Reconstructing Within-Farm Transmission of a PRRSV-1 Outbreak Using Structured and Unstructured Phylodynamic Models

AUTHOR: L. BALTHAUS (2672860)
DAILY SUPERVISOR: DR. M. MEESTER
VU EXAMINOR: PROF. DR. F. BRUGGEMAN
COURSE ID: XM_0072 (MINOR RESEARCH PROJECT M BSB)
EC POINTS: 30
DATE: 30/01/2026

Abstract

Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) causes significant economic losses and animal welfare concerns globally. While its rapid evolution makes it ideal for phylodynamic inference, the accuracy of these models at the fine-scale, non-random contact environments of individual farms remains unvalidated. By integrating genetic and epidemiological data, phylodynamics offer a distinct advantage over traditional methods, enabling the estimation of transmission parameters during unsampled periods, the quantification of unreported cases, and the reconstruction of the outbreak's origin, independent of clinical observation. This study investigates whether Bayesian birth-death models can accurately reconstruct within-farm transmission dynamics and distinguish spread across spatially distinct units. Building on the original study by Clilverd and colleagues (2023), which used longitudinal incidence monitoring and phylogenetic contact tracing to characterize the epidemic, we applied an unstructured Birth-Death Skyline (BDSKY) model to infer temporal trajectories and a structured Birth-Death with Migration (BDMM) model to resolve pen- and room-level spread.

The BDSKY analysis successfully recovered the established three-phase epidemic trajectory (outbreak, dormancy, resurgence) by the original study and independently validated the timing of the resurgence event. At finer resolution, the BDMM analysis indicated that within-pen transmission was efficient (Re=2.61, 95% Highest Posterior Density (HPD): 0.52–5.21), while spread between pens was significantly constrained within rooms (Re=0.46, 95% HPD: 0.09–0.94) and across rooms (Re=0.06 , 95% HPD: 0.02–0.12), reflecting effective physical compartmentalization. Sensitivity analyses revealed that high-density Open Reading Frame 5 (ORF5) sampling enabled detection of an intense superspreading event, whereas whole-genome sequencing (WGS) provided improved precision for molecular clock calibration.

These results demonstrate that integrating genetic data into birth–death models enables precise reconstruction of epidemic trajectories and robust estimation of compartment-specific transmission rates. This insight supports precision animal health management, complementing traditional epidemiological approaches and potentially reducing the economic impact of PRRSV while improving herd welfare.

Introduction

Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) is globally recognized as one of the most economically significant pathogens in the swine industry, causing severe respiratory and reproductive disorders that greatly compromise the health and welfare of affected pigs (Chandra et al., 2025; Clilverd et al., 2023; Sha et al., 2025). The virus is classified into two distinct species based on their geographic origins: PRRSV-1, first identified in Europe (Wensvoort et al., 1991) and PRRSV-2, first identified in North America (Collins et al., 1992). However, both species share a conserved ~15 kb positive-sense RNA genome with at least 10 open reading frames (ORFs), with ORF1a/b encoding non-structural proteins and ORF2–7 encoding structural proteins (Clilverd et al., 2023). ORF5 sequencing, encoding the highly variable glycoprotein 5, is still commonly used for viral classification (Murtaugh et al., 2010; Weng et al., 2025) and offers a key advantage in its ease of amplification and Sanger sequencing from routine clinical samples (Zhang et al., 2017), despite representing only ~4% of the genome and thus limiting capture of genome-wide diversity (Murtaugh et al., 2010). In contrast, whole-genome sequencing (WGS) provides the advantage of detecting variation across both structural and non-structural regions for comprehensive epidemiological insight (Li et al., 2022).

It is known that PRRSV infection dynamics are highly age-dependent in both genetic variants: neonatally infected pigs act as long-term reservoirs capable of shedding virus for up to 250 days (Butler et al., 2014; Pileri & Mateu, 2016; Rowland et al., 2003; Wills et al., 2003; You et al., 2022), whereas pigs infected post-birth typically clear infection within 42 days (Charpin et al., 2012). Therefore, while vertical transmission often initiates circulation, prolonged shedding by neonatally infected pigs bridges production stages and sustains endemicity through subsequent horizontal spread (Pileri & Mateu, 2016). Furthermore, both species exhibit substantial genetic variability driven by high mutation and recombination rates typical of positive-sense RNA viruses (Franzo et al., 2021; Kappes & Faaberg, 2015; Parisio et al., 2024; Wu et al., 2024; Zhou et al., 2024). Given this exceptional genetic diversity and rate heterogeneity, universal substitution models are often unsuitable for characterizing PRRSV evolution (Balka et al., 2018; Gong et al., 2024). While this fast evolution complicates disease control in PRRSV (Caserta et al., 2023), it renders it particularly suitable for phylodynamic modelling (Duchene et al., 2020; Featherstone et al., 2023). Unlike phylogenetics, which primarily focuses on evolutionary relationships and viral classification (Frias-De-Diego et al., 2021), phylodynamics combines genetic and epidemiological data to quantify pathogen transmission dynamics (Stadler et al., 2024). Central to these dynamics is the effective reproductive number (Re), defined as the expected number of secondary infections caused by a single infected individual at a specific point in time (Stadler et al., 2024). By estimating the Re through time, it becomes possible to assess the epidemic trajectory relative to the epidemiological threshold of 1, distinguishing between periods of growth ($Re > 1$) and decline ($Re < 1$) (Andrade & Duggan, 2022).

In the field of animal health, phylodynamics is particularly valuable for addressing epidemiological questions that are difficult to answer with traditional surveillance data alone, which relies primarily on observational methods like case reporting and contact tracing (Guinat et al., 2021). In particular, phylodynamics can estimate cross-species transmission rates between wild and domestic animals and retrospectively assess the impact of control strategies, such as vaccination, by modeling changes in transmission dynamics over time (Guinat et al., 2021)

Extensive research has successfully characterized transmission dynamics within large-scale farming networks (Sequeira et al., 2025). For PRRSV-2, phylodynamic models have been utilized to infer farm-to-farm transmission pathways and quantify farm-level transmissibility (Pamornchainavakul et al., 2023), as well as to trace the historical migration of viral lineages across different production companies (Alkhamis et al., 2016). Similarly, for PRRSV-1, migration of

lineages has been reconstructed across Europe to identify determinants of spread (Franzo et al., 2022), and the role of pig movements in shaping viral diversity in Central Eastern Europe has been investigated (Balka et al., 2018). Furthermore, Franzo et al. (2021) showed how phylodynamics can track transmission within integrated pig companies, where farrowing, nursery, and finishing units are managed by a single organization.

Complementing these studies, research at the within-farm resolution has largely relied on phylogenetic and evolutionary approaches**.** For instance, phylogenetic analysis has been used to investigate specific introduction and circulation events within a single farm (Li et al., 2022), while the evolutionary dynamics and selection pressures facilitating persistence in vaccinated breeding herds have also been characterized (Clilverd et al., 2024).

Despite these advances, quantitative phylodynamic estimates for transmission parameters at the within-farm level remain limited (Meester et al., 2025). Applying these models to a single herd presents distinct challenges, primarily because farm populations do not mix randomly. Instead, they are highly structured environments where contact is constrained by physical barriers, such as separation into specific pens and rooms. Consequently, the suitability of phylodynamic models to resolve transmission dynamics within such a compartmentalized system, especially where viral genetic diversity is constrained and sparse sampling can obscure rapid transmission chains, remains to be fully validated (Alkhamis et al., 2016; Guinat et al., 2023; Meester et al., 2025).

To address these knowledge gaps, this study leverages the dataset from Clilverd and colleagues (2023) to evaluate whether phylodynamic models can reliably infer within-farm transmission dynamics. Unlike the original study, which characterized the epidemic utilizing longitudinal incidence monitoring and phylogenetic contact tracing, this analysis validates the utility of phylodynamic models by investigating their ability to reconstruct the established epidemiological trajectory by Clilverd and colleagues (2023), differentiate within-pen transmission from transmission between pens sharing the same room versus those in separate rooms, detect differences in infectious periods between production stages (farrowing vs. nursery), and assess the impact of genomic resolution (WGS vs. ORF5) on inference reliability. Following validation of the temporal signal, absence of recombination and excluding viral introductions from sources outside the farm during the study period, unstructured Birth-Death Skyline (BDSKY) and structured Birth-Death with Migration (BDMM) models were applied to quantify these long-term and fine-scale transmission parameters, respectively.

Methods

Study Design and Sequence Acquisition

The virological and epidemiological data used for this phylodynamic analysis were obtained from a longitudinal study by Clilverd et al. (2023), which monitored a PRRSV-1 outbreak on a Spanish 300-sow farrow-to-wean farm. The original study followed three distinct batches of piglets from birth to nine weeks of age at 1.5 months (Batch 1), 8 months (Batch 2), and 12 months (Batch 3) after the clinical detection of the outbreak, which was characterized by the sudden appearance of abortions, stillbirths, and weak-born animals. For all batches, umbilical cords were collected at birth, and blood samples were drawn from individually ear-tagged piglets at 2, 4, 6, and 9 weeks of age. Viral RNA was extracted from collected samples and screened for PRRSV-1 using RT-qPCR. For samples with sufficient viral load (Ct $\leq$ 32.0), viral sequencing was performed. This generated ORF5 sequences via Sanger sequencing for animals in Batch 1 and Batch 3. Additionally, WGS were generated from viral isolates using an Illumina MiSeq platform. No sequences could be obtained from Batch 2. This phase was characterized by an absence of clinical symptoms and only sporadic PCR-positive results (8/74 animals) with high Ct values (>31.5), which prevented the recovery of sufficient genetic material for sequencing. All consensus sequences corresponding to the WGS and ORF5, regions were deposited by Clilverd and colleagues (2023) in GenBank under accession numbers OP688189–OP688223 and OP688224-OP688357, respectively.

In the current study, two primary datasets were compiled: a WGS dataset and an ORF5 dataset (figure 1). Depending on the analysis performed, these were utilized either as restricted 'cohort' datasets or comprehensive 'extended' datasets. The cohort WGS and ORF5 datasets consisted exclusively of the consensus sequences derived directly from the outbreak study described by Clilverd and colleagues (2023). These cohort-specific datasets were used for the phylodynamic analyses. The extended WGS dataset combined the farm consensus sequences with ten regional strains, five Spanish commercial PRRSV vaccines, and all GenBank WGS (taxid 1965066), while the extended ORF5 dataset included farm sequences and all GenBank ORF5 entries. These datasets were used to screen for recombination, and to rule out viral introductions from an outside source during the study period.

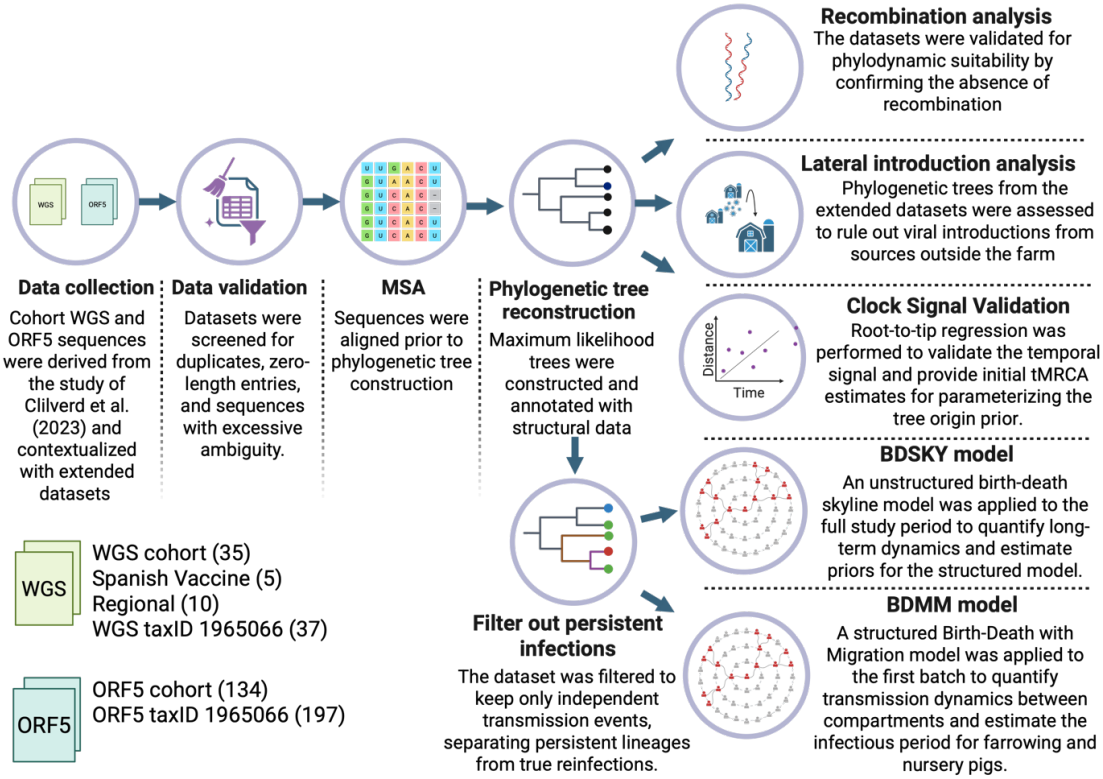


Figure 1. Schematic overview of the study workflow. *The diagram shows computational pipeline used to analyze the PRRSV-1 outbreak. The process moves from Data Collection of Whole Genome Sequences (WGS) and ORF5 sequences (with sequence counts shown in brackets) to Data Validation and Multiple Sequence Alignment (MSA). Following Phylogenetic Tree Reconstruction, the dataset underwent validation steps, including recombination analysis, lateral introduction checks, and clock signal validation. Finally, the data were subjected to Birth-Death Skyline (BDSKY) and Birth-Death with Migration (BDMM) models to quantify long-term and short-term structured transmission dynamics, respectively. tMRCA=Time to Most Recent Common Ancestor. Created with BioRender.com.*

Data processing and validation

All datasets were screened for duplicate or zero-length entries, invalid characters, and International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes (Johnson, 2010). Sequences displaying excessive ambiguity, indicative of poor sequencing quality, were excluded from the datasets. However, sporadic ambiguity codes were retained, as these may represent biological signals of intra-host viral diversity rather than sequencing errors. Additionally, a pairwise identity analysis was performed on the cohort WGS and ORF5 datasets to identify and cluster 100% identical sequences.

