

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

1 **Abstract**

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

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

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

29 **Introduction**

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

45 It is known that PRRSV infection dynamics are highly age-dependent in both genetic variants:
46 neonatally infected pigs act as long-term reservoirs capable of shedding virus for up to 250 days
47 (Butler et al., 2014; Pileri & Mateu, 2016; Rowland et al., 2003; Wills et al., 2003; You et al.,
48 2022), whereas pigs infected post-birth typically clear infection within 42 days (Charpin et al.,
49 2012). Therefore, while vertical transmission often initiates circulation, prolonged shedding by
50 neonatally infected pigs bridges production stages and sustains endemicity through subsequent
51 horizontal spread (Pileri & Mateu, 2016). Furthermore, both species exhibit substantial genetic
52 variability driven by high mutation and recombination rates typical of positive-sense RNA viruses
53 (Franzo et al., 2021; Kappes & Faaberg, 2015; Parisio et al., 2024; Wu et al., 2024; Zhou et al.,
54 2024). Given this exceptional genetic diversity and rate heterogeneity, universal substitution
55 models are often unsuitable for characterizing PRRSV evolution (Balka et al., 2018; Gong et al.,
56 2024). While this fast evolution complicates disease control in PRRSV (Caserta et al., 2023), it
57 renders it particularly suitable for phylodynamic modelling (Duchene et al., 2020; Featherstone et
58 al., 2023). Unlike phylogenetics, which primarily focuses on evolutionary relationships and viral
59 classification (Frias-De-Diego et al., 2021), phylodynamics combines genetic and epidemiological
60 data to quantify pathogen transmission dynamics (Stadler et al., 2024). Central to these dynamics
61 is the effective reproductive number (R_e), defined as the expected number of secondary infections
62 caused by a single infected individual at a specific point in time (Stadler et al., 2024). By
63 estimating the R_e through time, it becomes possible to assess the epidemic trajectory relative to the
64 epidemiological threshold of 1, distinguishing between periods of growth ($R_e > 1$) and decline ($R_e < 1$)
65 (Andrade & Duggan, 2022).

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

72 Extensive research has successfully characterized transmission dynamics within large-scale
73 farming networks (Sequeira et al., 2025). For PRRSV-2, phylodynamic models have been utilized
74 to infer farm-to-farm transmission pathways and quantify farm-level transmissibility
75 (Pamornchainavakul et al., 2023), as well as to trace the historical migration of viral lineages across
76 different production companies (Alkhamis et al., 2016). Similarly, for PRRSV-1, migration of
77 lineages has been reconstructed across Europe to identify determinants of spread (Franzo et al.,
78 2021).

81 2022), and the role of pig movements in shaping viral diversity in Central Eastern Europe has been
82 investigated (Balka et al., 2018). Furthermore, Franzo et al. (2021) showed how phylodynamics
83 can track transmission within integrated pig companies, where farrowing, nursery, and finishing
84 units are managed by a single organization.

85
86 Complementing these studies, research at the within-farm resolution has largely relied on
87 phylogenetic and evolutionary approaches. For instance, phylogenetic analysis has been used to
88 investigate specific introduction and circulation events within a single farm (Li et al., 2022), while
89 the evolutionary dynamics and selection pressures facilitating persistence in vaccinated breeding
90 herds have also been characterized (Clilverd et al., 2024).

91
92 Despite these advances, quantitative phylodynamic estimates for transmission parameters at the
93 within-farm level remain limited (Meester et al., 2025). Applying these models to a single herd
94 presents distinct challenges, primarily because farm populations do not mix randomly. Instead, they
95 are highly structured environments where contact is constrained by physical barriers, such as
96 separation into specific pens and rooms. Consequently, the suitability of phylodynamic models to
97 resolve transmission dynamics within such a compartmentalized system, especially where viral
98 genetic diversity is constrained and sparse sampling can obscure rapid transmission chains, remains
99 to be fully validated (Alkhamis et al., 2016; Guinat et al., 2023; Meester et al., 2025).

100
101 To address these knowledge gaps, this study leverages the dataset from Clilverd and colleagues
102 (2023) to evaluate whether phylodynamic models can reliably infer within-farm transmission
103 dynamics. Unlike the original study, which characterized the epidemic utilizing longitudinal
104 incidence monitoring and phylogenetic contact tracing, this analysis validates the utility of
105 phylodynamic models by investigating their ability to reconstruct the established epidemiological
106 trajectory by Clilverd and colleagues (2023), differentiate within-pen transmission from horizontal
107 spread between pens sharing the same room versus those in separate rooms, detect differences in
108 infectious periods between production stages, and assess the impact of genomic resolution (WGS
109 vs. ORF5) on inference reliability. Following validation of the temporal signal and absence of
110 recombination, unstructured Birth-Death Skyline (BDSKY) and structured Birth-Death with
111 Migration (BDMM) models were applied to quantify these long-term and fine-scale transmission
112 parameters, respectively.

113 Methods

114 Study Design and Sequence Acquisition

115 The virological and epidemiological data used for this phylodynamic analysis were obtained from
 116 a longitudinal study by Clilverd et al. (2023), which monitored a PRRSV-1 outbreak on a Spanish
 117 300-sow farrow-to-wean farm. The original study followed three distinct batches of piglets from
 118 birth to nine weeks of age at 1.5 months (Batch 1), 8 months (Batch 2), and 12 months (Batch 3)
 119 after the clinical detection of the outbreak, which was characterized by the sudden appearance of
 120 abortions, stillbirths, and weak-born animals. For all batches, umbilical cords were collected at
 121 birth, and blood samples were drawn from individually ear-tagged piglets at 2, 4, 6, and 9 weeks
 122 of age. Viral RNA was extracted from collected samples and screened for PRRSV-1 using RT-
 123 qPCR. For samples with sufficient viral load ($C_t \leq 32.0$), viral sequencing was performed. This
 124 generated ORF5 sequences via Sanger sequencing for animals in Batch 1 and Batch 3. Additionally,
 125 WGS were generated from viral isolates using an Illumina MiSeq platform. No sequences could be
 126 obtained from Batch 2. This phase was characterized by an absence of clinical symptoms and only
 127 sporadic PCR-positive results (8/74 animals) with high C_t values (>31.5), which prevented the
 128 recovery of sufficient genetic material for sequencing. All consensus sequences corresponding to
 129 the WGS and ORF5, regions were deposited by Clilverd and colleagues (2023) in GenBank under
 130 accession numbers OP688189–OP688223 and OP688224–OP688357, respectively.

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

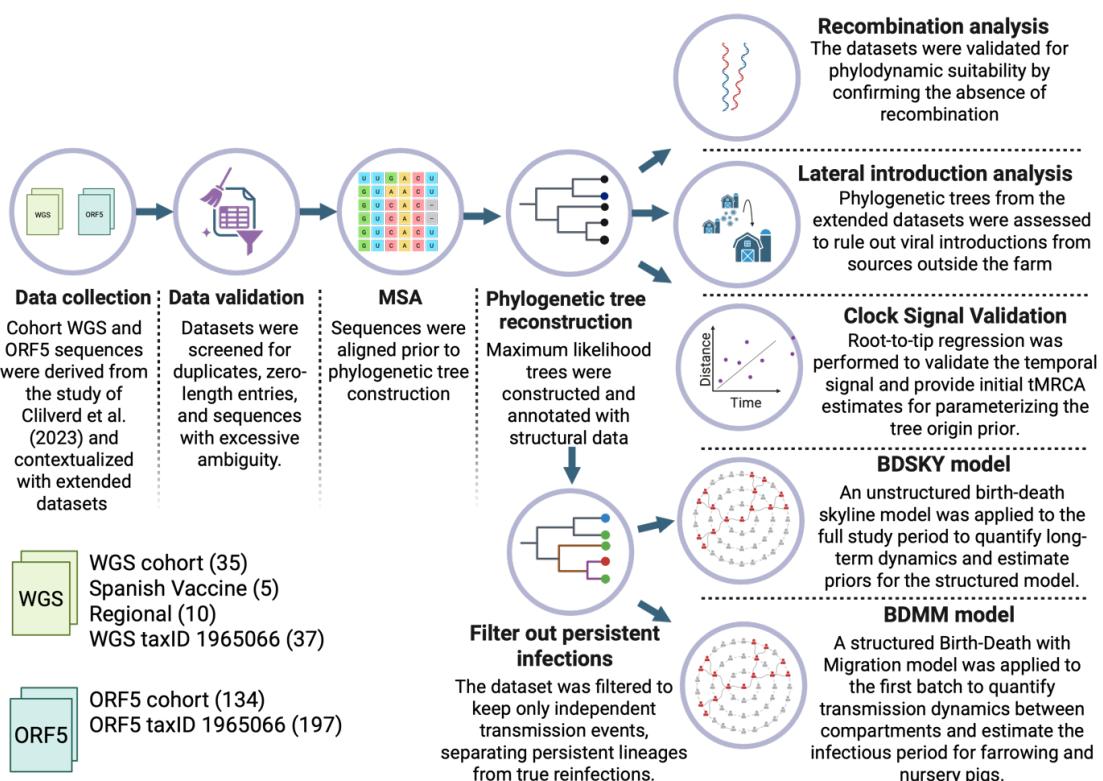


Figure 1. Schematic overview of the study workflow. The diagram shows computational pipeline used to analyze the PRRSV-1 outbreak. The process moves from Data Collection of WGS and ORF5 sequences (with sequence counts shown in brackets) to Data Validation and MSA (Multiple Sequence Alignment). Following 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. Created with BioRender.com.

