–908. Epub 1998/07/22. PubMed Central PMCID: PMC1170725. <https://doi.org/10.1093/emboj/17.14.3899> PMID: [9670007](https://pubmed.ncbi.nlm.nih.gov/9670007/)
6. Lamb RA. Paramyxovirus fusion: a hypothesis for changes. *Virology*. 1993; 197(1):1–11. Epub 1993/11/01. <https://doi.org/10.1006/viro.1993.1561> PMID: [8212546](https://pubmed.ncbi.nlm.nih.gov/8212546/)
7. He CQ, Xie ZX, Han GZ, Dong JB, Wang D, Liu JB, et al. Homologous recombination as an evolutionary force in the avian influenza A virus. *Molecular biology and evolution*. 2009; 26(1):177–87. Epub 2008/10/22. <https://doi.org/10.1093/molbev/msn238> PMID: [18931384](https://pubmed.ncbi.nlm.nih.gov/18931384/)

8. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23(21):2947–8. Epub 2007/09/12. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
9. Phillips ML, Hall TA, Sekar K, Tomey JL. Assessment of medical personnel exposure to nitrogen oxides during inhaled nitric oxide treatment of neonatal and pediatric patients. *Pediatrics*. 1999; 104(5 Pt 1):1095–100. Epub 1999/11/05.
10. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*. 2011; 28(10):2731–9. Epub 2011/05/07. PubMed Central PMCID: PMC3203626. <https://doi.org/10.1093/molbev/msr121> PMID: 21546353
11. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998; 14(9):817–8. Epub 1999/01/27. PMID: 9918953
12. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19(12):1572–4. Epub 2003/08/13. PMID: 12912839
13. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010; 26(19):2462–3. Epub 2010/08/28. PubMed Central PMCID: PMC2944210. <https://doi.org/10.1093/bioinformatics/btq467> PMID: 20798170
14. Liang Z, Kumar AS, Jones MS, Knowles NJ, Lipton HL. Phylogenetic analysis of the species Theilovirus: emerging murine and human pathogens. *Journal of virology*. 2008; 82(23):11545–54. Epub 2008/09/26. PubMed Central PMCID: PMC2583687. <https://doi.org/10.1128/JVI.01160-08> PMID: 18815294
15. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*. 2006; 23(2):254–67. Epub 2005/10/14. <https://doi.org/10.1093/molbev/msj030> PMID: 16221896
16. Aris-Brosou S. Least and most powerful phylogenetic tests to elucidate the origin of the seed plants in the presence of conflicting signals under misspecified models. *Systematic biology*. 2003; 52(6):781–93. Epub 2003/12/12. PMID: 14668117
17. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*. 2005; 21(10):2531–3. Epub 2005/02/17. <https://doi.org/10.1093/bioinformatics/bti320> PMID: 15713735
18. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution*. 2005; 22(5):1208–22. Epub 2005/02/11. <https://doi.org/10.1093/molbev/msi105> PMID: 15703242
19. Boni MF, de Jong MD, van Doorn HR, Holmes EC. Guidelines for identifying homologous recombination events in influenza A virus. *PloS one*. 2010; 5(5):e10434. Epub 2010/05/11. PubMed Central PMCID: PMC2862710. <https://doi.org/10.1371/journal.pone.0010434> PMID: 20454662
20. He HB, Lv XX, Yu QW, Feng YQ. Multiresidue determination of (fluoro)quinolone antibiotics in chicken by polymer monolith microextraction and field-amplified sample stacking procedures coupled to CE-UV. *Talanta*. 2010; 82(4):1562–70. Epub 2010/08/31. <https://doi.org/10.1016/j.talanta.2010.07.055> PMID: 20801373
21. Wang FX, Zhang XJ, Pan L, Qiao SK, Guo XL, Dong ZR. [Angioimmunoblastic T-cell lymphoma with autoimmune hemolytic anemia and pure red cell aplasia]. *Zhongguo shi yan xue ye xue za zhi*. 2007; 15(4):862–5. Epub 2007/08/22. PMID: 17708820
22. Bolt G, Jensen TD, Gottschalck E, Arctander P, Appel MJ, Buckland R, et al. Genetic diversity of the attachment (H) protein gene of current field isolates of canine distemper virus. *The Journal of general virology*. 1997; 78 (Pt 2):367–72. Epub 1997/02/01.
