

# Complete genome sequences of porcine reproductive and respiratory syndrome viruses: perspectives on their temporal and spatial dynamics

Sook Hee Yoon · Hyekwon Kim · Jaemin Kim ·  
Hak-Kyo Lee · Bongkyun Park · Heebal Kim

Received: 2 November 2012 / Accepted: 26 September 2013 / Published online: 12 October 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** Porcine reproductive and respiratory syndrome virus (PRRSV) has been investigated extensively at the molecular level. Nevertheless, genome wide study on the temporal and spatial dynamics of the virus is non-existent. To explore this topic, we determined complete coding genome sequences for four PRRSV isolates and analyzed them together with 122 global published ones using the Bayesian coalescent approach as well as Bayesian inferences and maximum likelihood methods. All EU-type viruses belonged to one of two groups or were unclassified (5 isolates), and all NA-type isolates were divided into one of three major

groups or were unclassified (1 isolate). Here, there was no apparent association between temporal or geographic origin and heterogeneity of global PRRSVs. Of the eight ORFs, ORF1a showed the most powerful evolutionary signal. Our findings also indicated that the PRRS virus evolved at a rate of  $1.98 \times 10^{-3}$  substitutions/site/year, and the most recent common ancestor of the virus existed 786.4 years ago. Here, EU-type viruses segregated 115.7 years ago, while NA-type isolates diverged 179.8 years before the present. In addition, our reconstruction of the effective population size depicted five phases of epidemic growth: an initial constant, followed sequentially by slow decrease, slight increase, sharp decline, and then a rapid expansion approaching the present.

Sook Hee Yoon and Hyekwon Kim have contributed in equal measure to the work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11033-013-2802-1) contains supplementary material, which is available to authorized users.

S. H. Yoon · H. Kim (✉)  
Department of Agricultural Biotechnology and Research  
Institute for Agriculture and Life Sciences, Seoul National  
University, Seoul 151-742, Republic of Korea  
e-mail: heebal@snu.ac.kr

H. Kim · B. Park (✉)  
Department of Veterinary Virology, College of Veterinary  
Medicine, Seoul National University, Seoul 151-742,  
Republic of Korea  
e-mail: parkx026@snu.ac.kr

J. Kim  
Interdisciplinary Program in Bioinformatics, Seoul National  
University, Seoul 151-742, Republic of Korea

H.-K. Lee  
Genomic Informatics Center, Hankyong National University,  
Anseong 456-749, Republic of Korea

**Keywords** Genome sequences · Epidemiology ·  
Phylogenomics · Evolution · Porcine reproductive  
and respiratory syndrome virus

## Introduction

Infection with porcine reproductive and respiratory syndrome virus (PRRSV) causes reproductive failure in sows and respiratory illness in piglets [20], and has led to huge economic losses in the global swine industry [5]. The causative agent of this disease, PRRSV, is an enveloped, positive-sense, and single stranded RNA virus. The viral genome is approximately 15 kb in length, and possesses nine open reading frames (ORFs); the first two ORFs (ORF1a and ORF1b) encode an ORF1ab replicase poly-protein, and the remaining seven ones (ORF2a, 2b, 3, 4, 5, 6, and 7) code for minor structural proteins (GP2, GP3, and GP4) and major structural proteins (GP5, M, and N) [30].

The PRRSVs were first reported in North America in 1987 [17] and in Western Europe in 1990 [37]. They are recognized as two distinct genotypes, type 1 (European, EU) and type 2 (North American, NA), according to the geographic origin of their prototype strains (Lelystad and VR-2332), respectively [37]. Although type 1 and 2 PRRSV isolates share the same viral morphology and give rise to similar disease symptoms, they were genetically and antigenically very different [22].

Although there have been many comparative publications of the genetic characterization, epidemiology, phylogenetics, and evolution of the PRRSV at the molecular level e.g., [3, 21, 23, 36, 39], the evolution of PRRSV is still controversial e.g., [11, 12, 15, 24, 28, 29, 31]. Above all things, to date, evolutionary studies on the PRRSV have focused on only one or a few genes such as ORF5 or ORF7. In order to fully understand the evolutionary patterns and dynamics of PRRSV as well as to aid prevention-and-control policies against the disease, it is necessary to perform a genomic scale analysis. So, we conducted here genome wide study on the temporal and spatial dynamics of the virus. We first determined coding genome sequences for four PRRSVs and then analyzed them together with 122 publicly available genome sequences sampled during a time period of 22 years (1989–2010). Specific aims of the present study are (1) to describe the sequence features of PRRSV coding genome sequences; (2) to reconstruct the genome wide phylogeny of the global PRRSVs using two different analyses methods, Bayesian inferences (BI) and maximum likelihood (ML); (3) to examine the key mechanisms that drive the evolution of this virus, such as selection pressure; and (4) to investigate substitution rates, divergence times, and population size changes using Bayesian coalescent approach.

## Materials and methods

### Samples and data collection

Using MARC-145 cells, four NA type PRRSV strains were isolated from three Korean prefectures (Gyeonggi, Gangwon, and Chungnam) in 2007, 2009, and 2010. The PRRSV isolates, CP07-401-9 (passage 9), CP07-626-2 (passage 13), e417-2 (passage 5) and A4699 (passage 9), were used for the genomic sequencing. Their genomic sequences were analyzed together with published sequences with available genomic sequences including 15 EU type and 107 NA type PRRSVs. At the first stage of our making the data matrix, all known recombinant sequences were excluded. Namely, all sequences were screened for the presence of recombination events via the RDP 3.0b41 package [19] with the default parameters. This package

incorporates six different recombination detection programs: RDP, GENECONV, MaxChi, BOOTSCAN, Chimera, and SiScan. Sequences would be considered as recombinants if they were identified as recombinant by at least 3 out of 6 methods. In all analyses,  $P < 0.05$  was taken to indicate statistical significance. The PRRSVs utilized in the present study are listed in Table S1.

### RNA extraction, RT-PCR, and sequencing

PRRSV isolates were cultured in MARC-145 cells in Dulbecco's modified Eagle's medium (5 % fetal bovine serum) for 5 days and harvested by two-times 'freezing and thawing' and centrifugation (3,000 rpm for 20 min), when the cytopathic effects were observed in approximately 80 % of cells. The harvested viruses were then used for RNA extraction. Total RNA was extracted from the virus stocks using Trizol LS (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. The cDNA was synthesized with random hexamers using SuperScript<sup>®</sup> III First Strand Synthesis kit (Invitrogen), following the manufacturer's instructions. For the sequencing, primer sets from a previous study were selected [40]. The PCR was performed separately with each primer set using Maxime PCR premix kit (Intron, Korea). The PCR amplification conditions were: 94 °C for 2 min; 40 cycles of 94 °C for 10 s, 58 °C for 30 s, and 72 °C for 30 s; and a final elongation step at 72 °C for 7 min. The amplified fragments were purified with a QIAquick Gel Extraction Kit (Qiagen) and sequenced in both the forward and reverse orientations by submitting to Macrogen company (Korea), using the corresponding primer sets. If the primers couldn't amplify the target region in some isolates or the ends of the neighboring contigs were not sufficient to link each other, the additional primer sets were designed based on the obtained contigs.

