Saudi Journal of Biological Sciences (2014) 21, 311–316

King Saud University

Saudi Journal of Biological Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Extensive homologous recombination in classical swine fever virus: A re-evaluation of homologous recombination events in the strain AF407339

Youhua Chen [a,*], You-Fang Chen [b]

[a] Department of Zoology, University of British Columbia, Vancouver V6T 1Z4, Canada
[b] School of Software, Harbin Normal University, Heilongjiang Province, China

Received 6 October 2013; revised 20 December 2013; accepted 25 December 2013
Available online 4 January 2014

KEYWORDS

Sequence analysis;
Homologous recombination;
Virus evolution;
PCR contamination

Abstract In this short report, the genome-wide homologous recombination events were re-evaluated for classical swine fever virus (CSFV) strain AF407339. We challenged a previous study which suggested only one recombination event in AF407339 based on 25 CSFV genomes. Through our re-analysis on the 25 genomes in the previous study and the 41 genomes used in the present study, we argued that there should be possibly at least two clear recombination events happening in AF407339 through genome-wide scanning. The reasons for identifying only one recombination event in the previous study might be due to the limited number of available CSFV genome sequences at that time and the limited usage of detection methods. In contrast, as identified by most detection methods using all available CSFV genome sequences, two major recombination events were found at the starting and ending zones of the genome AF407339, respectively. The first one has two parents AF333000 (minor) and AY554397 (major) with beginning and ending breakpoints located at 19 and 607 nt of the genome respectively. The second one has two parents AF531433 (minor) and GQ902941 (major) with beginning and ending breakpoints at 8397 and 11,078 nt of the genome respectively. Phylogenetic incongruence analysis using neighbor-joining algorithm with 1000 bootstrapping replicates further supported the existence of these two recombination events. In addition, we also identified additional 18 recombination events on the available CSFV strains. Some of them may be trivial and can be ignored. In conclusion, CSFV might have relatively high

* Corresponding author. Tel.: +1 5877786218.
E-mail address: haydi@126.com (Y. Chen).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.sjbs.2013.12.004

frequency of homologous recombination events. Genome-wide scanning of identifying recombination events should utilize multiple detection methods so as to reduce the risk of misidentification.

1. Introduction

Recombination, as the exchange of nucleotide segments among sequences, is believed to be ubiquitous in many viruses (Posada and Crandall, 2001; Worobey, 2000). However, in some viruses, the recombination events based on available genomic sequences were rarely detected, for example, influenza A, B and C viruses (Han and Worobey, 2011; He et al., 2009; Liu et al., 2010); Canine distemper virus (Han et al., 2008; McCarthy et al., 2007).

Classical swine fever virus (CSFV) is an ssRNA positive-strand virus that causes the contagious disease of swine. Gene evolution of the virus has been widely explored, from synonymous codon usage pattern (Cao et al., 2012; Tao et al., 2009) to selection pressure analysis (Perez et al., 2012; Shen et al., 2011; Tang et al., 2008). The putative homologous recombination of CSFV has been evaluated as well in an early paper (He et al., 2007), which only identified one recombination evidence. This seminal report would give us a first impression that the frequency of homologous recombination in CSFV might be low. However, given the increasing genome data of CSFV published in NCBI and the common recombination of many viruses (Posada and Crandall, 2001; Van der Walt et al., 2009; Worobey, 2000; Wu et al., 2009), we may expect to see some new recombination evidence happening on CSFV genomes when new sequences data are included. A comprehensive recombination analysis of CSFV has been performed in a previous study (He et al., 2007). The central objective of the present study is to re-evaluate and quantify the genome-wide homologous recombination evidence of CSFV. We would perform extensive recombination detection analyses using all available CSFV genome sequences in order to detect all possible recombination events and compare them to the previous study (He et al., 2007).

2. Materials and methods

2.1. Sequences

42 non-redundant complete genomes of CSFV were obtained from GenBank (http://www.ncbi.nlm.nih.gov/). The coding regions of the genomes were extracted and retained for subsequent analysis. The accession numbers for the genomes and other information were available in Appendix I. The sequences were aligned using MUSCLE software (Edgar, 2004) and available from the authors upon request.

