Title

Phylodynamic analysis of an emergent *Mycobacterium bovis* outbreak in an area with no previously known wildlife infections

Authors

Gianluigi Rossi[1], Joseph Crispell[2], Tanis Brough[3], Samantha J. Lycett[1], Piran C. L. White[4], Adrian Allen[5], Richard J. Ellis[6], Stephen V. Gordon[2,7], Roland Harwood[8], Eleftheria Palkopoulou[9], Eleanor L. Presho[5], Robin Skuce[5], Graham C. Smith[10], and Rowland R. Kao[1*]

[1] Roslin Institute and Royal (Dick) School of veterinary Studies, University of Edinburgh, EH25 9RG, UK

[2] School of Veterinary Medicine, University College Dublin, Dublin D04, Ireland

[3] Advice Services Team, Service Delivery Directorate, APHA, Merrythought, Calthwaite, Penrith, Cumbria, CA11 9RR, UK

[4] Department of Environment and Geography, University of York, Wentworth Way, York YO10 5NG, UK

[5] Bacteriology Branch, Veterinary Sciences Division, Agri-food and Biosciences Institute, Belfast, BT4 3SD, UK

[6] Surveillance and Laboratory Services Department, APHA Weybridge, New Haw, Surrey, KT15 3NB, UK

[7] UCD Conway Institute, University College Dublin, Dublin D04, Ireland

[8] Veterinary Service, Department of Agriculture, Environment and Rural Affairs, Northern Ireland, UK

[9] Department of Bacteriology, APHA, Weybridge, New Haw, Surrey, KT15 3NB, UK

[10] National Wildlife Management Centre, APHA, Sand Hutton, York, YO41 1LZ, UK

*Corresponding author: Rowland.Kao@ed.ac.uk

Abstract

Understanding how an emergent pathogen successfully establishes itself and persists in a previously unaffected population is a crucial problem in disease ecology. In multi-host pathogen systems this problem is particularly difficult, as the importance of each host species to transmission is often poorly characterised, and the epidemiology of the disease is complex. Opportunities to observe and analyse such emergent scenarios are few.

Here, we exploit a unique dataset combining densely-collected data on the epidemiological and evolutionary characteristics of an outbreak of *Mycobacterium bovis* (*M. bovis,* the causative agent of bovine tuberculosis, bTB) in a population of cattle and badgers in an area considered low-risk for bTB, that has no previous record of either persistent infection in cattle, or of any infection in wildlife.

We analyse the outbreak dynamics using a combination of mathematical modelling, machine learning and Bayesian evolutionary analyses. Comparison to *M. bovis* whole-genome sequences from Northern Ireland confirmed this to be a single introduction of the pathogen from the latter region, with evolutionary analysis supporting an introduction directly into the local cattle population at least six years prior to its first discovery in badgers. Once introduced, the evidence supports *M. bovis* epidemiological dynamics passing through two phases, the first dominated by cattle-to-cattle transmission before becoming established in the local badger population.

These findings emphasise the importance of disease surveillance for early containment of outbreaks, in particular for pathogens not causing immediately evident symptoms in the infected host, and highlight the utility of combining dynamic modelling and phylogenetic analyses for understanding the often complex infection dynamics associated with emergent outbreaks.

1. Introduction

Pathogens able to spread at the interfaces between livestock, wildlife, and humans are one of the most serious threats to human health, wildlife conservation, and livestock economic sustainability [1,2]. Generally, the spread of a pathogen is enhanced when it co-circulates in multiple sympatric host species, as interspecific and intraspecific transmissions can complement each other, resulting in pathogen persistence [3,4].

M. bovis, a member of the *Mycobacterium tuberculosis* complex (MTBC) [5], is responsible for bovine (or animal) tuberculosis (bTB) in domestic cattle and a range of wild mammals [6], including European badgers and deer in Great Britain and Ireland [7–9], deer and wild boar in the Iberian Peninsula [10,11], deer and elk in Michigan, US [12], possums in New Zealand [13,14], and water buffalo in South Africa [15].

In Great Britain and Ireland several studies have established an association between the presence of infected badger populations and the persistence of bTB in cattle [16–18]. More recently, researchers have been able to demonstrate that the same *M. bovis* strains are co-circulating in domestic cattle and sympatric badgers in endemic bTB areas, first using the pathogen's DNA genotyping techniques [19,20], and later using whole-genome sequencing [21,22]. Despite the efforts made by governments to control and eradicate bTB, the last decades have seen an increase in the number of cases and substantial expansion of bTB endemic areas, in particular in England and Wales [5,23]. The eradication efforts of this disease in England alone, costs the UK government around 100 million pounds per year [8,24,25].

Collecting reliable and up-to-date data about wildlife populations can present many challenges at broader scales [26] and while over the years many studies focused on characterising specific badger populations (see [27], and references therein), the lack of information in some areas might prevent the design of effective disease control practices

when bTB is introduced. In addition, broad surveys across England have shown that, since the mid-1980s, the estimated number of badger social groups has been increasing by 2.6% annually, contributing to the uncertainty around their level of contributions potential as bTB reservoir in different regions [28]. However, reliable estimates of badger density, movement, and potentially infectious contact patterns are poorly recorded in most regions of Great Britain and Ireland, with only few populations such as at Woodchester Park (Gloucestershire, England) subject to denser sampling, in this case since the 1980s [29].

