

Genetic diversity, spatiotemporal dynamics and evolution of porcine reproductive and respiratory syndrome virus 2 (PRRSV-2) lineage 1 in China

Chenglin Chi

Yangzhou University

Naisu Yang

Yangzhou University

Pingping Zhang

Yangzhou University

Sen Jiang

Yangzhou University

Wanglong Zheng

Yangzhou University

Nanhua Chen

Yangzhou University

Jianzhong Zhu

jzzhu@yzu.edu.cn

Yangzhou University

Research Article

Keywords: PRRSV-2, ORF5, Genetic classification, Lineage, Sublineage

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11287843/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Porcine reproductive and respiratory syndrome virus 2 (PRRSV-2) lineage 1 (L1) has become the dominant epidemic lineage worldwide, posing a serious threat to the swine industry. However, the fine-scale evolutionary history, spatiotemporal transmission dynamics, and molecular drivers of L1 in China remain poorly understood. In this study, 32,215 global PRRSV-2 ORF5 sequences including 1,213 ORF5 sequences from Chinese L1 strains were systemically analyzed. Phylogenetic analysis revealed a unipolar dominance of the L1C sublineage (NADC30-like strains), accounting for 86.12% of Chinese L1 strains and forming a well-supported monophyletic clade. Transmission network analysis identified Henan Province as the primary hub for cross-regional spread of L1C, with Shandong Province as the secondary hub and Fujian, Guangdong and Sichuan Provinces serving as tertiary transmission nodes. Positive selection analysis of ORF5 revealed that L1C accumulated a broader set of positively selected sites, highly enriched in B-cell epitopes and glycosylation sites, indicating a strong selective advantage in immune evasion. Furthermore, whole-genome recombination analysis of L1 identified major recombination hotspots concentrated in certain viral non-structural protein genes (nsp2, nsp9) and structural protein genes (ORF5, ORF6). Together, these findings demonstrate that the unipolar dominance of L1C in China is sustained by a synergistic cycle of transmission, antigenic evolution, and recombination, providing an important basis for precision molecular surveillance and region-specific control of PRRSV-2.

Background

Porcine Reproductive and Respiratory Syndrome (PRRS) is one of the most significant viral diseases affecting the global swine industry. The causative pathogen, PRRSV, can cause severe reproductive disorders in sows and respiratory diseases in pigs of all ages, leading to substantial economic losses [1]. PRRSV is an enveloped, positive-sense, single-stranded RNA virus belonging to the *Arteriviridae* family. Its genome is approximately 15 kilobases in length and is characterized by marked genetic diversity and high mutation potential [2]. PRRSV is primarily divided into two species: PRRSV-1 (European genotype, prototype strain Lelystad virus) [3], first detected in Europe in 1991; and PRRSV-2 (North American genotype, prototype strain VR-2332) [4], first identified in North America in 1992. Since its initial emergence in the United States in the late 1980s [5] and its first reported outbreak in China in 1995 [6], PRRSV-2 has become the dominant circulating genotype in Chinese swine herds.

The rapid development of China's swine industry and its intensive farming practices have created a complex environment conducive to the transmission and evolution of PRRSV. The progression of PRRS in China has involved several critical phases: the emergence of highly pathogenic PRRSV (HP-PRRSV) in 2006 triggered a severe nationwide epidemic [7, 8]; around 2013, NADC30-like strains were introduced into China and rapidly became dominant [9]; and since 2017, the emergence and spread of NADC34-like strains have raised renewed concerns [10, 11]. These events highlight the ongoing dynamic process of continuous evolution and variation of PRRSV-2 within China.

Phylogenetic analysis based on the open reading frame 5 (ORF5) gene sequence, which encodes the major envelope glycoprotein GP5, is a widely used method for investigating the genetic diversity and lineage classification of PRRSV-2 [5]. Globally, PRRSV-2 strains are classified into multiple lineages (lineages 1–9, with recent proposals for lineages 10 and 11) [12, 13]. Notably, lineage 1 (L1) has emerged as the dominant epidemic lineage both globally and in China, and is further subdivided into multiple sublineages or evolutionary clusters (e.g., L1A-L1J), reflecting a complex genetic background and spatiotemporal distribution pattern [13]. In particular, the L1C sublineage has become one of the predominant circulating branches in China today [14]. However, despite global classification studies, there remains a lack of systematic and in-depth analysis of the fine-scale evolutionary history, transmission dynamics, recombination characteristics, and underlying driving factors of this key lineage, L1, within China.

This study aims to conduct a comprehensive phylogeographic and evolutionary dynamics analysis of a large collection of publicly available PRRSV-2 ORF5 sequences (totaling over 32,215 entries) spanning the period from 1990 to 2025, with a primary focus on the epidemic status of the L1 lineage in China. The research objectives include: 1) constructing an up-to-date phylogenetic tree of PRRSV-2 L1 lineage to clarify the spatiotemporal distribution and dominant populations of L1 and its sublineages within China; 2) uncovering the complex transmission pathways and key hubs of L1 strains circulating in China; 3) conducting in-depth analysis of the recombination frequency, hotspot regions and their correlations with geographical areas and time periods among L1 strains; and 4) evaluating positive selection pressure on key antigenic sites to explore their potential association with immune pressure. The findings of this research will provide critical insights into the molecular evolutionary patterns of the PRRSV-2 L1 lineage in China and offer direct scientific support for outbreak tracing, risk early warning, and the optimization of prevention and control strategies.

Results

Genetic diversity and dominant lineage distribution of PRRSV-2

Phylogenetic analysis based on 32,215 global PRRSV-2 ORF5 sequences revealed that the strains could be clearly classified into 11 major genetic lineages (lineage 1–8, lineage 9A, 9B, and 9C) (Fig. 1A). Among these, lineage 1 (L1) was the dominant epidemic lineage both globally and in China, accounting for 50.09% (11,780/32,215) of all sequences. Its circulation spanned from 1991 to 2025, with a geographical distribution covering multiple regions including North America, Asia, and Europe. Detailed information on the other lineages is provided in Tables S1 and S2. Further high-resolution phylogenetic analysis of the L1 lineage showed that it could be subdivided into 10 stable sublineages (L1A-L1J) (Fig. 1B). Detailed characteristics of each sublineage are presented in Table S3. The classification of lineages and sublineages in this study followed the methods established by Shi [12], Zeller [15] and Wannarat [13], respectively.

Among the L1 strains circulating in China (Tables S3), the L1C sublineage (represented by NADC30-like strains) predominated, with a total of 1,247 sequences detected (accounting for 86.12% of Chinese L1 strains). These strains were mainly distributed in major swine-producing provinces such as Henan, Hubei, and Shandong of China. The L1A (represented by NADC34-like strains) and L1B sublineages were represented by 191 and 10 sequences,

respectively. In contrast, the L1D–L1J sublineages were relatively rare within China. This distribution pattern is consistent with the findings reported by Tian et al[16]. The phylogenetic topology showed that the L1C strains prevalent in China formed a distinct, highly supported (bootstrap value > 90%) monophyletic clade, suggesting that this sublineage may have been introduced into China through one or more independent cross-regional transmission events, followed by local adaptation and subsequent expansion to become the dominant population.

PRRSV-2 L1 prevalence timeframe and regional statistics

To elucidate the temporal evolutionary patterns of PRRSV-2 L1 sublineages L1A, L1B, and L1C, we performed a Sankey diagram-based visualization of the annual distribution of L1A, L1B, and L1C sublineages from 1996 to 2025 (Fig. 2).

At the global level, the L1 lineage exhibited distinct region-specific evolutionary patterns. The United States, as the region with the highest diversity of the L1 sublineages, sustains high prevalence of all three sublineages-L1A, L1B, and L1C. In contrast, Asia shows a pattern of dominance by a single sublineage, with both China and South Korea being overwhelmingly dominated by the L1C sublineage, reflecting consistent regional epidemiological characteristics (Fig. 2).

