AASV Foundation Research Report – Final Report

Title: Refining PRRSV-2 ORF5-based genetic classification system to better characterize genetic diversity and relatedness of PRRSV

Investigators: Jianqiang Zhang, Phillip Gauger, Karen Harmon, Rodger Main, Wannarat Yim-im. Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University.

Report

1. Statement of the problem

Porcine reproductive and respiratory syndrome (PRRS) is one of the most devastating swine diseases causing tremendous economic losses to the swine industry worldwide. Porcine reproductive and respiratory syndrome virus (PRRSV), the causative agent of PRRS, is an enveloped, single-stranded, and positive-sense RNA virus in the family *Arterividae* (1). According to the new virus taxonomy, PRRSV includes two species: *Betaarterivirus suid 1* (with virus name PRRSV-1, previously known as the European genotype) and *Betaarterivirus suid 2* (with virus name PRRSV-2, previously known as the North American genotype). The PRRSV genome is ~15kb in length and is composed of 11 open reading frames (ORF), including ORF1a, ORF1b, ORF2a, ORF2b, ORF3, ORF4, ORF5a, ORF5, ORF6, ORF7, and a short transframe (TF) ORF in the nsp2 region. These ORFs encode 16 non-structural proteins (nsp1α, nsp1β, nsp2–nsp6, nsp7α, nsp7β, nsp8–nsp12, nsp2TF, and nsp2N) and eight structural proteins (GP2, E, GP3–GP5, GP5a, M, and N) (2, 3).

Due to its high genetic variation and harboring neutralization epitopes, ORF5 encoding major envelope glycoprotein 5 (GP5) has been widely used to study genetic diversity of PRRSV (4-7). Restriction fragment length polymorphism (RFLP) typing based on three restriction enzymes *Mlu*I, *Hinc*II, and *Sac*II on PRRSV-2 ORF5 was first introduced in 1998 to distinguish Ingelvac PRRS MLV vaccine (RFLP 2-5-2) and wild-type strains (8). Since then, PRRSV-2 genetic diversity and emerging PRRSV-2 strains in North America have been widely reported according to RFLP patterns (9, 10). Recently, 228 distinct RFLP patterns from over 40,000 sequences were reported in the USA during 2007–2019; temporal and geographic distributions of different RFLP patterns reflected extensive PRRSV-2 genetic diversity (11). However, RFLP typing cannot capture PRRSV-2 genetic changes outside the cutting sites of the three restriction enzymes and single nucleotide change at the cutting sites can change the RFLP pattern; hence, RFLP typing has limited utility as a tool to define genetic relatedness of PRRSV-2. Also, RFLP typing has mainly been used in North America and has not been widely adopted in other parts of the world.

To overcome limitations of RFLP typing, phylogenetic classification based on 8,624 global ORF5 sequences was proposed in 2010 to describe PRRSV-2 genetic diversity; PRRSV-2 was divided into nine lineages (L1–L9) and 37 sublineages (sublineages 1.1–1.9 in lineage 1, sublineages 5.1–5.2 in lineage 5, sublineages 8.1–8.9 in lineage 8, and sublineages 9.1–9.17 in lineage 9) (6). The 841 reference sequences representing 9 lineages have been widely used for epidemiological studies to investigate PRRSV-2 genetic diversity in many different countries (4, 12-27). However, the proposed 37-sublineage system has not been globally used because the

reference sequences representing 37 sublineages are not made available to the public. Due to the tremendous genetic expansion of PRRSV-2 in lineage 1 in the past decade, Paploski et al divided lineage 1 into nine sublineages (L1A-L1C, L1Dalpha, L1Dbeta, L1E-L1H) based on analysis of >20,000 PRRSV-2 ORF5 sequences collected in the USA (7, 28). However, for over ten years, PRRSV-2 sequences at global levels have not been thoroughly analyzed to determine if new lineages have emerged; nor have the sublineages within lineages other than the lineage 1 been evaluated and updated.

This study aimed to fulfill multiple objectives as described below.

2. Objective(s)

1) Refine the PRRSV-2 genetic classification system based on analysis of a large dataset of global ORF5 sequences and establish reference sequences for epidemiological and diagnostic applications.
2) Characterize the relationships of PRRSV-2 RFLP typing and ORF5-based phylogenetic lineages/sublineages.
3) Analyze the geographic distributions of global PRRSV-2 ORF5-based lineages and sublineages
4) Characterize geographic and temporal dynamic changes of PRRSV-2 in the USA, including both wild-type and vaccine-like virus strains.

3. Brief materials and methods

Datasets. A dataset of 82,237 global PRRSV-2 ORF5 sequences from two sources was used in this study: 57,260 sequences from the Iowa State University Veterinary Diagnostic Laboratory (ISU VDL) and 24,977 PRRSV-2 ORF5 sequences from the United States Swine Pathogen Database (US-SPD, https://swinepathogendb.org) created by the Agricultural Research Service National Animal Disease Center, United States Department of Agriculture (29). The US-SPD database collects all sequences deposited in GenBank, representing sequences from global countries (29). For the sequences from the US-SPD database, virus strain or isolate name, GenBank accession number, sample collection date, sample collection location, and RFLP pattern information was compiled. For the ISU VDL PRRSV-2 ORF5 sequences, the information including case submission ID, sample collection date, site location, and RFLP pattern was compiled. The information of 82,237 PRRSV-2 ORF5 sequences including collection region/country and collection year is summarized in Table 1. The 57,260 sequences from ISU VDL accumulated from 2003 to 2021 included samples collected in the USA (n=55,524), Canada (n=2), Mexico (n=1,733), and Venezuela (n=1). The 24,977 sequences from the US-SPD database were collected during 1989-2021 from seven global regions, including East Asia, Southeast Asia, South Asia, Europe, North America (n=18,943; Canada, the United States, and Mexico), Central America, South America, and unidentified countries.

Redundant ORF5 sequences with 100% nucleotide (nt) similarity, ORF5 sequences with more than five ambiguous nucleotides, and incomplete ORF5 sequences were removed from ORF5 sequences retrieved from ISU VDL and US-SPD databases using mothur v.1.44.3 (30). Previous studies analyzing 355 and 949 PRRSV-2 whole genome sequences (19, 27) indicate that the recombination hot spots were mainly located in nsp9 and the ORF2–ORF4 genomic regions. Recombination within ORF5 rarely occurred. Even if recombination within ORF5 occurs under rare circumstances, it may create a phylogenetic clade that is located between the two lineages to which the two parental PRRSV-2 sequences belong. That will still be interesting for investigation.

Therefore, in the current study, ORF5 sequences with the potential within-ORF5 recombination were not excluded for analysis. Eventually, 40,601 PRRSV-2 ORF5 sequences from ISU VDL, 16,851 PRRSV-2 ORF5 sequences from US-SPD database, and 841 ORF5 reference sequences from Shi et al 2010 (6) were included to analyze and refine the PRRSV-2 ORF5-based genetic classification system as described in the section "Phylogenetic analyses of PRRSV-2 ORF5 sequences" below.

After the new PRRSV-2 lineage/sublineage classification system was defined, 1,100 reference sequences were randomly selected from different clusters of phylogenetic trees to represent all lineages and sublineages. Subsequently, the established reference sequences were used to analyze all of the 82,237 sequences to determine their lineage/sublineage information.

Phylogenetic analysis of PRRSV-2 ORF5 sequences. Multiple sequence alignment was performed by the progressive method (FFT-NS-1) in MAFFT v7.407 (31). The phylogenetic tree from the multiple sequence alignment result was constructed using maximum likelihood with stochastic algorithm, general time-reversible nucleotide substitution with 10 categories of FreeRate heterogeneity model (GTR+F+R10) and 1,000 bootstrap replicates in IQ-TREE v1.6.12 (32). Simple manipulation of tree (smot) v0.14.2 (https://github.com/flu-crew/smot) was performed to reduce the dataset based on genetic relatedness from the large phylogenetic tree. Multiple sequence alignment of the remaining ORF5 sequences after smot analysis was performed by the progressive method (FFT-NS-2) in MAFFT v7.407 (31) and four independent phylogenetic trees were built using maximum likelihood with stochastic algorithm, general time-reversible nucleotide substitution with 10 categories of FreeRate heterogeneity model (GTR+F+R10) and 1,000 bootstrap replicates in IQ-TREE v1.6.12 (32). Tree topology tests were performed to statistically compare four independent phylogenetic trees using RELL approximation with 10,000 replicates (33), weighted Kishino-Hasegawa (KH) test (34), weighted Shimodaira-Hasegawa (SH) test (35), and approximately unbiased (AU) (36). Then, the best tree topology was identified from amongst the four trees.

Genetic lineages and sublineages of PRRSV-2 ORF5 sequences were classified according to the previously described procedure (6). Phylogenetic trees were visualized in Archaeopteryx.js v0.9928 beta-2018-07-05 (37) and lineages and sublineages were annotated in ggtree v3.1.4.991 (38).

Nucleotide identities were calculated by the MegAlign 15 program in the DNASTAR Lasergene 15 software. Average pairwise raw genetic distance within and between lineages and sublineages were calculated by MEGA X (39) to investigate genetic diversity. For calculating the pairwise distances at the lineage level (L1–L11), 1,100 reference sequences and 9,419 additional sequences randomly selected from the phylogenetic clusters representing the 11 lineages were used. For calculating the pairwise distances among the sublineages L1A–L1F and L1H–L1J, 466 reference sequences representing the nine sublineages within L1 and 10,170 additional sequences randomly selected from the phylogenetic clusters representing nine L1 sublineages were used. For calculating the pairwise distances in L5, all of the 16,845 sequences within L5 including 67 reference sequences representing L5A and L5B were used. For calculating the pairwise distances in L8, all of the 10,611 sequences within L8 including 219 reference sequences representing L8A–L8E were used. For calculating the pairwise distances in L9, all of the 6,052 sequences within L9 including 133 reference sequences representing L9A–L9E were used.

Genetic diversity analyses and visualization regarding RFLPs and genetic lineages and sublineages. Discriminant analysis of principal component (DAPC) is a combination method between discriminant analysis and principal component analysis. The DAPC method (40) in the

adegenent package (41) in R was implemented to analyze and visualize genetic diversity of PRRSV-2 ORF5 sequences. To analyze genetic difference of a RFLP among different genetic lineages, ORF5 sequences were aligned using the progressive method (FFT-NS-1) in MAFFT v7.407 (31) and imported to R and then DAPC version 2.1.5 (40) was performed to visualize genetic difference.

PRRSV-2 ORF5 reference sequences for the refined genetic classification system and GenBank accession numbers. The 1,100 PRRSV-2 ORF5 reference sequences for the refined genetic classification system included 59 L1A, 34 L1B, 29 L1C.1, 19 L1C.2, 30 L1C.3, 26 L1C.4, 30 L1C.5, 30 L1C-Others, 37 L1D, 32 L1E, 41 L1F, 40 L1H, 20 L1I, 39 L1J, 30 L2, 39 L3, 23 L4, 47 L5A, 20 L5B, 31 L6, 41 L7, 35 L8A, 38 L8B, 67 L8C, 39 L8D, 40 L8E, 28 L9A, 25 L9B, 30 L9C, 30 L9D, 20 L9E, 27 L10, and 24 L11 sequences (Supplemental fasta file). Among them, 478 sequences were already present in GenBank and we deposited the remaining 622 sequences into GenBank with the accession numbers from OR293404 to OR294025.

4. Significant results

Refine PRRSV-2 genetic classification system based on global ORF5 sequences. A dataset of 82,237 PRRSV-2 ORF5 sequences from various countries was used in this study (Table 1). Following the procedures described in Materials and Methods, global PRRSV-2 ORF5 sequences were classified into 11 genetic lineages (L1 to L11) on the basis of the nine genetic lineage classification system previously proposed by Shi, et al (6) with the addition of two new lineages (L10 and L11). As exemplified in Fig 1, there were over 85% bootstrap supports in lineages L1–L8 and L10–L11 while there was 50.6% bootstrap support in lineage L9.

In this study, 21 sublineages were proposed. These include 9 sublineages in lineage L1 (L1A–L1F, L1H–L1J) that refines those described in Paploski et al (28), 2 sublineages in lineage L5 (L5A and L5B), 5 sublineages in lineage L8 (L8A–L8E), and 5 sublineages in lineage L9 (L9A–L9E) (Fig. 2a–d).

Average pairwise genetic distances between and within lineages or sublineages are summarized in Table 2. PRRSV-2 pairwise nucleotide distances at the inter-lineage levels were typically >11% but were in the range of 9.06% (between L5 and L7) to 17.18% (between L3 and L6). Sequences detected in L5 and L7 are mainly vaccine-like sequences. If L5 and L7 sequences are excluded from the analysis, the pairwise nt distances at the inter-lineage levels would be in the range of 10.88% (between L8 and L9) to 17.18% (between L3 and L6). The average pairwise nt distances at the intra-lineage levels were typically <11% but ranged from 0.46% within lineage L7 to 11.61% within lineage L3. L7 and L5 mainly comprised vaccine viruses and had low intra-lineage pairwise nt distance (0.46%–2.69%); L10, comprising sequences solely from Thailand, also had low intra-lineage pairwise nt distance of 2.22%. Average within-sublineage distances were typically <8.5%, and sublineages belonging to the same lineage were typically >9% divergent.

The newly proposed PRRSV-2 ORF5-based phylogenetic classification in this study was cross-tabulated with two PRRSV-2 ORF5-based phylogenetic classification systems proposed in previous studies (6, 7, 28) as summarized in Table 3. Lineages L10 and L11 in our proposed classification system were previously undescribed and they have been only detected in Thailand and South Korea, respectively. Our proposed PRRSV-2 sublineage names are different from the 37 sublineages distributed in lineages 1, 5, 8 and 9 proposed by Shi et al in 2010 (6), in part because of newly evolved genetic diversity in the past decade and in part because reference sequences needed to apply the sublineages proposed by Shi et al. are not available. Recently,

Paploski et al proposed 9 sublineages within lineage L1 (L1A–L1C, L1Dalpha, L1Dbeta, L1E–L1H), but did not propose sublineages for other lineages (7, 28). Our proposed refinement of L1 sublineages attempt to be consistent with those described by Paploski et al. However, some modifications are proposed in the current study and a summary is provided in Table 3. For example, the sublineages L1B and L1G proposed by Paploski et al generally cluster together and it is difficult to consistently distinguish them; therefore, L1B and L1G are combined into L1B in our new system and the use of L1G is discontinued. Similarly, L1Dalpha and L1E proposed by Paploski et al cluster together and it is difficult to accurately differentiate them in the tree; therefore, L1Dalpha and L1E are combined into L1E in our new system and the use of L1Dalpha is discontinued. L1Dbeta proposed by Paploski et al is therefore simplified to L1D in our new system and the use of L1Dbeta is discontinued. In this study, we also described two new sublineages: L1I and L1J. Hence, the new proposed sublineages in L1 include L1A–L1F, and L1H–L1J. The proposed sublineages L5A and L5B respectively correspond to L5.1 and L5.2 described by Shi et al (6), with L5A including the PRRSV-2 prototype virus VR-2332 and the commonly used Ingelvac PRRS MLV vaccine virus. Sublineages L8.1–L8.9 and L9.1–L9.17 were proposed by Shi et al (6); however, we have refined lineage 8 to include sublineages L8A–L8E and lineage 9 to include sublineages L9A–L9E.