Further complications arise from the difficulty of estimating the true prevalence of *M. bovis* infection in badgers, as well as in domestic cattle. First is the elusive nature of *M. bovis*: the bacillus is characterised by slow replication with the potential for latent periods of variable length within the host [30,31]. In addition, the accuracy of currently available diagnostic tests is suboptimal in both cattle [32] and badgers [33]. These factors contribute to obscuring the relative roles of the two species in bTB maintenance and spread, and hampering the control and surveillance strategies. In particular, if both species are able to maintain the disease, control efforts focused on only one will be ineffective in achieving eradication [22].

One of the main goals of the current control and eradication strategy in GB is to prevent bTB from becoming established in new non-endemic areas, in particular those officially categorised as “low-risk” [34]. Within the Low Risk Area of England (LRA), the eastern part of the Cumbria county in north-west England (hereafter referred to as ‘East Cumbria’), has recently experienced a bTB outbreak of unusual magnitude and duration for this area. The outbreak began in 2014 and by mid-2019, through enhanced TB surveillance testing of all cattle herds in the affected area of East Cumbria, it had resulted in the detection of 39 breakdowns (positive cattle herds) across 33 premises [35]. The outbreak was caused by a strain of *M. bovis* (genotype) new to England, but previously observed in Northern Ireland, which was shown by preliminary molecular analyses to be the likely area of origin (Skuce,

103 personal communication). Surveillance of 'found dead' wildlife (badgers and wild deer) for
104 *M. bovis* infection was initiated in September 2016 in the area by the Animal and Plant
105 Health Agency (APHA) [35]. By August 2018 three (out of 52 inspected) roadkill badgers
106 had been found to be infected, all of them with the same bacterium genotype previously
107 isolated from local cattle herds. As a result of this epidemiological link, a badger cull was
108 licensed in a defined area within the affected part of East Cumbria with the aim of eradicating
109 the disease in badgers and cattle. During the first season of culling operations in the autumn
110 of 2018, 11% of all the removed badgers were found to be infected, all animals with the same
111 genotype, suggesting that *M. bovis* infection has become established in the local badger
112 population.
113 The aim of this study was to shed light on the dynamic spread of *M. bovis* when introduced in
114 a two-host system in a non-endemic area, for which the outbreak in East Cumbria provided
115 us with a unique opportunity to closely study. Here, we describe the East Cumbria outbreak's
116 spatial and temporal characteristics, and identify the factors which led *M. bovis* infection
117 becoming established in a wildlife population, while estimating the number of intraspecies
118 and interspecies transmission events. Our approach includes the use of forensic molecular
119 epidemiology [36], since over 60 isolates of *M. bovis* from the outbreak with usable whole
120 genomic sequences were available at the time of writing; complete with precise metadata
121 including dates, locations and, for cattle, animal and farm identifications.
122 Results of this study are an important step toward a deeper understanding of bTB
123 introduction and establishment into non-endemic areas, thus assisting with the process of
124 disease risk assignment and future policy decision making by the animal health authorities.

2. Results

2.1.Outbreak description

In November 2014, typical bTB lesions were detected via routine slaughterhouse inspection of a seven month-old male calf from a dairy herd in East Cumbria. Bacteriological culture of the lesions yielded an unusual genotype of *M. bovis*, designated 17:z by APHA (Figure 1, A). Following this first report, 23 more cattle were confirmed to be infected with the same strain in East Cumbria, with the last of these detected in November 2018 (at the time of writing). Further cattle were declared bTB positive (using the tuberculin skin test and/or supplementary interferon-gamma blood tests) during this period in the outbreak area, although an *M. bovis* bacilli could not be isolated. The 24 cattle infected with *M. bovis* 17:z genotype included three animals detected outside the outbreak area but still in Cumbria, as well as three in the neighbouring counties of Lancashire (two) and Yorkshire (one), all deemed likely to be part of the same outbreak due to associations through contact tracing. The index animal in this outbreak was a homebred calf that had never left its birth farm until it was moved to slaughter. Therefore, this animal could not have been the "case zero" of this outbreak. We attempted to trace back the first infected individual introduced in the outbreak area by analysing all the Cattle Trace System dataset records that included animals born in Northern Ireland or in the Republic of Ireland from 2009 to 2014. Tracing back the direct movements from Northern Ireland to the outbreak area indicated a limited number of "first arrival" premises (on average 9.3 per year, range 5-15), but unfortunately it did not reveal an obvious first introduction. Conversely, searching for indirect links between Northern Irish farms by selecting other British farms with links to the Cumbrian outbreak which previously imported animals from Northern Ireland, provided too many potential "arrival" premises (on average 216.5 and range 165-247, farms in the outbreak area per year).