Temporally, the global PRRSV L1 sublineages displayed clear succession trends. The L1B sublineage maintained a consistently high proportion from approximately 2006 to 2015, representing the dominant circulating sublineage during this period. The L1A sublineage, by contrast, persisted at moderate levels from 2001 to the present without reaching dominance. The L1C sublineage began to increase in prevalence around 2006 and rose to dominance after approximately 2011, thereafter becoming the predominant circulating sublineage in key regions including the United States, China, and South Korea (Fig. 2).

Temporal dynamics of PRRSV-2 L1 China and the United States

To explore the epidemic and evolutionary patterns of the PRRSV-2 L1 lineage across different geographical regions, we compared the annual proportional distribution of L1A, L1B, and L1C sublineages between China and the United States. The results revealed distinct evolutionary trajectories of the sublineages in the two countries, each displaying unique spatiotemporal dynamics.

The L1 lineage in China exhibits a distinct phased succession pattern from 2013 onward, L1C expanded rapidly and remained dominant until the present day (Fig. 3A). L1A, first detected in China in 2017[11], was in a phase of rapid expansion in 2020, but showed a declining trend by 2023, while L1B remained at low prevalence (< 5%) throughout (Fig. 3A). Further analysis of the spatial distribution of L1 sublineages across Chinese provinces (2020–2025), showed that L1C predominates in major pig-producing regions including Henan (HEN), Heilongjiang (HLJ), and Hebei (HEB) (Fig. 3B). The proportion of L1A is relatively low across the country, but it is relatively high in provinces such as Heilongjiang (HLJ). This is similar to the result of Jian[17]. Notably, Taiwan province exclusively harbors L1A strains, with no evidence of L1B or L1C circulation (Fig. 3B). The L1 lineage in the United States exhibits a more complex, alternating dominance pattern (Fig. 3C). The L1A sublineage held a relatively high proportion from 2002 to 2005, but was subsequently replaced by the L1B sublineage, which dominated from 2006 to 2013. Between 2010 and 2014, the L1C sublineage gradually supplanted L1B. From 2014 to 2019, the L1A sublineage experienced another large-scale epidemic wave (Fig. 3C).

Unlike the single dominant sublineage pattern observed in China, the three L1 sublineages (L1A, L1B, and L1C) in the United States have been engaged in continuous cyclical competition. There is significant spatiotemporal heterogeneity in the composition of the L1 lineage between the two countries. China is characterized by a unipolar dominance of the L1C sublineage, while the United States exhibits a co-circulation equilibrium of multiple sublineages. This difference likely reflects the disparities in swine production systems, immunological backgrounds, and historical patterns of viral introduction between the two nations. Notably, the rapid rise of the L1C sublineage in China (2013) occurred about three years later than its large-scale outbreak in the United States (2010), and the rapid expansion of the L1A sublineage in China (2020) lagged approximately six years behind its large-scale epidemic wave in the United States (2014), indicating the possible introduction events of specific strains. In comparison, the more dynamic turnover of dominant sublineages in the United States may reflect more complex animal movement patterns and a longer history of virus-host co-evolution.

The transmission routes of PRRSV-L1 in China

Phylogenetic and spatiotemporal analyses showed that the L1A sublineage was first detected in Taiwan in 2004, forming an independent evolutionary branch genetically distinct from mainland strains, with no evidence of cross-strait transmission (Fig. 4A, Supp Video 1). In mainland China, the transmission network indicated that this sublineage first emerged in Liaoning in 2017, consistent with reports by Zhang[11], subsequently establishing stable regional transmission chains among Liaoning, Jilin, and Heilongjiang, before further disseminating to Hebei, Shandong, Beijing, and other provinces. The overall pattern was characterized by radial expansion from Northeast China toward North and East China (Fig. 4A, Supp Video 1). Ancestral state reconstruction suggested that the common ancestor of Taiwanese L1A strains clustered with the U.S. strain PRRSV-2/IP50/USA/2018, whereas the common ancestor of mainland L1A strains clustered with U.S. strains PRRSV-2/USA/AP1977/2014, 13217, 13113 and 13103.

The spatiotemporal transmission network showed that the L1B sublineage exhibited a relatively sparse structure in mainland China, displaying a scattered distribution pattern with limited transmission scale and pronounced regional circulation (Fig. 4B). Li et al. [18] first reported the detection of an L1B strain (HLJDZD55) in Heilongjiang in 2023. In contrast, the present study traced its earliest occurrence to Shandong in 2015, suggesting an earlier introduction than previously recognized. Transmission pathway analysis indicated that its spread was largely confined to northern and northeastern regions, with relatively short transmission distances. Most transmission events occurred between neighboring provinces, implying that its dissemination likely depended primarily on regional swine movement. The temporal window of transmission spanned from 2015 to 2023 (Fig. 4B). Common ancestor analysis showed that mainland L1B strains clustered with U.S. strains Minnesota11A, CB1035, and LS367.

The L1C sublineage exhibited the most complex nationwide transmission network (Fig. 4C, Supp Video 2). This study traced its earliest occurrence in mainland China to Guangxi Zhuang Autonomous Region in 2006, approximately six years earlier than the previously reported date of 2012 [9]. The sublineage was subsequently detected in Hubei Province in 2011 and Henan Province in 2012. However, large-scale epidemics did not begin until after 2013, suggesting a prolonged latent adaptation period. The period from 2013 to 2024 represented a phase of rapid expansion, during which a nationwide transmission network was established, covering major swine-producing regions in Northeast, North, Central, and East China. Transmission pathway analysis identified Henan Province as the core transmission hub and Shandong Province as the secondary hub, with Fujian, Guangdong, and Sichuan Provinces serving as tertiary transmission nodes (Fig. 4C, Supp Video 2). The transmission pattern featured both multi-center radiation and long-distance dissemination. Common ancestor analysis indicated that mainland L1C strains clustered with U.S. strains PRRSV2/USA/LS1658/2014, NP337/2012, LS880/2008, and NP307/2010.

Characteristics of positive selection sites in the ORF5 protein and their association with antigenic epitopes

To explore the functional significance of adaptive evolution in the ORF5 proteins of the PRRSV-2 L1 lineage and its association with immune pressure, this study systematically analyzed the spatial distribution relationship between positive selection sites and key functional domains (B-cell epitopes, T-cell epitopes and glycosylation sites) using the SLAC, MEME, and combined SLAC & MEME models. As illustrated in Fig. 5A, the linear schematic of GP5 highlights the precise locations of glycosylation sites (blue), B-cell epitopes (green), and T-cell epitopes (orange) along the ORF5-encoded proteins, while Fig. 5B provides detailed sequence-positional information for these regions—including the exact amino acid positions and their immunological annotations (e.g., dominant neutralizing epitope, predicted MHC-I binder, and key glycan shield).

Through comparative analysis of three major sublineages—L1A (226 sequences), L1B (10 sequences), and L1C (469 sequences)—distinct patterns of antigen-driven evolution among different sublineages were revealed (Table 1). The analysis revealed that the L1A and L1C sublineages underwent strong positive selection in the ORF5 protein, whereas the L1B sublineage showed no significant positive selection signal in SLAC analysis, likely due to its limited sample size (n = 10). Notably, the L1C sublineage exhibited the most extensive positive selection pattern, with significantly more positively selected sites (MEME: 38 sites; SLAC: 28 sites) than the L1A sublineage (MEME: 19 sites; SLAC: 10 sites). The number of shared positively selected sites in L1C (27 sites) was 2.7 times that of L1A (10 sites) (Table 1).