The classification system proposed in this study is flexible for growth if additional lineages, sublineages, or more granular classifications are needed in the future. Here, we use the newly emergent PRRSV-2 L1C variant as an example. Beginning in October 2020, high mortality and morbidity of pigs associated with PRRSV was observed in Iowa and Minnesota swine farms and the virus was subsequently detected in other states. ORF5 sequence analysis suggested that the PRRS viruses from these cases formed a distinct cluster within the sublineage L1C and most of them had a 1-4-4 RFLP pattern - for which reason these PRRS viruses have been referred to as “PRRSV L1C 1-4-4 variant” or sometimes “PRRSV L1C variant” (42, 43). However, sublineage L1C still includes numerous other clusters and it remained undetermined how to denote these clusters with a distinct name that distinguishes them for epidemiological applications. In this study, we performed phylogenetic analysis using 10,961 sequences classified in sublineage L1C and proposed five groups L1C.1–L1C.5 with high bootstrap support to describe sequences that consistently formed unique monophyletic clusters. “L1C variant” sequences (n=822) were annotated as L1C.5, several other L1C clusters were annotated as L1C.1 (n=1,644), L1C.2 (n=270), L1C.3 (n=1,358), and L1C.4 (n=2,157). A large number of L1C sequences that did not consistently form distinct genetic clusters were annotated as L1C-Others (n=4,710) (Fig. 3a and Table 3). The pairwise distances within each of L1C.1 to L1C.5 were 1.29%–3.99% and the pairwise distances between L1C.1–L1C.5 were 5.51%–9.99% (Fig. 3b).

Relationship of PRRSV-2 RFLP tying and ORF5-based lineage/sublineage classification. Among 82,237 PRRSV-2 ORF5 sequences, 341 RFLP patterns were determined. The analyses revealed that the number of RFLP patterns detected in lineages or sublineages did not correspond to genetic differences within lineages or sublineages. For example, sublineage L1A (n=16,625 sequences) had the lowest average pairwise genetic distance (3.64%) among all sublineages in lineage L1 but 86 distinct RFLP patterns were detected in L1A (Fig. 4a). In contrast, sublineage L1E (n=2,134 sequences) had the highest average pairwise genetic distance (10.06%) among all sublineages in lineage L1 but only 50 distinct RFLP patterns were identified in L1E (Fig. 4b). These observations suggest that the number of RFLP patterns in these two sublineages is not a good indicator of virus genetic diversity.

In Table 3, the three most frequently detected RFLP patterns in each lineage and sublineage are presented. Some RFLP patterns were strictly detected in a particular lineage or sublineage. For example, all sequences with RFLP 2-5-2 (n=13,519) were only detected in sublineage L5A (mainly a vaccine virus sublineage) with 0.8% average pairwise genetic distance; among 10,043 sequences with RFLP 1-7-4, most sequences (n=9,872; 98.3%) were classified in sublineage L1A with 3.4% average pairwise genetic distance. In contrast, some RFLP patterns were detected in multiple lineages and sublineages. For example, nine of the 21 sublineages (and six of nine sublineages within L1 specifically) included RFLP 1-4-4 as one of the three most frequent patterns. Sequences with RFLP 1-4-4 (n=10,726) were mostly detected in L1C (n=6,974; 65.0%) and L1A (n=1,948; 18.2%), followed by L1H (n=469; 4.4%), L1E (n=308; 2.9%), and L1F (n=150; 1.4%) with 9.4% average pairwise genetic distance (Fig. 4c). Similarly, sequences with RFLP 1-8-4 (n=8,286) were mostly detected in L1H, L1F, L1A, and L1D with 9.7% average pairwise genetic distance (Fig. 4d), and this RFLP type was listed in the three most common patterns for six of nine L1 sublineages. Sequences with RFLP 1-4-2 (n=6,769) were mostly detected in L8A, L9A, and L9C, followed by L1C and L1E with 9.3% average pairwise genetic distance (Fig. 4e). Sequences with RFLP 1-3-2 (n=3,998) were mostly detected in L8C followed by L1C, L9A, L1E, L8A, and L3 with 10.0% average pairwise genetic distance (Fig. 4f). Altogether, these data suggest that sequences with RFLP 2-5-2 or 1-7-4 have relatively lower genetic diversity but sequences with RFLP 1-4-4, 1-8-4, 1-4-2, or 1-3-2 have relatively higher genetic diversity (i.e., sequences with the same RFLP type can have genetic distances of >9%). Therefore, RFLP typing alone cannot accurately reflect genetic diversity and relatedness of different PRRSV-2 strains in most scenarios.

Genetic lineages/sublineages, RFLPs, and detection frequency of vaccine-like viruses. Six commercial PRRSV-2 modified live virus (MLV) vaccines including Ingelvac PRRS MLV (Boehringer Ingelheim), Ingelvac PRRS ATP (Boehringer Ingelheim), Fostera PRRS (Zoetis), Prime Pac PRRS RR (Merck), Prevacent PRRS (Elanco), and PRRSGard (Pharmgate) vaccines were launched or relaunched in 1994, 1997, 2012, 2018, 2018, and 2020, respectively, to control PRRSV in swine herds. In the USA, all six commercial vaccines are officially approved for use. The lineage/sublineage and RFLP information of the six commercial PRRSV-2 MLV vaccines as well as their ORF5 nt identity ranges to PRRSV-2 viruses in each lineage and sublineage is summarized in Table 4. Each vaccine has variable nt identities to PRRSV-2 viruses in different lineages and sublineages. However, nt identities are commonly found in the 80–90% range, which potentially makes it challenging to estimate the protective efficacy of vaccines against viruses in different lineages and sub-lineages using exclusively ORF5 nt identity.

As no standard cutoff of nt identity was defined to distinguish between vaccine-like and wild-type viruses, in the current study, we arbitrarily defined any sequence with ≥98% ORF5 nt identity to a vaccine virus to be vaccine-like, 95-<98% to be vaccine-like suspect, and <95% ORF5 nt identity to be wild-type virus. Sequences considered as vaccine-like suspects could possibly be vaccine-like or wild-type descendants and more data are required to define the true status of such sequences, which was beyond the focus of this study.

Ingelvac PRRS MLV vaccine virus belongs to the sublineage L5A and has the RFLP 2-5-2. As shown in Table 5a, among 16,518 ORF5 sequences in sublineage L5A, 14,614 sequences (88.5%), 1,257 sequences (7.6%), and 647 sequences (3.9%) were considered Ingelvac PRRS MLV vaccine-like viruses, vaccine-like suspects, and wild-type viruses, respectively. Among 13,519 sequences with RFLP 2-5-2 in L5A, most of them (n=13,162, 97.4%) were Ingelvac PRRS MLV vaccine-like viruses and only 333 (2.5%) and 24 (0.2%) sequences were vaccine-

like suspects and wild-type viruses, respectively (Table 5b). The 14,614 Ingelvac PRRS MLV vaccine-like viruses had very low 0.8% average pairwise genetic distance but included 38 different RFLP patterns, with 13,162 sequences (90.1%) having RFLP 2-5-2 and 1,452 sequences (9.9%) having RFLPs other than 2-5-2 (Table 5c).

Ingelvac PRRS ATP vaccine virus belongs to the sublineage L8A with RFLP 1-4-2, Fostera PRRS vaccine virus is in the sublineage L8C with RFLP 1-3-2, Prime Pac RR vaccine virus belongs to the lineage L7 with RFLP 1-4-4, and Prevacent PRRS vaccine virus is in the sublineage L1D with RFLP 1-8-4. The detection frequency of these MLV vaccine-like viruses among the sequences in the lineage or sublineage to which they respectively belong is summarized in Table 5a, 5b, and 5c in a similar way as for Ingelvac PRRS MLV. As PRRSGard vaccine is a chimeric vaccine between two different PRRSV-2 strains, ORF5 sequence alone cannot distinguish PRRSGard-like and wild-type strains. Therefore, PRRSGard vaccine-like virus data were not included in Table 5.

Global geographic distribution of PRRSV-2 ORF5-based lineages and sublineages. Ninety-seven ORF5 sequences without country information for sample collection were excluded from the investigation of PRRSV-2 geographic distribution. Based on the available information, historical circulations of PRRSV-2 based on ORF5 genetic lineages and sublineages in global regions are summarized in Table 3 and Figure 5.

At the lineage level, L1, L5, L8 and L9 sequences were detected in numerous countries whereas L2, L3, L4, L6, L7, L10 and L11 sequences were only detected in a limited number of countries (Table 3). For example, L3 sequences were mainly detected in Asian countries and regions such as mainland China, Taiwan, Hong Kong, and South Korea. L4, L10, and L11 sequences were only detected in Japan, Thailand, and South Korea, respectively (Table 3). At the sublineage level, L1A, L1B, L1C, L1I, L5A, L5B, L8A, L8E, and L9D were detected in multiple countries whereas L1D and most of L1E, L1F, L1H, L1J, L8B, L8C, L8D, L9A, L9B, L9C, and L9E were detected in a limited number of countries and regions (Table 3).

Among 122 PRRSV-2 ORF5 sequences reported from Europe, 121 ORF5 sequences were in sublineage L5A, in which 77 ORF5 sequences collected during 1997-2017 were considered as Ingelvac PRRS MLV vaccine-like viruses while 39 and 5 ORF5 sequences collected during 2003-2018 were considered as vaccine-like suspects and wild-type viruses, respectively. These 121 L5A sequences were from Austria (n=4), Denmark (n=90), Germany (n=18), Hungary (n=3), Lithuania (n=1), Poland (n=4), and Spain (n=1) (Table 1 and Table 3). The only one non-L5A sequence was collected in 2012 from Hungary (GenBank accession number KM514315) and it belonged to L1E.

Among 571 PRRSV-2 ORF5 sequences from Southeast Asia, most sequences (n=551) were from Thailand and Vietnam and only 20 sequences were from other countries (Table 1). All 17 sequences reported in Cambodia (n=6), Laos (n=4), Myanmar (n=6), and Philippines (n=1) belonged to L8E; for two sequences from Malaysia, one belonged to L5A and the other one could not be assigned to a lineage; and one sequence from Singapore belonged to L5A. In Thailand, L10, L8E, L1I, and L5A sequences were detected with the number and percentage of sequences shown in Fig 5a. In addition, a few L5B and undefined sequences were detected in Thailand. In Vietnam, L8E and L5A sequences were detected (Fig 5b). In South Asia, 55 L8E sequences were reported from India in this study (Table 1 and Table 3). In these areas, L8E sequences mainly included the so-called HP-PRRSV-like strains that emerged in China in 2006 and subsequently spread to other Asian countries (22, 24, 26).

During 1991–2020, 5,157 PRRSV-2 ORF5 sequences were reported from East Asia in GenBank and most of them were from China (n=4,188) during 1996-2020 based on the available data in this study. In mainland China, the sequences from high to low frequency in this study mainly were classified as L8E (HP-PRRSV, HP-PRRSV-like, and CH-1a-like viruses) during 1996–2019, L1C (called NADC30-like strains in China) during 2013–2019, L3 during 2009–2019, and L5A during 1996–2018 (Fig 5c). Other sequences reported in China at <1% included L1A (called NADC34-like strains in China) during 2017–2019, L9B, L5B, L1B, and undefined lineages. In Hong Kong, a few sequences in L8E, L3 and L5A were reported. In Taiwan, 153 L3 sequences during 1991–2018, 13 L5A sequences during 1992–2011, one L1A sequence (year 2018), and one L5B sequence (year 2013) were reported. In Japan, 38 L4, 6 L5A, one for each of L1F, L5B, L6, L8, L9D sequences, and some sequences with unidentified lineage/sublineage were reported. In South Korea, the sequences from high to low frequency in this study mainly included L5A during 2001–2018, L11 during 1999–2018, L1J during 2005–2019, L5B during 2000–2015, L3 during 2002–2016, L1C during 2014–2017, and L1E (n=3) detected in 2005 (Fig 5d).

In this study, limited data were available from Central and South America. Only one PRRSV-2 sequence was reported from Guatemala in Central America which belonged to L8, and 32 PRRSV-2 sequences were reported in South American countries including Peru (20 L1A sequences), Chile (11 L1B sequences), and Venezuela (one L2 sequence).

A large number of PRRSV-2 ORF5 sequences (n=76,202) were included from North America in this study. Among them, most sequences (n=74,119) were collected in the USA, specifically from the Iowa State University Veterinary Diagnostic Laboratory database (n=55,524) and the US-SPD database (n=18,595) during 1989-2021, while fewer sequences were from Mexico (n=1,918) and Canada (n=165). Among ORF5 sequences reported from Canada, there were three main genetic lineages including lineage L1 (mainly L1H and L1C sequences and a few sequences in L1E, L1F and L1I) during 1991-2018, L5 (L5A) during 1995–2018, and L8 (L8A and L8C) during 2005–2018 (Fig 5e). Most sequences classified in L5A and L8A were Ingelvac PRRS MLV vaccine-like viruses and Ingelvac PRRS ATP vaccine-like viruses, respectively. In Mexico, the top three sequence lineages included lineage L1 (mainly L1B and L1A sequences and a few sequences in L1C, L1E and L1H) during 2009–2021, L5 (L5A) during 2004–2021, and L8 (mainly L8D sequences and a few sequences in L8A and L8C) during 2008–2021 (Fig 5f). In addition, seven L2, 20 L6, four L9A, and 51 L9D sequences were reported in Mexico. Among 432 L5A sequences from Mexico, most sequences were Ingelvac PRRS MLV vaccine-like viruses (n=394) reported during 2004–2021 while 19 and 19 sequences reported during 2006–2019 were vaccine-like suspects and wild-type viruses, respectively. Sequences in sublineage L8D had a distinct geographic distribution, found only in Mexico, and most ORF5 sequences (n=495 of 572) in L8D had a distinct nt length (609 bp). Lineage and sublineage information of PRRSV-2 in the USA is described in the section below.

Geographic distribution and temporal dynamic changes of PRRSV-2 in the USA. The 74,119 PRRSV-2 ORF5 sequences collected in the USA during 1989-2021 were classified into seven different genetic lineages, including the four major lineages L1, L5, L8 and L9 (Fig 5g), a few minor lineages L2 (n=99; 0.1%), L6 (n=303; 0.4%), and L7 (n=98; 0.1%), and undefined lineages (n=165; 0.2%). For the four major lineages L1, L5, L8 and L9 in the USA, sequences were further characterized at the sublineage level. A total of 45,265 sequences in lineage L1 were annotated into sublineages L1A, L1B, L1C, L1D, L1E, L1F, and L1H (Fig 5h) together with L1I (n=43; 0.1%), and undefined sublineage (n=111; 0.3%). L1J was not found in the USA.