To enable further downstream analyses, multiple sequence alignments (MSAs) were constructed using the DECIPHER v3.4.0 package (Wright, 2016) on the extended WGS and ORF5 datasets for recombination and phylogenetic analyses.

Classification of Independent Transmission Events

To ensure the phylodynamic model was informed by independent transmission events, a clade-based analysis was performed to differentiate between sequences representing prolonged viral persistence (within-host evolution) and those from new reinfection events in pigs sampled multiple times. This step was essential as standard birth-death models assume that every branching event in the phylogenetic tree represents a transmission event between hosts (Stadler et al., 2024). Including these sequences, which resemble intra-host evolutionary variants results in over-sampling and false

clusters that bias growth rate estimates upwards (Dearlove et al., 2017) and can lead to poor model convergence (Alkhamis et al., 2022)

First, a patristic distance matrix was calculated from the phylogenetic tree using the ape v5.8-1 package (Paradis & Schliep, 2019), and hierarchical clustering (Kaufman & Rousseeuw, 1990) was performed on this matrix. The resulting dendrogram was initially partitioned into five distinct clades, aligning with the classification described by Clilverd and colleagues (2023), who identified four co-circulating clades in Batch 1 and a single clade in Batch 3. For each pig with multiple samples, the clade assignment of its virus was tracked chronologically. If a pig's subsequent sample was assigned to the same clade as a previous sample, it was classified as a persistent infection; if it was assigned to a different clade, it was classified as a reinfection. Since hierarchical clustering requires a pre-determined number of clusters, a robustness test was performed by varying the defined clade sizes to ensure that the classification of infection status for individuals remained consistent across different thresholds. Based on this classification, we retained only the first samples of persistent lineages, defined as initial infections, as well as samples from reinfections for the main phylodynamic analyses

Phylogenetic Reconstruction and Topology

To assess viral diversity circulating on the farm, phylogenetic trees were constructed for the cohort WGS and ORF5 datasets using IQ-TREE v3 (Wong et al., 2025). The optimal nucleotide substitution model was inferred using the ModelFinder module (Kalyaanamoorthy et al., 2017), and nodal support was assessed via Ultrafast Bootstrap Approximation with 1,000 replicates (Ecker et al., 2024). These trees were rooted using phytools v2.5 (Revell, 2024) and subsequently annotated with spatial metadata. The structural data used for this annotation was inferred directly from the supplementary video provided by Clilverd and colleagues (2023). However, this spatial annotation was restricted to Batch 1 sequences, as structural data for Batch 3 was not available. For the Batch 1 sequences, the farm structure was defined as a farrowing unit comprising a single compartment with 18 pens, and a nursery unit consisting of four distinct compartments, each containing four pens.

Assessment of Recombination and Lateral Introduction

To rule out viral introductions from sources outside the farm during the study period, maximum likelihood trees were generated for both the extended WGS and ORF5 datasets. These trees were evaluated to determine whether farm-derived sequences clustered into a single, strongly supported monophyletic clade, consistent with a single introduction event.. Following this, recombination was assessed in both datasets, as it violates the bifurcation assumptions of standard phylogenetic models (Driebe et al., 2015). This was done by visual inspection using Neighbor-Net in SplitsTree (Huson & Bryant, 2024) and statistical validation via the Pairwise Homoplasy Index (Phi) test (Bruen et al., 2006).

Molecular Clock Validation and Temporal Signal

To assess the temporal signal over the 12 month study period, root-to-tip regression analyses were performed using Clockor2 (Featherstone et al., 2024). The analyses were also used to provide an initial estimate of the time to the Most Recent Common Ancestor (tMRCA) required to parameterize the tree origin prior for the subsequent BEAST analyses. In addition, the cohort ORF5 data were partitioned to assess whether Batch 1 and Batch 3 exhibited valid temporal signals under separate local clocks.

Phylodynamic Analysis

Phylodynamic analysis was performed using Bayesian inference in BEAST v2.7.7 (Bouckaert et al., 2019), allowing for the joint estimation of the phylogenetic tree, evolutionary parameters, and

epidemiological rates by integrating prior knowledge with the likelihood of the sequence data. Within this Bayesian framework, transmission dynamics were reconstructed using birth-death phylodynamic models, which simulate the epidemic as a forward-in-time stochastic process driven by rates of viral transmission (λ), recovery (μ), and sampling (ψ) (Stadler et al., 2024). To facilitate biological interpretation, the bdmm-prime v2.6.3 package (Kühnert et al., 2016; Scire et al., 2022; Vaughan & Stadler, 2025) was used, which re-parameterizes the raw stochastic rates (λ, μ and ψ) into derived epidemiological quantities: The Re ($Re = \lambda/\delta$), the sampling proportion ($p = \psi/\delta$) and the rate of becoming uninfectious ($\delta = \mu+r\psi$) which can be converted to the average infectious period by taking its reciprocal ($1/\delta$) (Stadler et al., 2024).

Since the resulting joint posterior distribution is analytically intractable (Stadler et al., 2024), Markov chain Monte Carlo (MCMC) sampling was employed to numerically approximate the posterior estimates for all parameters. MCMC chains were run until convergence was achieved; unless otherwise stated, sufficient mixing and the quality of state-space exploration were confirmed using Tracer v1.7.2 (Rambaut et al., 2018), ensuring that all continuous parameters reached an Effective Sample Size (ESS) > 200. The final log files were processed and summarized using the coda package v0.19.4.1 (Plummer et al., 2006) in R, with burn-in periods configured individually for each run to optimize posterior sampling.

Crucially, the Sampled Ancestors package (Gavryushkina et al., 2014) was used, with the removal probability upon sampling (r) set to 0 as animals remained in the herd after sampling. Unlike standard phylogenetic models, which typically treat all samples as terminal tips in the phylogenetic tree (Stadler et al., 2024), this approach allows sampled individuals to be direct ancestors of subsequent cases. This configuration is essential to capture within-farm dynamics, where high contact rates within small subpopulations create a high probability that sampled individuals are the direct ancestors of subsequently sampled pen-mates.

Long-term Temporal Dynamics (BDSKY)

To reconstruct the epidemic trajectory and estimate the Re over the 12-month study period, while evaluating the impact of genomic resolution (WGS vs. ORF5) on parameter estimates, an unstructured phylodynamic analysis was conducted using the BDSKY model (Stadler et al., 2013). Unlike constant-rate birth–death models, BDSKY allows epidemiological parameters to vary in a piecewise-constant manner across discrete time intervals, defined as “epochs”. This enables the reconstruction of the full trajectory of the outbreak by estimating the Re across distinct phases. Consequently, it allows for the precise identification of epidemiological shifts that may be missed by standard phylogenetic analyses, such as those previously applied by Clilverd and colleagues (2023).

To ensure robust inference, prior specifications were designed to address the parameter redundancy inherent in birth-death-sampling models, where diverse combinations of transmission, recovery, and sampling rates can produce identical likelihoods (Gavryushkina et al., 2014; Scire et al., 2022). Narrow priors were applied to the rate of becoming uninfectious and origin, informed by established PRRSV-1 duration of infectivity (Charpin et al., 2012) and the confirmed outbreak timeline (Clilverd et al., 2023) (Table 1). Priors for the Re, sampling proportion, clock rate, and epoch transition times were kept uninformative to allow these parameters to be estimated directly from the genomic data rather than prior assumptions. By avoiding overly informative priors, this method reduces over-smoothing and preserves the ability to detect genuine changes in epidemiological patterns, including abrupt shifts in transmission dynamics (Parag et al., 2022). To validate these specifications, we performed a ‘sample-from-prior’ analysis (Drummond & Bouckaert, 2015) in BEAST, running the MCMC chain without sequence data to verify that the posterior distributions of inferred parameters were driven by the data rather than artifacts of the joint prior constraints.

Table 1: Prior distributions and rationale for the Bayesian phylodynamic models. *Summary of the prior distributions applied to the Birth-Death Skyline (BDSKY) and Birth-Death with Migration (BDMM) models, including reference sources. HPD = Highest Posterior Density, tMRCA = time to Most Recent Common Ancestor*

Parameter	Model	Prior Distribution	Rationale	Reference
Effective Reproduction Number (Re)	Both	Lognormal(0, 1.0)	Loose prior centered at 1, reflecting the epidemiological threshold. Its broad range allows genomic data to drive the estimates while preventing biologically implausible rates.	Andrade & Duggan (2022); Boskova et al. (2018)
Rate of Becoming Uninfectious	BDSKY	Lognormal(8.0, 0.1)	Tight prior (mean infectious period of ~45 days, 95% CI ~37–65 days) based on experimental data for nursery-age infected pigs which made up the majority of the study population	Charpin et al. (2012)
	BDMM	Lognormal(8.0, 0.5)	Broader prior (mean infectious period of ~45 days, 95% CI ~20–140 days) to evaluate differences between infections in utero and after birth.	Charpin et al. (2012); Butler et al., (2010; Rowland et al., (2003); R. W. Wills et al., (2003); You et al., (2022)
Sampling Proportion	BDSKY	Beta(1.0, 1.0)	Non-informed (uniform) prior allowing free estimation across the range 0 to 1.	-
	BDMM	Normal(0.33, 0.09)	Confined to the 95% HPD interval derived from the Batch 1 posterior of the BDSKY analysis.	This study
Clock Rate (Mean)	BDSKY	Lognormal(0.001, 1.25)	Broad distribution capturing the range for observed rates for PRRSV-1 centered around 0.001 substitutions per site per year.	Boskova et al. (2018); Li et al. (2022); Parisio et al. (2024); Shin et al. (2022)
	BDMM	Lognormal(0.0136, 0.1)	Confined by the posterior estimates from the BDSKY analysis due to weak local temporal signal in the Batch 1 only.	This study
Clock Rate (Std. Dev.)	Both	Gamma(0.5396, 0.3819)	Designed to complement the mean clock rate prior centered at 0.001, inducing an effective prior that provides a biologically plausible baseline while retaining the flexibility to estimate rate heterogeneity	Boskova et al. (2018)
Tree Origin (tMRCA)	BDSKY	Lognormal(1.20, 0.05)	Parameterized relative to the last Batch 3 sample (mean ~438 days, 95% CI 398–482 days) based on root-to-tip regression and the expected 1-to-8-week subclinical phase.	This study; Pedro Mil-Homens et al. (2024)
	BDMM	Lognormal(0.45, 0.1)	Parameterized relative to the last Batch 1 sample (mean ~164 days, 95% CI 134–199 days) based on root-to-tip regression and the expected 1-to-8-week subclinical phase.	This study; Pedro Mil-Homens et al. (2024)
Change Times (Epochs)	BDSKY	Uniform(0, 1.2)	Uninformative prior allowing free estimation of epoch configurations between the tMRCA and the end of Batch 3.	This study
Migration Rate	BDMM	Exponential(1.0)	Uninformative prior that accommodates a broad range of biologically plausible rates and ensures robustness in the presence of limited signal for population structure	Seidel et al. (2024)
Substitution Model	Both	bModelTest	Co-estimated with default priors for invariable sites, rate heterogeneity, and for base frequencies.	Bouckaert & Drummond (2017)

To infer the optimal nucleotide substitution model directly from the data instead of relying on predefined universal substitution models, the bModelTest package (Bouckaert & Drummond, 2017) was used. Additionally, an uncorrelated relaxed clock model (Drummond et al., 2006) was assessed to evaluate whether a single mean molecular clock rate could be applied across the entire study period, or whether substitution rates varied among branches of the phylogeny.

Distinct epoch strategies were used for the Re and sampling proportion. The Re was modeled using flexible 3-, 4-, and 5-epoch configurations with uniformly estimated transition times over the full study period (Figure 2). The 3-epoch configuration was applied to test whether the inferred dynamics could recover the established three-phase timeline (outbreak, dormancy, and resurgence) defined by Clilverd and clolleagues (2023). Additionally, the 4- and 5-epoch configurations were used to investigate whether a higher temporal resolution could reveal further, previously unobserved epidemiological phases.