149 During 2016 and 2017, respectively, two and 35 roadkill badger carcasses were reported to
150 the local authorities within the designated outbreak area of East Cumbria. Three of the badger
151 carcasses, retrieved respectively in January, February, and April 2017, were positive for *M.*
152 *bovis* on culture (while two carcasses were unsuitable for inspection) (Figure S1.1), and all
153 three positive animals were infected with the 17:z genotype.
154 The identification of infected badgers led Defra, following a public consultation, to issue a
155 badger culling licence in a specified section of the outbreak area in the autumn of 2018.
156 Culling operations from September to November 2018 resulted in 602 culled badgers, of
157 which 369 were submitted for post mortem inspection and laboratory testing (Defra 2019). In
158 total, 42 were culture positive for *M. bovis* and of those 38 isolates yielded a whole genome
159 sequence of sufficient quality to enable phylogenetic analysis [37].
160 Data on found-dead surveillance of 2018 and 2019, but prior to the second culling season,
161 included an additional 42 retrieved carcasses (29 and 13, respectively), all negative to *M.*
162 *bovis*, except for 15 that were unsuitable for post-mortem inspection, and three still pending
163 at the time of data gathering.
164

165 **2.2. Outbreak phylogeny and Northern Ireland isolates**

166 The phylodynamic tree of the East Cumbria outbreak is reported in Figure 1 (B). Early
167 analyses identified a genotype usually found in Northern Ireland; further evidence showed
168 the existence of cattle movements from this area to England and Wales. Coincidentally, the
169 origin area included the recently completed Test, Vaccinate or Remove (TVR) trial area [38]
170 in Northern Ireland (Skuce, personal communication) where extensive *M. bovis* whole
171 genome sequencing had already been done – these isolates were included in the current
172 analyses.

173 The complete phylogenetic tree (Figure 2) confirmed the association between the *M. bovis*
174 circulating in the Northern Irish TVR area and in East Cumbria; thus it appeared that the East
175 Cumbrian outbreak likely originated from the dominant strain circulating in or around the
176 TVR area (Figure 2, orange branches) imported by movement of infected cattle, though the
177 first introduction was not identified.
178

179 **2.3. Epidemiological signatures in genetic data**

180 Following our previous approach [22], epidemiological signatures in the sequence distances
181 were identified using Boosted Regression Trees (BRT) [39], which combines decision trees
182 and boosting techniques [40].
183 As previously, the dependent variable was the genetic distance between *M. bovis* strains,
184 expressed as single nucleotide variants (SNVs); for explanatory variables we calculated 18
185 relational covariates for each pair of sampled animals. These covariates are listed in Table 1
186 and are divided into four categories: temporal (1 covariate), spatial (2 covariates), group (3
187 covariates), and contact networks (12 covariates).
188 The BRT model on the full dataset (1,711 observations) was able to explain 41% (pseudo R^2)
189 of the variability on the test dataset, while the Root Mean Squared Error (RMSE) was 0.94.
190 To test the robustness of this model we ran the same analysis on the same dataset but with an
191 increasing percentage of randomly reassigned values for the dependent variable (SNV
192 distance) observations. The results (Figure S3.1) showed that even for a limited percentage of
193 re-assigned data (10%) the model underperformed substantially, explaining only 28% of the
194 variation on average. The full model results (Figure 3) showed that the most important
195 covariate to explain the SNV distance between isolates was the time between isolate
196 samplings (25.2%), followed by the hosts populations' size (14.3%) and the geographical
197 distance (considering both isolate sampling locations and between isolates land parcels,

respectively 11.3 and 9.3%). Among the contact network covariates, the degree in multi-layer (9.2%) and land parcels networks (9.1%) were the most important, while the covariates related to the single-species networks were not important (all lower than 4%). Partial dependency plots are reported in Figure S3.1.

We ran a further three BRT models: two of them considered the within-species interactions only, cattle-to-cattle (153 observations), and badger-to-badger (820 observations). Conversely, the third model considered cross-species interactions (738 observations). These models were able to explain respectively 16%, 43% and 36% of the variation, while the respective RMSE values were 0.82, 1.04, and 0.96.

The covariates influence rankings in the badger and inter-species models were similar to the full model (sampling time distance the most influent, respectively, 19.6 and 26.7%), with the exception that in both cases the degree in the land parcels network was the third most influential covariate (11.4 and 12.0%). In the badger model the population size dropped to sixth (9.7%), while the geographic distance between sampling locations was second (14.5%). The cattle model showed the most differences with the full model, with group size and degree in the land-parcel network being the most influential covariates (22.2 and 12.9%), while sampling time distance was third (11.0%).

2.4. Pairwise transmission probability and most likely transmission tree

The pairwise transmission probability was calculated using the Kolmogorov Forward Equations (KFEs)[41], and the pairwise transmission probability matrix was reported in Figure 4. The KFEs methodology was used to calculate the probability of observing a pair of bTB infected hosts given their sampling time and the genetic divergence between their bacterial isolates, assuming that a direct transmission occurred between the pair.

