Correlation of the virulence of CSFV with evolutionary patterns of E2 glycoprotein

Zhiyin Wu[1, 2], Qin Wang[3], Qian Feng[1], Yingying Liu[1], JunlinTeng[1], Albert Cheunghai Yu[4], Jianguo Chen[1, 2]

[1]*The Key Laboratory of Cell Proliferation and Differentiation of Ministry of Education and The State Key Laboratory of Bio-membrane and Membrane Bio-engineering, Department of Cell Biology and Genetics, College of Life Sciences, Peking University, Beijing 100871, P. R. China,* [2]*The Center for Theoretical Biology, Peking University, Beijing 100871, P. R. China,* [3]*Department of Inspection Technology Research, China Institute of Veterinary Drug Control, Beijing 100081, P.R.China,* [4]*Neuroscience Research Institute and Infectious Disease Center, Health Science Center, Peking University, Beijing 100191, P.R.China*

TABLE OF CONTENTS

1. ABSTRACT

Infection with classical swine fever virus (CSFV) is costly to the livestock industry. Several genomic sequences including velogenic strains and low virulent strains have been identified. However, the reasons for the virulence of the virus have remained unclear. Based on selective pattern and pressure strength, we classified all genes of CSFV into three classes. Among these genes, the E2 gene was under the strongest positive selection. Based on the analysis of 85 representative E2 gene sequences, the location and intensity of positive selection in CSFV isolates from group one and group two were identified. These results suggest that these two groups employ evolutionary difference. Moreover, the mutations, potentially driven by positive selection, can be correlated with the virulence of CSFV by altering the conformation and function of E2 and/or changing its glycosylation pattern. Based on these results, a model for the evolution of virulence of CSFV is proposed. The results provide a link between epidemiology and the gene function of CSFV, and may shed light on the molecular mechanism underlying the variation of CSFV virulence.

2. INTRODUCTION

Classical swine fever virus (CSFV), a member of the Flaviviridae family, is a widespread viral pathogen that has been associated with outbreaks of classical swine fever (CSF) worldwide (1, 2). CSF is highly contagious and listed by the World Organization for Animal Health (OIE) (http://www.oie.int/eng/maladies/en_classification2008.htm). A number of cells including endothelial, dendritic, macrophages from the peripheral blood and monocytic cells from the bone marrow are the primary targets of CSFV. Highly debilitating and usually fatal hemorrhagic fever is observed in the host animal. It is characterized by bystander apoptosis of T lymphocytes, disseminated intravascular coagulation, fibrinolysis, and occasionally chronic infection (3). CSFV isolates share a serotype, and are segregated into three genetic groups based upon molecular epidemic data (4-6), and into four types according to clinical symptoms and virulence (2, 7). The virulence of CSFV is highly variable among strains, with newly emerging strains, often evolving from highly pathogenic to less pathogenic phenotypes in host populations of domesticated pigs (2). Control of the infection is predominantly thought to be *via* neutralizing

antibodies targeting the E2 surface protein and cell mediated immunity. While vaccines do exist, differentiating infected from vaccinated animals faces considerable technical difficulty and commonly leads to culling as a measure of outbreak control (8). Thus, a better understanding of the factors governing CSFV strain virulence might lead to better veterinary control measures against the disease.

CSFV, as an enveloped positive-sense RNA virus, has only one open reading frame (ORF) encoding a polypeptide of 3898 amino acids. The polypeptide is processed into 11 or 12 structural and nonstructural proteins, because the C-terminal border of E2 was not experimentally determined (9, 10). The structural proteins (C, E0, E1 and E2) play key roles in the CSFV's biological functions. E2, a transmembrane glycoprotein (55KD), is the most antigenic protein (11-13), and might be necessary in mediating CSFV entry into the host cell (14). Previous studies on the virulence of CSFV focused on E2 (15-19), and recent reports indicate that Npro (20, 21), E0 (9, 19, 22, 23), E1 (24, 25) are also correlated with the virulence of certain CSFV strains. However, the genetic basis of the diversity in virulence among CSFV isolates is not fully understood. Identification of the genetic determinants and key sites responsible for variation of CSFV virulence has traditionally relied on site-directed mutagenesis and/or reverse genetics of CSFV's cDNA infectious clones (9, 15, 17-19, 21, 25). These methods coupled with experiments in animals, are the most direct means of assessing the contribution of a particular motif, residue, or variation to the pathogenesis of CSF. However, such approaches are time-consuming and cannot account for synergistic interactions between sites. An alternative way is evolutionary analysis, which has been used successfully in the research for many other viruses including HIV (26-28), FMDV (29, 30) and Luteovirus (31). The evolutionary pressures on both virus and host drive their genetic and phenotypic diversification (32). The positive selective pressure on a specific gene or a specific site is pivotal to understanding the outcome of the virus-host interaction, such as the virulence of a virus. It has been demonstrated that various subtypes or lineages of the virus are dependent on the heterogeneous selective pressures correlating with its lifestyle (26, 33-35). Evolutionary analysis has also contributed in confirming some phenotypic variations of the virus associated with mutation at key sites driven by positive selection (36-38). Moreover, reverse genetic studies would benefit greatly from a detailed analysis of the evolutionary patterns, a virus has undergone in natural or experimental infections, as well as structural or biochemical experiments, suggesting the importance of particular regions for virulence.

The CSFV genome has a single stranded nature, high mutation rate and many strains making it an ideal candidate for evolutionary analysis. Thus, the importance of understanding its variability in the major targets of immune response might provide particular insight into virulence factors encoded by the CSFV genome (39). Candidate sites were identified by searching for signatures of positive selection that are potentially associated with the

variation of CSFV virulence; this shows that variation in virulence is itself an evolutionary adaptation of the virus (40-42). Specifically, we found that the E2 glycoprotein undergoes the highest positive selective pressure, and the substitution pattern of positively selective sites located on E2 glycoprotein corresponds with the changing virulence of CSFV.

3. MATERIALS AND METHODS

3.1. Dataset construction and phylogeny

Complete genomes of 36 CSFV isolates were downloaded from GenBank and aligned with ClustalW (43). Eighty-five additional complete sequences of E2 were identified by TBlastn searches on NCBI (see Additional file 1). A maximum likelihood (ML) tree and a neighbor joining (NJ) tree were constructed by Phylip 3.65 (44, 45). The topologies of these two trees are very similar, so only the results based on the ML tree were presented.

3.2. Estimate of mean omega (ω) across the genome and testing for positive selection

The rates of non-synonymous (dN) versus synonymous (dS) substitutions per site were compared. Substitution rate ratios dN/dS (ω) = 1.0, <1.0, and >1.0 indicate neutral evolution, negative (purifying) selection, and positive selection, respectively, at protein level (41). A maximum-likelihood approach was applied to compare models of evolution that allow dN/dS to vary across sites based on an ML tree (41). One model specifies a distribution of omega ratio (dN/dS) classes across sites that are constrained to be < or = 1.0, including model M0 (one-ratio), M1 (neutral) and M7 (beta), thereby specifying neutral (purifying) evolution. Whereas the more complex models including M3 (discrete), M2 (selection), M8 (beta&ù) incorporate additional class of codons where dN/dS can be > 1.0, allowing for positive selection to be modeled. The standard likelihood-ratio tests (LRT) were used to compare these models (M1 vs M2; M0 vs M3; and M7 vs M8). A Bayesian approach to generate posterior probabilities of a given dN/dS class for each amino acid site was applied to identify individual codons probably subject to positive selection. Sites with high probabilities (> 0.95 or 0.99) of falling into dN/dS >1 category are most likely to have been under positive selection (see additional file 2 for detailed parameters used).

The selection of individual viral lineages was also tested by comparing the MA1 model (in which each branch is assumed to have the same dN/dS ratio) with the MA model (in which each branch is allowed to have a different dN/dS ratio). LRT was used to compare MA with MA1 (41, 46, 47). All were performed using CODEML from the PAML package (48) (see Additional file 3 and 4). The physicochemical selective pressures of E2 were investigated as previously described by Wong *et al* (49) (see Additional file 5).

$$\dot{\omega}\,i = \sum_{k=1}^{11} f_k\,\omega_k$$

The strength of positive selection, ω, is denoted by the weighted mean (33, 50), where k is the number of ω categories under model i, and fk is the posterior probability that a site belongs to a particular category with positive selection ωk.

3.3. Structural and functional analysis of positively selected sites within E2

The physicochemical selective pressures on the E2 protein of group one and group two were compared (49, 51). The secondary structure of E2 predicted by three methods: PHD (52), Prof (53), and PSI-Pred (54-59). Additionally, the protein contact map of E2 in various CSFV isolates were predicted by GPCPRED (60) and ProfCon (61) and the CBS Prediction Server (http://www.cbs.dtu.dk/services/) was used to predict the glycosylated sites of E2.

4. RESULTS

4.1. Phylogenetic analysis of CSFV sequences

The phylogenetic tree was constructed to infer the phylogenetics of CSFV with our data by the ML method. The tree (Figure 1 and Figure 2) indicated that our dataset covered all three genetic groups of CSFV. In addition, the clustering of CSFV was not associated with geographical distribution. The main members of group one were earlier isolates and vaccine strains, while nearly all recently prevalent isolates such as 92TC1 and Alfort_T (see additional file 1) were clustered in group two. The isolates in group three were mainly found in strains distributed in Taiwan and Thailand.