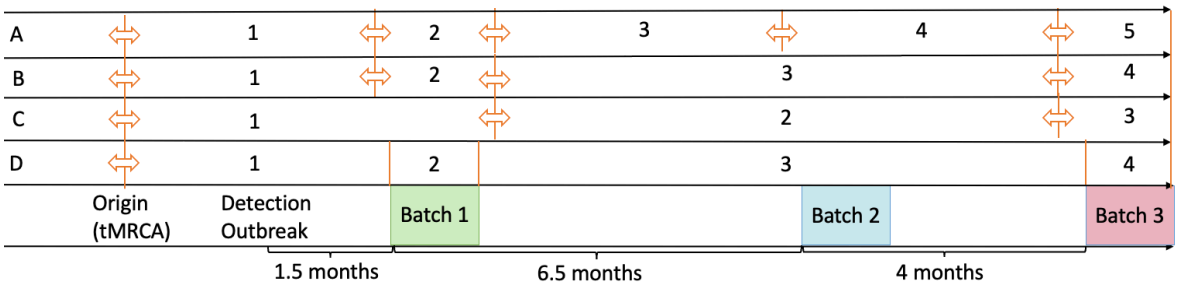


Figure 2: Schematic of temporal discretization strategies for the Birth-Death Skyline (BDSKY) model. *The diagram illustrates the flexible production stage configurations (A–C) applied to the effective reproductive number (Re), which utilized variable 3-, 4-, and 5-epoch intervals to evaluate if the model could reconstruct the outbreak phases. Orange arrows indicate estimated parameters, denoting both the flexible epoch transition times and the time of Most Recent Common Ancestor (tMRCA), reflecting the origin of the outbreak. This is contrasted with the sampling proportion configuration (D), where epoch intervals were fixed to strictly mirror the study schedule and sampling gaps. Here the sampling proportion of epoch 1 and 3 were fixed to 0 while this value was estimated for epoch 2 and 4.*

The sampling proportion (p) was modeled using a fixed 4-epoch structure to reflect the sampling schedule of the study. In the general birth-death framework, p is defined as the ratio of the sampling rate to the total rate of becoming uninfectious ($p=\psi/(\mu+r\psi)$) (Stadler et al., 2024). However, because the removal probability was set to zero ($r=0$) in the Sampled Ancestors model, this simplifies to the ratio of the sampling rate to the recovery rate ($p=\psi/\mu$). Furthermore, since longitudinal duplicates were excluded while distinct reinfection events were retained, p is functionally interpreted here as the proportion of distinct infection events that were successfully sequenced. Based on this definition, the parameter was constrained according to data availability: the first and third epochs, which contained no samples, were fixed to 0. Conversely, for the second and fourth epochs (covering Batch 1 and Batch 3), p was estimated using an uninformative prior, as the total number of infection events in the complete population could not be reliably inferred from the available metadata.

Fine-Scale Spatial Transmission Dynamics (BDMM)

To quantify Batch 1 dynamics, specifically estimating compartment-specific transmission rates and infectious periods of farrowing and nursery stage pigs, a structured analysis was conducted using the BDMM model (Kühnert et al., 2016). This framework extends the temporal flexibility of the BDSKY model to structured populations by allowing parameters to vary not only across time but also across compartments. Given the insufficient WGS coverage for Batch 1, this structured

analysis was performed on the ORF5 dataset by partitioning the outbreak into three epochs that reflected distinct production stages: the initial sow herd phase (tMRCA to farrowing), the farrowing phase (first 4 weeks), and the nursery phase (weeks 4–9) (figure 3).

The first epoch included a single unobserved ("ghost") deme with a Re representing the sow reservoir which aimed to capture transmission via unsampled hosts or reservoirs and avoids bias in structured phylodynamic models (Ewing & Rodrigo, 2006; Müller et al., 2025). The root location probability was fixed at 100% for this deme, reflecting that the cohort was unborn at viral introduction to the farm and that all ancestral lineages originated there before migrating into observed populations.

The spatial configuration of the demes was modeled directly on the pen-level topology described in the "Phylogenetic Reconstruction and Topology" section. However, because the BDMM model requires a consistent number of demes across all epochs, the overall structure was constrained to 17 demes, comprising 16 pens and one ghost deme. While this configuration excluded two specific farrowing pens, the impact on transmission inference was considered negligible due to data sparsity; the farrowing phase yielded only six sequences, all of which originated from just two pens already included in the model.

In this analysis, epoch change times were fixed but defined under two configurations: a production stage configuration and a migration configuration. The production stage configuration estimated the Re and rate of becoming uninfectious across three biological stages: reservoir, farrowing, and nursery. The migration configuration, based on metadata from Clilverd et al. (2023), defined epochs of per-lineage, forward-in-time migration rates, assuming discrete movements between compartment at each sampling period. The analysis focused on four key one-day transitions: from sow to farrowing, from farrowing to nursery at week 4, and within the nursery at weeks 6 and 9. Movement within the farrowing unit at week 2 was excluded due to insufficient sequence data, as only six sequences were available for week 0 and none for week 2, precluding reliable estimation of migration rates for this interval.

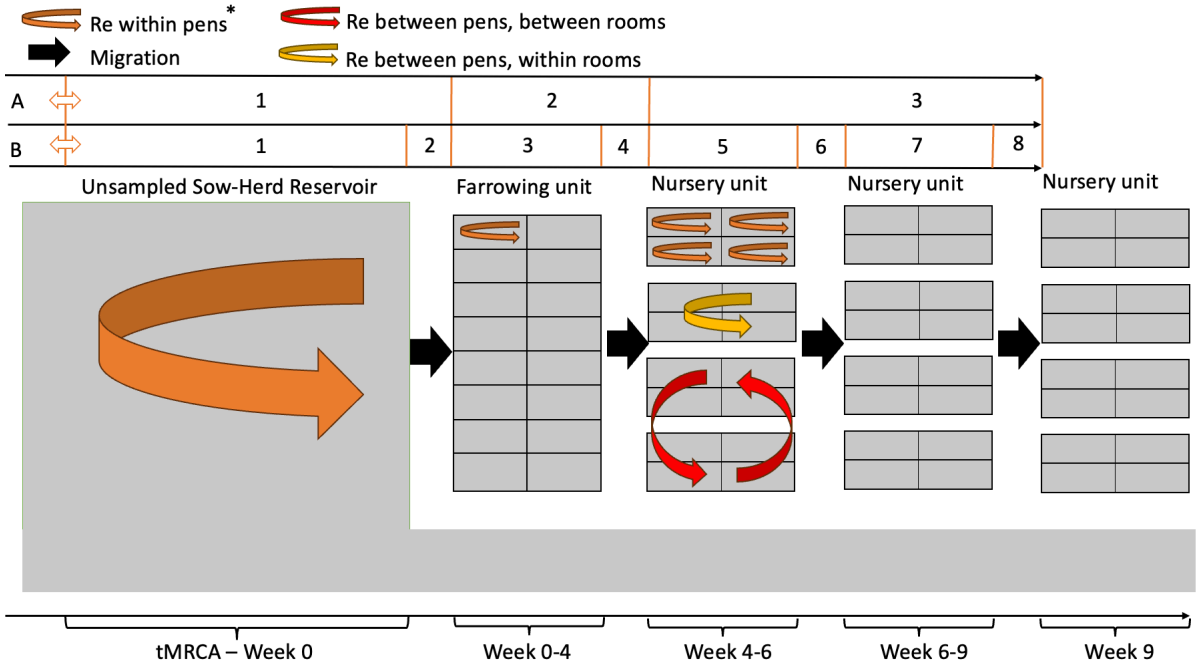


Figure 3: Schematic representation of the Birth-Death with Migration Model (BDMM) configuration. *The diagram outlines the spatial structure and migration pathways modeled for Batch 1, including the unsampled sow-herd reservoir, farrowing unit, and nursery unit*

compartments. Unlike the flexible unstructured model, the epoch intervals for the BDMM analysis were fixed based on defined production stages (A) and migration events (B).
**: For the Sow-Herd Reservoir, which is modeled as a single "ghost" deme, this parameter represents transmission within the reservoir population as a whole rather than within a specific physical pen.*

To characterize transmission dynamics at different spatial levels, a parameter linking strategy was used to estimate three distinct Re values. For the nursery phase, three linking strategies were applied: first, the Re was estimated within pens and linked across all pens; second, the Re between pens within rooms (e.g., pens 1–4, 5–8) were linked; and third, the Re between pens across rooms were linked. For the farrowing phase, only within-pen Re values were estimated and linked because horizontal transmission between pens could be excluded: all six sequences from this phase (five from one pen, one from another) were identified as vertical transmission (Clilverd et al., 2023), with no genetic evidence for horizontal transmission between pens. Furthermore, it was assumed that the farrowing and nursery units were isolated from the external farm environment after initial introduction, so the Re between pens and the ghost deme was set to zero, treating the ghost deme as an ancestral lineage source, not a site of ongoing transmission.

The molecular clock rate for Batch 1 could not be reliably inferred due to insufficient temporal signal (see results: “Molecular Clock Validation and Temporal Signal”). To address this, the clock rate prior was specified using the posterior estimates from the preceding BDSKY analysis fitted to Batch 1 and Batch 3 (table 1). This approach aligns with recent HEV phylodynamic protocols where informative priors from larger datasets are applied to short-term subsets (Meester et al., 2025), and is supported by evidence that evolutionary rates converge and remain broadly stable across longer epidemic timescales (Bryant et al., 2007).

Furthermore, to address the parameter redundancy inherent to birth–death–sampling models, both the sampling proportion prior (centered on the BDSKY posterior estimate for batch 1, with the sow herd phase fixed at 0) and the origin time prior were assigned narrow distributions, ensuring reliable inference of the remaining epidemiological parameters. Finally, the bModelTest package was used using the same configuration as was done in the BDSKY model to investigate whether the inferred substitution model changed by excluding batch 3 from the analysis.

A robustness test was performed to challenge the assumption of discrete one-day animal movements by testing a continuous migration model, where migration was permitted continuously throughout the nursery phase. Finally, to check whether the posterior estimates were driven by genetic data rather than prior assumptions, the BDMM model was run with the sample-from-prior setting for both the continuous and discrete migration configurations.

Data and Code Availability

To ensure transparency and facilitate the reproducibility of the analyses presented here, all computational code, processing scripts, and final phylodynamic model configurations have been deposited in a public GitHub repository: **leon1603/prrsv-withinfarm-phylodynamics** (https://github.com/leon1603/prrsv-withinfarm-phylodynamics). Within this repository, a dedicated directory named "thesis" contains the digital version of this document alongside all referenced supplementary files. This includes an extensive metadata file (provided as Supplementary File 1 in the "thesis" folder), which details the specific structural-level data, infection classifications, clade assignment and identical sequence clusters for all cohort sequences in addition to pigs that were not sequenced but had metadata available which could be extracted from the supplementary files of Clilverd et al. (2023)

Results

Study Population and Sequence Validation

The final Cohort datasets were assembled, comprising a Cohort WGS dataset of 35 sequences (partitioned into n=24 for Batch 1 and n=11 for Batch 3) and a Cohort ORF5 dataset of 134 sequences (partitioned into n=86 for Batch 1 and n=48 for Batch 3) (Table 2). Furthermore, the Extended datasets, consisted of 87 and 331 sequences for WGS and ORF5, respectively.

Sequence validation revealed minor and explainable inconsistencies in the cohort data; a 'Y' ambiguity code (C/T) was identified in two WGS, consistent with previously reported intra-host viral diversity in this dataset by Clilverd and colleagues (2023) (see Supplementary File 2 for detailed per-sequence ambiguity information). Furthermore, negligible ambiguity (<0.9%) was observed in all external reference strains which were retained to avoid compromising phylogenetic signal (Lozano-Fernandez, 2022).

Table 2: Summary of PRRSV-1 sequence datasets, quality metrics, and batch partitioning. *Overview of the Cohort and Extended datatsets for Whole Genome Sequences (WGS) and ORF5, detailing sequence counts, length ranges and ambiguity codes.*

Dataset	Sequence Count	Min Length	Max Length	Total Ambiguities	Ambiguity Characters
Cohort WGS Consensus	35	15098	15098	2	Y
Batch 1	24				
Batch 3	11				
Cohort ORF5 Consensus	134	606	606	0	None
Batch 1	86				
Batch 3	48				
Regional Sequences	10	14443	14910	132	K, M, N, R, S, W, Y
Licenced Vaccines in Spain	5	14758	15120	0	None
WGS (taxID 1965066)	37	14932	15428	225	K, N, R, S, Y
ORF5 (taxID 1965066)	197	606	606	97	K, M, R, S, W, Y

Assessment of genetic redundancy revealed that all sequences in the WGS cohort were unique. In contrast, the ORF5 dataset comprised only 47 unique sequences, corresponding to a redundancy of 64.93% and forming 20 groups of identical sequences (see Supplementary File 3 for detailed tables).

