



Canadian Journal of Microbiology  
Revue canadienne de microbiologie

## Different evolutionary patterns of classical swine fever virus envelope proteins

Journal:	<i>Canadian Journal of Microbiology</i>
Manuscript ID	cjm-2015-0709.R1
Manuscript Type:	Article
Date Submitted by the Author:	16-Nov-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 Zhang, Mingwang; Sichuan Agricultural University, College of Animal Science and Technology
Keyword:	Classical swine fever virus, Envelope proteins, Adaptive diversification, Different patterns

SCHOLARONE™  
Manuscripts

**Different evolutionary patterns of classical swine fever virus envelope proteins**

Yan Li, Zexiao Yang, and Mingwang Zhang

**Y. Li and M. Zhang.** 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](mailto:liyan@sicau.edu.cn)).

## Abstract

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. The proteins E<sup>ms</sup>, E1, and E2 are components of the viral envelope membrane. They are also implicated in virus attachment and entry, replication, and/or anti-immune response. Here, we studied the genetic variations of these envelope proteins in the evolution of CSFV. The results reveal that the envelope proteins underwent different evolutionary fates. In E<sup>ms</sup> and E1, but not E2, a number of amino acid sites experienced functional divergence. Furthermore, the diversification in E<sup>ms</sup> and E1 were generally episodic because the divergence-related changes of E1 have only occurred with the split between two major groups of CSFV and that of E<sup>ms</sup> have taken place with the division of one major group. The major divergence-related sites of E<sup>ms</sup> are located on one of the substrate-binding regions of RNase domain and C-terminal extension. These functional domains have been reported to block activation of the innate immune system and attachment and entry into host cells, respectively. Our results might shed some light on the divergent roles of the envelope proteins.

**Keywords:**

Classical swine fever virus,  
Envelope proteins,  
Adaptive diversification,  
Different patterns.

Draft

## Manuscript text

## Introduction

Classical swine fever virus (CSFV) is the pathogen that causes a highly infectious disease of pigs and outbreaks of classical swine fever (CSF) in domestic pigs resulted in devastating losses to the pig industry (Edwards et al. 2000; Meuwissen et al. 1999; Vandeputte and Chappuis 1999). Field CSFV strains show different degrees of virulence, resulting in clinical signs with different levels of severity (Artois et al. 2002; Kaden et al. 2000; Leifer et al. 2013; Moennig et al. 2003). Highly virulent field strains generally cause an acute hemorrhagic fever with nearly 100% mortality. High body temperature is often the first clinical trials, which is followed by respiratory, gastrointestinal and neurological symptoms. The terminal stage is characterized by skin haemorrhages and skin cyanosis on the tips of extremities (ears, tail, and vulva). Low-virulent and moderately virulent strains can give rise to sub-acute or chronic infections prior to death.

CSFV belongs to the genus *Pestivirus* of the family *Flaviviridae* with the closely related viruses Bovine viral diarrhoea virus genotypes 1 (BVDV-1) and 2 (BVDV-2), and Border disease virus (BDV) (Van Regenmortel et al. 2001). Based on genetic diversity, CSFV strains could be segregated into two major groups (1 and 2) and a rare group (3) (Lowings et al. 1996; Postel et al. 2012). Furthermore, clear segregation of subgroups can be observed within the major group 2 (Postel et al. 2012).

CSFV possesses a single-stranded, positive-sense RNA genome. The viral genome contains two untranslated regions (UTRs) at the 5' and 3' ends, and an open reading frame (ORF). The

ORF encodes a polyprotein, which is further co- and post-translationally processed into 12 mature polypeptides by viral and host cell proteases (reviewed in (Meyers and Thiel 1996)). These proteins comprise four structural proteins (capsid protein C and three envelope proteins  $E^{ms}$ , E1, and E2) and eight nonstructural proteins (Npro, P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B).

The proteins  $E^{ms}$ , E1, and E2 are parts of the viral envelope membrane of CSFV (Rumenapf et al. 1993). The protein  $E^{ms}$  has a RNase domain which is characterized by two nucleotide binding sites N1 and N2 (Krey et al. 2012), and a C-terminal extension (Aberle et al. 2014; Langedijk 2002). E1 and E2 are type I transmembrane proteins with an N-terminal ectodomain and a C-terminal hydrophobic anchor (Thiel et al. 1991).

The envelope proteins are involved in the formation of infectious virus particles, viral attachment and entry, and replication (Hulst and Moormann 1997; Rumenapf et al. 1993; Wang et al. 2004).  $E^{ms}$  is cytotoxic for lymphocytes and not for epithelial cells (Bruschke et al. 1997; Magkouras et al. 2008; van der Molen and van Oirschot 1981; von Freyburg et al. 2004). In addition,  $E^{ms}$  possesses ribonuclease activity (Hulst et al. 1994; Langedijk et al. 2002; Schneider et al. 1993), and soluble  $E^{ms}$  secreted from infected cells has been proposed to block induction of type I interferon (IFN) by degradation of circulating dsRNA that otherwise would trigger the IFN response via a Toll-like receptor (TLR)-dependent pathway (Magkouras et al. 2008; Matzener et al. 2009).

It is noteworthy that, although the phylogenetic splits of CSFV into groups and subgroups are well supported, the genetic diversification patterns of the envelope proteins during these segregation processes are not yet known. In this study, we analyzed the genetic variation of all envelope proteins in the evolution of CSFV. Our results might give some insights into the

functional diversification of the envelope proteins and molecular mechanisms relevant to CSFV virulence.

## Materials and methods

### Sequence data collection

We retrieved the full-length coding sequences for three envelope proteins ( $E^{ns}$ , E1, and E2) of field CSFV strains from the NCBI database. Because recombinant CSFV may not be diffused in natural populations (He et al. 2007; Lowings et al. 1996; Postel et al. 2012), the sequences of rare recombinant were not adopted in the present study. Seventy sequences of these envelope proteins were finally used for following evolutionary analyses (Table S1). A complete coding sequence of these proteins from closely related species BDV was chosen as an outgroup (Table S1).