A total of 15,571 ORF5 sequences in lineage L5 were annotated in L5A and L5B (Fig 5i). A total of 6,641 ORF5 sequences in lineage L8 were classified into L8A, L8B, and L8C (Fig 5j) together with L8E (n=1; 0.01%), and undefined sublineage (n=386; 5.8%). L8D was not found in the USA. The only one L8E sequence reported in the USA (GenBank #KU504046) was from the publication by Alkhamis et al. but no detailed information about that sequence was provided (44). A total of 5,977 ORF5 sequences in lineage L9 were classified into L9A, L9B, L9C, L9D, and L9E (Fig 5k), together with some undefined sequences (n=12; 0.2%).

Since 1,027 PRRSV-2 ORF5 sequences from the USA had no collection years, they were excluded from the temporal analysis and thus 73,092 PRRSV-2 ORF5 sequences collected in the USA during 1989–2021 were used for analyzing temporal dynamic changes.

Lineage L1, representing ~7.5% of sequences during 1989–2001 and increasing to 29.6% in 2002, became the dominant lineage from 2004–2021 in the USA with a dramatic increase from 33.3% to 74.4% (Fig. 6). The distribution of sublineages within L1 is shown in Fig. 7. Sublineage L1A increased from 8.5% in 2002 to 21.9% in 2003 and declined from 16.6 % in 2007 to 2.02% in 2013. Subsequently, L1A dramatically increased and became the dominant sublineage during 2015–2021, rising from 27.2% in 2014 to over 60% during 2015–2018 and then declining to ~40% in 2021, a pattern also shown in another study (28). Sublineage L1B, representing ~10% of the sequences during 2002–2005, increased to 24.0% in 2006 and became a dominant sublineage observed in over 40% during 2008–2009. After that, L1B decreased from 37.5% in 2013 to less than 2% during 2019–2021. Expansion and contraction were not distinctly observed in sublineage L1D representing less than 8% of the sequences during 2003–2021, probably because L1D has previously been described as the earliest documented L1 sub-lineage that likely reached its peak in 2003 (28). Prevacent PRRS vaccine is an MLV vaccine belonging to L1D and was launched in 2018. Coinciding with this, all ORF5 sequences in lineage L1 during 1989–2018 were wild-type viruses, with roughly over 10% ORF5 nt distance to the Prevacent PRRS vaccine, while Prevacent-like viruses in L1D were observed during 2019–2021 rising from 3.6% to 7.5% of sequences (Fig. 7). Sublineage L1E represented over 20% from 2002 to 2003 and then declined to less than 10% after 2005. L1F was a dominant sublineage representing ~40% during 2002–2006 then decreased from 33.8% in 2007 to 2.6% in 2014 and dropped to less than 1% during 2015–2021. Sublineage L1H rose from 1.69% in 2013 to over 20% during 2019–2021. L1I sequences were detected at very low levels from 2003 to 2010 and were rarely detected after 2010.

Sublineage L1C, which represented less than 7% during 2002–2007, became a dominant sublineage during 2010–2013 increasing to over 40% of sequences. L1C subsequently decreased to less than 20% during 2015–2020. In recent years, L1C rose from 11.7% in 2020 to 28.4% in 2021. At the L1C.1 to L1C.5 group level, L1C.1 sequences increased from 4.7% in 2010 to over 29% in 2014–2015 then dropped to less than 1% after 2018 (Fig 3c). L1C.2 sequences were overall at low detection levels for several years but it is noteworthy that L1C.2 sequences has had an increasing trend since 2018 (Fig 3c). L1C.3 dramatically increased from 6.5% in 2015 to 41.0% in 2016 and became a dominant L1C group during 2016-2020. Detection frequency of L1C.4 sequences increased from 30.4% detection in 2009 to over 50% of L1C sequences in 2010-2011 then decreased to 10.4% in 2014 and less than 1% after 2017. A few L1C.5 (L1C variant) sequences were first detected in 2018-2019 with little attention until 2020 when the detection frequency increased to ~5.8% in 2020 followed by a sharp increase to 59.4% in 2021 (Fig 3c).

Lineage L5 detection ranged roughly from 8.4%–16.6% of all sequences during 2002–2009 and rose to ~20% during 2010–2021 (Fig. 6). For the temporal changes of sublineages in L5 shown in Fig. 8a, L5A was the most detected sublineage from 2002 to 2021 while a small number of sequences belonging to L5B were observed only from 2002 to 2008. Among L5A sequences, wild-type viruses with <95% nt identity to Ingelvac PRRS MLV vaccine accounted for ~4.5–17.2% of the sequences from 2002 to 2005 and then dropped to less than 3% after 2012 (Fig 8b). Sequences with 95–<98% nt identity to Ingelvac PRRS MLV vaccine, considered as vaccine-like suspects, accounted for 17.4–39.1% of L5A sequences during 2002–2009 and reduced to ~3–6% during 2010–2021. Sequences with ≥98% nt identity to Ingelvac PRRS MLV vaccine expanded from 56.4% of L5A sequences in 2002 to 62.1% in 2008. Concurrent with the decrease in wild-type and vaccine-like suspect viruses in L5A, Ingelvac PRRS MLV-like viruses increased from 74.9% in 2009 to 93.5% of L5A sequences in 2021 (Fig 8b).

Lineage L8 sequences accounted for ~10% of sequences during 2002–2010 and in the range of 4.6% to 10% during 2011–2021 (Fig. 6). The temporal changes of sublineages within L8 indicate that L8A was the primary L8 sublineage from 2003 to 2013 (Fig. 9a), which increased from 49.3% of L8 sequences in 2003 to generally over 60% during 2004–2012, and then declined from 58.6% in 2013 to 12.3% in 2021. Sublineage L8B increased from 15.5% of L8 sequences in 2002 to 42.8% in 2008 then declined to 5.2% in 2015, and no sequences in L8B was reported during 2016–2021. Sublineage L8C, with 15.7% of L8 sequences in 2012, became the primary L8 sublineage during 2014–2021, increasing from 52.2% in 2014 to over 80% of L8 sequences during 2016–2021. As Ingelvac PRRS ATP vaccine (L8A) was launched in 1997, vaccine-like viruses with ≥98% nt identity to Ingelvac PRRS ATP vaccine accumulated to 36.2–66.7% of L8A sequences during 2002–2004 and up to ~78–98% during 2005–2020 and then reduced to ~61% in 2021 (Fig. 9b). Vaccine-like suspects with 95-<98% nt identity to Ingelvac PRRS ATP vaccine were reported in the range of ~3–13% of L8A sequences during 2003–2020 and increased to ~35% in 2021 while wild-type L8A viruses with <95% nt identity to Ingelvac PRRS ATP vaccine reduced from 63.8% in 2002 to less than 4% during 2009–2021 (Fig. 9b). Regarding Fostera PRRS vaccine (L8C) which was launched in 2012, only a vaccine parent strain was reported in 1995 and no other sequences in L8 was reported to have ≥98% nt identity to Fostera PRRS vaccine until 2012 (Fig 9c). Fostera vaccine-like viruses (≥98% nt identity to Fostera vaccine) were first observed in 2012 and were in the range of ~94–100% of L8C sequences during 2012–2019, and then dropped to 64.9% in 2020 and 30.4% in 2021 (Fig. 9c). Vaccine-like suspect sequences in L8C with 95-<98% nt identity to Fostera PRRS vaccine were observed at ~1–6% of L8C sequences during 2013–2019 and increased to 34.8% in 2020 and 68.6% in 2021 while wild-type L8C viruses with <95% nt identity to Fostera PRRS vaccine were observed less than 1% during 2012–2021 (Fig. 9c).

Lineage L9 was a dominant lineage before 2004 representing >30% of sequences and declined to ~20% during 2005–2009 and to less than 1% after 2013 (Fig 6). This finding corresponded to the previous studies that lineage 9 declined overtime (25, 28). The frequency of sequences reported in lineage L9 is shown in Fig. 8c. For example, L9A was the most common sublineage during 2002-2007, representing over 60% of L9 sequences before 2006, and fluctuated from ~24% to ~ 40% during 2008–2013 (Fig 8c). L9C increased from 4.9% of L9 sequences in 2002 to 44.5% in 2007 and then fluctuated from ~33% to 55% during 2008–2013 (Fig 8c).

Lineages L2, L6, and L7 accounted for a small number of sequences reported in the USA across years as shown in Fig. 6. L2 represented less than 1% of sequences from 2004 to 2011 and

has not been reported since 2012, while L6 represented less than 1% of sequences from 2009 to 2021. L7 was not reported during 2007–2013 and was observed in less than 1% of sequences during 2014–2021.

Eight states with over 1,000 ORF5 sequences in each state reported from 2015 to 2021 (Iowa, Illinois, Indiana, Minnesota, Missouri, North Carolina, Nebraska and Oklahoma) were selected to investigate whether the detection frequency of PRRSV-2 lineages/sublineages varied at the state level. During 2015–2021, L1 was the dominant lineage in all eight states, representing 60.2–79.1% of sequences, while lineages L5 and L8 were reported for 15.7–33.5% and 1.9–8.8%, respectively, across the eight states (Fig 10a), indicating a lack of major differences in PRRSV lineage distribution between states. When the detection frequency of PRRSV-2 L1 sublineages was compared across states (Fig. 10b), L1A was the dominant sublineage in seven of eight states (except Oklahoma) between 2015–2021, and the proportion of L1A sequences varied from 40.9–90.9% of sequences among states. Sublineage L1C occurrence ranged from 19.3–30.2% in Iowa, Illinois, Indiana, Minnesota, and Nebraska, while L1C accounted for a smaller proportion (~10% of sequences) in Missouri, North Carolina, and Oklahoma. Sublineage L1H was the dominant sublineage in Oklahoma (representing over 80% of sequences), while it accounted for less than 20% of sequences in other states and apparently absent in North Carolina. A small number of sequences from L1B, L1D, L1E, and L1F were reported in some states, with L1E being somewhat more common in Oklahoma.

5. Discussion of how results can be applied by practitioners

In this study, based on analysis of 82,237 global ORF5 sequences reported during 1989–2021, we classified PRRSV-2 into eleven genetic lineages (L1–L11) and 21 sublineages (L1A–L1F, L1H–L1J, L5A–L5B, L8A–L8E, and L9A–L9E). The lineages L1–L9 were overall similar to those proposed by Shi et al in 2010 (6). But, L10 and L11 were proposed for the first time in this study to describe PRRSV-2 sequences detected mostly in Thailand (L10) and South Korea (L11). The sublineages within L1 proposed in this study are overall consistent with what Paploski et al described in 2019 and 2021 (7, 28) with some exceptions. For example, merging the former L1B and L1G into a single sublineage L1B with discontinued use of term L1G, combining the former L1Dalpha and L1E into new L1E, renaming L1Dbeta to L1D with discontinued use of L1Dbeta, and proposing the new sublineages L1I and L1J. Compared to the nine sublineages (8.1–8.9) in L8 and 17 sublineages (9.1–9.17) in L9 originally proposed by Shi et al (6), we have proposed 5 sublineages for both L8 (L8A–L8E) and L9 (L9A–L9E) to simplify the classification and facilitate their use. The proposed classification system in this study is flexible for growth if additional lineages, sublineages, or more granular classifications are needed. For example, for more fine-scale epidemiological investigation, L1C was further divided into five groups (L1C.1–L1C.5), with L1C.5 corresponding to the recently emerged L1C variant. In the future, if there are demands for epidemiological studies, sequences in other genetically diverse sublineages can be similarly sub-divided into more smaller groups that may be more relevant for routine PRRSV monitoring and investigation by swine health professionals.

Using the refined PRRSV-2 ORF5 classification system, we analyzed the relationship between RFLP typing and lineages/sublineages. Although sequences with RFLP 2-5-2 were restrictedly detected in sublineage L5A and sequences with RFLP 1-7-4 were mostly classified in sublineage L1A, sequences with other RFLP patterns were detected in multiple lineages and sublineages. For example, sequences with RFLP 1-4-4 were detected in L1C, L1A, L1H, L1E, L1F, etc.; sequences with RFLP 1-8-4 were detected in L1H, L1F, L1A, L1D, etc.; sequences

with RFLP 1-4-2 were detected in L8A, L9A, L9C, L1C, L1E, etc.; and sequences with RFLP 1-3-2 were detected in L8C, L1C, L9A, L1E, L8A, L3, etc. Overall, the data clearly demonstrate that RFLP typing alone cannot accurately reflect genetic diversity and relatedness of different PRRSV-2 strains in most scenarios.

Commercial MLV vaccines are commonly used to control PRRSV infection. ORF5 sequencing via the Sanger method is still the most commonly used method to distinguish PRRSV-2 vaccine strains from wild-type strains although vaccine-like PCRs and next-generation sequencing technology have started to be used in recent years. However, there is no standard ORF5 nt identity cutoff value to define vaccine-like viruses and wild-type viruses. In the current study, we arbitrarily defined any sequence with ≥98% ORF5 nt identity to a vaccine virus to be vaccine-like, 95-<98% to be suspected vaccine-like, and <95% ORF5 nt identity to be wild-type virus. Genetic homology of six commercial PRRSV-2 vaccines to each lineage/sublineage and detection frequency of vaccine-like viruses was determined and summarized in Tables 4 and 5, serving as a valuable resource for swine practitioners. Notably, L8C viruses having 95–<98% ORF5 nt identity to the Fostera vaccine virus was increasingly detected in 2020–2021, coinciding with the decreasing percentage of L8C viruses having 98-100% ORF5 nt identity to the Fostera vaccine virus. It warrants further investigation and monitoring in the future to determine whether these L8C sequences having 95–<98% ORF5 nt identity to the Fostera vaccine virus are due to the further divergence of Fostera vaccine or not.

Global geographic distribution of each PRRSV-2 lineage/sublineage was presented in Table 3 and Figure 5. We have also investigated geographic and temporal dynamic changes of PRRSV-2 in the USA during 1989–2021 by analyzing 73,092 ORF5 sequences, as summarized in Figures 5–10. Thus far, this is the most thorough study describing the molecular epidemiology of global PRRSV-2. Such information is critical for understanding the circulation status of PRRSV-2 worldwide. After presenting the data at several conferences, we have received numerous requests from different parties in the USA asking us to share the PRRSV-2 distribution data. These data will be invaluable for future characterization of PRRSV-2.