Functional domain association analysis revealed significant spatial overlap between positively selected sites and key immune-related regions. In both the L1A and L1C sublineages, glycosylation sites represented the most prominent overlapping regions. Specifically, sites 32^G, 33^G, 34^G, 58^G, and 59^G were identified as positively selected in both sublineages, indicating conserved adaptive evolution at these sites across lineages (Table 1). Notably, the region encompassing sites 32–34 and 57–59 corresponds to the previously defined hypervariable regions 1 and 2 (HVR1 and HVR2) of the GP5 protein, which are known hotspots for genetic variation and are critical for modulating virus-host interactions [19]. The mutations observed at these sites, such as the N32D substitution in the putative HVR1 and the cluster of changes around A57N, N58K, and K59S in HVR2, as reported by [19], may directly influence antigenic properties and immune evasion. Furthermore, the L1C sublineage exhibited more extensive positive selection signals within B-cell epitopes (five sites: 27B, 29B, 30B, 41B, 189B), far exceeding the number observed in L1A (two sites) (Table 1). Additionally, the only T-cell epitope overlapping site detected in L1C, 120T, suggests that cellular immune pressure may also contribute to the evolution of this dominant sublineage (Table 1). The identification of positively selected sites within known variable regions (HVR1/HVR2) and key epitopes underscores the multifaceted strategy of PRRSV in adapting to host immune pressures, combining glycan shield modulation with direct antigenic drift. This integrated finding provides new insights for understanding the immune evasion mechanisms of PRRSV.

Table 1
Prediction of the positive selection sites for PRRSV L1A, L1B, and L1C

Dataset	Number of sequences	Length (AA)	Number of sites under positive selection			Positively selected sites MEME&SLAC
			MEME	SLAC	MEME&SLAC	
L1A	226	200	19	10	10	2,15,19,25 ^B ,30 ^B ,32 ^G ,33 ^G ,34 ^G , 58 ^G , 59 ^G
L1B	10	200	4	0	0	
L1C	469	200	38	28	27	2,3,6,11,13,14,15,16,17,19,25,26,27 ^B ,29 ^B ,30 ^B ,32 ^G ,33 ^G ,34 ^G ,35 ^G ,41 ^B ,58 ^G ,59 ^G ,102,104,112,189 ^B
Note: In this column, superscript B indicates that the site is located in a B-cell antigenic region; superscript G denotes a potential glycosylation site; and sup T signifies that the site is situated within a T-cell antigenic region.						

Spatiotemporal dynamics and genomic distribution patterns of recombination events

To elucidate the recombination characteristics of PRRSV-2 L1 lineage strains circulating in China, this study systematically analyzed 251 complete genome sequences of the L1 lineage from China. After quality control screening, 218 sequences with complete year information and 172 sequences with complete province information were included in the analysis. The results demonstrated that recombination is an important driver of L1 lineage evolution in China, with its frequency exhibiting significant spatiotemporal heterogeneity and being closely associated with specific regions of the viral genome (Fig. 6).

Specifically, temporal analysis revealed no clear annual pattern in the frequency of recombination events (Fig. 6A). The non-chronological arrangement of the time axis on the right may be attributed to variations in sampling intensity across different years, or it may reflect the inherent fluctuation of PRRSV

epidemics. This uneven yet occasionally clustered temporal distribution pattern is likely associated with the complex interplay of factors such as adjustments in vaccine immunization strategies, and seasonal peaks in swine movement.

Spatial analysis revealed significant geographical heterogeneity in the distribution of recombination signals (Fig. 6B). Recombination hotspots were localized in non-structural protein-coding regions (nsp4–nsp10) in Gansu (GS), Chongqing (CQ), Beijing (BJ), and Shandong (SD) provinces, while in Tianjin (TJ), Hebei (HEB), and Henan (HEN) provinces, they clustered in structural protein-coding regions (ORF2a–ORF6). While recombination signals were also detected in Jilin (JL), the limited sample size from this province hinders its reliable classification within this framework. This pattern may reflect differences in viral population structure and transmission intensity: non-structural regions (nsps) harbor core replication machinery, and their recombination may facilitate rapid adaptation under high transmission pressure (e.g., in densely farmed areas like SD); structural regions (ORFs 2a–6), involved in immune evasion and host interaction, may experience more recombination in regions with complex viral diversity (e.g., HEN, a major pig movement hub), potentially driven by immune selection or co-infection dynamics.

The distribution of recombination breakpoints across the PRRSV-2 genome exhibits marked regional preferences (Fig. 6C and 6D), consistent with previous large-scale genomic analyses. Notably, recombination hotspots are concentrated in specific regions, including spanning nsp1 to nsp2, nsp3 to nsp4, and nsp9 within the non-structural protein-coding regions, as reported by Yuan et al. [9, 20]. Our analysis corroborates the prominence of nsp2 and nsp9 as clear hotspots, aligning with the known high variability of these regions and their critical functional roles in the viral replication complex [5]. Within the structural protein-coding region, Yuan et al. identified a major hotspot spanning ORF2 to ORF6. While this broad interval includes the ORF2-ORF4 region highlighted by Cui et al. [21], our fine-scale analysis further pinpoints the ORF5 (GP5) and ORF6 (M) genes as particularly prominent and frequent sites of recombination.

Discussion

This study reveals a distinct epidemic pattern in China, contrasting with the “dynamic equilibrium” of long-term co-circulation among multiple sublineages (L1A, L1B, and L1C) observed in North America. China exhibits a scenario dominated by a single sublineage L1C, and this marked divergence may stem from the confluence of virological and husbandry factors. First, specific introduction and establishment events were pivotal. While the NADC30-like strains (L1C) were first reported in China around 2013 [9], our phylogenetic analysis identified their earliest presence in Guangxi (GX) as early as 2006, with subsequent detections in Hubei (HB) in 2011. This suggests a period of localized circulation and adaptation prior to widespread emergence. The introduced L1C sublineage likely possessed a fitness advantage that allowed it to rapidly outcompete local strains, consistent with a founder effect in viral population expansion. Second, and critically, China's intensive, centralized swine production system, characterized by frequent and large-scale inter-regional animal movements, created ideal conditions for the rapid nationwide dissemination of this successfully established strain. This high-connectivity network likely facilitated the sweep of L1C, simultaneously limiting the ecological space available for competing sublineages to persist and spread [14], thereby establishing and maintaining its unipolar dominance.

The spatiotemporal analyses revealed marked heterogeneity in the transmission hubs of PRRSV-2 L1 sublineages in mainland China, which is consistent with previous epidemiological studies. For L1A, the early concentration of spread in Northeast China (Heilongjiang, Liaoning, and Jilin) agrees with earlier reports that NADC34-like PRRSV first emerged in Liaoning in 2017 and was subsequently detected in Heilongjiang and neighboring northeastern provinces [11, 22]. The appearance of Henan in the L1A network more plausibly reflects its role as a national swine-transport transit node than a primary source of L1A introduction. The distinct evolutionary status of Taiwanese L1A strains is also in line with previous observations. For L1C, the present study defines a hub structure with Henan as the core and Shandong as the secondary hub. This is consistent with the established view that NADC30-like PRRSV was first isolated in Henan in 2012 and later disseminated nationwide [23]. The secondary status of Shandong further highlights the contribution of northern swine-logistics networks to L1C spread. Fujian, Guangdong, and Sichuan act as lower-tier nodes, whereas Heilongjiang's activity in the L1C network more likely reflects intensive sampling and northern connectivity than a founding-hub role. Taken together, different L1 sublineages are anchored by different geographic hubs. Recognizing this heterogeneity avoids treating any single province as a universal “national hub” for PRRSV-2 and provides a finer framework for understanding how swine-production scale and live-pig movement shape viral dispersal.