### Sequence alignment and phylogenetic analysis of genes coding for the envelope proteins

The coding sequences of the envelope proteins were aligned using ClustalW (Thompson et al. 1997) and checked manually. For each protein, maximum likelihood (ML) analysis in MEGA5 (Tamura et al. 2011) was used to infer evolutionary relationships among the coding sequences. The General Time Reversible model (Lanave et al. 1984) of evolution allowing gamma distribution plus a proportion of invariant sites for  $E^{ns}$  and E2 and Kimura 2-parameter model (Kimura 1980) with gamma distribution plus a proportion of invariant sites for E1 were implemented in ML analyses on these datasets. We used these models because they fit the

corresponding datasets with the lowest Bayesian Information Criterion (BIC) scores (Nei and Kumar 2000). One thousand bootstrap replications were conducted to evaluate the reliabilities of the reconstructed trees.

**The McDonald-Kreitman (MK) test of CSFV envelope proteins**

Because the reconstructed tree for each envelope protein is divided into well-supported clades (groups or subgroups) based on genetic diversity (see below), we adopted the MK test (McDonald and Kreitman 1991) to identify adaptive divergence in these proteins. If replacement (nonsynonymous) and synonymous differences were due to neutral mutations alone, the ratio of synonymous to replacement divergence between clades would be expected to be the same as the ratio of synonymous to replacement polymorphism within clades. A Fisher's exact test of independence was calculated and significant deviation ( $Probability \leq 0.05$ ) from this expectation can be used to reject the neutral mutation hypothesis (McDonald 2009).

**Tests for the functional divergence of CSFV envelope proteins between major groups and subgroups**

We adopted the statistical methods Gu-99 and Gu-2006 in DIVERGE 3.0 to detect type I and type II functional divergence at individual amino acid site of envelope proteins (Gu 1999, 2001, 2006). Type I functional divergence represents an evolutionary pattern where one amino acid is conserved in the one clade but highly variable in the other clade. Type II functional divergence means that the evolution of amino acid is conserved in both clades but their physiochemical properties are remarkably different. The coefficient of the type I functional divergence ( $\theta_I$ ) and that of type II functional divergence ( $\theta_{II}$ ) are calculated with methods Gu-99 and Gu-2006,

respectively. A likelihood ratio test (LRT) for the maximum likelihood estimate and Z-score test for the model-free method estimate (MFE) are adopted to statistically evaluate the type I functional divergence. For type II functional divergence, the Z-score test is performed for statistical evaluation (Gu 2001; Gu et al. 2013). The null hypothesis is  $\theta = 0$ , which means absence of functional divergence between clades. For the observed coefficient  $\theta_I$ , a rejection of the null hypothesis implies the alteration of functional constrain which results from the functional divergence. For  $\theta_{II}$ , a deviation from the null hypothesis indicates that a remarkable shift of amino acid physiochemical properties is likely to have occurred between these clades. Furthermore, the posterior probability is used to predict the amino acid residues that could be responsible for these functional differences (Gu 2006).

## Results

### Phylogenetic analysis of CSFV envelope proteins

To identify the evolutionary profiles of the envelope proteins used in present study, we reconstructed their nucleotide trees using the maximum likelihood method. The clustering structures of three envelop proteins are generally similar. For each protein, the obtained nucleotide tree comprises two major groups and one minor group (Fig. 1, 2, and 3). The division among these groups is well supported (bootstrap values  $\geq 90\%$ ) but the position of the minor group is variable. Because the clustering structures are totally consistent with that found in previous studies (Lowings et al. 1996; Postel et al. 2012), we adopted the same group numbering system as that used by Postel (Postel et al. 2012). The major groups (groups 1 and 2)

contain most of the available viruses (20 and 48 strains respectively), and minor group 3 is represented by only two strains. Since few group 3 sequences could be retrieved (Postel et al. 2012), further analyses were performed for the two major groups only. For each envelope protein, the group 2 comprises three well supported subgroups (subgroups 2.1, 2.2, and 2.3) although the location of 2.2 is variable (Fig. 1, 2, and 3), which is generally consistent with previously obtained results (Lowings et al. 1996). These subgroups comprise 18, 5, and 25 sequences, respectively (Fig. 1).

**Inter-group adaptive diversity between major groups of E1 but not between those of E<sup>rns</sup> and E2**

The clear genetic division of envelop proteins between the group 1 and group 2 raised question about the underlying driving forces during this evolutionary process. Using the MK test with the coding sequences of each envelope protein, we detected the occurrence of adaptive divergence with the split of the groups from their common ancestor. For each protein, a Fisher's exact test of independence showed that the ratio of replacement versus fixed synonymous differences was not significantly greater than that of replacement versus synonymous polymorphisms (Table 1). These results showed that no adaptive divergence took place on the envelope proteins in the dividing process of the CSFV progenitor into groups.

It has been widely accepted that the adaptive diversity typically occurs at few sites, as most amino acids in a protein are under functional constraints (LI 1997). It is also known that the changes of amino acid physiochemical properties at critical sites could contribute to the diversification associated with different environmental challenges (LI 1983; Nei 1987). Hence, we have examined the divergence between the group 1 and group 2 envelope proteins at

individual sites. We first implemented the method Gu-2006 in DIVERGE 3 (see Materials and methods) (Gu 2006; Gu et al. 2013) to calculate the coefficient of type II functional divergence ( $\theta_{II}$ ) on each protein dataset between two groups (Table 2). For each envelope protein, no  $\theta_{II}$  value is larger than 0 with a significant probability. Furthermore, the site-specific scores based on the posterior probability ratio showed that no site is related to the functional divergence with a probability over 0.90 (Table 2).

For type I functional divergence (see Materials and methods) (Gu 1999; Gu et al. 2013), however, the envelope proteins present different characteristics. The site-specific score suggested that a site (80) of E1 is related to the functional divergence with significant probability ( $P \geq 0.95$ ), and two sites (83 and 175) exhibit a deviation from the neutral mutation hypothesis in favor of functional divergence with marginal significance ( $0.95 > P \geq 0.90$ ) (Table 2). The amino acids at these type-I divergence-related sites harbor high level of polymorphisms within one of major group, but were conserved within another one (Table S2). These results implied that the amino acids at these sites experienced functional divergence with the split of group 1 and group 2 from their common ancestor. For  $E^{rns}$  and E2, in contrast, no site has been subjected to functional divergence although Z-score test and LRT indicated that type I functional divergence  $\theta_I$  is larger than 0 with  $\geq 0.90$  probability.

