

Reassessment of PRRSV-2 ORF5 sequences assigned to Lineage 4 in Japan identifies a previously unrecognized candidate lineage-level clade

Osamu Taira

o.taira@nibs.or.jp

General Incorporated Foundation Nippon Institute for Biological Science

Ryosuke Takai

General Incorporated Foundation Nippon Institute for Biological Science

Yoriko Yonezawa

Nisseiken Co

Nobuyuki Tsutsumi

General Incorporated Foundation Nippon Institute for Biological Science

Katsuaki Sugiura

General Incorporated Foundation Nippon Institute for Biological Science

Research Article

Keywords:

Posted Date: April 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9175809/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Archives of Virology on July 23rd, 2026.
See the published version at <https://doi.org/10.1007/s00705-026-06696-0>.

Abstract

Porcine reproductive and respiratory syndrome virus type 2 (PRRSV-2) exhibits extensive genetic diversity, and ORF5-based classification provides a standardized framework for global molecular surveillance. In a previous nationwide study in Japan, lineage 4 (L4) was identified as a dominant lineage exhibiting unexpectedly high genetic diversity. This observation motivated a focused reassessment of L4-assigned viruses to determine whether this heterogeneity reflected within-lineage diversification or deeper phylogenetic separation.

In this study, we re-evaluated PRRSV-2 ORF5 sequences previously assigned to L4 in Japan using maximum-likelihood phylogenetic reconstruction and pairwise nucleotide distance analyses consistent with the global ORF5 classification framework. Following sequence de-duplication, 321 L4-assigned ORF5 sequences were analyzed together with global reference datasets. Phylogenetic analysis revealed two divergent clades within the L4-assigned population. One clade contained the global L4 reference sequences, whereas the other formed a distinct monophyletic group composed predominantly of Japanese sequences and supported by bootstrap values $\geq 85\%$ at its defining node. Pairwise p-distance analysis showed that divergence between the two clades exceeded the 11% ORF5 lineage demarcation threshold (median between-clade p-distance = 0.133).

Based on combined phylogenetic and genetic distance evidence, we report this distinct clade, provisionally designated JPL-A, as a candidate lineage-level clade separate from the major L4 clade. These findings indicate that contemporary PRRSV-2 diversity in Japan includes a lineage-level group not fully represented in historical global references and highlight the value of regional surveillance data in refining standardized ORF5-based classification resources.

1. Introduction

Porcine reproductive and respiratory syndrome virus type 2 (PRRSV-2) is a major cause of economic losses in global swine production, and its rapid evolutionary dynamics have made molecular surveillance an essential component of control programs [1]. ORF5-based phylogenetic classification has become a cornerstone of PRRSV-2 molecular epidemiology because it provides a standardized framework for comparative surveillance across regions and time [2, 3].

Using a standardized ORF5-based lineage framework, we previously conducted a nationwide molecular epidemiological analysis of PRRSV-2 in Japan [4]. That study demonstrated co-circulation of multiple lineages and identified Lineage 4 (L4) as a prominent lineage in the domestic PRRSV-2 population. Importantly, the Japanese L4 population exhibited high genetic diversity comparable to Lineage 1 (L1), suggesting continuous local evolution and the potential existence of distinct internal substructures [4]. In other settings, fine-scale classification within major lineages (e.g., L1) has improved epidemiological resolution, including sequential dominance and temporal turnover of co-circulating sub-lineages [5, 6]. The unexpectedly high genetic diversity observed within the Japanese L4-assigned population motivated

a focused re-evaluation of L4-assigned viruses to determine whether the heterogeneity reflects within-lineage structure or deeper phylogenetic separation.

In our previous nationwide study, lineage assignment was performed using a standardized ORF5-based reference framework implemented in Nextclade [7]. While such reference-based tools enable rapid and reproducible classification, their outputs depend on the composition and temporal breadth of the reference datasets that anchor lineage definitions. In the global ORF5 classification framework, L4 is represented primarily by sequences from 1992–2010 [3], whereas Japanese sequences detected during 2020–2023 may include diversity not fully represented in historical references. Under such circumstances, divergent clades can be provisionally assigned to the closest available lineage label, motivating *de novo* phylogenetic re-evaluation.