- 142 **Data processing and validation**
143 All datasets were screened for duplicate or zero-length entries, invalid characters, and International
144 Union of Pure and Applied Chemistry (IUPAC) ambiguity codes (Johnson, 2010). Sequences
145 displaying excessive ambiguity, indicative of poor sequencing quality, were excluded from the
146 datasets. However, sporadic ambiguity codes were retained, as these may represent biological
147 signals of intra-host viral diversity rather than sequencing errors. Additionally, a pairwise identity
148 analysis was performed on the cohort WGS and ORF5 datasets to identify and cluster 100%
149 identical sequences.
150
151 To enable further downstream analyses, multiple sequence alignments (MSAs) were constructed
152 using the DECIPHER v3.4.0 package (Wright, 2016) on the extended WGS and ORF5 datasets for
153 recombination and phylogenetic analyses.
154
155 **Classification of Independent Transmission Events**
156 To ensure the phylodynamic model was informed by independent transmission events, a clade-
157 based analysis was performed to differentiate between sequences representing prolonged viral
158 persistence (within-host evolution) and those from new reinfection events in pigs sampled multiple
159 times. This step was essential as standard birth-death models assume that every branching event in
160 the phylogenetic tree represents a transmission event between hosts (Stadler et al., 2024). Including
161 these sequences, which resemble intra-host evolutionary variants results in over-sampling and false
162 clusters that bias growth rate estimates upwards (Dearlove et al., 2017) and can lead to poor model
163 convergence (Alkhamis et al., 2022)
164
165 First, a patristic distance matrix was calculated from the phylogenetic tree using the ape v5.8-1
166 package (Paradis & Schliep, 2019), and hierarchical clustering (Kaufman & Rousseeuw, 1990) was
167 performed on this matrix. The resulting dendrogram was initially partitioned into five distinct
168 clades, aligning with the classification described by Clilverd and colleagues (2023), who identified
169 four co-circulating clades in Batch 1 and a single clade in Batch 3. For each pig with multiple
170 samples, the clade assignment of its virus was tracked chronologically. If a pig's subsequent sample
171 was assigned to the same clade as a previous sample, it was classified as a persistent infection; if it
172 was assigned to a different clade, it was classified as a reinfection. Since hierarchical clustering
173 requires a pre-determined number of clusters, a robustness test was performed by varying the
174 defined clade sizes to ensure that the classification of infection status for individuals remained
175 consistent across different thresholds. Based on this classification, we retained only the first
176 samples of persistent lineages, defined as initial infections, as well as samples from reinfections for
177 the main phylodynamic analyses
178
179 **Phylogenetic Reconstruction and Topology**
180 To assess viral diversity circulating on the farm, phylogenetic trees were constructed for the cohort
181 WGS and ORF5 datasets using IQ-TREE v3 (Wong et al., 2025). The optimal nucleotide
182 substitution model was inferred using the ModelFinder module (Kalyaanamoorthy et al., 2017),
183 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, maximum likelihood trees were generated for both the extended WGS and ORF5 datasets. Following this, recombination was assessed in both datasets, as it violates the bifurcation assumptions of standard phylogenetic models (Driebe et al., 2015), using visual inspection with 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 uninfected ($\delta = \mu+r\psi$), 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

234 contact rates within small subpopulations create a high probability that sampled individuals are the
235 direct ancestors of subsequently sampled pen-mates.

236

237 Long-term Temporal Dynamics (BDSKY)

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

246

247 To ensure robust inference, prior specifications were designed to address the parameter redundancy
248 inherent in birth-death-sampling models, where diverse combinations of transmission, recovery,
249 and sampling rates can produce identical likelihoods (Gavryushkina et al., 2014; Scire et al., 2022).
250 Narrow priors were applied to the rate of becoming uninfected and origin, informed by
251 established PRRSV-1 duration of infectivity (Charpin et al., 2012) and the confirmed outbreak
252 timeline (Clilverd et al., 2023) (Table 1). Priors for the R_e , sampling proportion, clock rate, and
253 epoch transition times were kept uninformative to allow these parameters to be estimated directly
254 from the genomic data rather than prior assumptions. By avoiding overly informative priors, this
255 method reduces over-smoothing and preserves the ability to detect genuine changes in
256 epidemiological patterns, including abrupt shifts in transmission dynamics (Parag et al., 2022). To
257 validate these specifications, we performed a 'sample from prior' analysis (Drummond &
258 Bouckaert, 2015) in BEAST, running the MCMC chain without sequence data to verify that the
259 posterior distributions of inferred parameters were driven by the data rather than artifacts of the
260 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 ~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)
	BDMM	Lognormal(0.45, 0.1)	Parameterized relative to the last Batch 1 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)
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)

261 To infer the optimal nucleotide substitution model directly from the data instead of relying on
 262 predefined universal substitution models, the bModelTest package (Bouckaert & Drummond,
 263 2017) was used. Additionally, an uncorrelated relaxed clock model was evaluated to determine
 264 whether the clock rate varied between branches, providing insight into potential evolutionary rate
 265 differences between the observed clades

266
 267 Distinct epoch strategies were used for the R_e and sampling proportion. The R_e was modeled using
 268 flexible 3-, 4-, and 5-epoch configurations with uniformly estimated transition times over the full
 269 study period (Figure 2). The 3-epoch configuration was applied to test whether the inferred
 270 dynamics could recover the established three-phase timeline (outbreak, dormancy, and resurgence)
 271 defined by Clilverd and colleagues (2023). Additionally, the 4- and 5-epoch configurations were
 272 used to investigate whether a higher temporal resolution could reveal further, previously
 273 unobserved epidemiological phases.

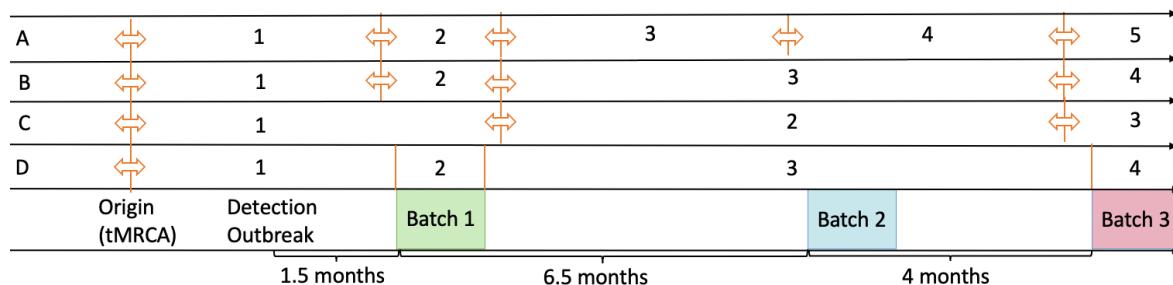


Figure 2: Schematic of temporal discretization strategies for the Birth-Death Skyline (BDSKY) model. The diagram illustrates the flexible production stage configurations (A–C) applied to the effective reproductive number (R_e), which utilized variable 3-, 4-, and 5-epoch intervals to evaluate if the 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.

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

286
 287 **Fine-Scale Spatial Transmission Dynamics (BDMM)**
 288 To quantify Batch 1 dynamics, specifically estimating compartment-specific transmission rates and
 289 infectious periods of farrowing and nursery stage pigs, a structured analysis was conducted using
 290 the BDMM model (Kühnert et al., 2016). This framework extends the temporal flexibility of the
 291 BDSKY model to structured populations by allowing parameters to vary not only across time but
 292 also across compartments. Given the insufficient WGS coverage for Batch 1, this structured

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

296
 297 The first epoch included a single unobserved (“ghost”) deme with a Re representing the sow
 298 reservoir which aimed to capture transmission via unsampled hosts or reservoirs and avoids bias in
 299 structured phylodynamic models (Ewing & Rodrigo, 2006; Müller et al., 2025). The root location
 300 probability was fixed at 100% for this deme, reflecting that the cohort was unborn at viral
 301 introduction to the farm and that all ancestral lineages originated there before migrating into
 302 observed populations.

303
 304 The spatial configuration of the demes was modeled directly on the pen-level topology described
 305 in the "Phylogenetic Reconstruction and Topology" section. However, because the BDMM model
 306 requires a consistent number of demes across all epochs, the overall structure was constrained to
 307 17 demes, comprising 16 pens and one ghost deme. While this configuration excluded two specific
 308 farrowing pens, the impact on transmission inference was considered negligible due to data
 309 sparsity; the farrowing phase yielded only six sequences, all of which originated from just two pens
 310 already included in the model.

311
 312 In this analysis, epoch change times were fixed but defined under two configurations: a production
 313 stage configuration and a migration configuration. The production stage configuration estimated
 314 the Re and rate of becoming uninfected across three biological stages: reservoir, farrowing, and
 315 nursery. The migration configuration, based on metadata from Clilverd et al. (2023), defined
 316 epochs of per-lineage, forward-in-time migration rates, assuming discrete movements between
 317 compartment at each sampling period. The analysis focused on four key one-day transitions: from
 318 sow to farrowing at the start of Epoch 2, from farrowing to nursery at week 4 (Epoch 3), and within
 319 the nursery at weeks 6 and 9. Movement within the farrowing unit at week 2 was excluded due to
 320 insufficient sequence data, as only six sequences were available for week 0 and none for week 2,
 321 precluding reliable estimation of migration rates for this interval.

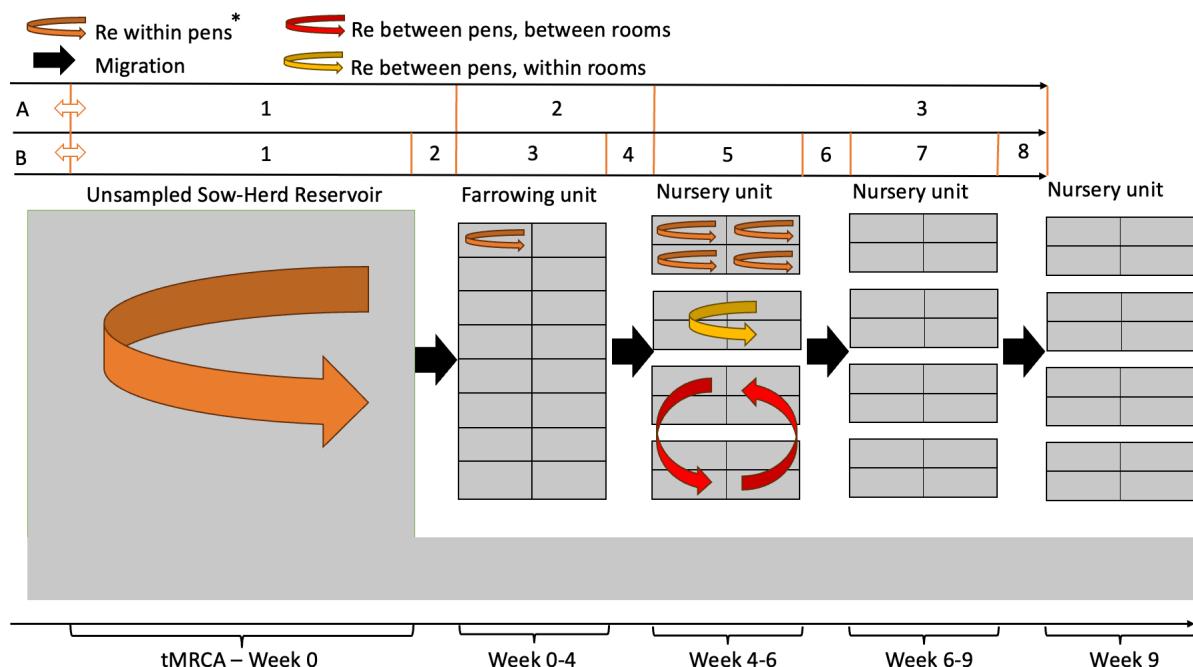


Figure 3: Schematic representation of the Birth-Death with Migration Model (BDMM) configuration. The diagram outlines the spatial structure and migration pathways modeled for Batch 1, including the unsampled sow-herd reservoir, farrowing unit, and nursery unit

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.

