Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A systematic review of whole-genome sequencing studies of Mycobacterium bovis that determine the direction of transmission between badgers and cattle

Paul Torgerson

paul.torgerson@uzh.ch

University of Zurich https://orcid.org/0000-0003-4277-9983

Giovanni Ghielmetti

University of Zurich

Thomas Langton

Herpetofauna Consultants

Systematic Review

Keywords: Bovine tuberculosis, Mycobacterium bovis, Whole-genome sequencing, Bayesian phylodynamics, Structured coalescent, Model misspecification, Surveillance lag, Diagnostic sensitivity

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11614101/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A systematic review of whole-genome sequencing studies of *Mycobacterium bovis* that determine the direction of transmission between badgers and cattle.

P. R. Torgerson[1] G. Ghielmetti[2] and T. E. S. Langton[3]

[1]Section of Epidemiology, Vetsuisse Faculty, University of Zürich, Zürich, Switzerland, email: paul.torgerson@uzh.ch

[2]Section of Veterinary Bacteriology, Institute for Food Safety and Hygiene, Vetsuisse Faculty, University of Zürich, Zürich, Switzerland

[3]Herpetofauna Consultants International Ltd, Suffolk, United Kingdom

Abstract

Background: Tuberculosis, caused by *Mycobacterium bovis*, is a bacterial disease affecting cattle and other domestic species, wild mammals, and with resilience to survive in the wider environment for many months. Whole-genome sequencing (WGS) methods can distinguish closely related isolates and local transmission clusters. However, the relatively slow evolutionary rate of *M. bovis*, with substitutions accumulating more slowly than individual transmission events occur, results in limited genetic divergence between epidemiologically linked isolates and a weak temporal signal, greatly limiting the ability to determine precisely 'who infected whom' and to estimate transmission rates. Differences in surveillance, sampling intensity, geographic coverage, and detection probability between cattle and the European badger *Meles meles* warrants a review to examine the strength of evidence, to better inform disease control considerations.

Methods: Following PRISMA guidance, we searched PubMed/MEDLINE, Web of Science, Scopus, and Embase for peer-reviewed studies using WGS data from sympatric cattle and European badgers, supplemented by citation searching to assemble a broader international layer of contextual descriptive-genomics studies.

Results: The reviewed studies consistently showed close genetic relationships and local sharing of *M. bovis* strains between cattle and badgers. However, estimates of transmission direction varied substantially between studies with large variation in sampling quality and analytical approaches. Some studies reported patterns compatible with greater badger-to-cattle transmission, while others indicated the opposite, and some emphasized within-species spread and cattle movement. Directional estimates were potentially sensitive to unequal host sampling, differences in surveillance processes, weak temporal signal, incomplete observation of infection chains, the treatment of environmental or unsampled transmission pathways, and especially model assumptions. Close genomic similarity supported shared local transmission histories but generally did not establish direct transmission or its direction. Collectively the evidence base is considered insufficiently rigorous for findings to be definitive and this is recognised by most authors.

Conclusions: Current WGS studies demonstrate shared local *M. bovis* strains and epidemiological connections, but directionality and reservoir status remain highly uncertain and particularly sensitive to sampling and modelling assumptions some of which show signs of confirmation bias. Existing caveated estimates of transmission are presently insufficient to inform generalised conclusions about variation in interspecies transmission rates and their use to validate badger intervention in disease control strategy.

Keywords: Bovine tuberculosis; *Mycobacterium bovis*; Whole-genome sequencing; Bayesian phylodynamics; Structured coalescent; Model misspecification; Surveillance lag; Diagnostic sensitivity.

Introduction

Bovine tuberculosis (bTB), is caused by *Mycobacterium bovis*. *M. bovis* is an ecotype of the *Mycobacterium tuberculosis* complex causing mammalian tuberculosis; a chronic bacterial disease that affects livestock, wildlife, and humans. Despite decades of control efforts including culling, vaccination, and cattle testing, new herd outbreaks of bTB persist which are often attributed to transmission from badgers (*Meles meles*) to cattle, thus complicating elimination campaigns.

Whole genome sequencing (WGS) has emerged as a powerful tool for molecular epidemiology, offering resolution far exceeding traditional typing methods such as spoligotyping and MIRU-VNTR (Wyllie et al., 2018). By comparing single nucleotide polymorphisms (SNPs) across isolates, researchers can reconstruct transmission chains, estimate mutation rates, and distinguish between recent and historical transmission events. A proposed threshold of five SNPs between isolates has been suggested as a cutoff for identifying direct transmission of *M. tuberculosis* complex organisms (Walker et al. 2013, Diel et al.,2019).

However, the application of WGS specifically to *M. bovis* isolates from both badgers and cattle to infer cross-species transmission remains poorly characterized in the literature. Mathematical models of the badger-cattle interface have largely relied on parameter estimation rather than direct genomic evidence of spillover. This systematic review aims to identify, appraise, and synthesize all published primary studies that employ WGS of *M. bovis* to track transmission between badgers and cattle, characterizing the methodological approaches, genomic findings, and transmission inferences reported.

Bayesian phylodynamic models map data onto structures where host species are treated as discrete traits (e.g., Badger vs. Cow). The algorithm calculates migration rates between these states based on the branch lengths of the resulting family trees. However, this framework is vulnerable to uneven sampling, including using samples taken from a considerable distance to increase sample sizes:

1. **Cattle Surveillance:** Regular, mandatory testing may ensure rapid detection. Bovine isolates are found early, appearing on the tree as short, fresh evolutionary branches. Some isolates may only be detected at abattoir surveillance from previous untested or test negative cattle.
2. **Badger Surveillance:** Badgers are sampled sporadically via roadkill or research captures. This may allow the pathogen to circulate for years before isolation. Badger-derived strains accumulate more mutations, resulting in long, deeply rooted branches.

This operational temporal lag drives a severe sampling window bias. Because the Bayesian algorithm misinterprets the long, sporadically sampled wildlife branches as deep evolutionary distance, the coalescent matrix forces an artificial inflation of the calculated migration rate out of the badger state to reconcile them with the short, fresh bovine branches. The resulting ten-to-one transmission asymmetry is most parsimoniously explained as an artefact of this surveillance deficit and sampling time distortion, not proof of biological directionality.

M. bovis is characterised by slow growth and a relatively long generation time (16–20 h). When cattle and badgers share isolates with 5 or fewer SNP distances, it confirms they belong to the same local outbreak. However, descriptive genomic studies (e.g., Menardo et al. 2019) stop where the data stops: because the mutation rate is slower than individual transmission events, the direction remains unresolvable. An identical strain can circulate silently within a cattle herd or environment for some time without generating a single SNP mutation; in the US survey of Orloski et al. (2018), 40% of within-herd isolates differed by zero SNPs and none by more than six.

This operational artefact is compounded by structural sensitivities to unprovable assumptions regarding within-host evolution. Proponents of the wildlife-reservoir hypothesis use discrete-state Bayesian phylodynamics and machine learning to map genetic distances against landscape covariates. However, as demonstrated by Wood et al. (2024), the slow and highly variable evolutionary rate of *M. bovis*—often yielding zero variations between hosts—renders definitive transmission tree reconstructions unreliable in low-density or biased sampling frameworks. While WGS remains a valuable adjunct for ruling out epidemiologically impossible pathways or identifying unobserved missing links, extracting transmission directionality from narrow SNP differences represents a mathematical over-extrapolation. By substituting code constraints for

granular spatial metadata, these models transform unresolvable outbreak clusters into artificial transmission asymmetries.

The methodological challenges of applying phylodynamic modelling to low-diversity pathogens are demonstrated by a genomic survey of Welsh *M. bovis* isolates (Healey et al. 2026). The study evaluates multi-host disease dynamics by comparing hundreds of cattle samples against a wildlife sample size of only 10 badgers and two fallow deer. This sampling asymmetry limits the capacity to resolve unobserved wildlife intermediaries, making inferences regarding transmission directionality statistically unsupported. Furthermore, the analysis assigns contemporary sequences to existing institutional clades constructed on restricted evolutionary priors, creating a circular calibration loop. Citing the badger reservoir as an established baseline fact at the outset introduces a confirmation bias, where shared local genetic profiles are automatically attributed to wildlife maintenance rather than being evaluated as potential cattle-to-cattle transmission.

This framework also features an analytical disconnect regarding variant functional profiles. The pipeline may mask hypervariable genomic loci to preserve a neutral molecular clock for phylogeny, yet utilizes these same coding regions to argue for localized host-pathogen adaptation. Notably, the highest abundance of non-synonymous variations within essential genes occurred in pathways governing intermediary metabolism and cellular respiration. Rather than indicating host adaptation, this variation may point to selection pressure linked to off-host survival.

Figure 1 represents the logical framework, with direct transmission between cattle, between badgers and between species, either directly or indirectly via the environment. It also includes the likely recrudescence of infection following periods when bTB remained undetected in a herd.

Transmission Pathway Framework

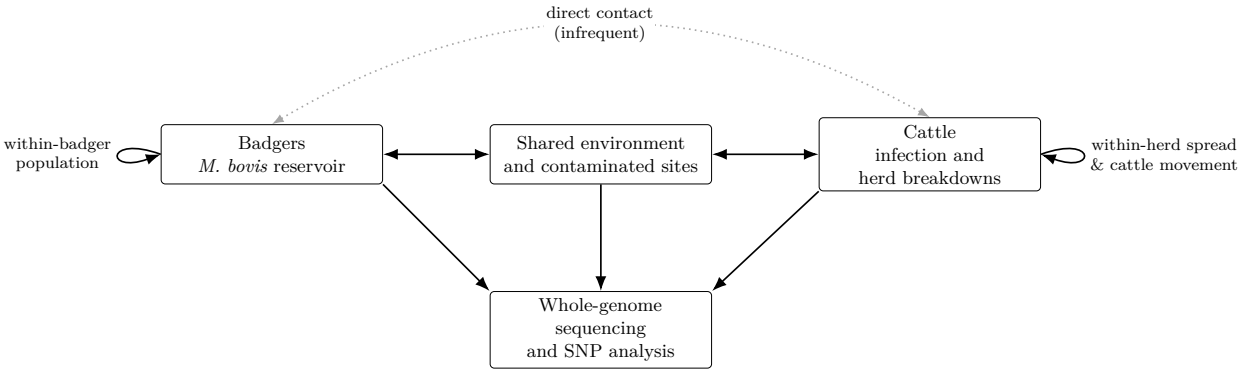

Figure 1: Conceptual framework for genomic investigation of badger–cattle transmission. Solid arrows denote the environmentally mediated and within-population routes that dominate maintenance and spread; the dotted arrow denotes infrequent direct badger–cattle contact. All routes feed isolates into whole-genome sequencing and SNP analysis, which cannot by itself separate direct contact from environmentally mediated transmission.

This study investigated the hypothesis that transmission of bTB between badgers and cattle is overstated in current phylodynamic models. This was undertaken by a systematic review and analysis of literature that has used WGS to infer interspecies transmission.

Materials and Methods

Search Strategy

A systematic literature search was conducted across four electronic databases: PubMed/MEDLINE, Web of Science Core Collection, Scopus, and Embase.Backward and forward citation searching of included studies

and relevant reviews was used to identify additional contextual (Tier 2) studies, including international and non-badger datasets; a final completeness check used the first 1000 Google Scholar hits. Searches were performed in September - October 2026 with no restriction on publication date or language. The core Boolean search combined terms for the pathogen and the sequencing technology only (AND, OR); host species and transmission direction were not used as Boolean filters but were applied during title/abstract and full-text screening, so that contextual multi-host and non-badger datasets were not excluded at the search stage. The full search string (Embase, Emtree syntax; equivalent free-text and MeSH strings were run in PubMed/MEDLINE, and adapted for Web of Science and Scopus) was:

```
('mycobacterium bovis'/exp OR 'mycobacterium bovis' OR 'm. bovis') AND
('whole genome sequencing'/exp OR 'whole genome sequencing' OR
'whole-genome sequencing' OR wgs OR (genome* AND sequenc*))
```

This string was adapted for each database's specific syntax and indexing terms. A final search, which included the first 1000 hits in google scholar was executed on 5th November 2026.

Eligibility Criteria

A two-tier inclusion structure was used. A study entered the core tier (Tier 1) only if it met both conditions: it analysed genome-wide SNP data from *M. bovis* isolates of both European badgers and cattle, and it used those data in a directional or phylodynamic/state-transition framework to infer or characterise interspecies transmission. Studies that analysed WGS of *M. bovis* in cattle, badgers, other hosts, or environmental samples but did not estimate badger–cattle direction were retained separately as contextual studies (Tier 2); this contextual tier was deliberately not restricted to Europe or to badger-containing datasets. Typing-only methods (spoligotyping, MIRU-VNTR/MLVA, RFLP) were ineligible for Tier 1 and admitted only to Tier 2 as context; preprints and other non-peer-reviewed sources were likewise eligible for Tier 2 only and flagged as such. Only Tier 1 studies underwent formal risk-of-bias appraisal.

Study Selection and Flow Characteristics

The systematic multi-database literature search identified a baseline pool of **345 unique records** following deduplication. Screening of titles and abstracts against the predefined inclusion rules resulted in the exclusion of **312 records** that did not utilize primary high-resolution whole-genome sequencing data or lacked explicit multi-host zoonotic transmission metrics.

Consequently, **33 full-text articles** were retrieved and rigorously evaluated for systemic synthesis. Of these, **8 studies** were categorized into the core quantitative synthesis layer because they deployed explicit, mathematically driven phylodynamic or state-transition transmission models to infer host-to-host directionality vectors. The remaining **25 studies** were retained in a separate descriptive tier as contextual molecular evidence to establish spatial baseline maps, clade home ranges, and regional dissemination footprints across global livestock ecosystems.

Data Extraction

Screening and extraction were performed by one reviewer with a large-language-model tool used as an independent cross-check of inclusion decisions and extracted fields; disagreements were resolved by the reviewer against the source papers. They were also cross checked and inconsistencies resolved by the other 2 authors The following variables were extracted: bibliographic details, study location and period, number and source of *M. bovis* isolates from badgers and cattle, sequencing platform and methods, bioinformatic pipelines, SNP thresholds used for transmission inference, phylogenetic or clustering approaches, and — for Tier 1 directional studies — the model class, the prior/root-state and symmetry assumptions, the population-structure (deme) specification, the sampling/observation model, and the type of directional quantity reported (e.g. migration-rate ratio, posterior transition probability, or event count); transmission inferences reported; and key conclusions.

Quality Assessment

Risk of bias was appraised only for the core (Tier 1) directional studies, using a prespecified framework of five transmission-inference domains: host-sampling balance, surveillance-lag control, environmental state-parameter handling, contiguous cattle-pathway modelling, and directionality-inference bias. Each domain was rated Low, Moderate, Serious or Critical and an overall rating assigned per study (Table 3). Contextual (Tier 2) studies were not formally appraised.

