Episodic adaptive diversification of classical swine fever virus RNA-dependent RNA polymerase NS5B

<table>
<thead>
<tr>
<th><strong>Journal:</strong></th>
<th>Canadian Journal of Microbiology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID:</strong></td>
<td>cjm-2015-0334.R1</td>
</tr>
<tr>
<td><strong>Manuscript Type:</strong></td>
<td>Article</td>
</tr>
<tr>
<td><strong>Date Submitted by the Author:</strong></td>
<td>30-Aug-2015</td>
</tr>
</tbody>
</table>
| **Complete List of Authors:** | Li, Yan; Sichuan Agricultural University, College of Animal Science and Technology  
Yang, Zexiao; Sichuan Agricultural University, College of Veterinary Medicine |
| **Keyword:** | Classical swine fever virus, RNA-dependent RNA polymerase, Episodic adaptation, Inter-group diversification |
Episodic adaptive diversification of classical swine fever virus

RNA-dependent RNA polymerase NS5B

Yan Li and Zexiao Yang

Y. Li. College of Animal Science and Technology, Sichuan Agricultural University, Wenjiang, P. R. China

Z. Yang. College of Veterinary Medicine, Sichuan Agricultural University, Wenjiang, P. R. China

Corresponding author: Yan Li (e-mail: liyan@sicau.edu.cn).
Abstract:

Classical swine fever virus (CSFV) is the pathogen that causes a highly infectious disease of pigs and has led to disastrous losses to pig farms and related industries. The RNA-dependent RNA polymerase (RdRp) NS5B is a central component of the replicase complex (RC) in some single stranded RNA (ssRNA) viruses, including CSFV. Based on genetic variation, the CSFV RdRps could be clearly divided into two major groups and a minor group, which is consistent with the phylogenetic relationships and virulence diversification of the CSFV isolates. However, the adaptive signature underlying such evolutionary profile of the polymerase and the virus is still an interesting open question. We analyzed the evolutionary trajectory of the CSFV RdRps over different timescales to evaluate the potential adaptation. We found that adaptive selection has driven the diversification of the RdRps between, but not within, CSFV major groups. Further, the major adaptive divergence related sites are located in the surfaces relevant to the interaction with other component(s) of RC and the entrance and exit of the template-binding channel. These results might shed some light on the nature of the RdRp in virulence diversification of CSFV groups.
Keywords:

Classical swine fever virus,

RNA-dependent RNA polymerase,

Episodic adaptation,

Inter-group diversification.
Manuscript text

Introduction

Classical swine fever virus (CSFV) is the causative agent of classical swine fever (CSF), which is a highly contagious disease of the domestic pig as well as wild boar. Outbreaks of CSF in domestic pigs resulted in devastating losses to the pig industry (Edwards et al. 2000; Meuwissen et al. 1999; Vandeputte and Chappuis 1999). CSFV makes up the genus Pestivirus of the family Flaviviridae along with the closely related viruses Bovine viral diarrhoea virus genotypes 1 (BVDV-1) and 2 (BVDV-2), and Border disease virus (BDV) (King et al. 2011). Besides Pestivirus, Flaviviridae consists of the genera Hepacivirus (e.g., hepatitis C virus (HCV)) and Flavivirus (e.g., dengue and West Nile (WNV) viruses).

CSFV possesses a single-stranded, positive-sense RNA genome. The virus-encoded RNA-dependent RNA polymerase (RdRp) is the central component in the replication of RNA viruses, including CSFV (Hansen et al. 1997), which is therefore an important target for antiviral drug development (Baginski et al. 2000; Musiu et al. 2014; Paeshuyse et al. 2006; Sun et al. 2003; Vrancken et al. 2009a; Vrancken et al. 2009b; Vrancken et al. 2008).

Based on genetic diversity, CSFV isolates could be subdivided into two major groups (1 and 2) with a rare group (3) (Lowings et al. 1996; Postel et al. 2012). It is noteworthy that the reported highly virulent strains and low and moderately
virulent isolates cluster in group 1 and group 2 (Li et al. 2006; Topfer et al. 2013), respectively. Low and moderately virulent isolates cause mild or often chronic pathological and clinical diseases, while highly virulent isolates cause an acute and hemorrhagic fever with nearly 100% mortality (Leifer et al. 2013; Moennig et al. 2003). Further, in terms of the function of RdRp, exchange of their coding sequences between CSFV isolates of distinct virulence can significantly influence the replication efficiency (Risager et al. 2013). The virulence and replication efficiency of a pathogen can be reflection of its fitness in a host. By this logic, these results indicate that the relevance between virulence divergence of CSFV groups and potential adaptive changes in RdRps should be considered.