322 To characterize transmission dynamics at different spatial levels, a parameter linking strategy was
323 used to estimate three distinct Re values. For the nursery phase, three linking strategies were
324 applied: first, the Re was estimated within pens and linked across all pens; second, the Re between
325 pens within rooms (e.g., pens 1–4, 5–8) were linked; and third, the Re between pens across rooms
326 was linked. For the farrowing phase, only within-pen Re values were estimated and linked because
327 horizontal transmission between pens could be excluded: all six sequences from this phase (five
328 from one pen, one from another) were identified as vertical transmission (Clilverd et al., 2023),
329 with no genetic evidence for horizontal transmission between pens. Furthermore, it was assumed
330 that the farrowing and nursery units were isolated from the external farm environment after initial
331 introduction, so the Re between pens and the ghost deme was set to zero, treating the ghost deme
332 as an ancestral lineage source, not a site of ongoing transmission.
333

334 The molecular clock rate for Batch 1 could not be reliably inferred due to insufficient temporal
335 signal (see results Molecular Clock Validation and Temporal Signal). To address this, the clock
336 rate prior was specified using the posterior estimates from the preceding BDSKY analysis fitted to
337 Batch 1 and Batch 3 (table 1). This approach aligns with recent HEV phylodynamic protocols
338 where informative priors from larger datasets are applied to short-term subsets (Meester et al.,
339 2025), and is supported by evidence that evolutionary rates converge and remain broadly stable
340 across longer epidemic timescales (Bryant et al., 2007).

341 Furthermore, to address the parameter redundancy inherent to birth–death–sampling models, both
342 the sampling proportion prior (centered on the BDSKY posterior estimate for batch 1, with the sow
343 herd phase fixed at 0) and the origin time prior were assigned narrow distributions, ensuring reliable
344 inference of the remaining epidemiological parameters. Finally, the bModelTest package was used
345 using the same configuration as was done in the BDSKY model to investigate whether the inferred
346 substitution model changed by excluding batch 3 (and therefore clade 3) from the analysis.
347

348 A robustness test was performed to challenge the assumption of discrete one-day animal
349 movements by testing a continuous migration model, where migration was permitted continuously
350 throughout the nursery phase. Finally, to check whether the posterior estimates were driven by
351 genetic data rather than prior assumptions, the BDMM model was run with the sample-from-prior
352 setting for both the continuous and discrete migration configurations.
353

354 Data and Code Availability

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

365 **Results**

366 **Study Population and Sequence Validation**

367 The final Cohort datasets were assembled, comprising a Cohort WGS dataset of 35 sequences
368 (partitioned into n=24 for Batch 1 and n=11 for Batch 3) and a Cohort ORF5 dataset of 134
369 sequences (partitioned into n=86 for Batch 1 and n=48 for Batch 3) (Table 2). Furthermore, the
370 Extended datasets, consisted of 87 and 331 sequences for WGS and ORF5, respectively.
371

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

378 *Table 2: Summary of PRRSV-1 sequence datasets, quality metrics, and batch partitioning.*
379 *Overview of the Cohort and Extended datatsets for Whole Genome Sequences (WGS) and ORF5,*
380 *detailling 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
Licensed 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

381 Assessment of genetic redundancy revealed that all sequences in the WGS cohort were unique. In
382 contrast, the ORF5 dataset comprised only 47 unique sequences, corresponding to a redundancy of
383 64.93% and forming 20 groups of identical sequences (see Supplementary File 3 for detailed
384 tables).

385 Additionally, an inconsistency in the sampling dates for the cohort ORF5 sequences was also
386 identified and corrected. The recorded metadata for Batch 1 listed identical sampling dates for week
387 6 and week 9 at 13/07/2017, except for a single week 9 sequence dated 25/07/2017. This was
388 determined to be a likely annotation error. Consequently, all Batch 1, week 9 sequences were
389 corrected to the 25/07/2017 date to accurately reflect the sampling schedule.

390 **Classification of Independent Transmission Events**

391 To differentiate persistent infections from reinfection events, a clade-based classification was
392 performed. The final classification for the WGS dataset identified 24 initial infections, 10 persistent
393 infections, and 1 reinfection. For the ORF5 dataset, the analysis identified 80 initial infections, 53
394 persistent infections, and 1 reinfection. Table 3 summarizes sequence inclusion after the exclusion
395 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

396 and seven for WGS, and between four and seven for the ORF5 dataset (Supplementary file 4).
 397 Additionally, both ORF5 and WGS analyses identified a reinfection event in animal 515. This event
 398 was recorded at week 6 in the ORF5 dataset (where the week 9 sample was unavailable) and at
 399 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

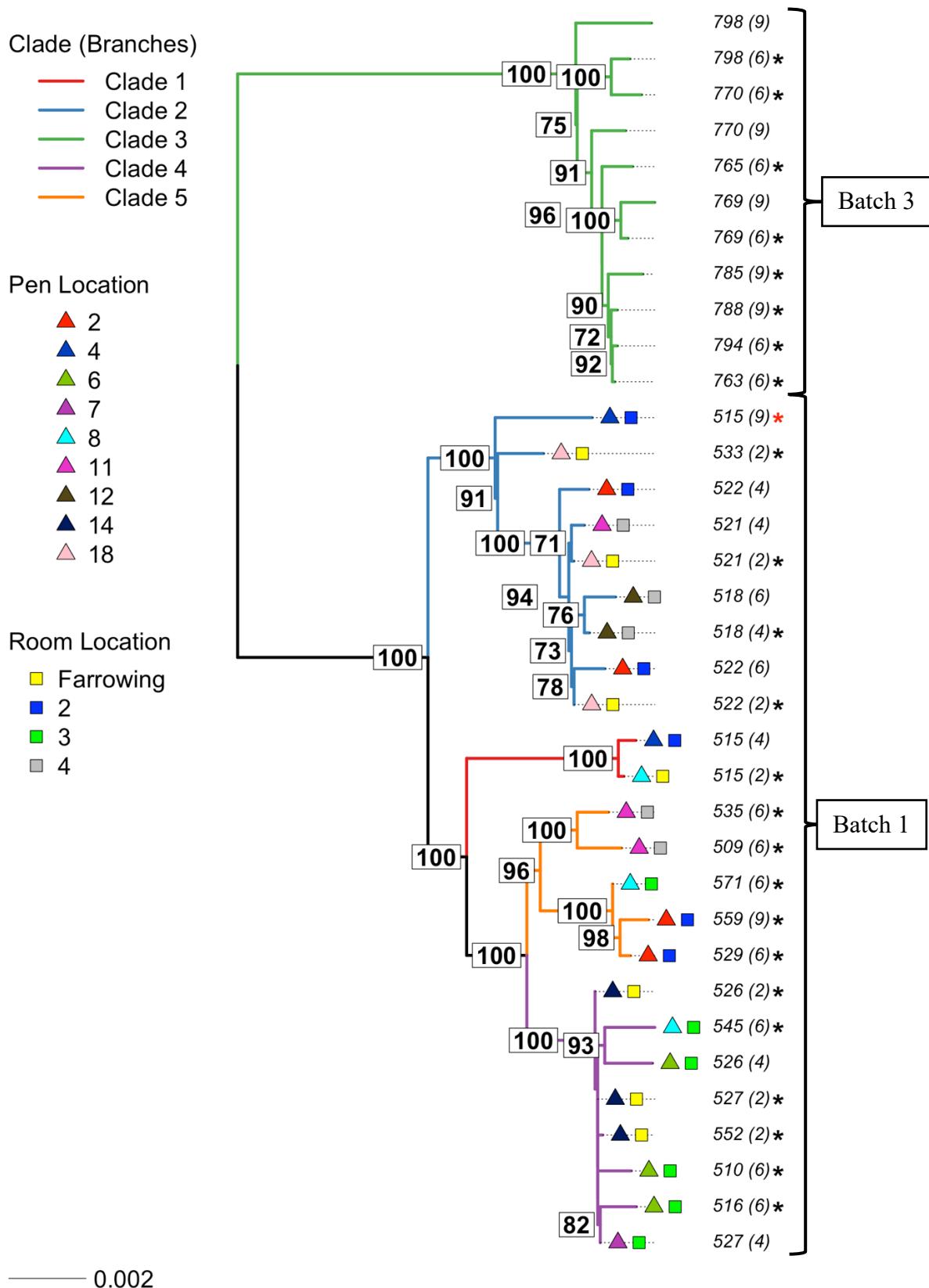
400 After all persistent infections were excluded from the ORF5 datasets, the number of total and
 401 unique sequences dropped from 134 to 81 and 47 to 35 sequences, respectively. The redundancy
 402 percentage improved from 64.93% to 56.59% and 11 identical sequence groups were observed
 403 compared to the 20 identical groups in the ORF5 dataset without exclusion. After removal of
 404 persistent infections from the WGS dataset, the total number of total sequences dropped from 35
 405 to 25 sequences.
 406

407 Phylogenetic Reconstruction and Topology

408 The phylogenetic reconstruction of the consensus WGS and ORF5 datasets yielded topologies
 409 characterized by high statistical confidence, as indicated by the high nodal bootstrap values
 410 observed across the backbone of both rooted trees (Figure 4).
 411

412 The clustering algorithm used for the reinfection analysis identified five monophyletic groups in
 413 the WGS dataset (figure 4A), yielding a cluster composition identical to the clades described by
 414 Clilverd and colleagues (2023). Farrowing unit sequences consistently occupied ancestral nodes
 415 relative to nursery unit sequences, which appeared as descendant taxa within the topology. In
 416 contrast, the ORF5 phylogeny failed to correctly recover this structure. Specifically, the clustering
 417 algorithm could not resolve one clade as a monophyletic group; instead, one lineage was split and
 418 interspersed among other clades (figure 4B). Additionally, the clear ancestral relationship observed
 419 in the genomic data, where farrowing sequences precede nursery sequences, was not evident in the
 420 ORF5 tree.
 421

422 Finally, a strong spatial signal is evident in the terminal branches of both phylogenies. It can be
 423 observed that the distinct sequence groups (clusters of 100% identical sequences) consist largely
 424 of samples originating from the same room within the nursery unit.