23. Mochizuki M, Hashimoto M, Hagiwara S, Yoshida Y, Ishiguro S. Genotypes of canine distemper virus determined by analysis of the hemagglutinin genes of recent isolates from dogs in Japan. *Journal of clinical microbiology*. 1999; 37(9):2936–42. Epub 1999/08/17. PubMed Central PMCID: PMC85418. PMID: 10449479
24. Pardo ID, Johnson GC, Kleiboeker SB. Phylogenetic characterization of canine distemper viruses detected in naturally infected dogs in North America. *Journal of clinical microbiology*. 2005; 43(10):5009–17. Epub 2005/10/07. PubMed Central PMCID: PMC1248462. <https://doi.org/10.1128/JCM.43.10.5009-5017.2005> PMID: 16207955
25. Martella V, Cirone F, Elia G, Lorusso E, Decaro N, Campolo M, et al. Heterogeneity within the hemagglutinin genes of canine distemper virus (CDV) strains detected in Italy. *Veterinary microbiology*. 2006; 116(4):301–9. Epub 2006/05/30. <https://doi.org/10.1016/j.vetmic.2006.04.019> PMID: 16730927
26. Martella V, Elia G, Buonavoglia C. Canine distemper virus. *The Veterinary clinics of North America Small animal practice*. 2008; 38(4):787–97, vii–viii. Epub 2008/05/27. <https://doi.org/10.1016/j.cvsm.2008.02.007> PMID: 18501278

27. Iwatsuki K, Tokiyoshi S, Hirayama N, Nakamura K, Ohashi K, Wakasa C, et al. Antigenic differences in the H proteins of canine distemper viruses. *Veterinary microbiology*. 2000; 71(3–4):281–6. Epub 2000/03/07. PMID: [10703710](#)
28. Lednicky JA, Dubach J, Kinsel MJ, Meehan TP, Bocchetta M, Hungerford LL, et al. Genetically distant American Canine distemper virus lineages have recently caused epizootics with somewhat different characteristics in raccoons living around a large suburban zoo in the USA. *Virology journal*. 2004; 1:2. Epub 2004/10/28. PubMed Central PMCID: PMC524033. <https://doi.org/10.1186/1743-422X-1-2> PMID: [15507154](#)
29. Lan NT, Yamaguchi R, Kien TT, Hirai T, Hidaka Y, Nam NH. First isolation and characterization of canine distemper virus in Vietnam with the immunohistochemical examination of the dog. *The Journal of veterinary medical science*. 2009; 71(2):155–62. Epub 2009/03/06. PMID: [19262025](#)
30. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annual review of microbiology*. 1997; 51:151–78. Epub 1997/01/01. <https://doi.org/10.1146/annurev.micro.51.1.151> PMID: [9343347](#)
31. McCarthy AJ, Shaw MA, Goodman SJ. Pathogen evolution and disease emergence in carnivores. *Proceedings Biological sciences*. 2007; 274(1629):3165–74. Epub 2007/10/25. PubMed Central PMCID: PMC2293938. <https://doi.org/10.1098/rspb.2007.0884> PMID: [17956850](#)
32. Lan NT, Yamaguchi R, Furuya Y, Inomata A, Ngamkala S, Naganobu K, et al. Pathogenesis and phylogenetic analyses of canine distemper virus strain 007Lm, a new isolate in dogs. *Veterinary microbiology*. 2005; 110(3–4):197–207. Epub 2005/09/08. <https://doi.org/10.1016/j.vetmic.2005.07.016> PMID: [16144749](#)
33. Lan NT, Yamaguchi R, Kawabata A, Uchida K, Kai K, Sugano S, et al. Stability of canine distemper virus (CDV) after 20 passages in Vero-DST cells expressing the receptor protein for CDV. *Veterinary microbiology*. 2006; 118(3–4):177–88. Epub 2006/09/20. <https://doi.org/10.1016/j.vetmic.2006.07.015> PMID: [16982161](#)