## **Inter-subgroup adaptive diversity within the groups 2 of $E^{rns}$ but not within those of E1 and E2**

Different selective pressures probably played roles in shaping the variations of the envelope proteins at different evolutionary stages. For all envelope proteins, since there are three well supported subgroups (2.1, 2.2, and 2.3) within the group 2 (Fig. 1, 2, and 3), we then measured

whether such division might be relevant to adaptive evolution (Gu 2006; Gu et al. 2013; McDonald and Kreitman 1991). Since only five sequences of subgroup 2.2 are available to date, which tends to falsely estimate the variations (LI 1997), the evolutionary analyses were only performed with the subgroups 2.1 and 2.3 sequences of each candidate protein.

Interestingly, for the subgroups 2.1 and 2.3  $E^{\text{ms}}$  coding sequences, the result of MK test indicated that the ratio of fixed amino acid replacement versus fixed synonymous differences was significantly greater than the rate of replacement polymorphisms versus synonymous differences (*Probability* = 0.028) (Table 3). In contrast, E1 and E2 didn't show a similar pattern of evolution (Table 3). The MK test is generally robust to most demographic assumptions (Aguileta et al. 2009; Eyre-Walker 2006). The exception is that a rapid expansion in population size can contribute to the fixation of slightly deleterious nonsynonymous mutations and thereby lead to artifactual evidence of adaptive evolution (Eyre-Walker 2002). A demographic factor should have similar effects on the different regions of the viral genome. Therefore, the opposite results of the MK test on the  $E^{\text{ms}}$ , E1, and E2 suggested that the effect of a demographic factor, such as population expansion, could be ruled out - that is, the significant deviation of MK test from the neutral mutation hypothesis could be a consequence of adaptive selection on  $E^{\text{ms}}$ .

We then repeated the method Gu-2006 in DIVERGE 3 to identify the potential type II functional divergence-related sites in each envelope protein. Consistent with that of MK test, a different pattern was found for  $E^{\text{ms}}$ , E1, and E2 coding sequences (Table 4).  $E^{\text{ms}}$ , but not E1 and E2, exhibited a deviation from neutrality in favor of functional divergence with a probability over 0.85 (Z-score = 1.13). More notably, the site-specific scores based on the posterior probability ratio showed that eight sites of  $E^{\text{ms}}$  are related to the functional divergence with a significant probability ( $P \geq 0.90$ ) (Table 4). The amino acids at these predicted sites exhibited

considerable difference of physiochemical properties (e.g., charge, hydrophobic, etc.) between the subgroups 2.1 and 2.3, while the majority of these sites were conserved within each of these subgroups (Table S3). Furthermore, because the strains were collected globally over the past several decades (Ji et al. 2014; Lowings et al. 1996), the substitutions at these sites had most likely occurred at the early stage of the subgroup split and maintained within these subgroups rather than resulted from fixation of mutant strains within the subsequent evolution of the subgroups.

Finally, we implemented the method Gu-99 to assess whether the major subgroups 2.1 and 2.3 of the envelope proteins underwent the type I functional divergence. For each protein, none of the coefficient of type I functional divergence ( $\theta_I$ ) is larger than 0 with a significant probability (Table 4). In addition, the site-specific posterior probability ratio suggested that no site has experienced functional divergence ( $P \geq 0.90$ ). On the whole, the accumulated results suggested that E<sup>ms</sup> had most likely experienced the type II, but not type I functional divergence with the split of subgroups 2.1 and 2.3 from an ancestor. For E1 and E2, different statistical tests indicated that no functional divergence took place with the division into these subgroups.

## Discussion

Based on genetic variation, the CSFV strains could be clearly divided into clades. In terms of the envelope proteins, the division raised questions about their diversity feature during this evolutionary process. In this study, we characterized the inter- and intragroup sequence variations of the envelope proteins in detail. The combined results of our evolutionary analyses

revealed that envelope proteins present different diversification patterns.

The highly virulent strains and low and moderately virulent strains cluster in major group 1 and group 2 (Li et al. 2006; Topfer et al. 2013), respectively. It has been reported that adaptive selection has driven the diversification of the RNA-dependent RNA polymerase NS5B between these groups (Li and Yang). In present work, our results suggested that the protein E1 had experienced functional divergence with the split of the group 1 and group 2. These results indicate that the diversification of multiple proteins which are involved in different critical functions might be related to the virulence of these groups. Notably, differently from the adaptive variations of NS5B which present the type II functional divergence (Li and Yang), the predicted sites of E1 had experienced the type I functional divergence (Table 2 and S2). These results suggest that the divergence-related proteins experienced different diversification in the split process of the CSFV progenitor into these major groups. Because no three-dimensional structural information of the protein E1 have to date been reported, and its function has been studied only to a limited extent, future investigation will be necessary to fully understand the significance of the variations.