An important step forward in detecting and understanding the functional adaptive mechanisms in CSFV RdRp includes the description of its evolutionary trajectories. In this study we analyze the variation of the polymerase over the evolutionary timescales of CSFV. Our results suggest that two major groups had experienced functional divergence at an early stage of their split. In intra-group evolution, however, no signature of the adaptive diversity has been found within each of the major groups of RdRp. This study shed some light on the evolutionary nature of RdRp in CSFV and the molecular adaptive mechanisms underlying the virulence diversification of the major groups.
Materials and methods

Sequence data collection and alignment

The complete coding sequences of RdRp (2,154 nt) of field CSFV isolates were collected from the NCBI database. Because recombination may not be extensive in natural populations of CSFV (He et al. 2007; Lowings et al. 1996; Postel et al. 2012), the sequences of rare isolates assigned to recombinants were not used in the present study. Forty-one complete coding sequences of RdRp were finally used for following evolutionary analyses (Table S1). A coding sequence of RdRp from a closely related species of BDV was chosen as an outgroup (Table S1). For detecting the influence of a potential expansion in CSFV population size, the complete coding sequences of NS3 (2,049 nt) and NS4B (1,041 nt) were collected from corresponding isolates (Meyers and Thiel 1996; Moennig and Plagemann 1992; Thiel et al. 1991). The coding sequences of RdRp, NS3, and NS4B were aligned in ClustalW (Thompson et al. 1997) respectively and checked manually.

Reconstruction of gene tree of RdRp

Maximum likelihood (ML) analysis (Saitou and Nei 1987) on the dataset in MEGA5 (Tamura et al. 2011) was used to infer relationships among the RdRp coding sequences. The General Time Reversible model (Lanave et al. 1984) of evolution allowing gamma distribution plus a proportion of invariant sites was implemented in ML analyses on the dataset. The support for the tree was evaluated.
The MK test of CSFV RdRp coding sequences

We adopted the McDonald and Kreitman (MK) test to identify adaptive evolution of RdRp, NS3, and NS4B in CSFV (McDonald and Kreitman 1991). If synonymous and replacement differences were due to neutral mutations alone, the ratio of synonymous to replacement divergence between lineages would be expected to be the same as the ratio of synonymous to replacement polymorphism within lineages. A Fisher's exact test (two tailed) of independence (McDonald 2009) was calculated and significant deviation from this expectation can therefore be used to reject the neutral mutation hypothesis.

Test for the functional divergence of RdRp coding sequences between major groups

We adopted the statistical method Gu-2006 in DIVERGE 3.0 to detect the functional divergence at the early stage after the groups or subgroups split from a common ancestor (type II functional divergence) (Gu 2001, 2006). Type II functional divergence means that the evolutionary pattern of amino acids is conserved within both groups but their physiochemical properties are remarkably different. The method Gu-2006 calculates the coefficient of type II functional divergence $\theta_{II}$ between the generated lineages (Gu 2006), and the Z-score test is performed for statistical evaluation based on the estimate of $\theta_{II}$ and its standard error.
The null hypothesis is $\theta_{II} = 0$, which means that the evolutionary rate is virtually the same between member lineages. Rejection of the null hypothesis means that a shift of amino acid physiochemical properties is likely to have occurred between these lineages. Further, the posterior probability is used to predict the amino acid residues that could be responsible for these functional differences with a posterior probability cutoff > 0.75 (Gu 2006).