222 The animal that infected the first detected cattle (C1) with *M. bovis* might have escaped
223 detection, since the transmission probabilities from other sampled animals are low (median
224 0.13×10^{-3}, range $0 – 1.98 \times 10^{-3}$). Similarly, two out of three roadkill badgers (B44 and B18)
225 had, respectively, the lowest and the third lowest average and maximum transmission
226 probability from other sampled animals (see Figure 5). This indicates that animals infected
227 early in the outbreak likely escaped detection (either the 4-year herd testing or carcass
228 inspection), in particular the "case zero" (i.e. the first cow imported infected with the 17:z
229 genotype of *M. bovis*).
230 In general, within-species transmission probabilities were higher than between-species ones
231 (Figure S4.1), and the phylogenetic root-to-tip temporal signal was strong as well ($R^2 = 0.39$,
232 p-value ~0; see Figure S4.2). This was consistent with the BRT model results, where the
233 temporal signal was identified as the most important factor to predict the SNV distance.
234 We computed 10,000 "random trees" built by selecting random transmissions (except the
235 ones for which the probability was zero because the cattle's lifespans did not overlap).
236 Results showed that random trees had a median [95th quantile] of 17[13–21] cattle-to-cattle,
237 26[21–31] badger-to-badger, 14[10–19] cattle-to-badger, and 6[2–10] badger-to-cattle
238 transmissions (Figure S4.3). We computed the most likely transmission tree by selecting the
239 transmissions with highest probability within the sampled animals while avoiding loops in
240 the tree (see [41]). The best tree (Figure 4) showed that most of the transmissions in this
241 system likely happened within-species, i.e. 20 cattle-to-cattle and 29 badger-to-badger,
242 respectively. Conversely, inter-species transmissions comprised 12 cattle-to-badgers
243 incidents, and 3 badgers-to-cattle. When comparing these results with the randomly
244 computed trees, the most-likely tree showed a lower number of cross-species transmissions
245 and a higher number of within-species transmissions, although the estimates fell in the 95th
246 quantile of the random trees ones.

247

248 **2.5. BASTA analysis**

249 The transmission rates estimated by Bayesian Structured coalescent Approximation
250 (BASTA) [42] on the 10 different sub-samples suggest that transmission from cattle-to-
251 badgers occurred much more frequently (at least an order of magnitude) than transmission
252 from badgers-to-cattle in Cumbria (Figure 7). In addition, there is little support for the
253 inclusion of badgers-to-cattle transmission in the structured population model. In contrast,
254 there is strong support for transmission of *M. bovis* from the sampled Northern Ireland area
255 into the Cumbria area via the cattle population. Figure S5.1 shows the rate estimates that
256 BASTA produced when no genomic data was provided, therefore only sampling dates were
257 used. These analyses were conducted to determine how much signal there was in the genomic
258 data to support the transmission rates being estimated. Given the contrasting rates shown in
259 Figure 7 and Figure S5.1, there was strong evidence that there was sufficient signal in the *M.*
260 *bovis* genomic data to estimate the transmission rates. Lastly, there is good agreement across
261 the 10 sub-samples, suggesting that the estimated rates were robust to any inter-sub-sample
262 variation.

263 Analyses in BASTA leveraging the temporal signal in the *M. bovis* genomic data were used
264 to estimate the timing of *M. bovis* transmission from the cattle in the TVR area to cattle in
265 Cumbria (Figure S4.4). Whilst credible intervals around these estimates were very broad, the
266 transmission event was estimated to have occurred in March 2011 (lower 2.5% bound
267 estimate: August 2001; upper 97.5% bound estimate: April 2014). This reflect the slow and
268 variable replication rate characteristic of *M. bovis*.

269

3. Discussion

While it is accepted that cattle movements are responsible for transmission on a wide spatial scale [23,43,44], uncertainty remains around the relative roles of cattle and badger in maintaining and spreading *M. bovis* infection at local scales, complicating the formulation and execution of control policies. Our analyses of the East Cumbria outbreak highlight this dynamic, with the introduction of *M. bovis* most likely being caused by a cattle movement from Northern Ireland, followed by a more complex spread among local cattle and badgers. Identifying the source of infection for the index case in cattle in this outbreak (marked as C1 in the figures) would be a crucial piece of information. Should a badger be the most likely infection source then this would imply an earlier establishment of the disease in wildlife (i.e. before or during 2014). Our results suggest another infected bovid as the most likely bTB source for the index case, but the transmission from known cases is poorly supported as the transmission probabilities from the other sampled animals to the index are generally low (Figure 5).

Overall, both the likely transmission tree selection (obtained with the KFEs) and the BASTA analysis indicate that most transmissions happened within-species, and that cattle-to-badger transmission has played a more important role than badger-to-cattle transmission, with similar outcomes obtained using two different analyses.