For E<sup>ms</sup>, our results indicated that eight sites may be related to the functional divergence between the subgroups 2.1 and 2.3. Structurally, the majority of adaptively selected sites are located on the nucleotide binding region N2 of RNase domain (sites 14 and 16) and the C-terminal extension (sites 175, 192, 211, and 218) based on the homologue of closely related viruses BVDV-1 (Aberle et al. 2014; Krey et al. 2012). Functionally, these domains play multiple roles in blocking the activation of the innate immune system, the retention/secretion of the protein, and membrane anchor (Fetzer et al. 2005; Hulst and Moormann 1997; Hulst et al. 2001; Magkouras et al. 2008; Matzener et al. 2009; Tews and Meyers 2007). It is becoming

increasingly clear that pathogens have evolved a range of anti-immune molecular mechanisms to overcome both innate and adaptive immunity (Finlay and McFadden 2006; Randall and Goodbourn 2008). Therefore, taking into consideration the structural and functional insights mentioned above, we suggested that the diversification at these adaptively selected sites might involve changes of (1) cytotoxicity against lymphocytes that relies on RNase activity of E<sup>ms</sup> (Bruschke et al. 1997; Magkouras et al. 2008; van der Molen and van Oirschot 1981; von Freyburg et al. 2004) and (2) retention/secretion of E<sup>ms</sup>, which is necessary to membrane anchoring (Aberle et al. 2014) and inhibit IFN response (Magkouras et al. 2008; Matzener et al. 2009). These changes can help E<sup>ms</sup> to counter the host immune strategies that target these important functions. Alternatively, for the predicted sites located in the C-terminal extension, the changes might favor their interactions with a wide number of membrane-associated molecules in host cells (Hulst et al. 2001; Kjellen and Lindahl 1991; Shukla et al. 1999; Stringer and Gallagher 1997), which might have constituted a functional adaptation to attach and enter into different types of cells. In this scenario, the subgroup splits of the major group 2 might be relevant to functional diversification and furthermore give a hint about the mechanism underlying the split process. It will be interesting to study the functional significance of these genetic variations and subgroup segregation. Finally, if the hypotheses are supported, these evolutionary insights might extend our understanding regarding the function and virulence nature of the envelope proteins, especially E1 and E<sup>ms</sup>, in CSFV.

## Acknowledgements

We thank anonymous referees for valuable comments and suggestions that led to improvement of the manuscript.

**References**

- Aberle, D., Muhle-Goll, C., Burck, J., Wolf, M., Reisser, S., Luy, B., Wenzel, W., Ulrich, A.S., and Meyers, G. 2014. Structure of the membrane anchor of pestivirus glycoprotein E(rns), a long tilted amphipathic helix. *PLoS pathogens* **10**(2): e1003973.
- Aguileta, G., Refregier, G., Yockteng, R., Fournier, E., and Giraud, T. 2009. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infect Genet Evol* **9**(4): 656-670.
- Artois, M., Depner, K.R., Guberti, V., Hars, J., Rossi, S., and Rutili, D. 2002. Classical swine fever (hog cholera) in wild boar in Europe. *Revue scientifique et technique (International Office of Epizootics)* **21**(2): 287-303.
- Bruschke, C.J., Hulst, M.M., Moormann, R.J., van Rijn, P.A., and van Oirschot, J.T. 1997. Glycoprotein Erns of pestiviruses induces apoptosis in lymphocytes of several species. *Journal of virology* **71**(9): 6692-6696.
- Edwards, S., Fukusho, A., Lefevre, P.C., Lipowski, A., Pejsak, Z., Roehe, P., and Westergaard, J. 2000. Classical swine fever: the global situation. *Veterinary microbiology* **73**(2-3): 103-119.
- Eyre-Walker, A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**(4): 2017-2024.
- Eyre-Walker, A. 2006. The genomic rate of adaptive evolution. *Trends in ecology & evolution* **21**(10): 569-575.
- Fetzer, C., Tews, B.A., and Meyers, G. 2005. The carboxy-terminal sequence of the pestivirus glycoprotein E(rns) represents an unusual type of membrane anchor. *Journal of virology* **79**(18): 11901-11913.

- 321 Finlay, B.B., and McFadden, G. 2006. Anti-immunology: evasion of the host immune system by  
322 bacterial and viral pathogens. *Cell* **124**(4): 767-782.
- 323 Gu, X. 1999. Statistical methods for testing functional divergence after gene duplication.  
324 *Molecular biology and evolution* **16**(12): 1664-1674.
- 325 Gu, X. 2001. Maximum-likelihood approach for gene family evolution under functional  
326 divergence. *Molecular biology and evolution* **18**(4): 453-464.
- 327 Gu, X. 2006. A simple statistical method for estimating type-II (cluster-specific) functional  
328 divergence of protein sequences. *Molecular biology and evolution* **23**(10): 1937-1945.
- 329 Gu, X., Zou, Y., Su, Z., Huang, W., Zhou, Z., Arendsee, Z., and Zeng, Y. 2013. An update of  
330 DIVERGE software for functional divergence analysis of protein family. *Molecular biology and*  
331 *evolution* **30**(7): 1713-1719.
- 332 He, C.Q., Ding, N.Z., Chen, J.G., and Li, Y.L. 2007. Evidence of natural recombination in  
333 classical swine fever virus. *Virus research* **126**(1-2): 179-185.
- 334 Hulst, M.M., Himes, G., Newbigin, E., and Moormann, R.J. 1994. Glycoprotein E2 of classical  
335 swine fever virus: expression in insect cells and identification as a ribonuclease. *Virology* **200**(2):  
336 558-565.
- 337 Hulst, M.M., and Moormann, R.J. 1997. Inhibition of pestivirus infection in cell culture by  
338 envelope proteins E(rns) and E2 of classical swine fever virus: E(rns) and E2 interact with  
339 different receptors. *The Journal of general virology* **78** ( Pt 11): 2779-2787.
- 340 Hulst, M.M., van Gennip, H.G., Vlot, A.C., Schooten, E., de Smit, A.J., and Moormann, R.J.  
341 2001. Interaction of classical swine fever virus with membrane-associated heparan sulfate: role  
342 for virus replication in vivo and virulence. *Journal of virology* **75**(20): 9585-9595.
- 343 Ji, W., Niu, D.D., Si, H.L., Ding, N.Z., and He, C.Q. 2014. Vaccination influences the evolution