Additionally, an inconsistency in the sampling dates for the cohort ORF5 sequences was also identified and corrected. The recorded metadata for Batch 1 listed identical sampling dates for week 6 and week 9 at 13/07/2017, except for a single week 9 sequence dated 25/07/2017. This was determined to be a likely annotation error. Consequently, all Batch 1, week 9 sequences were corrected to the 25/07/2017 date to accurately reflect the sampling schedule.

Classification of Independent Transmission Events

To differentiate persistent infections from reinfection events, a clade-based classification was performed. The final classification for the WGS dataset identified 24 initial infections, 10 persistent infections, and 1 reinfection. For the ORF5 dataset, the analysis identified 80 initial infections, 53 persistent infections, and 1 reinfection. Table 3 summarizes sequence inclusion after the exclusion of persistent infections. This analysis proved to be robust for both datasets: the number of persistent

and new infections remained stable when the number of clades was set to any value between three and seven for WGS, and between four and seven for the ORF5 dataset (Supplementary file 4). Additionally, both ORF5 and WGS analyses identified a reinfection event in animal 515. This event was recorded at week 6 in the ORF5 dataset (where the week 9 sample was unavailable) and at week 9 in the WGS dataset (where the week 6 sample was unavailable).

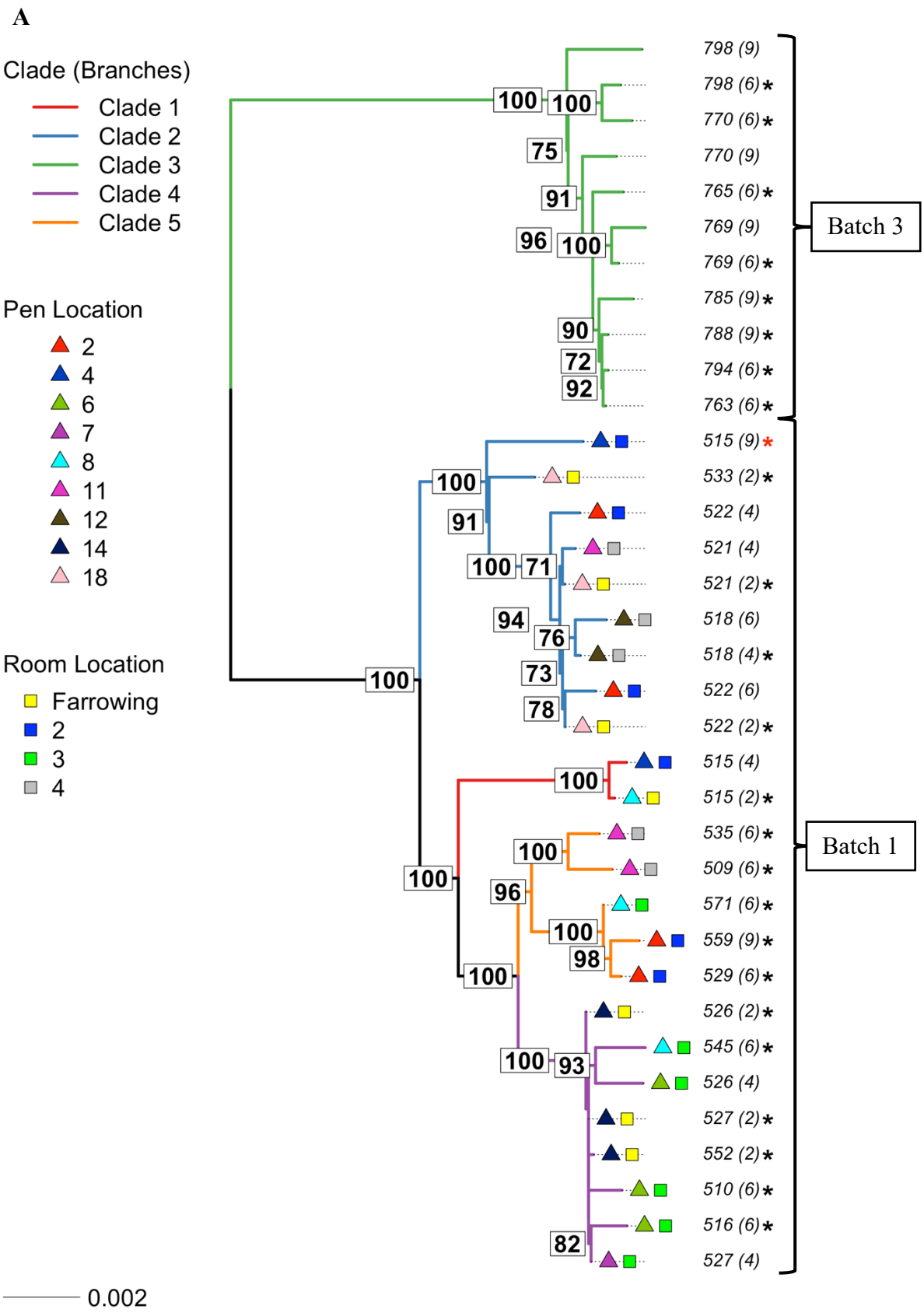
Table 3: Metadata summary of ORF5 sequences before and after filtering for persistent infections. *Data are presented as Total Sequences / Included Sequences.*

	General Cohorts		Spatial Distribution (Batch 1)				
Week	**Batch 1**	**Batch 3**	**Farrowing**	**Room 1**	**Room 2**	**Room 3**	**Room 4**
Week 0	6 / 6	0 / 0	6 / 6	0 / 0	0 / 0	0 / 0	0 / 0
Week 2	2 / 0	5 / 5	2 / 0	0 / 0	0 / 0	0 / 0	0 / 0
Week 4	8 / 8	0 / 0	0 / 0	0 / 0	3 / 3	5 / 5	0 / 0
Week 6	34 / 25	22 / 18	0 / 0	9 / 9	8 / 5	17 / 11	0 / 0
Week 9	36 / 14	21 / 5	0 / 0	12 / 5	4 / 1	11 / 2	9 / 6
Total	**86 / 53**	**48 / 28**	**8 / 6**	**21 / 14**	**15 / 9**	**33 / 18**	**9 / 6**

After all persistent infections were excluded from the ORF5 datasets, the number of total and unique sequences dropped from 134 to 81 and 47 to 35 sequences, respectively. The redundancy percentage improved from 64.93% to 56.59% and 11 identical sequence groups were observed compared to the 20 identical groups in the ORF5 dataset without exclusion. After removal of persistent infections from the WGS dataset, the total number of total sequences dropped from 35 to 25 sequences.

Phylogenetic Reconstruction and Topology

The phylogenetic reconstruction of the consensus WGS and ORF5 datasets yielded topologies characterized by high statistical confidence, as indicated by the high nodal bootstrap values observed across the backbone of both rooted trees (Figure 4). The clustering algorithm used for the reinfection analysis identified five monophyletic groups in the WGS dataset, yielding a cluster composition identical to the clades described by Clilverd and colleagues (2023). Farrowing unit sequences consistently occupied ancestral nodes relative to nursery unit sequences, which appeared as descendant taxa within the topology. In contrast, the ORF5 phylogeny failed to correctly recover this structure. Specifically, the clustering algorithm could not resolve one clade as a monophyletic group; instead, one lineage was split and interspersed among other clades. Additionally, the clear ancestral relationship observed in the genomic data, where farrowing sequences precede nursery sequences, was not evident in the ORF5 tree. Finally, a strong spatial signal is evident in the terminal branches of the ORF5 phylogeny. It can be observed that the clusters of 100% identical sequences consist largely of samples originating from the same room. While these identical sequence groups also show a clear tendency to cluster within the same physical pens, this signal is most pronounced at the room level.


A
Clade (Branches)
Clade 1
Clade 2
Clade 3
Clade 4
Clade 5
Pen Location
2
4
6
7
8
11
12
14
18
Room Location
Farrowing
2
3
4
0.002
798 (9)
798 (6)*
770 (6)*
770 (9)
765 (6)*
769 (9)
769 (6)*
785 (9)*
788 (9)*
794 (6)*
763 (6)*
515 (9)*
533 (2)*
522 (4)
521 (4)
521 (2)*
518 (6)
518 (4)*
522 (6)
522 (2)*
515 (4)
515 (2)*
535 (6)*
509 (6)*
571 (6)*
559 (9)*
529 (6)*
526 (2)*
545 (6)*
526 (4)
527 (2)*
552 (2)*
510 (6)*
516 (6)*
527 (4)
Batch 3
Batch 1

B

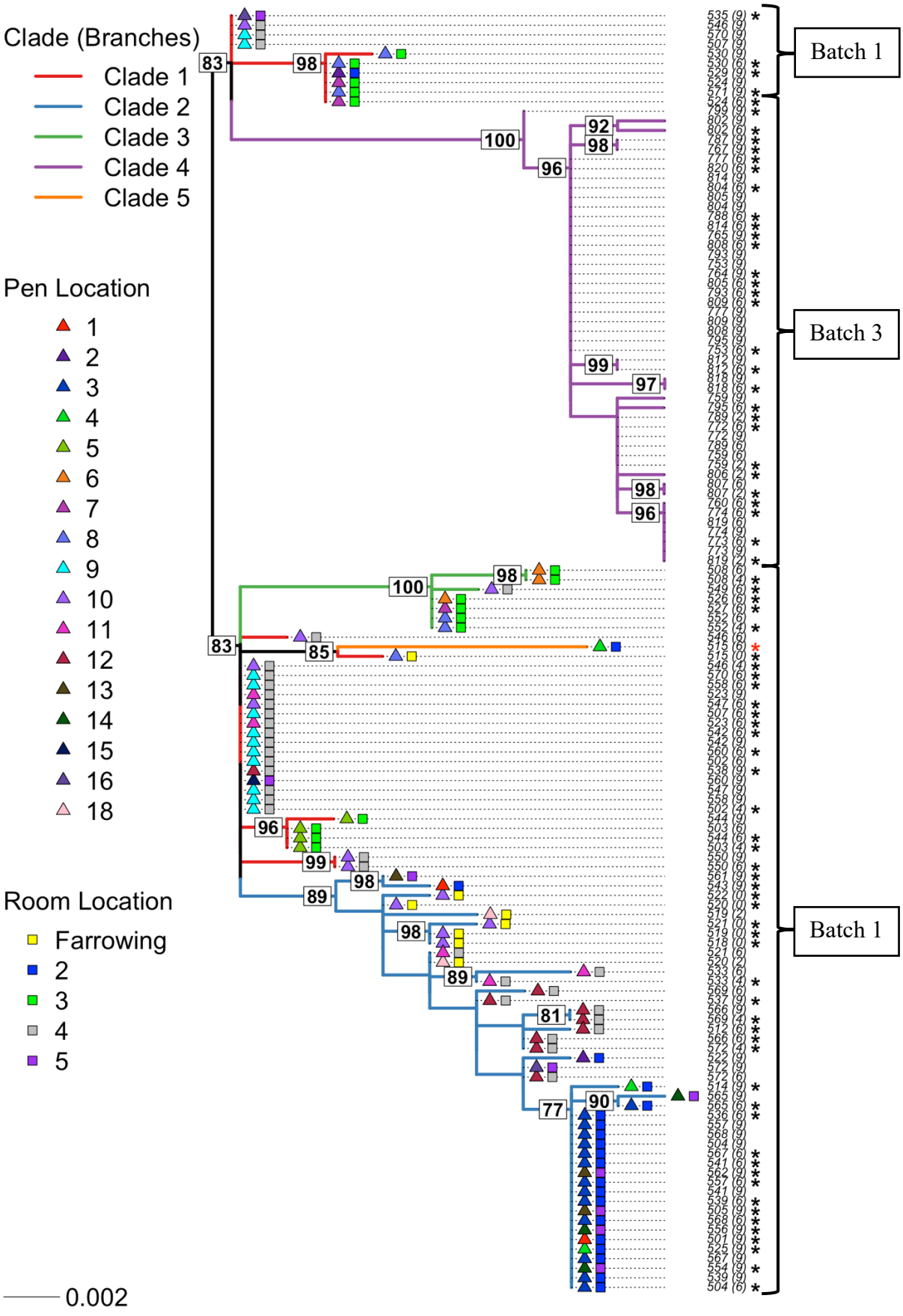

Clade (Branches)
Clade 1
Clade 2
Clade 3
Clade 4
Clade 5
Pen Location
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
18
Room Location
Farrowing
2
3
4
5
0.002
Batch 1
Batch 3
Batch 1

Figure 4: Maximum likelihood phylogenetic reconstruction of the 35 cohort Whole Genome Sequences (WGS) (A) and 134 consensus ORF5 sequences (B). *The tree displays the evolutionary relationships of the cohort sequences with branches colored according to the five-clade classification. The scale bar represents genetic distance expressed as nucleotide substitutions per site. Colored triangles and squares denote pen- and room-level locations, respectively. Nodal support values are indicated on the tree nodes (values below 70 are removed for clarity). An asterisk (*) following the sequence annotation denotes an initial infection, whereas a red asterisk indicates a reinfection event.*

Assessment of Recombination and Lateral Introduction

Phylogenetic reconstruction of the extended WGS and ORF5 datasets revealed strongly supported monophyletic clades of farm sequences in both datasets (Supplementary File 5), confirming a single introduction and validating that the outbreak was constrained to within-farm dynamics for the duration of the 12-month study period. Subsequent screening for recombination revealed no evidence of reticulation in Neighbor-Net networks (Supplementary file 6) nor significant signals in the Phi test (WGS: p=1.0; ORF5: p=0.84), validating its suitability for phylodynamic inference.