Accordingly, the objective of the present study was to re-evaluate PRRSV-2 ORF5 sequences assigned to L4 in Japan using a phylogeny- and distance-based approach consistent with the global ORF5 classification framework. Specifically, we aimed to (i) determine whether the L4-assigned Japanese population resolves into multiple reproducible major clades under maximum-likelihood inference, and (ii) quantify genetic separation among these clades using pairwise nucleotide distance patterns. Because no L4-derived live vaccine viruses are used in Japan, interpretation of L4-assigned diversity in this setting is not confounded by lineage-specific vaccine-like clusters. This study provides an ORF5-based framework to report and track contemporary PRRSV-2 diversity in Japan as a direct extension of our previous nationwide analysis.

2. Materials and methods

2.1. Sequence datasets

ORF5 nucleotide sequences of PRRSV-2 L4 were compiled from three sources. The primary dataset comprised 536 ORF5 sequences classified as L4 in our previous nationwide molecular epidemiological study in Japan [4]. To reduce redundancy and minimize potential bias in phylogenetic and genetic distance analyses, sequence de-duplication was performed following multiple sequence alignment. ORF5 sequences showing 100% nucleotide identity across the entire aligned region were grouped, and a single representative sequence was retained from each group. Identity was defined as complete identity across all aligned nucleotide positions, including gap characters. After removal of identical sequences, 287 non-redundant domestic L4 sequences were retained for downstream analyses.

To anchor the domestic dataset within a globally standardized ORF5-based classification framework, additional reference sequences were included. Twenty-three L4 ORF5 sequences were selected from the global dataset used by Yim-im et al. [3]. In addition, 11 ORF5 sequences were retrieved from GenBank to supplement temporal and geographical diversity. Only these GenBank sequences (n = 11) were assigned to L4 using Nextclade based on the ORF5-based global lineage classification system (<https://clades.nextstrain.org>) [7] and were subsequently processed using the same analytical pipeline

as the domestic dataset. In total, 321 L4-assigned ORF5 sequences (287 domestic, 23 from the Yim-im dataset, and 11 from GenBank) were included for reassessment of the L4-assigned population. For phylogenetic placement relative to established PRRSV-2 ORF5 lineages, non-L4 ORF5 sequences from the Yim-im global dataset were additionally included in the maximum-likelihood phylogenetic reconstruction. These non-L4 reference sequences were used to provide global lineage context for the tree topology and did not alter the definition of the L4-assigned dataset described above. Among the 287 non-redundant domestic L4-assigned ORF5 sequences, 233 were resolved within the candidate clade designated JPL-A in the ML phylogeny. These 233 domestic JPL-A sequences were newly generated in this study. Four of these sequences were excluded prior to sequence submission because three contained internal gaps within the ORF5 coding region and one contained a premature stop codon preventing translation. The remaining 229 complete ORF5 sequences were deposited in the DDBJ database under accession numbers LC921277–LC921505.

2.2. Multiple sequence alignment

Multiple sequence alignment of ORF5 nucleotide sequences was performed in MEGA version 12 using the MUSCLE algorithm [8]. All sequences were trimmed to 603 nt corresponding to the complete ORF5 coding region. The final alignment was visually inspected after translation to confirm the absence of frame shifts, internal gaps, premature stop codons, and excessive ambiguous positions. Alignment columns were not filtered prior to downstream analyses; sites containing gaps or missing data were handled according to the settings described below in the genetic distance analyses.