**The maximum likelihood analysis of RdRp coding sequences within major groups**

We tested for positive selection at individual sites within each groups with CODEML in the PAML package (Yang 2007). The program CODEML adopts codon-based models (M) for studying the evolutionary mechanism of amino acid sites by comparing synonymous and nonsynonymous substitution rates (Goldman and Yang 1994). For each candidate lineage, we calculated likelihood values with three pairs of implemented models (M0 vs. M2, M7 vs. M8, and M8a vs. M8). Null models M0, M7, and M8a assumed that all codons evolved neutrally or under purifying selection ($d_N / d_S$ values <= 1); alternative models M2 and M8 allowed a proportion of sites to be under positive selection with $d_N / d_S > 1$. For each paired comparison of models, a likelihood ratio test (LRT) was used to estimate the difference between them (Nielsen and Yang 1998; Swanson et al. 2003).
Results and discussion

Group division of RdRp in CSFV

We reconstructed a gene tree to identify the general relationships of RdRp complete coding sequences from field CSFV isolates used in the present study. The gene tree comprises two well-supported major groups and one minor group (Fig. 1). Our clade structure is consistent with that found in previous studies (Lowings et al. 1996; Postel et al. 2012) and we adopted the same group numbering system as that used by Postel et al. (Postel et al. 2012) (see also Table S1). The major groups (groups 1 and 2) consist of 16 and 24 isolates, respectively. Group 3 is represented by only one isolate. Also consistent with the phylogenetic trees as shown previously (Postel et al. 2012), our group 2 comprises two well supported lineages, one includes subgroups 2.1 and 2.2, and the other includes subgroup 2.3. These lineages comprise 14 and 10 sequences, respectively (Fig. 1). Further analyses were performed for the two major groups only, since only one group 3 sequence could be retrieved.

Adaptive diversity of RdRp at the early stage of the split from a common ancestor

Using the MK test with the RdRp coding sequences from the two clearly disparate groups (group 1 and group 2) (Fig. 1), we detected the occurrence of adaptive evolution at their split from a common ancestor. A Fisher's exact test of
independence (McDonald 2009) showed that the ratio of fixed amino acid replacement versus fixed synonymous differences was greater than that of replacement polymorphisms versus synonymous ones with significance (Probability = 0.02) (Table 1). We suggested from this result that the RdRps of group 1 and/or group 2 might have experienced adaptive evolution in the dividing process of the CSFV progenitor into groups.

The MK test is generally robust to most demographic assumptions (see (Aguileta et al. 2009; Eyre-Walker 2006) and references therein). The exception is that the slightly deleterious nonsynonymous mutations can be fixed during a rapid expansion in population size and thereby lead to artifactual evidence of adaptive evolution (Eyre-Walker 2002). Natural viral population expansion is difficult to measure thus its influence on viral evolution is largely unknown (reviewed in (Domingo et al. 2006)). With regard to CSFV, since there is no detailed information on the relevant factors influencing the evolution of CSFV (Bjorklund et al. 1999; Lowings et al. 1996), the appearance and role of expansion in its history is still unclear.

A demographic factor, such as expansion in population size, should have similar effects on the different genomic regions. Therefore, we conducted the MK test for NS3 and NS4B coding sequences, which are shared by Flaviviridae viruses, to validate the impact of the potential expansion in population size. For NS4B, the Fisher's exact test of independence indicated that the ratio of replacement to synonymous divergence fixed between groups is not statistically different from that of replacement to synonymous polymorphisms (Table 1). Result on NS3 showed
an excess of replacement polymorphisms within groups with significant probability

\((P < 0.01)\) (Table 1). The opposite results of the MK test on the NS5B, NS3, and

NS4B coding sequences suggested that the effect of the potential population

expansion could be roughly cancelled out - that is, the evidence of adaptive selection

in CSFV RdRp is unlikely to be a consequence of an expansion in population size.

We then attempted to identify the changes at the individual amino acid sites

between group 1 and group 2 RdRps that were relevant to their divergence. Using

the method Gu-2006 in DIVERGE 3 (Gu 2006; Gu et al. 2013) (see Materials and

Methods), we calculated the coefficient of type II functional divergence \((\theta_{II})\) between

these groups (Table 2). The Z-score test indicated that type II functional

divergence \(\theta_{II}\) is larger than 0 with a marginally significant probability \((P > 0.90)\).

Further, the site-specific score based on the posterior probability ratio showed that

eleven amino acid sites are related to the functional divergence with a probability

over 0.75 (Table 2). The amino acids at these predicted sites exhibited difference of

physiochemical properties (e.g., charge, hydrophobic, etc.) between the two major

groups, but were highly conserved within each of these groups (Table S2). Since

the isolates were collected globally over the past several decades (Ji et al. 2014;

Lowings et al. 1996), such evolutionary signature at these sites had most likely

occurred at the early stage of the group split. Notably, because of the high mutation

rate of CSFV (Lowings et al. 1996; Zhang et al. 2011) and manifest existence of

quasispecies rather than a defined genomic sequence in viruses (reviewed in

(Domingo et al. 2012)) including CSFV (Topfer et al. 2013), the assumption of type
II functional divergence may be over-strict in the present case and tend to underestimate the $\theta_{II}$ and the functional divergence related substitutions. Despite this, consistent with the insight from the MK test, the above results support the hypothesis that functional divergence between the two major groups is likely to have occurred.