The reference sequences (n=1,100) representing our refined lineages and sublineages are provided together with this paper. So far, the USDA Swine Pathogen Database (US-SPD), Iowa State University Veterinary Diagnostic Laboratory (ISU VDL), University of Minnesota, Swine Disease Reporting System (SDRS) have indicated the willingness to use this refined classification system. In fact, the ISU VDL and SDRS are almost ready to implement this new PRRSV-2 ORF5 classification system. In order to make it more convenient for use, we have worked with Nextclade to incorporate our classification system. A beta- testing version is available now at the link https://clades.nextstrain.org/?dataset-url=https://github.com/mazeller/NextClade_Datasets/tree/main/prrsv_yimim_2023. People around the world can enter their PRRSV-2 ORF5 sequences and obtain the lineage/sublineage information in a few seconds. This will not only be a useful and convenient tool for swine veterinarians, diagnosticians, researchers, and practitioners in the USA but also facilitate global standardization and application of PRRSV-2 genetic classification.

6. References

1. Cavanagh D. 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Arch Virol 142:629-633.
2. Kappes MA, Faaberg KS. 2015. PRRSV structure, replication and recombination: Origin of phenotype and genotype diversity. Virology 479-480:475-486.

3. Lunney JK, Fang Y, Ladinig A, Chen N, Li Y, Rowland B, Renukaradhya GJ. 2016. Porcine Reproductive and Respiratory Syndrome Virus (PRRSV): Pathogenesis and Interaction with the Immune System. Annu Rev Anim Biosci 4:129-154.
4. 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.
5. Stadejek T, Oleksiewicz MB, Scherbakov AV, Timina AM, Krabbe JS, Chabros K, Potapchuk D. 2008. Definition of subtypes in the European genotype of porcine reproductive and respiratory syndrome virus: nucleocapsid characteristics and geographical distribution in Europe. Arch Virol 153:1479-1488.
6. Shi M, Lam TT, Hon CC, Murtaugh MP, Davies PR, Hui RK, Li J, Wong LT, Yip CW, Jiang JW, Leung FC. 2010. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. J Virol 84:8700-11.
7. Paploski IAD, Corzo C, Rovira A, Murtaugh MP, Sanhueza JM, Vilalta C, Schroeder DC, VanderWaal K. 2019. Temporal Dynamics of Co-circulating Lineages of Porcine Reproductive and Respiratory Syndrome Virus. Front Microbiol 10:2486.
8. Wesley RD, Mengeling WL, Lager KM, Clouser DF, Landgraf JG, Frey ML. 1998. Differentiation of a porcine reproductive and respiratory syndrome virus vaccine strain from North American field strains by restriction fragment length polymorphism analysis of ORF 5. J Vet Diagn Invest 10:140-144.
9. Han J, Wang Y, Faaberg KS. 2006. Complete genome analysis of RFLP 184 isolates of porcine reproductive and respiratory syndrome virus. Virus Res 122:175-182.
10. van Geelen Albert GM, Anderson TK, Lager KM, Das PB, Otis NJ, Montiel NA, Miller LC, Kulshreshtha V, Buckley AC, Brockmeier SL, Zhang J, Gauger PC, Harmon KM, Faaberg KS. 2018. Porcine reproductive and respiratory disease virus: Evolution and recombination yields distinct ORF5 RFLP 1-7-4 viruses with individual pathogenicity. Virology 513:168-179.
11. Trevisan G, Sharma A, Gauger P, Harmon KM, Zhang J, Main R, Zeller M, Linhares LCM, Linhares DCL. 2021. PRRSV2 genetic diversity defined by RFLP patterns in the United States from 2007 to 2019. J Vet Diagn Invest 33:920-931.
12. Guo Z, Chen X-x, Li R, Qiao S, Zhang G. 2018. The prevalent status and genetic diversity of porcine reproductive and respiratory syndrome virus in China: a molecular epidemiological perspective. Virol J 15:2.
13. Neira V, Brito B, Mena J, Culhane M, Apel MI, Max V, Perez P, Moreno V, Mathieu C, Johow M, Badia C, Torremorell M, Medina R, Ortega R. 2017. Epidemiological investigations of the introduction of porcine reproductive and respiratory syndrome virus in Chile, 2013-2015. PLoS One 12:e0181569.
14. Ramirez M, Bauermann FV, Navarro D, Rojas M, Manchego A, Nelson EA, Diel DG, Rivera H. 2019. Detection of porcine reproductive and respiratory syndrome virus (PRRSV) 1-7-4-type strains in Peru. Transbound Emerg Dis 66:1107-1113.
15. Kvisgaard LK, Hjulsager CK, Brar MS, Leung FC, Larsen LE. 2013. Genetic dissection of complete genomes of Type 2 PRRS viruses isolated in Denmark over a period of 15 years. Vet Microbiol 167:334-44.
16. Shi M, Lemey P, Singh Brar M, Suchard MA, Murtaugh MP, Carman S, D'Allaire S, Delisle B, Lambert ME, Gagnon CA, Ge L, Qu Y, Yoo D, Holmes EC, Chi-Ching Leung F. 2013. The spread of type 2 Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in North America: a phylogeographic approach. Virology 447:146-54.
17. Brar MS, Shi M, Ge L, Carman S, Murtaugh MP, Leung FC. 2011. Porcine reproductive and respiratory syndrome virus in Ontario, Canada 1999 to 2010: genetic diversity and restriction fragment length polymorphisms. J Gen Virol 92:1391-7.
18. Brar MS, Shi M, Murtaugh MP, Leung FC. 2015. Evolutionary diversification of type 2 porcine reproductive and respiratory syndrome virus. J Gen Virol 96:1570-80.

19. Cui X, Xia D, Huang X, Sun Y, Shi M, Zhang J, Li G, Yang Y, Wang H, Cai X, An T. 2022. Analysis of Recombinant Characteristics Based on 949 PRRSV-2 Genomic Sequences Obtained from 1991 to 2021 Shows That Viral Multiplication Ability Contributes to Dominant Recombination. Microbiol Spectr 10:e0293422.
20. Rupasinghe R, Lee K, Liu X, Gauger PC, Zhang J, Martinez-Lopez B. 2022. Molecular Evolution of Porcine Reproductive and Respiratory Syndrome Virus Field Strains from Two Swine Production Systems in the Midwestern United States from 2001 to 2020. Microbiol Spectr 10:e0263421.
21. Xu H, Xiang L, Tang YD, Li C, Zhao J, Gong B, Sun Q, Leng C, Peng J, Wang Q, Zhou G, An T, Cai X, Tian ZJ, Zhang H, Song M. 2022. Genome-Wide Characterization of QYYZ-Like PRRSV During 2018-2021. Front Vet Sci 9:945381.
22. Do HQ, Trinh DT, Nguyen TL, Vu TT, Than DD, Van Lo T, Yeom M, Song D, Choe S, An DJ, Le VP. 2016. Molecular evolution of type 2 porcine reproductive and respiratory syndrome viruses circulating in Vietnam from 2007 to 2015. BMC Vet Res 12:256.
23. Guzman M, Melendez R, Jimenez C, Piche M, Jimenez E, Leon B, Cordero JM, Ramirez-Carvajal L, Uribe A, Van Nes A, Stegeman A, Romero JJ. 2021. Analysis of ORF5 sequences of Porcine Reproductive and Respiratory Syndrome virus (PRRSV) circulating within swine farms in Costa Rica. BMC Vet Res 17:217.
24. Jantafong T, Sangtong P, Saenglub W, Mungkundar C, Romlamduan N, Lekchareonsuk C, Lekcharoensuk P. 2015. Genetic diversity of porcine reproductive and respiratory syndrome virus in Thailand and Southeast Asia from 2008 to 2013. Vet Microbiol 176:229-38.
25. Kikuti M, Sanhueza J, Vilalta C, Paploski IAD, VanderWaal K, Corzo CA. 2021. Porcine reproductive and respiratory syndrome virus 2 (PRRSV-2) genetic diversity and occurrence of wild type and vaccine-like strains in the United States swine industry. PLoS One 16:e0259531.
26. Jiang Y, Li G, Yu L, Li L, Zhang Y, Zhou Y, Tong W, Liu C, Gao F, Tong G. 2020. Genetic Diversity of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) From 1996 to 2017 in China. Front Microbiol 11:618.
27. Yu F, Yan Y, Shi M, Liu HZ, Zhang HL, Yang YB, Huang XY, Gauger PC, Zhang J, Zhang YH, Tong GZ, Tian ZJ, Chen JJ, Cai XH, Liu D, Li G, An TQ. 2020. Phylogenetics, Genomic Recombination, and NSP2 Polymorphic Patterns of Porcine Reproductive and Respiratory Syndrome Virus in China and the United States in 2014-2018. J Virol 94:e01813-19.
28. Paploski I, Pamornchainavakul N, Makau DN, Rovira A, Corzo C, Schroeder DC, Cheeran M, Doeschl-Wilson A, Kao RR, Lycett S, VanderWaal K. 2021. Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States. Vaccines 9:608.
29. Anderson TK, Inderski B, Diel DG, Hause BM, Porter EG, Clement T, Nelson EA, Bai J, Christopher-Hennings J, Gauger PC, Zhang J, Harmon KM, Main R, Lager KM, Faaberg KS. 2021. The United States Swine Pathogen Database: integrating veterinary diagnostic laboratory sequence data to monitor emerging pathogens of swine. Database 2021.
30. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl Environ Mircobiol 75:7537-7541.
31. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772-780.
32. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268-274.
33. Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution 31:151-160.

34. Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170-179.
35. Shimodaira H, Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. Mol Biol Evol 16:1114-1116.
36. Shimodaira H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. Syst Biol 51:492-508.
37. Zmasek CM, Eddy SR. 2001. ATV: display and manipulation of annotated phylogenetic trees. Bioinformatics 17:383-384.
38. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 8:28-36.
39. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547-1549.
40. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94.
41. Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.
42. Trevisan G, Li G, Moura CAA, Coleman K, Thomas P, Zhang J, Gauger P, Zeller M, Linhares D. 2021. Complete Coding Genome Sequence of a Novel Porcine Reproductive and Respiratory Syndrome Virus 2 Restriction Fragment Length Polymorphism 1-4-4 Lineage 1C Variant Identified in Iowa, USA. Microbiol Resour Announc 10:e0044821.
43. Kikuti M, Paploski IAD, Pamornchainavakul N, Picasso-Risso C, Schwartz M, Yeske P, Leuwerke B, Bruner L, Murray D, Roggow BD, Thomas P, Feldmann L, Allerson M, Hensch M, Bauman T, Sexton B, Rovira A, VanderWaal K, Corzo CA. 2021. Emergence of a New Lineage 1C Variant of Porcine Reproductive and Respiratory Syndrome Virus 2 in the United States. Front Vet Sci 8:752938.
44. Alkhamis MA, Perez AM, Murtaugh MP, Wang X, Morrison RB. 2016. Applications of Bayesian Phylodynamic Methods in a Recent U.S. Porcine Reproductive and Respiratory Syndrome Virus Outbreak. Front Microbiol 7:67.

TABLE 1 Countries and collection years of PRRSV-2 ORF5 sequences from 1989–2021 used in this study

Region/Country	Collection years of PRRSV-2 ORF5 sequences							
	1989–2000	**2001–2005**	**2006–2010**	**2011–2015**	**2016–2018**	**2019–2021**	**Undefined collection year**	**Grand Total**
East Asia	**18**	**127**	**1,266**	**1,765**	**959**	**77**	**945**	**5,157**
Mainland China	1	40	1,058	1,397	828	75	789	4,188
Hong Kong, China		12						12
Taiwan, China	7	11	3	28	75		44	168
Japan	3		19	1			32	55
South Korea	7	64	186	339	56	2	80	734
Southeast Asia		**1**	**179**	**244**	**5**		**142**	**571**
Cambodia			2	4				6
Laos			1	3				4
Malaysia				2				2
Myanmar				6				6
Philippines			1					1
Singapore			1					1
Thailand		1	114	159	2		121	397
Vietnam			60	70	3		21	154
South Asia				**23**	**30**	**2**		**55**
India				23	30	2		55
Europe	**2**	**21**	**33**	**34**	**2**		**30**	**122**
Austria							4	4
Denmark	1	11	25	29			24	90
Germany	1	10	7					18
Hungary				2	2			4
Lithuania			1					1
Poland				3			1	4
Spain							1	1
North America	**972**	**6,910**	**11,878**	**23,357**	**15,502**	**16,532**	**1,051**	**76,202**
Canada	17	21	28	3	91		5	165
USA	955	6,862	11,741	22,725	15,143	15,665	1,028	74,119
Mexico		27	109	629	268	867	18	1,918
Central America	**1**							**1**
Guatemala	1							1
South America				**20**	**12**			**32**
Chile				11				11
Peru				8	12			20
Venezuela				1				1
Undefined country		1					96	97
Grand Total	**993**	**7,060**	**13,356**	**25,443**	**16,510**	**16,611**	**2,264**	**82,237**

TABLE 2 Average pairwise genetic distance (% difference) for PRRSV-2 ORF5 sequences classified into 11 lineages and 21 sublineages based on the newly proposed classification.

a) Average pairwise genetic distance (% difference) of 11 intra-lineages and inter-lineages (L1-L11)

Lineage	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
L1	10.96										
L2	14.87	10.46									
L3	15.87	16.58	11.61								
L4	13.27	13.56	14.27	9.08							
L5	13.79	12.69	15.03	11.01	2.69						
L6	15.33	15.02	17.18	13.53	11.42	6.87					
L7	13.27	12.85	14.26	9.73	9.06	11.55	0.46				
L8	14.52	13.85	15.39	12.07	11.03	12.47	9.90	8.31			
L9	14.92	14.41	15.84	12.25	11.67	12.31	9.60	10.88	9.59		
L10	14.25	13.75	15.18	11.47	11.18	13.37	9.30	11.59	11.62	2.22	
L11	14.15	14.17	15.12	12.53	11.97	13.92	12.08	13.22	13.29	13.26	9.64

b) Average pairwise genetic distance (% difference) of nine intra- and inter-sublineages (L1A-L1F, L1H-L1J)

Sublineage	L1A	L1B	L1C	L1D	L1E	L1F	L1H	L1I	L1J
L1A	3.95								
L1B	10.21	6.21							
L1C	12.21	13.20	7.01						
L1D	11.28	12.97	11.60	5.09					
L1E	12.26	13.53	13.76	12.78	10.32				
L1F	11.16	12.63	12.05	11.30	13.17	5.56			
L1H	12.92	14.18	13.53	11.15	14.03	12.27	6.36		
L1I	11.63	13.01	13.22	12.02	13.45	12.26	13.71	10.17	
L1J	12.05	14.18	13.55	12.49	14.15	11.66	13.99	13.11	8.49

c) Average pairwise genetic distance (% difference) of two intra- and inter-sublineages of L5A-L5B

Sublineage	L5A	L5B
L5A	1.49	
L5B	9.82	8.79

d) Average pairwise genetic distance (% difference) of five intra- and inter-sublineages of L8A-L8E

