Canadian Journal of Microbiology
Revue canadienne de microbiologie

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

Journal:	<i>Canadian Journal of Microbiology</i>
Manuscript ID	cjm-2015-0334.R1
Manuscript Type:	Article
Date Submitted by the Author:	30-Aug-2015
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

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

3
4
5 Yan Li and Zexiao Yang

6
7
8 **Y. Li.** College of Animal Science and Technology, Sichuan Agricultural University,
9 Wenjiang, P. R. China

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

12
13
14 **Corresponding author:** Yan Li (e-mail: liyan@sicau.edu.cn).
15

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.

35 **Keywords:**

36 Classical swine fever virus,

37 RNA-dependent RNA polymerase,

38 Episodic adaptation,

39 Inter-group diversification.

40

Draft

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

63 virulent isolates cluster in group 1 and group 2 (Li et al. 2006; Topfer et al. 2013),
64 respectively. Low and moderately virulent isolates cause mild or often chronic
65 pathological and clinical diseases, while highly virulent isolates cause an acute and
66 hemorrhagic fever with nearly 100% mortality (Leifer et al. 2013; Moennig et al.
67 2003). Further, in terms of the function of RdRp, exchange of their coding
68 sequences between CSFV isolates of distinct virulence can significantly influence
69 the replication efficiency (Risager et al. 2013). The virulence and replication
70 efficiency of a pathogen can be reflection of its fitness in a host. By this logic,
71 these results indicate that the relevance between virulence divergence of CSFV
72 groups and potential adaptive changes in RdRps should be considered.

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

82

83

84

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

107 by bootstrap analysis (1000 bootstraps).

108

109 **The MK test of CSFV RdRp coding sequences**

110 We adopted the McDonald and Kreitman (MK) test to identify adaptive evolution

111 of RdRp, NS3, and NS4B in CSFV (McDonald and Kreitman 1991). If

112 synonymous and replacement differences were due to neutral mutations alone, the

113 ratio of synonymous to replacement divergence between lineages would be expected

114 to be the same as the ratio of synonymous to replacement polymorphism within

115 lineages. A Fisher's exact test (two tailed) of independence (McDonald 2009) was

116 calculated and significant deviation from this expectation can therefore be used to

117 reject the neutral mutation hypothesis.

118

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

121 We adopted the statistical method Gu-2006 in DIVERGE 3.0 to detect the

122 functional divergence at the early stage after the groups or subgroups split from a

123 common ancestor (type II functional divergence) (Gu 2001, 2006). Type II

124 functional divergence means that the evolutionary pattern of amino acids is

125 conserved within both groups but their physiochemical properties are remarkably

126 different. The method Gu-2006 calculates the coefficient of type II functional

127 divergence θ_{II} between the generated lineages (Gu 2006), and the Z-score test is

128 performed for statistical evaluation based on the estimate of θ_{II} and its standard error

129 (θ_{II} SE) (Gu 2001; Gu et al. 2013). The null hypothesis is $\theta_{II} = 0$, which means that
130 the evolutionary rate is virtually the same between member lineages. Rejection of
131 the null hypothesis means that a shift of amino acid physiochemical properties is
132 likely to have occurred between these lineages. Further, the posterior probability is
133 used to predict the amino acid residues that could be responsible for these functional
134 differences with a posterior probability cutoff > 0.75 (Gu 2006).

135

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

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

149

150

151 **Results and discussion**

152

153 **Group division of RdRp in CSFV**

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

167

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

170 Using the MK test with the RdRp coding sequences from the two clearly
171 disparate groups (group 1 and group 2) (Fig. 1), we detected the occurrence of
172 adaptive evolution at their split from a common ancestor. A Fisher's exact test of

173 independence (McDonald 2009) showed that the ratio of fixed amino acid
174 replacement versus fixed synonymous differences was greater than that of
175 replacement polymorphisms versus synonymous ones with significance (*Probability*
176 = 0.02) (Table 1). We suggested from this result that the RdRps of group 1 and/or
177 group 2 might have experienced adaptive evolution in the dividing process of the
178 CSFV progenitor into groups.