2.3. Phylogenetic analysis

Phylogenetic relationships among PRRSV-2 ORF5 sequences were inferred using maximum-likelihood (ML) analysis implemented in IQ-TREE version 2.3.6. Following the analytical framework used in the global ORF5 classification study [3], a nucleotide substitution model (GTR + F + R10) was applied and branch support was assessed using 1,000 bootstrap replicates. Phylogenetic trees were visualized and annotated using the Interactive Tree of Life (iTOL) v7 (<https://itol.embl.de/>). Major monophyletic clades were defined based on support at the defining (basal) node (bootstrap $\geq 85\%$). These clades were treated as candidate lineage-level clades for downstream distance-based evaluation (Section 2.5). For clarity in figures and tables, the candidate clade that was distinct from the clade containing the global L4 reference sequences was provisionally labeled “JPL-A” (a provisional label); this label is used solely for reporting in this study and does not imply formal acceptance of a revised global nomenclature.

2.4. Genetic distance analysis

Genetic distances were calculated from the aligned ORF5 nucleotide sequences in MEGA version 12 to quantify divergence within and between candidate lineages. Pairwise nucleotide distances were calculated as uncorrected p-distances, with gaps and missing data handled in MEGA12 using the pairwise deletion option, and all codon positions (1st, 2nd, and 3rd) included. For distance-based analyses, only sequences assigned to either the major L4 clade or the candidate clade in the ML phylogeny were included. Sequences that did not consistently cluster within either of these clades were

excluded from distance distribution analyses. Although additional internal structure was observed within both clades, further subdivision was not applied because the corresponding internal nodes were not consistently supported by bootstrap values $\geq 85\%$. Accordingly, distance-based comparisons were restricted to clades defined by robustly supported backbone topology. Mean pairwise distances were summarized for (i) within-clade comparisons and (ii) between-clade comparisons and reported as a distance matrix (Table 1). In addition, the full distribution of between-clade pairwise distances was visualized to assess whether divergence between the major L4 clade and the candidate clade consistently exceeded the 11% lineage threshold beyond mean values alone (Fig. 2). All pairwise p-distance values were exported from MEGA and used to generate the distance distribution plot. The violin plot was produced in Python (version 3.10) using custom scripts based on the matplotlib library [9].

2.5. Criteria for proposing candidate novel lineages

Candidate lineage-level clades within the L4-assigned population were defined using conservative criteria integrating topology, support, and distance structure. Following phylogenetic reconstruction of the L4-assigned dataset together with the global reference dataset [3], a putative candidate clade was evaluated as a candidate lineage when it met all of the following: (i) monophyly in the ORF5 ML phylogeny; (ii) bootstrap support for the defining branch of $\geq 85\%$; (iii) mean within-clade p-distance of $< 11\%$; (iv) mean between-clade p-distance of $\geq 11\%$ from the major clade containing the global L4 reference sequences; and (v) between-clade ORF5 p-distance distributions consistently exceeding the 11% lineage threshold. The 11% divergence threshold was adopted in accordance with the global ORF5 classification framework [3]. For clarity, the clade designated as JPL-A in this study is referred to as JPL-A in figures and tables as a provisional reporting label; this label is used solely for consistency in this manuscript and does not constitute a formal nomenclature proposal. Sequences that did not consistently cluster within the major L4 clade or the candidate clade were classified as “unclassified/others.” Because this study is ORF5-focused, whole-genome confirmation and recombination assessment were considered outside scope and are noted as priorities for future work.

3. Results

3.1. Phylogenetic structure within the L4-assigned population

ML analysis of ORF5 nucleotide sequences previously assigned to lineage 4 identified at least two major monophyletic clades within the L4-assigned population (Fig. 1). One of these clades contained the global L4 reference sequences from the Yim-im dataset and was therefore defined as the major L4 clade. In contrast, a second monophyletic clade, composed predominantly of domestic Japanese sequences, formed a distinct clade clearly separated from the major L4 clade. This clade showed consistent bootstrap support of $\geq 85\%$ at its defining node and was provisionally designated as “JPL-A” in this study. Although additional internal structure was observed within both major clades, none of the corresponding internal nodes were consistently supported at the $\geq 85\%$ bootstrap level. Therefore, subsequent analyses

were conducted using the two backbone clades (the major L4 clade and the JPL-A clade) as operational groups, without further subdivision.

