

Article

Not peer-reviewed version

---

# Current Evolutionary Dynamics of Porcine Epidemic Diarrhea Virus (PEDV) in the U.S. a Decade After Introduction

---

[Joao Paulo Herrera da Silva](#)\*, [Nakarin Pamornchainavakul](#), [Mariana Kikuti](#), Xiaomei Yue, [Cesar A. Corzo](#), [Kim VanderWaal](#)\*

Posted Date: 20 March 2025

doi: 10.20944/preprints202503.1500.v1

Keywords: porcine epidemic diarrhea virus, PEDV; pigs; evolution; coronavirus; temporal dynamics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

*Article*

# Current Evolutionary Dynamics of Porcine Epidemic Diarrhea Virus (PEDV) in the U.S. a Decade After Introduction

Joao P. Herrera da Silva \*, Nakarin Pamornchainavakul, Mariana Kikuti, Xiaomei Yue, Cesar A. Corzo and Kimberly VanderWaal

Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA

\* Correspondence: herre466@umn.edu

**Abstract:** Porcine Epidemic Diarrhea Virus (PEDV) was introduced in the United States (U.S.) in 2013, spreading rapidly and leading to economic losses. Two strains, S-INDEL and non-S-INDEL, are present in the U.S. We analyzed 313 genomes and 556 Spike protein sequences generated since its introduction. PEDV case numbers were highest during the first two years after its introduction (epidemic phase), then declined and stabilized in the following years (endemic phase). Sequence availability was higher during the initial epidemic phase. Our results suggest non-S-INDEL strain is the predominant strain in U.S. The non-S-INDEL strain exhibits a nucleotide identity percentage above 97.6%. Most non-S-INDEL sequences sampled after 2017 clustered into two sub-clades. No descendants derived from other clades present in the epidemic period were identified in the contemporary data, suggesting that these clades are no longer circulating in the U.S. The two clades currently circulating are restricted to two respective geographic regions and our results suggest limited inter-regional spread. This insight helps determine the risk of re-introduction of PEDV if it were regionally eliminated. Continued molecular surveillance is key to confirming the extinction of older clades, mapping the distribution and spread of recent clades, and understanding PEDV's evolutionary diversification.

**Keywords:** porcine epidemic diarrhea virus, PEDV; pigs; evolution; coronavirus; temporal dynamics

## 1. Introduction

Porcine Epidemic Diarrhea Virus (PEDV) causes acute enteric disease in pigs, which is often lethal in suckling piglets (1-3). PEDV is widely dispersed throughout the world and is responsible for significant losses in the swine industry globally (4-7). PEDV was first reported in the United States (U.S.) around April of 2013, and was responsible for a major epidemic in the country (6, 8). After its introduction, the virus spread rapidly across the U.S. and has been circulating ever since in swine herds. The epidemic phase of PEDV in the U.S., a period characterized by rapid between-farm spread and high incidence, lasted April 2013 to August 2014. This has been followed by an endemic phase in which disease incidence stabilized up to the present (9-11). The primary mode of transmission for PEDV is through the fecal-oral route (12). The following activities have been reported as main modes of PEDV spread between farms: transportation of infected animals from one farm to another, fomites carried by movement of farm employees between farms and vehicles between different farms, and contaminated feed, among others (13-16)

The causative agent of this disease, Alphacoronavirus porci, is an enveloped, positive-sense, single-stranded RNA virus of the Coronaviridae family, Alphacoronavirus genus, with a genome of approximately 28 kb (17). Its genome consists of ORF1a and ORF1b, encoding 16 non-structural proteins involved in replication, transcription, and immune evasion, and other four ORFs that encodes for a structural protein: the spike protein (S), nucleocapsid (N), membrane (M), envelope (E),

as well as the accessory protein ORF3. RNA viruses are characterized by rapid evolutionary dynamics, which allows them to adapt quickly to changes in the niches they occupy (18, 19). Most RNA viruses have high mutation rates due to the absence of proofreading activity in their replicases (20, 21). Coronaviruses encode a proofreading exonuclease (nsp14), causing these viruses to exhibit lower mutation rates compared to other RNA viruses (22, 23). Recombination is another mechanism that drives the diversification of viral populations. Recombination enables the emergence of new haplotypes, enables new epistatic interactions, and can purge deleterious mutations (21, 24-26). Coronaviruses exhibit high rates of recombination, and recombination events have been reported in all seven ORFs of PEDV (27-31).

PEDV is divided into two major genogroups based on the phylogeny of the spike protein, and the G1 and G2 genogroups are further subdivided into five subgroups: G1a/b, and G2a/b/c (6, 32, 33). In the U.S., the two strains present are the S-INDEL strain (low pathogenic), which is phylogenetically related to genogroup G1b, and the non-S-INDEL strain (high pathogenic), which is related to genogroup G2b. These two strains are characterized by the presence or absence of deletions in the ORF-S gene, respectively (3, 8, 31, 34). The PEDV non-S-INDEL strain was first detected in April 2013, whereas the S-INDEL strain was reported in June 2013 (6). Phylogenetic analyses demonstrated that PEDV sequences in the U.S. were closely related to PEDV from China, suggesting a possible origin of PEDV found in the U.S. (3, 6). Recent studies using phylogeographic approaches have highlighted connectivity between several swine-producing countries, allowing the identification of key contributors to the transboundary transmission chain (3, 35-37). For example, several “US-like” PEDV strains have been documented in relation to outbreaks in other countries, suggesting that the United States has played a role in the spread of PEDV to countries, including South Korea, Japan, China, and Mexico (3, 35-37).

Most studies investigating the evolutionary dynamics of PEDV in the U.S. were carried out only at the early stages of the epidemic, shortly after its introduction (3, 6, 8, 30, 31). The evolutionary dynamics of PEDV during the endemic period in the country remains unexplored. Here, we aim to investigate the temporal dynamics of PEDV in the U.S. over more than a decade since its introduction. By estimating the genetic diversity and variability of PEDV and through Bayesian inference, we characterized the changes in the genetic composition of the PEDV population over time.

## 2. Materials and Methods

### *Data Source*

A dataset consisting of 556 PEDV spike protein sequences from the U.S., generated between 2013 and 2024 was assembled. Of these, 228 sequences were downloaded from GenBank (38), and 328 were obtained from the Morrison Swine Health Monitoring Project (MSHMP) (39). MSHMP is a voluntary disease monitoring program that aggregates data on swine diseases accounting for approximately 60% of the breeding herd population (9, 39). The project also curates PEDV sequences generated by participants through their routine monitoring efforts in breeding, gilt developing units, growing and finishing herds. A second additionally, another dataset comprising 313 PEDV full-length genomes from the U.S., generated between 2013 and 2017 was constructed which included 73 sequences obtained from MSHMP and 240 sequences retrieved from GenBank. The sequences included in this study originated from 19 states that include all major swine producing regions.

For the purpose of this study, the “epidemic period” was defined as spanning from May 2013 to August 2014, during which disease prevalence peaked at 31.6% of monitored sow farms (9-11). The “endemic period” was defined as the one beginning in August 2014 and continuing thereafter, the prevalence decreased and stabilized, reaching a peak of 5.5% in 2022 (9-11).

### *Sequence Alignment*

Using the default parameters, sequence alignment was performed using MAFFT V7.490 (40). The sequences were trimmed by removing the 5' and 3' UTR regions and only the coding regions of

the genomes were used in our analyses. Sequences were classified as the S-INDEL or non-S-INDEL strain based on the presence of insertions and deletions at specific positions in the coding region of the Spike protein, as described by Wang et al. (8).

### *Sequences Comparison*

Genetic distance for both the complete genome and the spike protein within and between clades was estimated as the pairwise nucleotide identity percentage using sequence tool demarcation SDT v.1.2 (41). Pairwise identity percentages were estimated for each strain, between strains, and between the current clades and the ancestral (epidemic period) clades of the non-S-INDEL strain.

### *Recombination Analysis*

Complete PEDV genomes and spike (S) protein sequences were scanned to infer putative recombination events using the Rdp (42), Geneconv (43), Bootscan (44), Max  $\chi^2$  (45), Chimaera (46), Siscan (47) and 3Seq (48) methods implemented in RDP5 (49) using default parameters. Only recombination events supported by at least 4 different methods and P-values lower than  $\alpha = 0.05$  (Bonferroni corrected) were considered to be reliable. Phylogenetic networks were inferred to capture patterns of non-tree-like evolution using the Neighbor-Net algorithm implemented in SplitsTree v. 4.19.2 (50). Distances were calculated using the GTR+G substitution model determined in jModelTest2, using 1,000 bootstrap replicates (50).

### *Phylogenetic Analysis*

Phylogenetic trees were inferred by maximum-likelihood implemented in RAxML-NG (51). The nucleotide substitution model was determined using ModelTest-NG (52). To assess whether the data exhibited any evidence of temporal signal, TempEst was used to perform a root-to-tip regression on ML phylogenetic trees (53). TempEst is useful for exploratory analyses of temporal signal and identifying outliers or sequences with incorrect dating (54, 55). However, several limitations have been discussed regarding the use of root-to-tip regression as a formal statistical test, such as the lack of data independence and the assumption of strict molecular clock-like behavior. Therefore, we employed the Bayesian Time Signal (BETS) algorithm (54, 55), implemented in BEAST V.1.10.4 (56), as a formal test to assess the temporal signal structure of PEDV populations. To determine the evidence of a temporal signal, we performed simulations where we provided sequence date information (heterochronous date) and where no date information was provided (isochronous date). The simulations were performed using both a strict molecular clock and a relaxed uncorrelated lognormal molecular clock. The performance of each simulation was evaluated by estimating the (log) Bayes factor (57), which was estimated based on the difference in the (log) maximum marginal likelihoods of the competing models. The marginal (log) likelihoods were estimated by generalized stepping-stone (GSS) sampling (58). We used an MCMC chain length of  $1 \times 10^8$  with 25% burnin, and a number of stepping stones of 100, with an subchain MCMC length of  $1 \times 10^6$ . The convergence of the analyses was assessed using Tracer v1.7 and was determined based on effective sample size of at least 200, and by the degree of interdependence of the samples assessed based on the degree of mixing of the parameters (59).