From a molecular evolutionary perspective, positive selection analysis provides a key explanation for the immune evasion mechanisms underlying the dominance of the L1C sublineage. This study found that positively selected sites in the ORF5 proteins of the L1C lineage are significantly enriched in known B-cell epitopes and glycosylation sites [24]. Particularly noteworthy, site 30^B and key glycosylation sites 34^G32^G33^G and 59^G are under strong selection in both the L1A and L1C lineages, indicating their pivotal functional and adaptive roles. Adaptive mutations at these sites may aid viral immune escape through two complementary mechanisms: first, direct variation in B-cell epitope regions can alter antigenicity, thereby diminishing the efficacy of pre-existing host immunity [25]; second, variation in glycosylation sites may reshape the surface “glycan shield”, physically masking underlying antigenic epitopes and interfering with antibody recognition and binding [26]. These findings align with studies by Ostrowski [25] and Das [26], collectively indicating that vaccine-induced immune pressure is a significant selective force driving the adaptive evolution of circulating strains. The more extensive and intense adaptive variations accumulated by the L1C lineage at these key immune-related sites likely conferred stronger survival advantages. Consequently, these persistently positively selected “hotspot” sites, especially those relatively unique to the L1C lineage (e.g., 35^G27^B41^B120^T189^B29^B, etc.), should be regarded as core targets for future vaccine strain selection and molecular surveillance of epidemic strains.

Within the dominant L1C sublineage in China, previous studies based on complete genome analyses have identified recombination hotspots concentrated in the hypervariable nsp2 region and the antigenic ORF5 gene [5]. The frequent recombination within ORF5, which encodes the major neutralizing antibody target, is of high concern due to its potential for “modular exchange” of antigenic epitopes, posing a continuous threat to vaccine efficacy. Spatially, recombination events reported in the literature exhibit significant geographical aggregation, overlapping with key transmission network hubs such as Henan (HEN) and Shandong (SD). This spatial congruence strongly implicates frequent animal movement in these areas as a primary facilitator of viral co-circulation, thereby creating an ideal environment for recombination. Major livestock transportation hubs (e.g., Beijing (BJ), Tianjin (TJ), Chongqing (CQ)

further amplified this signal, underscoring their role as critical “blending centers” despite lower local sequencing depth. Collectively, these findings support a “transmission-mixing-recombination-dissemination” cycle as a core mechanism sustaining the rapid evolution and strain turnover of PRRSV-2 in China’s intensive farming system. Therefore, future control strategies should be dual-focused: (1) enhanced surveillance targeting both identified genomic hotspots (nsp2, ORF5) and key geographical hotspots/provinces; and (2) implementation of stringent biosecurity and movement restrictions at major hubs to physically limit viral mixing and interrupt this evolutionary cycle.

Conclusion

This study delineates the driving forces behind the epidemic of PRRSV-2 lineage 1 in China. The currently dominant L1C sublineage spread nationwide via a transmission network centered on Henan (HEN) Province. Its rapid adaptation is driven by two key molecular mechanisms: frequent recombination in replication-related (nsp2) and key antigenic (ORF5) genes, and strong positive selection on neutralizing epitopes and glycosylation sites of the GP5 protein. These findings reveal a synergistic “transmission-recombination-antigenic evolution” cycle for PRRSV-2 within China’s intensive swine industry, providing a direct scientific basis for targeted surveillance of key genomic regions (e.g., ORF5, nsp2) and control measures in hotspot provinces.

Methods

Data collection and preprocessing

This study retrieved 32,215 global PRRSV ORF5 gene sequences from the NCBI database. To supplement sequence metadata such as sampling location and collection date, an automated Python script was developed. The script utilizes the European Bioinformatics Institute (EBI) ENA portal API to perform multi-threaded concurrent requests, extracting the country and collection date for each sequence. The retrieved data were parsed and stored in a structured format for subsequent phylogeographic analysis. Sequence filtering criteria were as follows: (1) sequences lacking collection dates or geographical information were removed (isolates from China were required to include at least provincial-level information); (2) unverified sequences, and laboratory-passaged strains were excluded; and (3) sequences containing ambiguous or missing residues were discarded.

Phylogenetic analysis

The filtered ORF5 sequences were manually trimmed to remove non-coding regions outside the ORF5 encoding gene using the Geneious Alignment plug-in within Geneious Prime software (v2022.1). Multiple sequence alignment was performed with MAFFT, followed by trimming of low-confidence regions using the TrimAL algorithm (v1.4). A maximum likelihood (ML) phylogenetic tree was constructed with IQ-TREE (v2.1.2), and node support was assessed with 1,000 bootstrap replicates. To optimize the visualization of genetic diversity, a stratified subsampling strategy was applied to simplify the original phylogenetic tree. Analyses were performed in R (v4.1.3) using packages including ape, ggtree, and dplyr were used to perform logarithmic scaling of sample sizes, thereby balancing the representation of dominant and rare lineages.

Construction of Sankey diagrams

To visualize the spatiotemporal flow patterns of L1 sublineages, a Sankey diagram was generated using the online tool SankeyMATIC (<https://sankeymatic.com/>). The input data were preprocessed by categorizing sequences based on their collection year and sublineage assignment (as determined in the phylogenetic analysis above).

Identification of Chinese L1 strains from the phylogenetic tree

A custom Python script was developed to automatically extract and compile the identifiers of all PRRSV-2 strains belonging to lineage 1 (L1) and originating from China. The script parsed the Newick-format file of the global maximum likelihood phylogenetic tree. It traversed the tree, identified leaf nodes (individual strain sequences), and filtered them based on two criteria: 1) the sequence name or associated metadata (extracted from the sequence label, typically in the format Accession|Country|Date) indicated the country of origin as “China”; and 2) the sequence was located within the L1 clade as defined by the global phylogenetic analysis (Section 2.2). The final output was a list of all Chinese L1 strain accession numbers, which served as the core dataset for all subsequent China-focused spatiotemporal and evolutionary analyses.

Geographic distribution and statistical analysis

The geographic distribution of PRRSV-2 L1 strains across China was analyzed and visualized using a custom Python program. The analysis was based on the list of Chinese L1 strains identified in Section 2.4 and their associated provincial data. The program utilized libraries including cartopy, matplotlib, pandas, and shapely to create high-quality thematic maps. Provincial boundaries of China were loaded as shapefiles. The geographic coordinates (longitude, latitude) of the centroid of each province, where L1 strains detected, were mapped. To analyze the annual distribution differences of sublineages (L1A-L1C) between China and the United States, statistical analyses were conducted in Excel. The annual proportion of sequences belonging to each sublineage was calculated for comparison.

Transmission route inference

Bayesian phylogenetic and phylogeographic analyses were performed separately for the PRRSV-2 L1A and L1C sublineages, which had sufficient sequence coverage. The ORF5 alignment was analyzed as a single nucleotide partition under the GTR + Γ substitution model with four discrete gamma rate categories (GTR + G); base frequencies were set to "Estimated". An uncorrelated lognormal relaxed clock model was used to accommodate branch rate variation, and the tree prior was set to the Bayesian Skyline coalescent model. For discrete phylogeographic inference, the sampling province of each sequence was treated as a discrete trait and analyzed using a continuous-time Markov chain (CTMC) model; non-zero transition rates between locations were assigned a Uniform[0.0, ∞] prior.

MCMC chains were run for 10 million generations, sampling every 10,000 generations. The initial 50% (5 million generations) was discarded as burn-in after inspecting trace plots and effective sample size (ESS) values in Tracer v1.7.2; ESS values for all parameters exceeded 200. Maximum clade credibility (MCC) trees were generated for each sublineage using TreeAnnotator with the same 50% burn-in. The resulting annotated MCC trees were then used as input for Spread3 v0.9.6 to visualize the inferred spatiotemporal diffusion. Finally, the inferred transmission network was rendered onto a geographic basemap of China using code written in VS Code and launched via the Live Server plugin, with province nodes marked as circular symbols and transmission directions indicated by directed arrows.