Sublineage	L8A	L8B	L8C	L8D	L8E
L8A	2.47				
L8B	9.75	6.01			
L8C	6.91	7.85	1.52		
L8D	14.11	14.68	12.81	6.29	
L8E	10.07	10.24	7.43	14.59	2.87

e) Average pairwise genetic distance (% difference) of five intra- and inter-sublineages of L9A-L9E

Sublineage	L9A	L9B	L9C	L9D	L9E
L9A	8.44				
L9B	10.10	7.45			
L9C	10.49	11.16	6.35		
L9D	10.47	10.87	10.59	6.82	
L9E	9.50	9.45	9.46	9.87	6.54

TABLE 3 The comparison between the newly proposed PRRSV-2 classification and two previously proposed PRRSV-2 classifications based on ORF5 sequences as well as RFLP typing and geographic distribution

Classification in this study	Number of ORF5 sequences	Sample collection year	Classification by Shi et al, 2010 [13]	Classification by Paploski et al, 2021 [10]	RFLP typing (only listing top 3 RFLPs)	Country/region distribution (number of sequences)	Distribution at continental level
Lineage L1	47,077	1991-2021; N/D (290)	Lineage 1	Lineage 1	1-7-4 (21.1%), 1-4-4 (21.1%), 1-8-4 (17.7%)	CAN (96), CHL (11), CHN (693), HUN (1), JPN (1), KOR (171), MEX (757), PER (20), THA (54), TWN (1), USA (45,265), Undefined (7)	East Asia, Southeast Asia, Europe, North America, South America, Undefined
Sublineage L1A	16,625	2000-2021; N/D (8)	L1.5?	L1A	1-7-4 (59.4%), 1-4-4 (11.7%), 1-8-4 (7.1%)	CHN (14), MEX (101), PER (20), TWN (1), USA (16,489)	East Asia, North America, South America
Sublineage L1B	7,548	1998-2021; N/D (10)	Undefined	L1B and L1G	1-26-2 (34.8%), 1-18-2 (31.5%), 1-18-4 (10.0%)	CAN (1), CHL (11), CHN (2), MEX (643), USA (6,891)	East Asia, North America, South America
Sublineage L1C	10,961	2000-2021; N/D (160)	L1.8?	L1C	1-4-4 (63.6%), 1-3-4 (12.9%), 1-3-2 (5.3%)	CAN (13), CHN (677), KOR (11), MEX (10), USA (10,249), Undefined (1)	East Asia, North America, Undefined
L1C.1	1,644	2005-2021; N/D (1)	Undefined	Undefined	1-4-4 (83.8%), 1-3-4 (5.7%), 1-8-4 (2.5%)	USA (1,644)	North America
L1C.2	270	2006-2021	Undefined	Undefined	1-2-4 (75.9%), 1-4-4 (15.6%), 1-1-4 (4.1%)	USA (270)	North America
L1C.3	1,358	2014-2021	Undefined	Undefined	1-3-2 (39.1%), 1-3-4 (31.9%), 1-4-4 (14.7%)	USA (1,358)	North America
L1C.4	2,157	2004-2020; N/D (1)	Undefined	Undefined	1-4-4 (86.0%), 1-4-3 (5.0), 1-3-4 (2.5%)	USA (2,157)	North America
L1C.5	822	2020-2021	Undefined	Undefined	1-4-4 (91.7%), 1-3-4 (1.8%), 1-4-3 (1.8%)	USA (822)	North America
L1C-Others	4,710	2000-2021; N/D (158)	Undefined	Undefined	1-4-4 (58.3%), 1-3-4 (17.3%), 1-4-3 (5.6%)	CAN (13), CHN (677), KOR (11), MEX (10), USA (3,998), Undefined (1)	East Asia, North America, Undefined
Sublineage L1D	1,428	2001-2021; N/D (4)	Undefined	L1Dbeta	1-8-4 (74.2%), 1-12-4 (6.9%), 1-8-3 (3.8%)	USA (1,428)	North America
Sublineage L1E	2,134	1998-2021; N/D (29)	Undefined	L1Dalpha and L1E	1-22-2 (21.8%), 1-3-2 (21.1%), 1-4-4 (14.4%)	CAN (2), HUN (1), KOR (3), MEX (1), USA (2,127)	East Asia, Europe, North America

Classification in this study	Number of ORF5 sequences	Sample collection year	Classification by Shi et al, 2010 [13]	Classification by Paploski et al, 2021 [10]	RFLP typing (only listing top 3 RFLPs)	Country/region distribution (number of sequences)	Distribution at continental level
Sublineage L1F	3,680	1999-2021; N/D (51)	L1.9	L1F	1-8-4 (74.5%), 1-16-4 (7.6%), 1-4-4 (4.1%)	CAN (3), JPN (1), USA (3,671), Undefined (5)	East Asia, North America, Undefined
Sublineage L1H	4,310	2004-2021	Undefined	L1H	1-8-4 (70.7%), 1-4-4 (10.9%), 1-12-4 (7.1%)	CAN (53), MEX (1), USA (4,256)	North America
Sublineage L1I	71	1991-2014; N/D (10)	Undefined	Undefined	1-8-4 (42.3%), 1-16-4 (11.3%), 1-16-1 (4.2%)	CAN (6), THA (22), USA (43)	North America, Southeast Asia
Sublineage L1J	158	2005-2019; N/D (9)	Undefined	Undefined	1-21-4 (19.0%), 1-8-4 (18.4%), 1-4-4 (10.8%)	KOR (157), Undefined (1)	East Asia, Undefined
Undefined sublineage in L1	162	1991-2020; N/D (9)	Undefined	Undefined	1-4-2 (25.3%), 1-4-1 (9.9%), 1-4-4 (9.9%)	CAN (18), MEX (1), THA (32), USA (111)	Southeast Asia, North America
Lineage L2	107	2001-2013; N/D (8)	Lineage 2	Same as Shi et al	1-2-4 (35.5%), 1-27-2 (19.6%), 1-20-4 (14.0%)	MEX (7), USA (99), VEN (1)	North America, South America
Lineage L3	455	1991-2019; N/D (69)	Lineage 3	Same as Shi et al	1-3-2(28.1%), 1-4-2 (9.2%), 1-4-4 (9.0%)	CHN (283), HKG (6), KOR (13), TWN (153)	East Asia
Lineage L4	38	1992-2010; N/D (21)	Lineage 4	Same as Shi et al	1-4-4 (44.7%), 1-2-4 (21.1%), 1-1-2 (5.3%)	JPN (38)	East Asia
Lineage L5	16,845	1989-2021; N/D (500)	Lineage 5	Same as Shi et al	2-5-2 (80.3%), 2-6-2 (3.7%), 1-5-2 (2.8%)	AUT (4), CAN (45), CHN (193), DEU (18), DNK (90), ESP (1), HKG (1), HUN (3), JPN (7), KOR (369), LTU (1), MEX (432), MYS (1), POL (4), SGP (1), THA (13), TWN (14), USA (15,571), VNM (6), Undefined (71)	East Asia, Southeast Asia, Europe, North America, Undefined
Sublineage L5A	16,518	1990-2021; N/D (421)	Lineage 5.1	Undefined	2-5-2 (81.8%), 2-6-2 (3.7%), 1-5-2 (2.8%)	AUT (4), CAN (45), CHN (191), DEU (18), DNK (90), ESP (1), HKG (1), HUN (3), JPN (6), KOR (269), LTU (1), MEX (432), MYS (1), POL (4), SGP (1), THA (11), TWN (13), USA (15,375), VNM (6), Undefined (46)	East Asia, Southeast Asia, Europe, North America, Undefined

Classification in this study	Number of ORF5 sequences	Sample collection year	Classification by Shi et al, 2010 [13]	Classification by Paploski et al, 2021 [10]	RFLP typing (only listing top 3 RFLPs)	Country/region distribution (number of sequences)	Distribution at continental level
Sublineage L5B	324	1989-2018; N/D (78)	Lineage 5.2	Undefined	1-4-2 (27.5%), 1-3-4 (19.4%), 1-27-4 (12.0%)	CHN (2), JPN (1), KOR (99), THA (2), TWN (1), USA (194), Undefined (25)	East Asia, Southeast Asia, North America, Undefined
Undefined sublineage in L5	3	1992, 2000; N/D (1)	Undefined	Undefined	1-10-4 (33.3%), 1-2-4 (33.3%), 1-4-2 (33.3%)	KOR (1), USA (2)	East Asia, North America
Lineage L6	324	1991-2019; N/D (5)	Lineage 6	Same as Shi et al	1-3-4 (54.3%), 1-3-2 (12.7%), 1-3-1 (8.3%)	JPN (1), MEX (20), USA (303)	East Asia, North America
Lineage L7	102	1996-2021; N/D (8)	Lineage 7	Same as Shi et al	1-4-4 (96.1%), 1-2-2 (1.0%), 1-3-4 (1.0%)	CHN (1), USA (98), Undefined (3)	East Asia, North America, Undefined
Lineage L8	10,611	1993-2021; N/D (936)	Lineage 8	Same as Shi et al	1-4-2 (29.9%), 1-3-2 (18.5%), 1-4-3 (17.1%)	CAN (20), CHN (3,012), GTM (1), HKG (5), IND (55), JPN (1), KHM (6), LAO (4), MEX (589), MMR (6), PHL (1), THA (113), USA (6,641), VNM (148), Undefined (9)	East Asia, Southeast Asia, South Asia, North America, Central America, Undefined
Sublineage L8A	2,935	1997-2021; N/D (272)	Lineage 8.9	Undefined	1-4-2 (84.4%), 1-3-2 (6.2%), 1-4-4 (2.7%)	CAN (14), MEX (11), USA (2,904), Undefined (6)	North America, Undefined
Sublineage L8B	820	1994-2015; N/D (33)	Undefined	Undefined	1-2-4 (30.9%), 1-4-4 (15.2%), 1-3-4 (10.6%)	USA (820)	North America
Sublineage L8C	2,540	2005-2021; N/D (3)	Undefined	Undefined	1-3-2 (67.6%), 1-4-2 (16.2%), 1-1-2 (9.7%)	CAN (6), CHN (1), MEX (3), USA (2,530)	North America, Undefined
Sublineage L8D	572	2008-2021; N/D (2)	Undefined	Undefined	1-6-3 (76.7%), 1-5-3 (6.4%), 1-2-3 (3.8%)	MEX (572)	North America
Sublineage L8E	3,311	2002-2020; N/D (590)	Lineage 8.7	Undefined	1-4-3 (54.8%), 1-3-3 (28.1%), 1-4-2 (4.4%)	CHN (2,974), HKG (5), IND (55), KHM (6), LAO (4), MMR (6), PHL (1), THA (108), USA (1), VNM (148), Undefined (3)	East Asia, Southeast Asia,
Undefined sublineage in L8	433	1993-1998; N/D (36)	Undefined	Undefined	1-4-2 (30.2%), 1-7-2 (16.2%), 1-4-1 (11.9%)	CHN (37), GTM (1), JPN (1), MEX (3), THA (5), USA (386)	South Asia, North America, Undefined
Lineage L9	6,052	1992-2021; N/D (288)	Lineage 9	Same as Shi et al	1-4-2 (44.8%), 1-3-2 (12.7%), 1-2-2 (11.8%)	CHN (6), JPN (6), KOR (1), MEX (59), THA (1), USA	East Asia, Southeast Asia,

Classification in this study	Number of ORF5 sequences	Sample collection year	Classification by Shi et al, 2010 [13]	Classification by Paploski et al, 2021 [10]	RFLP typing (only listing top 3 RFLPs)	Country/region distribution (number of sequences)	Distribution at continental level
						(5,977), Undefined (2)	North America, Undefined
Sublineage L9A	3,135	1996-2020; N/D (196)	Undefined	Undefined	1-4-2 (41.6%), 1-3-2 (18.0%). 1-1-2 (9.3%)	MEX (4), USA (3131)	North America
Sublineage L9B	429	1995-2015; N/D (29)	Undefined	Undefined	1-3-4 (38.5%), 1-4-2 (21.2%), 1-2-2 (13.5%)	CHN (6), KOR (1), USA (422)	East Asia, North America
Sublineage L9C	1,800	1999-2017; N/D (14)	Undefined	Undefined	1-4-2 (52.2%), 1-2-2 (21.8%), 1-3-2 (7.4%)	USA (1800)	North America
Sublineage L9D	486	1997-2017; N/D (12)	Undefined	Undefined	1-4-2 (52.1%), 1-2-4 (16.5%), 1-4-4 (10.1%)	JPN (1), MEX (51), USA (434)	East Asia, North America
Sublineage L9E	180	1990-2021; N/D (33)	Undefined	Undefined	1-4-2 (66.1%), 1-4-1 (6.7%), 1-3-2 (5.6%)	USA (178), Undefined (2)	North America, Undefined
Undefined sublineage in L9	22	1992-2016; N/D (4)	Undefined	Undefined	1-1-2 (22.7%), 1-4-2 (22.7%), 1-2-2 (13.6%)	JPN (5), MEX (4), THA (1), USA (12)	East Asia, North America, Southeast Asia
Lineage L10	216	2004-2011; N/D (90)	Undefined	Undefined	1-30-4 (77.3%), 1-16-4 (7.9%), 1-59-4 (3.7%)	THA (216)	Southeast Asia
Lineage L11	163	1999-2018; N/D (19)	Undefined	Undefined	1-10-3 (27.6%), 1-3-3 (20.9%), 1-3-4 (6.8%)	KOR (163)	East Asia
Undefined PRRSV-2 sequences	247	1992-2020; N/D (30)	Undefined	Undefined	1-4-4 (17.0%), 1-3-2 (14.2%), 1-3-4 (8.5%)	CAN (4), JPN (1), KOR (17), MEX (54), MYS (1), USA (165), Undefined (5)	East Asia, Southeast Asia, North America, Undefined
Grand total	82,237						

Notes:

1. Three-letter country codes were defined in ISO 3166-1 alpha-3 codes. AUT: Austria; CAN: Canada; CHL: Chile; CHN: China (mainland); DEU: Germany; DNK: Denmark; ESP: Spain; GTM: Guatemala; HKG: Hong Kong; HUN: Hungary; IND: India; JPN: Japan; KHM: Cambodia; KOR: South Korea; LAO: Laos; LTU: Lithuania; MEX: Mexico; MMR: Myanmar; MYS: Malaysia; PER: Peru; PHL: Philippines; POL: Poland; SGP: Singapore; THA: Thailand; TWN: Taiwan; USA: United States of America; VEN: Venezuela; VNM: Vietnam; Undefined: country information is unknown.
2. N/D: no data. 3. East Asia: Mainland China, Hong Kong China, Taiwan China, Japan, and South Korea. 4. Southeast Asia: Cambodia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam. 5. South Asia: India. 6. Europe: Austria, Denmark, Germany, Hungary, Lithuania, Poland, and Spain.
7. North America: Canada, USA, and Mexico. 8. Central America: Guatemala. 9. South America: Chile, Peru, and Venezuela.
10. RFLP patterns are the results of three-digit cutting patterns created by *MluI*, *HincII*, and *SacII* in PRRSV-2 ORF5 sequence.