### Sequence and selection pressure analyses

Both nucleotide and amino acid sequences of the coding regions for 126 PRRSVs were initially aligned using the CLUSTAL X 1.81 [35], and then adjusted manually to eliminate poorly aligned positions. The final alignment comprised 15,211 bps and can be obtained from the authors. The alignment of coding genome sequences was also divided into the corresponding individual genes: ORF1a, 1b, 2, 3, 4, 5, 6, and 7. All gene positions quoted here are with respect to PRRSV genome of strain Lelystad (GenBank accession no. M96262) for EU-type and strain VR-2332 (GenBank accession no. U87392) for NA-type.

For both nucleotide and amino acid sequences of the eight individual genes, as well as the entire coding region, following values were calculated using BIOEDIT 7.053 [14] and Modeltest 3.7 [25]: total sites (including gaps), conserved

sites, sequence homologies, base frequencies, and substitution matrix. In addition, we plotted the number of both nucleotide and amino acid variations at each position throughout the 126 PRRSV genome alignment. We first calculated the nucleotide differences by counting the number of minor nucleotides at each position across the alignment. Here, minor nucleotides were defined as the less frequent ones at each site. For instance, at the specific position of the alignment, if the nucleotide A appeared dominantly in 90 isolates, and C, G and T were observed in the remaining 36 isolates, then the nucleotide difference was 36. The same principle was applied for plotting the amino acid differences.

To perform the selective pressure study on the PRRSVs, the relative rates of non-synonymous and synonymous substitution ( $\omega = dN/dS$ ) across coding portion of the viral genome were also calculated with ClustalX 1.81 [35], PAL2NAL [34], and codeml program in the PAML 3.14.1 package [38]. A  $dN/dS$  ratio of  $<1$  is considered to indicate purifying selection,  $dN/dS = 1$  suggests an absence of selection (i.e. neutral evolution), and  $dN/dS > 1$  indicates positive or diversifying selection.

#### Phylogenomic reconstructions

Phylogenomic approaches involved two different analytical methods, Bayesian inferences (BI) and maximum likelihood (ML). For both analyses, we first determined the best fitting model of sequence evolution for the PRRSV coding genome. GTR+I+G was selected as the best-fit model under Akaike's information criterion (AIC) in Modeltest 3.7 [25]. BI analysis was performed using the program MrBayes 3.1.2 [27] with the following options: nst, 6; rates, gamma; number of generation, 20,000,000; sample frequency, 500; number of chains, 4; burn-in generation, 20,000. To estimate the node reliability, Bayesian posterior probability (BPP) values were shown on the BI tree. ML analysis was implemented in PHYML 3.0 [13], and was run under the following parameters: model of nucleotide substitution, GTR; initial tree, BIONJ; nonparametric bootstrap analysis, yes, 500 replicates; proportion of invariable sites, estimated; number of substitution rate categories, 6; gamma shape parameter, estimated by program; optimize tree topology, yes.

In order to screen for congruent tree topologies with genome tree topologies, each gene segment was analyzed in the same methods as coding genome data. Best fit models used in each gene analyses are presented in Table 2.

Co-estimation of substitutions rates, time of the most recent common ancestor (tMRCA), and population size changes

Sequences with an available year of collection were introduced to Bayesian coalescent approach in order to co-

estimate rates of nucleotide substitutions, times of the most recent common ancestor (tMRCA), and changes in population size. The analysis was conducted under GTR+I+G model, nst = 6, and rates = gamma derived from Akaike's information criterion (AIC) in Modeltest 3.7 [25]. We employed both strict and relaxed (both uncorrelated exponential and uncorrelated log-normal) molecular clocks [8] with five different demographic models (constant size, exponential growth, expansion growth, logistic growth, and Bayesian skyline) using BEAST 1.5.3 [7]. The datasets were each run for 400,000,000 generations to ensure convergence of all parameters (ESSs  $> 100$ ) with discarded burn-in of 10 %. By comparison of Bayes factors ( $\log_{10}$  Bayes Factors  $> 2$  in all cases) based on the relative marginal likelihoods of the models [33], the relaxed uncorrelated exponential clock and exponential population size model were selected as showing the best fit for the PRRSV coding genome dataset. The changes in effective population size over time were examined using the Bayesian skyline plot (BSP). The resulting convergence was analyzed using Tracer 1.5 (<http://beast.bio.ed.ac.uk/Tracer>) and the statistical uncertainties were summarized in the 95 % highest probability density (HPD) intervals. Trees were summarized as maximum clade credibility trees using the TreeAnnotator program, which forms part of the BEAST package, and visualized using FigTree 1.3.1 [26]. In addition, we carried out Bayesian coalescent approach for eight individual genes in order to trace the evolutionary dynamics of each of them. Evolutionary models selected for each gene analysis are presented in Table 2, as well.

## Results

### Sequence analyses and selection pressure analyses

The complete coding genome sequences from four Korean PRRSV strains were determined and were deposited in GenBank (accession nos. JX138233–JX138236). Each of them was classified as NA type virus but was not same to each other in length: 15,059 bps for CP07-401-9; 14,678 bps for e417-2; 14,672 bps for both CP07-626-2 and A4699. The configuration of the 126 PRRSV coding genome and individual gene sequences are summarized in Table 1. The coding genome alignment (including gaps) was 15,211 base pairs, and revealed a very low degree of genetic similarities; 5,358 (35.2 %) of the nucleotide positions were conserved. Pairwise comparisons also showed that the average identities among the complete coding genome sequences were 77.8 % for the nucleotide sequences. Of the eight individual genes, ORF1a was the most variable (average sequence identities of 72.8 and 73.0 % for nucleotides and amino acids, respectively),