Results

Study Selection

The literature search identified 345 records after deduplication, of which 33 were assessed in full text for relevance to the review question (Figure 2). Of these, 8 studies were included as the core evidence base because they directly used WGS of *M. bovis* from both cattle and badgers to investigate interspecies transmission, while the remaining 25 studies were retained as contextual or comparative evidence from other multi-host or surveillance settings.

The directly relevant studies were conducted in Northern Ireland, England, and south-western France, and ranged from tightly localized hotspot investigations to large retrospective population analyses.

Characteristics of Included Studies

Sample sizes varied substantially across the included badger–cattle studies. The earliest proof-of-concept study analyzed 31 isolates from 26 cattle and 5 from 4 badgers in a spatio-temporally linked cluster in Northern Ireland (Biek et al., 2012). Later studies used much denser sampling, including 619 isolates from a 100 km^2 hotspot in Northern Ireland, 167 strains from cattle and badgers in France, and 1,442 genomes from the Randomised Badger Culling Trial in south-west England (Akhmetova et al., 2023; Duault et al., 2022; Van Tonder et al., 2021).

The included literature also contained broader WGS studies from Ireland, New Zealand, Spain, Portugal, Italy, Michigan, and the United States, showing that WGS has been applied across a range of endemic and low-prevalence settings to investigate host sharing, persistence, spatial clustering, and outbreak tracing in *M. bovis* systems (O'Shaughnessy et al., 2026; Crispell et al., 2017; Pozo et al., 2022; Pereira et al., 2023; Scaltriti et al., 2024; Salvador et al., 2019; Orloski et al., 2018).

Use of WGS to Resolve Transmission

Across studies, WGS consistently provided greater discriminatory power than older typing methods and appear to resolve transmission patterns at finer spatial and temporal scales than spoligotyping, VNTR, or MLVA alone (O'Shaughnessy et al., 2026; Allen et al., 2024; Sandhu et al., 2025; Biek et al., 2012).

In Northern Ireland, closely related badger and cattle isolates supported recent interspecies transmission, with badger isolates differing by between zero and four SNPs from the nearest cattle isolate in one local study (Biek et al., 2012). The same study found that bacterial lineages could persist on or near farms for several years despite multiple clear whole-herd tests. This indicates that local persistence in cattle can occur even when routine surveillance temporarily fails to detect infection.

In a larger Northern Ireland hotspot, birth–death and TransPhylo analyses indicated that cattle were likely driving the local epidemic, with cattle-to-badger transmission more common than badger-to-cattle transmission (Akhmetova et al., 2023). In contrast, analysis of isolates from Woodchester Park in England suggested that transmission occurred more often from badgers to cattle than vice versa, while transmission within species exceeded between-species transmission (Crispell et al., 2019). The largest badger–cattle dataset, derived from the Randomised Badger Culling Trial, likewise found that badger-to-cattle transmission was broadly more common than cattle-to-badger transmission, but that most transmission occurred within herds or within badger populations rather than between host species (Van Tonder et al., 2021).

In south-west France, a Bayesian evolutionary model estimated a badger-to-cattle transition rate 52-fold higher than the cattle-to-badger rate, with a median of 64 badger-to-cattle transition events and a median of zero cattle-to-badger events, supporting long-term circulation of *M. bovis* in badgers in that setting (Duault et al., 2022).

In the absence of methodological audit, these studies indicate that the direction and magnitude of transmission between badgers and cattle are not uniform across regions, and instead appear to be strongly context dependent (Akhmetova et al., 2023; Van Tonder et al., 2021).

Spatial Structure, Persistence, and Alternative Transmission Pathways

Several studies found that genetic similarity was associated more strongly with geography than host species, indicating that local spatial structure is a major feature of *M. bovis* transmission dynamics (Pozo et al., 2022; Biek et al., 2012; Allen et al., 2024). In Great Britain and Northern Ireland, WGS-defined clusters had smaller core ranges than traditional MLVA types, improving outbreak tracing and helping distinguish local persistence from dissemination via cattle movement (Allen et al., 2024).

The Randomised Badger Culling Trial dataset identified recurrence of herd breakdowns within the same transmission clusters, as well as superspreading and long-distance transmission events linked to cattle movement (Van Tonder et al., 2021). Similar movement-associated dissemination was reported in Ireland, where a proof-of-concept decision-tree framework classified likely transmission pathways as local area transmission, within-herd transmission, and cattle movement-associated transmission, and suggested undetected carrier animals as a biosecurity concern (Harvey et al., 2025).

Beyond direct animal-to-animal transmission, environmental sequencing showed that animal and environmental genomes were highly intertwined within the same lineages, supporting environmental contamination and plausible indirect transmission routes in endemic settings (Pereira et al., 2024a).

Comparative Evidence From Other Multi-Host Systems

Comparative studies outside the badger–cattle system reinforced the general utility of WGS for identifying host sharing and inferring reservoir roles. In New Zealand, cattle and possum isolates were often indistinguishable, suggesting a high degree of exchange, although directional estimates were sensitive to sampling strategy (Crispell et al., 2017). In Michigan, WGS combined with spatial and Bayesian phylogenetic analyses showed strong spatiotemporal admixture among elk, deer, and cattle isolates, but found no support for substantial transmission between elk and cattle, with ranched deer a possible maintenance host (Salvador et al., 2019). In Ireland, cattle and deer shared highly similar *M. bovis* strains, and the diversity observed in deer suggested that deer may act as a source of infection for local cattle populations in some high-density settings (Crispell et al., 2020). In France, a whole-genome SNP analysis of a single dominant genotype (spoligotype SB0121/SB0295, MLVA 645311257) reconstructed a decades-long transmission network among farmed and park-held red deer, fallow deer, bison, wapiti, wild boar and cattle, resolving two co-circulating sub-clusters — one confined to deer reared as livestock and one to zoos and hunting parks — whose links were inferred from SNP proximity combined with animal-movement records rather than from any phylodynamic directional model (Michelet et al., 2019). Produced by the same national reference laboratory that later applied Bayesian migration-rate inference to the badger–cattle system, it illustrates the descriptive ceiling of SNP clustering for assigning transmission direction, and the relevance of farmed and captive deer, rather than only free-ranging wildlife, as a maintenance context for *M. bovis*.

Studies from Spain, Portugal, and Italy similarly found that closely related strains occurred across wildlife and livestock hosts, that genetic structure was often geographic rather than host-specific, and that persistent local clonal clusters could be detected over time (Pozo et al., 2022; Pereira et al., 2023; Scaltriti et al., 2024).In low-prevalence or surveillance-oriented settings, WGS also disclosed previously unrecognized links among herds, suggested undocumented cattle movement and importation as possible sources, and reduced the time and cost of epidemiological investigations (Orloski et al., 2018).

Methodological Limitations

Several studies emphasized that the low genomic variability and slow evolution of *M. bovis* constrain inference of transmission direction at fine scales, particularly when host sampling is imbalanced or when epidemiological metadata are sparse (Biek et al., 2012; Crispell et al., 2019; Harvey et al., 2025). Even where WGS showed clear host sharing, directionality was sometimes unresolved, and model outputs could be sensitive to sampling strategy or the analytical unit chosen (Crispell et al., 2017; Duault et al., 2022).In multi-host systems analysed by descriptive SNP clustering alone, transmission direction could not be assigned from the genomic data at all: a French network linking farmed and park-held deer, wild boar and cattle could only be resolved into epidemiologically coherent chains once SNP distances were combined with animal-movement and husbandry records, underscoring that WGS localises and dates closely related isolates but does not, by itself, establish who infected whom (Michelet et al., 2019).

Overall, the evidence supports WGS as a high-resolution tool for reconstructing local persistence, spatial clustering, and interspecies links in bTB. However the inference about "who infected whom" and transmission rates are less robust and depends multiple variables such as density of sampling of all relevant species and bacterial refuges, dense sampling, representative host coverage, and integration with uncertain movement, spatial, and ecological data (Biek et al., 2012; Crispell et al., 2019; Harvey et al., 2025).

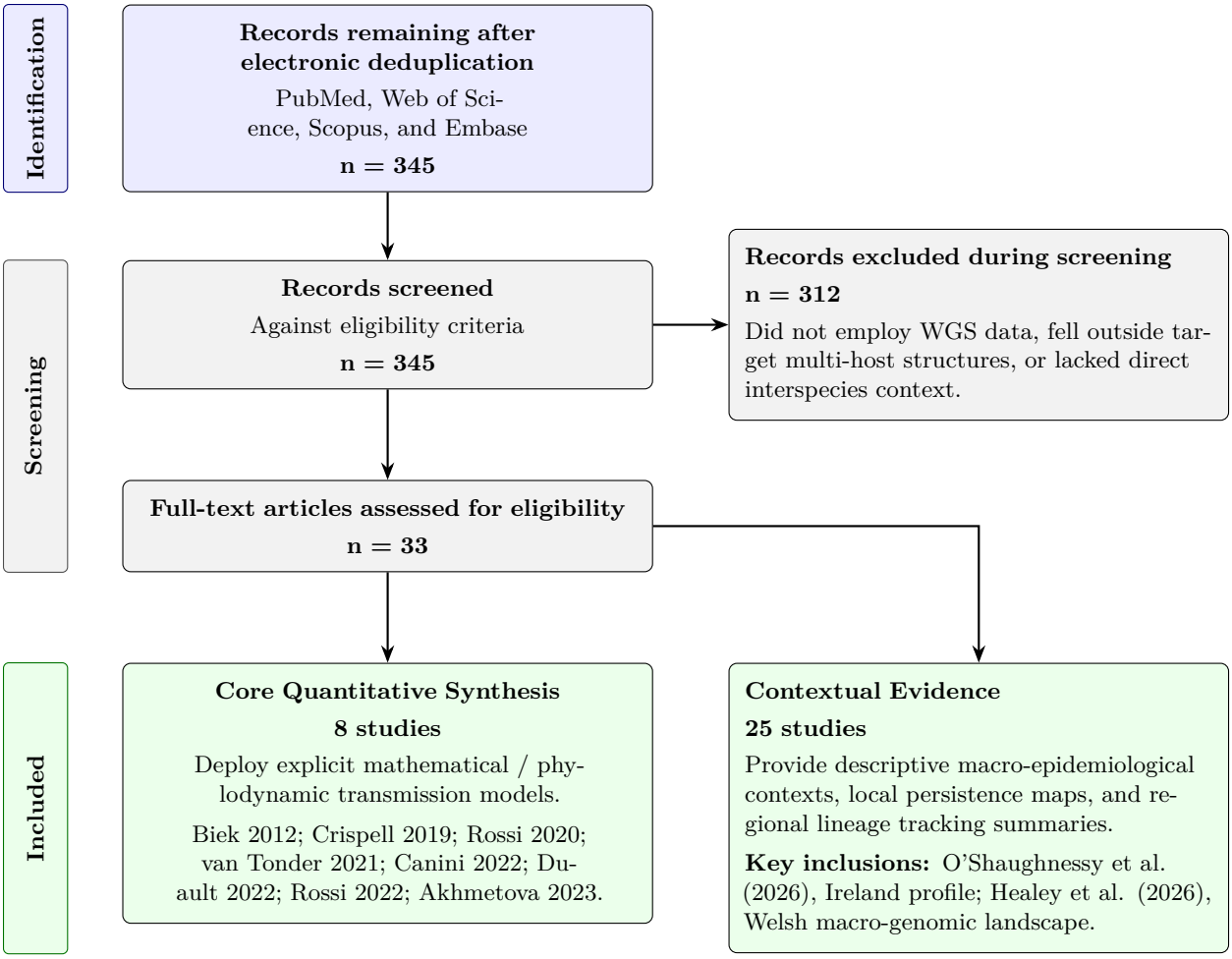

Figure 2: PRISMA-style flow diagram detailing identification and screening, followed by stratification into 8 core quantitative-synthesis studies and 25 contextual molecular mapping studies.

Table 1: Core Badger–Cattle Whole-Genome Sequencing Studies and High-Resolution Quantitative Models

#	Study	Setting / Region	n (Badger / Cattle)	Main Analytical Finding & Systematic Role
1	Biek et al. (2012)	Northern Ireland	5 / 26	Zero-to-four SNP gaps supported localized recent transmission links but authors appropriately refused to infer directionality due to slow clock limits.
2	Crispell et al. (2019)	England (Woodchester)	193 / 159	Structured-coalescent (BASTA) analysis estimated badger-to-cattle transmission $\sim 10.4\times$ the reverse in its best-supported two-deme model; sensitive to model structure (a four-deme model gave near-equal rates) and to sampling bias.
3	Rossi et al. (2020)	England (Woodchester)	191 / 69	Pairwise Kolmogorov Forward Equations (KFEs) mapped transmission networks but outputs were hyper-sensitive to unprovable latency bottleneck settings.
4	Van Tonder et al. (2021)	SW England	690 / 752	Structured-coalescent (BASTA) estimate of badger→cattle transmission exceeded the reverse (median $2.1\times$), but the interval crosses unity, within-species paths dominated, and long-range seeding was cattle-movement-driven.
5	Canini et al. (2022)	France	Multi-host co-hort	MASCOT structured coalescent model mapped rare, region-specific circuits; identified wild boar intermediaries but left $\sim$90% of transition nodes unknown.
6	Duault et al. (2022)	SW France	21 / 146	MASCOT structured coalescent model calculated a 52-fold higher badger-to-cattle rate, an artefact of a badger effective-population-size estimate $\sim 28\times$ that of cattle (from only 21 badger sequences) extrapolated into a pre-2013 period with no wildlife sampling.
7	Rossi et al. (2022)	England (East Cumbria)	41 / 24	BASTA structured coalescent model resolved a dominant cattle-to-badger vector in a naive, single-introduction outbreak, consistent with badgers as a downstream spill-over host rather than a reservoir.
8	Akhmetova et al. (2023)	Northern Ireland	769 badger microsatellites* / 619 WGS total	High-density sympatric TransPhylo and birth–death networks indicated a cattle-driven epidemic; cattle-to-badger vectors outpaced badger-to-cattle events.

Note: *Akhmetova et al. (2023) utilized 769 host badger microsatellite genotypes alongside circulating bacterial pathogen isolates to cross-examine host family structures against strain evolution.