TABLE 4 ORF5 nucleotide identity ranges of PRRSV-2 sequences in each lineage and sublineage compared to six commercial PRRSV-2 modified live virus vaccines.

Lineage	Commercial PRRSV-2 modified live virus vaccines (RFLP; Lineage)					
	Ingelvac MLV (2-5-2; L5A)	Ingelvac ATP (1-4-2; L8A)	Fostera (1-3-2; L8C)	Prime Pac (1-4-4; L7)	Prevacent (1-8-4; L1D)	PRRSGard (1-8-4; L1F)
Lineage L1	79.1-90.9%	79.3-91.5%	79.4-93%	79.4-90.5%	81.3-100%	81.7-100%
Sublineage L1A	79.1-90.4%	79.3-91.5%	79.4-91.7%	79.4-89.4%	81.8-90.5%	82.4-92%
Sublineage L1B	81.2-90%	80.8-90%	80.8-89.1%	81-89.6%	81.3-90.4%	81.7-91.5%
Sublineage L1C	81.6-87.9%	80.6-87.6%	81.3-89.2%	81.4-88.4%	84.6-91.4%	84.9-94.4%
Sublineage L1D	82.8-88.6%	83.5-88.4%	83.7-89.6%	84.2-89.4%	88.8-100%	87-92.2%
Sublineage L1E	83.7-90.9%	83.4-89.9%	84.2-93%	83.9-90.5%	84.4-90.4%	83.9-91.9%
Sublineage L1F	82.3-90.7%	81.8-88.4%	82.8-90%	82.4-90%	82.4-91.5%	86.7-100%
Sublineage L1H	81.3-87.1%	81.3-87.4%	81.8-87.6%	81.8-88.9%	84.7-94.7%	85.2-91.5%
Sublineage L1I	82.8-88.7%	82.5-88.4%	83.6-89.7%	83.9-89.4%	84.8-90.9%	84.8-92.4%
Sublineage L1J	81.8-87.2%	82.6-87.4%	82.1-88.1%	81.9-88.6%	84.4-89.9%	85.2-91.7%
Lineage L2	78.3-92%	77.1-90.5%	78.4-92.7%	77.4-92.9%	76.9-88.6%	78.4-88.2%
Lineage L3	81.6-90%	81.1-89.2%	82.6-91.2%	82.6-90.5%	81.1-88.4%	80.8-89.4%
Lineage L4	86.4-91.2%	85.9-89.4%	87.4-92%	87.9-92.9%	84.9-89.1%	85.4-90.2%
Lineage L5	84.9-100%	81.3-92.2%	81.9-93%	82.4-92.9%	78.6-92%	78.3-89.2%
Sublineage L5A	84.9-100%	81.3-92.2%	81.9-93%	82.4-92.9%	78.6-92%	78.3-88.4%
Sublineage L5B	85.2-93.9%	84.1-91.2%	85.4-91.7%	83.6-92.2%	81.6-88.2%	81.3-89.2%
Lineage L6	86.5-93.2%	84.7-91.2%	86.4-94.2%	85.6-93.9%	81.9-87.9%	83.1-87.7%
Lineage L7	90.7-91.9%	89.4-90.5%	92.4-93.4%	96.8-100%	87.5-88.1%	86.4-87.4%
Lineage L8	82.4-93.2%	83.9-100%	85.1-100%	82.8-93.5%	80-88.6%	80.3-91.4%
Sublineage L8A	84.1-91.9%	87.6-100%	87.1-95.5%	84.9-92.4%	80.9-87.2%	81.4-91.4%
Sublineage L8B	84.7-90.5%	85.1-92.4%	87.7-95.2%	85.7-92.2%	82.6-87.6%	83.1-87.9%
Sublineage L8C	88.2-92.2%	89.4-94.2%	93-100%	88.7-93.5%	84.2-88.6%	84.6-87.6%
Sublineage L8D	82.4-88.1%	84.1-90%	85.1-92.4%	83.4-89.7%	81.3-85.9%	80.3-86.2%
Sublineage L8E	82.6-93.2%	83.9-93.9%	86.2-96.7%	82.8-92.9%	80-88.1%	80.6-89%
Lineage L9	81.1-92.9%	80.1-93.5%	82.2-96.4%	81.9-94.7%	80.4-88.2%	81.1-89.7%
Sublineage L9A	81.1-91.5%	80.1-92.5%	82.3-95.7%	81.9-93.4%	80.4-88.2%	81.1-88.7%
Sublineage L9B	86.2-92.9%	85.9-93.5%	88.4-96.4%	87.9-94.7%	83.6-87.9%	83.6-88.7%
Sublineage L9C	85.2-90.5%	85.7-91.2%	87.2-93.7%	86.7-92.4%	83.4-88.4%	83.1-88.1%
Sublineage L9D	85.9-91%	86.7-93.4%	89.1-94%	88.4-92.4%	82.4-87.6%	83.4-88.1%
Sublineage L9E	87-91.7%	86.2-93%	89.2-96.2%	86.6-93.7%	82.9-87.9%	82.6-89.7%
Lineage L10	87.1-90.4%	86.1-89.6%	88.9-92%	87.9-91.2%	83.7-86.4%	84.9-88.1%
Lineage L11	84.9-91.7%	84-89.5%	84.7-91%	84.7-91%	82.9-88.5%	82.9-87.8%

Notes:

1. Ingelvac PRRS MLV vaccine GenBank number AF066183. Ingelvac PRRS ATP vaccine (AY656991). Prime Pac PRRS RR vaccine (AF066384). Prevacent PRRS vaccine (KU131568). Fostera PRRS and PRRSGard were sequenced by our group.
2. According to ORF5 analyses, Ingelvac PRRS MLV, Ingelvac PRRS ATP, FosteraPRRS, Prime Pac PRRS RR, Prevacent PRRS, and PRRSGard vaccines were classified in L5A (RFLP 2-5-2), L8A (RFLP 1-4-2), L8C (RFLP 1-3-2), L7 (RFLP 1-4-4), L1D (RFLP 1-8-4), and L1F (RFLP 1-8-4), respectively.

TABLE 5. Summary of RFLP and genetic lineage/sublineage information of PRRSV-2 MLV vaccine-like viruses

a) Frequency of vaccine-like viruses in lineage/sublineage to which each PRRSV-2 MLV vaccine belongs

Lineage/Sublineage	Number of sequences	Vaccine-like sequences	Vaccine-like suspect sequences	Wild-type sequences	Note
L5A	n=16,518	n=14,614 (88.47%)	n=1,257 (7.61%)	n=647 (3.92%)	Ingelvac PRRS MLV
L8A	n=2,935	n=2,412 (82.18%)	n=242 (8.25%)	n=281 (9.57%)	Ingelvac PRRS ATP
L8C	n=2,540	n=2,094 (82.44%)	n=439 (17.28%)	n=7 (0.28%)	Fostera PRRS
L7	n=102	n=101 (99.02%)	n=1 (0.98%)	n=0	Prime Pac PRRS
L1D	n=1,428	n=626 (43.84%)	n=186 (13.02%)	n=616 (43.14%)	Prevacent PRRS

b) Frequency of vaccine-like viruses among the sequences with the same RFLP pattern and lineage/sublineage as each PRRSV-2 MLV vaccine

RFLP	Number of sequences	Vaccine-like sequences	Vaccine-like suspect sequences	Wild-type sequences	Note
RFLP 2-5-2 in L5A	n=13,519	n=13,162 (97.36%)	n=333 (2.46%)	n=24 (0.18%)	Ingelvac PRRS MLV
RFLP 1-4-2 in L8A	n=2,476	n=2,271 (91.72%)	n=144 (5.82%)	n=61 (2.46%)	Ingelvac PRRS ATP
RFLP 1-3-2 in L8C	n=1,712	n=1,637 (95.62%)	n=72 (4.20%)	n=3 (0.18%)	Fostera PRRS
RFLP 1-4-4 in L7	n=98	n=97 (98.98%)	n=1 (1.02%)	n=0	Prime Pac PRRS
RFLP 1-8-4 in L1D	n=1,100	n= 581 (52.82%)	n=103 (9.36%)	n=416 (37.82%)	Prevacent PRRS

c) Distribution of RFLP patterns among each PRRSV-2 MLV vaccine-like viruses

Vaccine-like viruses	Number of sequences	RFLP	Other RFLPs
Ingelvac PRRS MLV-like	n=14,614	n=13,162 (90.06%) RFLP 2-5-2	n=1,452 (9.94%), 37 RFLPs other than 2-5-2
Ingelvac PRRS ATP-like	n=2,412	n=2,271 (94.15%) RFLP 1-4-2	n=141 (5.84%), 8 RFLPs other than 1-4-2
Fostera PRRS-like	n=2,094	n=1,637 (78.18%) RFLP 1-3-2	n=457 (21.82%), 13 RFLPs other than 1-3-2
Prime Pac PRRS-like	n=101	n=97 (96.04%) RFLP 1-4-4	n=4 (3.96%), 3 RFLPs other than 1-4-4
Prevacent PRRS-like	n=626	n=581 (92.81%)RFLP 1-8-4	n=45 (7.12%), 7 RFLPs other than 1-8-4

Note: In this study, we arbitrarily defined any sequence with ≥98%, 95-<98%, or <95% ORF5 nucleotide identity to a vaccine virus as a vaccine-like virus, vaccine-like suspect, or wild-type virus, respectively.

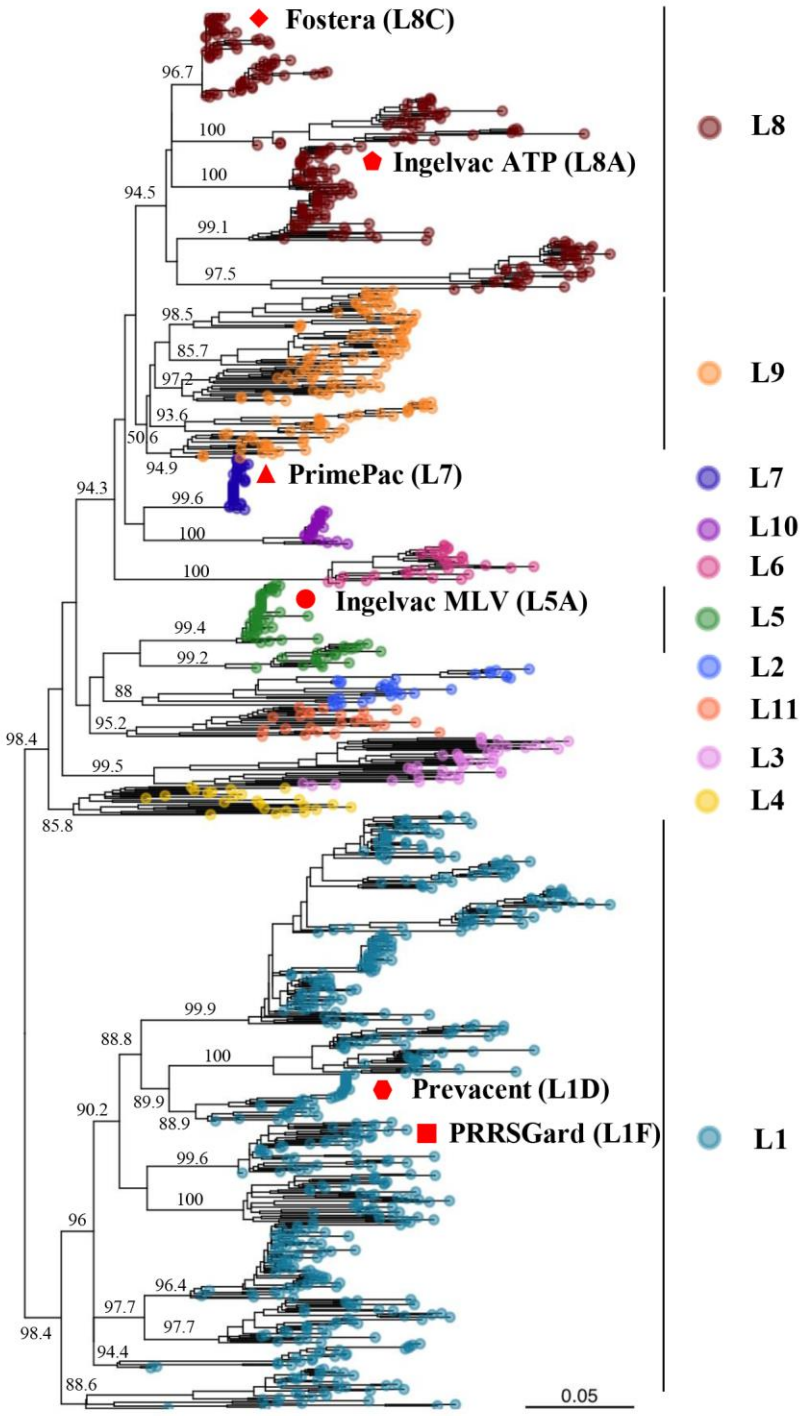


FIG 1. Phylogenetic tree showing the newly proposed PRRSV-2 lineages L1 to L11 based on global ORF5 sequences. Tip points are presented in different colors according to lineages. Bootstrap values are shown for the major clades. Six commercial PRRSV-2 modified live virus vaccines Ingelvac PRRS MLV, Ingelvac PRRS ATP, Fostera PRRS, Prime Pac PRRS RR, Prevacent PRRS, and PRRSGard are classified in the sublineages L5A, L8A, L8C, L7, L1D, and L1F, respectively.