3.2. Genetic distance patterns between the major L4 clade and the JPL-A clade

Pairwise nucleotide distance analysis of ORF5 sequences showed that the mean p-distance between the major L4 clade and the JPL-A clade exceeded the 11% divergence threshold used to demarcate PRRSV-2 lineages in the global ORF5 classification framework (Table 1). Visualization of the full distribution of between-clade pairwise distances showed that the distribution was centered above the 0.11 threshold, with a median distance of 0.133 (Fig. 2). Most comparisons exceeded this threshold.

4. Discussion

The present phylogeny- and distance-based reassessment of PRRSV-2 ORF5 sequences previously assigned to lineage 4 revealed two major backbone clades within the L4-assigned population in Japan. One clade contained the global L4 reference sequences defined in the updated ORF5 classification framework, whereas the other formed a distinct monophyletic clade supported by bootstrap values $\geq 85\%$ and characterized by between-clade ORF5 p-distances centered above the 11% lineage demarcation threshold [3, 7]. Based on these independent phylogenetic and genetic distance patterns, we report this latter group, provisionally designated JPL-A, as a candidate lineage-level clade distinct from the major L4 clade under the current ORF5-based framework.

A central question arising from this study is why a candidate lineage-level clade emerged from viruses that were initially classified as “L4” by a reference-based tool. Nextclade enables rapid, standardized, and reproducible lineage assignment by placing query sequences onto a fixed reference phylogeny and has become an important component of contemporary PRRSV molecular surveillance [7]. However, lineage labels produced by such tools are inherently dependent on the diversity represented in the reference datasets that anchor lineage definitions. In the global ORF5 framework, L4 is primarily represented by sequences sampled between the early 1990s and 2010 [3], whereas the Japanese sequences analyzed here were largely collected between 2020 and 2023. Thus, the observed phylogenetic separation likely reflects continued regional evolution and diversification of L4-related viruses beyond the temporal scope of historical reference datasets, rather than misclassification by the reference-based approach itself. When regionally evolving viral populations accumulate genetic diversity that is insufficiently represented in historical references, divergent clades may be provisionally assigned to the closest available lineage label, even when deeper phylogenetic separation exists. The observation that global L4 reference sequences clustered within only one of the two backbone clades identified here is consistent with this interpretation.

An important implication of this study concerns the future role of standardized, open, and reference-based classification platforms such as Nextclade and PRRSV-Loom. These tools provide a shared analytical framework in which sequences generated by different laboratories, regions, and time periods can be evaluated on a common phylogenetic background, thereby enhancing reproducibility, transparency, and comparability across studies [7, 10, 11]. Their long-term value, however, depends on the continual updating and critical reassessment of the reference datasets that define lineage boundaries. The identification of the JPL-A clade illustrates how national or regional surveillance programs can actively contribute to this process by detecting phylogenetically coherent groups that are not yet incorporated into global reference collections. Accordingly, our findings should be viewed as complementary to reference-based tools and supportive of future updates of ORF5 lineage resources as additional data become available.

Recently, a nationwide ORF5-based study from South Korea proposed a previously unrecognized Korean-specific clade as a novel PRRSV-2 lineage based on consistent phylogenetic separation and lineage-level genetic divergence [12]. This independent regional report provides a closely related example demonstrating how expanding national surveillance datasets can reveal lineage-level diversity not yet incorporated into global reference frameworks. Similarly, a recent nationwide Canadian preprint identified multiple highly divergent PRRSV-2 clades within established lineages and proposed two novel sub-lineages, at a finer classification level, based on standardized phylogenetic and genetic distance criteria, emphasizing the extent of previously unrecognized diversity even in well-characterized regions [13]. That study further highlighted that recognizing and naming such viral groups is not merely a taxonomic exercise, but a necessary step toward building a coherent global framework that supports consistent communication, epidemiological interpretation, and coordinated surveillance efforts. Together with the present study, these findings support the view that regional sequence resources play a critical role in the iterative refinement of PRRSV-2 genetic classification.