4.2. Strength of positive selection on each protein encoding sequence of CSFV

In order to differentiate the respective role of each gene in the evolution of CSFV, the likelihood parameters for each (putative) protein encoding genes in CSFV genomes was estimated (Table 1) and the positive selection strength of those sequences was measured (Table 2). It was found that all the putative genes could be divided into three classes based on natural selection patterns. Class A contains gene E1, E2, E2-P7, NS4B, and NS5A. Most sites in this class are under strong negative selection with ω close to 0, while some sites are under the moderate negative selection with a ω value between 0.22 and 0.43 (under model M3). However, in this class some sites (from 0.2% to 2.3%) are actually under positive selection with ω between 1.5 and 4.6 (under model M3), suggesting that the mutations at these sites were driven by adaptive selection. Class B was composed of gene N-pro, capsid, E0, NS2, NS3, NS2/3 and NS5B, and some sites in these genes were also under dominant negative selection with ω close to 0. All other sites were under relatively weaker negative selection with ω between 0.14 and 0.9. The heterogeneous selection pressures among sites in this class were supported by their respective LRT (Table 3). Class C only composed of P7 and NS4A, which showed the global negative selection and the LRT had no support for heterogeneous selection pressures among these sites (Table 3).

The selection strengths varied among putative genes (Table 2). The average ω of E0, E1, E2, E2-P7, N-pro and NS5A was between 0.11 and 0.16. Among them, E2 had the highest mean ω between 0.146 and 0.188 for the six substitutions models M0, M1, M2, M3, M7, and M8. The average ω of the rest of genes was between 0.033 and 0.1. NS3 has the lowest mean ω (0.033 ~ 0.037 for the six models), indicating that the structural and functional constraint of NS3 is more stringent than the others. In summary, the envelope proteins E0, E1, E2 and E2-P7, and the non-structural proteins N-pro and NS5A were under stronger positive selection. The positively selected sites were detected with a high confidence score of over 95%. The other non-structural proteins were under stronger negative selection implying structural and functional constraints.

4.3. Positively selected sites of each putative gene of CSFV genome

Positively selected sites in N-pro, capsid, E0, NS2, NS3, NS2/3, NS5B, P7 and NS4A were not selected. The sites, with high confidence of over 95%, were detected by the maximum likelihood estimates for E1, E2, E2-P7, NS4B and NS5A under models M3 and M8 (Table 1). Among them, only positively selected sites of E1 could be detected under model M2. The models M0, M1 and M2 could be rejected by M3 with over 95% confidence in the LRT for E1, E2, E2-P7, NS4B and NS5A. Model M7 was also rejected by M8 in the LRT for E1, E2 and E2-P7. Furthermore, the maximum likelihood estimates under model M3 that could give the greatest likelihood value was analyzed. E1, E2, E2-P7, NS4B and NS5A had 0.5%, 2.3%, 2.1%, 0.2% and 0.2% sites under positive selection with ω of 4.590, 1.759, 1.736, 1.536 and 3.436, respectively thus confirming the existence of positive selection. The maximum likelihood estimates under model M8 indicated that the positively selected sites with over 90% confidence of each putative gene were the 80th amino acid of E1, the 72nd and 200th amino acid of E2, the 55th amino acid of NS4B and the 163rd amino acid of NS5A (Table 4). Moreover, the positively selected sites identified both by models M3 and M8 for every gene had over 95% confidence. Taken together, our results indicate that E2 occupied the most positively selected sites and that E2, E2-P7 and NS4B are under moderate positively selective pressure with ω close to 2. Interestingly, the only positively selected sites of E1 and NS5A possessed a ω value twice that of E2.

4.4. Maximum likelihood estimates of omega (ω) and detection of physicochemical selective pressures of E2 along specific lineage of CSFV

E2 reflected the changes of selective pressure to the greatest extent among all the putative genes of CSFV. The E2 dataset containing 85 complete sequences was further analyzed to investigate selective pressure on specific lineage of CSFV. The evolutionary patterns of E2 were found to be completely different between group one and group two. The adaptive difference along the specific lineage under the branch-site model implemented in the program CODEML was also compared. When the E2 of group one was taken as the foreground lineage, eight positively selected sites (55V, 72E, 75P, 200L, 205R, 290R, 299A, 364I) in the E2 with a confidence score,
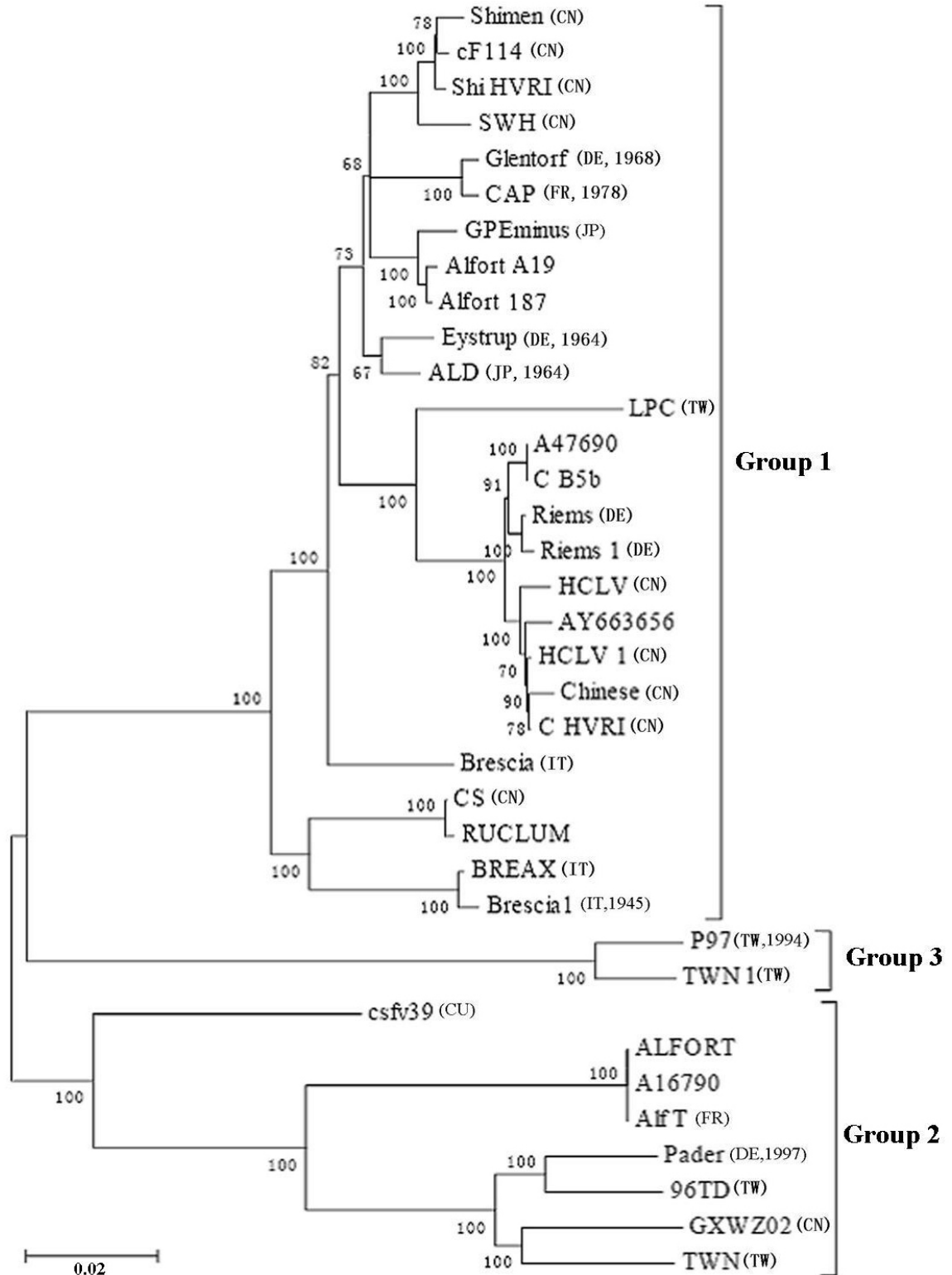
Figure 1. The ML tree of CSFV was constructed using the 36 complete CSFV genomic sequences. The genetic group is indicated and the known isolation region and date shown in brackets. The grouping refers partly to the information from Community Reference Laboratory for Classical Swine Fever (http://viro08.tiho-hannover.de/eg/csf/startCSF.cgi) and previous research. Bootstrap values are shown for the internal nodes. CN, China; IT, Italy; DE, Germany; TW, Taiwan; FR, France; JP, Japan; TH, Thailand; CU, Cuba.
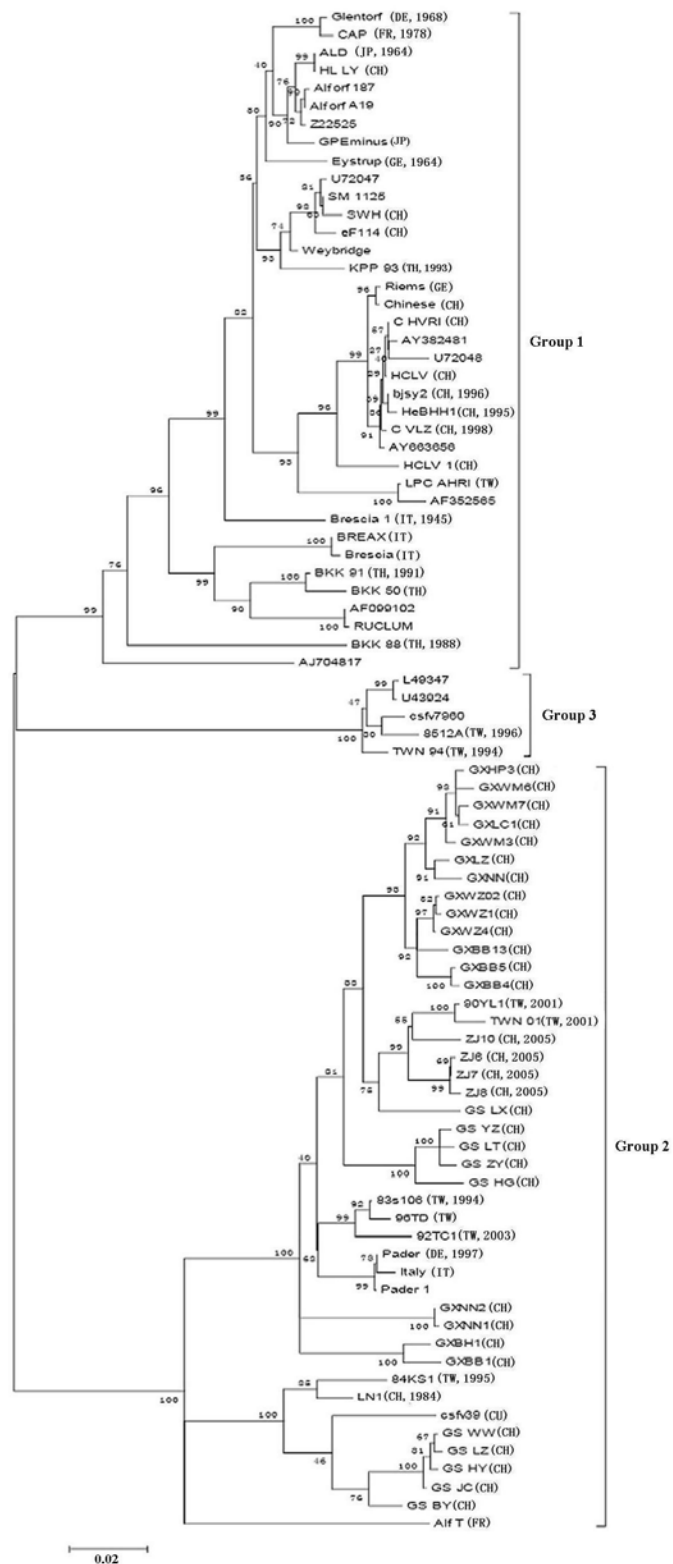
Figure 2. The ML tree of CSFV was constructed using the 85 complete E2 sequences.The genetic group is indicated and the known isolation region and date shown in brackets. The grouping reference and the abbreviations are same as Figure 1. Bootstrap values are shown for the internal nodes.