Table 2: Systematic Review Tier 2: Summary of the Contextual and Descriptive Whole-Genome Sequencing Studies (Chronological Order)

#	Study & Year	Setting / Country	Core Focus & Systematic Role
1	Trewby et al. (2016)	Northern Ireland	Continuous phylogeographic diffusion models show slow, heterogeneous spread (2 km/year) and expose key limits in tracing low-divergence lineages.
2	Glaser et al. (2016)	Minnesota, USA	Cattle and white tailed deer. Its own authors concede isolate dates $\neq$ infection dates — direct support for the sampling-time/latency critique and the short-time-depth resolution limit.
3	Crispell et al. (2017)	New Zealand	Large-scale genomic sequencing evaluating cattle-wildlife host sharing and temporal sampling sensitivities.
4	Dippenaar et al. (2017)	South Africa	African buffalo, lion, leopard, kudu, baboon; a few cattle as outgroups. Shared strain but no direction of transmission.
5	Kohl et al. (2018)	Germany	Captive cervids (red/fallow/sika deer), bison, lynx, pot-bellied pigs — no badgers, no cattle. Exemplar that WGS without a directional model yields relatedness/persistence only — supports the Pathway A vs B distinction; non-endemic captive-system boundary case.
6	Orloski et al. (2018)	USA	National descriptive survey of 1,248 cattle and farmed-cervid isolates (no badgers); traces cryptic herd links and undocumented-movement sources with minimal within-herd divergence.
7	Salvador et al. (2019)	Michigan, USA	Spatiotemporal admixture tracking across sympatric elk, deer, and cattle populations.
8	Michelet et al. (2019)	France	Cattle, farmed/park red and fallow deer, bison, wapiti, wild boar, 1 primate. Descriptive-genomics (Pathway A) exemplar in a multi-host captive/farmed system; illustrates the ceiling of SNP-clustering-plus-epidemiology for directionality and the deer/park–cattle interface.
9	Crispell et al. (2020)	Ireland (Wicklow)	Evaluates sika/fallow deer using WGS; suggests a potential role as a localized reservoir.
10	Perea et al. (2021)	Spain	WGS/SNP phylogenetic mapping of 125 *M. bovis* isolates from cattle and wild boar (no badgers); Establishes regional clade and cluster footprint.
11	Pozo et al. (2022)	Spain	Diversity mapping of prevalent spoligotypes in a high-density multi-host ecosystem.
12	Downs et al. (2022)	UK	Cattle (surveillance); badger surveys (verification only). Cattle clade-mapping only; no badger WGS available. Reservoir assumed badger-associated, not estimated.
13	Legall and Salvador (2022)	UK/NZ/USA (pooled, 700 public genomes)	Comparative genomics re-use of Crispell 2017/2019 and Salvador 2019 data. Genomic-signature discrimination only — no transmission-direction inference.
14	Roos et al. 2023	South Africa : Wildlife	Attribution-bias exemplar: WGS splits a shared matched pair "spillover" into two distinct clades. Lower-resolution methods overstate connectivity; also illustrates the isolate-date vs divergence-date caveat.

Continued on next page

Table 2 continued from previous page

#	Study & Year	Setting / Country	Core Focus & Systematic Role
15	Pereira et al. (2023)	Portugal	Phylogenetic mapping and regional dissemination footprints of Eu2 clades.
16	Pereira et al. (2024a)	Portugal (Environment)	Environmental WGS tracking transmission dynamics at the animal-environment interface.
17	Pereira et al. (2024b)	Portugal	Eco-phylodynamics of 170 *M. bovis* genomes (Eu1/Eu2/Eu3) in a cattle–red deer–wild boar system; Rigorous directional inference, but no badgers.
18	Allen et al. (2024)	Northern Ireland	Resolution comparison of MLVA/spoligotype vs WGS for routine bTB tracing in lineage 1.140 (2,108 typed isolates; 148 WGS, incl. 27 badgers). SNP15 phylogeographic core ranges, movement-informed source tracing, and BEAST skyline/Re population history. Descriptive phylogeography and lineage tracking; no badger–cattle directionality inference.
19	Scaltriti et al. (2024)	Italy	Genomic analysis of persistent local clonal clusters within the SB0120 spoligotype.
20	Wood et al. (2024)	Great Britain	Evaluates 241 cattle and badger isolates; uses pairwise numerical integration to prune implausible pathways but demonstrates extreme sensitivity to the molecular clock prior.
21	Harvey et al. (2025)	Ireland	Proof-of-concept decision-tree framework to classify local and movement-linked pathways.
22	Sandhu et al. (2025)	Great Britain	Enhanced genomic diversity analysis of *M. bovis* to aid national bTB control.
23	Healey et al. (2026)	Wales	Regional macro-genomic cluster analysis identifying six distinct multi-host clades.
24	O'Shaughnessy et al. (2026)	Island of Ireland	Country-wide spatial mapping; highlights trade-driven multi-strain seeding inside CFUs.
25	Akhmetova et al. (2026)	Northern Ireland (County Down)	MLVA/VNTR typing with kernel discriminant analysis of badger–cattle spatial concordance in the TVR zone; reports co-location, not transmission direction. Typing-based (not WGS) and an unrefereed preprint.

The 8 studies that met the inclusion criteria with respect to inferences about transmission between badgers and cattle are summarized in more detail.

Biek et al. 2012

Biek et al. (2012) applied whole-genome sequencing to 31 *Mycobacterium bovis* isolates — 26 from cattle and 5 from 4 road-kill badgers (one badger sequenced from two tissues) — across five neighbouring farms, no more than 5~km apart, in Northern Ireland. The study demonstrated that WGS provides finer resolution than VNTR typing for tracking local herd-level spread. The authors found strong evidence for local lineage persistence on or near farms over several years despite multiple clear whole-herd tests, and showed that spatial proximity correlated with genetic distance more closely than did the recorded network of cattle movements. Badger isolates lay between zero and four SNPs from the nearest cattle isolate — two were genetically indistinguishable from cattle isolates — providing evidence for recent cross-species transmission or a shared local source.

The study is constrained by a very small badger sample (four animals) and by the slow evolutionary rate of *M. bovis* (estimated at 3.40×10^{-8} substitutions per site per year), which together limit resolution at the within-herd scale and preclude any inference of transmission direction. The transmission model is a mechanistic, compartmental (S–E–T–I) formulation; its one acknowledged shortcut lies not in its mechanistic structure

but in the likelihood, which the authors note was approximated by a state-probability calculation rather than fully integrated. These are limitations the authors state plainly rather than work around.

Within those limits the analysis is methodologically sound and its reporting is appropriately restrained. On finding tight genetic clusters (0–4 SNPs) between sympatric cattle and badgers, the authors explicitly decline to infer directionality, judging it unresolvable given the small badger sample, the slow molecular clock, and the long, variable interval between infection and sampling, while noting that denser sampling would be needed to quantify it. The model does not pre-specify a wildlife vector: infections not explained by transmission among sequenced reactor cattle are absorbed into a single external force of infection (a_{ext}) — an unsampled "reservoir'' that may equally represent latent or undetected cattle infection, environmental persistence, unrecorded local movement, or badgers — so no directional pathway is hard-coded. The paper therefore supports its claims about local genomic structure and recent interspecies linkage, but makes, and warrants, no stronger claim about transmission direction or the role of badgers as a reservoir. On this review's framework that places it at **Moderate** overall risk of bias: a genuine host-sampling imbalance (D1, only four badgers), offset by restraint that stops the imbalance from driving any directional conclusion.

Crispell et al. (2019)

Crispell et al. (2019) combined whole-genome sequencing, detailed epidemiological metadata, Bayesian phylogenetics, and machine learning (random forest and boosted regression models) to quantify bi-directional transmission of *Mycobacterium bovis* between badgers and cattle at Woodchester Park, south-west England. After quality control the analysis used 193 badger-derived sequences (from 98 badgers sampled 2000–2011) and 159 cattle-derived sequences (1988–2013) — one of the densest multi-host *M. bovis* datasets available. Three complementary analyses were run: random-forest and boosted-regression models relating genetic to spatial, temporal, and network distances; a phylogenetic clade analysis; and a structured-coalescent (BASTA) migration model. The preferred two-deme BASTA model estimated badger-to-cattle transitions at 10.4 times the reverse rate, while all analyses found within-species transmission to exceed between-species transmission in both hosts.

Two features of the study's own framing bear directly on how its headline should be read. First, the 10.4-fold figure is a modelled population transition-rate ratio, not an individual-level transmission probability, a reproduction number, or the proportion of cattle infections attributable to badgers. Second, the authors do not present it as a case for badger removal: their stated conclusion, explicitly conditioned on the sample being representative, is that control should target *both* cattle and badgers, and the eLife summary frames culling only as something that could be made more targeted and less impactful on badgers if the results were confirmed. The analysis is also conducted, and described, as being carried out in the presence of known data biases, with an extensive mislabelling-detection and removal step applied before inference.

The directional estimate is nonetheless not robust. It is sensitive to model structure: the two-deme configuration yields the 10.4-fold asymmetry, whereas a four-deme configuration gives near-equal rates. It rests on a spoligotype shared by cattle and badgers, which can inflate inter-species transitions relative to the full transmission system, and it does not represent cattle-to-cattle, environmental, movement-mediated, other-wildlife, and unsampled pathways with equivalent detail. The cattle were sampled over a wider window than the badgers (1988–2013 versus 2000–2011), so the basal position of several cattle-derived genomes — which the authors note is consistent with a cattle origin for those clades — is confounded with sampling time and cannot be read cleanly either way. The slow, variable molecular clock of *M. bovis* limits transmission-tree reconstruction in low-diversity, unevenly sampled data, so extracting direction from narrow SNP differences carries a real risk of over-interpretation.

One revision step warrants specific comment and is documented in the journal's published peer-review record. In the first submission a phylogenetic ancestral-character estimation placed the ancestral nodes of most inter-species clades in cattle — a pattern the Reviewing Editor noted was "seemingly inconsistent with the badger-reservoir hypothesis." The editor asked the authors to interpret that result and to give "careful consideration of the contrasting observations…and how they can be reconciled"; removal was not requested. The authors instead removed the analysis entirely, stating that the method was "highly sensitive to sampling biases" and that the structured-coalescent (BASTA) approach, which can accommodate such

biases, was preferred. Dropping a sampling-sensitive analysis has a legitimate methodological rationale, and the exchange is transparently on the public record. But the same sampling imbalance was not shown to be innocuous for the retained BASTA result, and the authors were asked to reconcile the contradictory signal, not to delete it. Choosing deletion over reconciliation removed the one analysis pointing toward a cattle origin while retaining the one consistent with a badger-reservoir reading — an asymmetric treatment of the evidence, taken under the senior author's responsibility, that is consistent with confirmation bias.

Overall, Crispell et al. (2019) provide good evidence of local host sharing and of substantial within-species transmission, and the dataset and quality-control steps are strong. The directional headline, however, is not merely fragile but selectively supported. The two-deme 10.4-fold estimate is not robust to deme structure, and the single analysis that pointed toward a cattle origin — the ancestral-state reconstruction — was removed in revision while the equally sampling-sensitive BASTA result was retained. Because that curation leaves the published evidence base more consistent with a badger-reservoir reading than the authors' own full analysis supported, the hedged verbal conclusions ("if representative"; target both species) rest on a directional signal that has been over-stated by selective presentation. The estimate does not establish that badgers are the dominant maintenance host, that they cause the majority of cattle infection, or that badger removal would materially reduce cattle incidence. On this review's framework, which scores directionality on the evidential core and does not credit a study for restraint it does not, on inspection, display, the selective removal of a disconfirming analysis is a core-integrity failure that directly shapes the primary directional claim, placing the study at **Critical** overall risk of bias.

Rossi et al. (2020)

Rossi et al.(2020) used the Kolmogorov forward equations (KFEs) to compute pairwise transmission probabilities of *M. bovis* among cattle and badgers in the Woodchester Park system. The approach combines pathogen SNP distances, sampling dates, and a simple SEI model of disease progression to rank plausible host-to-host transmission links in a low-diversity, data-sparse setting. It is, by design, a methodological study (Figure 3): its stated objective is to demonstrate that transmission signal can be extracted when pathogen genetic diversity is low and metadata are limited, rather than to quantify the epidemiological balance of transmission between species. This framing is important, because it means the study neither attempts to, nor is structured to, estimate the relative rates of cattle-to-badger and badger-to-cattle transmission.

A first limitation is the sensitivity of inferred direction to assumptions about within-host evolution, transmission bottlenecks, and the timing of mutation during the latent and infectious periods. When the authors varied whether SNP substitution could occur during latency (bottleneck at infection) or only during the infectious stage (bottleneck at the end of latency), the inferred configuration changed for a subset of host pairs and triads: in some triads the reconstructed chain was restructured or the inferred direction reversed. Notably, this reversal arose from changing an unobserved within-host assumption while the genomic and temporal data were held fixed, rather than from the SNP data themselves. It should also be noted that this robustness test was performed for one clade only (clade 3), so the demonstration of model-dependence, while genuine, rests on a limited sample; only a minority of closely-related clusters were comparatively robust, and even among these the authors report that in one the positions of many hosts and the inferred direction of transmission still differed. The appropriate conclusion is therefore not that directionality is universally unidentifiable, but that for an appreciable subset of pairs it is not robustly identified once plausible, unobserved within-host parameters are varied.

A second limitation follows from the pairwise structure of the framework. It evaluates candidate host-to-host links rather than reconstructing a complete contact network, and the authors themselves note that pairwise approaches introduce dependency problems that may bias the transmission likelihood calculations. The method does, however, retain a useful capacity that should be acknowledged: through the difference in sampling times relative to the expected molecular clock, it can flag transmission links that are more plausibly mediated by an unsampled intermediate than by direct contact. What it cannot do is resolve the nature of that intermediate. It therefore has limited capacity to distinguish direct transmission from infection acquired through an unobserved cattle, badger, environmental. The absence of an explicit environmental state compounds this: an environmentally mediated exposure may be represented as a direct animal-to-animal link because no separate compartment is available, although the direction of any resulting bias

cannot be assumed without explicit simulation.

A third limitation, and the one most consequential for any inference about the direction of transmission between species, is the method's reliance on unbiased sampling. Cattle and badgers in this system are ascertained by fundamentally different processes: cattle through routine, scheduled skin-test and slaughterhouse surveillance, and badgers through capture-mark-recapture trapping with its own, distinct diagnostic sensitivity. Because the inference leans heavily on the ordering of sampling times, any species-specific difference in the interval between infection and detection can systematically bias the inferred direction of cross-species links. This concern is amplified by the long and highly variable latency of *M. bovis*, which weakens the correspondence between sampling-time ordering and infection-time ordering; indeed, the authors report cases in which the individual carrying additional SNPs is inferred as the source purely on the basis of sampling time. Differential ascertainment between cattle and badgers is thus a plausible source of directional error that the framework, as applied, cannot correct for.

A fourth limitation is that the model does not parametrise transmission in a species-specific or direction-specific manner. The contact-rate prior is a single uniform distribution applied identically to every host pair, and the latency and substitution-rate priors are likewise common to cattle and badgers. Consequently the inferred probabilities are driven by genetic distance and sampling time, not by any species-specific contact structure, and the framework as run cannot estimate an asymmetry between badger-to-cattle and cattle-to-badger transmission. The authors note that such structure could in principle be embedded, but it was not; this is the mechanistic reason the study cannot speak to a rate asymmetry between the two directions.