Beyond its classification implications, from a molecular epidemiological perspective, reporting JPL-A as a provisional candidate lineage-level clade has practical utility even prior to any formal update of global nomenclature. The availability of a consistent operational label facilitates longitudinal tracking of this clade within Japan and supports standardized reporting across surveillance efforts. Previous large-scale studies have demonstrated that structured genetic classification enhances the interpretability of PRRSV molecular surveillance data, enabling detection of lineage turnover, shifts in dominance, and spatiotemporal dissemination patterns [2, 5, 6]. In this context, provisional designation of this candidate clade provides a pragmatic framework for monitoring emerging diversity while formal nomenclature revisions are pending. The identification of JPL-A highlights the importance of periodically reassessing lineage structure as national sequence datasets expand and as contemporary viral diversity diverges from historical reference sets. From a practical standpoint, continued recognition and tracking of such a candidate lineage-level clade may support earlier detection of emerging PRRSV diversity with potential epidemiological or control relevance.

Several limitations should be acknowledged. This study focused exclusively on ORF5, and lineage-level inference is therefore restricted to the evolutionary signal contained within this single genomic region. Although large-scale analyses indicate that recombination within ORF5 is relatively uncommon compared with other regions of the PRRSV genome, discordance between ORF5 phylogeny and whole-genome evolutionary history cannot be excluded [3]. Accordingly, whole-genome sequencing, recombination screening, and phylogenetic confirmation using additional genomic regions will be essential next steps to evaluate whether JPL-A represents a coherent genome-wide lineage rather than an ORF5-defined grouping. Furthermore, we intentionally avoided further subdivision within the major L4 clade and within JPL-A because internal nodes did not consistently meet the bootstrap $\geq 85\%$ criterion. Future analyses incorporating expanded domestic and international sequence collections may resolve additional reproducible substructure that can be evaluated under established sublineage or variant frameworks [5, 11]. Integration of Japanese sequence data into international reference datasets will also be important to determine whether JPL-A or related clades are present but currently under-sampled in other regions.

Overall, combined evidence from ML phylogeny and ORF5 p-distance structure supports reporting JPL-A as a candidate lineage-level clade distinct from the major L4 clade within the current ORF5 classification framework. Continued PRRSV-2 sequencing in Japan, systematic phylogenetic reassessment of expanding datasets, integration of these data into global reference resources, and genome-wide follow-up analyses will be critical to determine whether formal recognition of this clade within future updates of PRRSV-2 nomenclature is warranted and to ensure that standardized classification tools continue to reflect contemporary viral diversity.

Declarations

Note: Nucleotide sequence data reported in this paper are available in the DDBJ database under accession numbers LC921277–LC921505.

Competing Interests

The authors declare no competing interests.

Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

O.T.: Conceptualization, Methodology, Investigation, Software, Formal analysis, Writing–original draft, Writing–review & editing. R.T.: Investigation, Software, Formal analysis, Data curation. Y.Y.: Resources, Data curation. N.T.: Resources. K.S.: Funding acquisition, Writing–review & editing.