179 The MK test is generally robust to most demographic assumptions (see (Aguileta
180 et al. 2009; Eyre-Walker 2006) and references therein). The exception is that the
181 slightly deleterious nonsynonymous mutations can be fixed during a rapid expansion
182 in population size and thereby lead to artifactual evidence of adaptive evolution
183 (Eyre-Walker 2002). Natural viral population expansion is difficult to measure
184 thus its influence on viral evolution is largely unknown (reviewed in (Domingo et al.
185 2006)). With regard to CSFV, since there is no detailed information on the relevant
186 factors influencing the evolution of CSFV (Bjorklund et al. 1999; Lowings et al.
187 1996), the appearance and role of expansion in its history is still unclear.

188 A demographic factor, such as expansion in population size, should have similar
189 effects on the different genomic regions. Therefore, we conducted the MK test for
190 NS3 and NS4B coding sequences, which are shared by *Flaviviridae* viruses, to
191 validate the impact of the potential expansion in population size. For NS4B, the
192 Fisher's exact test of independence indicated that the ratio of replacement to
193 synonymous divergence fixed between groups is not statistically different from that
194 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 (θ_{II}) between these groups (Table 2). The Z-score test indicated that type II functional divergence θ_{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

217 II functional divergence may be over-strict in the present case and tend to
218 underestimate the θ_{II} and the functional divergence related substitutions. Despite
219 this, consistent with the insight from the MK test, the above results support the
220 hypothesis that functional divergence between the two major groups is likely to have
221 occurred.

222

223 **Evolution of RdRp within groups 1 and 2**

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

238 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 d_N / d_S 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

261 CSFV.

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

283 between the polymerase and template and/or other RC component(s). Alternatively,
284 it is possible that such conversions constitute a functional adaptation that counters
285 host immune response that block the interaction between RC component(s), and/or
286 polymerase and template.

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

296

297

298 **Acknowledgements**

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

301

302 **References**

303

- 304 Aguileta, G., Refregier, G., Yockteng, R., Fournier, E., and Giraud, T. 2009. Rapidly
305 evolving genes in pathogens: methods for detecting positive selection and examples
306 among fungi, bacteria, viruses and protists. *Infect Genet Evol* **9**(4): 656-670.
- 307 Baginski, S.G., Pevear, D.C., Seipel, M., Sun, S.C., Benetatos, C.A., Chunduru, S.K.,
308 Rice, C.M., and Collett, M.S. 2000. Mechanism of action of a pestivirus antiviral
309 compound. *Proceedings of the National Academy of Sciences of the United States of*
310 *America* **97**(14): 7981-7986.
- 311 Bjorklund, H., Lowings, P., Stadejek, T., Vilcek, S., Greiser-Wilke, I., Paton, D., and
312 Belak, S. 1999. Phylogenetic comparison and molecular epidemiology of classical
313 swine fever virus. *Virus genes* **19**(3): 189-195.
- 314 Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., and Stuart, D.I. 2001. A
315 mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**(6825):
316 235-240.
- 317 Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevear, D.C., and
318 Rossmann, M.G. 2004. The structure of the RNA-dependent RNA polymerase from
319 bovine viral diarrhea virus establishes the role of GTP in de novo initiation.
320 *Proceedings of the National Academy of Sciences of the United States of America*
321 **101**(13): 4425-4430.
- 322 Domingo, E., Roossinck, M.J., and Schneider, W.L. 2006. Mutant Clouds and
323 Occupation of Sequence Space in Plant RNA Viruses. *In* *Quasispecies: Concept and*