Evolution of RdRp within groups 1 and 2

Because different selective pressures probably played roles in shaping the variation of RdRp over different timescales, we further evaluated the evolution of the polymerase within group 1 and groups 2, independently. Adaptive evolution typically occurs at few sites since most amino acids in a protein are under functional constraints (LI 1997). Hence, for group 1 RdRps coding sequences, we detected the individual sites that might be under adaptive selection using a codon-based test in a maximum likelihood framework (Yang 1997, 2007). The likelihood ratio test (LRT) results showed that the model for adaptive selection (M2, $d_N / d_S > 1$) fits the data significantly better than the neutral or purifying selection models (M0, $d_N / d_S \leq 1$) (Table S3). However, the difference between M8 and M7, and M8 and M8a are not significant, which suggested that adaptive selection should not have influenced the variation within group 1 RdRp. Further, both inferences of models M2 and M8 indicated that no amino acid sites exhibit positive selection based on $d_N / d_S$ ratios with $> 0.95$ probability level (Table S3).

Since there are two well supported subgroups (comprising 2.3; 2.1 and 2.2) within
group 2 RdRp (Fig. 1), we first performed the MK test and method Gu-2006 of
DIVERGE 3 to measure whether this division might be the result of adaptive
evolution (Gu 2006; Gu et al. 2013; McDonald and Kreitman 1991). The
combined results suggested that no functional divergence took place with the split
into these subgroups (Table 1 and 2). We then repeated the codon-based analysis to
detect adaptive variation within the whole of group 2. Although the LRT of M2 vs.
M0 showed that the model for adaptive selection (M2) fits the data significantly
better than the neutral or purifying selection model (M0), the difference in the
comparisons M8 vs. M7 and M8 vs. M8a are not significant (Table S3). For models
M2 and M8, no amino acid site had been subject to adaptive selection based on the
dN / dS ratio at P > 0.95 level. Thus, the accumulated statistical results of the MK
test, method Gu-2006, and the maximum likelihood analysis consistently indicated
that the no amino acid site was identified as subject to adaptive selection within
group 2.

In summary, these evidences led to the suggestion that the evolution of the RdRp
might be episodic, with adaptive diversification in the inter-group split but not
within the two major groups. The episodic diversification pattern of the RdRp is
compatible with the reported virulence framework of different isolates from different
groups (Li et al. 2006; Topfer et al. 2013). In spite of the consistent pattern,
however, it is important to note that other proteins of CSFV have been shown to be
involved in virulence determination (Leifer et al. 2013). Therefore, further studies
will be necessary to fully comprehend the mechanisms that govern the virulence of
CSFV.

The inter-group functional divergence related sites lie in N-terminal (sites 34, 60, and 89), fingers (sites 202 and 367), and thumb (sites 502, 574, 619, 622, 688, and 712) domains based on the three dimensional structures of BVDV homologue (Choi et al. 2004). Functionally, the majority of the functional divergence related sites cluster in two interesting regions: (1) on or nearby surfaces implicated in the interaction between NS5B and another RC component NS3, and enhance the RdRp activity of the polymerase (sites 60, 89, 619, and 622) (Wang et al. 2011), (2) at the entrance (sites 34, 60, 89, 502, 619, and 622) (Butcher et al. 2001; Choi et al. 2004; Wang et al. 2011) and exit (sites 574, 688, and 712) (Choi et al. 2004) of the template-binding channel. As previously reported, chimeric RNAs, in which RdRps are substituted with the equivalent sequences from different CSFV strains, present significantly distinct efficiency of replication (Risager et al. 2013). Thus, one potential explanation for this distribution is that the changes might contribute to the alternation of interacting behavior between RdRp and NS3 and/or template, which consequently results in such functional diversification between two major groups. However, the MK test on NS3 indicated that the ratio of replacement versus fixed synonymous differences was significantly less than that of replacement versus synonymous polymorphisms and few amino acid replacements were fixed between groups (Table 1). These results inferred that no adaptive divergence took place on NS3 in the dividing process of the CSFV progenitor into groups. In this scenario, the diversification on NS5B would most likely influence the interaction
between the polymerase and template and/or other RC component(s). Alternatively, it is possible that such conversions constitute a functional adaptation that counters host immune response that block the interaction between RC component(s), and/or polymerase and template.