References

1. Osemeke O, Silva GS, Corzo CA et al (2025) Economic impact of productivity losses attributable to porcine reproductive and respiratory syndrome virus in United States pork production, 2016–2020. *Prev Vet Med* 244:106627. <https://doi.org/10.1016/j.prevetmed.2025.106627>
2. Shi M, Lam TT-Y, Hon C-C et al (2010) Phylogeny-Based Evolutionary, Demographical, and Geographical Dissection of North American Type 2 Porcine Reproductive and Respiratory Syndrome Viruses. *J Virol* 84:8700–8711. <https://doi.org/10.1128/jvi.02551-09>
3. Yim-im W, Anderson TK, Paploski IAD et al (2023) Refining PRRSV-2 genetic classification based on global ORF5 sequences and investigation of their geographic distributions and temporal changes. *Microbiol Spectr* 11:e02916–e02923. <https://doi.org/10.1128/spectrum.02916-23>
4. Takai R, Sato T, Yonezawa Y et al (2025) Spatiotemporal distribution and genetic characteristics of porcine reproductive and respiratory syndrome virus-2 lineage 1 in Japan from 2020 to 2023. *Res Vet Sci* 106025. <https://doi.org/10.1016/j.rvsc.2025.106025>
5. Paploski IAD, Pamornchainavakul N, Makau DN et al (2021) Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States. *Vaccines* 9:608. <https://doi.org/10.3390/vaccines9060608>
6. Paploski IAD, Corzo C, Rovira A et al (2019) Temporal Dynamics of Co-circulating Lineages of Porcine Reproductive and Respiratory Syndrome Virus. *Front Microbiol* 10:2486. <https://doi.org/10.3389/fmicb.2019.02486>
7. Zeller MA, Chang J, Trevisan G et al (2024) Rapid PRRSV-2 ORF5-based lineage classification using Nextclade. *Front Vet Sci* 11:1419340. <https://doi.org/10.3389/fvets.2024.1419340>
8. Kumar S, Stecher G, Suleski M et al (2024) MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Mol Biol Evol* 41:msae263. <https://doi.org/10.1093/molbev/msae263>
9. Hunter JD (2007) Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* 9:90–95. <https://doi.org/10.1109/mcse.2007.55>
10. VanderWaal K, Pamornchainavakul N, Kikuti M et al (2024) Phylogenetic-based methods for fine-scale classification of PRRSV-2 ORF5 sequences: a comparison of their robustness and reproducibility. *Front Virol* 4:1433931. <https://doi.org/10.3389/fviro.2024.1433931>
11. VanderWaal K, Pamornchainavakul N, Kikuti M et al (2025) PRRSV-2 variant classification: a dynamic nomenclature for enhanced monitoring and surveillance. *mSphere* 10:e00709–e00724. <https://doi.org/10.1128/msphere.00709-24>

12. Kim S-C, Kang SC, Kim H-J et al (2025) Nationwide emergence and spread of highly virulent PRRSV-2 mutants in Korea. *Porc Heal Manag* 11:58. <https://doi.org/10.1186/s40813-025-00470-5>
13. da Silva JPH, Paploski IAD, Charette R et al (2026) Phylogenetic Lineages of PRRSV-2 from Canada Reveal Patterns of Transboundary Spread and Two Novel Sub-Lineages in North America. <https://doi.org/10.20944/preprints202601.0710.v1>. Preprints [Preprint]

Tables

Table 1 is available in the Supplementary Files section.