Table 1. Maximum likelihood estimates (MLEs) for putative genes coded by CSFV genome

Gene	Lc	M0	M1	M2	M3	M7	M8
N_pro	168	$\omega=0.133$	$P_0=0.925$ $\omega_0=0.086$ $P_1=0.075$ $\omega_1=1.000$	$P_0=0.925$ $\omega_0=0.086$ $P_1=0.059$ $\omega_1=1.000$ $P_2=0.016$ $\omega_2=1.000$	$P_0=0.440$ $\omega_0=0.000$ $P_1=0.513$ $\omega_1=0.186$ $P_2=0.047$ $\omega_2=0.935$	p=0.363 q=2.192	p=0.453 q=3.068 p1=0.008 $\omega=\mathbf{1.881}$
Capsid	99	$\omega=0.096$	$P_0=1.000$ $\omega_0=0.096$ $P_1=0.000$ $\omega_1=1.000$	$P0=1.000$ $\omega0=0.096$ $P1=0.000$ $\omega1=1.000$ $P2=0.000$ $\omega2=40.588$	$P0=0.150$ $\omega0=0.026$ $P1=0.473$ $\omega1=0.026$ $P2=0.377$ $\omega2=0.215$	p=0.729 q=6.587	p=0.729 q=6.587 p1=0.000 $\omega=1.000$
E0	225	$\omega=0.114$	$P_0=0.912$ $\omega_0=0.061$ $P_1=0.088$ $\omega_1=1.00$	$P_0=0.912$ $\omega_0=0.061$ $P_2=0.000$ $\omega_2=38.614$	$P_0=0.838$ $\omega_0=0.045$ $P_2=0.061$ $\omega_2=0.900$	p=0.258 q=1.667	p=0.680 q=7.924 p1=0.061 $\omega=1.000$
E1	195	$\omega=0.109$	$P_0=0.946$ $\omega_0=0.071$ $P_1=0.054$ $\omega_1=1.000$	$P_0=0.949$ $\omega_0=0.073$ $P_1=0.047$ $\omega_1=1.000$ $P_2=0.005$ $\omega_2=\mathbf{4.719}$	$P_0=0.816$ $\omega_0=0.044$ $P_1=0.179$ $\omega_1=0.365$ $P_2=0.005$ $\omega_2=\mathbf{4.590}$	p=0.325 q=2.429	p=0.445 q=3.784 p1=0.005 $\omega=\mathbf{4.590}$
E2	373	$\omega=0.146$	$P_0=0.861$ $\omega_0=0.057$ $P_1=0.139$ $\omega_1=1.000$	$P_0=0.861$ $\omega_0=0.057$ $P_1=0.110$ $\omega_1=1.000$ $P_2=0.029$ $\omega_2=1.000$	$P_0=0.685$ $\omega_0=0.014$ $P_1=0.292$ $\omega_1=0.371$ $P_2=0.023$ $\omega_2=\mathbf{1.759}$	p=0.183 q=0.979	p=0.233 q=1.534 p1=0.017 $\omega=\mathbf{1.916}$
P7	70	$\omega=0.078$	$P_0=1.000$ $\omega_0=0.078$ $P_1=0.000$ $\omega_1=1.000$	$P_0=1.000$ $\omega_0=0.078$ $P_1=0.000$ $\omega_1=1.000$ $P_2=0.000$ $\omega_2=1.493$	$P_0=0.535$ $\omega_0=0.000$ $P_1=0.280$ $\omega_1=0.170$ $P_2=0.186$ $\omega_2=0.170$	p=0.500 q=5.587	p=0.500 q=5.587 p1=0.000 $\omega=5.243$
E2-P7	443	$\omega=0.134$	$P_0=0.880$ $\omega_0=0.058$ $P_1=0.120$ $\omega_1=1.000$	$P_0=0.880$ $\omega_0=0.058$ $P_1=0.110$ $\omega_1=1.000$ $P_2=0.010$ $\omega_2=1.000$	$P_0=0.681$ $\omega_0=0.012$ $P_1=0.298$ $\omega_1=0.339$ $P_2=0.021$ $\omega_2=\mathbf{1.736}$	p=0.190 q=1.111	p=0.244 q=1.753 p1=0.016 $\omega=\mathbf{1.880}$
NS2	457	$\omega=0.090$	$P_0=0.954$ $\omega_0=0.068$ $P_1=0.046$ $\omega_1=1.000$	$P_0=0.954$ $\omega_0=0.068$ $P_1=0.046$ $\omega_1=1.000$ $P_2=0.000$ $\omega_2=29.176$	$P_0=0.605$ $\omega_0=0.000$ $P_1=0.376$ $\omega_1=0.220$ $P_2=0.020$ $\omega_2=0.677$	p=0.276 q=2.480	p=0.276 q=2.480 p1=0.000 $\omega=1.000$
NS3	682	$\omega=0.033$	$P_0=0.993$ $\omega_0=0.031$ $P_1=0.007$ $\omega_1=1.000$	$P_0=0.993$ $\omega_0=0.031$ $P_1=0.007$ $\omega_1=1.000$ $P_2=0.000$ $\omega_2=11.567$	$P_0=0.493$ $\omega_0=0.015$ $P_1=0.349$ $\omega_1=0.015$ $P_2=0.157$ $\omega_2=0.143$	p=0.401 q=10.459	p=0.399 q=10.425 p1=0.000 $\omega=2.421$
NS2/3	1139	$\omega=0.056$	$P_0=0.972$ $\omega_0=0.042$ $P_1=0.028$ $\omega_1=1.000$	$P_0=0.972$ $\omega_0=0.042$ $P_1=0.028$ $\omega_1=1.000$ $P_2=0.000$ $\omega_2=20.732$	$P_0=0.657$ $\omega_0=0.000$ $P_1=0.324$ $\omega_1=0.152$ $P_2=0.019$ $\omega_2=0.575$	p=0.218 q=3.176	p=0.218 q=3.175 p1=0.000 $\omega=2.450$
NS4A	63	$\omega=0.037$	$P_0=1.000$ $\omega_0=0.037$ $P_1=0.000$ $\omega_1=1.000$	$P_0=1.000$ $\omega_0=0.037$ $P_1=0.000$ $\omega_1=1.000$ $P_2=0.000$ $\omega_2=3.967$	$P_0=0.233$ $\omega_0=0.000$ $P_1=0.144$ $\omega_1=0.000$ $P_2=0.623$ $\omega_2=0.059$	p=1.419 q=36.329	p=1.419 q=36.330 p1=0.000 $\omega=2.982$
NS4B	347	$\omega=0.073$	$P_0=0.962$ $\omega_0=0.051$ $P_1=0.038$ $\omega_1=1.000$	$P_0=0.962$ $\omega_0=0.051$ $P_1=0.031$ $\omega_1=1.000$ $P_2=0.006$ $\omega_2=1.000$	$P_0=0.746$ $\omega_0=0.015$ $P_1=0.247$ $\omega_1=0.223$ $P_2=0.007$ $\omega_2=\mathbf{1.536}$	p=0.258 q=2.944	p=0.331 q=4.313 p1=0.006 $\omega=\mathbf{1.564}$
NS5A	497	$\omega=0.118$	$P_0=0.926$ $\omega_0=0.076$ $P_1=0.074$ $\omega_1=1.000$	$P_0=0.926$ $\omega_0=0.076$ $P_1=0.066$ $\omega_1=1.000$ $P_2=0.009$ $\omega_2=1.000$	$P_0=0.791$ $\omega_0=0.044$ $P_2=0.002$ $\omega_2=\mathbf{3.436}$	p=0.351 q=2.305	p=0.386 q=2.656 p1=0.002 $\omega=\mathbf{3.382}$
NS5B	718	$\omega=0.088$	$P_0=0.948$ $\omega_0=0.061$ $P_1=0.052$ $\omega_1=1.000$	$P_0=0.948$ $\omega_0=0.061$ $P_1=0.037$ $\omega_1=1.000$ $P_2=0.015$ $\omega_2=1.000$	$P_0=0.329$ $\omega_0=0.000$ $P_1=0.544$ $\omega_1=0.070$ $P_2=0.127$ $\omega_2=0.432$	p=0.325 q=3.027	p=0.325 q=3.028 p1=0.000 $\omega=3.634$
ORF	3894	$\omega=0.090$	$P_0=0.940$ $\omega_0=0.056$ $P_1=0.060$ $\omega_1=1.000$	$P_0=0.940$ $\omega_0=0.056$ $P_1=0.060$ $\omega_1=1.000$ $P_2=0.000$ $\omega_2=17.856$	$P_0=0.776$ $\omega_0=0.022$ $P_1=0.216$ $\omega_1=0.315$ $P_2=0.008$ $\omega_2=1.570$	p=0.239 q=2.131	p=0.274 q=2.674 p1=0.005 $\omega=\mathbf{1.798}$