For the L1B sublineage, Bayesian phylogeographic analysis was not performed because of the limited number of available sequences. Instead, its transmission route was manually reconstructed based on the phylogenetic tree topology and sampling years.

Positive selection detection

The Datamonkey online server (<https://datamonkey.org/>) was used to separately run the MEME and SLAC algorithms. MEME was employed to detect branch-specific signals of positive selection, while SLAC was applied to evaluate lineage-wide pervasive selective pressure. Subsequently, sites that showed statistical significance ($p\text{-value} \leq 0.05$) in both methods were extracted. The intersection of the results from the two approaches was taken as a set of high-confidence candidate positively selected sites, thereby cross-validating the findings and ensuring the robustness of the results.

Protein function prediction

To predict the functional domains of the ORF5-encoded protein, a comprehensive analysis was performed using the following bioinformatic tools and experimental evidence. N-linked glycosylation pattern prediction: Potential N-linked glycosylation sites were predicted using the NetNGlyc 1.0 server with a threshold of > 0.5 . Identification of B-cell antigenic regions: Possible B-cell epitope regions were determined based on experimental data published by Kim HK and Ostrowski M et al [24, 25]. Identification of T-cell antigenic regions: Potential T-cell epitope regions were identified according to research findings from Vashisht K et al [27].

Recombination analysis

Recombination events were systematically scanned using RDP4 software (v4.101), which integrates seven detection algorithms including RDP, GENECONV and BootScan. To enhance the reliability of the findings, only predicted recombination events supported by at least three methods with a statistical significance threshold of $p < 1 \times 10^{-6}$ were retained. The filtered recombination data were subsequently analyzed using the Pandas library in Python to characterize recombination breakpoint distributions and to quantify the contribution proportion of each parental strain. For visual representation of the recombination patterns, the Hplot platform (<https://hiplot.com.cn/>) was employed to generate a recombination heatmap, a breakpoint distribution plot, and a phylogenetic network based on recombination events, thereby illustrating recombination hotspot regions and the genetic exchange features among parental strains.

Declarations

Consent to Publish declaration

not applicable

Ethical approval

The data used in this study were sourced from open-access repositories (NCBI, EBI), whose terms of use explicitly permit unrestricted reuse for non-commercial academic research. Therefore, no additional permissions from the original data submitters were required. This study is a retrospective analysis of publicly available genome sequence data, and no primary data collection from human participants or animals was conducted. Therefore, in accordance with the research ethics guidelines of Yangzhou University, this type of research is exempt from ethical review approval by an institutional review board (IRB) committee of Yangzhou University.

Author Contribution

C.C, N.C and J.Z conceived and designed the experiments; C.C, P.Z, S.S and W.Z collected the data; C.C and N.Y performed the systemic analysis; C.C, N.C and J.Z wrote the paper. All authors contributed to the article and approved the submitted version.

Acknowledgment

not applicable

Data Availability

The supplementary raw data and codes of this study are all available online: <https://doi.org/10.6084/m9.figshare.31923054.v1>

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding Declaration

This work was partly supported by the National Natural Science Foundation of China (32473040; 32673871), the 111 Project under Grant D18007, and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). C.C was funded by Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX25_2382).

Author contribution statement

C.C, N.C and J.Z conceived and designed the experiments; C.C, P.Z, S.S and W.Z collected the data; C.C and N.Y performed the systemic analysis; C.C, N.C and J.Z wrote the paper. All authors contributed to the article and approved the submitted version.

Data availability statement

The supplementary raw data and codes of this study are all available online: <https://doi.org/10.6084/m9.figshare.31923054.v1> [28].

Declaration of generative AI use

Authors declare no use of generative AI tools in the manuscript preparation process upon submission of the paper.

References

1. Vu Manh T-P, Frias-De-Diego A, Williams A, Byrne J, Sirisereewan C, Hicks J, Liu H-C, Crisci E. Mitochondrial dysfunction in PRRSV-2 infected macrophages. *Front Immunol.* 2025;16:1670488.
2. Jiao S, Zhang J, Wang J, Ma X, Li G, Li J, Cui Z, Li D, Li P, Zeng Q, et al. Whole-genome analysis of the recombination and evolution of newly identified NADC30-like porcine reproductive and respiratory syndrome virus strains circulated in Gansu province of China in 2023. *Front Vet Sci.* 2024;11:1372032.
3. Wensvoort G, Terpstra C, Pol J, Ter Laak E, Bloemraad M, De Kluiver E, Kragten C, Van Ld B, Den Besten A, Wagenaar F. Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Veterinary Q.* 1991;13(3):121–30.
4. Collins JE, Benfield DA, Christianson WT, Harris L, Hennings JC, Shaw DP, Goyal SM, McCullough S, Morrison RB, Joo HS. Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs. *J Vet Diagn Invest.* 1992;4(2):117–26.
5. Murtaugh MP, Stadejek T, Abrahante JE, Lam TT, Leung FC. The ever-expanding diversity of porcine reproductive and respiratory syndrome virus. *Virus Res.* 2010;154(1–2):18–30.
6. Zhou L, Han J, Yang H. The evolution and diversity of porcine reproductive and respiratory syndrome virus in China. *Vet Microbiol.* 2024;298:110252.
7. Tian K, Yu X, Zhao T, Feng Y, Cao Z, Wang C, Hu Y, Chen X, Hu D, Tian X. Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. *PLoS ONE.* 2007;2(6):e526.
8. Zhou YJ, Hao XF, Tian ZJ, Tong GZ, Yoo D, An TQ, Zhou T, Li GX, Qiu HJ, Wei TC. Highly virulent porcine reproductive and respiratory syndrome virus emerged in China. *Transbound Emerg Dis.* 2008;55(3–4):152–64.
9. Zhou L, Wang Z, Ding Y, Ge X, Guo X, Yang H. NADC30-like strain of porcine reproductive and respiratory syndrome virus, China. *Emerg Infect Dis.* 2015;21(12):2256.
10. Xu H, Li C, Li W, Zhao J, Gong B, Sun Q, Tang YD, Xiang L, Leng C, Peng J. Novel characteristics of Chinese NADC34-like PRRSV during 2020–2021. *Transbound Emerg Dis.* 2022;69(5):e3215–24.
11. Zhang H-L, Zhang W-L, Xiang L-R, Leng C-L, Tian Z-J, Tang Y-D, Cai X-H. Emergence of novel porcine reproductive and respiratory syndrome viruses (ORF5 RFLP 1-7-4 viruses) in China. *Vet Microbiol.* 2018;222:105–8.
12. Shi M, Lam TT-Y, Hon C-C, Murtaugh MP, Davies PR, Hui RK-H, Li J, Wong LT-W, Yip C-W, Jiang J-W. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. *J Virol.* 2010;84(17):8700–11.
13. Yim-Im W, Anderson TK, Paploski IA, VanderWaal K, Gauger P, Krueger K, Shi M, Main R, Zhang J. Refining PRRSV-2 genetic classification based on global ORF5 sequences and investigation of their geographic distributions and temporal changes. *Microbiol Spectr.* 2023;11(6):e02916–02923.