Figures

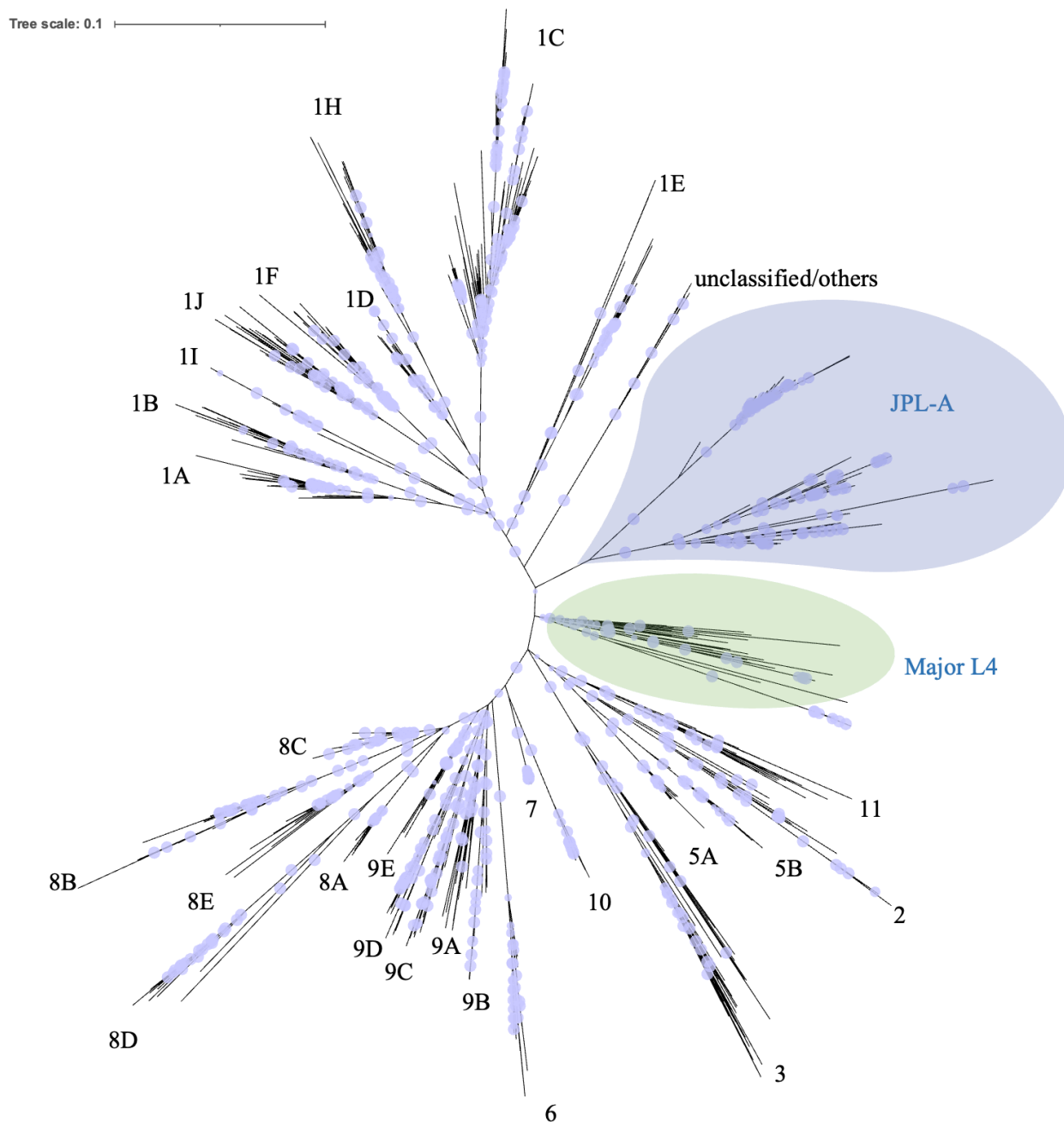


Figure 1

Maximum-likelihood phylogeny of PRRSV-2 ORF5 sequences showing the placement of the major L4 clade and the candidate clade (JPL-A).

Maximum-likelihood trees were inferred in IQ-TREE (GTR+F+R10) with 1,000 bootstrap replicates. Major ORF5 lineages are labeled. The clade containing global L4 reference sequences is indicated as the major L4 clade, and a distinct monophyletic clade predominantly composed of Japanese sequences is indicated as JPL-A (provisional designation in this study). Shaded areas highlight the two backbone clades defined by bootstrap support and used for distance-based analyses. Filled circles

indicate branches with bootstrap support $\geq 85\%$. Sequences not consistently clustering with either backbone clade are labeled as unclassified/others. The tree is shown unrooted for visualization.