A fifth limitation concerns the specification of the priors and, more importantly, the way the parameters are estimated. The epidemiological parameters are drawn from literature-informed distributions: a beta-PERT distribution for the substitution rate (mode 0.31, range 0.1 to 0.94 substitutions per genome per year, taken from other *M. bovis* systems), a beta-PERT distribution for the latency period (mode 348, range 116 to 827.5 days, from cattle-focused literature), and a uniform distribution on the contact rate over the interval (0, 0.1). Each of these choices is defensible as a summary of the available evidence, so the priors are not malformed; the difficulty lies in three specification decisions and one estimation decision. First, the parameters are not integrated over, as they would be in a Bayesian posterior, nor fixed at a central value: the authors perturb them with a Gaussian kernel for 1000 iterations and then select the combination that yields the highest transmission probability. Reporting the maximum achievable probability over the prior space imparts an upward, optimistic bias to every pairwise estimate, and because each pair and each direction is maximised independently the resulting probabilities are not evaluated at a common parameter set, which weakens their comparability across pairs and across the two directions of a pair. The wider the priors, the greater the scope for this inflation. Second, a single parameter set is applied to both species; in particular the latency prior is derived from cattle data yet applied to badgers, whose disease progression differs, so the timing logic on the badger side of any cross-species pair is mis-calibrated. Third, the substitution rate is imported from other systems rather than estimated locally and is treated as a single clonal clock with no lineage-to-lineage variation, despite the authors' own emphasis that the *M. bovis* clock is slow and highly variable; since directionality is inferred by reconciling SNP accumulation against sampling-time gaps, a mis-centred or homogeneous clock can systematically rescale that mapping and reverse the inferred direction, especially at differences of zero or one SNP. Finally, the upper bound of the uniform contact-rate prior is arbitrary. None of these issues is examined in the paper: the only robustness analysis reported is the structural bottleneck variation, and no sensitivity analysis of the priors or of the maximisation-based estimator is presented.

A further, more minor point concerns the construction of the candidate contact structure. Clades were defined by genetic distance (isolates within ten SNPs of one another) and then used as proxies for likely contact clusters within which transmission was inferred. Genetic similarity is thus used both to delimit the candidate contact groups and to infer transmission within them, introducing a degree of circularity that should be borne in mind when interpreting the within-clade trees.

These limitations do not invalidate the KFE approach as an exploratory method for ranking epidemiologically plausible links, and its ability to identify likely missing hosts in a low-diversity system is a genuine contribution. They do, however, constrain the interpretation of the output. The inferred transmission probabilities should be regarded as conditional likelihoods under specified within-host, bottleneck, and sampling

assumptions, not as direct observations of transmission or as unbiased estimates of cattle-to-badger and badger-to-cattle rates. The study demonstrates that, for a substantial subset of pairs in this low-diversity system, directionality is model-dependent, and it neither attempts to nor can establish a dominant wildlife-to-cattle pathway or justify badger intervention.

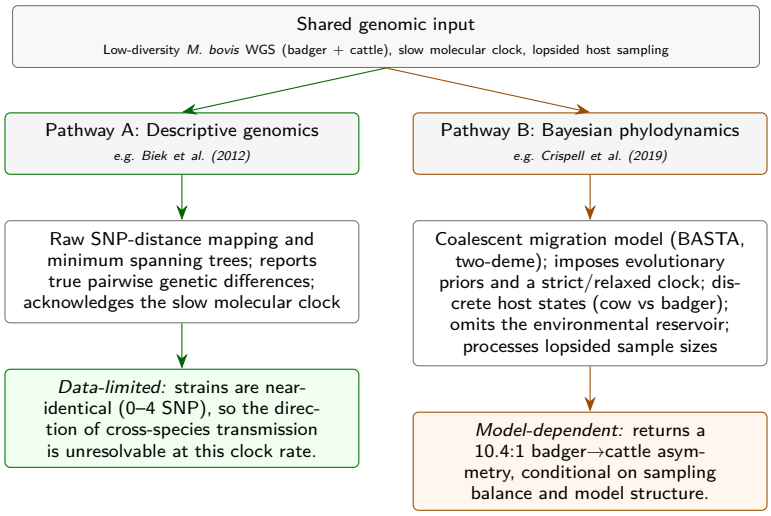

Figure 3: Methodological divergence from a shared genomic input. Descriptive approaches (Pathway A, e.g. Biek et al. 2012) report that low pairwise SNP distances leave the direction of badger–cattle transmission unresolvable, whereas model-dependent Bayesian phylodynamic approaches (Pathway B, e.g. Crispell et al. 2019) return a directional asymmetry that is conditional on sampling balance and model structure.

Van Tonder et al. (2021)

Van Tonder et al. (2021) combined whole-genome sequences from 1,442 *M. bovis* isolates, collected from contemporaneous badger (n = 690) and cattle (n = 752) populations across the ten Randomised Badger Culling Trial (RBCT) areas, with location and cattle-movement data to identify transmission clusters and infer rates and routes of spread. Using a structured-coalescent approach (BASTA), they estimated that badger-to-cattle transmission was a median of 2.1 times higher than cattle-to-badger transmission, that within-species transmission considerably exceeded cross-species transmission for both hosts, that long-distance spread and superspreading were driven by cattle movement rather than badgers, and that most trial areas contained a single cluster established shortly before sampling, often contemporaneous with the expansion of bTB in the 1980s. They concluded that these clusters were established by long-distance seeding events involving cattle movement, not by recrudescence from a long-established wildlife reservoir.

Several features of the study qualify how far its headline asymmetry can be read as evidence for a dominant badger-to-cattle pathway.

The headline directional asymmetry is not statistically robust. The pooled median badger-to-cattle transmission rate of 2.1 times the reverse direction carries a 95% highest posterior density interval of 0.8–3.8, which includes unity and extends below it. The estimate is therefore statistically compatible with no asymmetry, or with a slight reversal of direction. Only two of the twelve clusters (Cluster 1 and Cluster 9) showed a statistically significant directional difference, and four of the twelve clusters ran in the reverse direction (cattle-to-badger higher). The per-cluster ratios vary from 1.2 (Cluster 11) to 13.6 (Cluster 1), underscoring that the pooled figure is an unstable summary rather than a consistent directional signal. On the authors'

own admission, the study does not establish that badger-to-cattle transmission dominates.

The paper's own reconstruction shows within-species transmission dominates and cross-species spill-over is the least frequent category. Even taken at face value, the transmission reconstruction attributes 60 of 84 highly-supported pairs to within-species events; cattle-to-cattle events averaged roughly 17 times, and badger-to-badger events roughly 4.7 times, the number of badger-to-cattle events. Long-distance seeding and superspreading were cattle-driven, and clusters were attributed to cattle movement rather than a wildlife reservoir. Badger-to-cattle spill-over is thus the rarest transition in the authors' own analysis, which sits awkwardly with any interpretation of the study as support for wildlife-driven maintenance of infection in cattle.

The temporal signal underpinning the dating narrative is weak by the authors' own diagnostics. Root-to-tip regression yielded very low or negative R^2 values for every transmission cluster, and dated-tip randomisation gave a clean separation from randomised data for only five of the twelve clusters, with two clusters (2 and 4) effectively failing the test, with the remaining 5 clusters having some overlap. The molecular-dating estimates that support the claim that clusters were established shortly before sampling therefore rest on a temporal signal the authors acknowledge is limited by the slow substitution rate of *M. bovis* and the twelve-year sampling window.

Structured-coalescent rate estimates are sensitive to sampling and to the equal-population-size assumption. The BASTA analysis was run under a strict clock and an equal-population-size model for the two host demes. Migration-rate estimates from structured-coalescent methods are known to be sensitive both to sampling proportion and to mis-specification of the relative effective population sizes of the demes. The authors estimated that they sampled approximately 40% of infected cases across the clusters, and host effective population sizes were neither equal nor separately modelled. A directional ratio whose credible interval already crosses unity should not be read as a robust directional result under these conditions.

The RBCT culling-effect claim shaped the sample provenance but does not enter the inference. The study sequenced isolates drawn exclusively from the proactive-cull arm of the RBCT, an experiment whose stated purpose was to quantify the effect of badger culling on cattle bTB incidence; the data were accordingly restricted to the trial's proactive triplets. This provenance reflects the culling assumption at the level of study design, but the assumption is not a parameter of the analysis: the BASTA runs use a two-deme (badger/cattle) structure with no culling covariate, and the directionality estimate is driven by the sampling completeness and equal-population-size issues described above, not by any modelled culling effect. The appropriate inference is therefore about the direction of transmission in the sampled setting, and neither supports nor refutes claims about the effect of culling on cattle bTB incidence.

The estimates describe one place and period and do not, and cannot, establish a generalisable dominant pathway. The rates were inferred from a single twelve-year window in South-West England during an active culling experiment. The study neither attempts to, nor can, establish a generalisable, present-day transmission pathway for endemic, non-culled settings; its estimates are properly read as a description of that specific epidemiological context rather than as a general quantification of the badger contribution to cattle infection.

Finally, the value of the RBCT as a reference point for badger control is itself contested. Re-analyses report that, once baseline population-at-risk offsets and unconfirmed early-stage reactor herds are accounted for, several formulations show no statistically supported effect of proactive culling on the herd incidence of bTB, though these re-analyses have themselves been disputed. That debate concerns the epidemiological effect of culling and is independent of the phylodynamic estimates above, which are derived from genome sequences rather than from the trial's herd-breakdown statistics. Crucially, the present critique does not depend on either the presence or the absence of a culling effect: because no culling effect is assumed anywhere in the analysis, the directionality inference is unaffected by that contested literature. The RBCT context is relevant here only to the interpretation and generalisability of the estimates, not to their internal validity.

Rossi et al. (2022)

Rossi et al. (2022) analysed an emergent *M. bovis* outbreak in East Cumbria, an area with no recent record of infection in cattle or wildlife. Combining whole-genome sequences with epidemiological data, they identified the outbreak as a single introduction from Northern Ireland, most likely via movement of infected cattle, entering the cattle population roughly six years before it was first detected in badgers. Structured-coalescent analysis (BASTA), with the population split into four subpopulations by host and location, estimated cattle-to-badger transmission to be at least an order of magnitude more frequent than badger-to-cattle transmission, with little support for including badger-to-cattle transmission in the model at all. The authors concluded the outbreak was cattle-introduced and cattle-driven, with badgers infected only once it was established, and found "little evidence for substantial badger-to-cattle transmission".

This study is directly supportive of the present argument. In a naive setting, where the direction of establishment could be reconstructed with unusual clarity, the wildlife contribution was minimal and the driver was the movement and local spread of cattle. The same structured-coalescent methodology used elsewhere to argue for a badger-to-cattle pathway here points the other way. Its direction of inference rests on several convergent lines—the monophyly of the Cumbria clade within the Northern Irish phylogeny, the six-year head start in cattle, and the low diversity of badger isolates consistent with a recent, downstream badger outbreak. The authors are appropriately cautious, reporting only "little support" for badger-to-cattle transmission, wide credible intervals on the timing (point estimate March 2011; 95% bounds August 2001 to April 2014), and sparse early badger sampling that limits the earliest dynamics. The residual caveats are interpretive rather than substantive: BASTA migration rates are net directional flows of lineages between demes and cannot separate direct contact from environmentally mediated or third-party routes; and the Boosted Regression Tree analysis explained cattle genetic distances only poorly, which the authors attribute to hidden cattle contact patterns—unrecorded movement and grazing across non-contiguous parcels obscured by main-building coordinates—pointing to further unmeasured cattle-to-cattle connectivity rather than to any wildlife role. The authors further note that early cases, including the index introduction, likely escaped detection under the four-yearly herd testing and slaughterhouse inspection of a low-risk area; given the limited sensitivity of the routine tuberculin skin test in cattle, the infection may have circulated and been moved between herds undetected for a considerable period before discovery, with spill-over into badgers arising as a downstream consequence rather than a cause. Taken together, the study is consistent with badgers acting here as a spill-over host rather than a reservoir host, and stands as evidence against the general paradigm of wildlife-driven maintenance of bTB in cattle.

Duault et al. (2022)

Duault et al. (2022) used whole-genome sequencing of 167 *M. bovis* strains (21 badger, 146 cattle; spoligotype SB0821) from south-west France and the Marginal Approximation of the Structured Coalescent (MASCOT, BEAST2) to infer host-structured transmission. They reported a badger-to-cattle transition rate (median 2.2 transitions/lineage/year) some 52 times the reverse rate (median 0.042), reconstructed over 75% of lineages from 1989–2000 as badger-borne, and recorded a median 64 badger-to-cattle events against zero the other way, concluding that *M. bovis* had long been present in the local badger population. The authors are explicit about the limits of this, however: the model infers inter-species lineage transitions rather than intra-population transmission, does not represent farms or badger social groups, and "could not confirm badgers as possible intermediaries in farm-to-farm transmission". The study therefore neither attempts to, nor can, establish a badger-mediated farm-to-farm pathway.

Two features further qualify the headline asymmetry. First, it is largely a consequence of the estimated effective population sizes rather than of an observed transmission flux: the badger N_e (median 34) was put at roughly 28 times the cattle N_e (median 1.2) on the basis of only 21 more-diverse badger sequences, and in a structured coalescent ancestral lineages are preferentially assigned to the larger-N_e deme, so backward-in-time migration flows out of badgers almost by construction—and backward migration is not forward transmission intensity. Second, the deep-time badger dominance is extrapolation beyond the data: wildlife surveillance began only in 2012 with the first badger case in 2013, so there is no badger sequence before 2013, yet lineages are reconstructed as badger-borne back to the 1990s, and the "52-fold" is a ratio of medians whose credible intervals are wide and, for the reverse rate, bounded essentially at zero. The directional

signal, in short, is manufactured by the model's demographic priors and ancestral extrapolation, not read off the epidemiology. Strip these artefacts away and the same data are equally consistent with badgers as a downstream, spill-over host: Duault et al. (2022) establish only that *M. bovis* circulates in local badgers—not that badgers maintain it, still less that they drive and sustain infection in cattle. On this evidence the reservoir reading is an assumption of the analysis, not a finding of it.

Canini et al. (2022)

Canini et al. (2022) applied a structured coalescent model (MASCOT, with BSSVS, in BEAST2) to whole-genome sequences of two Eu3 *M. bovis* genotypes from sympatric cattle, badgers (*Meles meles*) and wild boars (*Sus scrofa*) in two French areas—Côte-d'Or (CdO; 144 isolates) and part of Nouvelle-Aquitaine (NAq; 219 isolates). They inferred a cattle→badger→wild boar→cattle circuit in CdO, with an additional badger→cattle rate in NAq, and proposed wild boars as an intermediary between badgers and cattle. The authors are, to their credit, markedly cautious: between-species transitions were infrequent (under 12% of known events), nearly 90% of transition events in NAq could not be labelled, the model cannot quantify within-species transmission, and they state the results "should therefore be considered with caution". The study thus neither attempts to, nor can, establish that wildlife maintains and drives infection in cattle.