All the ω values indicating the positive selection ($\omega>1$) are bold

Table 2. The strength of positive selection on each protein encoded by the genome of CSFV

Model	*N_pro*	*Capsid*	*E0*	**E1**	**E2**	<u>P7</u>	**E2-P7**	*NS2*	*NS3*	*NS2/3*	<u>NS4A</u>	**NS4B**	**NS5A**	*NS5B*	ORF
M0	0.133	0.096	0.114	0.109	0.146	0.078	0.134	0.090	0.033	0.056	0.037	0.073	0.118	0.088	0.090
M1	0.154	0.096	0.144	0.120	0.188	0.078	0.171	0.110	0.037	0.070	0.037	0.087	0.145	0.110	0.112
M2	0.154	0.096	0.144	0.138	0.188	0.078	0.171	0.110	0.037	0.070	0.037	0.087	0.145	0.110	0.112
M3	0.139	0.097	0.129	0.125	0.159	0.079	0.146	0.096	0.035	0.060	0.037	0.077	0.131	0.093	0.097
M7	0.138	0.097	0.131	0.114	0.155	0.079	0.143	0.096	0.035	0.060	0.037	0.076	0.128	0.093	0.097
M8	0.140	0.097	0.133	0.125	0.159	0.079	0.146	0.096	0.035	0.060	0.037	0.077	0.131	0.093	0.098

The E1, E2, E2-P7, NS4B, and NS5A in the Class A are bold, the N-pro, capsid, E0, NS2, NS3, NS2/3 and NS5B from the Class B are italic and the P7 and NS4A in the Class C are underlined

reliability of potential positive selection sites (46), of higher than 95% and the 72nd amino acid had a confidence score higher than 99%. The ω for the foreground lineage was between 1.2 and 1.3 (Table 5). However, when the E2 of group two was set as the foreground lineage, no positively selected sites with a confidence score of higher than 90% were detected using the same method. The 197th amino acid in group two possessed the highest ω value with a confidence of 85.2%, and the ω for the foreground lineage was 1.0 (Table 6); this suggests that the evolutionary rate of the 197th amino acid in group two was more rapid than that in group one. It is clear that the positively selective pressure of group one was much higher than that of group two. Among the eight positively selected sites, seven of them are located in the former 200 amino acids or flanked by the amino acids between 306 and 309, where the antigenic domain of E2 is located (11-13). However, the 364th amino acid position belongs to the cytoplasmic domain and could be associated with viral assembly. Maximum likelihood estimates for E2 of group one and group two under models M0, M1, M2, M3, M7, and M8 indicated that all average selective pressures on group one were higher than group two.

Table 3. The likelihood ratio test (LRT) for putative genes of CSFV

Gene	M0 vs. M3	M1 vs. M2	M1 vs. M3	M2 vs. M3	M7 vs. M8
	$\chi^2_{0.05,\,4}$= 9.49	$\chi^2_{0.05,\,2}$= 5.99	$\chi^2_{0.05,\,4}$= 9.49	$\chi^2_{0.05,\,2}$= 5.99	$\chi^2_{0.05,\,2}$= 5.99
N_pro	37.316	0.072	6.103	6.175	2.427
Capsid	5.726	0.000	5.726	5.726	0.000
E0	78.706	0.000	4.061	4.061	5.538
E1	44.781	3.307	9.537	6.229	11.104
E2	187.981	0.000	30.025	30.025	6.286
<u>P7</u>	4.815	0.000	4.815	4.815	0.000
E2-P7	198.270	0.000	35.778	35.778	7.730
NS2	88.204	0. 000	48.823	48.823	0.000
NS3	13.353	0.002	11.651	11.653	0.001
NS2/3	134.180	0.002	66.131	66.133	0.001
<u>NS4A</u>	0.223	0.000	0.224	0.223	0.000
NS4B	56.251	0.000	14.600	14.600	4.523
NS5A	91.617	0.000	16.556	16.556	5.040
NS5B	98.121	0.000	27.796	27.796	0.001
ORF	883.555	0.009	193.898	193.908	37.303

Notes: The likelihood ratio test (LRT) for putative genes under models M0, M1, M2, M3, M7, and M8. The E1, E2, E2-P7, NS4B, and NS5A in the Class A are bold, the N-pro, capsid, E0, NS2, NS3, NS2/3 and NS5B from the Class B are *italic* and the P7 and NS4A in the Class C are underlined. $_{0.05,\,4}$ means the chi-square value when the degree of freedom is 4 and the p value is 95%. $\chi^2_{0.05,\,2}$ is similar.

Table 4. Positive selection sites detected in genomic analysis

	E1	E2	NS4B	NS5A
M3	**80**	**72, 200**	**55**	**163**
M8	**80**	**72**, 75, 90, **200**	**55**	**163**

These sites (boldface) have a posterior probability greater than 95%, and that of the remaining (non bold) is between 90% and 95%.

The site-specific physicochemical selective pressures on E2 were investigated to elucidate any indications of the functional and structural changes. The composition of side chain, polarity, volume, polarity and/or volume, hydropathy and isoelectric point of the E2 protein were analyzed and the physicochemical selective pressures on E2 measured by altering the above-mentioned physicochemical properties of group one and group two (Tables 7 and 8). Clearly, LRTs for all physicochemical selective pressures supported the rejection of null model beside those that alter hydropathy. The ω values for physicochemical selective pressures on the E2 of group one were between 3.1 and 12.2, and the numeric order was ωc > ωv > ωp > ωpv > ω > ωh. On the other hand, those of group two were between 1.3 and 3.0, and the numeric order was ωh > ωv > ωp > ωpv > ω > ωc. Therefore, the physicochemical selective pressures of group one was far greater than those of group two and presented distinct evolutionary patterns between the two groups, 1.07% of the sites had a ω value of 5.015. In group one, 4.001%, 1.305%, 1.335%, 1.198%, 2.416%, and 0.325% of sites were under the physicochemical selective pressures to alter composition of the side chain, polarity, volume, polarity and/or volume, hydropathy and isoelectric point, respectively. This indicated that various proportions of sites were under specific physicochemical selective pressure. Moreover, the positively selected sites with a confidence score of higher than 95% were amino acid residues 72, 75 and 200. The amino acid 75 for altering composition of the side chain and hydropathy, 72 for altering polarity and isoelectric point, and all three for altering volume. Therefore, each positively selected site was under various physicochemical selective pressures. On the other hand, in group two, only the 197th amino acid was under positive selection with a confidence of 97.3%, and it was under the physicochemical selective pressures to alter volume and
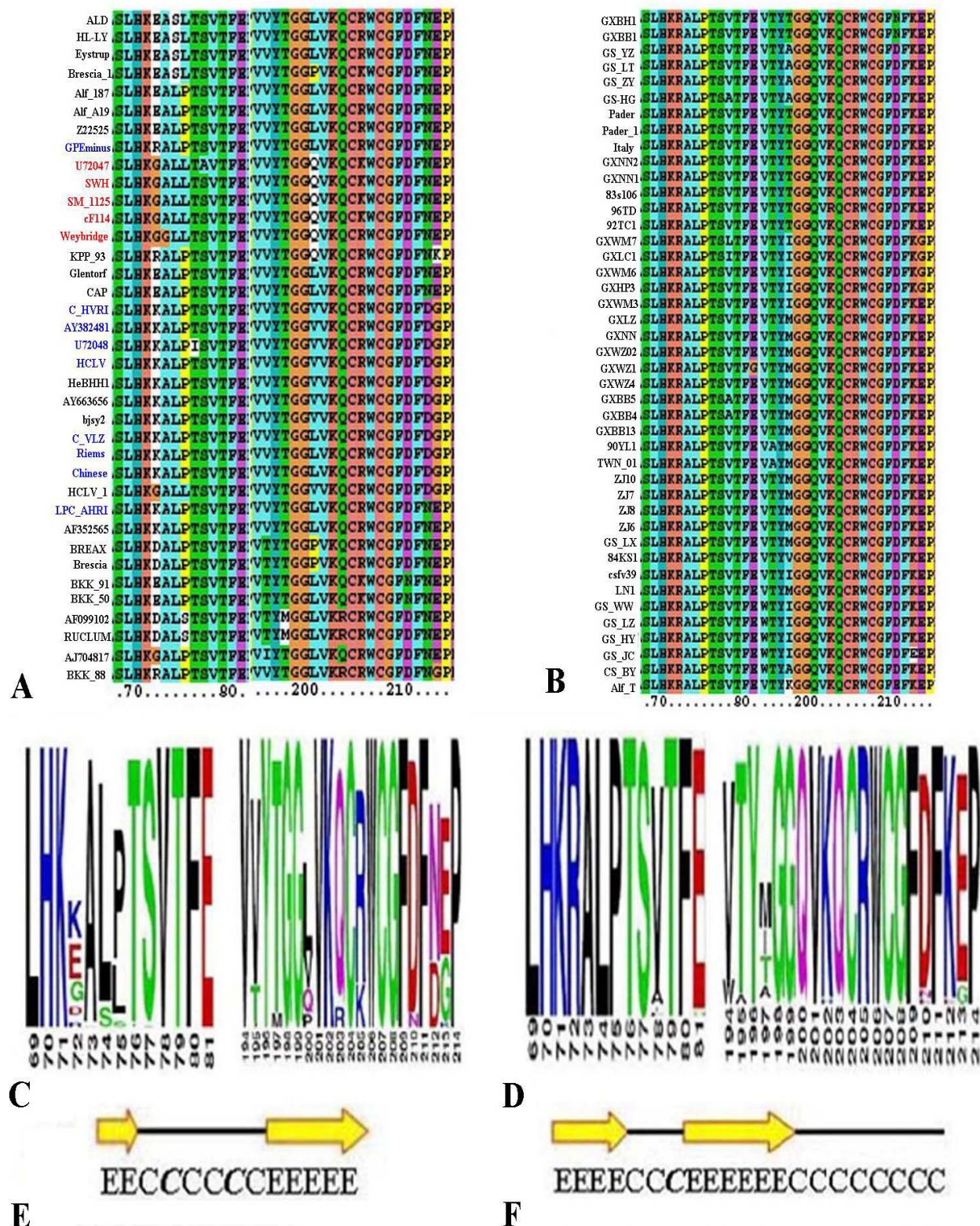
polarity and/or volume with a confidence of 90.6% and 92.5%, respectively; this supports that the selective pressure and the physicochemical selective pressures on the E2 of group one were far more higher than those of group two, and the patterns were distinct between group one and group two.