34. Martella V, Elia G, Lucente MS, Decaro N, Lorusso E, Banyai K, et al. Genotyping canine distemper virus (CDV) by a hemi-nested multiplex PCR provides a rapid approach for investigation of CDV outbreaks. *Veterinary microbiology*. 2007; 122(1–2):32–42. Epub 2007/02/06. <https://doi.org/10.1016/j.vetmic.2007.01.005> PMID: [17275219](#)
35. Han GZ, Liu XP, Li SS. Cross-species recombination in the haemagglutinin gene of canine distemper virus. *Virus research*. 2008; 136(1–2):198–201. Epub 2008/06/14. <https://doi.org/10.1016/j.virusres.2008.04.022> PMID: [18550189](#)
36. Bruen TC, Poss M. Recombination in feline immunodeficiency virus genomes from naturally infected cougars. *Virology*. 2007; 364(2):362–70. Epub 2007/04/21. PubMed Central PMCID: PMC1934564. <https://doi.org/10.1016/j.virol.2007.03.023> PMID: [17445856](#)
37. Herrewegh AA, Smeenk I, Horzinek MC, Rottier PJ, de Groot RJ. Feline coronavirus type II strains 79–1683 and 79–1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *Journal of virology*. 1998; 72(5):4508–14. Epub 1998/04/29. PubMed Central PMCID: PMC109693. PMID: [9557750](#)
38. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of virology*. 1999; 73(1):152–60. Epub 1998/12/16. PubMed Central PMCID: PMC103818. PMID: [9847317](#)
39. Raaben M, Grinwis GC, Rottier PJ, de Haan CA. The proteasome inhibitor Velcade enhances rather than reduces disease in mouse hepatitis coronavirus-infected mice. *Journal of virology*. 2010; 84(15):7880–5. Epub 2010/05/21. PubMed Central PMCID: PMC2897637. <https://doi.org/10.1128/JVI.00486-10> PMID: [20484516](#)
40. Demeter Z, Palade EA, Hornyak A, Rusvai M. Controversial results of the genetic analysis of a canine distemper vaccine strain. *Veterinary microbiology*. 2010; 142(3–4):420–6. Epub 2009/11/26. <https://doi.org/10.1016/j.vetmic.2009.10.017> PMID: [19931324](#)
41. Lednicky JA, Meehan TP, Kinsel MJ, Dubach J, Hungerford LL, Sarich NA, et al. Effective primary isolation of wild-type canine distemper virus in MDCK, MV1 Lu and Vero cells without nucleotide sequence changes within the entire haemagglutinin protein gene and in subgenomic sections of the fusion and phospho protein genes. *Journal of virological methods*. 2004; 118(2):147–57. Epub 2004/04/15. <https://doi.org/10.1016/j.jviromet.2004.02.004> PMID: [15081610](#)
42. Chare ER, Gould EA, Holmes EC. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *The Journal of general virology*. 2003; 84(Pt 10):2691–703. Epub 2003/09/19. <https://doi.org/10.1099/vir.0.19277-0> PMID: [13679603](#)
43. Cong Y, Zarlenga DS, Richt JA, Wang X, Wang Y, Suo S, et al. Evolution and homologous recombination of the hemagglutinin-esterase gene sequences from porcine torovirus. *Virus genes*. 2013; 47(1):66–74. Epub 2013/06/12. <https://doi.org/10.1007/s11262-013-0926-y> PMID: [23749172](#)

44. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *Journal of molecular evolution*. 2006; 62(5):551–63. Epub 2006/03/25. <https://doi.org/10.1007/s00239-005-0221-1> PMID: 16557338
45. Shackelton LA, Parrish CR, Truyen U, Holmes EC. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(2):379–84. Epub 2005/01/01. PubMed Central PMCID: PMC544290. <https://doi.org/10.1073/pnas.0406765102> PMID: 15626758