14. Guo Z, Chen X-x, Li R, Qiao S, Zhang G. The prevalent status and genetic diversity of porcine reproductive and respiratory syndrome virus in China: a molecular epidemiological perspective. *Viol J.* 2018;15(1):2.
15. Zeller MA, Chang J, Trevisan G, Main RG, Gauger PC, Zhang J. Rapid PRRSV-2 ORF5-based lineage classification using Nextclade. *Front Veterinary Sci.* 2024;11:1419340.
16. Tian X, Wei Z, Khan M, Zhou Z, Zhang J, Huang X, Yang Y, Wang S, Wang H, Cai X, et al. Refining Lineage Classification and Updated RFLP Patterns of PRRSV-2 Revealed Viral Spatiotemporal Distribution Characteristics in China in 1991–2023. *Transbound Emerg Dis.* 2025;2025:9977088.
17. Jian Y, Lu C, Shi Y, Kong X, Song J, Wang J. Genetic evolution analysis of PRRSV ORF5 gene in five provinces of Northern China in 2024. *BMC Vet Res.* 2025;21(1):242.
18. Li J, Xu H, Li C, Zhao J, Gong B, Sun Q, Guo Z, Zhang S, Zhang M, Xiang L et al. Genomic Characterization of HLJDZD55: The First L1B PRRSV in China. *Transbound Emerg Dis* 2024, 2024:2969771.
19. Zhao J, Xu Z, Xu T, Zhou Y, Li J, Deng H, Li F, Xu L, Sun X, Zhu L. Molecular Characterization of the Nsp2 and ORF5s of PRRSV Strains in Sichuan China during 2012–2020. *Animals: an open access journal from MDPI* 2022, 12(23).
20. Yuan N, Yang Z, Lv F, Dou L, Li X, Zhao B, Dong S. Molecular Epidemiology and Genetic Evolution of Porcine Reproductive and Respiratory Syndrome Virus in Northern China During 2021–2023. *Viruses* 2025, 17(1).
21. Cui X, Xia D, Huang X, Sun Y, Shi M, Zhang J, Li G, Yang Y, Wang H, Cai X, et al. 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.* 2022;10(5):e0293422.
22. Xu H, Song S, Zhao J, Leng C, Fu J, Li C, Tang YD, Xiang L, Peng J, Wang Q, et al. A potential endemic strain in China: NADC34-like porcine reproductive and respiratory syndrome virus. *Transbound Emerg Dis.* 2020;67(4):1730–8.
23. Zhang H, Leng C, Ding Y, Zhai H, Li Z, Xiang L, Zhang W, Liu C, Li M, Chen J, et al. Characterization of newly emerged NADC30-like strains of porcine reproductive and respiratory syndrome virus in China. *Arch Virol.* 2019;164(2):401–11.
24. Kim HK, Yang JS, Moon HJ, Park SJ, Luo Y, Lee CS, Song DS, Kang BK, Ann SK, Jun CH, et al. Genetic analysis of ORF5 of recent Korean porcine reproductive and respiratory syndrome viruses (PRRSVs) in viremic sera collected from MLV-vaccinating or non-vaccinating farms. *J Vet Sci.* 2009;10(2):121–30.
25. Ostrowski M, Galeota JA, Jar AM, Platt KB, Osorio FA, Lopez OJ. Identification of neutralizing and nonneutralizing epitopes in the porcine reproductive and respiratory syndrome virus GP5 ectodomain. *J Virol.* 2002;76(9):4241–50.
26. Das PB, Vu HL, Dinh PX, Cooney JL, Kwon B, Osorio FA, Pattnaik AK. Glycosylation of minor envelope glycoproteins of porcine reproductive and respiratory syndrome virus in infectious virus recovery, receptor interaction, and immune response. *Virology.* 2011;410(2):385–94.
27. Vashisht K, Goldberg TL, Husmann RJ, Schnitzlein W, Zuckermann FA. Identification of immunodominant T-cell epitopes present in glycoprotein 5 of the North American genotype of porcine reproductive and respiratory syndrome virus. *Vaccine.* 2008;26(36):4747–53.
28. Genetic diversity. spatiotemporal dynamics and evolution of porcine reproductive and respiratory syndrome virus 2 (PRRSV-2) lineage 1 in China.

Figures

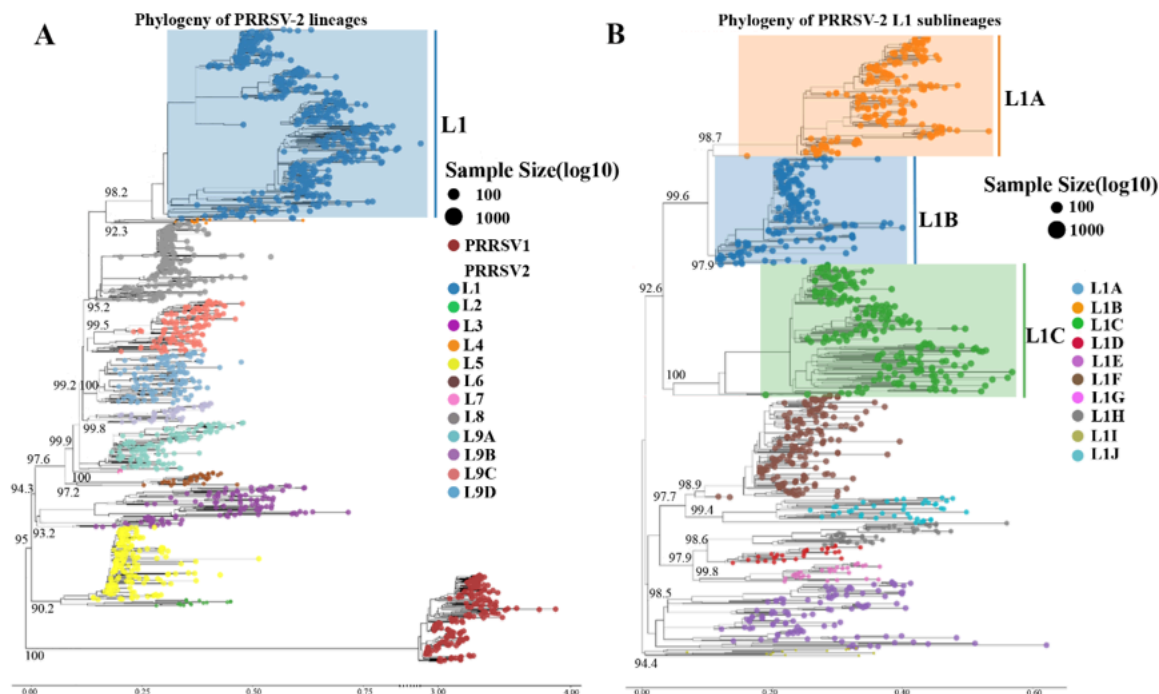


Figure 1

(A) Phylogenetic analysis and lineage classification of global PRRSV based on ORF5 sequences, with the L1 lineage shaded in light blue. (B) Phylogenetic analysis and sublineage classification of global PRRSV-2 lineage 1 (L1) based on ORF5 sequences, with the currently prevalent L1A (orange), L1B (light blue), and L1C (green) sublineages highlighted. These correspond to the major circulating strains in China.