### *Genetic Diversity*

The average nucleotide diversity index,  $\pi$  (60), was computed through pairwise sequence comparisons using a Python script (61) for both complete genomes and individual coding regions. We estimated the 95% confidence intervals for the mean  $\pi$  values using a bootstrap test with 1,000 non-parametric simulations (61), performed with the Simpleboot package in R software (62). This approach was used to assess the statistical significance among the  $\pi$  values. Nucleotide diversity per site across the genome was calculated using a sliding window of 100 nucleotides with a step size of 10 nucleotides, utilizing DnaSp v.5.10 (63).  $\pi$  values per site were estimated for each ORF individually

and for the whole genome set, as well as for each strain individually (PEDV S-INDEL and PEDV non-S-INDEL).

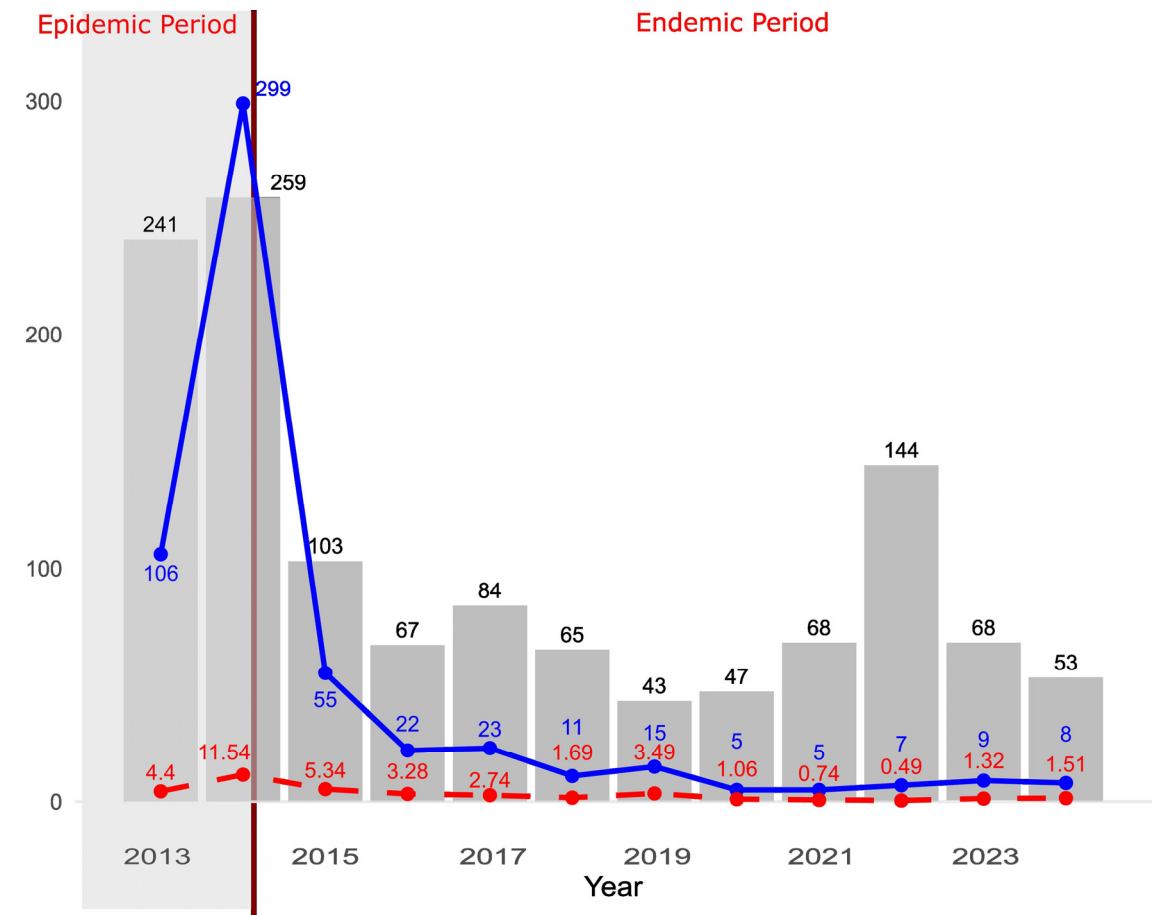
*Selections Analysis*

Selection analyses were performed to infer how natural selection acting on different ORFs encoding structural proteins of the PEDV genome. Recombination events can confound the search for sites under selection, so to avoid noise generated by recombination, a search for recombination breakpoints was performed for each ORF using the Genetic Algorithm Recombination Detection (GARD) (64). We then performed a search for potential sites under positive or negative selection in each of the ORFs encoding structural proteins. Three different methods were used to screen sites for evidence of selection: Single Likelihood Ancestor Counting (SLAC) (65) Mixed Effects Model of Evolution (MEME) (66) and Fast Unconstrained Bayesian Approximation (FUBAR) (67), all implemented in the DataMonkey web server (68). We also estimated the ratios between synonymous and nonsynonymous substitution rates ( $\omega$ ) using SLAC (65).

3. Results

*Stabilization of PEDV Case Numbers After the Epidemic Period*

The occurrence of PEDV in the country was evaluated based on the number of cases reported by production systems and participating in MSHMP. Here, “case numbers” refer to the number of sow farms reporting an outbreak. Following the introduction of PEDV, a massive number of cases were reported, especially during the first two years (epidemic period). After this initial surge, the number of cases decreased and became stable, with slight fluctuations, as expected in an endemic period. However, 2022 was an unusual year, showing a slight increase in the number of cases (Figure1).



**Figure 1.** Bars represent the number of cases per year reported to MSHMP. The solid blue line represents the number of spike protein sequences available. The dashed red line represents the number of sequences per 10 reported outbreaks. The brown line marks the boundary between the epidemic and endemic periods.

PEDV sequence surveillance was intense during the epidemic period. Afterwards, the number of sequences obtained decreased, particularly after 2019, when fewer than ten sequences per year were generated. In terms of sequencing effort, the ratio of sequences per number of cases was fairly constant as a median of 2.22 sequences per ten cases across 2013-2024 (with an interquartile range of 1.26 to 3.72). At least one sequence per ten cases was obtained, except in 2021 and 2022, when only 0.74 and 0.49 sequences per ten cases were generated, respectively (Figure 1). The ratio has been reported as the number of sequences available per number of cases reported. However, it is important to note that the case data available includes only those reported in breeding herds and exclusively those submitted to the MSHMP. As a result, the actual number of cases may be higher than what is presented here. Furthermore, the sequences we analyzed do not necessarily originate solely from breeding herds as some may have been obtained from grow-finish farms.

In total, we analyzed 556 spike protein sequences, with 499 corresponding to the non-S-INDEL strain and 57 to S-INDEL strain. Additionally, we analyzed 313 full genomes, of which 288 corresponded to non-S-INDEL strain and 25 to S-INDEL strain. These findings suggest that the non-INDEL strain is the predominant in the U.S.

#### *Recombination Analysis*

Recombination analyses were conducted to understand the contribution of horizontal transfer of genetic material to the diversification of PEDV populations in the U.S.. In total, we identified 12 recombination events (Table 1A) that were detected by at least four different methods implemented in RDP5 with a p-value less than 0.05.

Among these events, four warrant special attention due to their detection in a large number of sequences. Event 6 in Table 1A involved a recombinant block located in the ORF1b region, more specifically, located between nsp13 and nsp15. This event was detected in 106 PEDV genomes (Table 1a). Event 10 was detected in 116 sequences, with the recombination breakpoint located between nsp4 in ORF1a and nsp12 in ORF1b (Table 1A). Event 11 was detected in 223 sequences, with the recombination breakpoint located between nsp3 in ORF1a and nsp12 in ORF 1b (Table 1A). Event 12 was detected in 152 sequences, with the recombination breakpoint located between nsp15 in ORF1b and Spike (Table 1A).

Recombination analyses were also performed specifically for the S1 subunit of the Spike protein. Only two well-supported recombination events were identified in the S1 region (Table 1B). These results differ somewhat from the full-genome analysis, where we were able to identify a greater number of recombination events in the Spike protein. Some sequences that showed evidence of recombination in the spike protein region in the full-genome analysis were not detected as recombinants when analyzing only the Spike protein dataset. In the full-genome analysis, 179 sequences were identified as recombinants, with breakpoints covering the S1 region of the Spike protein (positions 20,342 to 22,562). However, when analyzing only the S1 region dataset (n = 566 sequences), only 61 sequences were detected as recombinants. We hypothesize that the loss of recombination signal in these datasets may be because this analysis was performed using partial sequences. Partial sequences differ from the context of the complete genome. The absence of regions where recombination events extend may weaken the signal, potentially leading to their non-detection. Since some of these recombination events located in the Spike region had their breakpoint initiation in other regions for instance, event 12, where the breakpoint began in nsp 15, this could have affected the sensitivity of event detection in the spike protein dataset. The majority of sequences that exhibited recombination events in the spike protein region were detected in samples collected before 2017. Only four sequences sampled from 2018 onward showed recombination events in spike protein. However, we cannot rule out the possibility that some of these more recent sequences may

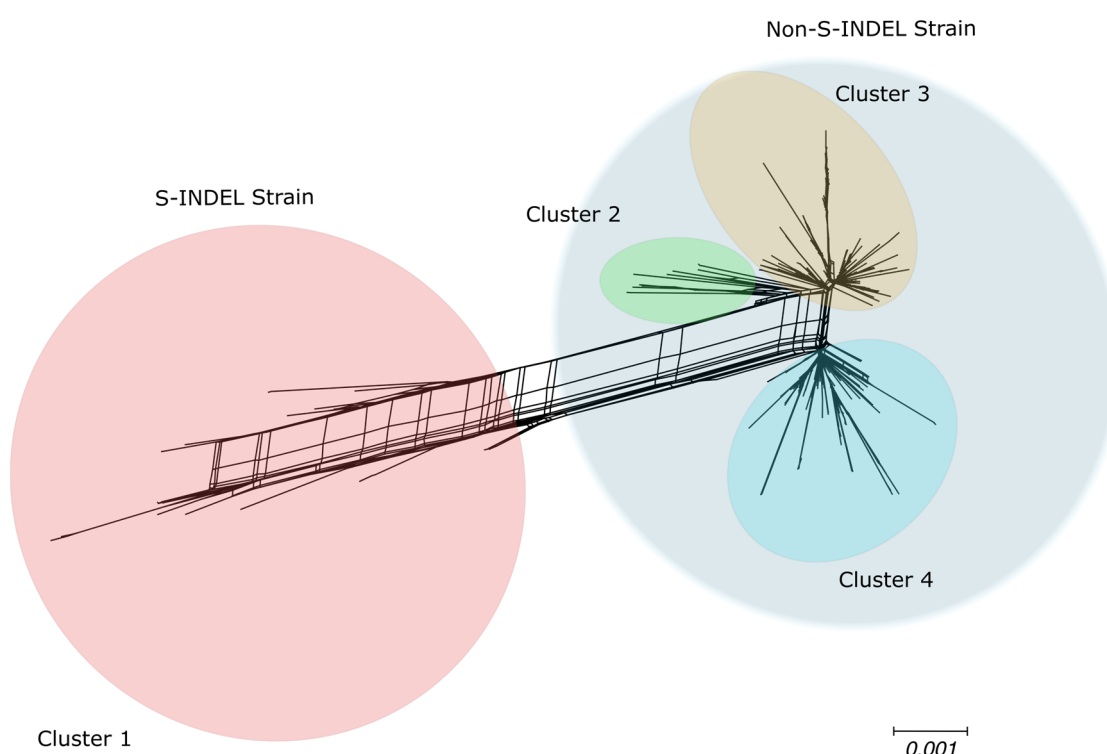
also contain undetectable recombination events using RDP, given that they are partial sequences. As previously highlighted, this can impact the ability to detect such events. This underscores the importance of complete genomes in obtaining a more comprehensive and accurate understanding of the underlying processes driving viral diversification.