Our hypothesis, which is supported by the described results, is that after an initial seeding of *M. bovis* into the local East Cumbria cattle population, the infection subsequently became established in local badgers. Once that happened, rapid spread of the pathogen led to the observed 11% prevalence in the autumn of 2018, with peaks up to 20.9% in the core area [37]. Given the historically low prevalence of *M. bovis* in this area [45], the immunological naivete of this specific badgers population might have facilitated a quicker spread, although more research in this direction is needed. From a disease dynamic point of view, the

295 relatively higher number of predicted cattle-to-badger transmission events suggests that the
296 establishment in badgers may happen when infection pressure from the sympatric cattle
297 population reaches a threshold, rather than a single transmission event.
298 A concern might be that transmission tree outcomes have been affected by the sampling
299 timeline, since early in the outbreak there are less *M. bovis* sampled sequences and most
300 come from cattle, while later on the majority of sampled isolates come from badgers. While
301 the BASTA analysis is designed to specifically reduce the effects of unbalanced sampling
302 [42], our conclusions are also supported by a low genetic diversity in the recovered badger *M.*
303 *bovis* isolates, which points to a relatively recent outbreak in this population. Nonetheless, the
304 BASTA analysis was able to exploit this low genetic diversity, as it is shown by the different
305 results of this model when neglecting the genetic information.
306 Similar to a previous study in a separate population [22], the Boosted Regression Trees
307 (BRT) analysis indicated temporal bacterial isolation differences, and spatial distance
308 between hosts, as the most important factors to predict inter-*M. bovis* genetic distance
309 (SNVs). Another important variable was host population sizes, with bigger groups (farms or
310 setts) linked to higher SNV distances. This result was particularly strong for the BRT models
311 which included cattle isolates, and it is consistent with farms size being a risk factor for bTB
312 [46,47].
313 Including different contact networks metrics variables in the BRT model can help to
314 understand which transmission mechanism is more relevant in this system. Among these
315 variables, the number of adjacent land parcels, corresponding to the spatial network's degree,
316 was equal (full model) or more important (cattle, badger and inter-species models) than other
317 metrics. This points to fine-scale spatial effects which cannot be entirely explained by the
318 simple distance between samples locations. On the contrary, the single-species network
319 metrics (animal movements for cattle and sett adjacency for badgers) were not significant in

320 any models. This might be the result of complex interactions between the two species that
321 cannot be explained when considered individually, as previously suggested [48]. However,
322 these interactions become evident when both species contact networks are included in a
323 single framework such as multi-layer network [49].
324 The badger-to-badger model was able to predict the SNV distances in the test dataset slightly
325 better than the full model (43% to 42%), despite reducing the sample size (from 1,711
326 observations in the full model, to 820 in the badger only model). Furthermore, we observed
327 that the SNV distance variability explained by the cattle only model was comparable to that
328 of the full model when 20% of the data were randomly shuffled. This was surprising since
329 more data are collected on domestic animals and this should provide a better picture than the
330 equivalent wildlife data. One potential explanation is that the selected variables for this
331 model might better explain *M. bovis* dynamics in badgers than in cattle, at the local scale.
332 However, when we consider this and the good explanatory power of the land parcels network
333 metrics, we could speculate that the cattle dataset might hide some contact patterns. This
334 could occur due to the inability to isolate *M. bovis* in all herd breakdowns or in all cattle, or
335 due to unrecorded movements, with cattle grazing in several land parcels belonging to the
336 same farm but not contiguous to the farm [50,51]. In general, the landscape and the farmland
337 fragmentation and distribution in space, which is ignored when considering farms' main
338 building locations only, might play a crucial role in disease dissemination. This calls for
339 surveillance and control strategies to be adapted to specific contexts and informed by detailed
340 veterinary investigation and spatial information, such as land parcel distribution.
341 By comparing the East Cumbria outbreak with other bTB in other contexts leads to some
342 important insights. In East Cumbria, badgers have played a lesser role in the local persistence
343 of the disease, the converse to what observed in a bTB endemic area in Woodchester Park
344 [22]. This may be due to the "age" of the outbreak. Whilst in newly infected areas, and

345 therefore non-endemic, cattle-to-cattle and cattle-to-wildlife transmissions may dominate, in
346 endemic areas the dynamic may have shifted towards a more complex dynamic, where
347 wildlife can play a more important role. While in both cases breaking the transmission at the
348 wildlife/livestock interface is critical, in outbreaks within low risk areas (non-endemic) it is
349 crucial to prevent the establishment of bTB.
350 Questions remains as to how likely bTB is to get established in non-endemic areas, and how
351 long it would take to detect it. From this perspective, the fact that the first introduction in this
352 area (estimated to be 2011), and that the initial infected cattle were not detected until
353 inspection at slaughter, indicates limitations of the bTB surveillance strategy in low risk areas
354 and the importance of good biosecurity to reduce the risk of onward transmission to wildlife
355 from introductions of cattle with undetected *M. bovis* infection. In order to mitigate this risk,
356 in April 2016 Defra adopted mandatory post movement bTB testing of all cattle moved from
357 high-risk to lower-risk areas [52].
358 To conclude, our analyses of the recent East Cumbria outbreak highlight how the
359 transmission dynamics of *M. bovis* can change during the establishment in a non-endemic
360 area, and how these changes affect the relative roles of wildlife and domestic animals in
361 establishing and maintaining the infection. Our results suggest how genomic data and
362 phylogenetic analyses are becoming fundamental to disentangle bacterial pathogen outbreaks,
363 in particular when combined with epidemiological and network models [53]. Therefore,
364 infrastructures for genomic surveillance can definitely help inform bTB and other endemic
365 diseases control policies.
366 Finally, we highlight how local spatial dynamics might affect pathogen spread during the
367 early phases of an outbreak. This makes the case for different diseases control strategies in
368 endemic and non-endemic areas that take into account detailed characteristics about the
369 system landscape and rearing practices, and for improving biosecurity, in order to achieve

370 minimal transmission at the cattle/wildlife interface and therefore preventing the

371 establishment of newly introduced diseases.