The study's directional inferences appear to be heavily influenced by the estimated effective population sizes rather than a clear signal of transmission direction. Because the coalescent rate is mathematically the inverse of N_e, ancestral lineages are naturally drawn toward the population pool where lineages merge most slowly. Here, badgers were assigned the highest N_e in both regions (CdO median 9.4; NAq 21.2), placing them well ahead of cattle (3.8; 5.0) and wild boar (2.7; 2.0). Backward in time, lineages tend to accumulate in this larger deme by model design, which can inadvertently create a forward-in-time "reservoir" effect. This structural tendency is particularly pronounced given the modest wildlife sample sizes (badger n = 52/41, wild boar n = 15/17) and the resulting wide credible intervals. Interestingly, looking past this population size effect reveals alternative signals in the data: in CdO, the most prominent between-species transition was actually cattle→badger, accounting for 95.2% of resolved events. While structured coalescent frameworks like MASCOT are valuable because they reduce standard sampling bias better than discrete migration models, the sharp asymmetry in host N_e here makes deep-node assignments highly sensitive. Consequently, the directional conclusions may reflect the model's underlying assumptions as much as the field epidemiology. Therefore the role of wildlife as a sustaining reservoir is possibly a by product of population scale rather than an independent empirical finding.

Akhmetova et al. (2023)

Akhmetova et al. (2023) investigated the genomic epidemiology of *M. bovis* in sympatric cattle and badger populations within a 100~km^2 high-incidence "hotspot" in Northern Ireland. The study sequenced an exceptional local dataset of 619 *M. bovis* isolates from cattle and badgers, targeting a single endemic lineage (6.263) persistent in the area for some 30–40 years, and additionally microsatellite-genotyped 769 badgers to test whether badger population genetic structure was associated with the spatial distribution of pathogen genetic diversity. Using a structured birth–death model (BDMM) and TransPhylo transmission-tree reconstruction, the authors concluded that cattle were likely driving the local epidemic, with cattle-to-badger transmission more common than badger-to-cattle, and that badger population genetic structure was not associated with how *M. bovis* diversity was partitioned across the landscape, implying that badger-to-badger transmission was not a major driver in this setting.

An Empirical Counter-Weight to Wildlife-Maintenance Narratives. Akhmetova et al. (2023) is a valuable empirical counter-weight within the whole-genome-sequencing literature. In contrast to the wildlife-driven directionality reported from model-dependent frameworks in England (Crispell et al., 2019, Woodchester Park) and France (Duault et al., 2022; Canini et al., 2022), this densely sampled study of a high-prevalence Northern Irish hotspot concluded that the regional epidemic was driven primarily by domestic cattle. Importantly, its directional inference does not rest on an effective-population-size-driven structured coalescent: the cattle-driven reading is supported independently by (i) a markedly higher cattle R_0 (BDMM

median 3.58, 95% HPD 1.59–6.51) than badger R_0 (median 1.01, 95% HPD 1.00–1.03, i.e. barely at the self-sustaining threshold); (ii) a cattle-to-badger BDMM transition rate of 32.37 per lineage per year against a badger-to-cattle rate of 0.04; (iii) TransPhylo transmission pairs that were overwhelmingly cattle-to-cattle (90–97% of high-posterior pairs), with no badger-to-badger pairs recovered in any run (and inter-species pairs rare, 3–10%); and (iv) the absence of any association between badger landscape genetic structure and pathogen genetic structure. The convergence of these independent lines of evidence, rather than a single modelled asymmetry, is what gives the cattle-driven conclusion its weight.

A Reduced, Not Necessarily Null, Role for Badgers. The microsatellite and isolation-by-distance analyses add an independent test seldom available in this literature: were badgers a self-maintaining reservoir, closely related badgers living in proximity should carry closely related *M. bovis*, yet no such association was found. Taken with a badger R_0 at unity and the absence of strongly supported badger-to-badger transmission events, this points to badgers as a comparatively minor, largely spill-over host in this epi-system rather than an engine of persistence. The authors are, however, appropriately cautious, and their caveats should be reproduced rather than elided: the BDMM maximum-clade-credibility tree had low posterior support, the cattle-to-badger transition rate and cattle R_0 were imprecisely estimated, TransPhylo inferred that only $\sim$ 40% of the epidemic had been sampled, and low-diversity, highly clonal pathogens such as *M. bovis* are intrinsically difficult for Bayesian phylodynamic methods. The study zone was also under active culling and vaccination (a Test–Vaccinate–Remove area) and experienced atypically high incidence, so the authors explicitly warn against generalisation.

The Amplification Caveat. The single route by which Akhmetova et al. (2023) leave open a larger role for badgers is a hypothesised "seeding-plus-amplification" effect: that even a small badger-to-cattle contribution might seed herd breakdowns that then amplify through cattle-to-cattle spread. It is important to be clear about the evidentiary status of this suggestion. Within the study it is explicitly a hypothesis, not a result—the authors state that their phylodynamic methods "lack the resolution" to identify such amplification events for a slowly evolving, monomorphic pathogen, so their own data can neither detect nor quantify it. The figure they invoke is not generated here but imported from a single modelling study of the Randomised Badger Culling Trial (Donnelly & Nouvellet, 2013), in which a 5.7% directly attributable badger-to-cattle contribution was *modelled* to amplify to 52%. That amplified estimate carries a bootstrap 95% confidence interval of 9.1–100%—an interval whose upper bound of 100% is effectively unidentified, leaving the data consistent with almost any contribution—and the estimate itself remains contested within the wider, unsettled re-analysis of the trial. The amplification caveat is therefore a model-dependent conjecture resting on an interval that cannot exclude anything, not an empirical demonstration that badgers contribute significantly cattle disease; it marks the limit of what the genomic data can see rather than a finding that offsets the study's central result.

Properly bounded in this way, Akhmetova et al. (2023) demonstrates that the wildlife-to-cattle directionality popularised elsewhere is not a uniform or transferable biological constant but a context-dependent quantity—here plausibly a function of host density, with Woodchester Park carrying 30–40 badgers~km^{-2} against 4–6 in the Northern Irish zone and some of western Europe's highest cattle densities (112 cattle~km^{-2}). In a setting of intense cattle transmission, the data are most parsimoniously explained by badgers as a downstream, spill-over host rather than a maintenance reservoir driving disease in cattle. The study neither claims, nor on these data could establish, that wildlife plays no role whatever; but it provides no support for the proposition that badgers maintain and drive the cattle epidemic, and it directly undercuts any assumption that broad-scale wildlife removal is a self-evident epidemiological imperative where the dominant means of transmission is the cattle network itself.

Table 3: Methodological Risk of Bias Matrix for Core Whole-Genome Sequencing Transmission Inferences

Methodological Domain	**Biek (2012)**	**Crispell (2019)**	**Rossi (2020)**	**Van Tonder (2021)**	**Canini (2022)**	**Duault (2022)**	**Rossi (2022)**	**Akhmetova (2023)**
Host Sampling Balance	Serious ($n = 4$ badgers)	Serious (wider cattle sampling window)	Serious (Differential cattle/badger ascertainment)	Serious (∼40% sampled; per-cluster 0.22–4.38)	Critical (Multi-host asymmetry)	Critical (21 vs 146 isolates)	Serious (Naïve emergent cohort)	Low (Imbalance not load-bearing)
Surveillance Lag Control	Low (Agnostic tracking)	Critical (Unadjusted)	Serious (Sampling-time reliance)	Serious (Weak temporal signal; 5/12 DTR)	Low (MASCOT; sampling-robust)	Low (MASCOT; sampling-robust)	Serious (Sampling-limited detection lag)	Low (Local baseline match)
Environmental State Parameter	Low (Bundled force)	Critical (Omitted)	Critical (Omitted)	Critical (Omitted)	Critical (Omitted)	Critical (Omitted)	Low (Omission not direction-bearing)	Low (Omission not direction-bearing)
Contiguous Cattle Pathway	Low (Model allowed)	Serious (Restricted)	Serious (Pairwise limits)	Low (Movement linked)	Critical (Homogeneous block)	Critical (Homogeneous block)	Serious (Omitted fragmentation)	Low (Mechanistic models)
Directionality Inference Bias	Low (Refused to infer)	Critical (Forced asymmetry)	Serious (Max-probability estimator; bottleneck sensitive)	Serious (Model dependent)	Critical (Ne-weighted; ∼90% unresolved)	Critical (Ne-forced asymmetry)	Low (Direction survives genome-free null)	Low (Data driven)
Overall Risk Rating	**Moderate**	**Critical**	**Serious**	**Serious**	**Critical**	**Critical**	**Moderate**	**Low**

Note: Bias ratings are defined according to modified genomic epidemiology screening protocols. A rating of "Critical" indicates that structural configuration flaws or severe uncorrected operational sampling asymmetries directly generate the study's primary headline transmission conclusions, making the associated inferences unreliable for policy application. Ne denotes the effective population size of a host deme in the structured coalescent; because the coalescent rate is the inverse of Ne, a larger Ne makes lineages coalesce more slowly and biases inferred ancestral assignment toward that deme. Ratings do not credit an undetected-infection compartment: no model represents cattle infected but missed by routine surveillance (infection may remain undetected for ∼15-years, Ghielmetti et al., 2017), so any *directionality* inference resting on sampling dates as infection times is inflated toward badger-to-cattle precedence.

Table 4: Whole-genome-sequencing studies of *M. bovis* transmission at the cattle–badger interface: findings, principal limitation, and bearing on the case for badger intervention.

Study	**Setting & data**	**Headline finding**	**Brief critique**	**Bearing on badger intervention**
Biek et al. (2012)	Five neighbouring farms, Northern Ireland; 31 WGS isolates (26 cattle, 5 from 4 road-kill badgers); mechanistic compartmental (S–E–T–I) model + ML network.	On-farm lineage persistence across clear herd tests; genetic distance tracks spatial proximity better than recorded cattle movements; badger–cattle isolates 0–4 SNPs apart (two indistinguishable), evidencing recent cross-species transmission.	Only four badgers; directionality explicitly not inferred (small sample, slow clock, long infection-to-detection lags), the authors anticipating that more sampling would resolve it; unexplained infections are absorbed into a single, agnostic external force of infection (an unsampled "reservoir" kept in quotes), and the likelihood is approximated rather than fully integrated, as the authors note.	The foundational, appropriately agnostic baseline: shows recent interspecies linkage but declines directional or reservoir claims. Provides no basis for badger removal.
Crispell et al. (2019)	Woodchester Park, Gloucestershire; cattle–badger WGS; structured coalescent (BASTA), two- vs four-deme.	Badger→cattle migration ∼10.4× the reverse rate in the two-deme model.	The 10.4× holds only in the best-supported two-deme model, whose AICM advantage overlapped that of a four-deme model giving near-equal rates. An ancestral-state analysis placing most clade origins in cattle — against the badger-reservoir reading — was dropped by the authors as sampling-bias-sensitive (the editor had asked them to interpret, not remove it), while the equally biased BASTA signal was retained and foregrounded. Directionality is sensitive to model structure, model selection, and selective reporting.	Provides no robust basis for intervention: the badger-to-cattle asymmetry is a property of the two-deme structure and collapses toward parity under four demes, and the host ancestral-state reconstruction was withdrawn at peer review, so no maintenance reservoir driving cattle disease is established.
Rossi et al. (2020)	Woodchester Park, UK; 191 badger / 69 cattle *M. bovis* isolates; pairwise host-to-host transmission inference accounting for within-host SNP accumulation (Kolmogorov forward equations, SEI).	Demonstrates that pairwise probabilities can rank likely vs unlikely transmission links — and flag links more plausibly mediated by an unsampled host — even at low diversity; inferred direction frequently sensitive to within-host (bottleneck) assumptions.	Each pairwise probability is taken as the maximum over 1000 Gaussian-kernel perturbations of the priors (the authors 'selected the combination of parameters which resulted in the highest transmission probability'), biasing estimates upward and removing a common parameter set, so values are not comparable across pairs or directions; a single cattle-derived parameter set is applied to both species; no environmental state; and the bottleneck-sensitivity test was run on clade 3 only (some closely-related clusters robust, others reversed). The authors concede reliance on unbiased sampling and pairwise-dependency effects.	A methods demonstration; neither attempts to, nor can, quantify the cattle-level impact of badger transmission or justify culling.

Study	Setting & data	Headline finding	Brief critique	Bearing on badger intervention
Van Tonder et al. (2021)	Randomised Badger Culling Trial isolates, England; cattle–badger WGS. 690 badger / 752 cattle isolates; 12 transmission clusters.	Overall badger→cattle rate ∼2.1× the reverse (95% HPD 0.8–3.8, crossing unity), but a cluster-level analysis found cattle→badger higher in 4 of 12 clusters.	Aggregate directionality masks substantial cluster-level heterogeneity: only 2/12 clusters were significantly directional and 4/12 ran the other way. Direction rests on a strict-clock, equal-population-size model with ∼40% sampling and a weak temporal signal (clean DTR in 5/12), not on any assumed culling effect.	Mixed, cluster-dependent direction; provides no consistent evidence of badgers as a significant source.
Rossi et al. (2022)	Emergent *M. bovis* outbreak in an area with no previously known wildlife infection (J. Appl. Ecol.).	Outbreak most consistent with a cattle-origin introduction, with wildlife as a spill-over rather than a reservoir.	Broadly supportive of the present argument; the main caveats are the authors' own — sparse early badger sampling, a weak temporal signal in the subsampled BASTA dataset, and a BRT that explains cattle genetic distance only poorly — the small, emergent setting is treated as a spill-over rather than reservoir.	Supports a spill-over-not-reservoir reading; argues against wildlife-driven maintenance.
Duault et al. (2022)	South-west France; cattle–badger WGS; structured coalescent (MASCOT).	Badger→cattle rate ∼52× the reverse; >75% of 1989–2000 lineages reconstructed as badger-borne.	Direction is driven by a badger effective population size ∼28× that of cattle, estimated from only 21 badger sequences; the deep-time badger dominance is extrapolated into a period with no wildlife sampling (none pre-2013); credible intervals are wide and backward migration is not forward transmission.	Directional/reservoir claims are an N_e artefact plus extrapolation; authors concede they could not confirm a farm-to-farm pathway.
Canini et al. (2022)	Two French areas (Côte-d'Or, Nouvelle-Aquitaine); cattle–badger–wild boar WGS; MASCOT + BSSVS.	A cattle→badger→wild boar→cattle circuit, plus badger→cattle in NAq; wild boar proposed as intermediary.	Direction is governed by the badger N_e (largest of the three hosts); ∼90% of transition events in NAq were unresolved; HPDs are wide and the 0.80 selection threshold arbitrary; in CdO the dominant between-species transition was in fact cattle→badger (95.2%).	Authors advise the results be "considered with caution"; the model cannot quantify within-species transmission and gives no support for wildlife maintenance.