**Table 1.** Summary of recombination events detected in the PEDV. A. The analysis was performed encompassing the entire coding region of the PEDV genome. B. The analysis was performed for PEDV Spike. Only recombination events detected by at least four different methods and with P-values lower than a Bonferroni-corrected  $\alpha = 0.05$  were considered significant. Sub-lineages are indicated in parentheses for both the major and minor parental sequences.

|                  |             | Breakpoint<br>Position        |       |                            |                   |                   |                        |                      |
|------------------|-------------|-------------------------------|-------|----------------------------|-------------------|-------------------|------------------------|----------------------|
| A                |             |                               |       |                            |                   |                   |                        |                      |
| Full<br>genome   |             | In<br>Recombinant<br>sequence |       |                            |                   |                   |                        |                      |
| event            | Found<br>in | Begin                         | End   | Recombination<br>sequences | Minor<br>Parental | Major<br>Parental | Detection<br>Methods # | P-<br>value*         |
| 1                | 6           | 16816                         | 22432 | KR265760                   | KM975738          | KF452322          | RGBMCSP3               | 2.64 e-<br>76        |
| 2                | 5           | 3910                          | 20375 | KR265761                   | KR265763          | KM975738          | RGMCS3                 | 6.01 e-<br>75        |
| 3                | 4           | 13244                         | 21566 | KM975740                   | KM975738          | KF452322          | RGMS3                  | 3.03 e-<br>53        |
| 4                | 2           | 13244                         | 17879 | KR265786                   | KJ645704          | KF468753          | RGMC3                  | 3.02 e-<br>46        |
| 5                | 2           | 11075                         | 27425 | KM975738                   | KR265759          | Unknown           | RGBMCS3                | 5.96 e-<br>37        |
| 6                | 106         | 16816                         | 19902 | KU558702                   | KJ645704          | 248               | RGMC3                  | 3.48 e-<br>32        |
| 7                | 12          | 20376                         | 21478 | KR265761                   | Unknown           | KJ645643          | RGBMS3                 | 1.48 e-<br>33        |
| 8                | 7           | 4545                          | 16815 | KR265759                   | Unknown           | 205               | RGBMS3                 | 7.98 e-<br>28        |
| 9                | 6           | 4377                          | 16815 | 264                        | Unknown           | KJ645641          | MCS3                   | 1.16 e-<br>22        |
| 10               | 116         | 9921                          | 14874 | KU893873                   | KF452322          | KJ645635          | RGMC3                  | 3.91 e-<br>19        |
| 11               | 223         | 7680                          | 13243 | KJ645704                   | KJ645635          | KR265844          | RGM3                   | 1.23 e-<br>14        |
| 12               | 152         | 19912                         | 21587 | KJ645641                   | MG837058          | Unknown           | RGMC3                  | 1.20 e <sup>-6</sup> |
| B                |             |                               |       |                            |                   |                   |                        |                      |
| Spike<br>Protein |             | Breakpoint<br>Position        |       |                            |                   |                   |                        |                      |
| event            | Found<br>in | Begin                         | End   | Recombination<br>sequences | Minor<br>Parental | Major<br>Parental | Detection<br>Methods # | P-<br>value*         |
| 1                | 3           | 8                             | 915   | KU982979                   | 265               | 340               | RMSP3                  | 1.75 e-<br>23        |
| 2                | 34          | 8                             | 955   | 340                        | Unknown           | 296               | RBMS3                  | 4.98 e-<br>24        |
| 3                | 24          | 8                             | 1056  | 262                        | Unknown           | KU982968          | MP3                    | 3.88 e-<br>23        |

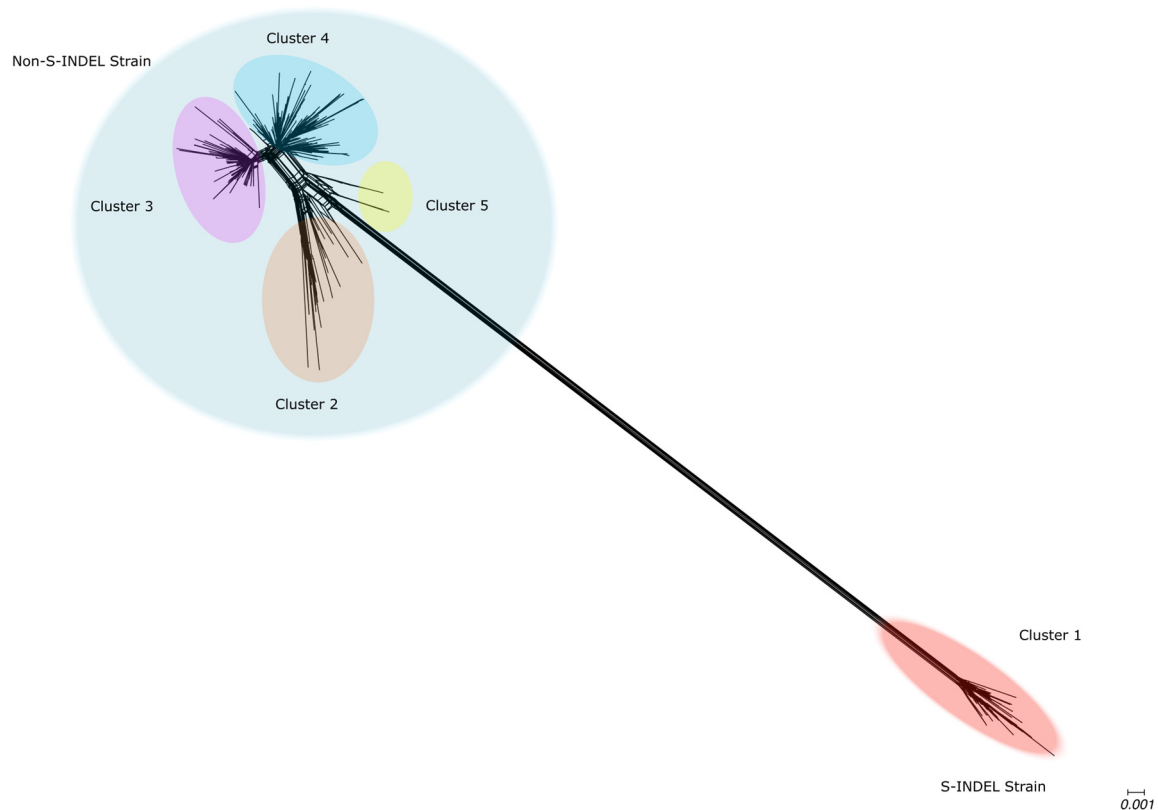
# Abbreviations of detection methods R, Rdp; G, Geneconv; B, Boostcan; M, Maxichi; C, Chimaera; S, Siscan; 3, 3Seq. \* Lowest P-value reported by the method in bold and underline.

Bifurcating phylogenetic trees are not always capable of capturing the evolutionary history of organisms, especially those that experience frequent recombination events or have undergone evolutionary convergence (50). To assess the impact of recombination events in shaping the genealogical trajectory of complete PEDV genomes, we reconstructed phylogenetic networks using SplitsTree4 (50) to capture the non-tree-like evolutionary patterns, which could potentially be explained by evidence of recombination events. The PEDV full genome were grouped into four clusters in the phylogenetic network: The cluster that includes the S-INDEL strain shows a long branch, reflecting the high divergence between these S-INDEL and non-S-INDEL, which is potentially due to evidence of recombination, and three other genetically more similar clusters comprising sequences of the PEDV non-S-INDEL strain (Figure 2). The perpendicular reticulations connecting the branches of the phylogenetic network provide evidence of lateral transfer of genome segments between groups (Figure 2). Additionally, we performed the PHI test, which detected significant evidence of recombination events among the PEDV full genome sequences ( $p < 0.001$ ).



**Figure 2.** Phylogenetic evidence of recombination between PEDV full genome sequences. Neighbor-Net network analysis was performed using SplitsTree4. The formation of a reticulated network rather than a strictly bifurcating tree suggests evidence of recombination. Large circles highlight the PEDV S-INDEL strain in red and the PEDV non-S-INDEL strain in gray. PEDV non-S-INDEL sequences are further divided into three smaller clusters: cluster 2 in green, cluster 3 in orange, and cluster 4 in blue.

Our phylogenetic networks based on the S1 region of the Spike protein displayed a topology similar to that observed in the networks reconstructed for complete genomes. These networks featured a long and highly divergent branch grouping PEDV S-INDEL strain, with a high degree of reticulation in the basal region of the network, suggesting occurrences of recombination inter-strain. A lower degree of reticulation was observed among the branches of the S-INDEL strain, suggesting that intra-strain recombination for S-INDEL are less frequent (Figure 3). The PEDV non-S-INDEL cluster was subdivided into four groups, each also exhibiting a high degree of reticulation (Figure 3). The PHI test further corroborated the evidence of recombination ( $p < 0.001$ ).