- 344 of classical swine fever virus. *Infect Genet Evol* **25**: 69-77.
- 345 Kaden, V., Ziegler, U., Lange, E., and Dedek, J. 2000. Classical swine fever virus: clinical,  
346 virological, serological and hematological findings after infection of domestic pigs and wild  
347 boars with the field isolate "Spante" originating from wild boar. *Berliner und Munchener  
348 tierarztliche Wochenschrift* **113**(11-12): 412-416.
- 349 Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions  
350 through comparative studies of nucleotide sequences. *Journal of molecular evolution* **16**(2):  
351 111-120.
- 352 Kjellen, L., and Lindahl, U. 1991. Proteoglycans: structures and interactions. *Annual review of  
353 biochemistry* **60**: 443-475.
- 354 Krey, T., Bontems, F., Vonnrhein, C., Vaney, M.C., Bricogne, G., Rumenapf, T., and Rey, F.A.  
355 2012. Crystal structure of the pestivirus envelope glycoprotein E(rns) and mechanistic analysis  
356 of its ribonuclease activity. *Structure* **20**(5): 862-873.
- 357 Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for calculating  
358 evolutionary substitution rates. *Journal of molecular evolution* **20**(1): 86-93.
- 359 Langedijk, J.P. 2002. Translocation activity of C-terminal domain of pestivirus Erns and  
360 ribotoxin L3 loop. *The Journal of biological chemistry* **277**(7): 5308-5314.
- 361 Langedijk, J.P., van Veelen, P.A., Schaaper, W.M., de Ru, A.H., Melen, R.H., and Hulst, M.M.  
362 2002. A structural model of pestivirus E(rns) based on disulfide bond connectivity and homology  
363 modeling reveals an extremely rare vicinal disulfide. *Journal of virology* **76**(20): 10383-10392.
- 364 Leifer, I., Ruggli, N., and Blome, S. 2013. Approaches to define the viral genetic basis of  
365 classical swine fever virus virulence. *Virology* **438**(2): 51-55.
- 366 LI, W.-H. 1983. Evolution of duplicate genes and pseudo genes. *In* *Evolution of Genes and*

- 367 Proteins. *Edited by* M. Nei and R.K. Koehn. Sinauer Associates, Sunderland, Mass. pp. 14-37.
- 368 LI, W.-H. 1997. Molecular evolution. . Sinauer Associates, Sunderland, MA.
- 369 Li, X., Xu, Z., He, Y., Yao, Q., Zhang, K., Jin, M., Chen, H., and Qian, P. 2006. Genome  
370 comparison of a novel classical swine fever virus isolated in China in 2004 with other CSFV  
371 strains. *Virus genes* **33**(2): 133-142.
- 372 Li, Y., and Yang, Z. Episodic adaptive diversification of classical swine fever virus  
373 RNA-dependent RNA polymerase NS5B. *Canadian Journal of Microbiology*.
- 374 Lowings, P., Ibata, G., Needham, J., and Paton, D. 1996. Classical swine fever virus diversity and  
375 evolution. *The Journal of general virology* **77** ( Pt 6): 1311-1321.
- 376 Magkouras, I., Matzener, P., Rumenapf, T., Peterhans, E., and Schweizer, M. 2008.  
377 RNase-dependent inhibition of extracellular, but not intracellular, dsRNA-induced interferon  
378 synthesis by Erns of pestiviruses. *The Journal of general virology* **89**(Pt 10): 2501-2506.
- 379 Matzener, P., Magkouras, I., Rumenapf, T., Peterhans, E., and Schweizer, M. 2009. The viral  
380 RNase E(rns) prevents IFN type-I triggering by pestiviral single- and double-stranded RNAs.  
381 *Virus research* **140**(1-2): 15-23.
- 382 McDonald, J.H. 2009. Handbook of biological statistics. Sparky House Publishing Baltimore,  
383 MD.
- 384 McDonald, J.H., and Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in  
385 *Drosophila*. *Nature* **351**(6328): 652-654.
- 386 Meuwissen, M.P., Horst, S.H., Huirne, R.B., and Dijkhuizen, A.A. 1999. A model to estimate the  
387 financial consequences of classical swine fever outbreaks: principles and outcomes. *Preventive*  
388 *veterinary medicine* **42**(3-4): 249-270.
- 389 Meyers, G., and Thiel, H.J. 1996. Molecular characterization of pestiviruses. *Advances in virus*

- research **47**: 53-118.
- Moennig, V., Floegel-Niesmann, G., and Greiser-Wilke, I. 2003. Clinical signs and epidemiology of classical swine fever: a review of new knowledge. *Vet J* **165**(1): 11-20.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei, M., and Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press.
- Postel, A., Schmeiser, S., Bernau, J., Meindl-Boehmer, A., Pridotkas, G., Dirbakova, Z., Mojzis, M., and Becher, P. 2012. Improved strategy for phylogenetic analysis of classical swine fever virus based on full-length E2 encoding sequences. *Veterinary research* **43**: 50.
- Randall, R.E., and Goodbourn, S. 2008. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *The Journal of general virology* **89**(Pt 1): 1-47.
- Rumenapf, T., Unger, G., Strauss, J.H., and Thiel, H.J. 1993. Processing of the envelope glycoproteins of pestiviruses. *Journal of virology* **67**(6): 3288-3294.
- Schneider, R., Unger, G., Stark, R., Schneider-Scherzer, E., and Thiel, H.J. 1993. Identification of a structural glycoprotein of an RNA virus as a ribonuclease. *Science (New York, N.Y)* **261**(5125): 1169-1171.
- Shukla, D., Liu, J., Blaiklock, P., Shworak, N.W., Bai, X., Esko, J.D., Cohen, G.H., Eisenberg, R.J., Rosenberg, R.D., and Spear, P.G. 1999. A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* **99**(1): 13-22.
- Stringer, S.E., and Gallagher, J.T. 1997. Heparan sulphate. *The international journal of biochemistry & cell biology* **29**(5): 709-714.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and