**Table 1** Summary of the genomic regions of entire PRRSV

Genomic region	Total sites including gaps, nt/aa	Conserved sites (%), nt/aa	Average identities (%), nt/aa	$\omega$ value (dN/dS)
ORF	15,211/NA	5,358 (35.2)/NA	77.8/NA	NA
ORF1a	7,585/2,529	2,142 (28.2)/595 (23.5)	72.8/73.0	0.24241
ORF1b	4,404/1,476	1,876 (42.6)/162 (11.0)	80.6/77.9	0.13251
ORF2	771/256	352 (45.7)/101 (39.5)	85.0/85.8	0.21361
ORF3	801/266	316 (39.5)/86 (32.3)	83.1/80.5	0.38301
ORF4	552/183	233 (42.2)/75 (41.0)	84.7/87.9	0.23765
ORF5	618/205	224 (36.2)/61 (29.8)	81.8/82.1	0.27130
ORF6	525/174	260 (49.5)/97 (55.7)	88.2/93.7	0.14340
ORF7	398/131	161 (40.5)/54 (41.2)	85.5/86.0	0.19081
EU-type				
ORF	14,763/NA	9,314 (63.1)/NA	89.2/NA	NA
ORF1a	7,191/2,396	4,186 (58.2)/1,566 (65.4)	87.5/89.1	0.19966
ORF1b	4,392/1,463	2,984 (67.9)/1,244 (85.0)	91.0/96.5	0.07685
ORF2	750/249	532 (70.9)/179 (71.9)	91.5/92.4	0.22400
ORF3	798/265	496 (62.2)/153 (57.7)	87.9/85.2	0.31792
ORF4	552/183	337 (61.1)/121 (66.1)	87.8/88.1	0.25309
ORF5	606/201	382 (63.0)/127 (63.2)	88.1/88.7	0.22332
ORF6	522/173	386 (73.9)/133 (76.9)	92.3/93.6	0.20148
ORF7	387/128	290 (74.9)/97 (75.8)	93.2/93.4	0.25684
NA-type				
ORF	15,185/NA	8,094 (53.3)/NA	93.7/NA	NA
ORF1a	7,513/2,514	3,469 (46.2)/1,042 (41.4)	91.2/91.1	0.30970
ORF1b	4,375/1,468	2,646 (60.5)/222 (15.1)	95.0/84.63	0.16993

**Table 1** continued

Genomic region	Total sites including gaps, nt/aa	Conserved sites (%), nt/aa	Average identities (%), nt/aa	$\omega$ value (dN/dS)
ORF2	771/256	495 (64.2)/156 (60.9)	95.7/95.3	0.25514
ORF3	765/254	453 (59.2)/142 (55.9)	94.1/92.8	0.42504
ORF4	537/178	338 (62.9)/108 (60.7)	94.7/94.7	0.33438
ORF5	603/200	350 (58.0)/106 (53.0)	93.8/92.9	0.37286
ORF6	525/174	364 (69.3)/137 (78.7)	97.2/98.2	0.18075
ORF7	372/123	235 (63.2)/78 (63.4)	96.0/96.4	0.22429

NA not applicable

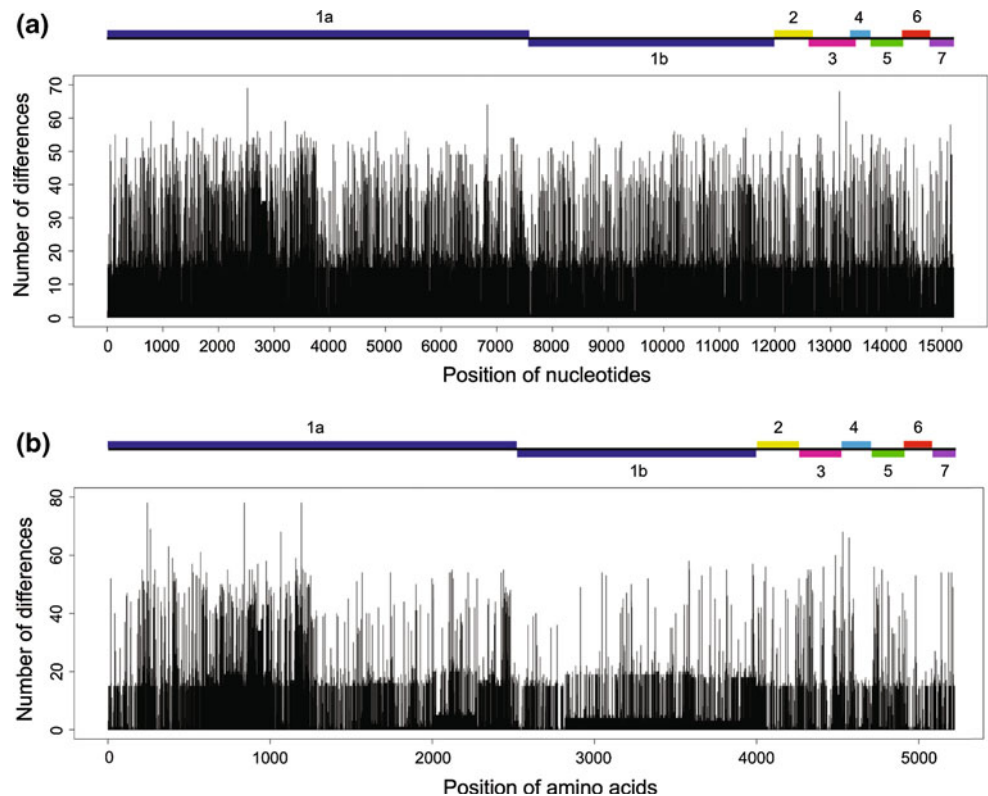
whereas ORF6 was the most conserved (average sequence similarities of 88.2 and 93.7 % for nucleotides and amino acids, respectively). Our plotting the number of both nucleotide and amino acid variations at each site throughout coding genome sequences of 126 global PRRSV isolates depicted that both nucleotide and amino acid alterations were evenly distributed throughout the genomes (Fig. 1). Especially, higher sequence diversities for both nucleotides and amino acids were showed in four regions, the 5' half of ORF1a, the 3' end part of ORF1b, ORF3, and ORF5.

The complete coding genome of the EU-type viruses included 14,763 nucleotides (including insertions), and revealed considerable genetic polymorphism: 9,314 (63.1 %) of the nucleotides were conserved. The average nucleotide sequence identity among this type isolates was 89.2 %. Of the eight genetic regions, ORF1a was the most variable (average sequence similarities, 87.5 % for nucleotides and 89.1 % for amino acid), whereas ORF7 was the most conserved (average sequence identities, 93.2 % for nucleotides and 93.4 % for amino acids).

The total sites (including gaps) of the NA-type coding genome had 15,185 base pairs. In comparison with EU-type sequences, they had similar configurations: 8,094 (53.3 %) of the nucleotide sites were conserved. Pairwise comparisons showed that the average identity between NA-type isolates was 93.7 % for the nucleotide sequences. ORF6, rather than ORF7 in the EU type, was most conserved (average sequence identities, 97.2 % for nucleotides and 98.2 % for amino acids).