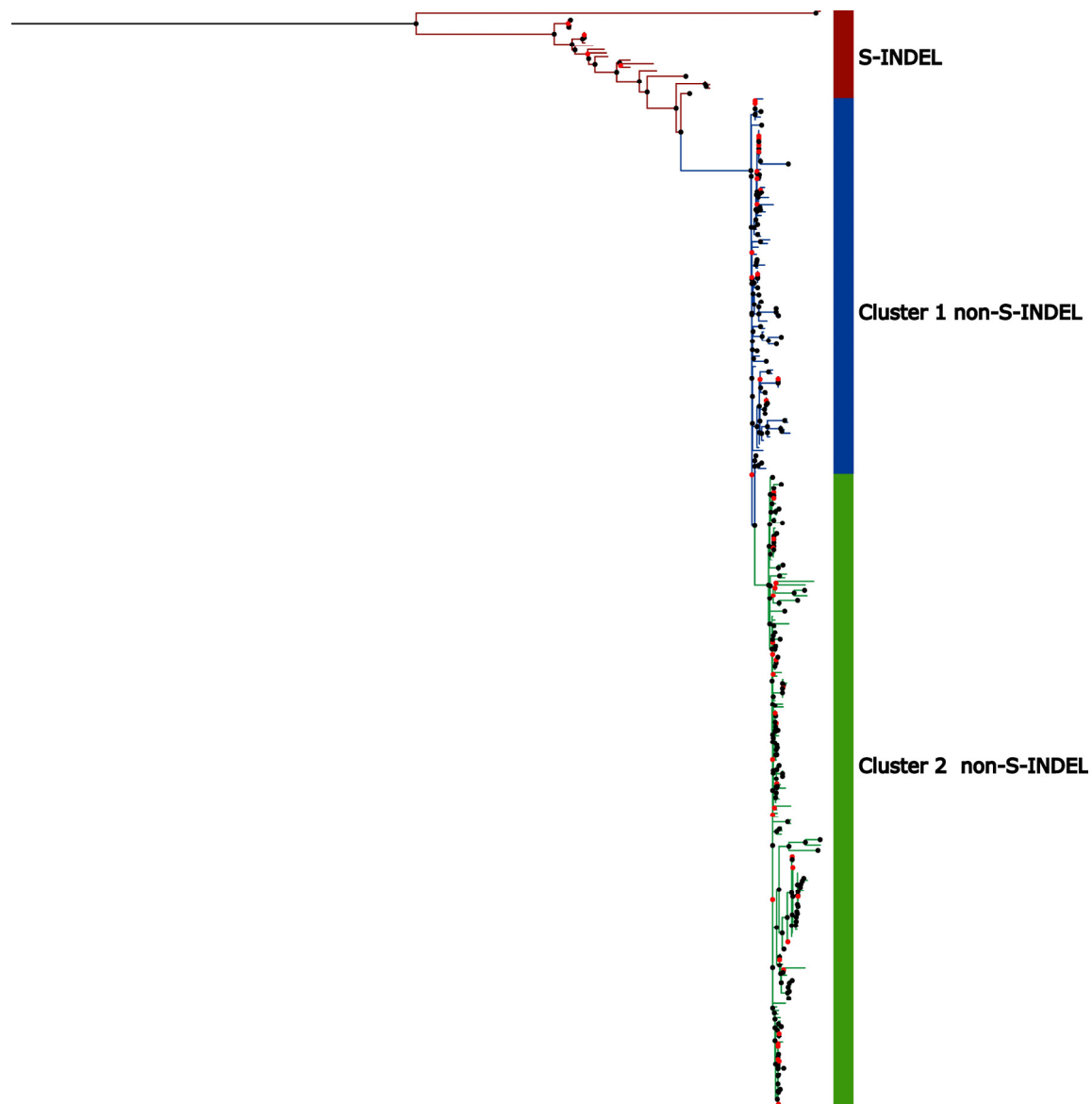


**Figure 3.** Phylogenetic Evidence of Recombination Between PEDV Spike Protein. Neighbor-Net network analysis was performed using SplitsTree4. The formation of a reticulated network, rather than a strictly bifurcating tree, suggests evidence of recombination. Large circles highlight the PEDV S-INDEL strain in red and the PEDV non-S-INDEL strain in gray. PEDV non-S-INDEL sequences are further divided into four smaller cluster: cluster 2 in orange, cluster 3 in pink, cluster 4 in blue, and cluster 5 in yellow.

#### *Temporal Signal and Nucleotide Substitution Rates*

Phylogenetic trees based on the full genomes and Spike protein were reconstructed using maximum likelihood in RAXML-NG(51), employing the GTR+I+G4 nucleotide substitution model, which was determined to be the best fit according to the Akaike Information Criterion (AIC). The phylogenetic tree reconstructed for complete genomes displayed a topology with a more basal group composed of sequences belonging to the PEDV S-INDEL strain and another clade containing the PEDV non-S-INDEL strain (Figure 4). The non-S-INDEL clade can be subdivided into two groups, one with several sub clades (Cluster 1) and another monophyletic clade (Cluster 2) (Figure 4). Due to the numerous recombination events throughout the complete genome, distributed across various regions, we were unable to perform time-scaled tree analyses for whole genomes. The occurrence of recombination events violates the assumptions of this analysis and affects coalescence times as well as evolutionary rates.

Recombination events were identified in the spike protein region, as previously mentioned. For subsequent analyses, all sequences presenting evidence of recombination were excluded, including those identified solely within the complete genome dataset as well as the Spike protein analysis. A particularly conservative approach was adopted, whereby any sequence showing recombination evidence was removed, irrespective of the number of detection methods that confirmed the events, resulting in a dataset of 320 sequences from the non-S-INDEL strain and 43 sequences from the S-INDEL strain. As mentioned in the previous section, only four of the most recent sequences showed evidence of recombination, while the others were all generated before 2017.



**Figure 4.** Maximal likelihood phylogenetic tree based on the full genome nucleotide sequences of PEDV. The colored circles on the nodes represent bootstrap support values for the branches: black circles indicate support values greater than 75%, while red circles represent branches with support values below 75%. Sequences belonging to PEDV S-INDEL strain are indicated by the red bar. The PEDV non-S-INDEL strains are divided into two clusters shown on the colored bar: Cluster 1 in blue and Cluster 2 in green.

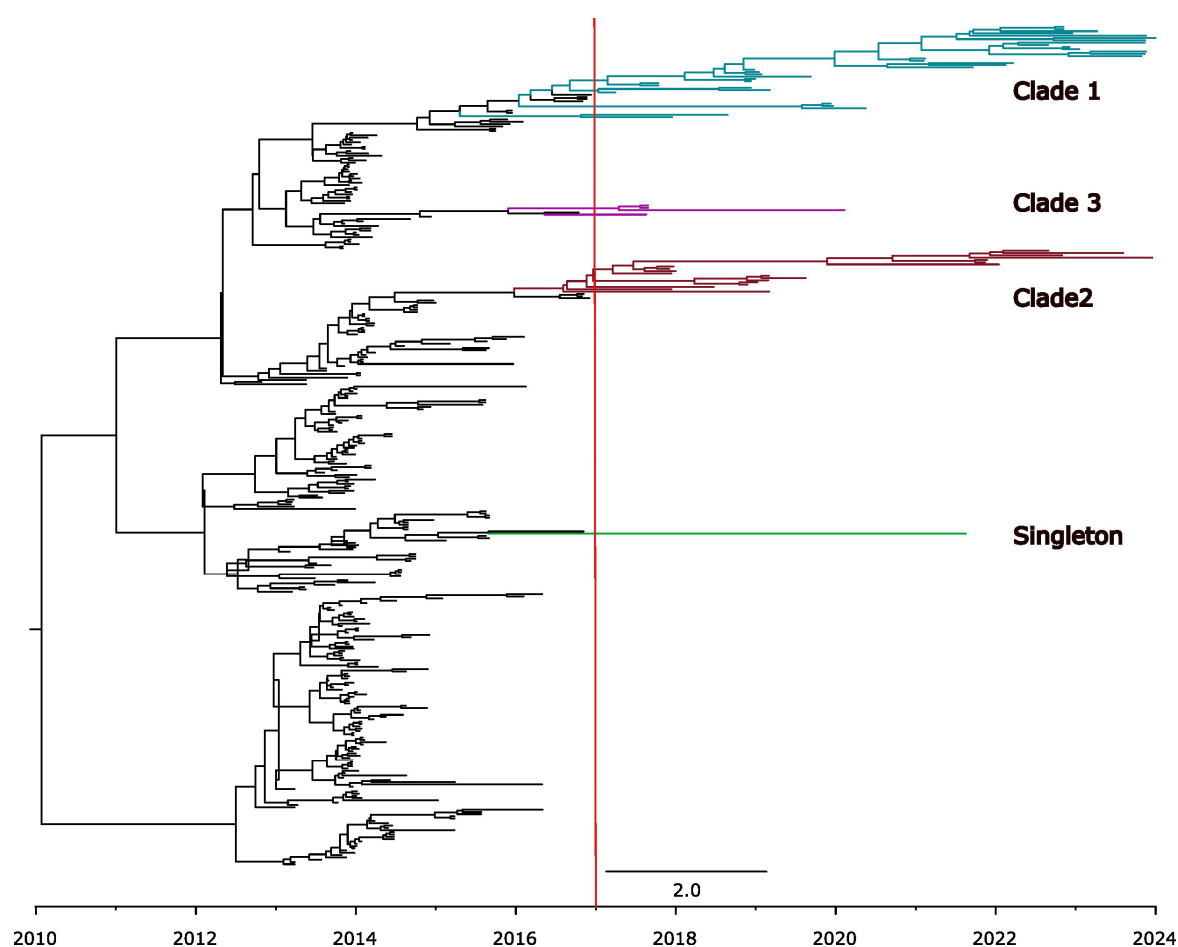
Maximum-likelihood trees were used to conduct root-to-tip regression analyses in TempEst (53) as an exploratory approach to assess evidence of temporal signal. Both strains demonstrated reasonable evidence of a temporal signal: the PEDV S-INDEL strain showed a correlation of 0.769 with an  $R^2$  value of 0.593, while the PEDV non-S-INDEL strain exhibited a correlation of 0.785 and an  $R^2$  value of 0.594.