A

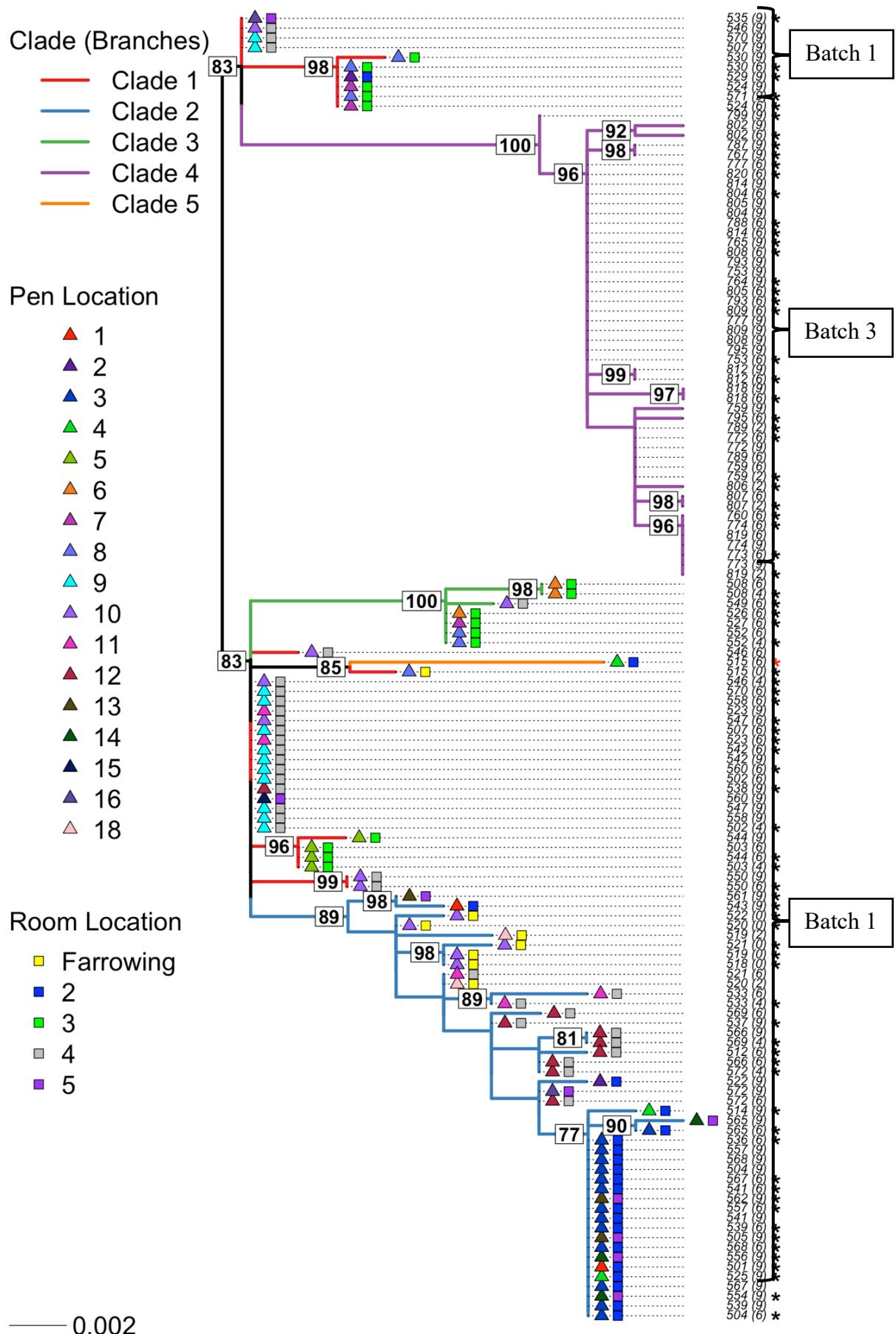
B

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.

425 [Assessment of Recombination and Lateral Introduction](#)

426 Phylogenetic reconstruction of the extended WGS and ORF5 datasets revealed strongly supported
427 monophyletic clades of farm sequences in both datasets (Supplementary File 5), confirming a single
428 introduction and validating that the outbreak was constrained to within-farm dynamics for the
429 duration of the 12-month study period. Subsequent screening for recombination revealed no
430 evidence of reticulation in Neighbor-Net networks (Supplementary file 6) nor significant signals in
431 the Phi test (WGS: $p=1.0$; ORF5: $p=0.84$), validating its suitability for phylodynamic inference.
432

433 [Molecular Clock Validation and Temporal Signal](#)

434 Root-to-tip regression analysis, performed using Clockor2 on the maximum likelihood
435 phylogenetic trees to obtain priors for the final BEAST models, yielded strong positive correlation
436 rates for both the cohort WGS ($R^2 = 0.992$) and ORF5 ($R^2 = 0.836$) datasets. The estimated
437 evolutionary rates were 1.354×10^{-2} and 1.698×10^{-2} substitutions/site/year for WGS and ORF5,
438 respectively. Furthermore, the tMRCA was estimated to be 2017/01/23 for the ORF5 dataset and
439 2017/02/24 for the WGS dataset, placing the origin of the outbreak approximately 8 weeks before
440 clinical detection at 2017/04/08. This confirmed the combined datasets' suitability for temporal
441 analysis.
442

443 However, the partitioned analysis (local clock model for ORF5 data, partitioned by Batch 1 and
444 Batch 3), revealed that Batch 1 exhibited a poor temporal signal on its own ($R^2 = 0.144$) with an
445 evolutionary rate of 2.693×10^{-2} substitutions/site/year, while Batch 3 displayed no temporal signal
446 (negative evolutionary rate). This finding indicated that a reliable clock rate could not be estimated
447 from the Batch 1 and 3 data separately.
448

449 [Long-term Temporal Dynamics \(BDSKY\)](#)

450 To reconstruct the viral spread and estimate evolutionary parameters over the 12-month study
451 period, the BDSKY model was applied to both the cohort ORF5 and WGS datasets. A direct
452 comparison of the inferred epidemiological trajectories showed that both datasets successfully
453 resolved the outbreak into the three distinct phases (initial outbreak, dormancy and resurgence)
454 previously described by Clilverd and colleagues (2023). Under the 3-epoch configuration, the
455 ORF5 and WGS datasets yielded consistent epidemiological patterns, though with slight variations
456 in magnitude (Figure 5 A and B). The ORF5 analysis estimated an initial outbreak phase with a Re
457 of 2.48 (95% Highest Posterior Density (HPD): 1.81–3.15), followed by a significant reduction in
458 transmission during the dormant phase ($Re = 0.20$; 95% HPD: 0.03–0.40), and a final resurgence
459 phase $Re = 3.17$; 95% HPD: 2.02–4.40). The WGS analysis mirrored this trajectory, estimating an
460 initial outbreak Re of 1.80 (95% HPD: 1.20–2.44), a dormant phase Re of 0.35 (95% HPD: 0.05–
461 0.71), and a resurgence Re of 3.85 (95% HPD: 1.62–6.42). Furthermore, the estimated transition
462 times between these phases aligned with the study's metadata for both datasets, placing Batch 1
463 within the outbreak phase, Batch 2 in the dormant phase, and Batch 3 in the resurgence phase.

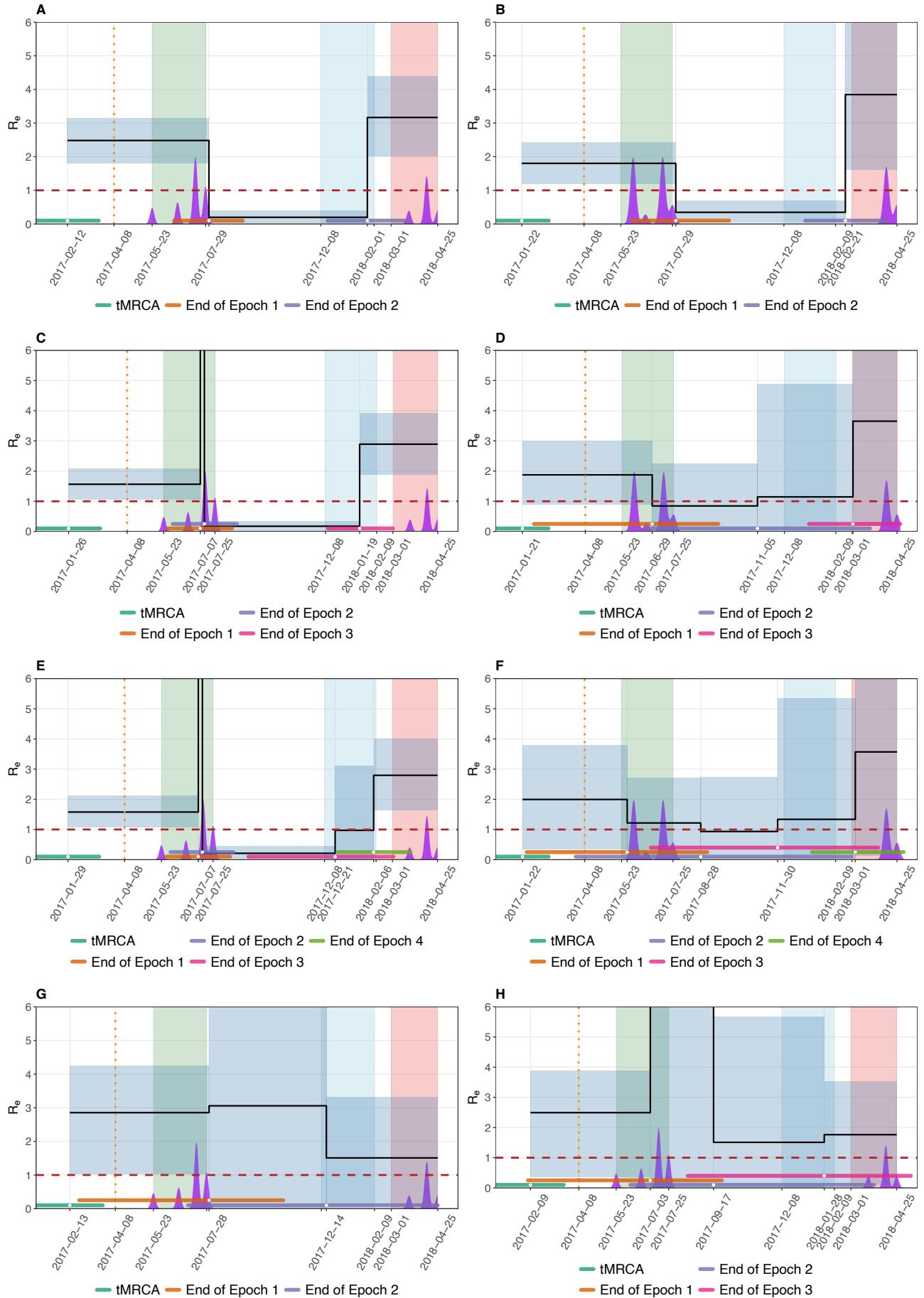


Figure 5: Comparison of the temporal trajectory of the Effective Reproductive Number (R_e) inferred from ORF5 and Whole Genome Sequence (WGS) datasets, alongside prior sensitivity analysis. The upper panels compare the epidemiological reconstruction for the ORF5 dataset (Left Column: A, C, E) and the 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 analysis sampled from the prior for the ORF5 dataset using the 3-epoch (G) and 4-epoch (H) configurations. The R_e is depicted by the solid black line (mean) and the shaded background regions indicate the 95% Highest Posterior Density 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 $R_e = 1$ 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). R_e estimates exceeding 6 are omitted from the visualization for clarity but are available in supplementary file 8.