The mean nonsynonymous/synonymous substitution ratio ( $\omega = dN/dS$ ) values for each component gene of the global, EU-, and NA-type PRRS viruses were lower than 1 (Table 1), which suggests that PRRSV genome was under

**Fig. 1** Plotting the nucleotide (a) and amino acid (b) differences throughout the complete coding genomes of 126 global PRRSVs. Number of differences was estimated as the total number of altered nucleotides at each site compared by the multiple sequence alignment method. Each ORF is indicated on top of each plot



purifying selection. The highest dN/dS ratios were observed in ORF3 (0.38301 for global, 0.31792 for EU-type, and 0.42504 for NA-type), whilst the lowest ones were showed in ORF1b (0.13251 for global, 0.07685 for EU-type, and 0.16993 for NA-type).

#### Phylogenomic analyses

Bayesian inferences (BI) and maximum likelihood (ML) methods produced consistent features regarding the phylogenomics of PRRSV, and supported the topology of the maximum clade credibility tree (Fig. 2). All of the global PRRSVs are divided into two different clades, EU and NA genotypes (all, posterior probabilities = 1.00 and bootstrap values = 100 %). Here, all EU-type viruses were either divided into one of two groups or were unclassified (5 isolates: EU076704, FJ349261, GU047344, JF276430 and JF276431). Group 1 contained six PRRSV isolates from the United States, the Netherlands, Thailand, and Germany during 1991–2001. Group 2 was grouped with four viruses from Spain, China, and Portugal between 2005 and 2009.

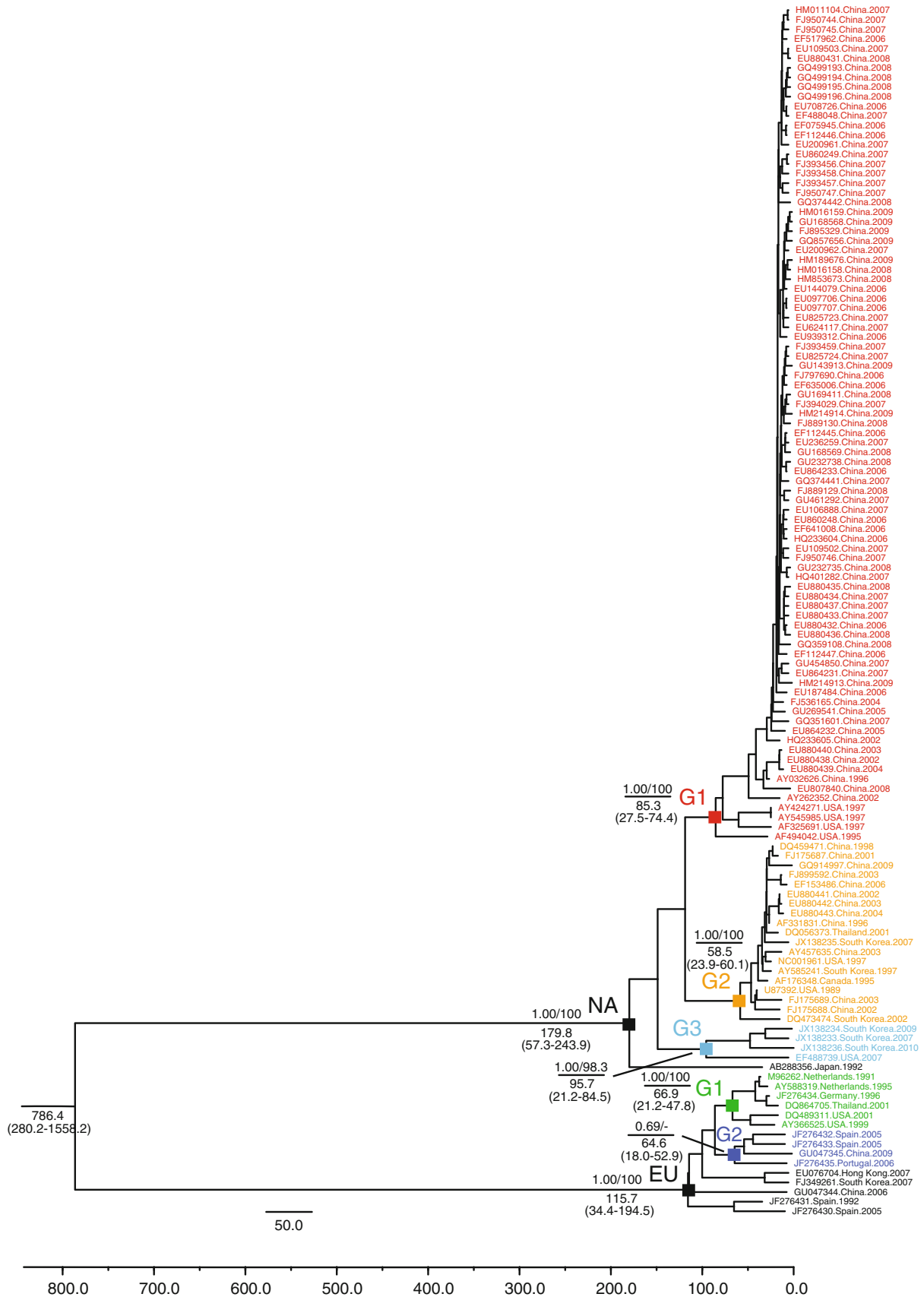
Within the NA-type clade, all members either belonged to one of three major groups or were unclassified (one isolate: AB288356). Group 1 consisted of 87 PRRSVs (78.4 % of all NA-type viruses), of which 83 were Chinese isolates and the remaining four were American viruses collected between 1995 and 2009. Group 2 was composed

of 19 isolates from Korea, China, Thailand, the United States, and Canada from 1989 to 2009. Viruses of Group 3 were collected from China and America during 2007–2010.

In addition to phylogenomic analyses, in order to elucidate the most significant phylogenetic marker among the eight PRRSV genes, we employed both BI and ML analyses at viral gene level under the best fit evolutionary model (Table 2). All individual trees were compared to the coding genome-based phylogenomic trees. Here, ORF1a gene trees (Fig. 3) most closely resembled the coding genome trees. The overall tree topologies of both ORF5 and ORF7 genes, which have been used as the molecular markers of the PRRSV investigation, were very different from those of genomic trees (Figs. S1, S2); all clades excluding NA-type Group 1 and NA-type Group 2 were collapsed in ORF5 trees, and all groupings were broken down in ORF7 trees as well.

#### Substitutions rates, divergence times, and population size changes

Under the relaxed uncorrelated exponential clock and exponential population size model showing the best fit for the PRRSV coding genome dataset, the evolutionary rate estimated was  $1.98 \times 10^{-3}$  (95 % HPD  $9.81 \times 10^{-3}$ – $3.04 \times 10^{-3}$ ) substitutions/site/year, and the tMRCA calculated was 786.4 (95 % HPD 280.2–1558.2) years ago.