To formally assess the temporal signal, we conducted a Bayesian Evaluation of Temporal Signal (BETS) (54, 55). In all BETS analyses, the (log) Bayes Factor exceeded 5, favoring the heterochronous model over the isochronous model and providing significant evidence for a temporal signal in both strains (Table S1). The strict molecular clock was the best-fit for the PEDV S-INDEL strain, while the uncorrelated relaxed clock lognormal model was preferred for the PEDV non-S-INDEL strain.

The mean nucleotide substitution rate was estimated at  $7.01 \times 10^{-4}$  subs/site/year (95% HPD:  $3.27 \times 10^{-4}$  to  $1.16 \times 10^{-3}$ ) for the PEDV S-INDEL strain spike protein and  $1.41 \times 10^{-3}$  subs/site/year (95% HPD:  $1.19 \times 10^{-3}$  to  $1.65 \times 10^{-3}$ ) for the PEDV non-S-INDEL strain. The PEDV S-INDEL strain has a

TMRCAs around 1991 (95% HPD: 1955.0 to 2012.6), while the PEDV non-S-INDEL strain presented a TMRCAs around 2010 (95% HPD: 2006.9 to 2012.2) (Figure S1).

In both analyses, TempEst and BETS were able to detect evidence of a temporal signal, reinforcing their strength. Interestingly, most sequences of the PEDV non-S-INDEL strain sampled after 2017 clustered into two monophyletic subclades (Figure 5). Clade 1 has a TMRCAs estimated at 2017.2 (95% HPD: 2017.4 to 2018.7), while Clade 2 has a TMRCAs estimated at 2016.0 (95% HPD: 2015.1 to 2016.6). We detected few descendants from other clades that were circulating during the epidemic period. Only singletons and other small Clade 3 were identified (Figure 5), both observed as radiating from other ancestral clades, but after 2021, they were no longer detected. We speculate that most ancestral clades may have been displaced by the more recent clades and are no longer circulating in the U.S.



**Figure 5.** Time-scaled tree of PEDV spike protein sequences from PEDV non-S-INDEL strains. The scale at the bottom of the tree represents the time-scale in years. The earlier clades from the epidemic-period are colored in black, while the clades currently circulating in the USA are highlighted in different colors: clade 1 in blue, clade 2 in red, clade 3 in pink, and singleton 1 in green. The red line indicates the year 2017, marking the point when the composition of the PEDV non-indele strain population changes, and the ancestral clades are displaced by the current ones.

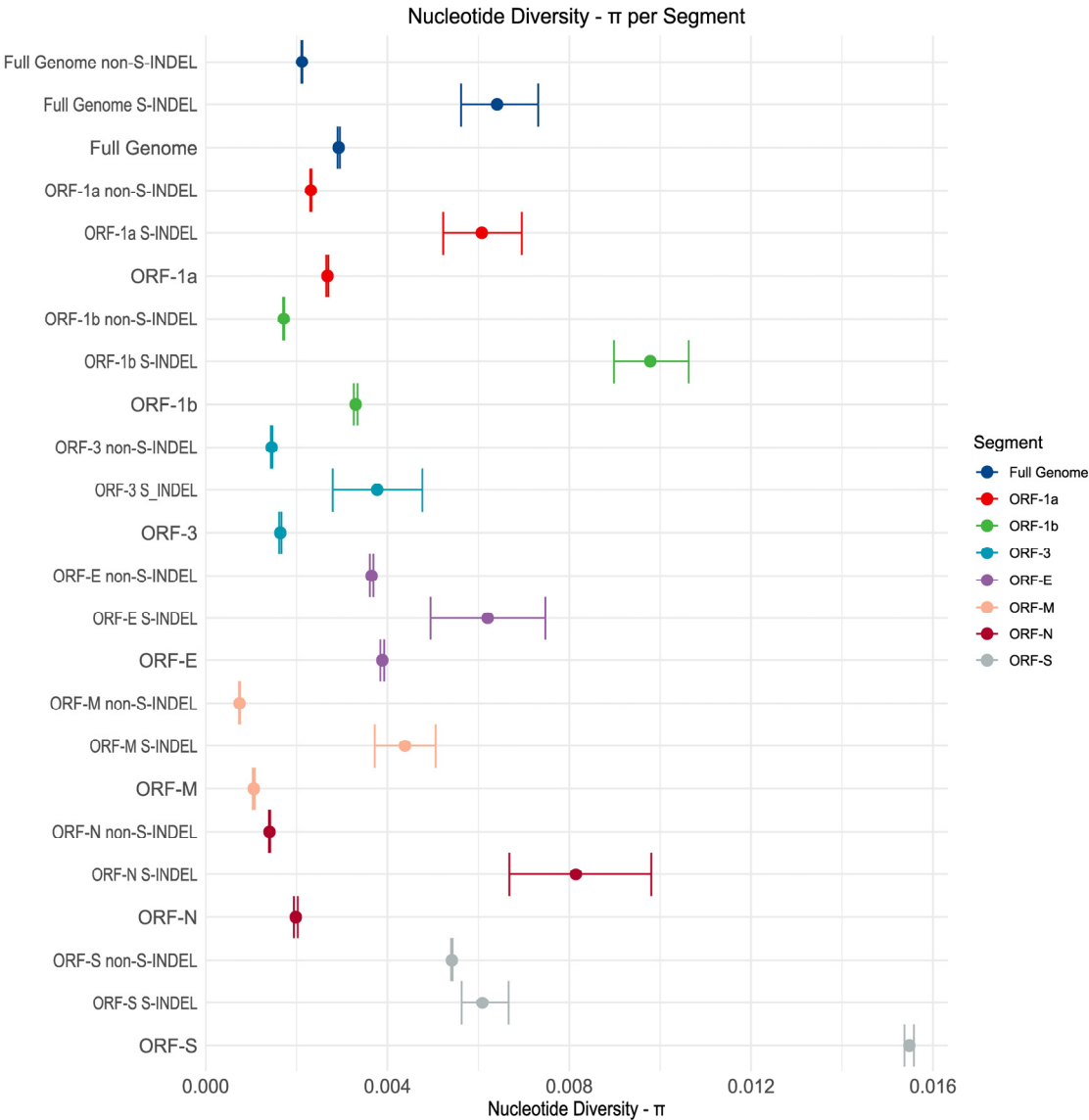
#### *Genetic Variability in U.S. Population*

Genetic variability was assessed for the complete set of sequences and for each strain individually. Nucleotide diversity was estimated using a Python script (61) for both complete genomes and individual coding regions (Figure 6). When analyzing the complete dataset (S-INDEL and non-S-INDEL combined), we found that the Spike (S) protein exhibited the highest nucleotide diversity by far (0.015), followed by ORF1b (0.003), ORF1a (0.002), Nucleocapsid (N) (0.002), ORF3

(0.001), and finally, the Matrix (M) protein (0.001). All showed significant differences in their mean  $\pi$  values (Figure 6).

The PEDV non-S-INDEL strain exhibited the highest nucleotide diversity in the Spike (S) protein (0.005), followed by the Envelope (E) protein (0.003), ORF1a (0.002), ORF1b (0.001718335), ORF3 (0.001), Nucleocapsid (N) (0.001), and finally, the Matrix (M) protein (0.0007). All of these differences were statistically significant (Figure 6).

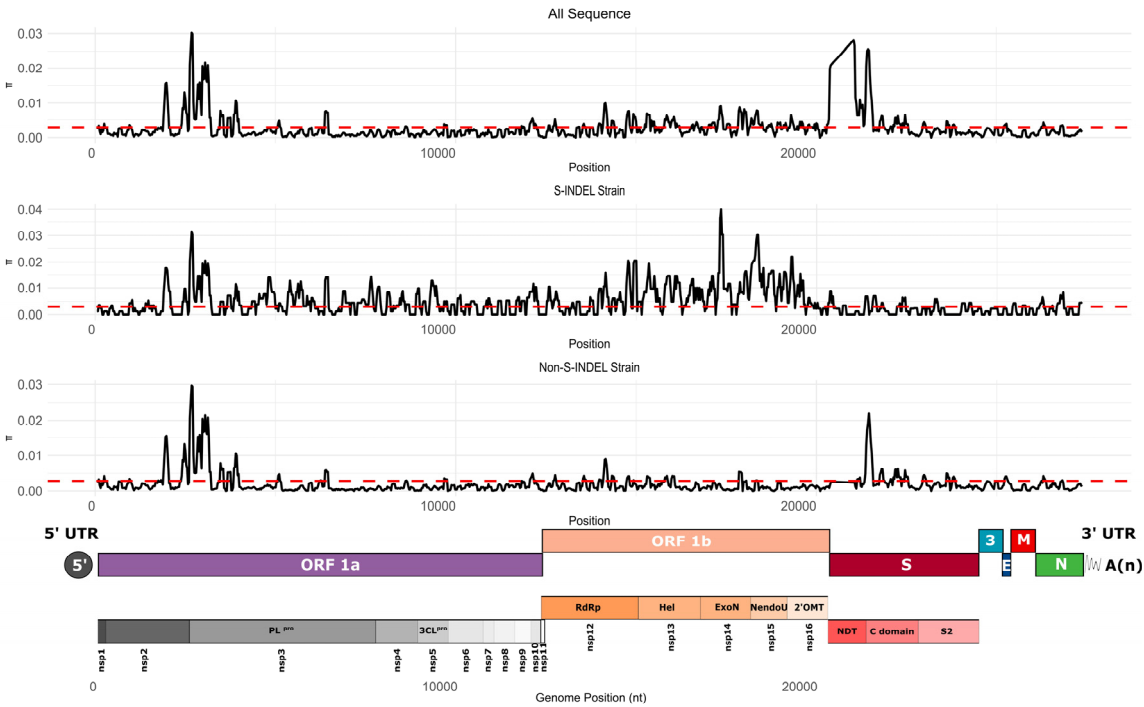
Although the number of sequences for the S-INDEL strain was low, it exhibited significantly higher nucleotide diversity compared to the non-S-INDEL strain, both at the full genome level and for each individual ORF analyzed (Figure 6). For the PEDV S-INDEL strain, the highest nucleotide diversity was observed in the ORF1b (0.009) and Nucleocapsid (N) (0.008) regions (Figure 6), with no statistically significant differences between them. These were followed by ORF1a (0.006), Envelope (E) (0.006), and Spike (S) (0.006), which also showed no statistically significant differences and overlapping confidence intervals with ORF N. The lowest nucleotide diversity was observed in the Matrix (M) (0.004) and ORF3 (0.003) regions (Figure 6), with no significant differences between these two genes either (Figure 6).