464 While the 3-epoch model yielded comparable parameter estimates across both datasets, increasing
 465 the temporal resolution to 4- and 5-epoch models resulted in differing model behavior between the
 466 datasets. In the ORF5 analysis, higher-resolution models resolved a short-duration, high-magnitude
 467 transmission peak during the early outbreak period, with estimated R_e values reaching
 468 approximately 20 (95% HPD: ~7–40) (Figure 5 C and E). In contrast, application of 4- and 5-epoch
 469 models to the WGS dataset resulted in increased uncertainty in parameter estimates. For these
 470 models, estimations for epoch transition times and R_e values exhibited overlapping 95% HPD
 471 intervals, and distinct temporal phases were not clearly resolved (Figure 5 D and F). Finally, the
 472 sampling proportion was estimated consistently across models for both Batches 1 and 3. Focusing
 473 on Batch 1, the WGS dataset yielded posterior estimates centered around 0.25 (95% HPD: 0.10–
 474 0.40), compared to approximately 0.30 (95% HPD: 0.15–0.45) for the ORF5 dataset (Figure 6).
 475

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

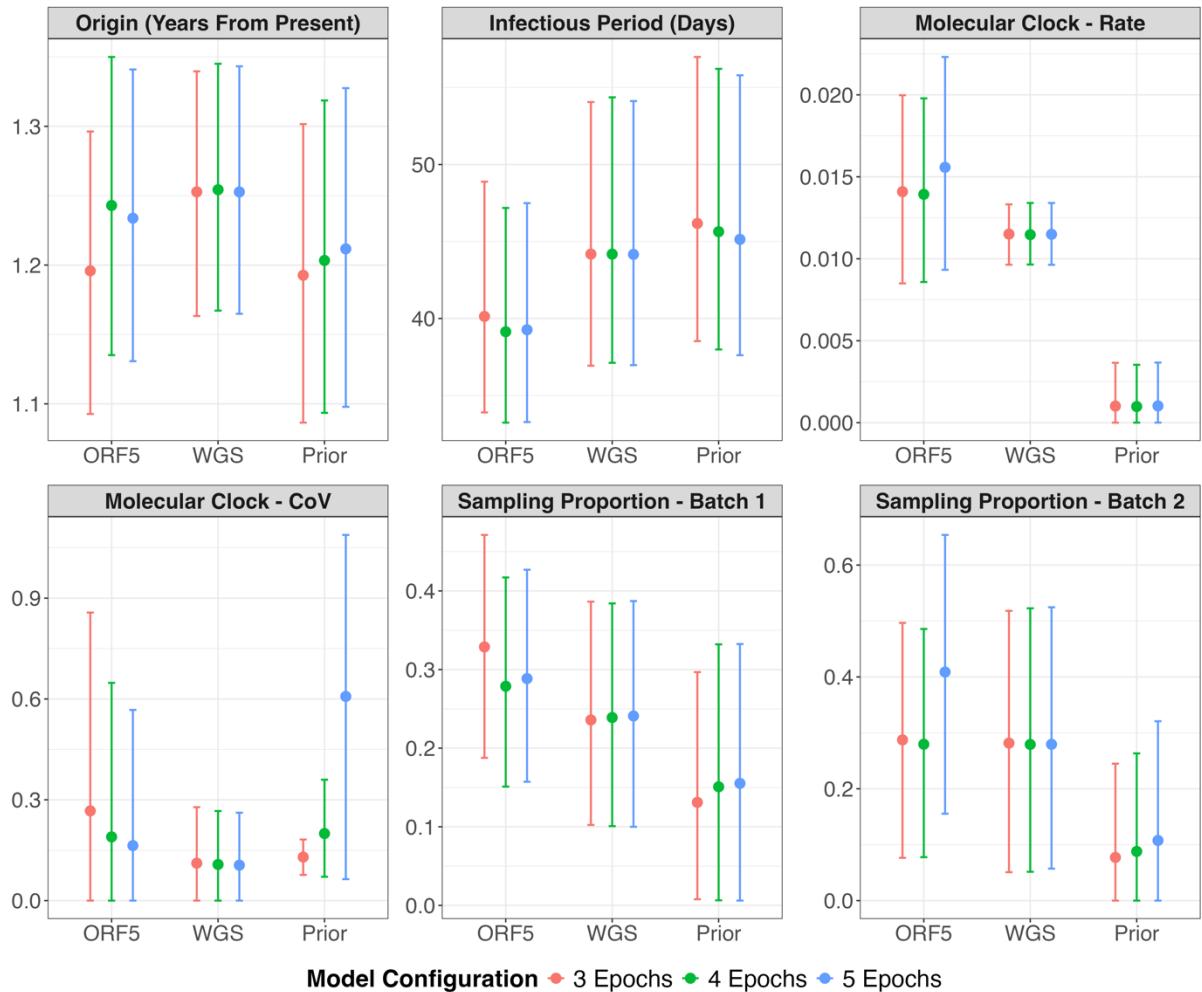


Figure 6: Sensitivity analysis of evolutionary and epidemiological parameter estimates across the Birth-Death Skyline (BDSKY) model configurations. The plots compare the posterior mean estimates and 95% Highest Posterior Density (HPD) intervals for parameters inferred from the ORF5 and WGS datasets, alongside estimates sampled from the Prior (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 uninfected, 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,

485 The comparison between the data-driven and sample-from-prior analysis revealed that the prior-
486 only analyses for both the 3- and 4-epoch configurations could not resolve distinct phases or
487 differences in the Re (Figure 5 G and H). Specifically, regarding the 4-epoch configuration, the
488 prior-only analysis showed no evidence of the explosive outbreak peak observed in the data-driven
489 model, exhibiting very wide 95% HPD intervals for both the Re and epoch transition estimates.
490 Notably, the 4- and 5-epoch sample-from-prior models failed to achieve an ESS of 200 for the
491 estimated epoch transition times, Re values, and the CoV (see Supplementary File 8). Tracer
492 diagnostics indicated poor mixing and an inability to distinguish between multiple local optima;
493 this was considered a stable limitation of the prior landscape that would not likely improve with
494 additional sampling. Regarding evolutionary parameters, significant deviations were observed

495 between the prior and data-driven results, particularly for the molecular clock rate and CoV,
496 confirming that the sequence data contained a strong phylogenetic signal that drove the inference.
497

498 Fine-Scale Spatial Transmission Dynamics (BDMM)

499 To quantify transmission dynamics at the pen and room levels and to determine whether the
500 infectious period differed between the farrowing and nursery phases, the BDMM analysis was
501 performed using the discrete migration configuration.
502

503 For the farrowing unit, the estimated R_e value was 1.67 (95% HPD: 0.24–3.26). This estimate did
504 not deviate substantially from the sample-from-prior analysis (Figure 7), indicating that these
505 results were largely influenced by prior assumptions rather than the genomic data. In contrast,
506 transmission dynamics within the nursery unit revealed a biological signal that diverged from the
507 prior distributions. While within-pen R_e was estimated to be 2.61 (95% HPD: 0.52–5.21), the R_e
508 between pens showed to significantly below the epidemiological threshold of 1. Crucially, the 95%
509 HPD intervals for between-pen transmission were considerably narrower than the analysis run only
510 on prior configurations. Specifically, the R_e within the same room was estimated at 0.46 (95%
511 HPD: 0.09–0.94), while transmission between rooms was even lower at 0.06 (95% HPD: 0.02–
512 0.12). This stands in contrast to the sample-from-prior analysis, where the within-room
513 transmission interval included 1 (mean 0.80; 95% HPD: 0.10–1.73), confirming that the discrete
514 model successfully resolved fine-scale constraints on viral spread that were not present in the prior
515 assumptions.
516

517 Furthermore, the model indicated no significant difference in the infectious period between the
518 farrowing units, with a mean of 44 days (95% HPD: 26–103 days), and nursery units, with a mean
519 of 37 days (95% HPD: 21–99 days). While the sample-from-prior analysis yielded a substantially
520 wider 95% HPD interval for the infectious period in the nursery phase compared to the data-driven
521 analysis, the interval for the farrowing phase remained comparable to the prior distribution.
522 Regarding movement dynamics, the migration rates estimated in the data-driven analysis closely
523 matched those obtained from the corresponding sample-from-prior analysis.
524

525 The posterior estimates for the sampling proportion, tree origin, and mean molecular clock rate
526 closely matched their respective prior distributions, consistent with the narrow priors applied to
527 these parameters. Additionally, consistent with the BDSKY analysis, the K81 model was identified
528 as the optimal nucleotide substitution model.
529

530 Finally, to challenge the assumption of discrete one-day animal movements, a robustness test was
531 performed using a continuous migration model. This analysis yielded highly consistent parameter
532 estimates with the discrete configuration described above, showing no major deviations in
533 epidemiological or evolutionary parameters (Supplementary File 7).