- maximum parsimony methods. *Molecular biology and evolution* **28**(10): 2731-2739.
- Tews, B.A., and Meyers, G. 2007. The pestivirus glycoprotein Erns is anchored in plane in the membrane via an amphipathic helix. *The Journal of biological chemistry* **282**(45): 32730-32741.
- Thiel, H.J., Stark, R., Weiland, E., Rumenapf, T., and Meyers, G. 1991. Hog cholera virus: molecular composition of virions from a pestivirus. *Journal of virology* **65**(9): 4705-4712.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25**(24): 4876-4882.
- Topfer, A., Hoper, D., Blome, S., Beer, M., Beerenwinkel, N., Ruggli, N., and Leifer, I. 2013. Sequencing approach to analyze the role of quasispecies for classical swine fever. *Virology* **438**(1): 14-19.
- van der Molen, E.J., and van Oirschot, J.T. 1981. Congenital persistent swine fever (hog cholera). I. Pathomorphological lesions in lymphoid tissues, kidney and adrenal. *Zentralblatt für Veterinärmedizin. Reihe B. Journal of veterinary medicine* **28**(2): 89-101.
- Van Regenmortel, M.H.V., Fauquet, C.M., and Bishop, D.H.L. (eds). 2001. *Seventh Report of the International Committee on Taxonomy Viruses*. Academic Press Inc, Oxford.
- Vandeputte, J., and Chappuis, G. 1999. Classical swine fever: the European experience and a guide for infected areas. *Revue scientifique et technique (International Office of Epizootics)* **18**(3): 638-647.
- von Freyburg, M., Ege, A., Saalmüller, A., and Meyers, G. 2004. Comparison of the effects of RNase-negative and wild-type classical swine fever virus on peripheral blood cells of infected pigs. *The Journal of general virology* **85**(Pt 7): 1899-1908.
- Wang, Z., Nie, Y., Wang, P., Ding, M., and Deng, H. 2004. Characterization of classical swine

436 fever virus entry by using pseudotyped viruses: E1 and E2 are sufficient to mediate viral entry.  
437 Virology **330**(1): 332-341.

438

439

440

Draft

## Tables

**Table 1.** McDonald-Kreitman test for detecting adaptive divergence of envelope proteins between CSFV groups 1 and 2.

	Fixed	Polymorphic
$E^{ms}$		
Replacement	3	56
Synonymous	5	220
<i>Probability</i> <sup>a</sup>		0.369
$E1$		
Replacement	0	35
Synonymous	7	209
<i>Probability</i>		0.598
$E2$		
Replacement	5	117
Synonymous	10	383
<i>Probability</i>		0.364

<sup>a</sup> *Probability*, the probability of deviation from the null hypothesis of MK test was calculated with a two-tailed Fisher's exact test.

447 **Table 2.** Estimation of functional divergence of envelope proteins between CSFV groups 1 and 2 calculated using DIVERGE.

Coding sequence	The type of functional divergence	Estimate (Z-score)	Estimate (LRT)	Divergence-related sites identified with posterior probability
E <sup>rns</sup>	Type-II	$\theta_{II}^a = 0.069 \pm 0.043$ (1.60)	13.067**	None (at $P_{II} \geq 0.90$ )
	Type-I	$\theta_I^b = 0.492 \pm 0.170$ (2.89**)		None (at $P_I \geq 0.90$ )
E1	Type-II	$\theta_{II} = 0.034 \pm 0.039$ (0.87)	14.507**	None (at $P_{II} \geq 0.90$ )
	Type-I	$\theta_I = 0.729 \pm 0.159$ (4.58**)		80 (at $P_I \geq 0.95$ ) 83, 175 ( $0.95 > P \geq 0.90$ )
E2	Type-II	$\theta_{II} = 0.003 \pm 0.038$ (0.08)	4.015	None (at $P_{II} \geq 0.90$ )
	Type-I	$\theta_I = 0.316 \pm 0.072$ (4.30**)		None (at $P_I \geq 0.90$ )

448 <sup>a</sup>  $\theta_{II}$ , coefficient of type-II functional divergence.

449 <sup>b</sup>  $\theta_I$ , coefficient of type-I functional divergence.

450 Double asterisks correspond to  $P < 0.01$  from the Z-score test.

451 .

452

**Table 3.** McDonald-Kreitman test for detecting adaptive divergence of envelope proteins between CSFV subgroups 2.1 and 2.3.

	Fixed	Polymorphic
$E^{rms}$		
Replacement	8	39
Synonymous	10	172
<i>Probability</i> <sup>a</sup>		0.028*
$E1$		
Replacement	1	22
Synonymous	12	150
<i>Probability</i>		1.000
$E2$		
Replacement	6	83
Synonymous	20	303
<i>Probability</i>		1.000

<sup>a</sup> *Probability*, the probability of deviation from the null hypothesis of MK test was calculated

with a two-tailed Fisher's exact test. \* Single asterisk corresponds to *Probability* < 0.05.

457 **Table 4.** Estimation of functional divergence of envelope proteins between CSFV subgroups 2.1 and 2.3 calculated using DIVERGE.

Coding sequence	The type of functional divergence	Estimate (Z-score)	Estimate (LRT)	Divergence-related sites identified with posterior probability
E <sup>rns</sup>	Type-II	$\theta_{II}^b = 0.035 \pm 0.031$ (1.13)		18, 99, 192, 211, 218 (at $P_{II} \geq 0.99$ ) 14, 52, 175 (at $0.99 > P_{II} \geq 0.95$ )
	Type-I	$\theta_I^a = 0.108 \pm 0.326$ (0.33)	0.005	None (at $P_I \geq 0.90$ )
E1	Type-II	$\theta_{II} = -0.008 \pm 0.028$ (0.28)		None (at $P_{II} \geq 0.90$ )
	Type-I	$\theta_I = -0.226 \pm 0.316$ (0.72)	0.032	None (at $P_I \geq 0.90$ )
E2	Type-II	$\theta_{II} = -0.017 \pm 0.029$ (0.59)		None (at $P_{II} \geq 0.90$ )
	Type-I	$\theta_I = -0.033 \pm 0.225$ (0.15)	0.047	None (at $P_I \geq 0.90$ )

458 <sup>a</sup>  $\theta_{II}$ , coefficient of type-II functional divergence.

459 <sup>b</sup>  $\theta_I$ , coefficient of type-I functional divergence.

## Figure Legends

**Fig. 1.** Maximum likelihood gene tree for the complete coding sequences of envelope protein E<sup>rns</sup> extracted from the 70 genomes or partial CDS 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. The bootstrap values below 90% are not reported.

**Fig. 2.** Maximum likelihood gene tree for the complete coding sequences of envelope protein E1 extracted from the 70 genomes or partial CDS 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. The bootstrap values below 90% are not reported.