**Figure 6.** Genetic variability based on complete genomes of PEDV. Nucleotide diversity ( $\pi$ ) values were estimated for the complete dataset and for each strain individually. We also estimated genetic diversity for each of the ORFs and for the complete dataset. In the graph, points represent the mean, and bars represent the 95% confidence intervals estimated through the bootstrap test.

To better understand how genetic diversity is distributed across the PEDV genome, we measured nucleotide diversity by site using a sliding window approach (100-nt window with 25-nt steps). The analyses were performed on the complete set of sequences as well as for each strain separately. We identified several peaks of nucleotide diversity that exceeded the mean across the genome (Figure 7). Overall, the distribution of nucleotide diversity along the genome was similar between the complete dataset and the PEDV non-S-INDEL strain. Both exhibited hypervariable regions in the 5' end of the genome, specifically within the nsp1 and nsp3 regions. Another hypervariable region was located in the S1 subunit of the Spike (S) protein, particularly in the C-domain (Figure 7). The dataset combining both PEDV S-INDEL and PEDV non-S-INDEL strains revealed an additional hypervariable region in the NTD of the S1 subunit of the Spike protein. This pattern was not observed in the PEDV non-S-INDEL strain, suggesting that hypervariable regions may vary by strain, reflecting distinct evolutionary trajectories between them (Figure 7).

The PEDV S-INDEL strain also displayed a hypervariable region within the nsp2 and nsp3 regions, as well as in ORF1b, particularly in the nsp14, nsp15, and nsp16 regions. In contrast, the hypervariable regions in the S1 subunit of the Spike protein observed in the PEDV non-S-INDEL strain were not detected in the S-INDEL strain (Figure 7).



**Figure 7.** Average nucleotide diversity per site across the full genome, computed using a sliding window of 100 nucleotides with steps of 25 nucleotides. The red dotted line represents the overall mean nucleotide diversity. Nucleotide diversity was calculated for each strain individually as well as for the complete dataset. At the bottom of the graph, the genomic organization map of PEDV is shown.

*High Sequence Similarity Between PEDV Strains and Among the Current Clades*

Complete genome analyses revealed a high nucleotide identity percentage for the two PEDV strains, above 98% for both intra-strain and inter-strain comparisons (Table 2). Sequence comparisons were also performed for datasets of the Spike protein. A high pairwise identity percentage was observed within the non-S-INDEL strain, with a mean of 99.1% (Table 2). We computed identity percentages within and between the two current clades, as well as between the clades that circulated before 2017. The mean pairwise identity percentages within the new clades was 99.3% for Clade 1 and 99.2% for Clade 2 (Table 2). However, comparisons between Clades 1 and 2, as well as between

these clades and those that circulated before 2017, showed mean pairwise identity values ranging between 98.6 and 99.3% , respectively (Table 2). The S-INDEL strain exhibited intra-strain identity percentages above 94.5%, with an average pairwise identity of 99.1%. Pairwise identity percentages between the S-INDEL and non-S-INDEL strains ranged from 91.1% to 94.3% (Table 2). These results indicate that the sequences within each clade are genetically very similar. However, there is a greater genetic distance between the S-INDEL and non-S-INDEL clades in the Spike protein.

**Table 2.** Pairwise nucleotide identity computed for the spike protein coding region, values were estimated using SDT.v1.2. Genetic distances were computed within and between clades. The table indicates the highest, lowest and average values, respectively.

| Clade                                                                | Max     | Min    | Average |
|----------------------------------------------------------------------|---------|--------|---------|
| Spike Protein                                                        |         |        |         |
| Non-S-INDEL                                                          | 99.8 %  | 97.6 % | 99.1%   |
| Clade1 <sup>non-S-INDEL</sup>                                        | 99.8 %  | 98.2 % | 99.3%   |
| Clade1 <sup>non-S-INDEL</sup> vs older clades <sup>non-S-INDEL</sup> | 99.6 %  | 97.7 % | 98.6%   |
| Clade2 <sup>non-S-INDEL</sup>                                        | 99.9 %  | 99.3 % | 99.2%   |
| Clade2 <sup>non-S-INDEL</sup> vs older clades <sup>non-S-INDEL</sup> | 99.7 %  | 97.8 % | 99.3%   |
| Clade1 <sup>non-S-INDEL</sup> vsc Clade 2 <sup>non-S-INDEL</sup>     | 99.0 %  | 97.9%  | 98.6%   |
| S-INDEL strain                                                       | 100.0 % | 94.5 % | 99.1%   |
| S-INDEL vs non-S-INDEL strains                                       | 94.25%  | 91.14% | 93,61%  |
| Full Genome                                                          |         |        |         |
| Non-S-INDEL strain                                                   | 100%    | 98.75% | 99.78%  |
| INDEL strain                                                         | 100%    | 99.21% | 99.55%  |
| INDEL vs Non-INDEL strains                                           | 99.85%  | 98.52% | 98.99%  |

### Positive Selection Evidence

Selection analyses were conducted to infer how natural selection modulates variation in regions encoding structural proteins, which are the main targets of the immune system (69). For this, we estimated the rates of synonymous and non-synonymous substitutions (dN/dS) for the ORFs S, E, M, and N. For both PEDV S-INDEL and PEDV non-S-INDEL strains, the dN/dS values were less than 1, highlighting the action of purifying selection on these genes (Table 3). The dN/dS ratios varied across the genome, with purifying selection acting more strictly on ORF M and more relaxed on ORFs S, E, and N, respectively. Although the dN/dS values were lower for the PEDV S-INDEL strain, the small sample size could have affected this estimation (Table 3).

The spike protein exhibited the highest number of sites under both purifying and diversifying selection, followed by the nucleocapsid protein. The matrix and envelope proteins exhibited very few sites under selection (Table 3). This pattern was observed for both strains. Of all the sites under diversifying selection, only one was detected in a known epitope of the spike protein (position 723), resulting in a serine-to-asparagine substitution.

**Table 3.** Nonsynonymous to Synonymous Substitution Ratios (dN/dS) and Selection Analysis of PEDV Structural Proteins. The table presents the dN/dS ratios calculated using the SLAC method and the number of negatively and positively selected sites identified using three distinct methods: SLAC, MEME, and FUBAR. These analyses were performed for each structural protein of PEDV. A dN/dS ratio < 1 suggests purifying selection, a dN/dS ratio = 1 indicates neutral evolution and a dN/dS ratio > 1 suggests positive selection.

| Gene           | dN/dS  | SLAC                                              |                                                                                                                             | MEME                                                                                             |           | FUBAR                                                                                                                                        |                                                                                                                                  |
|----------------|--------|---------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
|                |        | diversifying                                      | purifying                                                                                                                   | diversifying                                                                                     | purifying | diversifying                                                                                                                                 | purifying                                                                                                                        |
| S-INDEL strain |        |                                                   |                                                                                                                             |                                                                                                  |           |                                                                                                                                              |                                                                                                                                  |
| Spike          | 0.331  | -                                                 | 368, 659                                                                                                                    | 27, 83, 240,<br>351, 429, 500,<br>632                                                            | -         | 83, 196, 310,<br>351, 719                                                                                                                    | 21, 41, 44,<br>48,<br>73, 76, 92,<br>93,<br>125, 141,<br>149,<br>199, 226,<br>237,<br>238, 269,<br>312                           |
| Envelop        | 0.491  | -                                                 | -                                                                                                                           | -                                                                                                | -         | 66                                                                                                                                           | -                                                                                                                                |
| Matrix         | 0.0967 | -                                                 | -                                                                                                                           | -                                                                                                | -         | -                                                                                                                                            | 27, 41, 71,<br>78, 116,<br>121,<br>122, 188,<br>201                                                                              |
| Nucleo capsid  | 0.240  | -                                                 | -                                                                                                                           | 27, 54, 415                                                                                      | -         | 27                                                                                                                                           | 28, 43, 51,<br>140, 147,<br>207,<br>211, 249,<br>267,<br>271, 298,<br>327, 360,<br>364, 414                                      |
| Non-S-INDEL    |        |                                                   |                                                                                                                             |                                                                                                  |           |                                                                                                                                              |                                                                                                                                  |
| Spike          | 0.887  | 144, 380, 488,<br>525, 526, 568,<br>610, 614, 724 | 80, 94, 100,<br>101, 114,<br>139, 154,<br>209, 276,<br>359, 374,<br>462, 558,<br>582, 588,<br>621, 654,<br>664, 697,<br>729 | 24, 58, 144,<br>277, 355, 380,<br>412, 417, 433,<br>488, 495, 525,<br>526, 568, 610,<br>614, 724 | -         | 24, 55, 58, 70,<br>144, 157, 196,<br>277, 355, 380,<br>412, 433, 488,<br>494, 501, 525,<br>526, 568, 610,<br>614, 637, 653,<br>676, 695, 724 | 53, 94, 100,<br>101, 109,<br>139, 154,<br>244, 276,<br>359, 374,<br>394, 462,<br>468, 558,<br>582, 621,<br>625, 654,<br>697, 729 |
| Envelop        | 0.743  | -                                                 | -                                                                                                                           | -                                                                                                | -         | -                                                                                                                                            | -                                                                                                                                |
| Matrix         | 0.195  | -                                                 | -                                                                                                                           | -                                                                                                | 200       | -                                                                                                                                            | -                                                                                                                                |
| Nucleo capsid  | 0.585  | -                                                 | 240                                                                                                                         | -                                                                                                | -         | 193                                                                                                                                          | 55, 87, 100,<br>190, 240,<br>244,<br>252, 386                                                                                    |

4. Discussion

The introduction of PEDV into the U.S. caused significant economic losses to the U.S. swine industry and affected pork availability in the market (13, 70, 71). The high transmission rate, combined with the movement of animals within the country, enabled the virus to spread rapidly nationwide (72). Studies identifying factors that increase the risk of outbreaks have contributed to

improving biosecurity protocols guiding animal management strategies (9, 73-75). The reduction in breeding herd incidence cumulative reflects the results of these efforts (9); however, PEDV continues to circulate in swine herds in the U.S. particularly during the colder seasons of the year (7). Here, we reconstructed the evolutionary trajectory of PEDV over the 10 years following its introduction.