Molecular Clock Validation and Temporal Signal

Root-to-tip regression analysis, performed using Clockor2 on the maximum likelihood phylogenetic trees to obtain priors for the final BEAST models, yielded strong positive correlation rates for both the cohort WGS (R^2 = 0.992) and ORF5 (R^2 = 0.836) datasets. The estimated evolutionary rates were 1.354 x 10^{-2} and 1.698 x 10^{-2} substitutions/site/year for WGS and ORF5, respectively. Furthermore, the tMRCA was estimated to be 2017/01/23 for the ORF5 dataset and 2017/02/24 for the WGS dataset, placing the origin of the outbreak approximately 8 weeks before clinical detection at 2017/04/08. This confirmed the combined datasets' suitability for temporal analysis.

However, the partitioned analysis (local clock model for ORF5 data, partitioned by Batch 1 and Batch 3), revealed that Batch 1 exhibited a poor temporal signal on its own (R^2 = 0.144) with an evolutionary rate of 2.693 x 10^{-2} substitutions/site/year, while Batch 3 displayed no temporal signal (negative evolutionary rate). This finding indicated that a reliable clock rate could not be estimated from the Batch 1 and 3 data separately.

Long-term Temporal Dynamics (BDSKY)

To reconstruct the viral spread and estimate evolutionary parameters over the 12-month study period, the BDSKY model was applied to both the cohort ORF5 and WGS datasets. A direct comparison of the inferred epidemiological trajectories showed that both datasets successfully resolved the outbreak into the three distinct phases (initial outbreak, dormancy and resurgence) previously described by Clilverd and colleagues (2023). Under the 3-epoch configuration, the ORF5 and WGS datasets yielded consistent epidemiological patterns, though with slight variations in magnitude (Figure 5 A and B). The ORF5 analysis estimated an initial outbreak phase with a Re of 2.48 (95% Highest Posterior Density (HPD): 1.81–3.15), followed by a significant reduction in transmission during the dormant phase (Re = 0.20; 95% HPD: 0.03–0.40), and a final resurgence phase Re = 3.17; 95% HPD: 2.02–4.40). The WGS analysis mirrored this trajectory, estimating an initial outbreak Re of 1.80 (95% HPD: 1.20–2.44), a dormant phase Re of 0.35 (95% HPD: 0.05–0.71), and a resurgence Re of 3.85 (95% HPD: 1.62–6.42). Furthermore, the estimated transition times between these phases aligned with the study's metadata for both datasets, placing Batch 1 within the outbreak phase, Batch 2 in the dormant phase, and Batch 3 in the resurgence phase.

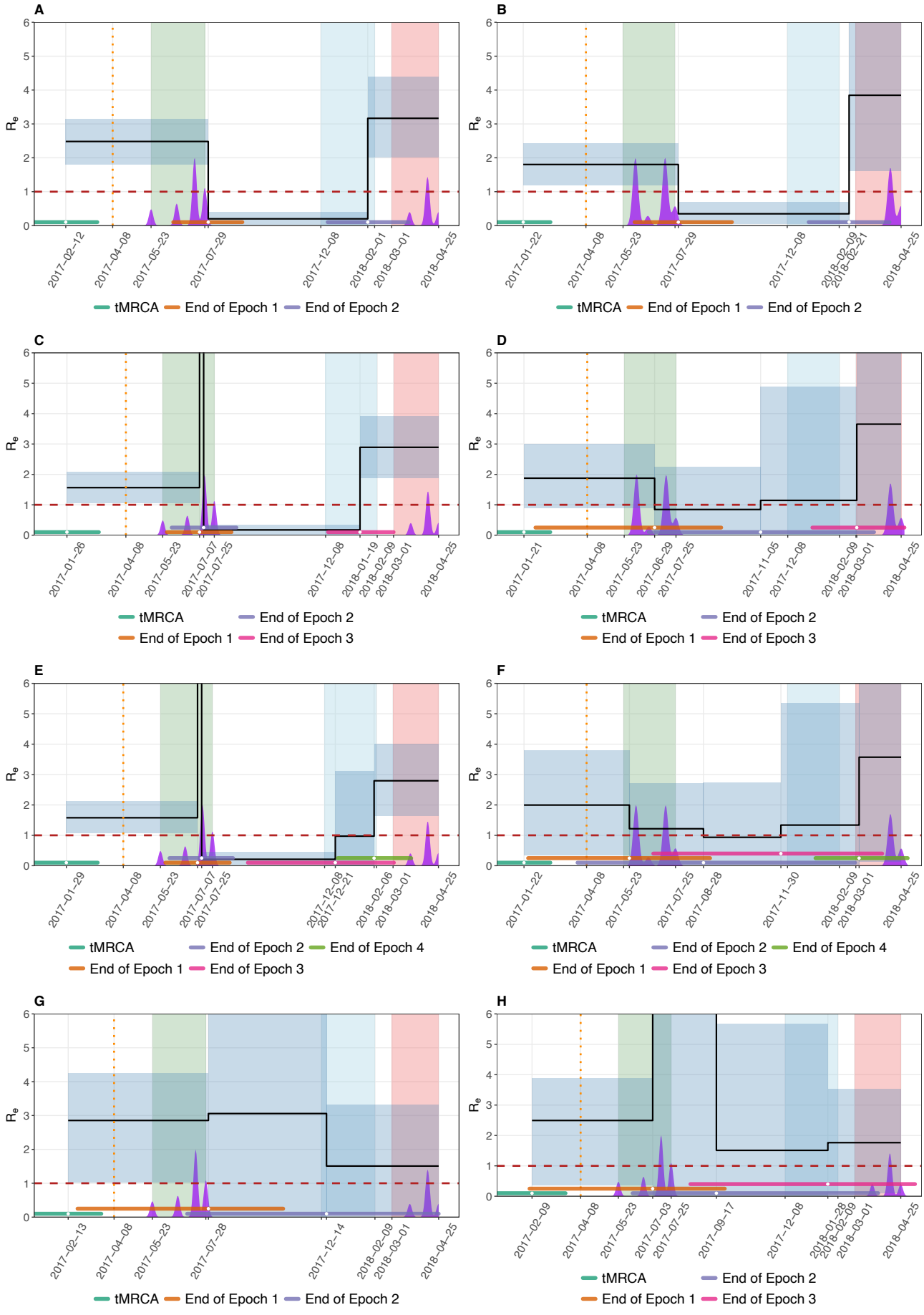

A
B
C
D
E
F
G
H
R_e
tMRCA
End of Epoch 1
End of Epoch 2
End of Epoch 3
End of Epoch 4

Figure 5: Comparison of the temporal trajectory of the Effective Reproductive Number (Re) inferred from ORF5 and Whole Genome Sequence (WGS) datasets, alongside prior sensitivity analysis. *The upper panels compare the epidemiological reconstruction for the ORF5 dataset (Left Column: A, C, E) and the WGS dataset (Right Column: B, D, F) across three temporal configurations: 3-epoch (A, B), 4-epoch (C, D), and 5-epoch (E, F). The bottom panels display the corresponding sample-from-prior analysis using the 3-epoch (G) and 4-epoch (H) configurations. The Re is depicted by the solid black line (mean) and the shaded background regions indicate the 95% Highest Posterior Density (HPD) intervals. The horizontal-colored bars positioned above the x-axis denote the 95% HPD intervals for the epoch transition times and the time to Most Recent Common Ancestor (tMRCA), while the purple density plots along the timeline indicate the sampling density. The red dashed line represents the epidemiological threshold of Re = 1 and the orange dotted vertical line indicates the detection of the outbreak on the farm. The colored vertical bars represent the sampling period for batch 1 (green), batch 2 (blue) and batch 3 (red). Re estimates exceeding 6 are omitted from the visualization for clarity but are available in supplementary file 7.*

While the 3-epoch model yielded comparable parameter estimates across both datasets, increasing the temporal resolution to 4- and 5-epoch models resulted in differing model behavior between the datasets. In the ORF5 analysis, higher-resolution models resolved a short-duration, high-magnitude transmission peak during the early outbreak period, with estimated Re values reaching approximately 20 (95% HPD: ~7–40) (Figure 5 C and E). In contrast, application of 4- and 5-epoch models to the WGS dataset resulted in increased uncertainty in parameter estimates. For these models, estimations for epoch transition times and Re values exhibited overlapping 95% HPD intervals, and distinct temporal phases were not clearly resolved (Figure 5D and F). Finally, the sampling proportion was estimated consistently across models for both Batches 1 and 3. Focusing on Batch 1, the WGS dataset yielded posterior estimates centered around 0.25 (95% HPD: 0.10–0.40), compared to approximately 0.30 (95% HPD: 0.15–0.45) for the ORF5 dataset (Figure 6).

Conversely, the comparison of evolutionary parameters showed that the WGS dataset offered higher precision than ORF5. The mean molecular clock rate for WGS was estimated at 1.12×10^{-2} substitutions/site/year (95% HPD: 0.97×10^{-2} – 1.33×10^{-2}), compared to the 1.41×10^{-2} (95% HPD: 0.85×10^{-2} – 2.00×10^{-2}) estimated for ORF5. Despite this difference in precision, the 95% HPD intervals for the Coefficient of Variation (CoV) consistently exceeded 0.1 across all models, supporting the use of an uncorrelated relaxed clock (Drummond & Bouckaert, 2015). Additionally, the analysis identified distinct substitution models: the K81 model for ORF5 and Model 24 for WGS, which couples A→T and G→T rates while distinguishing A→C from G→T transversion rates (Bouckaert & Drummond, 2017).

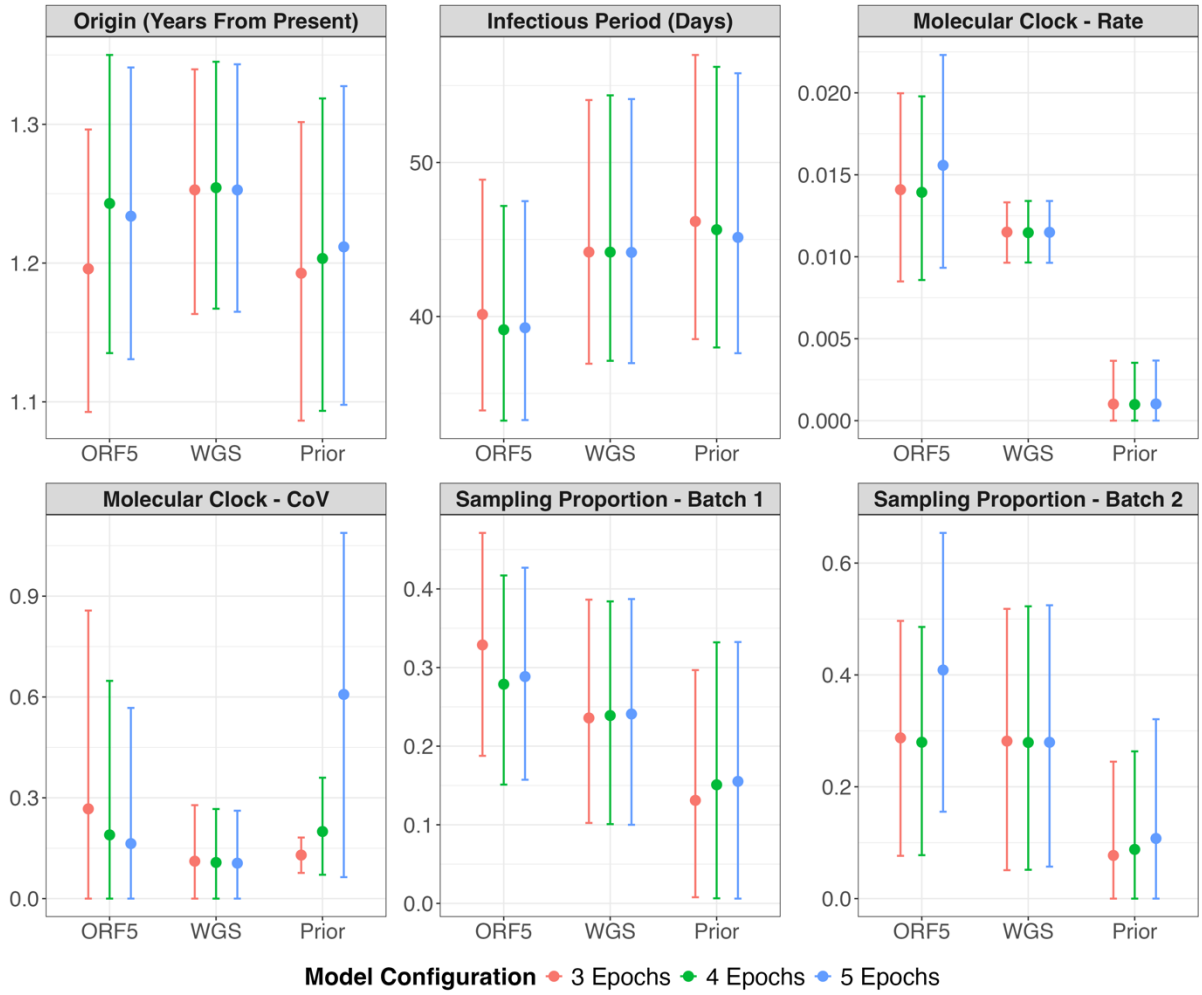


Figure 6: Sensitivity analysis of evolutionary and epidemiological parameter estimates across the Birth-Death Skyline (BDSKY) model configurations. *The plots compare the posterior mean estimates and 95% Highest Posterior Density (HPD) intervals for parameters inferred from the ORF5 and Whole Genome Sequence (WGS) datasets, alongside estimates resulting from the sample-from-prior analysis (indicated on the x-axis). All estimates are represented with colors differentiating the temporal resolutions applied: 3-epoch (orange), 4-epoch (green), and 5-epoch (blue) configurations. The comparison highlights the increased precision of the WGS dataset in estimating evolutionary parameters (Clock Rate and Coefficient of Variation (CoV)) relative to ORF5. Note that the posterior estimate of the rate of becoming uninfectious, as output by BEAST, was converted into the infectious period per animal, expressed in days. Additionally, the the 4- and 5-epoch sample-from-prior models failed to achieve an Effective Sample Size (ESS) of 200 for the CoV,*