372

Methods

4.1. Data description

4.1.1. Sequences and metadata

Test positive cattle, found-dead badgers and most culled badgers in the Cumbria area were subject to post-mortem and culture of suitable tissues at the Animal and Plant Health Agency (APHA), with positive cultures subjected to genotyping and whole genome sequencing at the Central Sequencing Unit in Weybridge. 65 *M. bovis* whole-genome sequences were available from East Cumbria (65 in total, 24 from cattle, 3 from roadkill badgers, and 38 from culled badgers). The sampling timeline of the available sequences is reported in Figure S1.1. The metadata included a unique identifier, the sampling date, location coordinates (for badger isolates), and the farm's county-parish-holding (CPH) code (for cattle sequences only). The isolates and raw sequence data were processed using the same pipeline as described by Crispell et al. [22].

The East Cumbria outbreak dataset included a further sequence from the same *M. bovis* genotype sampled in Scotland (Figure S1.2). The epidemiological investigation showed that the animal was imported from Northern Ireland for slaughter only, thus we did not consider it in the analyses.

4.1.2. Badger population

A total of 160 badger setts were identified in and around the outbreak area in 2017/18 by the APHA, 117 of them were in the 2018 culling permit area. Badgers were both shot and trapped, with trapped animals subjected to post-mortem analysis, and population data (number of badgers removed, TB positive, negative, and TB status not determined) were available for 99 setts.

398 ***4.1.3. Cattle population and outbreak area definition***

399 To obtain all infected cattle life histories (movements, birth and death) we first matched the
400 sequences' unique identifier in the SAM dataset, then we extracted the data from the Cattle
401 Tracing System (CTS) using the animal unique identifier.

402 The outbreak area was defined as the area within the minimum circle around the sequences
403 sampling locations and all 160 badger setts, and adding to that all the parcels assigned to the
404 infected farms which are contiguous to the above described circle (Figure 1, A). This study
405 included all the farms active between 01/01/2010 and 31/12/2018 (which reported any cattle
406 movements, births or deaths) directly located in the area or which owned a parcel in the
407 outbreak area. The total number of selected cattle farms was 336.

408

409 ***4.1.4. Northern Ireland TVR data***

410 The Test, Vaccinate or Remove (TVR) trial in Northern Ireland ran from 2014 to 2018, and it
411 was designed to determine whether a combination of vaccination and an animal side TB test
412 could be an effective means of controlling *M. bovis* infection in badger populations. During
413 this period *M. bovis* samples were taken from infected cattle and badgers in the area for
414 culturing.

415 All positive cultures were sequenced at the Agri-Food and Biosciences Institute in Belfast
416 (AFBI-NI). In addition, archived *M. bovis* isolates that were stored as part of routine
417 surveillance operations in the TVR area prior to the start of the trial were selected for
418 sequencing. These additional isolates were sourced from routine test and slaughter
419 surveillance of cattle or road killed badgers.

420 From the TVR area in Northern Ireland, there were 544 *M. bovis* genomes sourced from
421 infected cattle (479) and badgers (65), sampled from 1996 to 2017. The distribution of
422 sampling times for the *M. bovis* genomes is shown in Figure S1.1.

423

424 **4.2. Epidemiological information and machine learning analysis**

425 An important part of this analysis involved investigating the population, temporal, spatial,
426 and contact network signatures in the sampled *M. bovis* genomic data [22]. We computed the
427 inter-sequence Single Nucleotide Variants (SNV) distance and then we fitted a Poisson
428 regression model using the Boosted Regression Trees [39] model in R [54].

429 As mentioned above, we used population, temporal, spatial, and network covariates, which
430 are listed in Table 1. Network covariates (shortest distance, same community, degree, and
431 number of infected nodes in the shortest path) were calculated for three contact networks:
432 single species, spatial, and multi-layer.

433 The *single species network* accounted for only the within-species potential contacts. The
434 cattle population's nodes corresponded to farms, and edges correspond to movements. In
435 order to build these networks for each pair of cattle sequences, we only accounted for
436 movements spanning from the previous year of the first of the two sequences sampled, to the
437 year of the second one (i.e. if the two sequences were sampled respectively in 2015 and 2017,
438 we computed the network by using movements recorded between 2014 and 2017). The
439 badger population's nodes corresponded to setts, and since most of the trapped badgers
440 occurred near a sett, their sequences were assigned to the closest one. We used the Voronoi
441 partition (or Dirichlet tessellation) to create the edges between setts, in order to avoid relying
442 on an assumption for the distance at which two setts were connected. In this network
443 configuration the inter-species sequences are considered not connected.