Our results suggest that the PEDV non-S-INDEL is the predominant strain in the U.S. although it is important to consider the potential bias in our dataset. There is a substantial gap in sequence surveillance during the post-epidemic period since sequencing efforts declined drastically after 2015. Although there is no evidence to support this, the possibility that PEDV S-INDEL is silently circulating within livestock populations cannot be ruled out, given that the PEDV S-INDEL strain is less pathogenic. Previous studies show that the non-S-INDEL PEDV strain has much higher transmission rates compared to the S-INDEL PEDV strain, which may contribute to our finding that the S-INDEL strain is less widespread (72).

PEDV populations continue to differentiate over time. Even though PEDV Spike protein exhibits low intra-strain genetic variability, strong evidence of temporal signal was detected. Similar patterns have also been observed in SARS-CoV-2 populations (55). Our phylogenetic analyses based on the S1 subunit of the Spike protein suggest that many of the early clades circulating during the epidemic period are no longer present in the U.S. Many of older clades appear to have left no descendants. Based on the available data, nearly all sequences sampled after 2017 clustered into two current clades. The persistence of these new clades could be explained by differential adaptability or by successive population bottlenecks and expansions, typical of seasonal respiratory diseases, which randomly contribute to the fixation of certain genotypes and the loss of others within the population (76, 77). Information regarding the pathogenicity and transmissibility of the PEDV genotypes currently circulating in the United States is yet to be uncovered, but they deserve further investigation.

Interestingly, the circulation of these two contemporary clades is restricted to two separate geographic regions, suggesting compartmentalized circulation within those regions with limited spread between different regions during this time. This pattern seems plausible, as one of the main factors associated with virus dispersal in swine in the U.S. is animal movement between regions(78). The observation of compartmentalized spread within geographic regions suggests that if PEDV were to be regionally eliminated, then the risk of re-introduction is relatively small. This is particularly relevant given ongoing discussions of eliminating PEDV from the U.S. swine (79).

We identified several hypervariable regions throughout the genome of the U.S. PEDV population. When analyzing each strain separately, we observed intricate patterns of hypervariation specific to each strain. The first hypervariable region is located in ORF1a, identified in both strains, spanning the carboxy-terminal region of nsp2 and the amino-terminal region of nsp3. The high variability in this region can be explained by evidence of recombination events in this portion of the genome. Multiple recombination breakpoints were detected in this genomic region. Another hypervariable region was detected exclusively in the non-S-INDEL strain and is located in the S1 subunit of the Spike protein. A similar pattern of genetic variation in this region, also associated with recombination events, has been previously reported (80, 81). The S-INDEL strain, on the other hand, exhibited a hypervariable region in the ORF1b region. The high nucleotide diversity in this region is associated with evidence of recombination in that region (Figure 7 Table 1A).

Recombination may play a central role in the diversification of PEDV. Recombination events located in regions associated with immune response deserve special attention, as they may result in the formation of new epitope profiles, leading to immune escape, or potentially contributing to increased pathogenicity and transmission rates, as well as threatening the durability of the vaccines available on the market and vaccine development, as well (80-85). The complete genome analysis revealed some recombination events that were not detected in the Spike protein-only analysis, highlighting the power of complete genomes to more accurately capture recombination events.

Several sites of positive selection were identified in the ORF S region; however, only one of these sites is located within a B-cell epitope (SE16, 722SSTFNSTREL731), warranting special attention as this epitope has been described as having a high antigenic index as well as hydrophilicity, making it

likely to interact with the host's immune system (86). Furthermore, the relevance of the N723S amino acid substitution extends as this is a potential region for vaccine development (86), and amino acid changes in this region may pose a direct challenge to vaccine stability (87).

A limitation of our findings is that PEDV sequence data is not routinely generated, and thus we do not have a full picture of PEDV diversity across the U.S. While sequencing effort during the endemic period has been relatively constant (e.g., 1-3 sequences available per 10 reported sow farm outbreaks), it is not clear whether sequencing efforts are proportional to regional incidence patterns (9), or whether sequencing is unevenly pursued across regions or in different production systems. Continued sequence surveillance is vital for the swine industry to advance toward disease eradication and is key to a) confirming the extinction of older clades, b) mapping the distribution of recent clades, and c) understanding PEDV's evolutionary diversification.

## 5. Conclusions

Our results provide an updated perspective on the evolutionary state of PEDV in the United States over the past decade. The PEDV strains currently circulating in the U.S. are slightly distinct from those present during the epidemic period. Several recombination events have been detected, some of which are widely fixed within North American populations, suggesting an adaptive advantage. The two recent clades reported here are geographically restricted to specific regions, indicating limited epidemiological connectivity leading to exchange of PEDV viruses between these areas. However, the relatively small number of available sequences post-epidemic may represent a potential bias in our dataset. It is crucial to resume more systematic PEDV sequence surveillance, particularly to generate additional complete genomes. This will provide a clearer understanding of the distribution of new clades across different regions and confirm the extinction of pre-existing clades. Such information is vital to achieving the eradication of PEDV in the United States.

## 6. Acknowledgements

The authors would like to thank the Morrison Swine Health Monitoring Project (MSHMP) participants and the Veterinary Diagnostic Laboratory, University of Minnesota, for sharing its PEDV genetic sequences, as well as the Swine health Information Center (SHIC) for funding the MSHMP.

## References

1. Wang X-m, Niu B-b, Yan H, Gao D-s, Yang X, Chen L, Chang H-t, Zhao J, Wang C-q. 2013. Genetic properties of endemic Chinese porcine epidemic diarrhea virus strains isolated since 2010. *Archives of virology* 158:2487-2494.
2. Stadler J, Zoels S, Fux R, Hanke D, Pohlmann A, Blome S, Weissenböck H, Weissenbacher-Lang C, Ritzmann M, Ladinig A. 2015. Emergence of porcine epidemic diarrhea virus in southern Germany. *BMC veterinary research* 11:1-8.
3. Stevenson GW, Hoang H, Schwartz KJ, Burrough ER, Sun D, Madson D, Cooper VL, Pillatzki A, Gauger P, Schmitt BJ. 2013. Emergence of Porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *Journal of veterinary diagnostic investigation* 25:649-654.
4. Zhang H, Zou C, Peng O, Ashraf U, Xu Q, Gong L, Fan B, Zhang Y, Xu Z, Xue C. 2023. Global dynamics of porcine enteric coronavirus PEDV epidemiology, evolution, and transmission. *Molecular Biology and Evolution* 40:msad052.
5. Ojkic D, Hazlett M, Fairles J, Marom A, Slavic D, Maxie G, Alexandersen S, Pasick J, Alsop J, Burlatschenko S. 2015. The first case of porcine epidemic diarrhea in Canada. *The Canadian Veterinary Journal* 56:149.
6. Vlasova AN, Marthaler D, Wang Q, Culhane MR, Rossow KD, Rovira A, Collins J, Saif LJ. 2014. Distinct characteristics and complex evolution of PEDV strains, North America, May 2013–February 2014. *Emerging infectious diseases* 20:1620.
7. Song D, Park B. 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus genes* 44:167-175.

8. Wang L, Byrum B, Zhang Y. 2014. New variant of porcine epidemic diarrhea virus, United States, 2014. *Emerging infectious diseases* 20:917.
9. Makau DN, Pamornchainavakul N, VanderWaal K, Kikuti M, Picasso-Risso C, Geary E, Corzo CA. 2024. Postepidemic Epidemiology of Porcine Epidemic Diarrhea Virus in the United States. *Transboundary and Emerging Diseases* 2024:5531899.
10. Yue X, Kikuti, M., Melini, C.M., Corzo, C. A. 2024. PEDV: Sow herds are eliminating the virus faster. *National Hog Farmer*.
11. Project MSHM. 2024. PED aggregate prevalence of sow herd status (last updated April 2024). Report | Morrison Swine Health Monitoring Project (2024) <https://mshmpumnedu/reports#Charts> [Accessed June 21, 2024].
12. Jung K, Saif LJ. 2015. Porcine epidemic diarrhea virus infection: Etiology, epidemiology, pathogenesis and immunoprophylaxis. *The Veterinary Journal* 204:134-143.
13. O'Dea EB, Snelson H, Bansal S. 2016. Using heterogeneity in the population structure of US swine farms to compare transmission models for porcine epidemic diarrhoea. *Scientific Reports* 6:22248.
14. Schumacher LL, Huss AR, Cochrane RA, Stark CR, Woodworth JC, Bai J, Poulsen EG, Chen Q, Main RG, Zhang J. 2017. Characterizing the rapid spread of porcine epidemic diarrhea virus (PEDV) through an animal food manufacturing facility. *PLoS One* 12:e0187309.
15. Kim Y, Yang M, Goyal SM, Cheeran MC, Torremorell M. 2017. Evaluation of biosecurity measures to prevent indirect transmission of porcine epidemic diarrhea virus. *BMC veterinary research* 13:1-9.
16. Dee S, Clement T, Schelkopf A, Nerem J, Knudsen D, Christopher-Hennings J, Nelson E. 2014. An evaluation of contaminated complete feed as a vehicle for porcine epidemic diarrhea virus infection of naive pigs following consumption via natural feeding behavior: proof of concept. *BMC veterinary research* 10:1-9.
17. Woo PC, de Groot RJ, Haagmans B, Lau SK, Neuman BW, Perlman S, Sola I, van der Hoek L, Wong AC, Yeh S-H. 2023. ICTV Virus Taxonomy Profile: Coronaviridae 2023: This article is part of the ICTV Virus Taxonomy Profiles collection. *Journal of general virology* 104:001843.
18. Elena SF, Sanjuán R. 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *Journal of virology* 79:11555-11558.
19. Makau DN, Lycett S, Michalska-Smith M, Paploski IA, Cheeran MC-J, Craft ME, Kao RR, Schroeder DC, Doeschl-Wilson A, VanderWaal K. 2022. Ecological and evolutionary dynamics of multi-strain RNA viruses. *Nature Ecology & Evolution* 6:1414-1422.
20. Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics* 9:267-276.
21. Dolan PT, Whitfield ZJ, Andino R. 2018. Mechanisms and concepts in RNA virus population dynamics and evolution. *Annual Review of Virology* 5:69-92.
22. Minskaia E, Hertzog T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J. 2006. Discovery of an RNA virus 3' → 5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proceedings of the National Academy of Sciences* 103:5108-5113.
23. Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *Journal of virology* 81:12135-12144.
24. Brown DW. 1997. Threat to humans from virus infections of non-human primates. *Reviews in medical virology* 7:239-246.
25. Gibbs MJ, Weiller GF. 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proceedings of the National Academy of Sciences* 96:8022-8027.
26. Simon-Loriere E, Holmes EC. 2011. Why do RNA viruses recombine? *Nature Reviews Microbiology* 9:617-626.
27. Wells HL, Bonavita CM, Navarrete-Macias I, Vilchez B, Rasmussen AL, Anthony SJ. 2023. The coronavirus recombination pathway. *Cell host & microbe* 31:874-889.
28. Müller NF, Kistler KE, Bedford T. 2022. A Bayesian approach to infer recombination patterns in coronaviruses. *Nature communications* 13:4186.