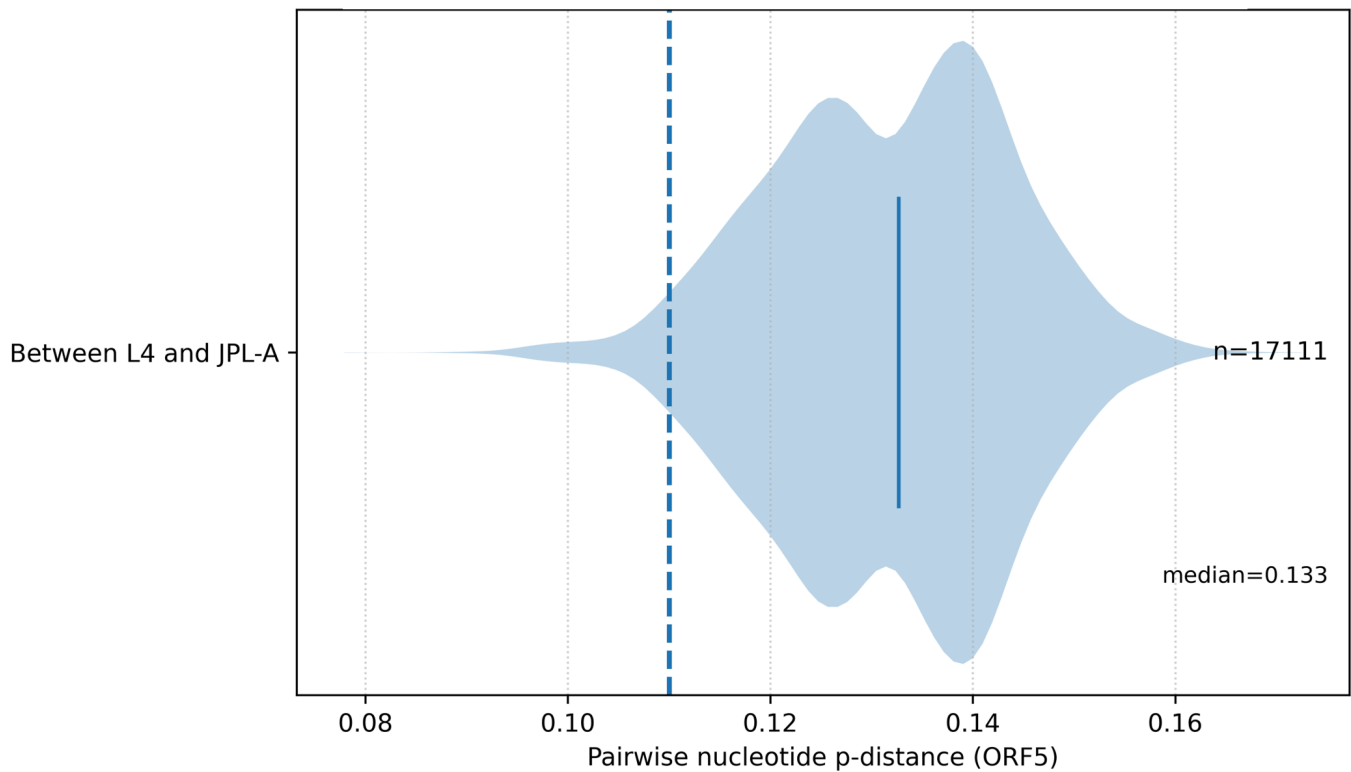


Figure 2

Between-group distribution of pairwise nucleotide p-distances between the major L4 clade and the candidate clade (JPL-A).

Pairwise nucleotide p-distances were calculated in MEGA12 from the aligned ORF5 nucleotide dataset using pairwise deletion for gaps/missing data and including all codon positions. Only sequences assigned to either the major L4 clade or the JPL-A clade in the maximum-likelihood phylogeny were included. All between-clade pairwise p-distance values were exported from MEGA and visualized as a violin plot in Python (v3.10) using the matplotlib library. The solid vertical bar indicates the median. n represents the total number of between-clade pairwise comparisons ($n_{L4} \times n_{JPL-A}$). The dashed vertical line indicates the 0.11 divergence threshold used for ORF5-based lineage demarcation.

Supplementary Files

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

- [Table.1.xlsx](#)