Study	**Setting & data**	**Headline finding**	**Brief critique**	**Bearing on badger intervention**
Akhmetova et al. (2023)	100 km^2 hotspot, Northern Ireland; 619 cattle–badger isolates + 769 badger microsatellites; BDMM + TransPhylo.	Cattle-driven epidemic: cattle→badger (BDMM 32.37) ≫ badger→cattle (0.04); badger R_0 ∼1.01; no association between badger genetic structure and pathogen diversity.	Positions badgers as a spill-over host, not a reservoir; directly undercuts a broad wildlife-removal rationale.	Argues directly against badger removal: transmission is overwhelmingly cattle-driven ($c \to b$ 32.37 vs $b \to c$ 0.04) and the badger R_0 (∼1.01) sits at the self-sustaining threshold, so badgers act as a downstream spill-over host dependent on cattle seeding rather than a self-maintaining reservoir.

Across all eight studies the same pattern recurs. The foundational WGS analysis (Biek et al. 2012) establishes recent cattle–badger linkage but holds strictly to the data, declining to infer either direction or a specific reservoir. Where later work reports strong wildlife-to-cattle directionality (Crispell 2019; Duault 2022; Canini 2022), it is a property of the model's structure or of an effective-population-size asymmetry estimated from sparse wildlife samples, is accompanied by wide credible intervals and large unresolved fractions, and is explicitly hedged by the authors. Where transmission is reconstructed at the level of actual events in densely sampled settings (Van Tonder 2021; Akhmetova 2023), the dominant signal is within-cattle transmission with badgers as a downstream, spill-over host. None of these studies quantifies the cattle-level benefit that badger removal would deliver—indeed none attempts to, nor on these data could. Taken together they therefore do not provide sufficient evidence that badgers maintain and drive the bTB epidemic in cattle, and so do not support badger intervention as a means of reducing tuberculosis incidence in cattle.

Several studies acknowledged sampling and model limitations, but few evaluated the full set of competing cattle, badger, environmental, and unsampled transmission pathways within a common sensitivity framework. Consequently, the extent to which reported directional estimates reflected biological transmission rather than analytical assumptions could not be determined consistently. Badger-to-cattle transmission is not ruled out, but the large directional asymmetries reported in several WGS studies may be mathematical artefacts of the models and observation processes, and they do not establish that badgers are the dominant or intervention-relevant driver of cattle bTB.

Discussion

The Epistemological Fragility of Genomic Directionality: A Comparative Synthesis

When viewed horizontally, the eight core whole-genome sequencing (WGS) studies move from providing independent biological consensus to exposing a substantial methodological divergence. There is an apparent conflict in headline directionality estimates. This ranges from the 52-fold badger-to-cattle transition rate in southwest France reported by Duault et al. (2022) and the 10.4-fold asymmetry in Woodchester Park calculated by Crispell et al. (2019), to the two-fold ratio across the English RBCT triplets asserted by Van Tonder et al. (2021) and the complete reversal to cattle-dominant transmission in Northern Ireland (Akhmetova et al. 2023).

Horizontally expanding the core quantitative synthesis layer, Canini et al. (2022) inferred region-specific cross-species transition chains that implicate wildlife reservoirs while highlighting local variation; specifically, their structured coalescent framework identified a direct cattle-to-badger-to-wild boar-to-cattle transmission circuit in Côte-d'Or, along with an additional direct badger-to-cattle transition vector operating in the more fragmented landscape of Nouvelle-Aquitaine. This is not a reflection of shifting regional biology. Instead, these lopsided metrics are mathematical artefacts dictated by localized host sampling densities, mathematical model constraints, and systemic surveillance lag distortions.

As detailed in the companion analysis by Torgerson and Langton (2026), this analytical bias may result directly from a 50-year history of institutional path dependency and empirical misdirection. The foundational structural flaw across the models claiming badger-to-cattle dominance is an uncorrected operational temporal lag between host testing frameworks. Regular domestic herd screening relies on the Single Intradermal Comparative Cervical Tuberculin (SICCT) skin test, which operates with a sensitivity of only 50% to 80% (Nuñez-Garcia et al., 2018). Consequently, *M. bovis* can remain entirely undetected within a domestic herd for years. This allows an infected farm to maintain an Officially Tuberculosis Free status and hence continue to trade cattle and hence horizontal transmission to other herds. Here the Swiss experience is instructive. Switzerland has been officially bTB free for several decades. The most recent outbreak was in 2013. It was shown that it was linked to the previous outbreak in 1998. *M. bovis* had been maintained undetected in cattle for 15 years between the two outbreaks (Ghielmetti et al. 2017).

When a breakdown is finally disclosed bovine isolates are intercepted rapidly via intensified containment measures, appearing on the phylogenetic tree as short, fresh branches. Conversely, wildlife populations

are sampled sporadically via passive road kill or research captures, allowing lineages to circulate slowly in badgers. These long-circulating wildlife strains accumulate a dense array of neutral mutations, entering the tree as deep, long-branched ancestral lineages. When processed by discrete-trait Bayesian phylodynamic or discrete-trait migration pipelines, the algorithm misinterprets this operational surveillance lag as evolutionary distance. (The MASCOT-based analyses are, in fact, insulated from this particular bias; their directional behaviour is governed instead by inferred effective population sizes—see below). To reconcile the long-branched wildlife lineages with the fresh, rapid-turnover cattle clusters, the migration matrix is mathematically forced to inject a severe inflation into the rate *out* of the wildlife state. As early proof-of-concept models explicitly recognized (Biek et al. 2012), the exceptionally slow molecular clock of *M. bovis* ($\leq$ 1 SNP per genome per year) means that direct transmission chains are unresolvable. By substituting code constraints for these unresolvable transmission links and omitting an explicit Environmental Migration State compartment to absorb indirect, pasture-mediated pathways, model-dependent architectures structurally force a false appearance of wildlife reservoir maintenance. Because *M. bovis* is shed in badger faeces and cattle slurry and can persist on pasture and in soil for months (Allen et al. 2021), this collapse of the environmental state — together with the parallel misattribution of long-undetected cattle infection — is illustrated in Figure 4.

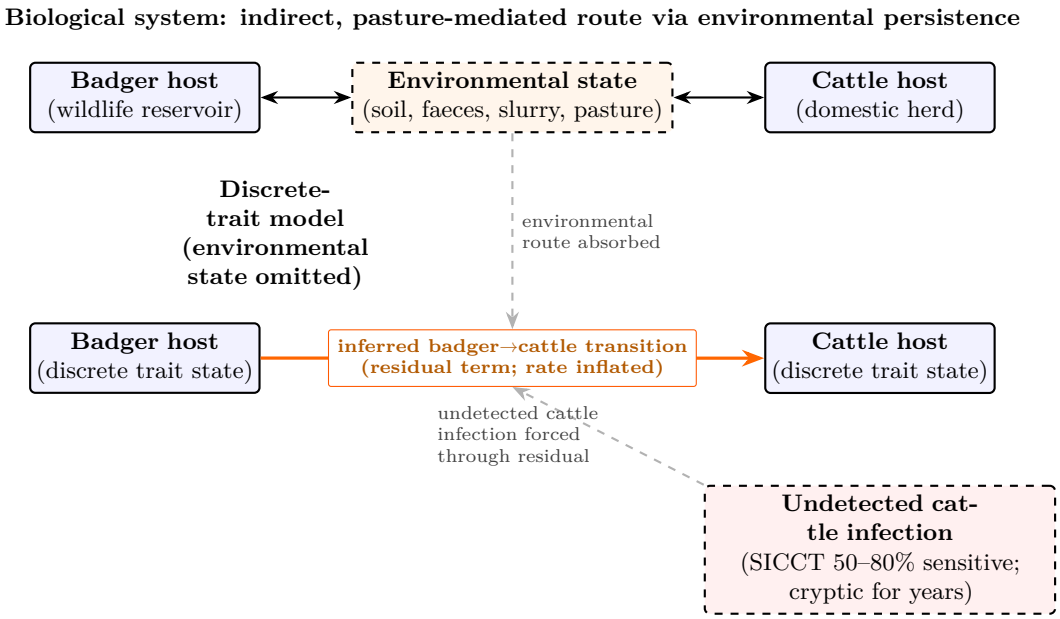

Figure 4: Residual absorption in discrete-trait models. In the biological system (upper) *M. bovis* can pass between badgers and cattle indirectly, via environmental persistence on pasture. A discrete-trait phylodynamic or migration model carries no environmental compartment (lower), so that indirect route has nowhere to reside and is absorbed into the inferred host-to-host (residual) rate. Cattle infections that remain undetected for years under imperfect skin-test surveillance (SICCT sensitivity 50–80%) are likewise not represented as such, and are forced through the same residual term; both inflate the apparent badger-to-cattle transition. Structured-coalescent analyses driven by effective population size (e.g. MASCOT) are insulated from this particular effect.

While traditional PCR-based typing methods like spoligotyping and MIRU-VNTR have historically aided macro-spatial tracking, they offer insufficient resolution at the farm-to-farm or individual host scale due to homoplasy and the highly clonal nature of the pathogen (Guimaraes and Zimpel, 2020), further underscoring the necessity of high-depth whole-genome sequencing to expose localized persistence and indirect, pasture-mediated contact networks.

To bypass the data biases inherent to lopsided wildlife sequencing, alternative frameworks have abandoned discrete-trait host mapping in favour of agent-based simulation approaches. Notably, O'Hare et al. (2021) demonstrated that by combining dense cattle demographic and movement registries with an agnostic, unobserved reservoir compartment, researchers can fit simulated outbreaks directly to empirical cattle SNP distance histograms. Their partial-likelihood model structurally avoids the branch-length illusion by treating the reservoir as a localized, non-diversity generating vector—concluding that while cattle-to-cattle transmission can maintain infection during herd breakdowns ($R_0 > 1$), a spatially constrained hidden reservoir is parsimoniously required to reconcile observed genetic clustering on a local farm scale.

While this simulation approach bypasses direct sequencing biases, its headline conclusion—that infected cattle cannot sustain circulation independently—remains structurally fragile. By restricting livestock parameters to recorded cattle trade paths and failing to account for the large numbers of infected cattle allowed by sub-optimal skin test sensitivity, the model's cattle-only scenario artificially collapses. Consequently, the algorithm is forced to rely on an external reservoir to sustain localized outbreaks, effectively misclassifying unrecorded contiguous cattle pathways (such as fence-line contact and farm fragmentation) as wildlife-mediated transmission vectors. Comparable cryptic pathways have been documented empirically: Orloski et al. (2018) attributed otherwise-unlinked US herd clusters to undocumented cattle movement, importation, and fence-line contact rather than to wildlife. Ultimately, the manuscript's title—claiming to "uncover" the role of an unobserved reservoir—presents a misleading inferential overreach, as the model's assertion that cattle cannot independently sustain circulation is a downstream consequence of structural parameter constraints that ignore hidden domestic latency by design.

Adding a further layer of complexity to traditional typing indices, Trewby et al. (2016) inferred from the phylogenetic distribution of the VNTR types that *M. bovis* can undergo molecular switching within a single evolutionary lineage, causing separate branches to alternate between VNTR-1 and VNTR-10 configurations over extended outbreak timelines. Consequently, surveillance pipelines reliant exclusively on rigid tandem-repeat criteria risk generating severe under-sampling blind spots by systematically misclassifying related outbreak clades due to homoplasy. The converse error is equally telling. Roos et al. (2023) found South African warthog and Kruger National Park wildlife isolates sharing an identical SB0121 spoligotype yet resolving into two whole-genome clades 19–47 SNPs apart, a divergence dating to the estate's establishment rather than to recent transmission. Where Trewby et al. (2016) show one lineage splintering across tandem-repeat types, Roos et al. (2023) show distinct lineages collapsing into one spoligotype; either way, conventional typing overstates transmission links, which is why whole-genome resolution must precede any directional inference.

The Legacy of the Non-Infectious Cattle Paradigm. This mathematical inflation remains deeply entangled with outdated veterinary pathology. Torgerson and Langton (2026) demonstrate that early policy misdirection originated from unverified 1970s pathomorphological assumptions—specifically the Gallagher view inside the Zuckerman (1980) report. This asserted that cattle with early-stage, sub-clinical micro-lesions were non-infectious. This unproven assumption allowed successive generations of field researchers to treat domestic livestock as an inert, background variable and to disproportionately externalize primary disease drivers to wildlife.

This historical paradigm directly underpins the convention of modern phylodynamic pipelines filtering out unconfirmed Officially Tuberculosis Free-Suspended (OTF-S) reactor herds, restricting active datasets exclusively to macroscopic, visible-lesion (OTF-W) cohorts. By excluding the early-stage infection pressure that drives horizontal transmission, studies such as Duault et al. (2022) and Crispell et al. (2019) trap themselves within a circular calibration loop. They calibrate their models to historical trials like the RBCT or the Irish Four Areas Project (FAP), whose claimed culling effects are themselves shown by Torgerson and Langton (2026) to be statistical artefacts of unadjusted multiplicity inflation, model overfitting, and the omission of

unconfirmed breakdowns.

The Concluding Resolution of the Evidence. When local host data balance and dense sampling are actually achieved in the field, as in the 100~km^2 high-prevalence hotspot audited by Akhmetova et al. (2023), the ancestral state artefacts disappear. When the data are representative, the networks reveal that the epidemic is likely to be overwhelmingly cattle-driven. The collective whole-genome sequencing literature provides little evidence of a wildlife reservoir driver. Instead, as argued by Torgerson and Langton (2026), it demonstrates that the badger functions as an incidental, localized spillover host fed continuously by hidden diagnostic test failures and uncorrected livestock trading mechanics internal to the domestic agricultural network.

Methodological Artifacts as Causal Inferences: Deconstructing the Wildlife Paradigm

A comprehensive, epidemiological appraisal of the whole-genome sequencing (WGS) evidence base demonstrates that the headline assertions of significant, unidirectional *M. bovis* transmission from European badgers to domestic cattle are epidemiologically unsupported and structurally compromised. Proponents of broad-scale wildlife interventions frequently cite high-profile phylodynamic models as independent, cumulative biological proof of a wildlife maintenance reservoir (Godfray et al. 2025). However, a horizontal synthesis of the eight core studies reveals that none provide unconfounded or methodologically sound evidence of significant wildlife-to-bovine transmission. Instead, their headline directionality fractions are downstream artefacts of unvalidated software assumptions, severe host tracking imbalances, and explicit code constraints that substitute computational shortcuts for biologically unresolvable field transmission links.