- 324 Implications for Virology. Springer Berlin Heidelberg. pp. 337-348.
- 325 Domingo, E., Sheldon, J., and Perales, C. 2012. Viral quasispecies evolution.
- 326 Microbiol Mol Biol Rev **76**(2): 159-216.
- 327 Edwards, S., Fukusho, A., Lefevre, P.C., Lipowski, A., Pejsak, Z., Roehe, P., and
- 328 Westergaard, J. 2000. Classical swine fever: the global situation. Veterinary
- 329 microbiology **73**(2-3): 103-119.
- 330 Eyre-Walker, A. 2002. Changing effective population size and the
- 331 McDonald-Kreitman test. Genetics **162**(4): 2017-2024.
- 332 Eyre-Walker, A. 2006. The genomic rate of adaptive evolution. Trends in ecology &
- 333 evolution **21**(10): 569-575.
- 334 Goldman, N., and Yang, Z. 1994. A codon-based model of nucleotide substitution for
- 335 protein-coding DNA sequences. Molecular biology and evolution **11**(5): 725-736.
- 336 Gu, X. 2001. Maximum-likelihood approach for gene family evolution under
- 337 functional divergence. Molecular biology and evolution **18**(4): 453-464.
- 338 Gu, X. 2006. A simple statistical method for estimating type-II (cluster-specific)
- 339 functional divergence of protein sequences. Molecular biology and evolution **23**(10):
- 340 1937-1945.
- 341 Gu, X., Zou, Y., Su, Z., Huang, W., Zhou, Z., Arendsee, Z., and Zeng, Y. 2013. An
- 342 update of DIVERGE software for functional divergence analysis of protein family.
- 343 Molecular biology and evolution **30**(7): 1713-1719.
- 344 Hansen, J.L., Long, A.M., and Schultz, S.C. 1997. Structure of the RNA-dependent
- 345 RNA polymerase of poliovirus. Structure **5**(8): 1109-1122.

- 346 He, C.Q., Ding, N.Z., Chen, J.G., and Li, Y.L. 2007. Evidence of natural
347 recombination in classical swine fever virus. *Virus research* **126**(1-2): 179-185.
- 348 Ji, W., Niu, D.D., Si, H.L., Ding, N.Z., and He, C.Q. 2014. Vaccination influences
349 the evolution of classical swine fever virus. *Infect Genet Evol* **25**: 69-77.
- 350 King, A.M.Q., Adams, M.J., Lefkowitz, E.J., and Carstens, E.B. 2011. *Virus*
351 *taxonomy: classification and nomenclature of viruses: Ninth Report of the*
352 *International Committee on Taxonomy of Viruses*. Elsevier.
- 353 Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for
354 calculating evolutionary substitution rates. *Journal of molecular evolution* **20**(1):
355 86-93.
- 356 Leifer, I., Ruggli, N., and Blome, S. 2013. Approaches to define the viral genetic
357 basis of classical swine fever virus virulence. *Virology* **438**(2): 51-55.
- 358 LI, W.-H. 1997. *Molecular evolution*. . Sinauer Associates, Sunderland, MA.
- 359 Li, X., Xu, Z., He, Y., Yao, Q., Zhang, K., Jin, M., Chen, H., and Qian, P. 2006.
360 Genome comparison of a novel classical swine fever virus isolated in China in 2004
361 with other CSFV strains. *Virus genes* **33**(2): 133-142.
- 362 Lowings, P., Ibata, G., Needham, J., and Paton, D. 1996. Classical swine fever virus
363 diversity and evolution. *The Journal of general virology* **77 (Pt 6)**: 1311-1321.
- 364 McDonald, J.H. 2009. *Handbook of biological statistics*. Sparky House Publishing
365 Baltimore, MD.
- 366 McDonald, J.H., and Kreitman, M. 1991. Adaptive protein evolution at the Adh
367 locus in *Drosophila*. *Nature* **351**(6328): 652-654.