◀ **Fig. 2** Bayesian maximum clade credibility phylogenomic tree derived from complete coding genome sequences of 126 global PRRSVs. The data set (15,211 base pairs) was also phylogenetically analyzed with BI and ML methods and identical topology was produced. The robustness of the phylogenetic analysis is presented above the nodes: left numbers represent Bayesian posterior probabilities ( $\geq 0.80$ ) and right ones represent ML bootstrap values ( $\geq 70\%$ ). Divergence times (in years) are positioned below the nodes; the 95% HPD intervals are indicated in brackets. Groups are marked by a “G”

Here, EU-type viruses diverged 115.7 (95% HPD 34.4–194.5) years ago and NA-type ones segregated 179.8 (95% HPD 57.3–243.9) years ago. The Bayesian skyline plot analysis of PRRSV isolates (Fig. 4) depicted five phases of epidemic growth. The viruses evolved under almost constant population size until the late 1400s, when they experienced a population decrease until the late 1600s. The population size then increased again until the early 1900s, when a rapid decline followed by a sharp expansion approaching the present.

In order to track the evolutionary marker of the eight gene segments, we additionally carried out Bayesian coalescent approaches under the best fit model (Table 2) for each gene. As shown Table 3, tMRCA estimated from ORF1a sequences was most similar to that from the coding genome: 709.7 (299.1–1259.9) years.

## Discussion

Our results based on coding genome sequences indicated that PRRSV heterogeneity is very high and this point is in agreement with viewpoints of previous investigators e.g., [1, 32]. Average nucleotide identity of inter-genotypic viruses was 77.8%; of the eight individual genes, ORF1a was most variable, while ORF6 was most conserve. Of the six structural genes, ORF5 which is the main target for virus neutralizing antibodies and phylogenetic evolution was most polymorphic. In addition, within each genotype,

similar configurations were presented as well. Those of intra-genotypic isolates were 89.2% for EU-type and 93.7% for NA-type. Interestingly, in spite of the considerable genetic polymorphisms between and within two genotypes, the viral morphology is same and their disease symptoms are very similar. Because the sequence variability may affect PCR diagnosis and vaccination efficiency as well as studies on immunology, epidemiology, and phylogenetic evolution, it is essential to continuously monitor their variation.

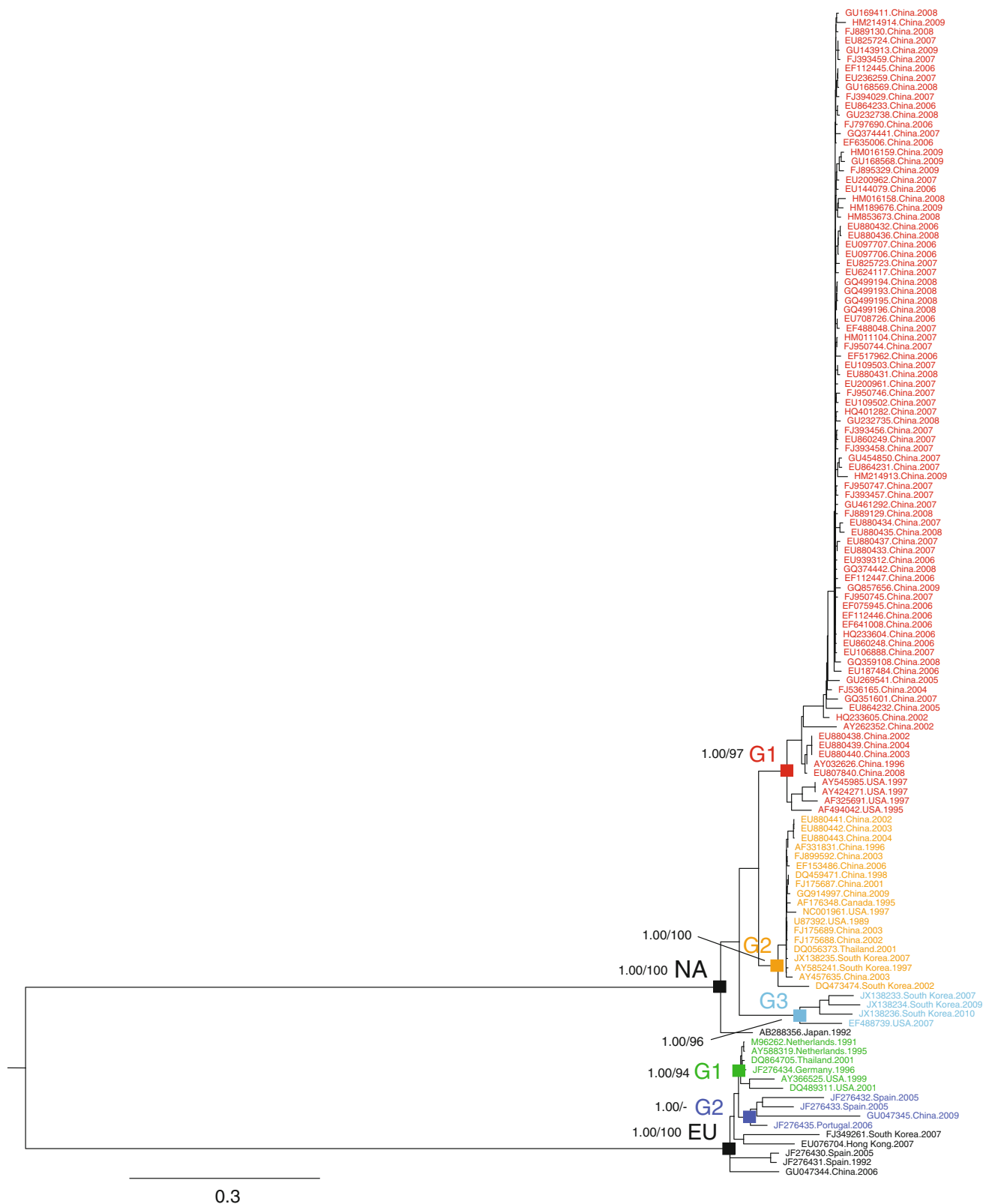
In regard to the selection pressure on PRRSVs, our findings based on complete coding genome sequences exhibited that purifying selection is acting on the viruses; the mean nonsynonymous/synonymous substitution ratio ( $\omega = dN/dS$ ) values for each genomic segment were low in all cases (all,  $\omega < 1$ ). Of which, the ORF3 showed the highest dN/dS values, while the ORF1b showed the lowest ones. The purifying selection on the PRRSV genome is in line with the suggestions of Pesch et al. [23], while it contradicts the viewpoints of others such as Hanada et al. [15] and Song et al. [31] who mentioned that the PRRSV evolution was affected by positive selection.