This systematic erosion of inferential reliability can be traced across three areas of model misspecification:

1. The Surveillance Lag Paradox (The Branch-Length Illusion). The massive transition asymmetries calculated across model-dependent architectures—such as the 10.4-fold ratio in Crispell et al. (2019) and the 52-fold rate in Duault et al (2022) —are driven by an uncorrected tracking imbalance inherent to national veterinary operations. Domestic cattle are subjected to regular, mandatory, and intensive skin and blood testing regimes, meaning infected individuals can be intercepted at an early stage of pathogenesis. On a phylogenetic tree, these bovine isolates register as short, fresh branches. Conversely, wildlife populations are sampled sporadically and passively (predominantly via roadkill or restricted research zones), allowing lineages to circulate silently in badgers for multi-year intervals before isolation. These wildlife strains consequently accumulate a dense array of neutral mutations, entering the tree as deep, long-branched ancestral lineages.

When processed by discrete-trait Bayesian phylodynamic models or structured coalescent matrices (such as MASCOT), the software misinterprets this operational surveillance lag as evolutionary distance. To reconcile the long-branched wildlife lineages with the fresh, rapid-turnover cattle isolates, the migration matrix is mathematically forced to systematically inflate the migration rate *out* of the wildlife state. As early proof-of-concept models explicitly recognized (Biek et al., 2012), the exceptionally slow molecular clock of *M. bovis* ($\leq$ 1 SNP per genome per year) means that identical strains routinely circulate horizontally within high-density herds for multiple seasons without mutating, rendering direct transmission chains unresolvable. The calculated asymmetry is a mathematical artefact of unbalanced tracking windows, not evidence of real biological directionality.

In BASTA/discrete-trait pipelines the bias enters through tip-sampling imbalance; in MASCOT (Duault et al. 2022) the host sample counts are deliberately not treated as data, so the directional bias instead enters through the inferred effective population sizes—here a badger N_e some 28$\times$ the cattle N_e, from 21 badger sequences—which pulls ancestral lineages into the badger deme and forces forward-time flow out of it. The same lever operates in Canini et al. (2022), where the badger N_e was again largest of the three hosts (median 9.4 in Côte-d'Or and 21.2 in Nouvelle-Aquitaine, against cattle 3.8/5.0 and wild boar 2.7/2.0)—so much of the wild boar-intermediary circuit is likewise a demographic-prior effect rather than an observed flux, and in Côte-d'Or the dominant between-species transition was in fact cattle$\rightarrow$badger (95.2%).

A documented Swiss re-emergence shows the scale of this hidden-reservoir problem directly: after fifteen years in which bTB was considered absent, the 2013 *M. bovis* SB0120 outbreak was traced to an archival isolate from a cow slaughtered in 1998 carrying an identical MIRU-VNTR profile, even though the source herd had passed repeated intradermal skin testing in the interval (Ghielmetti et al., 2017). Because the routine tuberculin test detects only 50–80% of infected animals, a lineage can circulate in cattle for a decade or more before it surfaces, so the sampling (detection) date that anchors every time-calibrated phylogeny here is not a reliable proxy for the time of infection. Where this detection lag is longer and heavier-tailed in cattle than in the wildlife sampling frame, cattle tips are placed too recently and ancestral lineages are displaced toward badgers, manufacturing an apparent badger-to-cattle precedence—and a spurious absence of cattle-to-badger events—in precisely those studies (Crispell et al., 2019; and, as a compounding confounder, Van Tonder et al., 2021) whose headline runs in that direction. The same bias cannot inflate the cattle-to-badger conclusions of Rossi et al. (2022) and Akhmetova et al. (2023), which therefore remain unaffected by this critique.

This is not merely an external critique: the same caveat is volunteered by investigators applying the method. In a well-resourced WGS investigation of an *M. bovis* outbreak in beef cattle and white-tailed deer in Minnesota, Glaser et al. (2016) cautioned that "the collection date does not likely correlate to infection status of either active replication or latent state," conceding that *M. bovis* latency makes the isolate sampling date "a source of error in the Bayesian molecular clock analysis." That a non-badger outbreak team independently reaches this conclusion confirms that the mismatch between detection date and infection date is a general property of *M. bovis* epidemiology rather than an artefact peculiar to the British and Irish badger–cattle studies.

2. The Environmental Vacuum and Forced Transitioning. A major deterministic flaw inside the primary mathematical pipelines is the reduction of a multi-host, environmentally connected ecosystem into a simple, binary host-to-host interaction matrix. Because *M. bovis* bacilli can survive on grass, within soil, and inside stored communal farm slurry setups for multiple months, a shared 0-SNP identical profile between a sympatric cow and badger may be mediated by indirect environmental persistence.

However, the mathematical pipelines of Crispell et al (2019), Van Tonder et al. (2021) and Duault et al. (2022) completely omit an explicit Environmental Migration State compartment from their equations. Because the model's code boundaries lack a physical space to absorb these indirect pathways, the algorithm is structurally forced to misclassify every instance of pasture-mediated contamination as a direct, face-to-face cross-species spillover event. This structural forcing tilts the causal attribution heavily toward wildlife, rendering the study's directionality fractions highly speculative and epidemiologically unreliable.

3. Homogeneous Trait Boxing and Omission of Contiguous Cattle Routes. To execute complex phylodynamic equations under data-sparse conditions, models like Duault et al. (2022) simplify the entire livestock population into a single well-mixed, homogeneous trait block. This setup completely ignores local contiguous cattle-to-cattle transmission pathways, such as fence-line nose-to-nose contact, shared agricultural equipment, or undetected movements across fragmented land parcels. Because no local horizontal livestock pathway exists within the model, any regional transmission that slips past the formal cattle movement registry is automatically assigned as badger mediated. The finding that badgers maintain regional infection is an inevitable consequence of a closed model structure that made alternative cattle-focused routes impossible by design.

This attribution bias is institutional as much as analytical: where direct badger infection data are too sparse to interrogate, cattle surveillance signatures — multi-year persistence, a predominance of confirmed (OTF-W) over unconfirmed incidents, and low cattle-movement attribution — are themselves adopted as the operational definition of a "badger-associated" reservoir (Downs et al., 2022). However transparent and pragmatically justified, such a definition embeds the badger-source hypothesis in the measurement, so a pattern it generates cannot subsequently be read as independent confirmation of that hypothesis.

Expanding the wildlife-reservoir critique to alternative sympatric hosts, Crispell et al. (2020) utilized descriptive genomics to demonstrate that cattle and wild sika deer share highly similar *M. bovis* strains within

localized hotspots like County Wicklow in Ireland. However, their downstream hypothesis—that high genomic diversity within the deer cohort implicates them as an active maintenance host rather than a passive spillover sink— is fundamentally constrained by severe sampling design imbalances. The data pipeline evaluated an extensive cross-sectional wildlife cull while restricting livestock data to a single representative isolate per herd, a structural asymmetry that mathematically inflates observed wildlife diversity metrics relative to domestic hosts. Furthermore, due to a severe biobank recovery failure that left the local badger population completely unrepresented ($n = 2$), the descriptive framework lacks the epidemiological parameter depth to rule out hidden multi-host intermediaries, illustrating that descriptive wildlife-source inferences remain highly sensitive to imbalanced host coverage. Informatively, where the same ancestral-state methods are applied by an independent group to a non-badger multi-host system, the direction reverses: Pereira et al. (2024b) dated the *M. bovis* Eu3 most recent common ancestor in Portugal to cattle, with onward transitions to red deer and wild boar rather than the reverse — the cattle-origin, spill-to-wildlife pattern that surveillance-lag bias tends to mask in the badger–cattle studies.

The same inferential ceiling is visible well outside the endemic badger–cattle interface. Kohl et al. (2018) applied whole-genome sequencing, alongside spoligotyping and MIRU-VNTR typing, to a protracted *M. bovis* outbreak in a German wildlife park and a captive deer farm some 100 km distant — a multi-host system of red, fallow and sika deer, bison, lynx and free-roaming pot-bellied pigs, but one containing neither badgers nor cattle. A single strain was shown to have persisted from 2002 to 2015, with a maximum of twelve SNPs separating any two isolates and most grouping within five, and the genomic data confirmed epidemiological linkage between the two facilities. Critically, the analysis establishes relatedness and long-term persistence; it does not, and does not attempt to, reconstruct a direction of transmission between the host species involved. This is instructive on two counts. First, it is a clean demonstration of the descriptive-genomics ceiling formalised in Figure 3: even rigorous, densely sampled phylogenetic WGS yields host sharing and persistence rather than directionality unless an explicit state-transition model is imposed. Second, a single lineage circulating undetected across a captive multi-host population for well over a decade reinforces the central caution of this review — that genomic proximity and shared strains index persistence and common exposure, not recent directional transmission, and a reminder that long cryptic persistence makes shared strains a marker of common exposure rather than recent directional spread (cf. Ghielmetti et al., 2017).

Acknowledging the inferential dangers of forcing definitive tree topologies on incomplete data, Wood et al. (2024) deployed a pairwise numerical integration framework to calculate relative transmission likelihoods across a sparse historical dataset from Woodchester Park. Rather than asserting explicit transmission vectors, their model operates as a exclusionary filter to prune epidemiologically impossible connections. However, a machine-learning Boruta sensitivity audit within their pipeline uncovers that the resulting likelihood configurations are overwhelmingly driven by the chosen pathogen substitution rate prior (μ). Because this molecular clock remains highly volatile and slow, the framework's filtering power is fundamentally limited by unprovable evolutionary parameters. This constraint is severely exacerbated by a profound sampling deficit, where their sequenced cohorts captured fewer than 5% of the 5,852 confirmed herd infections circulating within the local radius, ensuring that the remaining plausible pathways are highly uncertain due to a massive unobserved data void.

Evidence and advice: the science–policy interface in bovine TB

These appraisals also raise a structural feature of the evidence base that deserves comment at the science–policy interface. Several of the studies weighed here were produced by researchers who simultaneously hold formal government advisory roles on the question the evidence informs. Kao, for example, is senior (last) author and a corresponding author on Crispell *et al.* (2019), and Rossi et al. (2020, 2022); three of the genomic transmission studies appraised above. He has held successive bovine-TB policy-advisory positions: member of the Defra Science Advisory Council (2018–2024), chair of its Exotic and Emerging Diseases subgroup (2021–2025), member of the Defra Bovine TB Partnership, and Chair of the Science Advisory Council from January 2025. This dual position is declared openly. It is stated verbatim in the conflict-of-interest declaration of Rossi *et al.* (2022). The point here is that the same individuals both generate the primary transmission evidence and advise on how that evidence should shape culling policy, which concentrates responsibility for the evidence base in a small number of hands and makes independent

replication and adversarial scrutiny especially important.

The Crispell *et al.* revision history (2019) illustrates why that scrutiny matters, and why it should rest on the documented record. As set out in the appraisal above and in the journal's published peer-review correspondence, a phylogenetic ancestral-character estimation in the first submission placed the ancestral nodes of most inter-species clades in cattle; a result the Reviewing Editor flagged as "seemingly inconsistent with the badger-reservoir hypothesis" and asked the authors to reconcile rather than remove. The authors, under the senior author's responsibility, removed the analysis instead, citing its sensitivity to sampling bias. That decision has a defensible methodological rationale and is transparently on the public record; it is nonetheless an author-initiated choice to delete, rather than reconcile, the one analysis pointing away from the study's headline reading. The appropriate response to a concentrated and policy-connected evidence base is therefore procedural; for example routine publication of sensitivity analyses and independent re-analysis.

A Concluding Synthesis

When horizontally synthesized across epidemiological contexts, the WGS literature reveals a profound divergence between raw molecular data and model-dependent interpretations. The popularized paradigm designating the European badger as the primary driver of bTB maintenance is built upon a highly fragile computational foundation. As demonstrated across the core quantitative synthesis tier, high-precision transmission asymmetries are consistently exposed as mathematical artifacts rather than independent biological insights. Whether operating as a 52-fold badger-to-cattle transition dominance in endemic France (Duault et al., 2022) or an inverse cattle-dominant vector in emergent Cumbria (Rossi et al., 2022), these lopsided migration fractions are fundamentally dictated by uncorrected field surveillance lags, code constraints, and the structural forcing of the branch-length illusion. This surveillance lag is not merely quantitative but temporal: because the routine skin test misses 50–80% of infected cattle and a single lineage can persist cryptically in a herd for up to fifteen years (Ghielmetti et al., 2017), the sampling date that calibrates each molecular clock is not the date of infection, so a strain recovered from badgers before cattle need not have arisen in badgers first. Directionality read from relative tip timing is therefore exposed to inversion, not just to widened uncertainty, in every model that omits an undetected cattle compartment.

This inferential fragility is systematically unified by two fatal areas of model misspecification. First, by compressing complex, multi-host systems into closed, binary mammalian interaction matrices, frameworks like Crispell et al. (2019) and Rossi et al. (2020) operate within an environmental vacuum. Lacking a physical compartment to absorb off-host survival and indirect pasture-mediated contamination—proven to be highly intertwined by environmental sequencing (Pereira et al., 2024a)—algorithms are structurally forced to misclassify field-mediated exposure as direct face-to-face cross-species spillovers. Second, by treating the domestic herd population as a single, well-mixed homogeneous trait box, these models completely ignore unrecorded livestock mechanics. Crucially, the mathematical necessity of an external reservoir asserted by agent-based frameworks (O'Hare et al., 2021) collapses when parameters are adjusted to account for the massive latent infection pressure and horizontal trade circulation allowed by sub-optimal diagnostic test sensitivity.

Ultimately, when local host data balance and high-density sampling are actually achieved in the field, the ancestral state artefacts disappear entirely. As demonstrated by the exceptionally dense, sympatric validation grid of Akhmetova et al. (2023), representative data networks consistently uncover an overwhelmingly cattle-driven epidemic. The collective WGS evidence base supports a decisive paradigm shift away from wildlife-reservoir maintenance hypotheses. The data suggests that badgers are most likely an incidental spillover host. This is facilitated by hidden diagnostic test failures, unrecorded contiguous contact, and uncorrected livestock trading mechanics internal to the domestic agricultural network. Future animal health resources must abandon broad-scale wildlife interventions and focus on improving domestic diagnostic sensitivity, movement control, biosecurity and herd level disease prevention.

References

Akhmetova, A., Guerrero, J., McAdam, P., Salvador, L.C., Crispell, J., Lavery, J., Presho, E., Kao, R.R., Biek, R., Menzies, F. and Trimble, N., 2023. Genomic epidemiology of *Mycobacterium bovis* infection in sympatric badger and cattle populations in Northern Ireland. *Microbial Genomics*, 9(5), p.001023.

Akhmetova, A., Salvador, L.C., Oravcova, K., Presho, E., McComick, C., Thompson, S., Wright, L., Menzies, F., Trimble, N., Harwood, R. and Kao, R.R., 2026. Scale-and genotype-dependent spatial concordance of *Mycobacterium bovis* in cattle and badgers. *bioRxiv*, pp.2026-10.