4.5. Evolutionary analysis of CSFV virulence

To investigate the roles of positively selected sites in the evolution of CSFV virulence, the complete E2 sequences of group one and group two (Figure 3A, B, C and D) were aligned. It indicated that the patterns of positively selected sites were [G72L75Q200], [K72P75 (VL) 200] and [R72P75Q200] for virulent strains, attenuated strains and the isolates of group two, respectively. The amino acid sites 72 and 75 of attenuated strains and the group two CSFV shared similar physicochemical properties, which were radically different from that of virulent strains. However, the virulent strains possessed the same 200th amino acid with the isolates of group two, but the attenuated strains had their specific ones which suffered the alteration of physicochemical properties. Further, we analyzed the effect, the structure and the function of E2 imposed by the positively selected sites described above. The secondary structure of E2 was predicted showing that all three sites were located in the loop region (Figure 3E and F). Due to the mobility of the loop region, the radical changes of physicochemical properties endowed all three sites with the feasibility to change the conformation of E2. In addition, the predictive numbers of contact networks within E2 were also changed with positive selection pressure on E2 (Figure 4). The contact map data showed conformational difference of E2 in various isolates. Furthermore, the relationship between positively selected sites and the glycosylation status of E2 was clarified to adequately understand the adaptive

Figure 3. The alignment of E2 on specific sites (amino acid 69-81,194-214)in group one (A) and group two (B). The identity of virulent strains and avirulent strains are red and blue, respectively. The logo display of A is C and that of B is D. The secondary structure of E2 on above-mentioned sites of group one (E) and group two (F) are shown, and the amino acid position 72, 75 and 200 are labeled *bold italic*.
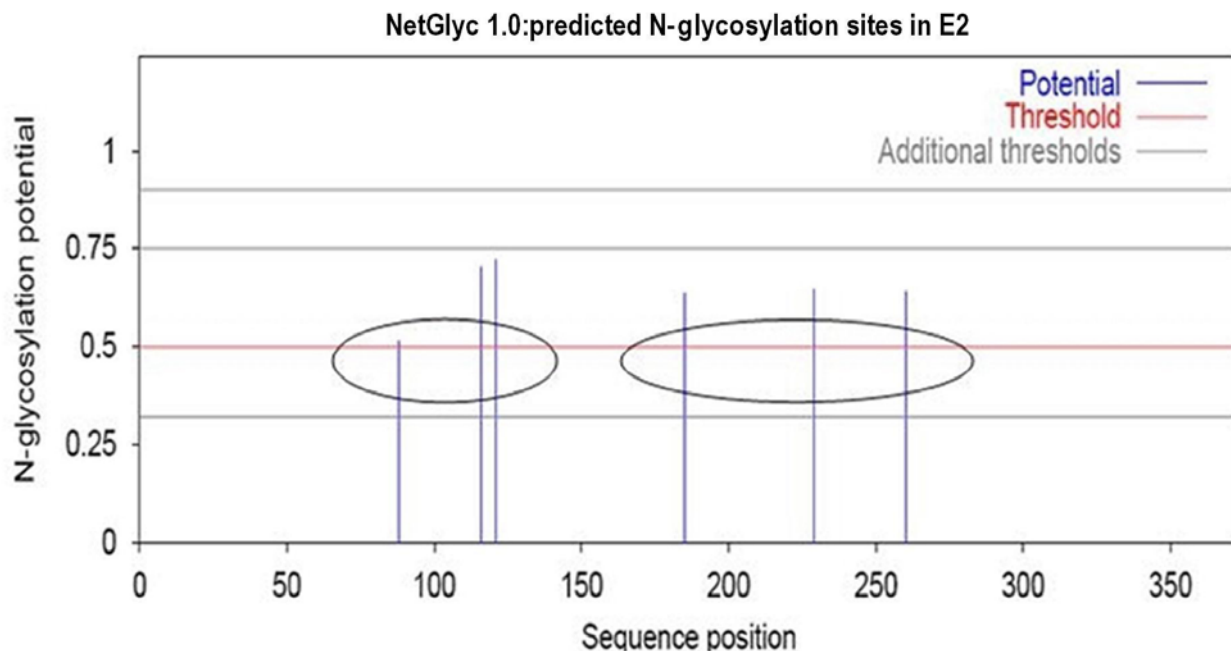
Figure 4. The contact map of E2 representative for avirulent isolate (ALD, Riems, and GS_HY), virulent isolate (cF114) and mild isolates (Pader, ZJ_7) are shown. The contact maps of additional isolates (8512A, csfv7960) from group 3 were also predicted. The number of contact networks and the name of each isolates are shown following the sequence, respectively.

Figure 5. The predicted N-glycosylation sites of E2 envelope protein. The N-glycosylation sites of E2 envelope protein are identified by a vertical bar. All these sites (amino acid 88,116,121,185,229,260,297) can be clustered into two groups indicated by the ellipse.

functional difference by predicting the N-glycosylation sites. The sites were found to be adjacent to the positively selected sites (Figure 5) and exposed in the full-chain structure of E2, predicted by the Robetta server, due to the absence of a suitable template structure (see additional figure 1). This indicated that the physicochemical changes of positively selected sites could change the glycosylation status of E2. The N-glycosylation sites have the potential to prevent the epitope from immune defense, by mediating the binding of carbohydrates to the viral envelope. The N-linked glycosylation status of E2 glycoprotein influences virulence of CSFV in swine (15). Therefore, it is reasonable to assume that positive selection has the ability to change the structure and function of E2 by remodeling its glycosylation. The radical changes of physicochemical properties driven by positive selection modify the structure and function of E2; this is a key determinant of CSFV virulence (17-19), validated by the fact that the reaction activity with specific antibody of E2 with the motif [G72L75] was different from that of E2 with the motif [(KR) 72P75] (62).

5. DISCUSSION

Using evolutionary analysis to investigate a novel aspect of CSFV virulence, statistically significant variations in selective pressures on CSFV among groups and among sites were found, suggesting that positive selection played a role in the evolution of CSFV virulence. Based on the dataset 85E2, the phylogenetic tree is highly similar to what was previously reported (4-6), indicating that the distribution of CSFV is in conflict with its geographical origin; this is possibly due to the exchanging

commerce of live pigs and related products between countries and districts suggesting the effect of human activities on CSFV.

Recombination is also an important mechanism of virus evolution and adaptation. However, CSFV has a serotype (4-6), which results only in the small probability of cross-infection for various CSFV isolates. In addition, CSFV is a single strand RNA virus. Both of the two factors could block the recombination between various isolates, and the recombination is still not supported by the phylogenetic analysis (4). Moreover, it has been suggested that the recombination, if present in our sequence data, was likely to be infrequent, which hardly affect the inference of positive selection (63, 64). So the recombination could not be important in the CSFV evolution and adaptation.

The likelihood parameters and strength of each protein encoding sequence (putative genes) of 36 CSFV genomic sequences indicated that both envelope proteins (E0, E1, E2, E2-P7) and nonstructural proteins (N-pro and NS5A) have a significant number of positively selected sites (Table 1 and 4). Thus, it suggested that the selective pressures from both immune system and host cell could affect the adaptation of CSFV. The lowest mean ω occupied by NS3 indicates that its structure and function were highly constrained and could be essential for the life cycle of CSFV (Table 1 and 2). Furthermore, the ω value of E1 and NS5A (4.590 and 3.436 respectively) were twice as high as that of E2, E2-P7, or NS4B with ω values close to 2, although the number of positively selected sites was less than E2, E2-P7, or NS4B (Table 1). The positively selected sites of E1 and NS5A might therefore play some special

Table 5. Maximum likelihood estimates (MLEs) for E2 gene on group 1 lineage

	Site Class	*Proportion*	Background ω	Foreground ω	Site for foreground lineage	ts/tv
MA	0	0.815	0.085	0.085	1R, 55V*, 72E**, 75P*,200L*,205R*, 290R*, 299A*,364I*	8.062
	1	0.081	1.000	1.000		
	2a	0.095	0.085	1.286		
	2b	0.009	1.000	1.286		
MA1	0	0.790	0.082	0.082		8.018
	1	0.078	1.000	1.000		
	2a	0.120	0.082	1.000		
	2b	0.012	1.000	1.000		
LRT 9062.645259-9063.385143=1.480 $\chi^2_{0.25, 1}$=1.323 $\chi^2_{0.05, 1}$=3.841 $\chi^2_{0.01, 1}$= 6.634						

Maximum likelihood estimates (MLEs) for E2 gene coming from CSFV group 1 as the foreground lineage. These sites with the superscript "**" or "*" have a posterior probability greater than 99% and 95%, respectively. The posterior probability of the remaining is between 90% and 95%.