To date, molecular phylogenetics has been contributed to the elucidation of the genomics, emergence, epidemiology, and evolutionary mechanisms of the PRRSV, due mainly to the many insights these analyses can provide. Although PRRSV data have increased rapidly, most phylogenetic studies describing the virus have been restricted to ORF5 or ORF7. Genome-level works of global PRRSVs have been performed using the only limited number of samples, as well. For example, to date, the most extensive PRRSV phylogenomics were inferred from only 25 (15 EU and 10 NA) and 30 (6 EU and 24 NA) isolates by Darwich et al. [4] and Zhu et al. [40], respectively. Therefore, our findings based on 126 coding genome sequences have advantages relative to results based on single gene or limited number of genome sequences and may be more

**Table 2** Models of evolution estimated for the genomic regions of entire PRRSV with Modeltest

Genomic region	Model	−lnL	Base frequencies (A, C, G)	Substitution matrix (A–C, A–G, A–T, C–G, C–T)	Pinvar	Nst	Shape
ORF	GTR+I+G	166254.56	0.21, 0.28, 0.25	1.13, 12.60, 1.51, 0.98, 13.85	0.31	6	1.66
ORF1a	TVM+I+G	89719.80	0.20, 0.28, 0.26	1.15, 13.20, 1.57, 0.93, 13.20	0.25	6	1.95
ORF1b	GTR+I+G	41793.23	0.23, 0.28, 0.25	0.83, 13.80, 1.50, 0.88, 18.10	0.39	6	1.63
ORF2	K81uf+I+G	7072.27	0.21, 0.27, 0.23,	1.00, 8.44, 0.68, 0.68, 8.44	8.43	6	1.30
ORF3	GTR+I+G	8455.15	0.20, 0.30, 0.23	1.54, 12.08, 1.48, 0.79, 9.30	0.34	6	1.46
ORF4	TVM+I+G	5418.04	0.21, 0.29, 0.25	1.43, 10.64, 1.21, 0.67, 10.64	0.34	6	1.23
ORF5	HKY+I+G	6577.39	0.22, 0.27, 0.24	TRatio 4.59	0.25	2	1.04
ORF6	SYM+G	4020.05	Equal	2.97, 18.88, 3.75, 2.80, 34.28	0	6	0.34
ORF7	TI-Mef+I+G	3288.97	Equal	1.00, 11.57, 1.62, 1.62, 17.56	0.34	6	1.39

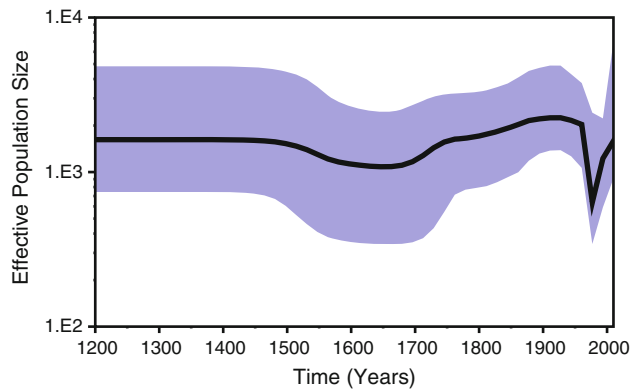
−lnL log likelihood scores, Pinvar proportion of invariable sites, Shape gamma distribution shape parameter



**Fig. 3** Bayesian Inference tree ( $-\ln L = 89719.80$ ) based on complete ORF1a nucleotide sequences of 126 global PRRSVs. Evolutionary best fit model is TVM+I+G. The robustness of the

phylogenetic analysis is shown above the nodes: left numbers are Bayesian posterior probabilities ( $\geq 0.80$ ) and right ones are ML bootstrap values ( $\geq 70\%$ ). Groups are marked by a “G”





**Fig. 4** Bayesian skyline plot on the basis of coding genome sequences from 126 global PRRSVs sampled between 1989 and 2010. The dark line in Bayesian skyline plot shows the estimated effective population size through time. The gray area represents the 95 % highest posterior density confidence intervals for this estimate

robust. Our results depicted two distinct phylogenetic clades, EU and NA genotypes. Although EU-type and NA-type PRRSVs were originated from Europe and North America, respectively, the mixture of two genotype isolates in a nation has been reported worldwide. Owing to the coexistence of two genotypes, the prevention and control strategies of this acute disease is much more complex and difficult.

Next, our phylogenomic study focused on geographic and/or period influences on the evolution of PRRSVs. On this topic, recent workers such as Cha et al. [2] and Zhu et al. [40] claimed that geographic separation is a factor influencing the PRRSV evolution, based on ORF5 and genome sequences, respectively. In contrast to their opinions, however, there were different suggestions by our [39] and other [16] previous studies derived from ORF7 sequences. The ORF7 data revealed that there was no immediate relationship between the date or place of collection and the topological distribution of PRRSVs, which is confirmed by the present study on the basis of complete

coding genome sequences. Within the EU-type clade, Group 1 members were collected from four countries (United States, the Netherlands, Thailand, and Germany) during 1995–2009, and Group 2 viruses were sampled from three countries (Spain, China, and Portugal) between 2005 and 2009. This configuration appeared similarly in the NA-type clade as well. Group 1 isolates occurred in China and America between 1995 and 2009. Group 2 members broke out in five countries (Korea, China, Thailand, the United States, and Canada) from 1989 to 2009. Group 3 PRRSVs were derived from China and America during 2007–2010. These features may be largely due to the rapid expansion and diversification of PRRSVs over a relatively short period of time, and their rapid spread via the frequent international trade in livestock. This intermix can make vaccine strategies more difficult. Thus, continuously screening for changes in the mixed population structure of this virus is needed.

We also tried to determine which region(s) has stronger phylogenetic signal by comparing eight individual gene trees with the coding genome trees. As a result, topologies of ORF1a trees, rather than ORF5 or ORF7 trees, much more closely resembled those of the coding genome trees. Accordingly, we propose that ORF1a rather than ORF5 or ORF7 is more useful marker to reconstruct the PRRSV phylogeny.