444 The *spatial network* was built by considering each land parcel as a single node. Two land
445 parcels were considered connected by the proximity criterion: if they shared a border or if the
446 borders were closer than 100 meters. In this case, the degree of a node in this network (i.e.
447 land parcel) corresponds to the number of neighbouring parcels.

448 Finally, the *multi-layer network* accounted for all the previously described contacts
449 combined. In this case, nodes are defined as land parcels as well, but on top of the proximity
450 criterion, two land parcels could also be connected if they include two farms or setts which
451 were connected in the single species network.
452 The network communities were defined using the Louvain method which optimises the
453 modularity, as provided in the R package *igraph* [55].
454 The model was run on the complete dataset (All Samples, Table 1 and Figure 3), on the cattle
455 and badger isolates only, and on the inter-species isolates. In all cases the train and test
456 dataset were half of the observations, while in the cattle dataset we used 60% for training and
457 40% for testing, given the reduced observations compared to the others. We evaluated the
458 models using the pseudo-R^2 calculation comparing the observed and predicted SNV distances
459 on the test datasets, and the Root Mean Squared Error (RMSE). These were both calculated
460 using the package *caret* [56]. For BRT the relative influence of the covariates is determined
461 by the times each variable is selected to split the data in a decision tree, which in turn is
462 weighted by the improvement to the model fit that resulted from that variable being used at
463 each split [39]. All models were fitted with a 10-fold cross validation.
464 Preliminary runs of the models were used to tune the BRT parameters in order to improve the
465 predictions. These parameters were the learning rate, which controls the contribution of each
466 tree to the final model, and the tree complexity, which corresponds to the number of nodes in
467 the tree. For the full dataset model the learning rate was set to 0.005 and the tree complexity
468 to 5. For the other model we set the learning rate to 0.003, 0.009, and 0.0075 for the cattle,
469 badger and inter-species dataset, respectively and the tree complexity to 7, 1, and 10.
470

471 **4.3. Pairwise KFE and transmission tree reconstruction**

472 We calculated the transmission probability between pairs of animals where the *M. bovis*
473 sequences have been sampled. We calculated this probability using the Kolmogorov-Forward
474 Equations (KFE) methodology [41]. The KFEs consist of a set of ordinary differential
475 equations which track the probability of a system to be in a given state through time [57–59].
476 In the pairwise transmission case, the system state was given by the combination of the two
477 hosts disease progression state and, once an individual is infected, by the number of SNVs on
478 its *M. bovis* strain. The underlying assumption was that, at the time of infection, the two
479 pathogen strains found in the source and recipient hosts are identical. After the infection, the
480 two strains start to replicate, and thus substitution on the pathogen DNA happens at a rate μ
481 (*substitution rate*), generating SNVs. Because the two strains are diverging, we will call the
482 SNVs found in a strain sampled in the source host A [recipient host B] as *divergent* SNVs, or
483 *divA*[*divB*]. The sum of *divA* and *divB* results in the SNV distance between the two strains.
484 In order to use this methodology, we have to provide three main pieces of information: the
485 pathogens sequences, the two hosts sampling time, and an underlying model representing the
486 disease progression. For bTB, we chose to use a simple Susceptible-Exposed-Infectious
487 model [23,60], where susceptible can become exposed (or latent) after contact with an
488 infectious host with *infection rate* β, and exposed hosts move to the infectious state with
489 *transition rate* σ (or after a latency period of average $1/\sigma$). In this study, we provided the
490 birth date of the cattle as well, in order to limit the time span where each cattle could have
491 been infected first. For badgers we assumed a constant death rate, based on the observation
492 that less than 0.1% of individuals would survive past 8-years of age [27]. The
493 epidemiological parameters were chosen according the most recent literatures (see [41]), and
494 in order to account for their variability, for each pairwise transmission we tested a 1'000
495 combination of randomly selected parameters combinations and chose the one returning the
496 highest probability.

497 Following Rossi et al. [41], we assembled the most likely transmission tree by progressively
498 selecting the pairs with the highest transmission probability, and excluding those not possible
499 given the previously selected ones (e.g. if A→B and B→C are selected, B→A, C→B and
500 C→A were going to be a priori excluded).
501
502 **4.4. BASTA analysis**
503 Given that the phylogenetic evidence suggests a large amount of inter-species transmission is
504 occurring in East Cumbria, the next critical question is in what direction? We used the
505 Bayesian Structured coalescent Approximation (BASTA) package [42] with the Bayesian
506 evolutionary analysis platform BEAST2 (Bayesian Evolutionary analysis by Sampling Trees;
507 [61]) to estimate *M. bovis* inter-species transmission rates. The BASTA package was able to
508 estimate these rates whilst accounting for the known structure and sampling biases in the
509 study population. In the current study, the sampled *M. bovis* population was split into four
510 different sub-populations based on host species (badger or cow) and location (Cumbria or
511 TVR) (Figure S4.1). In addition, to estimate transmission rates in a structured population,
512 BASTA is robust to sampling biases, which are likely to be present in the current *M. bovis*
513 dataset. Importantly, it is assumed that transmission from TVR to Cumbria only occurred in
514 one direction, as the Cumbria clade is monophyletic within the larger TVR phylogeny
515 (Figure 2).
516 The evolutionary analyses using BASTA require that there is a temporal signal in the
517 *M. bovis* genomic data. With the presence of a temporal signal, the accumulation of
518 substitutions will be tied to the evolutionary processes of the sampled population, making it
519 possible to leverage genetic variation to estimate evolutionary dynamics such as transmission
520 rates. A root-to-tip versus sampling time regression was used to determine whether a
521 measurable temporal signal was present in the *M. bovis* genomic data (Figure S4.2). The