Table 6. Maximum likelihood estimates (MLEs) for E2 gene on group 2 lineage

	Site Class	Proportion	Background ω	Foreground ω	Positively selected sites for foreground lineage	ts/tv
MA	0	0.863	0.097	0.097	78 V (0.655), 192 E (0.602), 197 T (0.852), 235 I (0.547)	8.121
	1	0.131	1.000	1.000		
	2a	0.005	0.097	1.000		
	2b	0.001	1.000	1.000		
MA1	0	0.863	0.097	0.097		8.121
	1	0.131	1.000	1.000		
	2a	0.005	0.097	1.000		
	2b	0.001	1.000	1.000		
LRT 9089.049224-9089.048871=0.001 $\chi^2_{0.25, 1}$=1.323 $\chi^2_{0.05, 1}$=3.841 $\chi^2_{0.01, 1}$= 6.635						

Maximum likelihood estimates (MLEs) for E2 gene of CSFV group 2 as the foreground. The posterior probabilities of the positively selected sites are in the bracket

Table 7. Physicochemical selective pressures on E2 gene of CSFV group 1 lineage

Selective pressure	Model	Lnl	Parameter estimates	-2ΔLnl	Positive select sites
Nonsynonymous	M7	-4379.072	p=0.244 q=0.728 κ=8.192	23.841	72D**,75P*, 200P*
	M8	-4367.152	p=0.337 q=1.122 κ=8.556 p1=0.011 **ω=5.015**		
Polarity-altering	M7	-4398.739	p=0.076 q=0.274 κ=8.270	8.802	72D**
	M8	-4394.338	p=0.271 q=1.211 κ=8.441 p1=0.013 **ω=5.799**		
Volume-altering	M7	-4378.403	p=0.075 q=0.226 κ=8.054	22.302	72D**, 75P**, 200P*
	M8	-4367.252	p=0.219 q=0.798 κ=8.280 p1=0.013 **ω=6.288**		
Polarity-and/or Volume-altering	M7	-4370.863	p=0.085 q=0.290 κ=8.125	15.693	72D**,75P*, 200P*
	M8	-4363.016	p=0.266 q=1.079 κ=8.364 p1=0.012 **ω=5.241**		
Hydrophobicity-altering	M7	-4415.493	p=0.085 q=0.254 κ=8.106	4.547	75P*
	M8	-4413.220	p=0.255 q=1.060 κ=8.200 p1=0.024 **ω=3.161**		
Composition-altering	M7	-4403.635	p=0.088 q=0.270 κ=8.025	2.723	75P*
	M8	-4402.273	p=0.271 q=1.267 κ=8.236 p1=0.040 **ω=2.428**		
IsoelectricPoint-altering	M7	-4412.282	p=0.113 q=0.306 κ=8.718	15.055	72D*
	M8	-4404.754	p=0.287 q=0.996 κ=8.693 p1=0.003 **ω=12.354**		
$\chi^2_{0.05,2}$=5.99146 $\chi^2_{0.025,2}$=7.37776 $\chi^2_{0.01,2}$=9.21034 $\chi^2_{0.005,2}$=10.59663					

Maximum likelihood estimates (MLEs) and likelihood ratio tests for physicochemical selective pressures on E2 gene of CSFV group 1. All the ω values (boldface) are greater than 1 indicating the positive selection on a physicochemical property. These sites with the superscript "**" or "*" have a posterior probability greater than 99% and 95%, respectively.

Table 8. Physicochemical selective pressures on E2 gene of CSFV group 2 lineage

Selective pressure	Model	Lnl	Parameter estimates	-2ΔLnl	Positive select sites
Nonsynonymous	M7	-4914.188	p=0.376 q=1.954 κ=7.196	10.700	197T (0.973*)
	M8	-4908.838	p=0.766 q=5.308 κ=7.219 p1=0.027 **ω=1.460**		
Polarity-altering	M7	-4926.414	p=0.043 q=0.164 κ=7.518	5.724	
	M8	-4923.552	p=0.161 q=1.227 κ=7.508 p1=0.039 **ω=1.967**		
Volume-altering	M7	-4939.500	p=0.056 q=0.201 κ=7.531	22.687	197T (0.906)
	M8	-4928.156	p=0.179 q=1.014 κ=7.185 p1=0.002 **ω=2.250**		
Polarity-and/or Volume-altering	M7	-4933.745	p=0.064 q=0.221 κ=7.666	25.168	197T (0.925)
	M8	-4921.161	p=0.189 q=1.041 κ=7.291 p1=0.005 **ω=1.915**		
Hydrophobicity-altering	M7	-4932.223	p=0.038 q=0.148 κ=7.488	0.5419	
	M8	-4931.953	p=0.143 q=0.673 κ=7.295 p1=0.000 **ω=2.908**		
Composition-altering	M7	-4920.312	p=0.041 q=0.147 κ=7.548	3.691	
	M8	-4918.467	p=0.139 q=0.978 κ=7.298 p1=0.062 ω=1.000		
IsoelectricPoint-altering	M7	-4937.321	p=0.053 q=0.198 κ=7.642	16.061	
	M8	-4929.291	p=0.180 q=1.053 κ=7.363 p1=0.000 ω=1.000		

$\chi^2_{0.05,2}$=5.99146 $\chi^2_{0.025,2}$=7.37776 $\chi^2_{0.01,2}$=9.21034 $\chi^2_{0.005,2}$=10.59663

Maximum likelihood estimates (MLEs) and likelihood ratio tests for physicochemical selective pressures on E2 gene of group 2 dataset. All the ω values indicating the positive selection on a physicochemical property are bold. The posterior probabilities of the positively selected sites are in the bracket. The site with the superscript "*" have a posterior probability greater than 95%.

roles other than that of E2, E2-P7 or NS4B in the adaptation of CSFV. On the other hand, existing studies suggest that Npro, E0, E1 and E2, with higher positively selective pressure than the other proteins encoded by CSFV, can influence the virulence of CSFV significantly (9, 15, 17-19, 21). Hence, the evolution of CSFV virulence could be a way of molecular adaptation.

The structural and functional role of an amino acid residue within a protein is determined by its position and physicochemical properties. The site-specific physicochemical selective pressures are important to investigate the evolutionary mechanism of a protein. The envelope protein E2 can reflect the greatest extent among all the putative genes of CSFV (Table 2) for changes in selection pressures. Therefore, the respective pattern of selective pressures on the specific lineage of CSFV was further investigated, by detecting physicochemical selective pressures of E2 along the specific lineage. The composition of side chain, polarity, volume, polarity and/or volume, hydropathy and isoelectric point were included in the analysis (49, 51, 65). It was found that the positively selected sites along group one (amino acid 55, 72, 75, 200, 205, 290, 299 and 364) and group two (amino acid 197) are almost fully overlapped with the antigenic domain of E2. Since the antigenic domain might be the interacting domain for the binding of virus with the host cell, the difference among the positively selected sites could lead to the dissimilar antigenic patterns and the distinct ability for receptor binding (66-71). The selective pressure and the physicochemical selective pressures on the E2 of group one was found to be much higher than that of group two (Table 5, 6, 7 and 8). Considering the variation of CSFV virulence (Figure 1 and see also additional file 1), it was demonstrated that the selective pressure on E2 decreased with the decline of CSFV virulence. Thus, attenuation driven by positive selection could contribute to the adaptation of CSFV.

The variation of virulence is believed to be crucial in understanding the adaptation of CSFV. There is also evidence that the Npro, E0, E1 and E2 may contribute to the attenuation of CSFV virulence (9, 16-25). Risatti *et al.* have previously found that E2 possesses the most significance in attenuation from virulent strain Brescia to vaccine strain CS. Furthermore, substitutions of amino acid residue 197 and 200 of E2 positively selected in our study, as it has been previously described (18). The site directed mutation of an epitope between amino acid residue 140 and 149 of E2, routinely used for CSF diagnostics, attenuated the virulent strain Brescia (17). This epitope is conservative among known CSFV isolates and differs greatly from homologous sequences of other pestiviruses. Thus, E2 glycoprotein is pivotal to maintenance of the virulence of the CSFV isolates. On the other hand, glycosylation could be essential for the correct folding and subsequent secretion

of E2. Recently, it has been demonstrated that N-linked glycosylation status of E2 influences the virulence of Brescia strain in swine (15). Furthermore, the infectivity and cell tropism of chimeric pestiviruses are consistent with the E2 gene donor (72, 73). E2 is also the most immunogenic glycoprotein (74) and the major ligand for target cell receptor (14, 75). Although the mechanism underlying the evolution of CSFV virulence remains obscure, it is conceivable that the attenuation is involved in the process of virus attachment and/or entry into the host cell. Thus, the influence of E2 on virulence could be more significant than any other protein encoded by CSFV. Here, E2 was regarded as a probe for investigating the evolution of CSFV virulence, because it is also potentially under the highest positively selective pressure.

Based on findings from this study and previous work on the attenuated mechanism of CSFV, a model is proposed. Positive selection drives the physicochemical alterations of amino acid residues 72, 75 and 200 within E2; These alterations influence the folding and glycosylation of E2, especially that of the N-terminal, leading to disappearance and rebirth of specific domain or epitope. The conformational and glycosylated changes of E2 alter the interaction between E2 and proteins from the immune system or other host factors, especially for cellular receptors. Similar evolution may happen in part or for all of Npro, E0, and E1. All the above changes modify the protein interaction network among proteins from CSFV and host. Ultimately, the network modifications drive the attenuation of CSFV.