On the evolutionary rates of the RNA viruses including the PRRSV, there have been several publications. Domingo et al. [6] stated that RNA viruses have high mutation rates of  $10^{-3}$ – $10^{-5}$  nucleotide substitutions per site per replication cycle. Similarly, Duffy et al. [9] also reported that RNA viruses evolve at an approximate rate of  $10^{-3}$  nucleotide substitutions/site/year. Especially, regarding the evolutionary rates of PRRSV, Forsberg et al. [12] estimated mean evolutionary rates of  $5.8 \times 10^{-3}$  (95 % HPD =  $4.8$ – $6.9 \times 10^{-3}$ ) substitutions/site/year on the basis of the ORF3 sequences of EU isolates. Hanada et al. [15] suggested the evolutionary rate of  $4.17$ – $9.8 \times 10^{-2}$  substitutions/site/year in their investigations

**Table 3** Bayesian estimates of divergence times and evolutionary parameters for PRRSV genomic region

Genomic region	Demographic model	Substitution rates (substitutions/site/year)	Age (years)
ORF	Exponential	$1.98 \times 10^{-3}$ ( $9.81 \times 10^{-3}$ – $3.04 \times 10^{-3}$ )	786.4 (280.2–1558.2)
ORF1a	Logistic	$2.19 \times 10^{-3}$ ( $1.13 \times 10^{-3}$ – $3.30 \times 10^{-3}$ )	709.7 (299.1–1259.9)
ORF1b	Exponential	$2.21 \times 10^{-3}$ ( $1.28 \times 10^{-3}$ – $3.19 \times 10^{-3}$ )	869.1 (413.1–1424.0)
ORF2	Constant	$2.55 \times 10^{-3}$ ( $1.38 \times 10^{-3}$ – $3.83 \times 10^{-3}$ )	414.1 (168.1–718.0)
ORF3	Logistic	$1.77 \times 10^{-3}$ ( $1.03 \times 10^{-3}$ – $2.56 \times 10^{-3}$ )	408.7 (169.3–708.6)
ORF4	Logistic	$1.39 \times 10^{-3}$ ( $7.32 \times 10^{-4}$ – $2.12 \times 10^{-3}$ )	540.7 (209.6–974.5)
ORF5	Logistic	$2.02 \times 10^{-3}$ ( $1.17 \times 10^{-3}$ – $2.94 \times 10^{-3}$ )	395.5 (162.8–682.2)
ORF6	Constant	$2.25 \times 10^{-3}$ ( $6.55 \times 10^{-4}$ – $1.86 \times 10^{-3}$ )	447.7 (176.0–809.5)
ORF7	Exponential	$2.82 \times 10^{-3}$ ( $1.23 \times 10^{-3}$ – $4.66 \times 10^{-3}$ )	251.2 (60.2–515.5)

For all parameter estimates the 95 % HPD values are given in parenthesis

based on ORFs 3–5 sequences of two genotype PRRSVs. Most recently, Shi et al. [29] and Song et al. [31] used ORF5 sequence data in their studies, and calculated the nucleotide substitution rates of  $1.46 \times 10^{-3}$  for NA-type viruses and  $3.29 \times 10^{-3}$  substitutions/site/year for two genotype isolates, respectively. Consistently with these previous investigations, our findings also indicate that PRRSV has very high evolutionary rate; the average substitution rate was  $1.98 \times 10^{-3}$  (95 % HPD  $9.81 \times 10^{-3}$ – $3.04 \times 10^{-3}$ ) substitutions/site/year. These fast evolutionary rates of RNA viruses including PRRSV are due to several factors such as their small genomes, short generation times, rapid mutation, and lack of polymerase proof-reading [10]. As a result, it is possible that RNA virus raises viral population adaptation, survival, and fitness, allowing them to rapidly spread to new hosts and novel environments [18].

Despite huge advances both in evolutionary theory and in sequencing technology, the divergence times of PRRSVs is still controversial. All of the PRRSV studies on this topic have performed using only individual gene sequences. Forsberg et al. [12] analyzed ORF3 sequences of EU isolates and dated the most recent common ancestor to 1979, more than 10 years before European outbreak. Plagemann [24] treated three genes (ORF1b, ORF5, and ORF7) and the tMRCA postulated by him was approximately 100 years ago. He thought that a mutant of LDV (lactate dehydrogenase-elevating virus) infected wild boars in central Europe in 1912 and that this intermediate host spread the virus to North Carolina imports; the virus then evolved independently on the European and American continents in the wild boar populations for about 70 years until independently entering the domestic swine population. Subsequently, two debates [11, 15] on the time scale of PRRSV have occurred; Forsberg's [11] analyses based on ORF3 sequences of two genotypes suggested that PRRSV diverged around the year 1880, whereas Hanada et al. [15] results on the basis of ORFs 3–5 data of two genotypes presented that the virus evolved in about 1980. Most recently, there were two additional works on this topic. Shi et al. [29] noted that the most recent common ancestor of NA-type virus appeared approximately in 1979 and Song et al. [31] demonstrated that the most recent common ancestor of two genotype PRRSVs diverged in 1894. In this study, in order to further infer the PRRSV time scale, we conducted a Bayesian coalescent approach using the most extended coding genome sequences from 126 PRRSVs collected during the past 22 years (1989–2010). As a result, in contrast to the previous works, our molecular dating suggests that PRRSV group has a much older origin in spite of its recent emergence. The tMRCA of PRRSV was 786.4 (95 % HPD 280.2–1558.2) years; they were originated approximately in the year 1,225.6 (95 % HPD 1731.8–453.8). Then, EU-type viruses diverged 115.7 years ago (95 % HPD 34.4–194.5), while the NA-type isolates

were segregated 179.8 years ago (95 % HPD 57.3–243.9); this configuration is in concordance with the conclusion of Nelsen et al. [23] who considered that the two lineages must have evolved separately from a very distant common ancestor prior to their emergence into the pig populations on the two continents.

In terms of tracking the evolutionary dynamics of each of the eight individual genes, our estimation showed that tMRCA of ORF1a is most close to that of the coding genome; tMRCA of ORF1a was 709.7 (299.1–1259.9) years ago and the mean evolutionary rate was  $2.19 \times 10^{-3}$  ( $1.13 \times 10^{-3}$ – $3.30 \times 10^{-3}$ ) substitutions/site/year. So, we suggest that ORF1a, of the eight individual regions, is the most appropriate marker for tracing the evolutionary history of PRRSV at the gene level.

Regarding the effective population size changes of PRRSVs, our Bayesian skyline plot (BSP) analysis (Fig. 4) revealed that the viruses appear to have evolved under almost constant population size until the late 1400s, when they experienced a population decrease until the late 1600s. Their population size then increased again until the early 1900s, when a rapid, sharp decline followed by a sharp expansion approaching the present.

To date, despite intensive efforts into PRRSV immunological works worldwide, porcine reproductive and respiratory syndrome is still the most notorious disease in the global swine industry. Accordingly, both national and global policies are necessary to prevent and control this acute disease. The expanding data and more extensive information of evolutionary dynamics resulting from the present study might be very useful for the prevention and control of this virus as well as for improving our knowledge about its infection tracking and evolution. This is the first genome wide study on the temporal and spatial dynamics of PRRSV.