The comparison between the data-driven and sample-from-prior analysis revealed that the prior-only analyses for both the 3- and 4-epoch configurations could not resolve distinct phases or differences in the Re (Figure 5G and H). Specifically, regarding the 4-epoch configuration, the prior-only analysis showed no evidence of the explosive outbreak peak observed in the data-driven model, exhibiting very wide 95% HPD intervals for both the Re and epoch transition estimates. Notably, the 4- and 5-epoch sample-from-prior models failed to achieve an ESS of 200 for the estimated epoch transition times, Re values, and the CoV (see Supplementary File 7). Tracer diagnostics indicated poor mixing and an inability to distinguish between multiple local optima; this was considered a stable limitation of the prior landscape that would not likely improve with additional sampling. Regarding evolutionary parameters, significant deviations were observed

between the prior and data-driven results, particularly for the molecular clock rate and CoV, confirming that the sequence data contained a strong phylogenetic signal that drove the inference.

Fine-Scale Spatial Transmission Dynamics (BDMM)

To quantify transmission dynamics at the pen and room levels and to determine whether the infectious period differed between the farrowing and nursery phases, the BDMM analysis was performed using the discrete migration configuration.

For the farrowing unit, the estimated Re value was 1.67 (95% HPD: 0.24–3.26). This estimate did not deviate substantially from the sample-from-prior analysis (Figure 7), indicating that these results were largely influenced by prior assumptions rather than the genomic data. In contrast, transmission dynamics within the nursery unit revealed a biological signal that diverged from the prior distributions. While within-pen Re was estimated to be 2.61 (95% HPD: 0.52–5.21), the Re between pens showed to significantly below the epidemiological threshold of 1. Crucially, the 95% HPD intervals for between-pen transmission were considerably narrower than the analysis run only on prior configurations. Specifically, the Re within the same room was estimated at 0.46 (95% HPD: 0.09–0.94), while transmission between rooms was even lower at 0.06 (95% HPD: 0.02–0.12). This stands in contrast to the sample-from-prior analysis, where the within-room transmission interval included 1 (mean 0.80; 95% HPD: 0.10–1.73), confirming that the discrete model successfully resolved fine-scale constraints on viral spread that were not present in the prior assumptions.

Furthermore, the model indicated no significant difference in the infectious period between the farrowing units, with a mean of 44 days (95% HPD: 26–103 days), and nursery units, with a mean of 37 days (95% HPD: 21–99 days). While the sample-from-prior analysis yielded a substantially wider 95% HPD interval for the infectious period in the nursery phase compared to the data-driven analysis, the interval for the farrowing phase remained comparable to the prior distribution. Regarding movement dynamics, the migration rates estimated in the data-driven analysis closely matched those obtained from the corresponding sample-from-prior analysis.

The posterior estimates for the sampling proportion, tree origin, and mean molecular clock rate closely matched their respective prior distributions, consistent with the narrow priors applied to these parameters. Additionally, consistent with the BDSKY analysis, the K81 model was identified as the optimal nucleotide substitution model.

Finally, to challenge the assumption of discrete one-day animal movements, a robustness test was performed using a continuous migration model. This analysis yielded highly consistent parameter estimates with the discrete configuration described above, showing no major deviations in epidemiological or evolutionary parameters (Supplementary File 8).

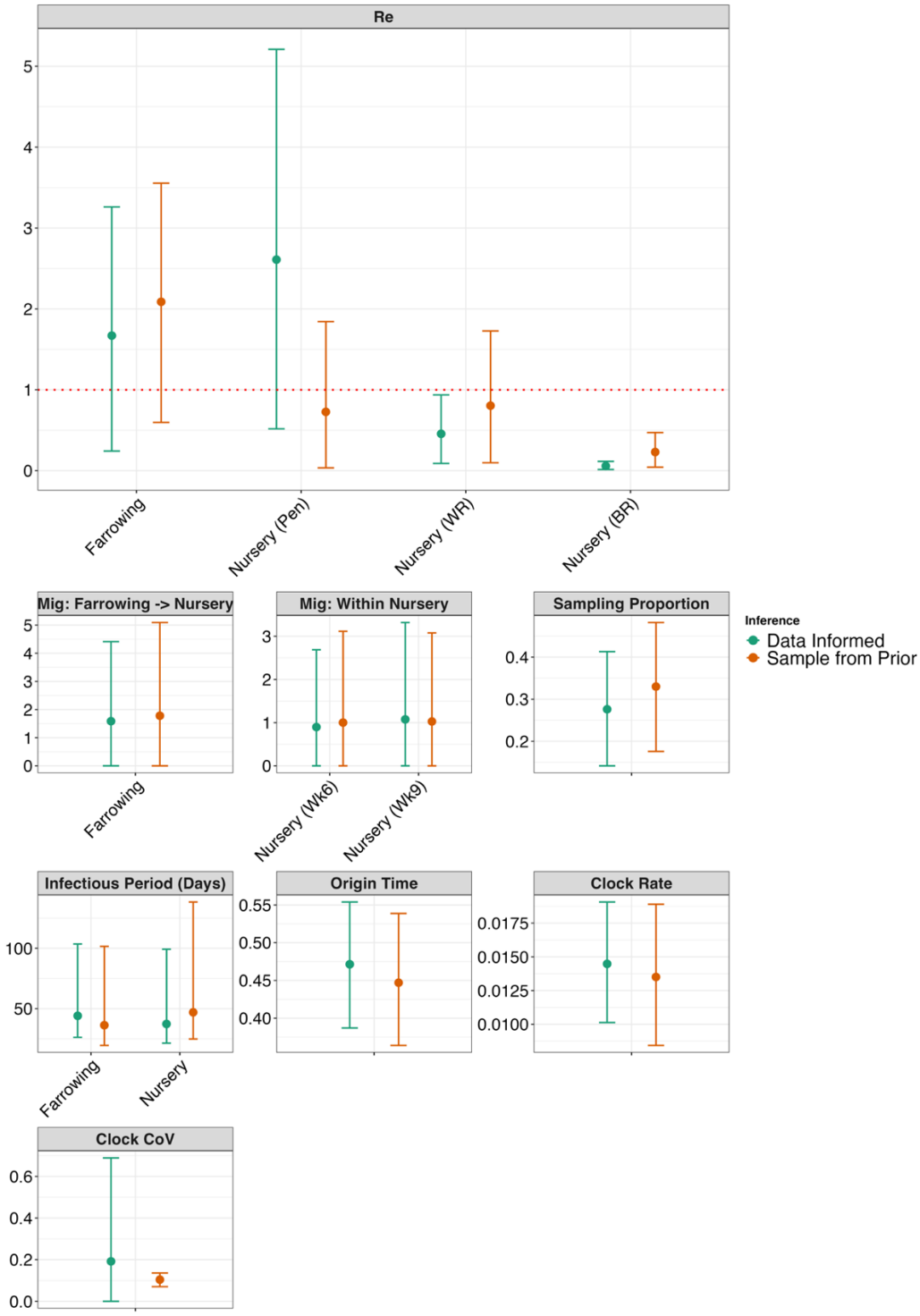


Figure 7: Comparison of posterior parameter estimates for the Birth-Death with Migration (BDMM) model using the discrete migration configuration against the sample-from-prior analysis. *The plots display the mean posterior estimates and 95% Highest Posterior Density (HPD) intervals for epidemiological and evolutionary parameters. The analysis compares the Data Informed analysis (blue) with the* sample-from-prior *analysis (orange) to validate the influence of the genetic data on the posterior estimates. The red dotted horizontal line represents the epidemiological threshold of Re = 1. Note that the posterior estimate of the rate of becoming uninfectious, as output by BEAST, was converted into the infectious period per animal, expressed in days. , BR=Between Room, CoV=Coefficient of Variation, Mig=Migration rate, Re = Effective reproduction number, WK = Week, WR = Within Room*

Discussion

The primary aim of this study was to evaluate the capacity of phylodynamic models to reliably infer previously unquantified epidemiological parameters and distinguish transmission potential across spatially distinct farm units. By leveraging genetic data to reconstruct the transmission dynamics of a PRRSV-1 outbreak, this analysis recovered the established three-phase epidemiological trajectory of outbreak, dormancy, and resurgence described by Clilverd et al. (2023). Crucially, the BDSKY model dated the resurgence event to shortly before Batch 3, supporting the hypothesis proposed by Clilverd and colleagues (2023) that this resurgence was driven by a selection process immediately preceding Batch 3 sampling, as evidenced by the limited genetic diversity of the emerging escape variant.

At the fine-scale resolution, the BDMM successfully distinguished between within- and between pen transmission in the nursery unit. While the within-pen Re were estimated to be 2.61(95% HPD: 0.52–5.21), estimates for transmission between pens were significantly below the epidemiological threshold of 1, dropping to 0.46 (95% HPD: 0.09–0.94) for pens within the same room and 0.06 (95% HPD: 0.02–0.12) for pens across different rooms. This drop in transmission efficiency is validated by our observed phylogenetic structure which displayed a strong spatial signal where clusters of 100% identical sequences consist of samples originating from the same pen and room. The 100% sequence similarity suggests that these individuals likely infected one another, forming localized transmission chains that are effectively constrained by the farm's physical compartmentalization. Additionally, experimental data for related swine viruses also support our findings. Research on Classical Swine Fever Virus (CSFV) shows that transmission within a pen is nearly 20 times faster than between pens separated by open barriers (Weesendorp et al., 2014). Similarly, the negligible Re for between-room spread aligns with PRRSV aerosol studies where the virus, though detectable in barn air, often fails to infect pigs in separate compartments despite exposure to contaminated air current (Otake et al., 2002; Trincado et al., 2004). This finding demonstrates that phylodynamic models can successfully quantify the structured transmission dynamics characteristic of farm environments. Specifically, the models reveal that while direct contact drives efficient transmission among pen-mates, horizontal spread between spatially separated units remains a rare event.

However, the Re estimated for the farrowing unit requires a distinct interpretation. Unlike the nursery estimates, this value does not reflect horizontal transmission between pen mates, as genetic analysis confirmed these infections were acquired in-utero (Clilverd et al., 2023). Instead, similar to the "ghost deme" sow reservoir, the farrowing Re serves a structural function: it bridges the temporal gap between introduction and the nursery outbreak, allowing the lineage to persist in the model without quantifying local horizontal spread.

Notably, the estimated infectious period did not differ between the farrowing and nursery phases. While this may reflect a genuine similarity between phases, the ability to detect a difference was likely constrained due to sampling sparsity: in the reconstructed tree, farrowing-phase sequences did not appear as ancestors of nursery-derived lineages, indicating that nursery infections likely arose from unsampled farrowing or nursery stage pigs or the sow reservoir. Consequently, the model could not interpret farrowing infections as long-term reservoirs bridging the nursery stage. This highlights a significant limitation of the model's "isolated" assumption, which treated the farrowing and nursery units as closed systems after the initial introduction from the sow herd. Moreover, the large 95% HPD intervals associated with the infectious period estimates in both phases obscured the detection of a statistically meaningful difference. This uncertainty can be explained by a parameter non-identifiability issue, in which the model cannot distinguish between “fast turnover” (high transmission with rapid recovery) and “slow turnover” (low transmission with prolonged infectiousness) scenarios because both yield indistinguishable likelihoods (Zarebski et

al., 2022). The impact of this limitation is apparent in the nursery unit results, where both within-pen Re and the infectious period were associated with wide 95% HPD intervals. A commonly adopted approach to alleviate such identifiability issues is to fix the rate of becoming uninfectious using a well-supported prior (Weber et al., 2023).