In conclusion, our analyses provided evidence that adaptive divergence has occurred at the early stage of the group split, but not within each group. In addition, the majority of adaptively selected sites are located in two important functional regions. Our results draw parallels among the current knowledge about virulence divergence between CSFV isolates of two groups, inter-group evolutionary diversification of the RdRps, and the cluster location of inter-group divergent sites in key functional regions. Certainly, future structural and functional studies are needed to determine the effects of these adaptive variations. If the hypotheses are supported, it might shed more insights into the nature of RdRp in replication.

Acknowledgements

We thank Dr. Judith H. Robins for her helpful comments and English editing of the manuscript.
**References**


Domingo, E., Roossinck, M.J., and Schneider, W.L. 2006. Mutant Clouds and Occupation of Sequence Space in Plant RNA Viruses. In Quasispecies: Concept and


Antiviral research 77(2): 114-119.


### Table 1. MK test on the RdRp NS5B and NS4B coding sequences.

<table>
<thead>
<tr>
<th>Coding sequence</th>
<th>Fixed</th>
<th>Polymorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NS5B</strong>&lt;br&gt;Group 1 vs. Group 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replacement</td>
<td>20</td>
<td>143</td>
</tr>
<tr>
<td>Synonymous</td>
<td>41</td>
<td>601</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>0.02*</td>
</tr>
<tr>
<td><strong>NS3</strong>&lt;br&gt;Group 1 vs. Group 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replacement</td>
<td>2</td>
<td>158</td>
</tr>
<tr>
<td>Synonymous</td>
<td>48</td>
<td>576</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>&lt; 0.01**</td>
</tr>
<tr>
<td><strong>NS4B</strong>&lt;br&gt;Group 1 vs. Group 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replacement</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>Synonymous</td>
<td>21</td>
<td>323</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>0.78</td>
</tr>
<tr>
<td><strong>NS5B</strong>&lt;br&gt;Subgroup 2.1 or 2.2 vs. Subgroup 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replacement</td>
<td>12</td>
<td>111</td>
</tr>
<tr>
<td>Synonymous</td>
<td>49</td>
<td>499</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>0.86</td>
</tr>
</tbody>
</table>

* Single and double asterisks correspond to two-tailed *Probability* < 0.05 or *Probability* < 0.01 (Fisher's exact test).
### Table 2. The estimates of the type-II functional divergences between two major groups and subgroups of group 2.

<table>
<thead>
<tr>
<th>Pair of groups or subgroups</th>
<th>Method</th>
<th>Estimate (Z-score)</th>
<th>Divergence-related sites identified with posterior probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 vs. Group 2</td>
<td>Gu-2006</td>
<td>$\theta_{II}^a = 0.029 \pm 0.020$ (1.45)</td>
<td>34, 60, 89, 202, 367, 502, 574, 619, 622, 688, 712 (at $P_{II} \geq 0.75$)</td>
</tr>
<tr>
<td>Subgroup 2.1 or 2.2 vs. Subgroup 2.3</td>
<td>Gu-2006</td>
<td>$\theta_{II} = 0.004 \pm 0.017$ (0.24)</td>
<td>None (at $P_{II} \geq 0.75$)</td>
</tr>
</tbody>
</table>

$^a\theta_{II}$, coefficient of type-II functional divergence. See Materials and Methods for details on functional divergence test.
Figure legends

Fig. 1. Maximum likelihood gene tree for the RdRp coding sequences extracted from each of the 41 genomes represented in Table S1. The numbering of groups and subgroups follows that of previous results (Postel et al. 2012) and see also Table S1. The tips are labeled with GenBank accession numbers and isolate names. The bootstrap values below 90% are not reported.
Appendix A. Supplementary data

Three supplementary tables are available:

Table S1. CSFV and BDV sequences used for this study.

Table S2. The amino acids at the Type-II divergence-related sites of the group 1 and group 2 RdRps.

Table S3. Maximum likelihood analysis of RdRps within major groups of RdRp coding sequence.
Fig. 1