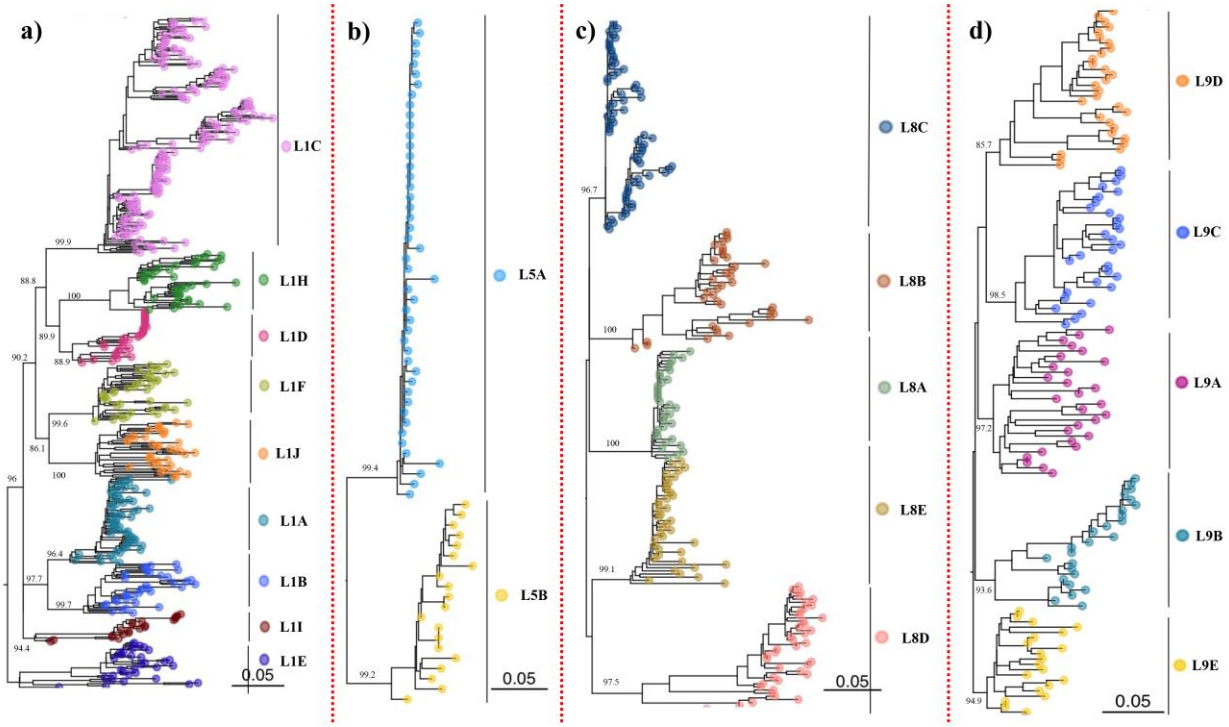


FIG 2. Phylogenetic trees showing PRRSV-2 ORF5-based sublineages within lineages L1, L5, L8 and L9. (a) Nine sublineages L1A, L1B, L1C, L1D, L1E, L1F, L1H, L1I, and L1J in lineage L1. (b) Two sublineages L5A and L5B in lineage L5. (c) Five sublineages L8A, L8B, L8C, L8D, and L8E in lineage L8. (d) Five sublineages L9A, L9B, L9C, L9D, and L9E in lineage L9. Tip points are presented in different colors according to sublineages classified in a particular lineage.

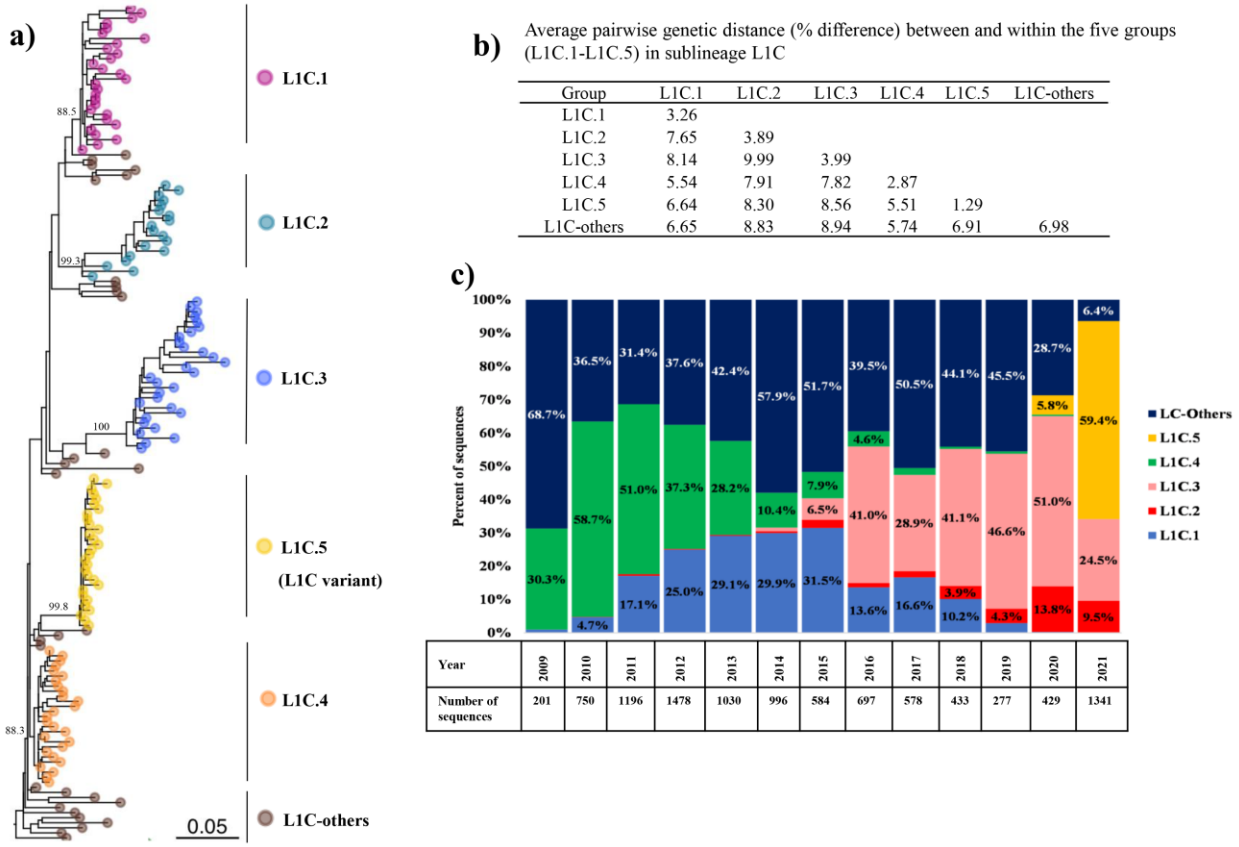


FIG 3. Phylogenetic tree and temporal dynamics of ORF5 sequences classified in sublineage of L1C. (a) Phylogenetic tree of ORF5 sequences classified in groups L1C.1 to L1C.5. (b) Average pairwise genetic distance (% difference) between and within the five groups L1C.1 to L1C.5 in sublineage L1C. (c) Temporal dynamics of U.S. PRRSV-2 ORF5 sequences classified in L1C.1–L1C.5 during 2009-2021. Total number of L1C sequences and percent of sequences classified in L1C.1–L1C.5 are indicated below the graph. Sequences not classified in L1C.1–L1C.5 were defined as L1C-Others.

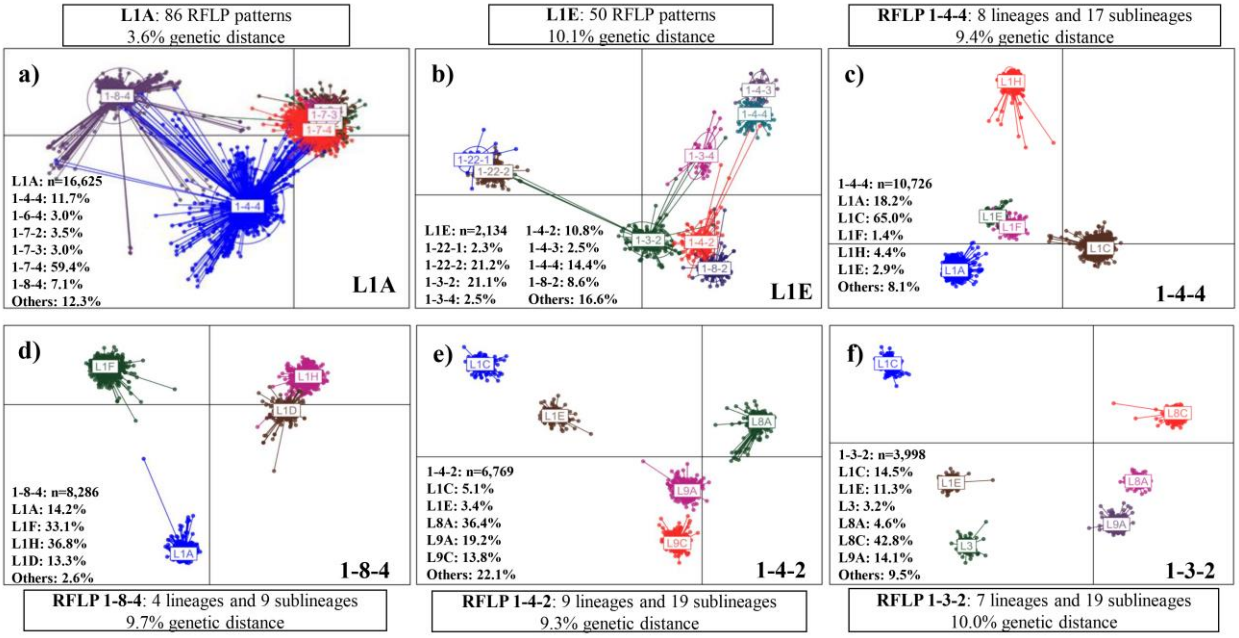


FIG 4. Discriminant analysis of principal components (DAPC) of different sublineages and restriction fragment length polymorphism (RFLP) of PRRSV-2. (a) DAPC of sequences in sublineage L1A. Eighty-six distinct RFLP patterns were detected in 16,625 sequences classified in L1A with 3.64% average pairwise genetic distance. Some of the most frequently detected RFLPs and their percentage in L1A are shown. (b) DAPC of sequences in sublineage L1E. (c) DAPC of sequences with RFLP 1-4-4. Among 10,726 sequences with RFLP 1-4-4 classified into 8 lineages and 17 sublineages, the top five frequently detected sublineages and their percentage are shown. (d) DAPC of sequences with RFLP 1-8-4. (e) DAPC of sequences with RFLP 1-4-2. (f) DAPC of sequences with RFLP 1-3-2.

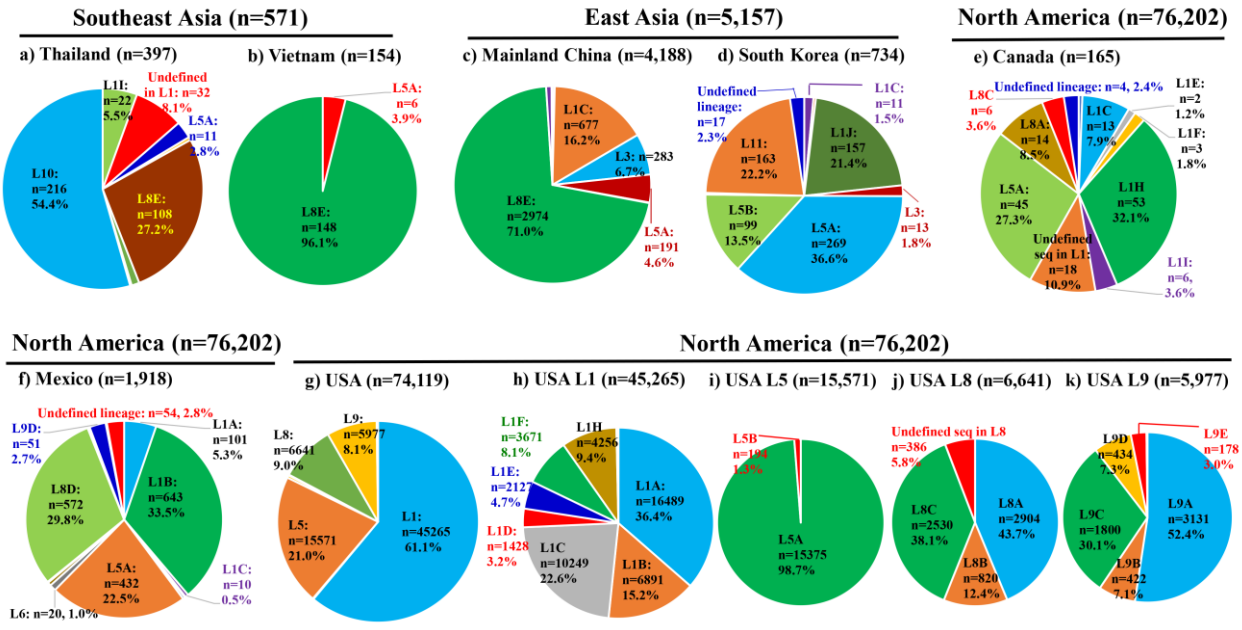
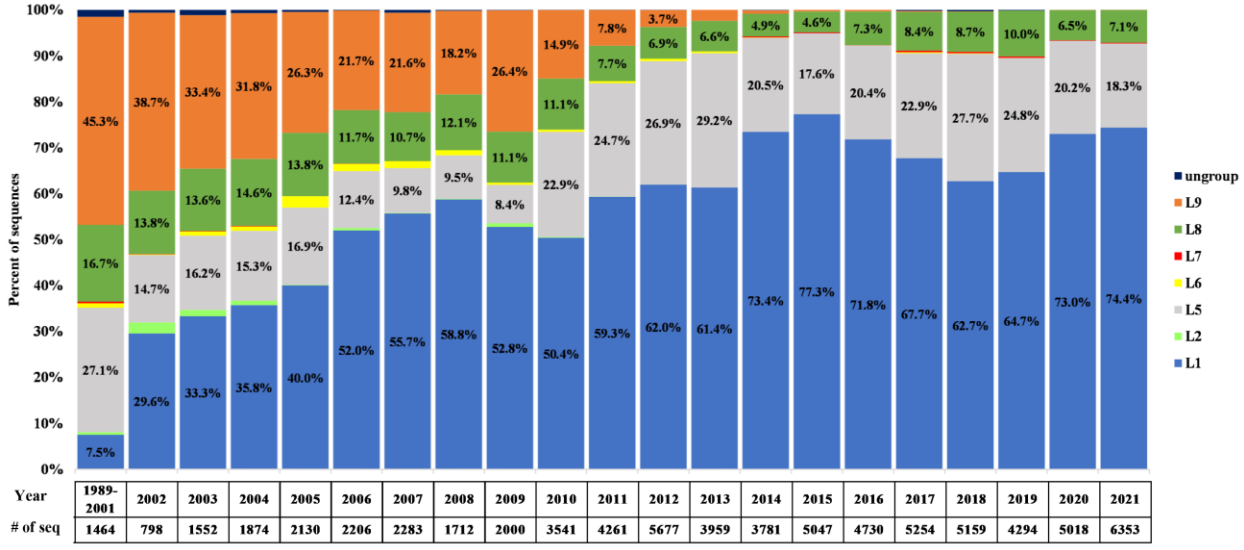
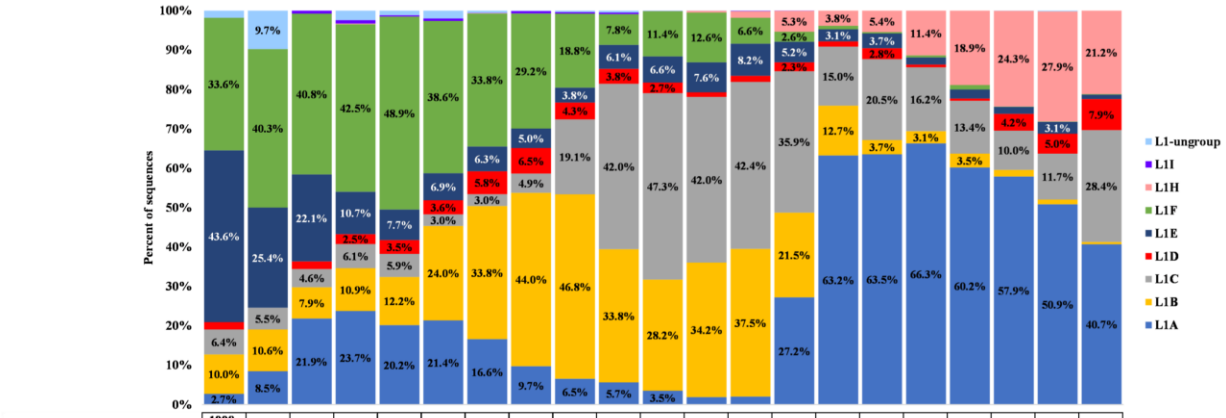


FIG 5. Detection frequency and distribution of PRRSV-2 ORF5-based lineages and sublineages in representative countries based on the dataset in this study. The number and percentage of sequences belonging to the major PRRSV-2 lineages and sublineages in Thailand, Vietnam, mainland China, South Korea, Canada, and Mexico are shown in a) to f). The number and percentage of sequences at the lineage level in the USA are shown in g) and at the sublineage level in the USA are shown in h) to k). Two L5 sequences with undefined sublineage are not shown in i). One L8E sequence is not shown in j).