A critical outcome of this study is the evaluation of genomic resolution on model performance, revealing complementary strengths between targeted gene sequencing and whole-genome approaches. The larger ORF5 dataset produced more confident epidemiological estimates for the Re and transition times than the sparse WGS dataset, supporting the argument that high taxon density is required to distinguish epidemiological rate shifts from background noise in BDSKY models (Culshaw et al., 2019). Notably, the higher resolution of the ORF5 dataset resolved a brief, intense transmission peak during the early outbreak that the WGS dataset failed to capture. We hypothesize that this peak reflects the superspreading event reported by Clilverd and colleagues (2023), in which a single pig accounted for 28.6% of all traced horizontal transmission events. Conversely, the WGS dataset provided superior resolution for evolutionary parameters, yielding narrower 95% HPD intervals for the clock rate and CoV, and a stronger temporal signal (R^2=0.992) compared to ORF5 (R^2=0.836). This aligns with the 'temporal horizon' concept (Dudas & Bedford, 2019), where longer sequences provide the mutational density necessary for precise clock calibration. This increased information density likely also drove the inference of a more complex substitution model for the WGS dataset (Model 24) compared to ORF5 (K81), suggesting that the exceptional genetic diversity of PRRSV requires data-driven models rather than reliance on default assumptions. Consequently, researchers should select genomic resolution based on specific objectives: high-density, targeted sequencing is recommended for monitoring fine-scale epidemiological dynamics and detecting superspreading events. Conversely, longer sequences should be prioritized for evolutionary precision, as increased mutational density enables accurate molecular clock calibration. While not the primary focus of this study, such precision is essential for analyses aiming to determine the origin of an outbreak by estimating the tMRCA.

A key strength of the BDSKY analysis lies in the strict parameterization of the sampling proportion to mirror the sampling schedule. This fixed configuration was necessary to distinguish sampling gaps from low viral prevalence and prevent poor statistical identifiability (Featherstone et al., 2021; Parag et al., 2022). By preventing the model from inherently interpreting the lack of samples during the dormant phase as a halt of transmission, this strategy ensured that parameter estimates remained robust and driven by genomic signals rather than prior configurations. This was further validated by sample-from-prior analyses, which confirmed that the prior configuration of the model alone could not resolve the epidemiological phases. However, a key limitation of the BDSKY model is its assumption of an unstructured population which is violated according to our BDMM analysis that confirmed that transmission was dependent on spatial units. Consequently, ignoring this structure can lead to biased transmission rate estimates, as the model incorrectly assumes equal contact probabilities across the entire herd (Seidel et al., 2024). Specifically, the model may misinterpret localized viral persistence within a single pen as broader, herd-wide transmission, potentially inflating the Re estimate (Seidel et al., 2024; Volz & Frost, 2014). Accordingly, it should be noted that the Re values reported in the BDSKY analysis may represent overestimates of the true transmission intensity.

Furthermore, the study faced several further strengths and limitations. The clade-based analysis effectively differentiated persistent infections from reinfections, supported by the recovery of the single reinfection event (pig 515) reported by Clilverd et al. (2023) and robustness checks confirming that infection classifications remained consistent across varying numbers of estimated clades. However, the ORF5 dataset failed to maintain monophyly for one lineage, probably due to a high proportion of identical sequences. This aligns with the finding that high sequence identity increases "monophyly-violating interclass coalescence," increasing the likelihood that lineages mix

before coalescing (Mehta et al., 2016). To overcome the limitation of low unique genetic variation, future studies could leverage intra-host diversity as demonstrated by Ortiz et al. (2023). By tracking shared minor variants that are typically discarded in consensus sequencing, this method resolves transmission links between hosts with identical consensus sequences, effectively distinguishing specific transmission pathways within otherwise unresolved clusters. Furthermore, the finding that transmission is predominantly confined to within-pen dynamics introduces a challenge regarding the "burnout" phenomenon (Parsons et al., 2024). In these small groups, the virus quickly exhausts the limited pool of susceptible pigs, causing outbreaks to terminate abruptly. This premature truncation of transmission chains can bias Re estimates, as standard models assume ongoing spread in a larger susceptible population rather than extinction due to host depletion.

To overcome the limitations associated with data sparsity and structure, future study designs targeting within-farm transmission should be optimized for phylodynamic inference. We recommend a shift from fixed longitudinal sampling of the same individuals to a rotational strategy. By replacing repeated sampling of already infected individuals at subsequent time points with sampling of previously unsampled pen-mates or animals that have not yet tested positive, researchers can avoid redundancy arising from persistent infections and instead maximize the recovery of distinct transmission events. Furthermore, sampling must extend beyond the farrowing and nursery units to include the sow herd to replace the "ghost deme" assumption with empirical data, enabling the direct quantification of vertical transmission and spillover events. Finally, future phylodynamic studies that aim to characterize transmission between isolated compartments should incorporate metadata on personnel movement and ventilation to better distinguish transmission pathways. Arruda and colleagues (2019) highlight that distinguishing between aerosol transmission and local spread via fomites is often confounded by a lack of data on internal biosecurity measures and contact networks. Integrating these factors is methodologically feasible in phylodynamic models, as Meester et al. (2025) recently used BDMM to successfully distinguish environmental from direct transmission routes. Therefore, detailed metadata would allow models to explicitly parameterize the physical and procedural barriers defining subpopulation interactions.

In conclusion, this study demonstrates that phylodynamic models can successfully reconstruct within-farm PRRSV-1 outbreak dynamics and quantify fine-scale transmission parameters using standard surveillance data. By confirming the three-phase epidemic trajectory and distinguishing within-pen spread from between-pen transmission, our findings validate the utility of phylodynamics for resolving epidemiological events that traditional methods may miss. The analysis also highlights critical trade-offs: while targeted ORF5 sequencing offers the high sampling density necessary to resolve rapid transmission shifts, whole-genome sequencing provides the mutational resolution required for robust evolutionary clock calibration. Ultimately, moving from observational surveillance to quantitative, genetics-driven modeling allows for a more nuanced understanding of outbreak mechanics, providing the foundation needed to refine internal biosecurity and targeted intervention strategies in the swine industry

Acknowledgements

I would like to express my sincere gratitude to Dr. M. Meester for providing excellent daily supervision and guidance throughout this project. This research was made possible through the support of the Faculty of Veterinary Medicine, Department of Population Health Sciences at Utrecht University, which provided the opportunity to perform this internship within their research environment. I am also grateful to all the staff members of the department for their valuable insights and feedback during our regular meetups and my presentation.

Special thanks are extended to Dr. Timothy Vaughan from ETH Zürich, one of the primary developers of BEAST. I am deeply appreciative of the time he took to meet with me personally to discuss the specific model settings required for the Birth-Death with Migration (BDMM) analysis. I would also like to thank Harsh Bardhan Gupta, a PhD student at Utrecht University, for our in-person discussions regarding the interpretation of results and for his helpful suggestions on the layout possibilities for the phylogenetic trees.

Finally, I acknowledge the use of Gemini, an artificial intelligence tool, which assisted in generating the basic layout for the computational code and served as a resource for debugging throughout the data analysis process.

References

Alkhamis, M. A., Fountain-Jones, N. M., Khajah, M. M., Alghounaim, M., & Al-Sabah, S. K. (2022). Comparative phylodynamics reveals the evolutionary history of SARS-CoV-2 emerging variants in the Arabian Peninsula. *Virus Evolution*, *8*(1), veac040.

Alkhamis, M. A., Perez, A. M., Murtaugh, M. P., Wang, X., & Morrison, R. B. (2016). Applications of Bayesian phylodynamic methods in a recent US porcine reproductive and respiratory syndrome virus outbreak. *Frontiers in Microbiology*, *7*, 67.

Andrade, J., & Duggan, J. (2022). Inferring the effective reproductive number from deterministic and semi-deterministic compartmental models using incidence and mobility data. *PLoS Computational Biology*, *18*(6), e1010206.

Arruda, A. G., Tousignant, S., Sanhueza, J., Vilalta, C., Poljak, Z., Torremorell, M., Alonso, C., & Corzo, C. A. (2019). Aerosol detection and transmission of porcine reproductive and respiratory syndrome virus (PRRSV): What is the evidence, and what are the knowledge gaps? *Viruses*, *11*(8), 712.

Balka, G., Podgórska, K., Brar, M. S., Bálint, Á., Cadar, D., Celer, V., Dénes, L., Dirbakova, Z., Jedryczko, A., Márton, L., & others. (2018). Genetic diversity of PRRSV 1 in Central Eastern Europe in 1994–2014: Origin and evolution of the virus in the region. *Scientific Reports*, *8*(1), 7811.

Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, *17*(1), 42.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., & others. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), e1006650.

Bruen, T. C., Philippe, H., & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics*, *172*(4), 2665–2681.

Bryant, J. E., Holmes, E. C., & Barrett, A. D. T. (2007). Out of Africa: A molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathogens*, *3*(5), e75.

Butler, J., Lager, K., Golde, W., Faaberg, K. S., Sinkora, M., Loving, C., & Zhang, Y. (2014). Porcine reproductive and respiratory syndrome (PRRS): An immune dysregulatory pandemic. *Immunologic Research*, *59*(1), 81–108.

Caserta, L. C., Zhang, J., Piñeyro, P., & Diel, D. G. (2023). Rapid genotyping of porcine reproductive and respiratory syndrome virus (PRRSV) using MinION nanopore sequencing. *PLoS One*, *18*(5), e0282767.

Chandra, S., Cezar, G., Rupasinghe, K., Magalhães, E., Silva, G. S., Almeida, M., Crim, B., Burrough, E., Gauger, P., Madson, D., & others. (2025). Harnessing sequencing data for porcine reproductive and respiratory syndrome virus (PRRSV): Tracking genetic evolution dynamics and emerging sequences in US swine industry. *Frontiers in Veterinary Science*, *12*, 1571020.

Charpin, C., Mahé, S., Keranflec'h, A., Belloc, C., Cariolet, R., Le Potier, M.-F., & Rose, N. (2012). Infectiousness of pigs infected by the Porcine Reproductive and Respiratory Syndrome virus (PRRSV) is time-dependent. *Veterinary Research*, *43*(1), 69.

Clilverd, H., Martín-Valls, G., Li, Y., Martín, M., Cortey, M., & Mateu, E. (2023). Infection dynamics, transmission, and evolution after an outbreak of porcine reproductive and respiratory syndrome virus. *Frontiers in Microbiology*, *14*, 1109881.

Collins, J. E., Benfield, D. A., Christianson, W. T., Harris, L., Hennings, J. C., Shaw, D. P., Goyal, S. M., McCullough, S., Morrison, R. B., Joo, H. S., & others. (1992). Isolation of swine

750 infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America
751 and experimental reproduction of the disease in gnotobiotic pigs. *Journal of*
752 *Veterinary Diagnostic Investigation*, *4*(2), 117–126.
753 Culshaw, V., Stadler, T., & Sanmartín, I. (2019). Exploring the power of Bayesian birth-death
754 skyline models to detect mass extinction events from phylogenies with only extant
755 taxa. *Evolution*, *73*(6), 1133–1150.
756 Dearlove, B. L., Xiang, F., & Frost, S. D. (2017). Biased phylodynamic inferences from
757 analysing clusters of viral sequences. *Virus Evolution*, *3*(2), vex020.
758 Driebe, E. M., Sahl, J. W., Roe, C., Bowers, J. R., Schupp, J. M., Gillece, J. D., Kelley, E., Price, L.
759 B., Pearson, T. R., Hepp, C. M., & others. (2015). Using whole genome analysis to
760 examine recombination across diverse sequence types of Staphylococcus aureus.
761 *PLoS One*, *10*(7), e0130955.
762 Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian evolutionary analysis with BEAST*.
763 Cambridge University Press.
764 Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics
765 and dating with confidence. *PLoS Biology*, *4*(5), e88.
766 Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., & Baele,
767 G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus*
768 *Evolution*, *6*(2), veaa061.
769 Dudas, G., & Bedford, T. (2019). The ability of single genes vs full genomes to resolve time
770 and space in outbreak analysis. *BMC Evolutionary Biology*, *19*(1), 232.
771 Ecker, N., Huchon, D., Mansour, Y., Mayrose, I., & Pupko, T. (2024). A machine-learning-based
772 alternative to phylogenetic bootstrap. *Bioinformatics*, *40*(Supplement_1), i208–i217.