29. Mei X, Guo J, Fang P, Ma J, Li M, Fang L. 2023. The characterization and pathogenicity of a recombinant porcine epidemic diarrhea virus variant ECQ1. *Viruses* 15:1492.
30. Jarvis MC, Lam HC, Zhang Y, Wang L, Hesse RA, Hause BM, Vlasova A, Wang Q, Zhang J, Nelson MI. 2016. Genomic and evolutionary inferences between American and global strains of porcine epidemic diarrhea virus. *Preventive Veterinary Medicine* 123:175-184.
31. Huang Y-W, Dickerman AW, Piñeyro P, Li L, Fang L, Kiehne R, Opriessnig T, Meng X-J. 2013. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *MBio* 4:10.1128/mbio.00737-13.
32. Lee C. 2015. Porcine epidemic diarrhea virus: an emerging and re-emerging epizootic swine virus. *Virology journal* 12:1-16.
33. Wang Q, Vlasova AN, Kenney SP, Saif LJ. 2019. Emerging and re-emerging coronaviruses in pigs. *Current opinion in virology* 34:39-49.
34. Jung K, Saif LJ, Wang Q. 2020. Porcine epidemic diarrhea virus (PEDV): An update on etiology, transmission, pathogenesis, and prevention and control. *Virus research* 286:198045.
35. Lee S, Lee C. 2014. Outbreak-related porcine epidemic diarrhea virus strains similar to US strains, South Korea, 2013. *Emerging Infectious Diseases* 20:1223.
36. Lin C-N, Chung W-B, Chang S-W, Wen C-C, Liu H, Chien C-H, Chiou M-T. 2014. US-like strain of porcine epidemic diarrhea virus outbreaks in Taiwan, 2013–2014. *Journal of Veterinary Medical Science* 76:1297-1299.
37. He W-T, Bollen N, Xu Y, Zhao J, Dellicour S, Yan Z, Gong W, Zhang C, Zhang L, Lu M. 2022. Phylogeography reveals association between swine trade and the spread of porcine epidemic diarrhea virus in China and across the world. *Molecular Biology and Evolution* 39:msab364.
38. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2019. GenBank. *Nucleic acids research* 47:D94-D99.
39. Perez AM, Linhares DC, Arruda AG, VanderWaal K, Machado G, Vilalta C, Sanhueza JM, Torrison J, Torremorell M, Corzo CA. 2019. Individual or common good? Voluntary data sharing to inform disease surveillance systems in food animals. *Frontiers in Veterinary Science* 6:194.
40. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30:772-780.
41. Muhire BM, Varsani A, Martin DP. 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PloS one* 9:e108277.
42. Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562-563.
43. Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265:218-225.
44. Martin DP, Posada D, Crandall K, Williamson C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research & Human Retroviruses* 21:98-102.
45. Smith JM. 1992. Analyzing the mosaic structure of genes. *Journal of molecular evolution* 34:126-129.
46. Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences* 98:13757-13762.
47. Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16:573-582.
48. Lam HM, Ratmann O, Boni MF. 2018. Improved algorithmic complexity for the 3SEQ recombination detection algorithm. *Molecular biology and evolution* 35:247-251.
49. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B. 2021. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution* 7:veaa087.
50. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* 23:254-267.

51. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453-4455.
52. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular biology and evolution* 37:291-294.
53. Rambaut A, Lam T, Max Carvalho L, Pybus O. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2: vew007.
54. Duchene S, Lemey P, Stadler T, Ho SY, Duchene DA, Dhanasekaran V, Baele G. 2020. Bayesian evaluation of temporal signal in measurably evolving populations. *Molecular Biology and Evolution* 37:3363-3379.
55. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus evolution* 6:veaa061.
56. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016.
57. Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the american statistical association* 90:773-795.
58. Baele G, Lemey P, Suchard MA. 2016. Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. *Systematic Biology* 65:250-264.
59. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology* 67:901-904.
60. Nei M. 1987. *Molecular evolutionary genetics*. Columbia university press.
61. Lima AT, Silva JC, Silva FN, Castillo-Urquiza GP, Silva FF, Seah YM, Mizubuti ES, Duffy S, Zerbini FM. 2017. The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus evolution* 3:vex005.
62. Peng RD. 2008. simpleboot: Simple Bootstrap Routines.
63. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34:3299-3302.
64. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22:3096-3098.
65. Kosakovsky Pond SL, Frost SD. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* 22:1208-1222.
66. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS genetics* 8:e1002764.
67. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution* 30:1196-1205.
68. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution* 35:773-777.
69. Li X, Wu Y, Yan Z, Li G, Luo J, Huang S, Guo X. 2024. A Comprehensive View on the Protein Functions of Porcine Epidemic Diarrhea Virus. *Genes* 15:165.
70. Weng L, Weersink A, Poljak Z, de Lange K, von Massow M. 2016. An economic evaluation of intervention strategies for Porcine Epidemic Diarrhea (PED). *Preventive Veterinary Medicine* 134:58-68.
71. Kimberly V, Andres P, Montse T, Meggan C. 2018. Role of animal movement and indirect contact among farms in transmission of porcine epidemic diarrhea virus.
72. Gallien S, Andraud M, Moro A, Lediguerher G, Morin N, Gauger PC, Bigault L, Paboeuf F, Berri M, Rose N. 2018. Better horizontal transmission of a US non-InDel strain compared with a French InDel strain of porcine epidemic diarrhoea virus. *Transboundary and emerging diseases* 65:1720-1732.
73. Kikuti M, Drebes D, Robbins R, Dufresne L, Sanhueza JM, Corzo CA. 2022. Growing pig incidence rate, control and prevention of porcine epidemic diarrhea virus in a large pig production system in the United States. *Porcine Health Management* 8:23.

74. Lowe J, Gauger P, Harmon K, Zhang J, Connor J, Yeske P, Loula T, Levis I, Dufresne L, Main R. 2014. Role of transportation in spread of porcine epidemic diarrhea virus infection, United States. *Emerging infectious diseases* 20:872.
75. Bowman AS, Krogwold RA, Price T, Davis M, Moeller SJ. 2015. Investigating the introduction of porcine epidemic diarrhea virus into an Ohio swine operation. *BMC veterinary research* 11:1-7.
76. Narechania A, Bobo D, Deitz K, DeSalle R, Planet PJ, Mathema B. 2024. Rapid SARS-CoV-2 surveillance using clinical, pooled, or wastewater sequence as a sensor for population change. *Genome Research* 34:1651-1660.
77. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B. 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182:812-827. e19.
78. Cardenas NC, Valencio A, Sanchez F, O'Hara KC, Machado G. 2024. Analyzing the intrastate and interstate swine movement network in the United States. *Preventive Veterinary Medicine* 230:106264.
79. Niederwerder M, Hesse R. 2018. Swine enteric coronavirus disease: a review of 4 years with porcine epidemic diarrhoea virus and porcine deltacoronavirus in the United States and Canada. *Transboundary and emerging diseases* 65:660-675.
80. Sun M, Ma J, Wang Y, Wang M, Song W, Zhang W, Lu C, Yao H. 2015. Genomic and epidemiological characteristics provide new insights into the phylogeographical and spatiotemporal spread of porcine epidemic diarrhea virus in Asia. *Journal of clinical microbiology* 53:1484-1492.
81. Tian Y, Yu Z, Cheng K, Liu Y, Huang J, Xin Y, Li Y, Fan S, Wang T, Huang G. 2013. Molecular characterization and phylogenetic analysis of new variants of the porcine epidemic diarrhea virus in Gansu, China in 2012. *Viruses* 5:1991-2004.
82. Sun R, Leng Z, Zhai S-L, Chen D, Song C. 2014. Genetic variability and phylogeny of current Chinese porcine epidemic diarrhea virus strains based on spike, ORF3, and membrane genes. *The scientific world Journal* 2014:208439.
83. Peng Q, Fan B, Song X, He W, Wang C, Zhao Y, Guo W, Zhang X, Liu S, Gao J. 2023. Genetic signatures associated with the virulence of porcine epidemic diarrhea virus AH2012/12. *Journal of Virology* 97:e01063-23.
84. Li D, Li Y, Liu Y, Chen Y, Jiao W, Feng H, Wei Q, Wang J, Zhang Y, Zhang G. 2021. Isolation and identification of a recombinant porcine epidemic diarrhea virus with a novel insertion in S1 domain. *Frontiers in microbiology* 12:667084.
85. Wang H, Zhang L, Shang Y, Tan R, Ji M, Yue X, Wang N, Liu J, Wang C, Li Y. 2020. Emergence and evolution of highly pathogenic porcine epidemic diarrhea virus by natural recombination of a low pathogenic vaccine isolate and a highly pathogenic strain in the spike gene. *Virus evolution* 6:veaa049.
86. Kong N, Meng Q, Jiao Y, Wu Y, Zuo Y, Wang H, Sun D, Dong S, Zhai H, Tong W. 2020. Identification of a novel B-cell epitope in the spike protein of porcine epidemic diarrhea virus. *Virology journal* 17:1-9.
87. Servín-Blanco R, Zamora-Alvarado R, Gevorkian G, Manoutcharian K. 2016. Antigenic variability: obstacles on the road to vaccines against traditionally difficult targets. *Human vaccines & immunotherapeutics* 12:2640-2648.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.