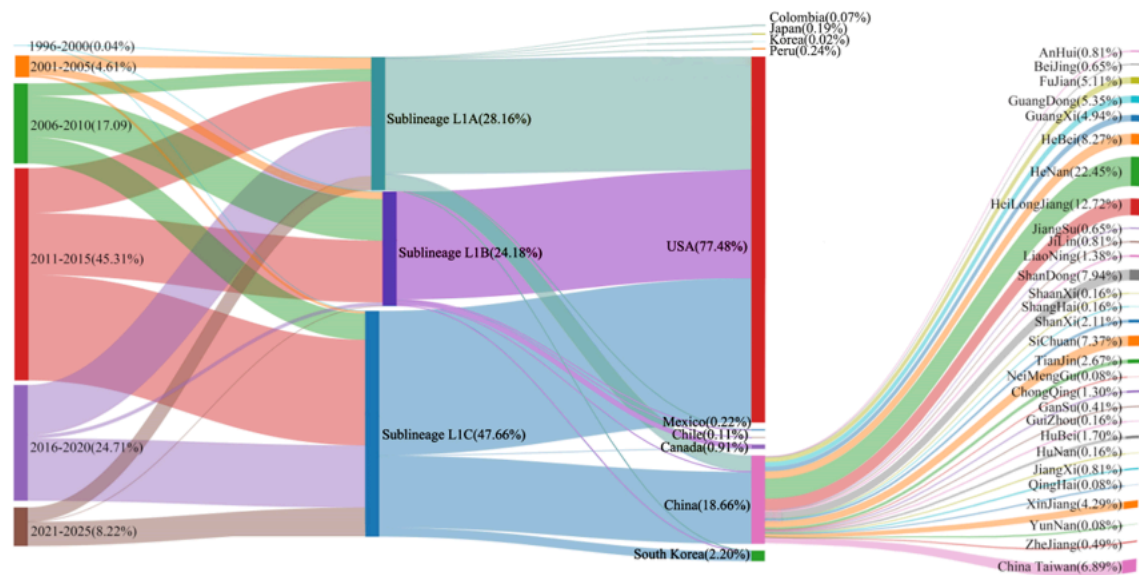


Figure 2
Annual distribution and succession Sankey diagram of global PRRSV-2 L1 sublineages A, B and C (2006-2025).
The left axis indicates the sequence collection year; the central axis represents the L1 representative sublineages (L1A, L1B, L1C); and the right axis indicates the countries or regions where the strains were identified. The detailed regions of China were shown in the most right.

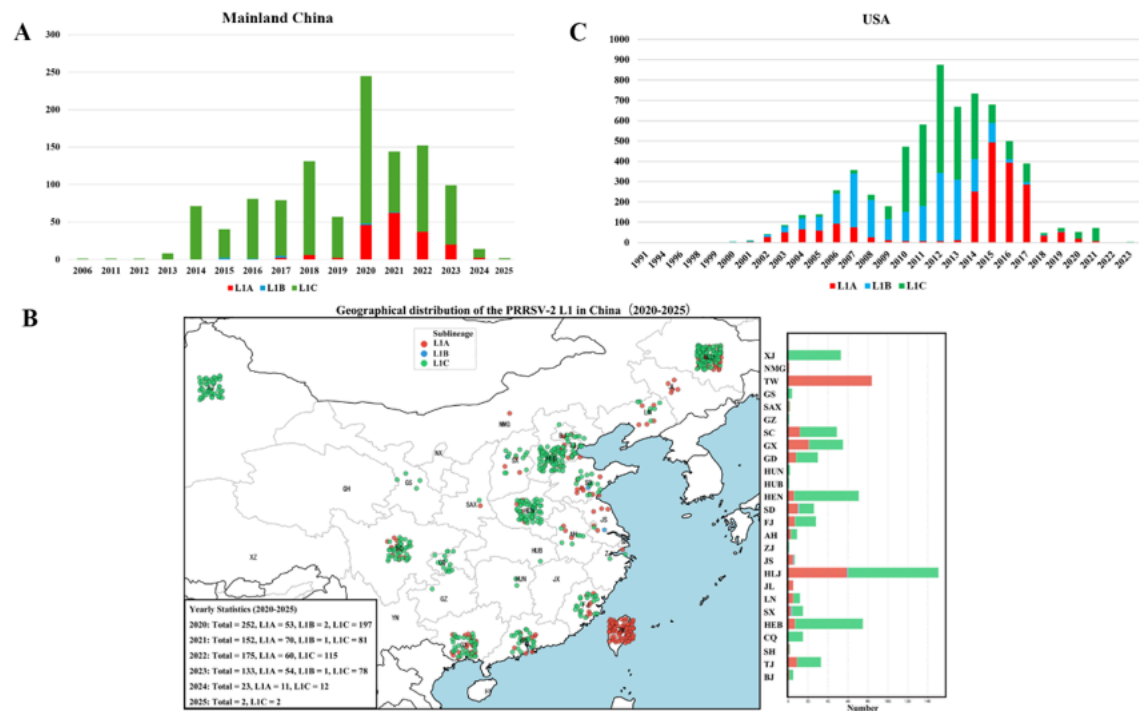


Figure 3
Temporal dynamics and geographical distribution of PRRSV-2 L1 sublineages (L1A, L1B, L1C) in China and the United States.
(A) Distribution proportions of L1A, L1B, and L1C sublineages across different years in China. (B) This map illustrates the distribution of PRRSV-L1 in each province of China from 2020 to 2025, while the adjacent bar chart shows the relative proportion of each sub-lineage in each province. (C) Distribution

proportions of L1A, L1B, and L1C sublineages across different years in the United States. The X-axis represents sampling year, and the Y-axis indicates the number of strains.

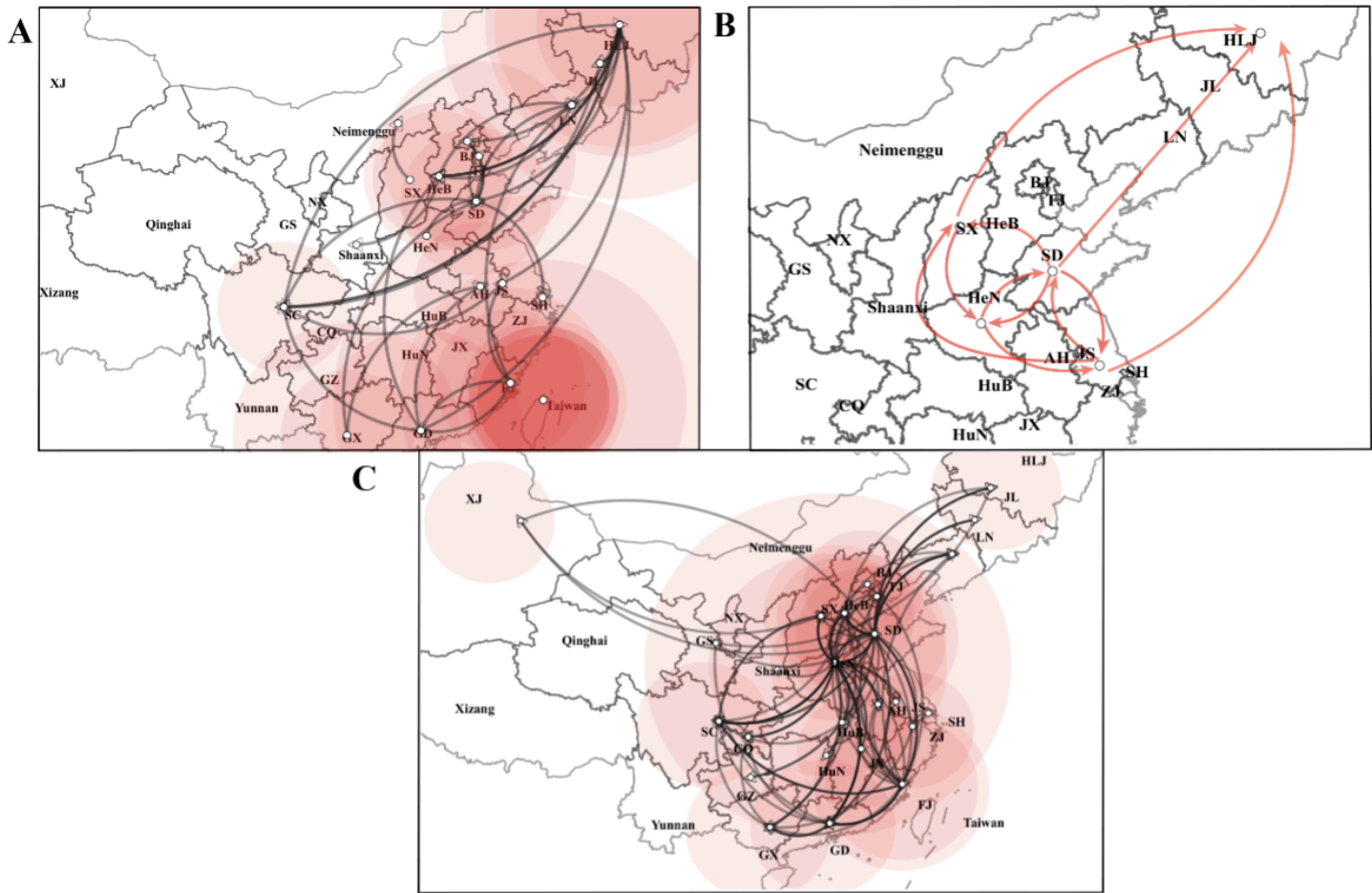
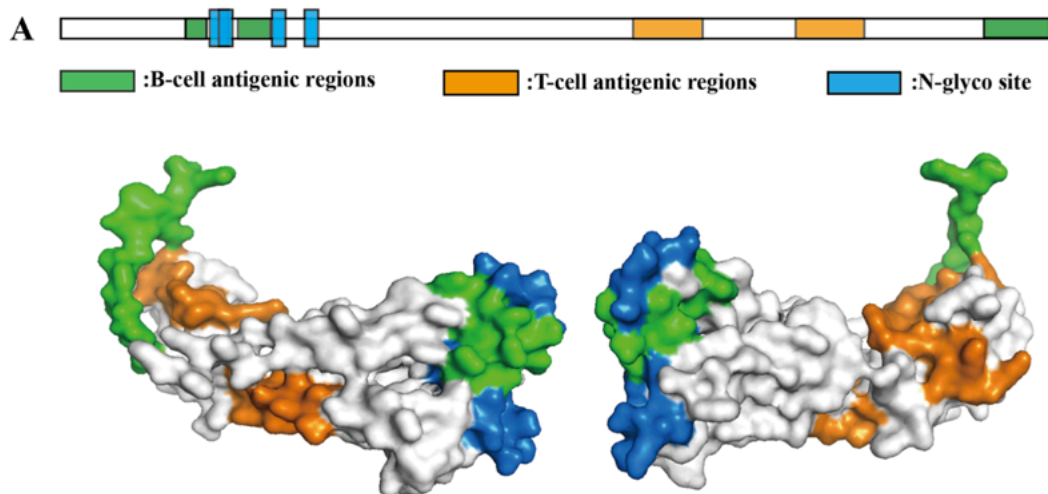