Previous work supports this model (62). The research investigated the reaction activities of a monoclonal antibody against Shimen strain, and Chinese strain with 33 various fields of CSFV isolates from 17 provinces in China. Isolates with stronger reaction (ELISA value higher than 0.3) possessed the [G72L75] motif, which was consistent with the pattern of positively selected sites in the E2 sequence of virulent strains (Figure 3A). Moreover, isolates showing weaker reaction (ELISA value) with the monoclonal antibody possessed the [(KR) 72P75] motif; this was consistent with the pattern of positively selected sites in the E2 sequence of avirulent strains in group one and the isolates in group two (Figure 3A and B). Simultaneously, all strains, except for three isolates which were detected with much weaker reaction, could not react with the monoclonal antibody against the Chinese strain. Chinese strain is the outcome of the serial passage of Shimen strain. It is thought that the evolution of amino acids 72 and 75, corresponding to the reaction ability of E2 with monoclonal antibody and the epitope recognized by the monoclonal antibody against Chinese strain, was not present in the remaining 30 isolates. Alternatively, the Chinese strain presented an interesting epitope not detected in other isolates, by the selective process of avirulent strains accompanied by "disappearance and rebirth" of specific epitope. In addition, replacing the complete E2 sequence with homologous sequence from the CS strain could attenuate the Brescia strain, but the region coding for all the structural proteins of Brescia could not rescue the CS strain (18). The epitope between amino acid residues

140 and 148, highly conserved across all known CSFV isolates, was found to be a virulence determinant within the E2 structural protein (17). These results indicate the occurrence of "disappearance and rebirth" of specific domains (epitopes included) during the attenuation of CSFV. The antibody binding could also be responsible for changes in the efficiency of neutralization and entry of the enveloped virus, as it is the case in Human Immunodeficiency Virus Type 1 (76). This suggests that the structural and functional changes of E2 are driven by positive selection on amino acids 72 and 75, associated with the binding activity of monoclonal antibody with CSFV isolates, leading to variation of CSFV virulence.

According to this model, positive selection plays key roles in the evolution of CSFV virulence. It is well known that virulence of the Chinese strain is not rescued when it underwent more than 20 serial passages in swine. It is depicted in the model that the attenuation of CSFV suffers many mutations within various components, driven by positive selection, because it would be impossible to rescue virulence given the number of variables in the attenuation of CSFV. Interestingly, all the avirulent strains in group one originate from the serial passage of the virulent strains, and the process may be driven through positive selection (table1, 2 and 3). Previously, the epidemic strains were primarily the virulent strains of group one, but the moderate and low virulent strains of group two dominate the recent CSF outbreaks (Figure 2 and also see additional file 1); they undergo the least number of passages compared to the avirulent strains. And the influence of positive selection on them is less than avirulent CSFV. It is commonly understood that attenuation is advantageous for viruses facing natural selection, since it increases the probability of the virus infecting a new host. Therefore, the attenuating process of epidemic strains may possibly be driven through positive selection. The recent CSFV epidemic of group two were perhaps derived from group one, and require further investigation .

7. REFERENCES

1. C. R. Pringle: Virus taxonomy--1999. The universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. *Arch Virol*, 144(2), 421-429 (1999)

2. G. Floegel-Niesmann, C. Bunzenthal, S. Fischer and V. Moennig: Virulence of recent and former classical swine fever virus isolates evaluated by their clinical and pathological signs. *Journal of Veterinary Medicine Series*

B-Infectious Diseases and Veterinary Public Health, 50(5), 214-220 (2003)

3. V. Moennig, G. Floegel-Niesmann and I. Greiser-Wilke: Clinical signs and epidemiology of classical swine fever: A review of new knowledge. *Veterinary Journal*, 165(1), 11-20 (2003)

4. D. J. Paton, A. McGoldrick, I. Greiser-Wilke, S. Parchariyanon, J. Y. Song, P. P. Liou, T. Stadejek, J. P. Lowings, H. Bjorklund and S. Belak: Genetic typing of classical swine fever virus. *Veterinary Microbiology*, 73(2-3), 137-157 (2000)

5. P. Lowings, G. Ibata, J. Needham and D. Paton: Classical swine fever virus diversity and evolution. *Journal of General Virology*, 77, 1311-1321 (1996)

6. S. Kumar, H. K. Panda, B. C. Kar and B. K. Panda: Adaptation and physicochemical characterization of "MB" strain of infectious bursal-disease virus. *Indian Veterinary Journal*, 82(9), 923-925 (2005)

7. C. Mittelholzer, C. Moser, J. D. Tratschin and M. A. Hofmann: Analysis of classical swine fever virus replication kinetics allows differentiation of highly virulent from avirulent strains. *Veterinary Microbiology*, 74(4), 293-308 (2000)

8. V. Moennig: Introduction to classical swine fever: virus, disease and control policy. *Vet Microbiol*, 73(2-3), 93-102 (2000)

9. G. Meyers, A. Saalmuller and M. Buttner: Mutations Abrogating the RNase Activity in Glycoprotein Erns of the Pestivirus Classical Swine Fever Virus Lead to Virus Attenuation. *J. Virol.*, 73(12), 10224-10235 (1999)

10. T. Rumenapf, G. Meyers, R. Stark and H. J. Thiel: Molecular characterization of hog cholera virus. *Arch Virol Suppl*, 3, 7-18 (1991)

11. M. Yu, L. F. Wang, B. J. Shiell, C. J. Morrissy and H. A. Westbury: Fine mapping of a c-terminal linear epitope highly conserved among the major envelope glycoprotein E2 (gp51 to gp54) of different pestiviruses. *Virology*, 222(1), 289-292 (1996)

12. P. A. van Rijn, G. K. Miedema, G. Wensvoort, H. G. van Gennip and R. J. Moormann: Antigenic structure of envelope glycoprotein E1 of hog cholera virus. *The Journal of Virology*, 68(6), 3934-3942 (1994)

13. P. A. Vanrijn, R. G. P. Vangennip, E. J. Demeijer and R. J. M. Moormann: A Preliminary Map of Epitopes on Envelope Glycoprotein-E1 of Hcv Strain Brescia. *Veterinary Microbiology*, 33(1-4), 221-230 (1992)

14. Z. Wang, Y. C. Nie, P. G. Wang, M. X. Ding and H. K. Deng: Characterization of classical swine fever virus entry by using pseudotyped viruses: E1 and E2 are sufficient to mediate viral entry. *Virology*, 330(1), 332-341 (2004)

15. G. R. Risatti, L. G. Holinka, I. F. Sainz, C. Carrillo, Z. Lu and M. V. Borca: N-linked glycosylation status of classical swine fever virus strain Brescia E2 glycoprotein influences virulence in swine. *Journal of Virology*, 81(2), 924-933 (2007)

16. G. R. Risatti, L. G. Holinka, I. Fernandez Sainz, C. Carrillo, G. F. Kutish, Z. Lu, J. Zhu, D. L. Rock and M. V. Borca: Mutations in the carboxyl terminal region of E2 glycoprotein of classical swine fever virus are responsible for viral attenuation in swine. *Virology*, 364(2), 371-382 (2007)

17. G. R. Risatti, L. G. Holinka, C. Carrillo, G. F. Kutish, Z. Lu, E. R. Tulman, I. F. Sainz and M. V. Borca: Identification of a novel virulence determinant within the E2 structural glycoprotein of classical swine fever virus. *Virology*, 355(1), 94-101 (2006)

18. G. R. Risatti, M. V. Borca, G. F. Kutish, Z. Lu, L. G. Holinka, R. A. French, E. R. Tulman and D. L. Rock: The E2 glycoprotein of classical swine fever virus is a virulence determinant in swine. *Journal of Virology*, 79(6), 3787-3796 (2005)

19. H. G. P. van Gennip, A. C. Vlot, M. M. Hulst, A. J. de Smit and R. J. M. Moormann: Determinants of virulence of classical swine fever virus strain Brescia. *Journal of Virology*, 78(16), 8812-8823 (2004)

20. N. Ruggli, A. Summerfield, A. R. Fiebach, L. Guzylack-Piriou, O. Bauhofer, C. G. Lamm, S. Waltersperger, K. Matsuno, L. Liu, M. Gerber, K. H. Choi, M. A. Hofmann, Y. Sakoda and J. D. Tratschin: Classical swine fever virus can remain virulent after specific elimination of the interferon regulatory factor 3-degrading function of Npro. *J Virol*, 83(2), 817-29 (2009)

21. D. Mayer, M. A. Hofmann and J. D. Tratschin: Attenuation of classical swine fever virus by deletion of the viral N-pro gene. *Vaccine*, 22(3-4), 317-328 (2004)

22. B. A. Tews, E. M. Schurmann and G. Meyers: Mutation of cysteine 171 of pestivirus Erns RNase prevents homodimer formation and leads to attenuation of classical swine fever virus. *J Virol* (2009)

23. I. F. Sainz, L. G. Holinka, Z. Lu, G. R. Risatti and M. V. Borca: Removal of a N-linked glycosylation site of classical swine fever virus strain Brescia Erns glycoprotein affects virulence in swine. *Virology*, 370(1), 122-129 (2008)

24. I. Fernandez-Sainz, L. G. Holinka, B. K. Gavrilov, M. V. Prarat, D. Gladue, Z. Lu, W. Jia, G. R. Risatti and M. V. Borca: Alteration of the N-linked glycosylation condition in E1 glycoprotein of Classical Swine Fever Virus strain Brescia alters virulence in swine. *Virology*, 386(1), 210-6 (2009)

25. G. R. Risatti, L. G. Holinka, Z. Lu, G. F. Kutish, E. R. Tulman, R. A. French, J. H. Sur, D. L. Rock and M. V.

Borca: Mutation of El glycoprotein of classical swine fever virus affects viral virulence in swine. *Virology*, 343(1), 116-127 (2005)

26. C. Casado, S. Garcia, C. Rodriguez, J. del Romero, G. Bello and C. Lopez-Galindez: Different evolutionary patterns are found within human immunodeficiency virus type 1-infected patients. *J Gen Virol*, 82(Pt 10), 2495-2508 (2001)

27. H. Barroso and N. Taveira: Evidence for negative selective pressure in HIV-2 evolution *in vivo*. *Infect Genet Evol*, 5(3), 239-246 (2005)