**Acknowledgments** This work was supported by a grant from the Next-Generation BioGreen 21 Program (Nos. PJ008196 and PJ009019), Rural Development Administration, Republic of Korea.

## References

1. An TQ, Zhou YJ, Liu GQ, Tian ZJ, Li J, Qiu HJ, Tong GZ (2007) Genetic diversity and phylogenetic analysis of glycoprotein 5 of PRRSV isolates in mainland China from 1996 to 2006: coexistence of two NA-subgenotypes with great diversity. *Vet Microbiol* 123:43–52
2. Cha SH, Choi EJ, Park JH, Yoon SR, Song JY, Kwon JH, Song HJ, Yoon KJ (2006) Molecular characterization of recent Korean porcine reproductive and respiratory syndrome (PRRS) viruses and comparison to other Asian PRRS viruses. *Vet Microbiol* 117:248–257
3. Chen N, Cao Z, Yu X, Deng X, Zhao T, Wang L, Liu Q, Li X, Tian K (2011) Emergence of novel European genotype porcine reproductive and respiratory syndrome virus in mainland China. *J Gen Virol* 92:880–892

4. Darwich L, Gimeno M, Sibila M, Diaz I, de la Torre E, Dotti S, Kuzemtseva L, Martin M, Pujols J, Mateu E (2011) Genetic and immunological diversities of porcine reproductive and respiratory syndrome genotype I strains. *Vet Microbiol* 150:49–62
5. Dea S, Gagnon CA, Mardassi H, Pirzadeh B, Rogan D (2000) Current knowledge on the structural proteins of porcine reproductive and respiratory syndrome (PRRS) virus: comparison of the North American and European isolates. *Arch Virol* 145: 659–688
6. Domingo E, Baranowski E, Escarmis C, Sobrino F (2002) Foot and mouth disease virus. *Comp Immunol Microbiol Infect Dis* 25:297–308
7. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
8. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88. doi:10.1371/journal.pbio.0040088
9. Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9:267–276
10. Elena SF, Sanjuan R (2005) Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol* 79:11555–11558
11. Forsberg R (2005) Divergence time of porcine reproductive and respiratory syndrome virus subtypes. *Mol Biol Evol* 22: 2131–2134
12. Forsberg R, Oleksiewicz MB, Petersen AM, Hein J, Bøtner A, Storgaard T (2001) A molecular clock dates the common ancestor of European-type porcine reproductive and respiratory syndrome virus at more than 10 years before the emergence of disease. *Virology* 289:174–179
13. Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
14. Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
15. Hanada K, Suzuki Y, Nakane T, Hirose O, Gojobori T (2005) The origin and evolution of porcine reproductive and respiratory syndrome viruses. *Mol Biol Evol* 22:1024–1031
16. Hao X, Lu Z, Kuang W, Sun P, Fu Y, Wu L, Zhao Q, Bao H, Fu Y, Cao Y, Li P, Bai X, Li D, Liu Z (2011) Polymorphic genetic characterization of the ORF7 gene of porcine reproductive and respiratory syndrome virus (PRRSV) in China. *Virol J* 8:73
17. Keffaber K (1989) Reproductive failure of unknown etiology. *Am Assoc Swine Pract Newsl* 1:1–10
18. Lauring AS, Andino R (2010) Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 6:e1001005
19. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21:260–262
20. Meulenber JM (1998) Porcine reproductive and respiratory syndrome virus (PRRSV) molecular characterization of the agent. In: *Proceedings of 15th IPVS congress, Birmingham*, pp 149–157
21. Nam E, Park CK, Kim SH, Joo YS, Yeo SG, Lee C (2009) Complete genomic characterization of a European type I porcine reproductive and respiratory syndrome virus isolate in Korea. *Arch Virol* 154:629–638
22. Nelsen CJ, Murtaugh MP, Faaberg KS (1999) Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents. *J Virol* 73:270–280
23. Pesch S, Meyer C, Ohlinger VF (2005) New insights into the genetic diversity of European porcine reproductive and respiratory syndrome virus (PRRSV). *Vet Microbiol* 107:31–48
24. Plagemann PG (2003) Porcine reproductive and respiratory syndrome virus: origin hypothesis. *Emerg Infect Dis* 9:903–908
25. Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818
26. Rambaut A (2006) <http://evolve.zoo.ox.ac.uk/beast>
27. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574
28. Shi M, Lam TT, Hon CC, Hui RK, Faaberg KS, Wennblom T, Murtaugh MP, Stadejek T, Leung FC (2010) Molecular epidemiology of PRRSV: a phylogenetic perspective. *Virus Res* 154:7–17
29. Shi M, Lam TT, Hon C-C, Murtaugh MP, Davies PR, Hui RK, Li J, Wong LT, Yip CW, Jiang JW, Leung FC-C (2010) A phylogeny-based evolutionary, demographical and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. *J Virol* 84:8700–8711
30. Snijder EJ, Meulenber JJ (1998) The molecular biology of arteriviruses. *J Gen Virol* 79:961–979
31. Song J, Shen D, Cui J, Zhao B (2010) Accelerated evolution of PRRSV during recent outbreaks in China. *Virus Genes* 41:241–245
32. Stadejek T, Oleksiewicz MB, Potapchuk D, Podgorska K (2006) Porcine reproductive and respiratory syndrome virus strains of exceptional diversity in eastern Europe support the definition of new genetic subtypes. *J Gen Virol* 87:1835–1841
33. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18:1001–1013
34. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612
35. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal-windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 22:4673–4680
36. Wang C, Wu B, Amer S, Luo J, Zhang H, Guo Y, Dong G, Zhao B, He H (2010) Phylogenetic analysis and molecular characteristics of seven variant Chinese field isolates of PRRSV. *BMC Microbiol* 10:146
37. Wensvoort G, Terpstra C, Pol JM, ter Laak EA, Bloemraad M, de Kluyver EP, Kragten C, van Buiten L, den Besten A, Wagenaar F et al (1991) Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Vet Q* 13:121–130
38. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556
39. Yoon SH, Song JY, Lee CH, Choi EJ, Cho IS, Kim B (2008) Genetic characterization of the Korean porcine reproductive and respiratory syndrome viruses based on the nucleocapsid protein gene (ORF7) sequences. *Arch Virol* 153:627–635
40. Zhu L, Zhang G, Ma J, He X, Xie Q, Bee Y, Gong SZ (2011) Complete genomic characterization of a Chinese isolate of porcine reproductive and respiratory syndrome virus. *Vet Microbiol* 147:274–282