Allen, A.R., Ford, T. and Skuce, R.A., 2021. Does *Mycobacterium tuberculosis* var. bovis survival in the environment confound bovine tuberculosis control and eradication? A literature review. *Veterinary Medicine International*, 2021(1), p.8812898.

Allen, A., Magee, R., Devaney, R., Ardis, T., McNally, C., McCormick, C., Presho, E., Doyle, M., Ranasinghe, P., Johnston, P. and Kirke, R., 2024. Whole-Genome sequencing in routine *Mycobacterium bovis* epidemiology–scoping the potential. *Microbial Genomics*, 10(2), p.001185.

Biek, R., O'Hare, A., Wright, D., Mallon, T., McCormick, C., Orton, R.J., McDowell, S., Trewby, H., Skuce, R.A. and Kao, R.R., 2012. Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS pathogen*, 8(11), p.e1003008.

Canini, L., Modenesi, G., Courcoul, A., Boschiroli, M.L., Durand, B. and Michelet, L., 2022. Deciphering the role of host species for two *Mycobacterium bovis* genotypes from the European 3 clonal complex circulation within a cattle-badger-wild boar multihost system. *MicrobiologyOpen*, 12(1), p.e1331.

Crispell J, Zadoks RN, Harris SR, Paterson B, Collins DM, de-Lisle GW, Livingstone P, Neill MA, Biek R, Lycett SJ, Kao RR, Price-Carter M., 2017. Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. *BMC Genomics*. 2017 Feb 16;18(1):180. doi: 10.1186/s12864-017-3569-x.

Crispell, J., Benton, C.H., Balaz, D., De Maio, N., Ahkmetova, A., Allen, A., Biek, R., Presho, E.L., Dale, J., Hewinson, G. and Lycett, S.J., Nunez-Garcia, J., Skuce, R. A., Trewby, H., Wilson, D. J., Zadoks, R. N., Delahay, R.J., Kao, R. R. 2019. Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. *elife*, 8, p.e45833.

Crispell J, et al. revision history, 2019. Peer review of "Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis in cattle and badgers*." eLife. Available from: https://elifesciences.org/articles/45833/peer-reviews

Crispell J, Cassidy S, Kenny K, McGrath G, Warde S, Cameron H, Rossi G, MacWhite T, White PCL, Lycett S, Kao RR, Moriarty J, Gordon SV., 2020. *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. *Microb Genom*. 2020 Aug;6(8):mgen000388. doi: 10.1099/mgen.0.000388.

Diel, R., Kohl, T.A., Maurer, F.P., Merker, M., Walter, K.M., Hannemann, J., Nienhaus, A., Supply, P. and Niemann, S., 2019. Accuracy of whole-genome sequencing to determine recent tuberculosis transmission: an 11-year population-based study in Hamburg, Germany. European Respiratory Journal, 54(5).

Dippenaar, A., Parsons, S.D.C., Miller, M.A., Hlokwe, T., van Pittius, N.C.G., Adroub, S.A., Abdallah, A.M., Pain, A., Warren, R.M., Michel, A.L. and Van Helden, P.D., 2017. Progenitor strain introduction of *Mycobacterium bovis* at the wildlife-livestock interface can lead to clonal expansion of the disease in a single ecosystem. *Infection, Genetics and Evolution*, 51, pp.235-238.

Donnelly, C.A., Nouvellet, P., 2013. Evaluating the impact of badger management on bovine tuberculosis in Great Britain. *PLoS Currents: Outbreaks* 1, doi:10.1371/currents.outbreaks.

Downs, S.H., Ashfield, S., Arnold, M., Roberts, T., Prosser, A., Robertson, A., Frost, S., Harris, K., Avigad, R. and Smith, G.C., 2022. Detection of a local *Mycobacterium bovis* reservoir using cattle surveillance data. *Transboundary and Emerging Diseases*, 69(4), pp.e104-e118.

Duault, H., Michelet, L., Boschiroli, M.L., Durand, B. and Canini, L., 2022. A Bayesian evolutionary model towards understanding wildlife contribution to F4-family *Mycobacterium bovis* transmission in the South-West of France. *Veterinary Research*, 53(1), p.28.

Ghielmetti, G., Scherrer, S., Friedel, U., Frei, D., Suter, D., Perler, L. and Wittenbrink, M.M., 2017. Epidemiological tracing of bovine tuberculosis in Switzerland, multilocus variable number of tandem repeat analysis of *Mycobacterium bovis* and *Mycobacterium caprae*. *PLoS One*, 12(2), p.e0172474.

Glaser, L., Carstensen, M., Shaw, S., Robbe-Austerman, S., Wunschmann, A., Grear, D., Stuber, T. and Thomsen, B., 2016. Descriptive epidemiology and whole genome sequencing analysis for an outbreak of bovine tuberculosis in beef cattle and white-tailed deer in northwestern Minnesota. *PloS One*, 11(1), p.e0145735.

Godfray, C., Hewinson, G., Silverman, B., Winter, M., Wood, J., 2025. Bovine TB strategy review update 2025. Department for Environment, Food & Rural Affairs, London. Available online at: https://assets.publishing.service.gov.uk/media/68d563169ce370a7e0a0fcfe/Godfray_bTB_Evidence_Review_Update_2025.pdf.

Guimaraes, A.M. and Zimpel, C.K., 2020. *Mycobacterium bovis*: from genotyping to genome sequencing. *Microorganisms*, 8(5), p.667.

Harvey, N., McGrath, G., O'Shaughnessy, J., Tratalos, J.A., Byrne, B., Allen, A., Skuce, R., Kenny, K., Gordon, S.V. and Farrell, D., 2025. Integrating whole-genome sequencing and epidemiology to characterise *Mycobacterium bovis* transmission in Ireland: a proof of concept. *Irish Veterinary Journal*, 79(1), p.3.

Healey, A. J. E., Williams, C. L., Dimonaco, N. J., Galloway, T., Ellis, R. J., Palkopoulou, E., Strong, J., Gibson, A. J., Hewinson, R. G., & Friedersdorff, J. C. A. 2026. Uncovering the genomic landscape of *Mycobacterium bovis* in Wales. *Scientific Reports*, 16, 28546.

Kohl, T.A., Utpatel, C., Niemann, S. and Moser, I., 2018. *Mycobacterium bovis* persistence in two different captive wild animal populations in Germany: a longitudinal molecular epidemiological study revealing pathogen transmission by whole-genome sequencing. *Journal of Clinical Microbiology*, 56(9), pp.10-1128.

Legall, N. and Salvador, L.C., 2022. Selective sweep sites and SNP dense regions differentiate *Mycobacterium bovis* isolates across scales. *Frontiers in Microbiology*, 13, p.787856.

Menardo, F., Duchêne, S., Brites, D. and Gagneux, S., 2019. The molecular clock of *Mycobacterium tuberculosis*. *PLoS Pathogens*, 15(9), p.e1008067.

Michelet, L., Conde, C., Branger, M., Cochard, T., Biet, F. and Boschiroli, M.L., 2019. Transmission network of deer-borne *Mycobacterium bovis* infection revealed by a WGS approach. *Microorganisms*, 7(12), p.687.

Nuñez-Garcia, J., Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron, A.R., Rolfe, S., Meyer, R.A.A., Clifton-Hadley, R.S., Greiner, M., 2018. Meta-analyses of the sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland. *Preventative Veterinary Medicine* 153, 94–107. doi:10.1016/j.prevetmed.2017.02.017

O'Hare A, Balaz D, Wright DM, McCormick C, McDowell S, Trewby H, et al.,2021. A new phylodynamic model of *Mycobacterium bovis* transmission in a multi-host system uncovers the role of the unobserved reservoir. *PLoS Comput Biol* 17(6): e1009005. https://doi.org/10.1371/journal.pcbi.1009005

Orloski, K., Robbe-Austerman, S., Stuber, T., Hench, B. and Schoenbaum, M., 2018. Whole genome sequencing of *Mycobacterium bovis* isolated from livestock in the United States, 1989–2018. *Frontiers in Veterinary Science*, 5, p.253.

O'Shaughnessy, J., Harvey, N., Byrne, B., McElroy, M., Gutierrez, M., Murphy, D., Kenny, K., Cameron, H., Prendergast, D., Cupial, R., Goggin, M., Dygico, L. K., Smith, J., Tratalos, J. A., Devaney, R., Ranasinghe, P., Ardis, T., Allen, A., McGrath, G., Gordon, S. V., & Farrell, D. 2026. The genomic diversity and spatial patterns of *Mycobacterium bovis* in Ireland revealed by whole genome sequencing. *Irish Veterinary Journal*, 79(1), 6–19.

Perea, C., Ciaravino, G., Stuber, T., Thacker, T.C., Robbe-Austerman, S., Allepuz, A. and Val, B.P.D., 2021. Whole-genome SNP analysis identifies putative *Mycobacterium bovis* transmission clusters in livestock and wildlife in Catalonia, Spain. *Microorganisms*, 9(8), p.1629.

Pereira, A.C., Reis, A.C. and Cunha, M.V., 2023. Genomic epidemiology sheds light on the emergence and spread of *Mycobacterium bovis* Eu2 Clonal Complex in Portugal. *Emerging Microbes & Infections*, 12(2), p.2253340.

Pereira, A.C., Pinto, D. and Cunha, M.V., 2024a. First time whole genome sequencing of *Mycobacterium bovis* from the environment supports transmission at the animal-environment interface. *Journal of Hazardous Materials*, 472, p.134473.

Pereira, A.C., Lourenço, J., Themudo, G., Botelho, A. and Cunha, M.V., 2024b. Population structure and history of *Mycobacterium bovis* European 3 clonal complex reveal transmission across ecological corridors of unrecognized importance in Portugal. *Microbiology spectrum*, 12(7), pp.e03829-23.

Pozo, P., Lorente-Leal, V., Robbe-Austerman, S., Hicks, J., Stuber, T., Bezos, J., de Juan, L., Saez, J.L., Romero, B., Alvarez, J. and Spanish Network on Surveillance Monitoring of Animal Tuberculosis, 2022. Use of whole-genome sequencing to unravel the genetic diversity of a prevalent *Mycobacterium bovis* spoligotype in a multi-host scenario in Spain. *Frontiers in Microbiology*, 13, p.915843.

Roos, E.O., Loubser, J., Kerr, T.J., Dippenaar, A., Streicher, E., Olea-Popelka, F., Robbe-Austerman, S., Stuber, T., Buss, P., de Klerk-Lorist, L.M. and Warren, R.M., 2023. Whole genome sequencing improves the discrimination between *Mycobacterium bovis* strains on the southern border of Kruger National Park, South Africa. *One Health*, 17, p.100654.

Rossi G, Crispell J, Balaz D, Lycett SJ, Benton CH, Delahay RJ, Kao RR., 2020. Identifying likely transmissions in *Mycobacterium bovis* infected populations of cattle and badgers using the Kolmogorov Forward Equations. *Scientific Reports*. 15;10(1):21980. doi: 10.1038/s41598-020-78900-3.

Rossi, G., Crispell, J., Brough, T., Lycett, S. J., White, P. C. L., Allen, A., Ellis, R. J., Gordon, S. V., Harwood, R., Palkopoulou, E., Presho, E. L., Skuce, R., Smith, G. C., & Kao, R. R., 2022. Phylodynamic analysis of an emergent *Mycobacterium bovis* outbreak in an area with no previously known wildlife infections. *Journal of Applied Ecology*, 59, 210–222.

Salvador, L.C., O'Brien, D.J., Cosgrove, M.K., Stuber, T.P., Schooley, A.M., Crispell, J., Church, S.V., Gröhn, Y.T., Robbe-Austerman, S. and Kao, R.R., 2019. Disease management at the wildlife-livestock interface: Using whole-genome sequencing to study the role of elk in *Mycobacterium bovis* transmission in Michigan, USA. *Molecular Ecology*, 28(9), pp.2192-2205.

Sandhu, P., Nunez-Garcia, J., Berg, S., Wheeler, J., Dale, J., Upton, P., Gibbens, J., Hewinson, R.G., Downs, S.H., Ellis, R.J. and Palkopoulou, E., 2025. Enhanced analysis of the genomic diversity of *Mycobacterium bovis* in Great Britain to aid control of bovine tuberculosis. *Frontiers in Microbiology*, 16, p.1515906.

Scaltriti, E., Iyad, K., Boniotti, M.B., Menozzi, I., Bolzoni, L., Ippolito, D., Ciarello, F.P., Loda, D., D'Incau, M., Zanoni, M. and Presti, V.D.M.L., 2024. Inside *Mycobacterium bovis* SB0120 spoligotype circulating in Italy: analysis of the most frequent genotypes by whole genome sequencing. *Frontiers in Microbiology*, 15, p.1416605.

Torgerson, P. R. and Langton, T. E. S., 2026. Badger Interventions and Bovine Tuberculosis Control in Britain and Ireland: A Systematic Review of Field Studies and Transmission Models. (preprint) https://dx.doi.org/10.2139/ssrn.7639878

Trewby H, Wright D, Breadon EL, Lycett SJ, Mallon TR, McCormick C, Johnson P, Orton RJ, Allen AR, Galbraith J, Herzyk P, Skuce RA, Biek R, Kao RR, 2016. Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. *Epidemics*. 14:26-35. doi: 10.1016/j.epidem.2015.08.003.

Van Tonder, A.J., Thornton, M.J., Conlan, A.J., Jolley, K.A., Goolding, L., Mitchell, A.P., Dale, J., Palkopoulou, E., Hogarth, P.J., Hewinson, R.G. and Wood, J.L., 2021. Inferring *Mycobacterium bo-*

vis transmission between cattle and badgers using isolates from the Randomised Badger Culling Trial. *PLoS Pathogens*, 17(11), p.e1010075.

Walker, T.M., Ip, C.L., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W. and Parkhill, J., 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet Infectious Diseases*, 13(2), pp.137-146.

Wood AJ, Benton CH, Delahay RJ, Marion G, Palkopoulou E, Pooley CM, Smith GC, Kao RR. The utility of whole-genome sequencing to identify likely transmission pairs for pathogens with slow and variable evolution. *Epidemics*. 2024 Sep;48:100787. doi: 10.1016/j.epidem.2024.100787.

Wyllie, D.H., Davidson, J.A., Smith, E.G., Rathod, P., Crook, D.W., Peto, T.E., Robinson, E., Walker, T. and Campbell, C., 2018. A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying Mycobacterium tuberculosis transmission: a prospective observational cohort study. EBioMedicine, 34, pp.122-130.

Zuckerman, S., 1980. *Badgers, Cattle and Tuberculosis: A Report to the Minister of Agriculture, Fisheries and Food*. Her Majesty's Stationery Office (HMSO), London.