**Fig. 3.** Maximum likelihood gene tree for the complete coding sequences of envelope protein E2 extracted from the 70 genomes or partial CDS 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. 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-I divergence-related sites of group 1 and group 2 E1.

**Table S3.** The amino acids at the Type-II divergence-related sites of subgroups 2.1 and 2.3 E<sup>ms</sup>.

Draft

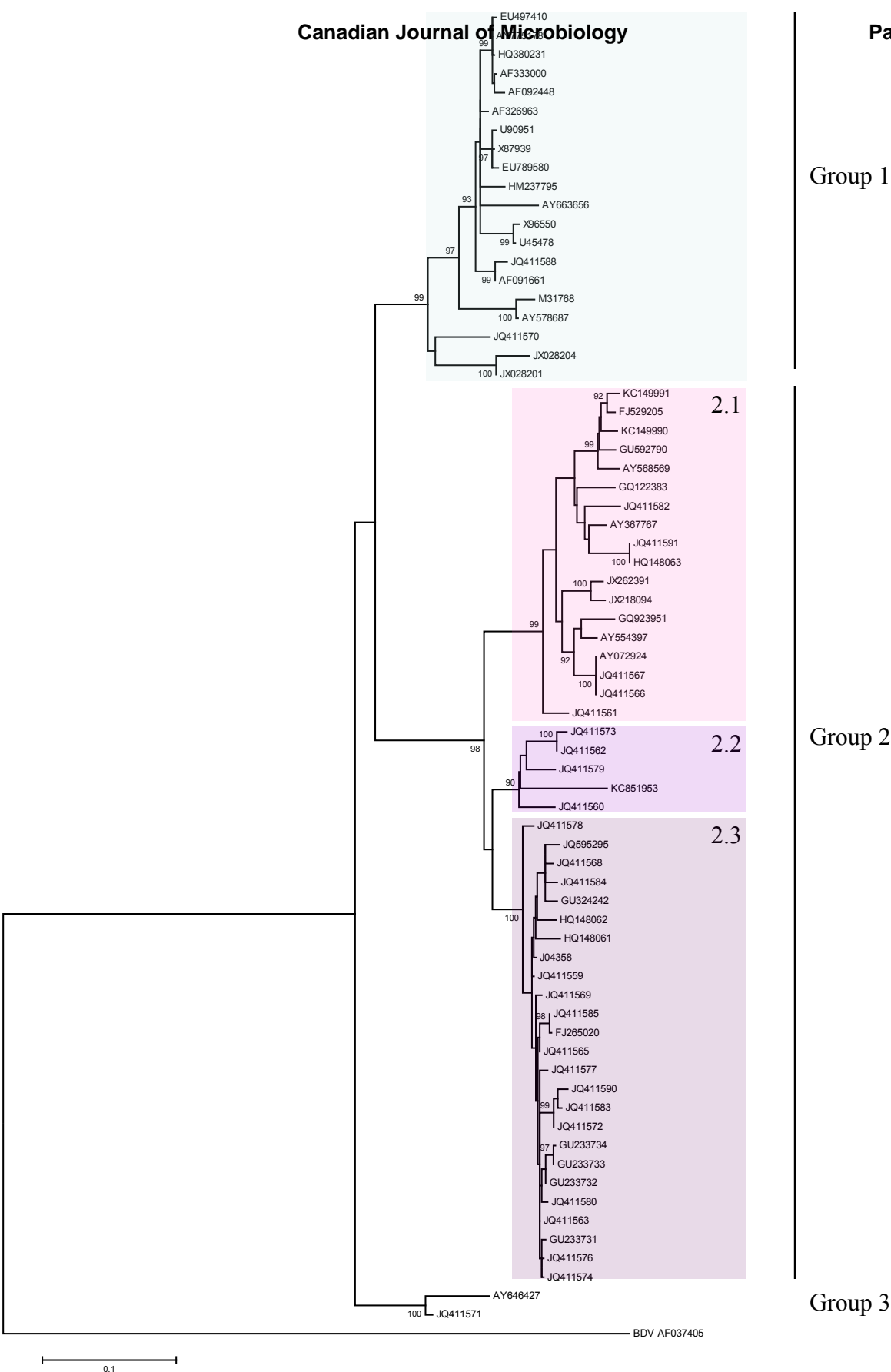


Figure 1

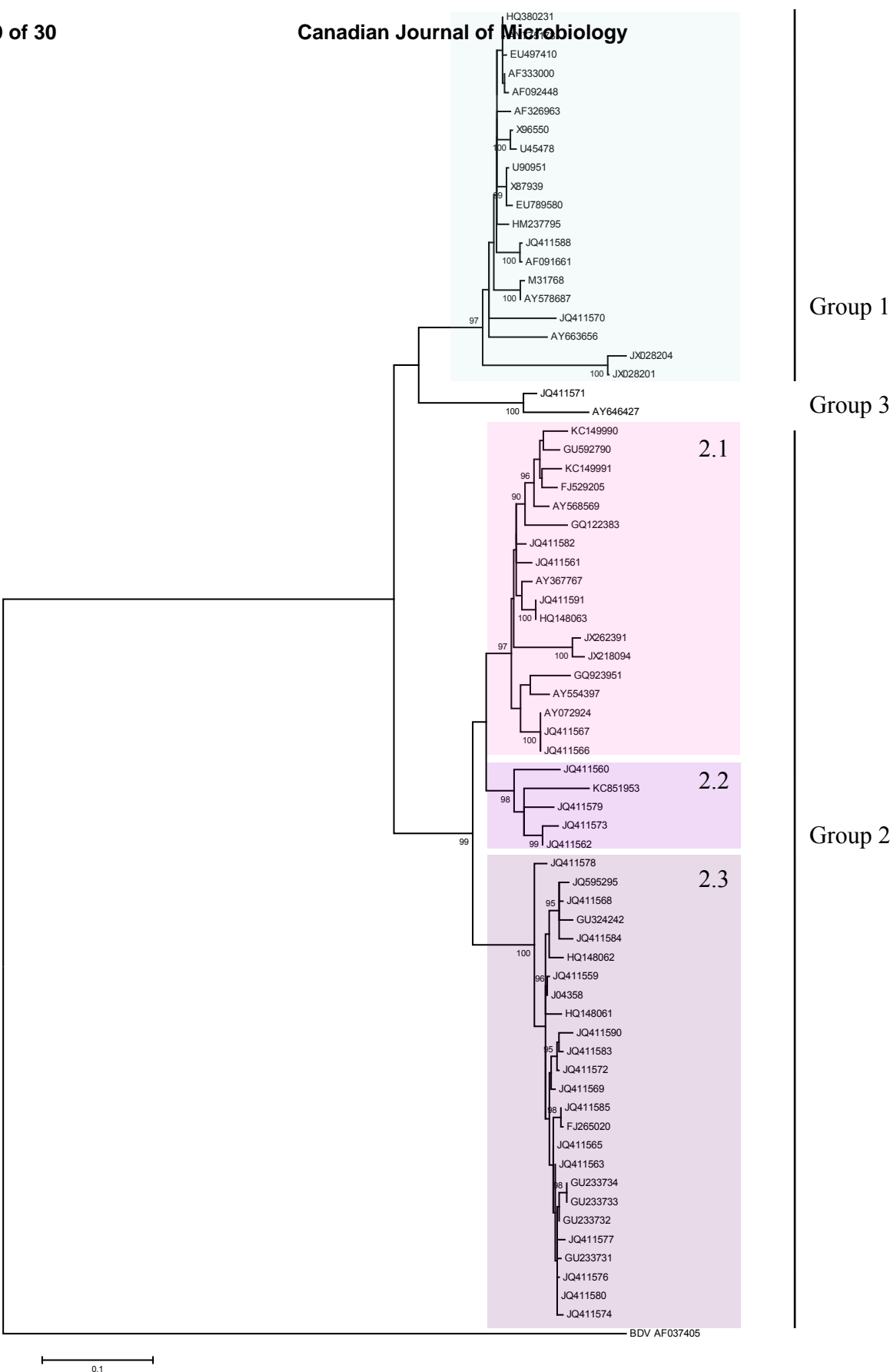
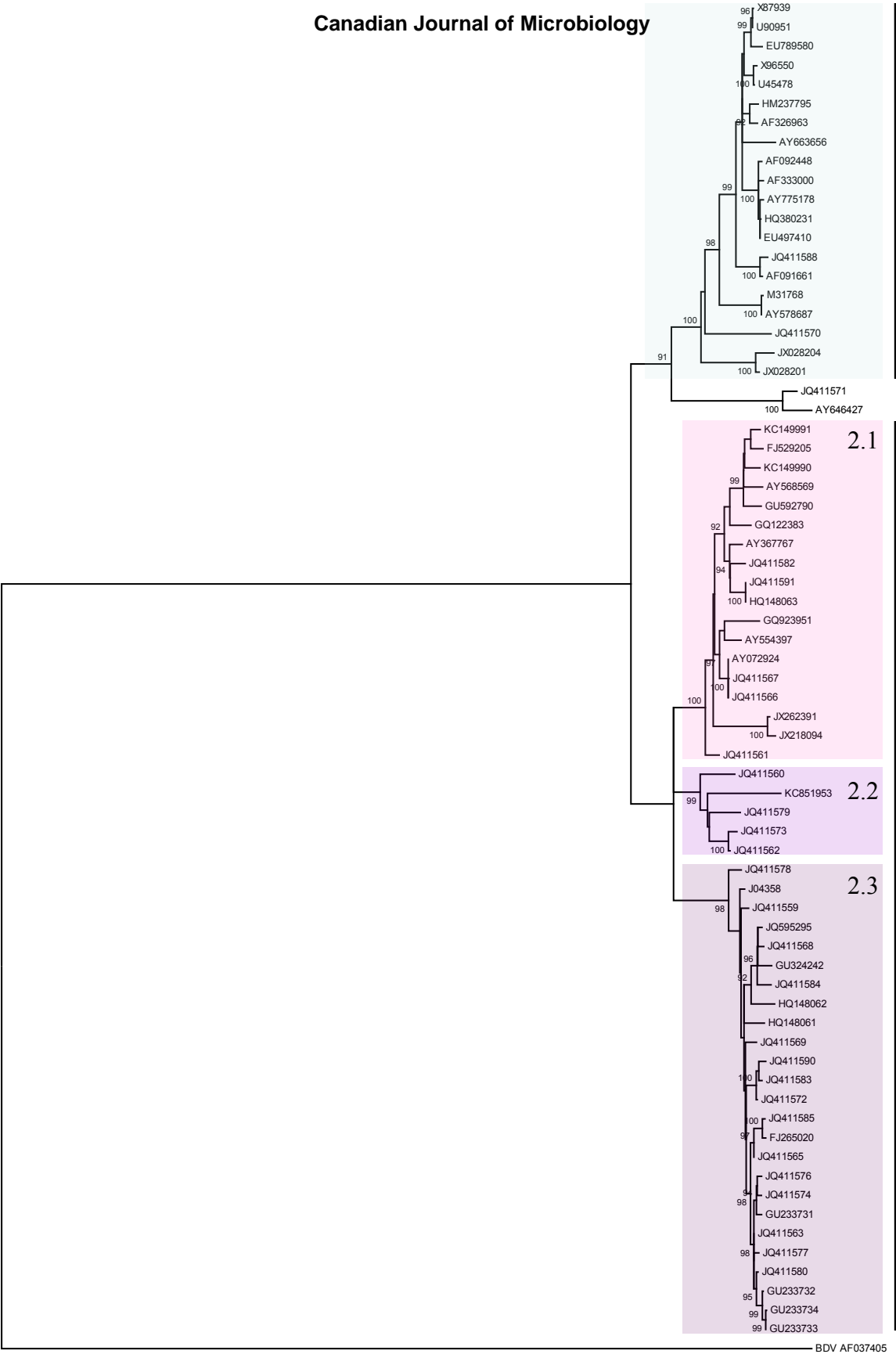


Figure 2



0.2

Figure 3