Figure 4

Phylogenetic tree and inter-provincial transmission network of major PRRSV-2 L1 sublineages in China.

The inter-provincial transmission networks of L1A (**A**) and L1C (**C**) were inferred using Bayesian phylogeographic analysis based on phylogenetic, geographic, and temporal data. The L1B network (**B**) was manually reconstructed based on phylogenetic tree topology and sampling year, due to the limited number of available sequences. Arrows indicate the direction of transmission.



B

Feature / Region	Position	Potential	Notes
N-glycol site	32-35	NNSS	Key glycan shield
	33-36	NSSS	
	44-47	NLTI	
	51-54	NGTD	
	57-60	NKSF	
B-cell antigenic regions	27-30	ALVN	Dominant neutralizing epitope
	37-45	SHLQLIYNL	
	187-200	TPVTKVSAEQWGRP	
T-cell antigenic regions	117-131	LAALTCFVIRLTKNC	Predicted MHC-I binder
	149-163	KGKLYRWRSPVIEK	

Figure 5

Mapping of glycosylation sites, B-cell epitopes, and predicted T-cell epitopes in the GP5 proteins of PRRSV-2 L1 lineage in China.

(A) Schematic representation of glycosylation sites, B-cell epitopes, and T-cell epitopes in ORF5. (B) Sequence and positional information of glycosylation sites, B-cell epitopes, and T-cell epitopes. Note: B-cell epitopes were annotated based on reference [24, 25]; T-cell epitopes were predicted according to reference [27]; N-glycosylation sites were predicted using the NetNGlyc 1.0 server (<https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/>).

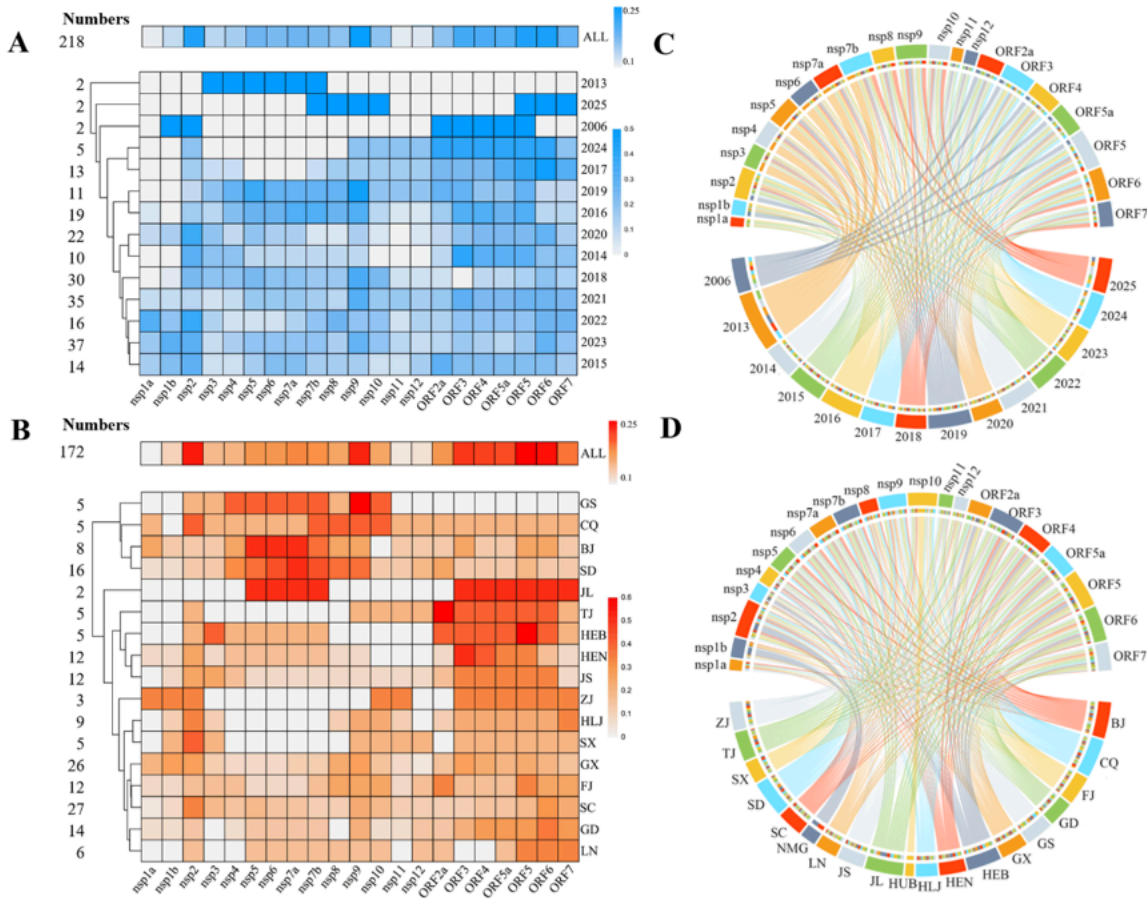


Figure 6

Spatiotemporal distribution of recombination events in the PRRSV-2 L1 lineages of China.

(A) Annual distribution heatmap of recombination events. (B) Provincial distribution heatmap of recombination events. (C) Annual distribution spectrum of structural and non-structural proteins in PRRSV. (D) Provincial distribution spectrum of structural and non-structural proteins in PRRSV. Information from year and province groups with a count of 1 was excluded.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SuppVideo1.L1Atransmissionroutes.mp4](#)
- [SuppVideo2.L1Ctransmissionroutes.mp4](#)
- [TableS123.docx](#)