773 Ewing, G., & Rodrigo, A. (2006). Estimating population parameters using the structured serial
774 coalescent with Bayesian MCMC inference when some demes are hidden.
775 *Evolutionary Bioinformatics*, *2*, 117693430600200026.
776 Featherstone, L. A., Di Giallonardo, F., Holmes, E. C., Vaughan, T. G., & Duchêne, S. (2021).
777 Infectious disease phylodynamics with occurrence data. *Methods in Ecology and*
778 *Evolution*, *12*(8), 1498–1507.
779 Featherstone, L. A., Duchene, S., & Vaughan, T. G. (2023). Decoding the fundamental drivers
780 of phylodynamic inference. *Molecular Biology and Evolution*, *40*(6), msad132.
781 Featherstone, L. A., Rambaut, A., Duchene, S., & Wirth, W. (2024). Clockor2: Inferring global
782 and local strict molecular clocks using root-to-tip regression. *Systematic Biology*,
783 *73*(3), 623–628.
784 Franzo, G., Barbierato, G., Pesente, P., Legnardi, M., Tucciarone, C. M., Sandri, G., & Drigo, M.
785 (2021). Porcine reproductive and respiratory syndrome (Prrs) epidemiology in an
786 integrated pig company of northern Italy: A multilevel threat requiring multilevel
787 interventions. *Viruses*, *13*(12), 2510.
788 Franzo, G., Faustini, G., Legnardi, M., Cecchinato, M., Drigo, M., & Tucciarone, C. M. (2022).
789 Phylodynamic and phylogeographic reconstruction of porcine reproductive and
790 respiratory syndrome virus (PRRSV) in Europe: Patterns and determinants.
791 *Transboundary and Emerging Diseases*, *69*(5), e2175–e2184.
792 Frias-De-Diego, A., Jara, M., Pecoraro, B. M., & Crisci, E. (2021). Whole genome or single
793 genes? A phylodynamic and bibliometric analysis of PRRSV. *Frontiers in Veterinary*
794 *Science*, *8*, 658512.

795 Gavryushkina, A., Welch, D., Stadler, T., & Drummond, A. J. (2014). Bayesian inference of
796 sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational*
797 *Biology*, *10*(12), e1003919.
798 Guinat, C., Tang, H., Yang, Q., Valenzuela Agüí, C., Vaughan, T. G., Scire, J., Yu, H., Wang, W.,
799 Chen, Z., Ducatez, M. F., & others. (2023). Bayesian phylodynamics reveals the
800 transmission dynamics of avian influenza A (H7N9) virus at the human–live bird
801 market interface in China. *Proceedings of the National Academy of Sciences*, *120*(17),
802 e2215610120.
803 Guinat, C., Vergne, T., Kocher, A., Chakraborty, D., Paul, M. C., Ducatez, M., & Stadler, T.
804 (2021). What can phylodynamics bring to animal health research? *Trends in Ecology*
805 *& Evolution*, *36*(9), 837–847.
806 Huson, D. H., & Bryant, D. (2024). The SplitsTree App: Interactive analysis and visualization
807 using phylogenetic trees and networks. *Nature Methods*, *21*(10), 1773–1774.
808 Johnson, A. D. (2010). An extended IUPAC nomenclature code for polymorphic nucleic acids.
809 *Bioinformatics*, *26*(10), 1386–1389.
810 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017).
811 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature*
812 *Methods*, *14*(6), 587–589.
813 Kappes, M. A., & Faaberg, K. S. (2015). PRRSV structure, replication and recombination:
814 Origin of phenotype and genotype diversity. *Virology*, *479*, 475–486.
815 Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster*
816 *analysis*. John Wiley & Sons.

817 Kühnert, D., Stadler, T., Vaughan, T. G., & Drummond, A. J. (2016). Phylodynamics with
818 migration: A computational framework to quantify population structure from
819 genomic data. *Molecular Biology and Evolution*, *33*(8), 2102–2116.
820 Li, C., Xu, H., Zhao, J., Gong, B., Sun, Q., Xiang, L., Li, W., Guo, Z., Li, J., Tang, Y., & others.
821 (2022). Epidemiological investigation and genetic evolutionary analysis of PRRSV-1 on
822 a pig farm in China. *Frontiers in Microbiology*, *13*, 1067173.
823 Lozano-Fernandez, J. (2022). A practical guide to design and assess a phylogenomic study.
824 *Genome Biology and Evolution*, *14*(9), evac129.
825 Meester, M., Valenzuela Agüí, C., Tobias, T. J., Honing, R. W. H. van der, Guinat, C.,
826 Bouwknegt, M., du Plessis, L., Fischer, E. A., Spaninks, M., Stadler, T., & others.
827 (2025). Zoonotic hepatitis E virus spreads through environmental routes in pig herds–
828 A phylodynamic analysis. *PLoS Pathogens*, *21*(11), e1013710.
829 Mehta, R. S., Bryant, D., & Rosenberg, N. A. (2016). The probability of monophyly of a
830 sample of gene lineages on a species tree. *Proceedings of the National Academy of*
831 *Sciences*, *113*(29), 8002–8009.
832 Müller, N. F., Bouckaert, R. R., Wu, C.-H., & Bedford, T. (2025). MASCOT-Skyline integrates
833 population and migration dynamics to enhance phylogeographic reconstructions.
834 *PLOS Computational Biology*, *21*(9), e1013421.
835 Murtaugh, M. P., Stadejek, T., Abrahante, J. E., Lam, T. T., & Leung, F. C.-C. (2010). The ever-
836 expanding diversity of porcine reproductive and respiratory syndrome virus. *Virus*
837 *Research*, *154*(1–2), 18–30.
838 Ortiz, A. T., Kendall, M., Storey, N., Hatcher, J., Dunn, H., Roy, S., Williams, R., Williams, C.,
839 Goldstein, R. A., Didelot, X., & others. (2023). Within-host diversity improves

840 phylogenetic and transmission reconstruction of SARS-CoV-2 outbreaks. *Elife*, *12*,
841 e84384.
842 Otake, S., Dee, S., Jacobson, L., Pijoan, C., & Torremorell, M. (2002). Evaluation of aerosol
843 transmission of porcine reproductive and respiratory syndrome virus under
844 controlled field conditions. *Veterinary Record*, *150*(26), 804–808.
845 Pamornchainavakul, N., Makau, D. N., Paploski, I. A., Corzo, C. A., & VanderWaal, K. (2023).
846 Unveiling invisible farm-to-farm PRRSV-2 transmission links and routes through
847 transmission tree and network analysis. *Evolutionary Applications*, *16*(10), 1721–
848 1734.
849 Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and
850 evolutionary analyses in R. *Bioinformatics*, *35*, 526–528.
851 https://doi.org/10.1093/bioinformatics/bty633
852 Parag, K. V., Pybus, O. G., & Wu, C.-H. (2022). Are skyline plot-based demographic estimates
853 overly dependent on smoothing prior assumptions? *Systematic Biology*, *71*(1), 121–
854 138.
855 Parisio, G., Franzo, G., Barbieri, I., Carta, V., Stadejek, T., Manenti, S., Campagna, D., Faccini,
856 S., Vignola, G., Alborali, G. L., & others. (2024). Evolutionary dynamics of PRRS virus
857 in Italian Pig farms: A retrospective study. *Virology Journal*, *21*(1), 326.
858 Parsons, T. L., Bolker, B. M., Dushoff, J., & Earn, D. J. (2024). The probability of epidemic
859 burnout in the stochastic SIR model with vital dynamics. *Proceedings of the National*
860 *Academy of Sciences*, *121*(5), e2313708120.
861 Pileri, E., & Mateu, E. (2016). Review on the transmission porcine reproductive and
862 respiratory syndrome virus between pigs and farms and impact on vaccination.
863 *Veterinary Research*, *47*(1), 108.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, *6*(1), 7–11.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Revell, L. J. (2024). phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, *12*, e16505. https://doi.org/10.7717/peerj.16505

Rowland, R. R., Lawson, S., Rossow, K., & Benfield, D. A. (2003). Lymphoid tissue tropism of porcine reproductive and respiratory syndrome virus replication during persistent infection of pigs originally exposed to virus in utero. *Veterinary Microbiology*, *96*(3), 219–235.

Scire, J., Barido-Sottani, J., Kühnert, D., Vaughan, T. G., & Stadler, T. (2022). Robust phylodynamic analysis of genetic sequencing data from structured populations. *Viruses*, *14*(8), 1648.

Seidel, S., Stadler, T., & Vaughan, T. G. (2024). Estimating pathogen spread using structured coalescent and birth–death models: A quantitative comparison. *Epidemics*, *49*, 100795.

Sequeira, S. C., Sebunia, N., Page, J. R., Lasisi, T., Habing, G., & Arruda, A. G. (2025). A systematic scoping review and thematic analysis: How can livestock and poultry movement networks inform disease surveillance and control at the global scale? *Plos One*, *20*(7), e0328518.

Sha, H., Lan, X., Yang, Z., Lv, C., Zhang, H., Luo, Q., Zheng, Y., Li, G., Kong, W., Huang, L., & others. (2025). Genetic variation in NSP4 of type 1 porcine reproductive and respiratory syndrome virus in China. *BMC Veterinary Research*, *21*(1), 374.

Stadler, T., Kühnert, D., Bonhoeffer, S., & Drummond, A. J. (2013). Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*, *110*(1), 228–233.

Stadler, T., Magnus, C., Vaughan, T. G., Barido-Sottani, J., Bošková, V., Huisman, J., & Pečerska, J. (2024). *Decoding genomes: From sequences to phylodynamics*. ETH Zurich.

Trincado, C., Dee, S., Otake, S., Pijoan, C., Jacobson, L., & Rossow, K. (2004). Attempts to transmit porcine reproductive and respiratory syndrome virus by aerosols under controlled field conditions. *Veterinary Record*, *154*(10), 294–297.

Vaughan, T. G., & Stadler, T. (2025). Bayesian Phylodynamic Inference of Multitype Population Trajectories Using Genomic Data. *Molecular Biology and Evolution*, *42*(6), msaf130.

Volz, E. M., & Frost, S. D. (2014). Sampling through time and phylodynamic inference with coalescent and birth–death models. *Journal of The Royal Society Interface*, *11*(101), 20140945.

Weber, A., Översti, S., & Kühnert, D. (2023). Reconstructing relative transmission rates in Bayesian phylodynamics: Two-fold transmission advantage of Omicron in Berlin, Germany during December 2021. *Virus Evolution*, *9*(2), vead070.

Weesendorp, E., Backer, J., & Loeffen, W. (2014). Quantification of different classical swine fever virus transmission routes within a single compartment. *Veterinary Microbiology*, *174*(3–4), 353–361.

Weng, C., Huang, X., Chen, Z., He, M., Zhang, B., Li, H., Xie, J., Chen, M., Qiu, L., Li, X., & others. (2025). Genetic evolution, epidemic trends, and recombination dynamics of PRRSV-1 in China. *Frontiers in Veterinary Science*, *12*, 1632917.

912 Wensvoort, G., Terpstra, C., Pol, J., Ter Laak, E., Bloemraad, M., De Kluyver, E., Kragten, C.,
913 Van Buiten, L. den, Den Besten, A., Wagenaar, F., & others. (1991). Mystery swine
914 disease in The Netherlands: The isolation of Lelystad virus. *Veterinary Quarterly*,
915 *13*(3), 121–130.
916 Wills, R. W., Doster, A. R., Galeota, J. A., Sur, J.-H., & Osorio, F. A. (2003). Duration of infection
917 and proportion of pigs persistently infected with porcine reproductive and
918 respiratory syndrome virus. *Journal of Clinical Microbiology*, *41*(1), 58–62.
919 Wong, T. K., Ly-Trong, N., Ren, H., Baños, H., Roger, A. J., Susko, E., Bielow, C., De Maio, N.,
920 Goldman, N., Hahn, M. W., & others. (2025). *IQ-TREE 3: Phylogenomic Inference*
921 *Software using Complex Evolutionary Models*. https://doi.org/10.32942/X2P62N
922 Wright, E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R*
923 *Journal*, *8*(1), 352–359.
924 Wu, Z., Chang, T., Wang, D., Zhang, H., Liu, H., Huang, X., Tian, Z., Tian, X., Liu, D., An, T., &
925 others. (2024). Genomic surveillance and evolutionary dynamics of type 2 porcine
926 reproductive and respiratory syndrome virus in China spanning the African swine
927 fever outbreak. *Virus Evolution*, *10*(1), veae016.
928 You, X., Lei, Y., Zhang, P., Xu, D., Ahmed, Z., & Yang, Y. (2022). Role of transcription factors in
929 porcine reproductive and respiratory syndrome virus infection: A review. *Frontiers in*
930 *Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.924004
931 Zarebski, A. E., du Plessis, L., Parag, K. V., & Pybus, O. G. (2022). A computationally tractable
932 birth-death model that combines phylogenetic and epidemiological data. *PLOS*
933 *Computational Biology*, *18*(2), e1009805.
934 Zhang, J., Zheng, Y., Xia, X.-Q., Chen, Q., Bade, S. A., Yoon, K.-J., Harmon, K. M., Gauger, P. C.,
935 Main, R. G., & Li, G. (2017). High-throughput whole genome sequencing of porcine

reproductive and respiratory syndrome virus from cell culture materials and clinical specimens using next-generation sequencing technology. *Journal of Veterinary Diagnostic Investigation*, *29*(1), 41–50.

Zhou, L., Han, J., & Yang, H. (2024). The evolution and diversity of porcine reproductive and respiratory syndrome virus in China. *Veterinary Microbiology*, *298*, 110252.