Besides, since our study is to compare the results with the previous study (He et al., 2007), we also downloaded the 25 genomes listed in He et al. (2007)'s paper. Some genomes have been included in the above 42 genome dataset. To avoid confusions, we defined He et al. (2007)'s dataset as 'He25.' And, a full genome dataset with 54 sequences of CSFV, which is made by combining the genomes from the above 42 ones and 'He25' together after removing duplicated genomes, is defined as 'Tot54' dataset for comparative study.

2.2. Recombination analysis

The aligned sequences were then subjected to recombination analysis. The RDP software was used to detect homologous recombination events (Martin and Rybicki, 2010; Martin et al., 2010). The software contained a series of recombination detection algorithms, including GENECOV (Padidam et al., 1999), Bootscan/Rescan (Martin et al., 2005), Chimaera (Posada and Crandall, 2001), MaxChi (Maynard Smith, 1992), SiScan (Gibbs et al., 2000), 3Seq (Boni et al., 2007), and RDP (Martin and Rybicki, 2010). All these methods are utilized and compared so as to obtain consensus results. A putative recombination event was retained to subsequent analysis only when it was consistently identified by at least three of the above-mentioned seven algorithms (Liu et al., 2010). Here we defined the minor parent as the one contributing the smaller fraction of the recombinant, while the major parent as the one contributing the larger fraction of the recombinant (Martin et al., 2010).

To avoid false-positive results, phylogenetic analysis of recombination was performed (Boni et al., 2008; Liu et al., 2010). For each putative recombinant, the entire dataset alignment was divided at the breakpoint positions. If two recombination breakpoints were found in a single sequence, the sequence region between the breakpoints was denoted the "minor" region, generated by the minor parent, and the remainder is the "major" region, generated by the major parent. Neighbor-joining phylogenetic trees were constructed to show topological shifts of specific sequences. Phylogenetic incongruence is reflected by a putative recombinant whose

Table 1 Recombinant events for CSFV genomes found in He25 dataset.

Recombinant	Major	Minor	Starting	Ending	Supporting methods
D49532	AF326963	X87939	1386	2375	RDP, GENECONV, Bootscan, SiScan, 3Seq
D49532	AF326963	U90951	2754	3571	RDP, GENECONV, Bootscan, SiScan
DQ127910	Unknown (U90951)	U45478	9787	10,739	RDP, GENECONV, Bootscan, SiScan, 3Seq
AF333000	AY072924	Z46258	88	12,260	All methods
AF407339	AY554397	AF333000	19	639	RDP, GENECONV, Bootscan, SiScan
AF407339	AY072924	Z46258	8421	12,260	All methods

Table 2 Recombinant events for CSFV genomes found in Tot54 combined dataset.

Recombinant	Major	Minor	Starting	Ending	Supporting methods
JQ268754	FJ529205	GU592790	10,678	11,514	RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan
AF407339	AY554397	AF333000	19	607	RDP, GENECONV, Bootscan, SiScan
AF407339	GQ122383	AF333000	Undetermined (11,079)	12,268	RDP, GENECONV, Bootscan, Chimaera, SiScan, 3Seq
AF407339	GQ902941	AF531433	8397	11,078	All methods
AF407339	HQ380231	Unknown (DQ127910)	Undetermined (8397)	8633	RDP, GENECONV, MaxChi, SiScan, 3Seq
DQ127910	GQ902941	AF531433	57	12,280	All methods
EU497410	GQ902941	AF531433	57	12,281	All methods
AY775178	GQ902941	AF531433	57	12,280	All methods
D49533	GQ902941	AF531433	172	12,281	All methods
HM237795	GQ902941	AF531433	172	12,281	All methods
AF326963	GQ902941	AF531433	172	12,284	All methods
D49532	AF326963	U90951	2754	3473	GENECONV, Bootscan, SiScan, 3Seq
D49532	HM237795	U90951	1111	2397	RDP, GENECONV, Bootscan, SiScan, 3Seq
D49532	GQ902941	AF531433	172	12,281	All methods
X87939	GQ902941	AF531433	172	12,281	All methods
U90951	GQ902941	AF531433	172	12,281	All methods
EU789580	HM237795	U90951	12,244	2369	RDP, GENECONV, Bootscan, SiScan, 3Seq
EU789580	GQ902941	AF531433	172	12,281	All methods
EU915211	HM237795	U90951	163	2369	RDP, GENECONV, Bootscan, SiScan, 3Seq
EU915211	GQ902941	AF531433	172	12,281	All methods

distance in the phylogeny is clearly close to one parent and far from another one for each sequence segment. Also, the recombination signal "weak" if one of the components received only low bootstrap support (Boni et al., 2008; Liu et al., 2010).