- 368 Meuwissen, M.P., Horst, S.H., Huirne, R.B., and Dijkhuizen, A.A. 1999. A model to
369 estimate the financial consequences of classical swine fever outbreaks: principles
370 and outcomes. *Preventive veterinary medicine* **42**(3-4): 249-270.
- 371 Meyers, G., and Thiel, H.J. 1996. Molecular characterization of pestiviruses.
372 *Advances in virus research* **47**: 53-118.
- 373 Moennig, V., Floegel-Niesmann, G., and Greiser-Wilke, I. 2003. Clinical signs and
374 epidemiology of classical swine fever: a review of new knowledge. *Vet J* **165**(1):
375 11-20.
- 376 Moennig, V., and Plagemann, P.G. 1992. The pestiviruses. *Advances in virus research*
377 **41**: 53-98.
- 378 Musiu, S., Purstinger, G., Stallinger, S., Vrancken, R., Haegeman, A., Koenen, F.,
379 Leyssen, P., Froeyen, M., Neyts, J., and Paeshuyse, J. 2014. Substituted
380 2,6-bis(benzimidazol-2-yl)pyridines: a novel chemical class of pestivirus inhibitors
381 that targets a hot spot for inhibition of pestivirus replication in the RNA-dependent
382 RNA polymerase. *Antiviral research* **106**: 71-79.
- 383 Nielsen, R., and Yang, Z. 1998. Likelihood models for detecting positively selected
384 amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**(3):
385 929-936.
- 386 Paeshuyse, J., Leyssen, P., Mabery, E., Boddeker, N., Vrancken, R., Froeyen, M.,
387 Ansari, I.H., Dutartre, H., Rozenski, J., Gil, L.H., Letellier, C., Lanford, R., Canard,
388 B., Koenen, F., Kerkhofs, P., Donis, R.O., Herdewijn, P., Watson, J., De Clercq, E.,
389 Puerstinger, G., and Neyts, J. 2006. A novel, highly selective inhibitor of pestivirus

- 390 replication that targets the viral RNA-dependent RNA polymerase. Journal of
391 virology **80**(1): 149-160.
- 392 Postel, A., Schmeiser, S., Bernau, J., Meindl-Boehmer, A., Pridotkas, G., Dirbakova,
393 Z., Mojzis, M., and Becher, P. 2012. Improved strategy for phylogenetic analysis of
394 classical swine fever virus based on full-length E2 encoding sequences. Veterinary
395 research **43**: 50.
- 396 Risager, P.C., Fahnoe, U., Gullberg, M., Rasmussen, T.B., and Belsham, G.J. 2013.
397 Analysis of classical swine fever virus RNA replication determinants using replicons.
398 The Journal of general virology **94**(Pt 8): 1739-1748.
- 399 Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for
400 reconstructing phylogenetic trees. Molecular biology and evolution **4**(4): 406-425.
- 401 Sokal, R.R., and Rohlf, F.J. 1981. Biometry W. H. Freeman, San Francisco.
- 402 Sun, J.H., Lemm, J.A., O'Boyle, D.R., 2nd, Racela, J., Colonno, R., and Gao, M.
403 2003. Specific inhibition of bovine viral diarrhea virus replicase. Journal of virology
404 **77**(12): 6753-6760.
- 405 Swanson, W.J., Nielsen, R., and Yang, Q. 2003. Pervasive adaptive evolution in
406 mammalian fertilization proteins. Molecular biology and evolution **20**(1): 18-20.
- 407 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011.
408 MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
409 evolutionary distance, and maximum parsimony methods. Molecular biology and
410 evolution **28**(10): 2731-2739.
- 411 Thiel, H.J., Stark, R., Weiland, E., Rumenapf, T., and Meyers, G. 1991. Hog cholera