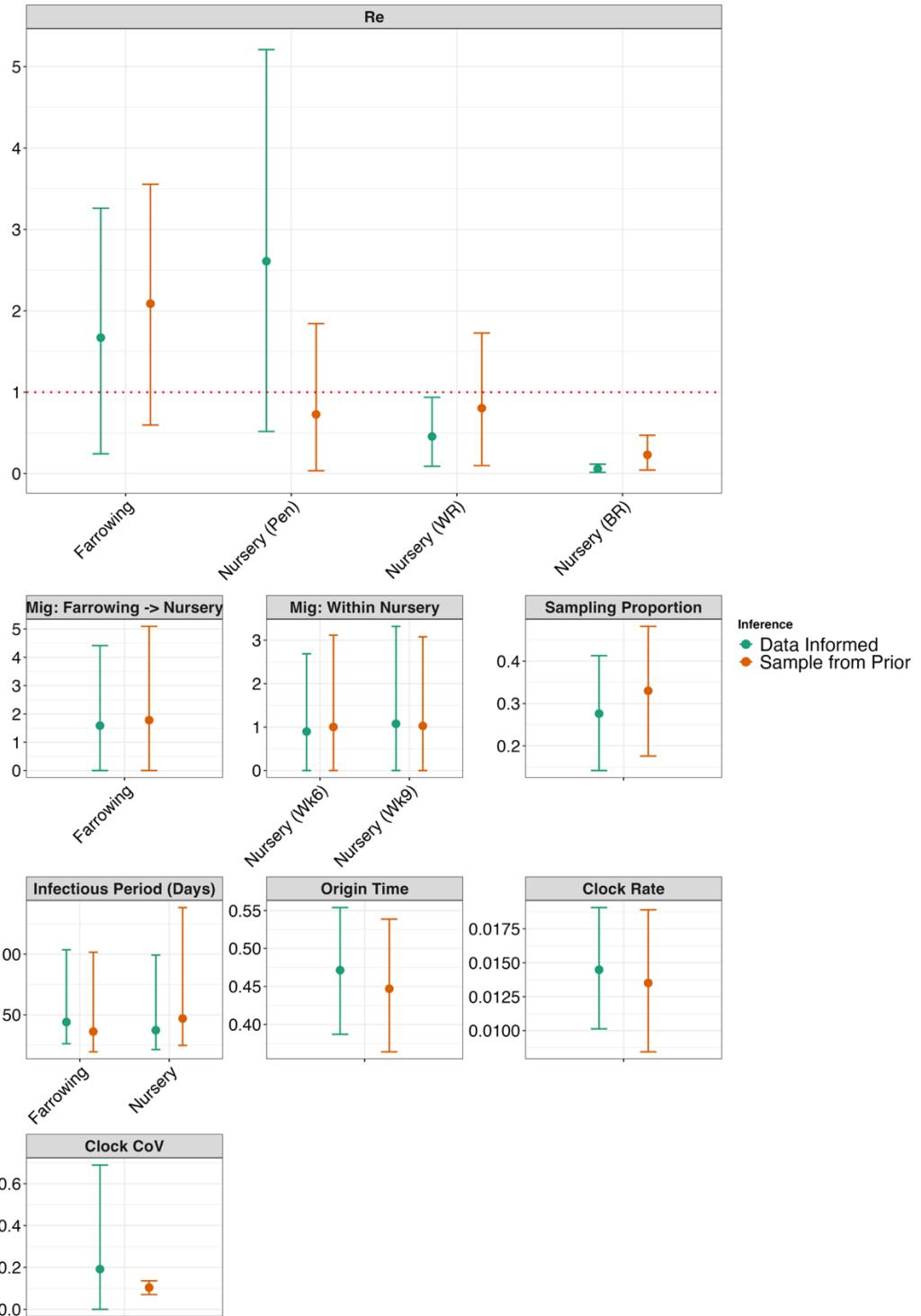


Figure 7: Comparison of posterior parameter estimates for the Birth-Death with Migration (BDMM) model using the discrete migration configuration against the sample-from-prior analysis. The plots display the mean posterior estimates and 95% Highest Posterior Density (HPD) intervals for epidemiological 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 uninfected, as output by BEAST, was converted into the infectious period per animal, expressed in days. Re = Effective reproduction number, CoV = Coefficient of Variation, Mig = Migration rate, WR = Within Room, BR = Between Room, WK = Week.

534 **Discussion**

535 The primary aim of this study was to evaluate the capacity of phylodynamic models to reliably
536 infer previously unquantified epidemiological parameters and distinguish transmission potential
537 across spatially distinct farm units. By leveraging genetic data to reconstruct the transmission
538 dynamics of a PRRSV-1 outbreak, this analysis recovered the established three-phase
539 epidemiological trajectory of outbreak, dormancy, and resurgence described by Clilverd et al.
540 (2023). Crucially, the BDSKY model dated the resurgence event to shortly before Batch 3,
541 supporting the hypothesis proposed by Clilverd and colleagues (2023) that this resurgence was
542 driven by a selection process immediately preceding Batch 3 sampling, as evidenced by the limited
543 genetic diversity of the emerging escape variant.

544

545 At the fine-scale resolution, the BDMM successfully distinguished between within- and between- pen transmission in the nursery unit. While the within-pen R_e were estimated to be 2.61(95% HPD:
546 0.52–5.21), estimates for transmission between pens were significantly below the epidemiological
547 threshold of 1, dropping to 0.46 (95% HPD: 0.09–0.94) for pens within the same room and 0.06
548 (95% HPD: 0.02–0.12) for pens across different rooms. This drop in transmission efficiency is
549 validated by experimental data for related swine viruses. Research on Classical Swine Fever Virus
550 (CSFV) shows that transmission within a pen is nearly 20 times faster than between pens separated
551 by open barriers (Weesendorp et al., 2014). Similarly, the negligible R_e for between-room spread
552 aligns with PRRSV aerosol studies where the virus, though detectable in barn air, often fails to
553 infect pigs in separate compartments despite exposure to contaminated air current (Otake et al.,
554 2002; Trincado et al., 2004). This finding demonstrates that phylodynamic models can successfully
555 quantify the highly structured transmission dynamics characteristic of farm
556 environments. Specifically, the models reveal that while direct contact drives efficient transmission
557 among pen-mates, horizontal spread between spatially separated units remains a rare event.

558

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

565

566 Notably, the estimated infectious period did not differ between the farrowing and nursery phases.
567 While this may reflect a genuine similarity between phases, the ability to detect a difference was
568 likely constrained due to sampling sparsity: in the reconstructed tree, farrowing-phase sequences
569 did not appear as ancestors of nursery-derived lineages, indicating that nursery infections likely
570 arose from unsampled farrowing or nursery stage pigs or the sow reservoir. Consequently, the
571 model could not interpret farrowing infections as long-term reservoirs bridging the nursery stage.
572 This highlights a significant limitation of the model's "isolated" assumption, which treated the
573 farrowing and nursery units as closed systems after the initial introduction from the sow herd.
574 Moreover, the large 95% HPD intervals associated with the infectious period estimates in both
575 phases obscured the detection of a statistically meaningful difference. This uncertainty can be
576 explained by a parameter non-identifiability issue, in which the model cannot distinguish between
577 "fast turnover" (high transmission with rapid recovery) and "slow turnover" (low transmission with
578 prolonged infectiousness) scenarios because both yield indistinguishable likelihoods (Zarebski et
579 al., 2022). The impact of this limitation is apparent in the nursery unit results, where both within-
580 pen R_e and the infectious period were associated with wide 95% HPD intervals. A commonly
581 adopted approach to alleviate such identifiability issues is to fix the rate of becoming uninfected
582 using a well-supported prior (Weber et al., 2023).

583

584

585 A critical outcome of this study is the evaluation of genomic resolution on model performance,
586 revealing complementary strengths between targeted gene sequencing and whole-genome
587 approaches. The larger ORF5 dataset produced more confident epidemiological estimates for
588 the Re and transition times than the sparse WGS dataset, supporting the argument that high taxon
589 density is required to distinguish epidemiological rate shifts from background noise in BDSKY
590 models (Culshaw et al., 2019). Notably, the higher resolution of the ORF5 dataset resolved a brief,
591 intense transmission peak during the early outbreak that the WGS dataset failed to capture. We
592 hypothesize that this peak reflects the superspreading event reported by Clilverd and colleagues
593 (2023), in which a single pig accounted for 28.6% of all traced horizontal transmission
594 events. Conversely, the WGS dataset provided superior resolution for evolutionary parameters,
595 yielding narrower 95% HPD intervals for the clock rate and CoV, and a stronger temporal signal
596 ($R^2=0.992$) compared to ORF5 ($R^2=0.836$). This aligns with the 'temporal horizon' concept (Dudas
597 & Bedford, 2019), where longer sequences provide the mutational density necessary for precise
598 clock calibration. This increased information density likely also drove the inference of a more
599 complex substitution model for the WGS dataset (Model 24) compared to ORF5 (K81), suggesting
600 that the exceptional genetic diversity of PRRSV requires data-driven models rather than reliance
601 on default assumptions. Consequently, researchers should select genomic resolution based on
602 specific objectives: high-density, targeted sequencing is recommended for monitoring fine-scale
603 epidemiological dynamics and detecting superspreading events. Conversely, longer sequences
604 should be prioritized for evolutionary precision, as increased mutational density enables accurate
605 molecular clock calibration. While not the primary focus of this study, such precision is essential
606 for analyses aiming to determine the origin of an outbreak by estimating the tMRCA.
607

608 A key strength of the BDSKY analysis lies in the strict parameterization of the sampling proportion
609 to mirror the sampling schedule. This fixed configuration was necessary to distinguish sampling
610 gaps from low viral prevalence and prevent poor statistical identifiability (Featherstone et al., 2021;
611 Parag et al., 2022). By preventing the model from inherently interpreting the lack of samples during
612 the dormant phase as a halt of transmission, this strategy ensured that parameter estimates remained
613 robust and driven by genomic signals rather than prior configurations. This was further validated
614 by sample-from-prior analyses, which confirmed that the prior configuration of the model alone
615 could not resolve the epidemiological phases. However, the BDSKY model is limited by its
616 assumption of an unstructured population which is violated according to our BDMM analysis that
617 confirmed that transmission was dependent on spatial units. Consequently, ignoring this structure
618 can lead to biased transmission rate estimates, as the model incorrectly assumes equal contact
619 probabilities across the entire herd (Seidel et al., 2024). Specifically, the model may misinterpret
620 localized viral persistence within a single pen as broader, herd-wide transmission, potentially
621 inflating the Re estimate (Seidel et al., 2024; Volz & Frost, 2014). Accordingly, it should be noted
622 that the Re values reported in the BDSKY analysis may represent overestimates of the true
623 transmission intensity.
624

625 Furthermore, the study faced several further strengths and limitations. The clade-based analysis
626 effectively differentiated persistent infections from reinfections, supported by the recovery of the
627 single reinfection event (pig 515) reported by Clilverd et al. (2023) and robustness checks
628 confirming that infection classifications remained consistent across varying numbers of estimated
629 clades. However, the ORF5 dataset failed to maintain monophyly for one lineage, probably due to
630 a high proportion of identical sequences. This aligns with the finding that high sequence identity
631 increases "monophyly-violating interclass coalescence," increasing the likelihood that lineages mix
632 before coalescing (Mehta et al., 2016). To overcome the limitation of low unique genetic variation,
633 future studies could leverage intra-host diversity as demonstrated by Ortiz et al. (2023). By tracking
634 shared minor variants that are typically discarded in consensus sequencing, this method resolves
635 transmission links between hosts with identical consensus sequences, effectively distinguishing
636 specific transmission pathways within otherwise unresolved clusters. Furthermore, the finding that

637 transmission is predominantly confined to within-pen dynamics introduces a challenge regarding
638 the "burnout" phenomenon (Parsons et al., 2024). In these small groups, the virus quickly exhausts
639 the limited pool of susceptible pigs, causing outbreaks to terminate abruptly. This premature
640 truncation of transmission chains can bias R_0 estimates, as standard models assume ongoing spread
641 in a larger susceptible population rather than extinction due to host depletion.