3. Results and discussion

3.1. Recombination events in CSFV using 'He25' dataset

When redoing the 25 genomes used in He et al. (2007), six unique recombination events were supported by at least 4 out of 7 detection methods and we did find that AF407339 was a recombinant which had two most significant recombination events (the other 4 ones are listed in Table 1). The first one has corresponding parents Z46258 (minor) and AY072924 (major), which is supported by all the 7 recombination methods with starting and ending breakpoints at 8421 and 12,260 nt of the genome (Table 2). The second one has corresponding parents AF333000 (minor) and AY554397 (major) with starting and ending breakpoints at 19 and 639 nt of the genome respectively. This event was only supported by 4 methods (Table 1). Also, it was found that the location of breakpoints for the event which took AF333000 as one parent was different between our re-analysis and He et al. (2007)'s report.

3.2. Recombination events in CSFV using 'Tot54' dataset

When analyzing all the available CSFV genomes (Table 2), it was found that 20 recombination events were identified with the support from at least four out of seven detection methods. As such, it was concluded that the recombination frequency in CSFV genomes is quite high.

Four recombination events have been identified on the recombinant AF407339 (Table 2). The first one had two parents AF333000 (minor) and AY554397 (major) with beginning and ending breakpoints located at 19 and 607 nt of the genome. The second one had two parents AF531433 (minor) and GQ902941 (major) with beginning and ending breakpoints at 8397 and 11,078 nt of the genome respectively. The phylogenetic analyses confirmed the existence of these two recombination events with strong bootstrapping support (Figs. 1 and 2). Finally, the last two ones had very short recombination segments and one/or both of the parents are unknown. Thus, the last two ones were ignored for subsequent discussion.

3.3. A systematic comparison and analysis of the recombination events on AF407339

We explored the three recombination-related sequences AF333000, AY367767 and AF407339 identified by He et al. (2007) by checking possible signals of recombination in RDP program. However, no recombination signals were found for these three sequences. However, when we added another three sequences AY554397, AF531433 and GQ902941 into the analysis, the two recombination events discovered above by using 'He25' and 'Tot54' datasets were clearly present. As such, it is concluded that the recombination event identified by He et al. (2007) might not be robust enough.

We argued that the limitation of He et al. (2007)'s work may be due to the limited sampling of the genome sequences