- 412 virus: molecular composition of virions from a pestivirus. *Journal of virology* **65**(9):
413 4705-4712.
- 414 Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997.
415 The CLUSTAL_X windows interface: flexible strategies for multiple sequence
416 alignment aided by quality analysis tools. *Nucleic acids research* **25**(24): 4876-4882.
- 417 Topfer, A., Hoper, D., Blome, S., Beer, M., Beerenwinkel, N., Ruggli, N., and Leifer,
418 I. 2013. Sequencing approach to analyze the role of quasispecies for classical swine
419 fever. *Virology* **438**(1): 14-19.
- 420 Vandeputte, J., and Chappuis, G. 1999. Classical swine fever: the European
421 experience and a guide for infected areas. *Revue scientifique et technique*
422 (International Office of Epizootics) **18**(3): 638-647.
- 423 Vrancken, R., Haegeman, A., Dewulf, J., Paeshuyse, J., Puerstinger, G., Tignon, M.,
424 Le Potier, M.F., Neyts, J., and Koenen, F. 2009a. The reduction of CSFV
425 transmission to untreated pigs by the pestivirus inhibitor BPIP: a proof of concept.
426 *Veterinary microbiology* **139**(3-4): 365-368.
- 427 Vrancken, R., Haegeman, A., Paeshuyse, J., Puerstinger, G., Rozenski, J., Wright, M.,
428 Tignon, M., Le Potier, M.F., Neyts, J., and Koenen, F. 2009b. Proof of concept for
429 the reduction of classical swine fever infection in pigs by a novel viral polymerase
430 inhibitor. *The Journal of general virology* **90**(Pt 6): 1335-1342.
- 431 Vrancken, R., Paeshuyse, J., Haegeman, A., Puerstinger, G., Froeyen, M., Herdewijn,
432 P., Kerkhofs, P., Neyts, J., and Koenen, F. 2008. Imidazo[4,5-c]pyridines inhibit the
433 in vitro replication of the classical swine fever virus and target the viral polymerase.

- 434 Antiviral research **77**(2): 114-119.
- 435 Wang, Y., Zhu, Z., Wang, P., Yu, J., Wan, L., Chen, J., and Xiao, M. 2011.
- 436 Characterisation of interaction between NS3 and NS5B protein of classical swine
- 437 fever virus by deletion of terminal sequences of NS5B. Virus research **156**(1-2):
- 438 98-106.
- 439 Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum
- 440 likelihood. Comput Appl Biosci **13**(5): 555-556.
- 441 Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular
- 442 biology and evolution **24**(8): 1586-1591.
- 443 Zhang, H., Cao, H.W., Wu, Z.J., and Cui, Y.D. 2011. Evolutionary rate of E2 genes
- 444 of classical swine fever virus in China. Isr J Vet Med **66**: 161-163.
- 445
- 446
- 447

448 **Tables**

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

Coding sequence	Fixed	Polymorphic
NS5B Group 1 vs. Group 2		
Replacement	20	143
Synonymous	41	601
<i>Probability</i> ^a		0.02*
NS3 Group 1 vs. Group 2		
Replacement	2	158
Synonymous	48	576
<i>Probability</i>		< 0.01**
NS4B Group 1 vs. Group 2		
Replacement	5	60
Synonymous	21	323
<i>Probability</i>		0.78
NS5B Subgroup 2.1 or 2.2 vs. Subgroup 2.3		
Replacement	12	111
Synonymous	49	499
<i>Probability</i>		0.86

450 ^a Single and double asterisks correspond to two-tailed *Probability* < 0.05 or

451 *Probability* < 0.01 (Fisher's exact test).

452

Table 2. The estimates of the type-II functional divergences between two major groups and subgroups of group 2.

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

^a θ_{II} , coefficient of type-II functional divergence. See Materials and Methods for details on functional divergence test.

458 **Figure legends**

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

464

Draft

465 **Appendix A. Supplementary data**

466 Three supplementary tables are available:

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

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

470 **Table S3.** Maximum likelihood analysis of RdRps within major groups of RdRp
471 coding sequence.

Draft

Fig. 1