642
643 To overcome the limitations associated with data sparsity and structure, future study designs
644 targeting within-farm transmission should be optimized for phylodynamic inference. We
645 recommend a shift from fixed longitudinal sampling of the same individuals to a rotational
646 strategy. By replacing repeated sampling of already infected individuals at subsequent time points
647 with sampling of previously unsampled pen-mates or animals that have not yet tested positive,
648 researchers can avoid redundancy arising from persistent infections and instead maximize the
649 recovery of distinct transmission events. Furthermore, sampling must extend beyond the farrowing
650 and nursery units to include the sow herd to replace the "ghost deme" assumption with empirical
651 data, enabling the direct quantification of vertical transmission and spillover events. Finally, future
652 phylodynamic studies that aim to characterize transmission between isolated compartments should
653 incorporate metadata on personnel movement and ventilation to better distinguish transmission
654 pathways. Arruda and colleagues (2019) highlight that distinguishing between aerosol transmission
655 and local spread via fomites is often confounded by a lack of data on internal biosecurity measures
656 and contact networks. Integrating these factors is methodologically feasible in phylodynamic
657 models, as Meester et al. (2025) recently used BDMM to successfully distinguish environmental
658 from direct transmission routes. Therefore, detailed metadata would allow models to explicitly
659 parameterize the physical and procedural barriers defining subpopulation interactions.
660

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

672 **Acknowledgements**

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

679

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

686

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

690 References

- 691 Alkhamis, M. A., Fountain-Jones, N. M., Khajah, M. M., Alghounaim, M., & Al-Sabah, S. K.
692 (2022). Comparative phylodynamics reveals the evolutionary history of SARS-CoV-2
693 emerging variants in the Arabian Peninsula. *Virus Evolution*, 8(1), veac040.
- 694 Alkhamis, M. A., Perez, A. M., Murtaugh, M. P., Wang, X., & Morrison, R. B. (2016).
695 Applications of Bayesian phylodynamic methods in a recent US porcine reproductive
696 and respiratory syndrome virus outbreak. *Frontiers in Microbiology*, 7, 67.
- 697 Andrade, J., & Duggan, J. (2022). Inferring the effective reproductive number from
698 deterministic and semi-deterministic compartmental models using incidence and
699 mobility data. *PLoS Computational Biology*, 18(6), e1010206.
- 700 Arruda, A. G., Tousignant, S., Sanhueza, J., Vilalta, C., Poljak, Z., Torremorell, M., Alonso,
701 C., & Corzo, C. A. (2019). Aerosol detection and transmission of porcine reproductive
702 and respiratory syndrome virus (PRRSV): What is the evidence, and what are the
703 knowledge gaps? *Viruses*, 11(8), 712.
- 704 Balka, G., Podgórska, K., Brar, M. S., Bálint, Á., Cedar, D., Celer, V., Dénes, L., Dirbakova,
705 Z., Jedryczko, A., Márton, L., & others. (2018). Genetic diversity of PRRSV 1 in
706 Central Eastern Europe in 1994–2014: Origin and evolution of the virus in the region.
707 *Scientific Reports*, 8(1), 7811.
- 708 Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model
709 averaging and model comparison. *BMC Evolutionary Biology*, 17(1), 42.
- 710 Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina,
711 A., Heled, J., Jones, G., Kühnert, D., De Maio, N., & others. (2019). BEAST 2.5: An
712 advanced software platform for Bayesian evolutionary analysis. *PLoS Computational
713 Biology*, 15(4), e1006650.
- 714 Bruen, T. C., Philippe, H., & Bryant, D. (2006). A simple and robust statistical test for
715 detecting the presence of recombination. *Genetics*, 172(4), 2665–2681.
- 716 Bryant, J. E., Holmes, E. C., & Barrett, A. D. T. (2007). Out of Africa: A molecular
717 perspective on the introduction of yellow fever virus into the Americas. *PLoS
718 Pathogens*, 3(5), e75.
- 719 Butler, J., Lager, K., Golde, W., Faaberg, K. S., Sinkora, M., Loving, C., & Zhang, Y. (2014).
720 Porcine reproductive and respiratory syndrome (PRRS): An immune dysregulatory
721 pandemic. *Immunologic Research*, 59(1), 81–108.
- 722 Caserta, L. C., Zhang, J., Piñeyro, P., & Diel, D. G. (2023). Rapid genotyping of porcine
723 reproductive and respiratory syndrome virus (PRRSV) using MinION nanopore
724 sequencing. *PLoS One*, 18(5), e0282767.
- 725 Chandra, S., Cezar, G., Rupasinghe, K., Magalhães, E., Silva, G. S., Almeida, M., Crim, B.,
726 Burrough, E., Gauger, P., Madson, D., & others. (2025). Harnessing sequencing data
727 for porcine reproductive and respiratory syndrome virus (PRRSV): Tracking genetic
728 evolution dynamics and emerging sequences in US swine industry. *Frontiers in
729 Veterinary Science*, 12, 1571020.
- 730 Charpin, C., Mahé, S., Keranflec'h, A., Belloc, C., Cariolet, R., Le Potier, M.-F., & Rose, N.
731 (2012). Infectiousness of pigs infected by the Porcine Reproductive and Respiratory
732 Syndrome virus (PRRSV) is time-dependent. *Veterinary Research*, 43(1), 69.
- 733 Clilverd, H., Martín-Valls, G., Li, Y., Martín, M., Cortey, M., & Mateu, E. (2023). Infection
734 dynamics, transmission, and evolution after an outbreak of porcine reproductive and
735 respiratory syndrome virus. *Frontiers in Microbiology*, 14, 1109881.
- 736 Collins, J. E., Benfield, D. A., Christianson, W. T., Harris, L., Hennings, J. C., Shaw, D. P.,
737 Goyal, S. M., McCullough, S., Morrison, R. B., Joo, H. S., & others. (1992). Isolation
738 of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North

- 739 America and experimental reproduction of the disease in gnotobiotic pigs. *Journal of*
740 *Veterinary Diagnostic Investigation*, 4(2), 117–126.
- 741 Culshaw, V., Stadler, T., & Sanmartín, I. (2019). Exploring the power of Bayesian birth-death
742 skyline models to detect mass extinction events from phylogenies with only extant
743 taxa. *Evolution*, 73(6), 1133–1150.
- 744 Dearlove, B. L., Xiang, F., & Frost, S. D. (2017). Biased phylodynamic inferences from
745 analysing clusters of viral sequences. *Virus Evolution*, 3(2), vex020.
- 746 Driebe, E. M., Sahl, J. W., Roe, C., Bowers, J. R., Schupp, J. M., Gillece, J. D., Kelley, E.,
747 Price, L. B., Pearson, T. R., Hepp, C. M., & others. (2015). Using whole genome
748 analysis to examine recombination across diverse sequence types of *Staphylococcus*
749 *aureus*. *PLoS One*, 10(7), e0130955.
- 750 Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian evolutionary analysis with BEAST*.
751 Cambridge University Press.
- 752 Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., &
753 Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2.
754 *Virus Evolution*, 6(2), veaa061.
- 755 Dudas, G., & Bedford, T. (2019). The ability of single genes vs full genomes to resolve time
756 and space in outbreak analysis. *BMC Evolutionary Biology*, 19(1), 232.
- 757 Ecker, N., Huchon, D., Mansour, Y., Mayrose, I., & Pupko, T. (2024). A machine-learning-
758 based alternative to phylogenetic bootstrap. *Bioinformatics*, 40(Supplement_1), i208–
759 i217.
- 760 Ewing, G., & Rodrigo, A. (2006). Estimating population parameters using the structured
761 serial coalescent with Bayesian MCMC inference when some demes are hidden.
762 *Evolutionary Bioinformatics*, 2, 117693430600200026.
- 763 Featherstone, L. A., Di Giallonardo, F., Holmes, E. C., Vaughan, T. G., & Duchêne, S. (2021).
764 Infectious disease phylodynamics with occurrence data. *Methods in Ecology and*
765 *Evolution*, 12(8), 1498–1507.
- 766 Featherstone, L. A., Duchene, S., & Vaughan, T. G. (2023). Decoding the fundamental drivers
767 of phylodynamic inference. *Molecular Biology and Evolution*, 40(6), msad132.
- 768 Featherstone, L. A., Rambaut, A., Duchene, S., & Wirth, W. (2024). Clockor2: Inferring
769 global and local strict molecular clocks using root-to-tip regression. *Systematic*
770 *Biology*, 73(3), 623–628.
- 771 Franzo, G., Barbierato, G., Pesente, P., Legnardi, M., Tucciarone, C. M., Sandri, G., & Drigo,
772 M. (2021). Porcine reproductive and respiratory syndrome (Prrs) epidemiology in an
773 integrated pig company of northern Italy: A multilevel threat requiring multilevel
774 interventions. *Viruses*, 13(12), 2510.
- 775 Franzo, G., Faustini, G., Legnardi, M., Cecchinato, M., Drigo, M., & Tucciarone, C. M.
776 (2022). Phylodynamic and phylogeographic reconstruction of porcine reproductive
777 and respiratory syndrome virus (PRRSV) in Europe: Patterns and determinants.
778 *Transboundary and Emerging Diseases*, 69(5), e2175–e2184.
- 779 Frias-De-Diego, A., Jara, M., Pecoraro, B. M., & Crisci, E. (2021). Whole genome or single
780 genes? A phylodynamic and bibliometric analysis of PRRSV. *Frontiers in Veterinary*
781 *Science*, 8, 658512.
- 782 Gavryushkina, A., Welch, D., Stadler, T., & Drummond, A. J. (2014). Bayesian inference of
783 sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational*
784 *Biology*, 10(12), e1003919.
- 785 Guinat, C., Tang, H., Yang, Q., Valenzuela Agüí, C., Vaughan, T. G., Scire, J., Yu, H., Wang,
786 W., Chen, Z., Ducatez, M. F., & others. (2023). Bayesian phylodynamics reveals the
787 transmission dynamics of avian influenza A (H7N9) virus at the human–live bird