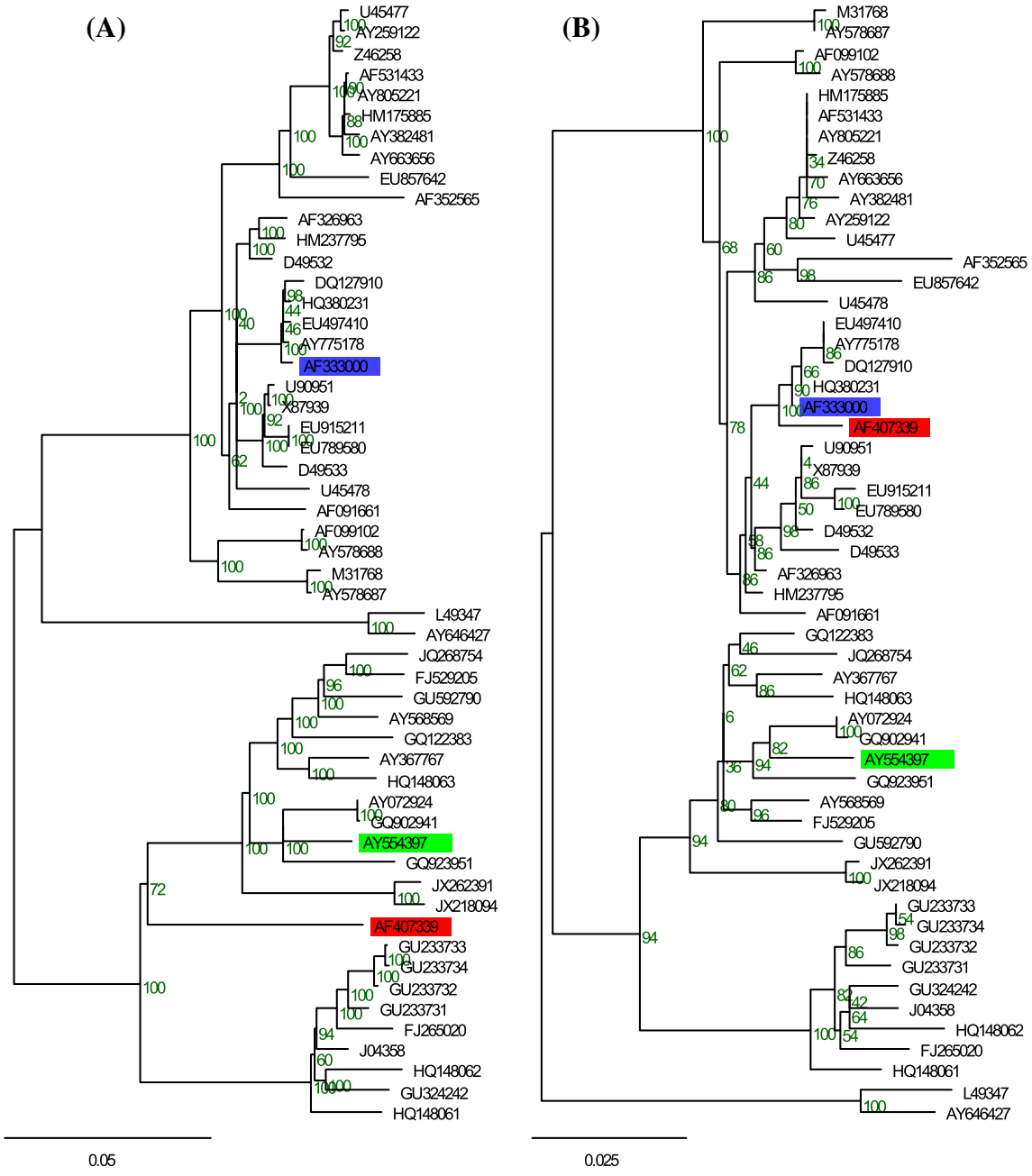

Figure 1 Neighbor-joining trees for the CSFV genomes, showing the evidence for the first major recombination event on the recombinant AF407339. The tree on subplot A was inferred from the major region, while the one on the subplot B was inferred from the minor region. All branch lengths are drawn to a scale of nucleotide substitutions per site. Numbers on the nodes indicate the bootstrapping support scores. The sequences marked with red, blue and green colors indicate the recombinant, the minor parent and major parent, respectively.

at the year 2007 and the limited usage of recombination methods (He et al., 2007). At that time, other CSFV sequences have become available online, but not been considered in the analyses in the previous study. These sequences included AF531433 (available since the year of 2002), AY663656 (available since the year of 2004), and AY805221 (available since the year of 2004). In particular, the exclusion of the significantly putative recombination parent AF531433 in the analysis can largely limit the identification sensitivity of recombination events in CSFV genomes. As seen in the present results (Table 2), AF531433 was found to be one of the parents for many recombinants. Moreover, in their study, only one method was utilized, SimPlot (Lole et al., 1999), therefore largely reducing the reliability and accuracy of the identification of recombination events. SimPlot only employed one method Bootscan for detecting recombination events. In contrast, in our study, we utilized all 7 statistical methods available in RDP program, plus the phylogenetic analysis to identify the