28. S. D. W. Frost, H. F. Gunthard, J. K. Wong, D. Havlir, D. D. Richman and A. J. L. Brown: Evidence for positive selection driving the evolution of HIV-1 env under potent antiviral therapy. *Virology*, 284(2), 250-258 (2001)

29. D. T. Haydon, A. D. Bastos, N. J. Knowles and A. R. Samuel: Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics*, 157(1), 7-15 (2001)

30. M. A. Fares, A. Moya, C. Escarmis, E. Baranowski, E. Domingo and E. Barrio: Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. *Molecular Biology and Evolution*, 18(1), 10-21 (2001)

31. M. W. Torres, R. L. Correa and C. G. Schrago: Analysis of differential selective forces acting on the coat protein (P3) of the plant virus family Luteoviridae. *Genet Mol Res*, 4(4), 790-802 (2005)

32. J. C. Ameisen, J. D. Lelievre and O. Pleskoff: HIV/host interactions: new lessons from the Red Queen's country. *Aids*, 16, S25-S31 (2002)

33. M. Choisy, C. H. Woelk, J.-F. Guegan and D. L. Robertson: Comparative Study of Adaptive Molecular Evolution in Different Human Immunodeficiency Virus Groups and Subtypes. *J. Virol.*, 78(4), 1962-1970 (2004)

34. S. A. A. Travers, M. J. O'Connell, G. P. McCormack and J. O. McInerney: Evidence for Heterogeneous Selective Pressures in the Evolution of the env Gene in Different Human Immunodeficiency Virus Type 1 Subtypes. *J. Virol.*, 79(3), 1836-1841 (2005)

35. B. Moury: Differential selection of genes of cucumber mosaic virus subgroups. *Molecular Biology and Evolution*, 21(8), 1602-1611 (2004)

36. C.-Y. Zhang, J.-F. Wei and S.-H. He: Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups. *BMC Microbiology*, 6(1), 88 (2006)

37. X. X. Qu, P. Hao, X. J. Song, S. M. Jiang, Y. X. Liu, P. G. Wang, X. Rao, H. D. Song, S. Y. Wang, Y. Zuo, A. H. Zheng, M. Luo, H. L. Wang, F. Deng, H. Z. Wang, Z. H. Hu, M. X. Ding, G. P. Zhao and H. K. Deng: Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *Journal of Biological Chemistry*, 280(33), 29588-29595 (2005)

38. K. V. Holmes: Adaptation of SARS coronavirus to humans. *Science*, 309(5742), 1822-1823 (2005)

39. S. Bonhoeffer and P. Sniegowski: Virus evolution - The importance of being erroneous. *Nature*, 420(6914), 367-369 (2002)

40. Z. Yang and R. Nielsen: Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Mol Biol Evol*, 19(6), 908-917 (2002)

41. Z. Yang, R. Nielsen, N. Goldman and A.-M. K. Pedersen: Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics*, 155(1), 431-449 (2000)

42. Z. H. Yang and J. P. Bielawski: Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15(12), 496-503 (2000)

43. J. D. Thompson, D. G. Higgins and T. J. Gibson: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22), 4673-4680 (1994)

44. J. Felsenstein: PHYLIP (Phylogeny Inference Package) version 3.6. *Distributed by the author Department of Genome Sciences, University of Washington, Seattle.* (2004)

45. J. Felsenstein: PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164-166 (1989)

46. Z. Yang, W. S. Wong and R. Nielsen: Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol*, 22(4), 1107-1118 (2005)

47. J. Zhang, R. Nielsen and Z. Yang: Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*, 22(12), 2472-2479 (2005)

48. Z. H. Yang: PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, 13(5), 555-556 (1997)

49. W. S. Wong, R. Sainudiin and R. Nielsen: Identification of physicochemical selective pressure on protein encoding nucleotide sequences. *BMC Bioinformatics*, 7, 148 (2006)

50. B. Gaschen, J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B. H. Hahn, T.

Bhattacharya and B. Korber: Diversity Considerations in HIV-1 Vaccine Selection. *Science*, 296(5577), 2354-2360 (2002)

51. R. Sainudiin, W. S. Wong, K. Yogeeswaran, J. B. Nasrallah, Z. Yang and R. Nielsen: Detecting site-specific physicochemical selective pressures: applications to the Class I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. *J Mol Evol*, 60(3), 315-326 (2005)

52. B. Rost: PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol*, 266, 525-539 (1996)

53. M. Ouali and R. D. King: Cascaded multiple classifiers for secondary structure prediction. *Protein Sci*, 9(6), 1162-1176 (2000)

54. L. J. McGuffin, K. Bryson and D. T. Jones: The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404-405 (2000)

55. D. T. Jones: Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, 292(2), 195-202 (1999)

56. L. J. McGuffin and D. T. Jones: Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics*, 19(7), 874-881 (2003)

57. D. T. Jones: GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol*, 287(4), 797-815 (1999)

58. D. T. Jones: Do transmembrane protein superfolds exist? *FEBS Lett*, 423(3), 281-5 (1998)

59. D. T. Jones, W. R. Taylor and J. M. Thornton: A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, 33(10), 3038-3049 (1994)

60. R. M. MacCallum: Striped sheets and protein contact prediction. *Bioinformatics*, 20 Suppl 1, I224-I231 (2004)

61. M. Punta and B. Rost: PROFcon: novel prediction of long-range contacts. *Bioinformatics*, 21(13), 2960-2968 (2005)

62. Q. Wang: Characterization of Pathogenicity of Classical Swine Fever Virus Isolates and Establishment of Classical Swine Fever Epidemiology Information System in China. *Ph.D. thesis.*, College of Veterinary Medicine, China Agriculural University. Beijing, P.R.China (2006)

63. M. Anisimova, R. Nielsen and Z. Yang: Effect of Recombination on the Accuracy of the Likelihood Method for Detecting Positive Selection at Amino Acid Sites. *Genetics*, 164(3), 1229-1236 (2003)

64. M. R. Pie: The influence of phylogenetic uncertainty on the detection of positive darwinian selection. *Molecular Biology and Evolution*, 23(12), 2274-2278 (2006)

65. X. Xia and W. H. Li: What amino acid properties affect protein evolution? *J Mol Evol*, 47(5), 557-564 (1998)

66. Y. X. He, J. J. Li, L. Y. Du, X. X. Yan, G. G. Hu, Y. S. Zhou and S. B. Jiang: Identification and characterization of novel neutralizing epitopes in the receptor-binding domain of SARS-CoV spike protein: Revealing the critical antigenic determinants in inactivated SARS-CoV vaccine. *Vaccine*, 24(26), 5498-5508 (2006)

67. Y. X. He, J. J. Li and S. B. Jiang: A single amino acid substitution (R441A) in the receptor-binding domain of SARS coronavirus spike protein disrupts the antigenic structure and binding activity. *Biochemical and Biophysical Research Communications*, 344(1), 106-113 (2006)

68. E. E. Fry, J. W. I. Newman, S. Curry, S. Najjam, T. Jackson, W. Blakemore, S. M. Lea, L. Miller, A. Burman, A. M. Q. King and D. I. Stuart: Structure of Foot-and-mouth disease virus serotype A10(61) alone and complexed with oligosaccharide receptor: receptor conservation in the face of antigenic variation. *Journal of General Virology*, 86, 1909-1920 (2005)

69. J. C. Whitbeck, M. I. Muggeridge, A. H. Rux, W. F. Hou, C. Krummenacher, H. Lou, A. van Geelen, R. J. Eisenberg and G. H. Cohen: The major neutralizing antigenic site on herpes simplex virus glycoprotein D overlaps a receptor-binding domain. *Journal of Virology*, 73(12), 9879-9890 (1999)

70. C. M. Ruiz-Jarabo, N. Sevilla, M. Davila, G. Gomez-Mariano, E. Baranowski and E. Domingo: Antigenic properties and population stability of a foot-and-mouth disease virus with an altered Arg-Gly-Asp receptor-recognition motif. *Journal of General Virology*, 80, 1899-1909 (1999)

71. H. Liebermann, R. Mentel, U. Bauer, P. Pring-Akerblom, R. Dolling, S. Modrow and W. Seidel: Receptor binding sites and antigenic epitopes on the fiber knob of human adenovirus serotype 3. *Journal of Virology*, 72(11), 9121-9130 (1998)

72. D. L. Liang, I. F. Sainz, I. H. Ansari, L. Gil, V. Vassilev and R. O. Donis: The envelope glycoprotein E2 is a determinant of cell culture tropism in ruminant pestiviruses. *Journal of General Virology*, 84, 1269-1274 (2003)

73. H. G. P. van Gennip, P. A. van Rijn, M. N. Widjojoatmodjo, A. J. de Smit and R. J. M. Moormann: Chimeric classical swine fever viruses containing envelope protein E-RNS or E2 of bovine viral diarrhoea virus protect pigs against challenge with CSFV and induce a distinguishable antibody response. *Vaccine*, 19(4-5), 447-459 (2000)

74. M. M. Hulst, D. F. Westra, G. Wensvoort and R. J. Moormann: Glycoprotein E1 of hog cholera virus expressed in insect cells protects swine from hog cholera. *J. Virol.*, 67(9), 5435-5442 (1993)

75. M. M. Hulst and R. J. M. Moormann: Inhibition of pestivirus infection in cell culture by envelope proteins E-rns and E2 of classical swine fever virus: E-rns and E2 interact with different receptors. *Journal of General Virology*, 78, 2779-2787 (1997)

76. X. Yang, I. Lipchina, S. Cocklin, I. Chaiken and J. Sodroski: Antibody Binding Is a Dominant Determinant of the Efficiency of Human Immunodeficiency Virus Type 1 Neutralization. *J. Virol.*, 80(22), 11404-11408 (2006)

Send correspondence to: Jianguo Chen, Department of Cell Biology and Genetics, College of Life Sciences, Peking University, Beijing 100871, China, Tel: 86-10-62755786, Fax: 86-10-62754427, E-mail: chenjg@pku.edu.cn

http://www.bioscience.org/current/vol2E.htm