- 788 market interface in China. *Proceedings of the National Academy of Sciences*, 120(17),
789 e2215610120.
- 790 Guinat, C., Vergne, T., Kocher, A., Chakraborty, D., Paul, M. C., Ducatez, M., & Stadler, T.
791 (2021). What can phylodynamics bring to animal health research? *Trends in Ecology
792 & Evolution*, 36(9), 837–847.
- 793 Huson, D. H., & Bryant, D. (2024). The SplitsTree App: Interactive analysis and visualization
794 using phylogenetic trees and networks. *Nature Methods*, 21(10), 1773–1774.
- 795 Johnson, A. D. (2010). An extended IUPAC nomenclature code for polymorphic nucleic
796 acids. *Bioinformatics*, 26(10), 1386–1389.
- 797 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017).
798 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature
799 Methods*, 14(6), 587–589.
- 800 Kappes, M. A., & Faaberg, K. S. (2015). PRRSV structure, replication and recombination:
801 Origin of phenotype and genotype diversity. *Virology*, 479, 475–486.
- 802 Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster
803 analysis*. John Wiley & Sons.
- 804 Kühnert, D., Stadler, T., Vaughan, T. G., & Drummond, A. J. (2016). Phylodynamics with
805 migration: A computational framework to quantify population structure from genomic
806 data. *Molecular Biology and Evolution*, 33(8), 2102–2116.
- 807 Li, C., Xu, H., Zhao, J., Gong, B., Sun, Q., Xiang, L., Li, W., Guo, Z., Li, J., Tang, Y., &
808 others. (2022). Epidemiological investigation and genetic evolutionary analysis of
809 PRRSV-1 on a pig farm in China. *Frontiers in Microbiology*, 13, 1067173.
- 810 Lozano-Fernandez, J. (2022). A practical guide to design and assess a phylogenomic study.
811 *Genome Biology and Evolution*, 14(9), evac129.
- 812 Meester, M., Valenzuela Agüí, C., Tobias, T. J., Honing, R. W. H. van der, Guinat, C.,
813 Bouwknegt, M., du Plessis, L., Fischer, E. A., Spaninks, M., Stadler, T., & others.
814 (2025). Zoonotic hepatitis E virus spreads through environmental routes in pig herds—
815 A phylodynamic analysis. *PLoS Pathogens*, 21(11), e1013710.
- 816 Mehta, R. S., Bryant, D., & Rosenberg, N. A. (2016). The probability of monophyly of a
817 sample of gene lineages on a species tree. *Proceedings of the National Academy of
818 Sciences*, 113(29), 8002–8009.
- 819 Müller, N. F., Bouckaert, R. R., Wu, C.-H., & Bedford, T. (2025). MASCOT-Skyline
820 integrates population and migration dynamics to enhance phylogeographic
821 reconstructions. *PLOS Computational Biology*, 21(9), e1013421.
- 822 Murtaugh, M. P., Stadejek, T., Abrahante, J. E., Lam, T. T., & Leung, F. C.-C. (2010). The
823 ever-expanding diversity of porcine reproductive and respiratory syndrome virus.
824 *Virus Research*, 154(1–2), 18–30.
- 825 Ortiz, A. T., Kendall, M., Storey, N., Hatcher, J., Dunn, H., Roy, S., Williams, R., Williams,
826 C., Goldstein, R. A., Didelot, X., & others. (2023). Within-host diversity improves
827 phylogenetic and transmission reconstruction of SARS-CoV-2 outbreaks. *Elife*, 12,
828 e84384.
- 829 Otake, S., Dee, S., Jacobson, L., Pijoan, C., & Torremorell, M. (2002). Evaluation of aerosol
830 transmission of porcine reproductive and respiratory syndrome virus under controlled
831 field conditions. *Veterinary Record*, 150(26), 804–808.
- 832 Pamornchainavakul, N., Makau, D. N., Paploski, I. A., Corzo, C. A., & VanderWaal, K.
833 (2023). Unveiling invisible farm-to-farm PRRSV-2 transmission links and routes
834 through transmission tree and network analysis. *Evolutionary Applications*, 16(10),
835 1721–1734.

- 836 Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and
837 evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
838 <https://doi.org/10.1093/bioinformatics/bty633>
- 839 Parag, K. V., Pybus, O. G., & Wu, C.-H. (2022). Are skyline plot-based demographic
840 estimates overly dependent on smoothing prior assumptions? *Systematic Biology*,
841 71(1), 121–138.
- 842 Parisio, G., Franzo, G., Barbieri, I., Carta, V., Stadejek, T., Manenti, S., Campagna, D.,
843 Faccini, S., Vignola, G., Alborali, G. L., & others. (2024). Evolutionary dynamics of
844 PRRS virus in Italian Pig farms: A retrospective study. *Virology Journal*, 21(1), 326.
- 845 Parsons, T. L., Bolker, B. M., Dushoff, J., & Earn, D. J. (2024). The probability of epidemic
846 burnout in the stochastic SIR model with vital dynamics. *Proceedings of the National
847 Academy of Sciences*, 121(5), e2313708120.
- 848 Pileri, E., & Mateu, E. (2016). Review on the transmission porcine reproductive and
849 respiratory syndrome virus between pigs and farms and impact on vaccination.
850 *Veterinary Research*, 47(1), 108.
- 851 Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and
852 Output Analysis for MCMC. *R News*, 6(1), 7–11.
- 853 Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior
854 summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5),
855 901–904.
- 856 Revell, L. J. (2024). phytools 2.0: An updated R ecosystem for phylogenetic comparative
857 methods (and other things). *PeerJ*, 12, e16505. <https://doi.org/10.7717/peerj.16505>
- 858 Rowland, R. R., Lawson, S., Rossow, K., & Benfield, D. A. (2003). Lymphoid tissue tropism
859 of porcine reproductive and respiratory syndrome virus replication during persistent
860 infection of pigs originally exposed to virus in utero. *Veterinary Microbiology*, 96(3),
861 219–235.
- 862 Scire, J., Barido-Sottani, J., Kühnert, D., Vaughan, T. G., & Stadler, T. (2022). Robust
863 phylodynamic analysis of genetic sequencing data from structured populations.
864 *Viruses*, 14(8), 1648.
- 865 Seidel, S., Stadler, T., & Vaughan, T. G. (2024). Estimating pathogen spread using structured
866 coalescent and birth–death models: A quantitative comparison. *Epidemics*, 49,
867 100795.
- 868 Sequeira, S. C., Sebunia, N., Page, J. R., Lasisi, T., Habing, G., & Arruda, A. G. (2025). A
869 systematic scoping review and thematic analysis: How can livestock and poultry
870 movement networks inform disease surveillance and control at the global scale? *Plos
871 One*, 20(7), e0328518.
- 872 Sha, H., Lan, X., Yang, Z., Lv, C., Zhang, H., Luo, Q., Zheng, Y., Li, G., Kong, W., Huang,
873 L., & others. (2025). Genetic variation in NSP4 of type 1 porcine reproductive and
874 respiratory syndrome virus in China. *BMC Veterinary Research*, 21(1), 374.
- 875 Stadler, T., Kühnert, D., Bonhoeffer, S., & Drummond, A. J. (2013). Birth–death skyline plot
876 reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV).
877 *Proceedings of the National Academy of Sciences*, 110(1), 228–233.
- 878 Stadler, T., Magnus, C., Vaughan, T. G., Barido-Sottani, J., Bošková, V., Huisman, J., &
879 Pečerska, J. (2024). *Decoding genomes: From sequences to phylodynamics*. ETH
880 Zurich.
- 881 Trincado, C., Dee, S., Otake, S., Pijoan, C., Jacobson, L., & Rossow, K. (2004). Attempts to
882 transmit porcine reproductive and respiratory syndrome virus by aerosols under
883 controlled field conditions. *Veterinary Record*, 154(10), 294–297.

- 884 Vaughan, T. G., & Stadler, T. (2025). Bayesian Phylodynamic Inference of Multitype
885 Population Trajectories Using Genomic Data. *Molecular Biology and Evolution*,
886 42(6), msaf130.
- 887 Volz, E. M., & Frost, S. D. (2014). Sampling through time and phylodynamic inference with
888 coalescent and birth–death models. *Journal of The Royal Society Interface*, 11(101),
889 20140945.
- 890 Weber, A., Översti, S., & Kühnert, D. (2023). Reconstructing relative transmission rates in
891 Bayesian phylodynamics: Two-fold transmission advantage of Omicron in Berlin,
892 Germany during December 2021. *Virus Evolution*, 9(2), vead070.
- 893 Weesendorp, E., Backer, J., & Loeffen, W. (2014). Quantification of different classical swine
894 fever virus transmission routes within a single compartment. *Veterinary Microbiology*,
895 174(3–4), 353–361.
- 896 Weng, C., Huang, X., Chen, Z., He, M., Zhang, B., Li, H., Xie, J., Chen, M., Qiu, L., Li, X.,
897 & others. (2025). Genetic evolution, epidemic trends, and recombination dynamics of
898 PRRSV-1 in China. *Frontiers in Veterinary Science*, 12, 1632917.
- 899 Wensvoort, G., Terpstra, C., Pol, J., Ter Laak, E., Bloemraad, M., De Kluyver, E., Kragten,
900 C., Van Buiten, L. den, Den Besten, A., Wagenaar, F., & others. (1991). Mystery
901 swine disease in The Netherlands: The isolation of Lelystad virus. *Veterinary
902 Quarterly*, 13(3), 121–130.
- 903 Wills, R. W., Doster, A. R., Galeota, J. A., Sur, J.-H., & Osorio, F. A. (2003). Duration of
904 infection and proportion of pigs persistently infected with porcine reproductive and
905 respiratory syndrome virus. *Journal of Clinical Microbiology*, 41(1), 58–62.
- 906 Wong, T. K., Ly-Trong, N., Ren, H., Baños, H., Roger, A. J., Susko, E., Bielow, C., De Maio,
907 N., Goldman, N., Hahn, M. W., & others. (2025). *IQ-TREE 3: Phylogenomic
908 Inference Software using Complex Evolutionary Models*.
909 <https://doi.org/10.32942/X2P62N>
- 910 Wright, E. S. (2016). Using DECIPIER v2.0 to Analyze Big Biological Sequence Data in R.
911 *The R Journal*, 8(1), 352–359.
- 912 Wu, Z., Chang, T., Wang, D., Zhang, H., Liu, H., Huang, X., Tian, Z., Tian, X., Liu, D., An,
913 T., & others. (2024). Genomic surveillance and evolutionary dynamics of type 2
914 porcine reproductive and respiratory syndrome virus in China spanning the African
915 swine fever outbreak. *Virus Evolution*, 10(1), veae016.
- 916 You, X., Lei, Y., Zhang, P., Xu, D., Ahmed, Z., & Yang, Y. (2022). Role of transcription
917 factors in porcine reproductive and respiratory syndrome virus infection: A review.
918 *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.924004>
- 919 Zarebski, A. E., du Plessis, L., Parag, K. V., & Pybus, O. G. (2022). A computationally
920 tractable birth-death model that combines phylogenetic and epidemiological data.
921 *PLOS Computational Biology*, 18(2), e1009805.
- 922 Zhang, J., Zheng, Y., Xia, X.-Q., Chen, Q., Bade, S. A., Yoon, K.-J., Harmon, K. M., Gauger,
923 P. C., Main, R. G., & Li, G. (2017). High-throughput whole genome sequencing of
924 porcine reproductive and respiratory syndrome virus from cell culture materials and
925 clinical specimens using next-generation sequencing technology. *Journal of
926 Veterinary Diagnostic Investigation*, 29(1), 41–50.
- 927 Zhou, L., Han, J., & Yang, H. (2024). The evolution and diversity of porcine reproductive and
928 respiratory syndrome virus in China. *Veterinary Microbiology*, 298, 110252.