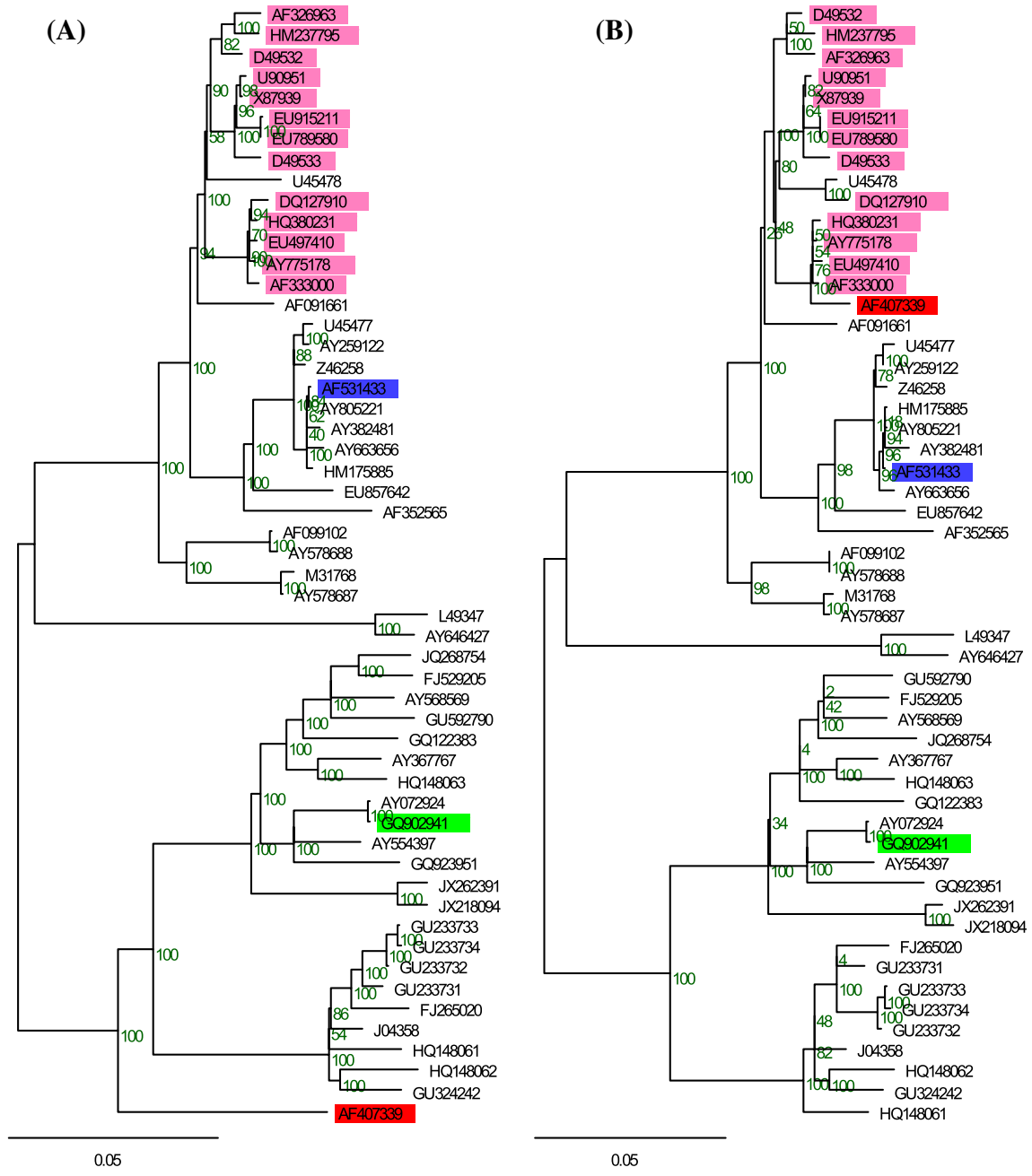

Figure 2 Neighbor-joining trees for the CSFV genomes, showing the evidence for the second major recombination event on the recombinant AF407339. The tree on subplot A was inferred from the major region, while the one on the subplot B was inferred from the minor region. All branch lengths are drawn to a scale of nucleotide substitutions per site. Numbers on the nodes indicate the bootstrapping support scores. The sequences marked with red, blue and green colors indicate the recombinant, the minor parent and major parent respectively. Sequences marked in pink indicate the ones with partial evidence of the same recombination event.

recombination parents. As such, we find no support of AF333000 as the minor parent of the recombination event.

At last, although the current recombination detection methods yielded high significance levels for the recombination events for CSFV, we have to acknowledge that they have the risk of false-positive detection of recombination (Leal et al., 2012). Further lab experiments might be required to verify these recombination events proposed in our study.

As the implications, the present study suggested that multiple statistical methods should be utilized so as to accurately identify the recombination events (Attoui et al., 2007; Shin et al., 2013). Finally, as seen, CSFV possesses relatively high frequency of homologous recombination events.