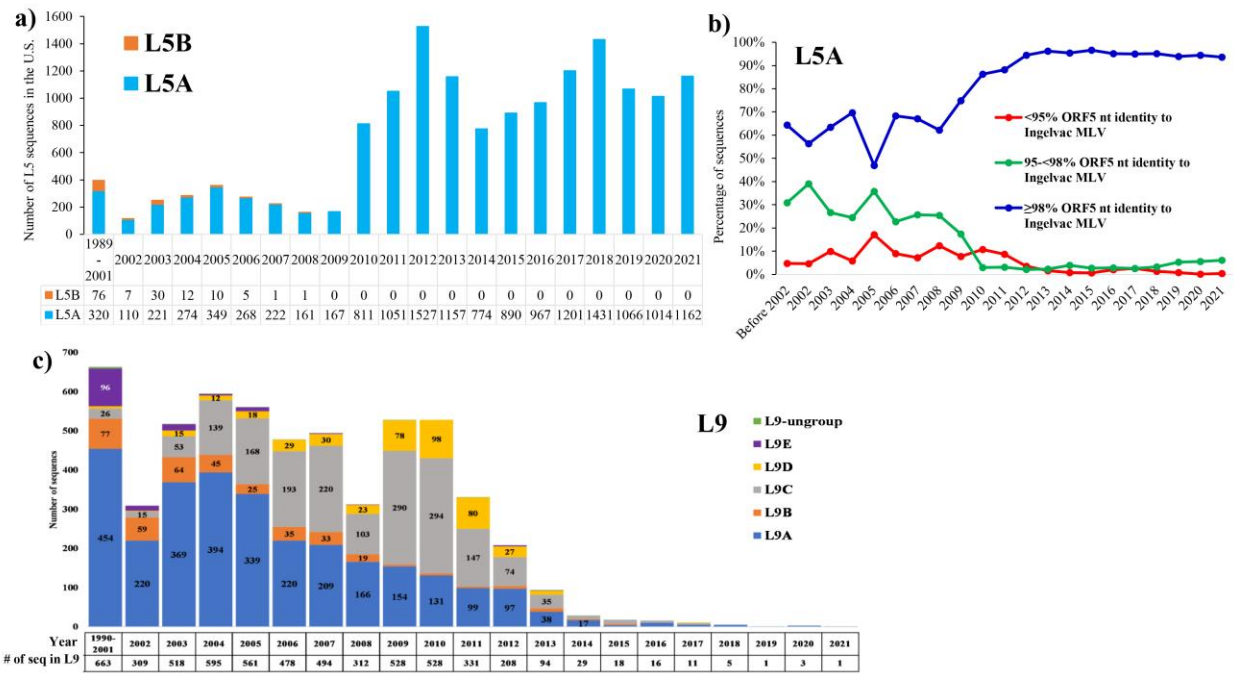
Year	1989-2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
# of seq	1464	798	1552	1874	2130	2206	2283	1712	2000	3541	4261	5677	3959	3781	5047	4730	5254	5159	4294	5018	6353

FIG 6. Temporal dynamics at the lineage level of 73,092 PRRSV-2 ORF5 sequences with samples collected in the USA during 1989–2021. Percent distribution of each lineage is indicated in the graph and the total number of sequences reported in a particular year is indicated in the table below the graph.



	1998-2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Total (Number of sequences) in L1	110	236	517	670	852	1147	1271	1006	1055	1785	2527	3517	2429	2777	3902	3397	3558	3235	2778	3665	4727
<95% to Prevacent in L1D	1.8%	0.0%	1.4%	1.5%	2.8%	2.2%	4.0%	4.3%	3.3%	2.9%	2.0%	1.0%	1.1%	1.9%	1.2%	2.6%	0.7%	0.4%	0.5%	0.2%	0.1%
95-<98% to Prevacent in L1D	0.0%	0.0%	0.6%	1.0%	0.7%	1.4%	1.8%	2.2%	0.9%	0.8%	0.7%	0.2%	0.4%	0.4%	0.2%	0.1%	0.0%	0.2%	0.1%	0.2%	0.3%
>98% to Prevacent in L1D	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%	4.6%	7.5%

FIG 7. Temporal dynamics at the sublineage level of U.S. PRRSV-2 ORF5 sequences classified in lineage L1 during 1998–2021. Percent of each sublineage is shown in the graph and the total number of sequences in the lineage reported in a particular year is indicated in the table below the graph. Percentages of sequences in L1D with <95%, 95-<98%, and ≥98% ORF5 nt identity to Prevacent PRRS vaccine were calculated against the total number of sequences in L1 in a particular year and the data are indicated in the table below the graph.



	1989-2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
L5B	76	7	30	12	10	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
L5A	320	110	221	274	349	268	222	161	167	811	1051	1527	1157	774	890	967	1201	1431	1066	1014	1162

Year	1990-2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
# of seq in L9	663	309	518	595	561	478	494	312	528	528	331	208	94	29	18	16	11	5	1	3	1

FIG 8. Temporal dynamics at the sublineage level of U.S. PRRSV-2 ORF5 sequences classified in lineage L5 or L9 during 1989–2021. Number of sequences in each sublineage is shown in the graph and

the total number of sequences in the lineage reported in a particular year is indicated in the table below the graph. (a) Temporal dynamics of sublineages in lineage L5. (b) Among all L5A sequences in the USA, percentages of L5A sequences with <95%, 95-<98%, and ≥98% ORF5 nt identity to Ingelvac PRRS MLV vaccine in a particular year are shown. (c) Temporal dynamics of sublineages in lineage L9.

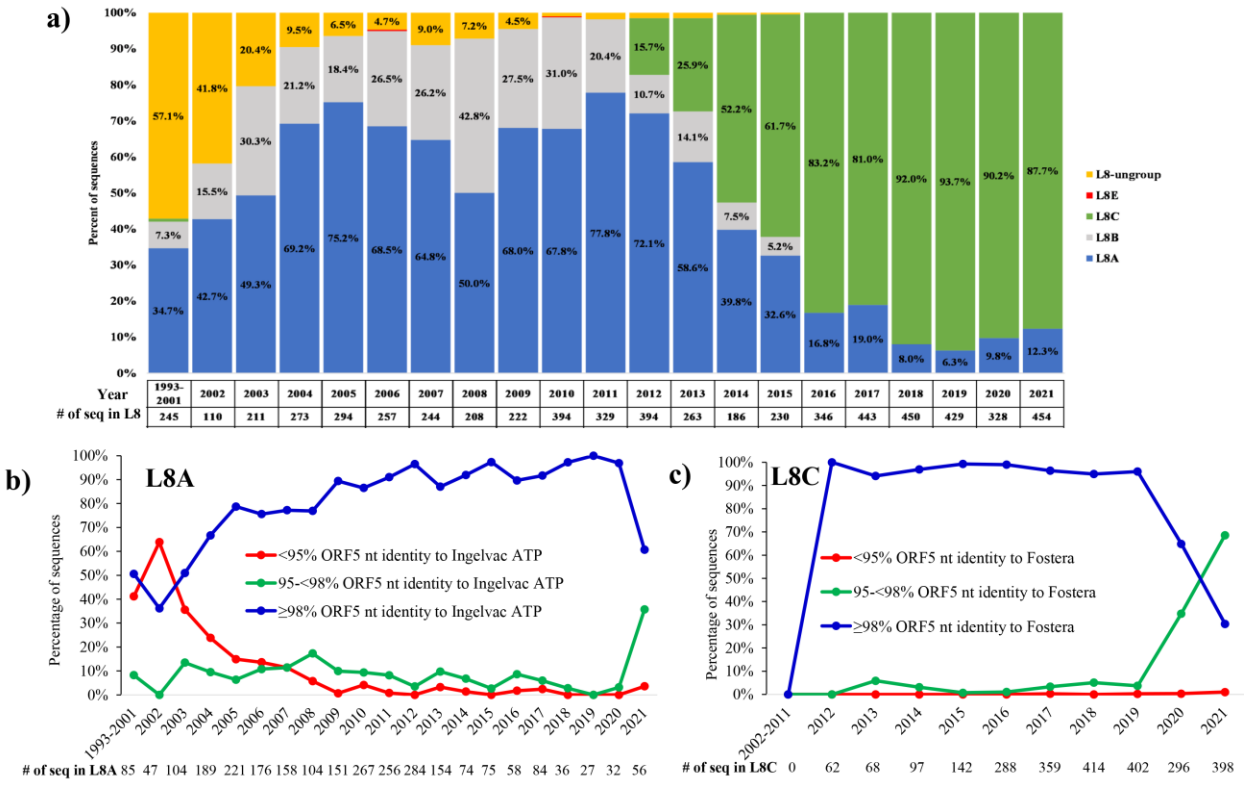
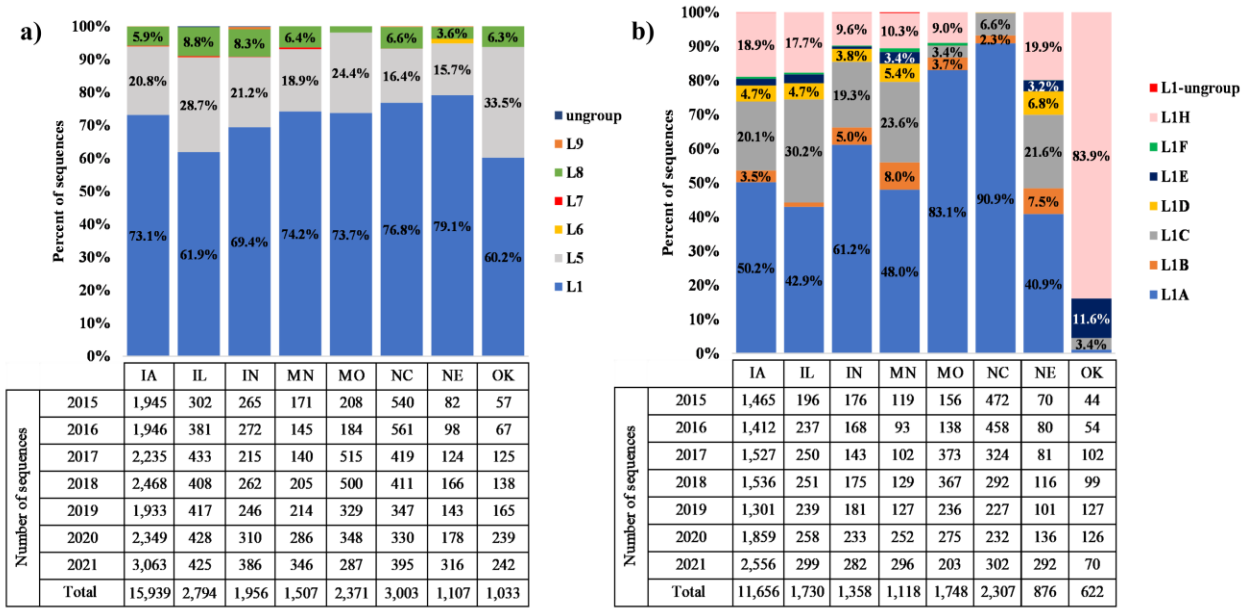


FIG 9. Temporal dynamics at the sublineage level of U.S. PRRSV-2 ORF5 sequences classified in lineage L8 during 1993–2021. (a) Percent of each sublineage is shown in the graph and total number of sequences in the lineage reported in a particular year is indicated in the table below the graph. Among all L8A sequences in the USA from 1993–2021, percentages of L8A sequences with <95%, 95-<98%, and ≥98% ORF5 nt identity to Ingelvac PRRS ATP vaccine in a particular year or period are shown in (b). Among all L8C sequences in the USA from 2002–2021, percentages of L8C sequences with <95%, 95-<98%, and ≥98% ORF5 nt identity to Fostera PRRS vaccine in a particular year or period are shown in (c).



Number of sequences	IA	IL	IN	MN	MO	NC	NE	OK
2015	1,945	302	265	171	208	540	82	57
2016	1,946	381	272	145	184	561	98	67
2017	2,235	433	215	140	515	419	124	125
2018	2,468	408	262	205	500	411	166	138
2019	1,933	417	246	214	329	347	143	165
2020	2,349	428	310	286	348	330	178	239
2021	3,063	425	386	346	287	395	316	242
Total	15,939	2,794	1,956	1,507	2,371	3,003	1,107	1,033

Number of sequences	IA	IL	IN	MN	MO	NC	NE	OK
2015	1,465	196	176	119	156	472	70	44
2016	1,412	237	168	93	138	458	80	54
2017	1,527	250	143	102	373	324	81	102
2018	1,536	251	175	129	367	292	116	99
2019	1,301	239	181	127	236	227	101	127
2020	1,859	258	233	252	275	232	136	126
2021	2,556	299	282	296	203	302	292	70
Total	11,656	1,730	1,358	1,118	1,748	2,307	876	622

FIG 10. Temporal dynamics of U.S. PRRSV-2 ORF5 sequences classified in lineages and sublineages in L1 reported in eight U.S. states during 2015-2021. Eight states are represented by two-letter state codes and the number of sequences reported in each state during 2015-2021 is indicated below the graph. (a) Temporal dynamics of lineages. The percentages were calculated against the total number of sequences during 2015-2021 for each state, respectively. (B) Temporal dynamics of sublineages in L1. The percentages were calculated against the total number of L1 sequences during 2015-2021 for each state, respectively.