References

Attoui, H., Sailleau, C., Jaafar, F., Belhouchet, M., Biagini, P., Cantaloube, J., Micco, P., Mertens, P., Zientara, S., 2007. Complete nucleotide sequence of Middelburg virus, isolated from

the spleen of a horse with severe clinical disease in Zimbabwe. J. Gen. Virol. 88, 3078–3088.
Boni, M., Posada, D., Feldman, M., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.
Boni, M., Zhou, Y., Taubenberger, J., Holmes, E., 2008. Homologous recombination is very rare or absent in human influenza A virus. J. Virol. 82, 4807–4811.
Cao, H., Zhang, H., Cui, Y., 2012. Synonymous codon usage bias of E2 genes of classical swine fever virus. Israel J. Vet. Med. 67, 253–258.
Edgar, R., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Gibbs, M., Armstrong, J., Gibbs, A., 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.
Han, G., Liu, X., Li, S., 2008. Cross-species recombination in the haemagglutinin gene of canine distemper virus. Virus Res. 136, 198–201.
Han, G., Worobey, M., 2011. Homologous recombination in negative sense RNA viruses. Viruses 3, 1358–1373.
He, C., Ding, N., Chen, J., Li, Y., 2007. Evidence of natural recombination in classical swine fever virus. Virus Res. 126, 179–185.
He, C., Xie, Z., Han, G., Dong, J., Wang, D., Liu, J., Ma, L., Tang, X., Liu, X., Pang, Y., Li, G., 2009. Homologous recombination as an evolutionary force in the avian influenza A virus. Mol. Biol. Evol. 26, 177–187.
Leal, E., Villanova, F., Lin, W., Hu, F., Liu, Q., Liu, Y., Cui, S., 2012. Interclade recombination in porcine parvovirus strains. J. Gen. Virol. 93, 2692–2704.
Liu, X., Wu, C., Chen, A., 2010. Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. Arch. Virol. 155, 685–693.
Lole, K., Bollinger, R., Paranjape, R., Gadkari, D., Kulkarni, S., Novak, N., Ingersoll, R., Sheppard, H., Ray, S., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.
Martin, D., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.
Martin, D., Posada, D., Crandall, K., Williamson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses 21, 98–102.
Martin, D., Rybicki, E., 2010. RDP: detection of recombination amongst aligned sequences. Bioinformatics 26, 2462–2463.
Maynard Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.
McCarthy, A., Shaw, M., Goodman, S., 2007. Pathogen evolution and disease emergence in carnivores. Proc. R. Soc. B Biol. Sci. 274, 3165–3174.
Padidam, M., Sawyer, S., Fauquet, C., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.
Perez, L., Arce, H., Perera, C., Rosell, R., Frias, M., Percedo, M., Tarradas, J., Doinguez, P., Nuenez, J., Ganges, L., 2012. Positive selection pressure on the B/C domains of the E2-gene of classical swine fever virus in endemic areas under C-strain vaccination. Infect. Genet. Evol. 12, 1405–1412.
Posada, D., Crandall, K., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. 98, 13757–13762.
Shen, H., Pei, J., Bai, J., Zhao, M., Ju, C., Yi, L., Kang, Y., 2011. Genetic diversity and positive selection analysis of classical swine fever virus isolates in south China. Virus Genes 43, 234–242.
Shin, D., Richards, S., Alto, B., Bettinardi, D., Smartt, C., 2013. Genome sequence analysis of Dengue virus 1 isolated in Key West, Florida. PLoS One 8, e74582.
Tang, F., Pan, Z., Zhang, C., 2008. The selection pressure analysis of classical swine fever virus envelope protein genes Erns and E2. Virus Res. 131, 132–135.
Tao, P., Dai, L., Luo, M., Tang, F., Tien, P., Pan, Z., 2009. Analysis of synonymous codon usage in classical swine fever virus. Virus Genes 38, 104–112.
Van der Walt, E., Rybicki, E., Varsani, A., Polston, J., Billharz, R., Donaldson, L., Monjane, A., Martin, D., 2009. Rapid host adaptation by extensive recombination. J. Gen. Virol. 90, 734–746.
Worobey, M., 2000. Extensive homologous recombination among widely divergent TT viruses. J. Virol. 74, 7666–7670.
Wu, J., Yu, T., Bao, Q., Zhao, F., 2009. Evidence of extensive homologous recombination in the core genome of Rickettsia. Comp. Funct. Genomics 2009. http://dx.doi.org/10.1155/2009/510270.