522 positive trend observed in this regression indicates the presence of a weak temporal signal,
523 therefore it was possible to proceed with the evolutionary analyses in BASTA.
524 The computational complexity of the analyses to be conducted using BASTA meant that the
525 large number of *M. bovis* genomes available within the outbreak clade (orange clade in
526 Figure 2) had to be sub-sampled.
527 In addition, it was only from 2014 to 2017 that the sampling of cattle and badgers in the TVR
528 area can be considered approximately equal (in terms of effort). Therefore, only genomes
529 sourced from infected cattle and badgers sampled from 2014 to 2018 were included in the
530 BASTA analyses. The window was extended to 2018 to include the genomes sourced from
531 Cumbria. The sub-sampling was then weighted to include samples from as many years as
532 possible and from equal numbers of cattle and badgers. The sub-sampling was conducted 10
533 times, each time selecting 20 badgers and 20 cattle-derived *M. bovis* genomes from Cumbria
534 and 40 badger and 40 cattle-derived genomes from the TVR area (an example of a sub-
535 sample is shown in Figure S4.3). A root-to-tip versus sampling time regression was
536 conducted for each sub-sample and found to be positive in all cases. Each sub-sample was
537 then analysed separately in BASTA.
538

539 **References**

540 1. Gortazar C, Diez-Delgado I, Barasona JA, Vicente J, De La Fuente J, Boadella M. The
541 Wild Side of Disease Control at the Wildlife-Livestock-Human Interface: A Review.
542 Front Vet Sci. 2015;1: 1–12. doi:10.3389/fvets.2014.00027
543 2. Wiethoelter AK, Beltrán-Alcrudo D, Kock R, Mor SM. Global trends in infectious
544 diseases at the wildlife–livestock interface. Proc Natl Acad Sci. 2015;112: 9662–9667.
545 doi:10.1073/pnas.1422741112
546 3. Haydon DT, Cleaveland S, Taylor LH, Laurenson MK. Identifying reservoirs of
547 infection: A conceptual and practical challenge. Emerg Infect Dis. 2002;8: 1468–1473.
548 doi:10.3201/eid0812.010317
549 4. Craft ME, Hawthorne PL, Packer C, Dobson AP. Dynamics of a multihost pathogen in
550 a carnivore community. J Anim Ecol. 2008;77: 1257–1264. doi:10.1111/j.1365-
551 2656.2008.01410.x
552 5. Smith NH, Gordon S V., de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG.
553 Bottlenecks and broomsticks: The molecular evolution of *Mycobacterium bovis*. Nat
554 Rev Microbiol. 2006;4: 670–681. doi:10.1038/nrmicro1472
555 6. Palmer M V. *Mycobacterium bovis*: Characteristics of wildlife reservoir hosts.
556 Transbound Emerg Dis. 2013;60: 1–13. doi:10.1111/tbed.12115
557 7. Skuce R, Breadon E, Allen A, Milne G, McCormick C, Hughes C, et al. Longitudinal
558 dynamics of herd-level *Mycobacterium bovis* MLVA type surveillance in cattle in
559 Northern Ireland 2003–2016. Infect Genet Evol. 2020;79: 104131.
560 doi:10.1016/j.meegid.2019.104131
561 8. Godfray HCJ, Donnelly CA, Kao RR, Macdonald DW, McDonald R, Petrokofsky G,
562 et al. A restatement of the natural science evidence base relevant to the control of
563 bovine tuberculosis in Great Britain. Proc R Soc B Biol Sci. 2013;280: 1–18.

564 doi:10.1098/rspb.2013.1634
565 9. Crispell J, Cassidy S, Kenny K, McGrath G, Warde S, Cameron H, et al.
566 *Mycobacterium bovis* genomics reveals transmission of infection between cattle and
567 deer in Ireland. Microb genomics. 2020;6. doi:10.1099/mgen.0.000388
568 10. Anderson LG, Gortázar C, Vicente J, Hutchings MR, White PCL. Modelling the
569 effectiveness of vaccination in controlling bovine tuberculosis in wild boar. Wildl Res.
570 2013;40: 367–376. doi:10.1071/WR12139
571 11. Madeira S, Manteigas A, Ribeiro R, Otte J, Fonseca AP, Caetano P, et al. Factors that
572 Influence *Mycobacterium bovis* Infection in Red Deer and Wild Boar in an
573 Epidemiological Risk Area for Tuberculosis of Game Species in Portugal. Transbound
574 Emerg Dis. 2017;64: 793–804. doi:10.1111/tbed.12439
575 12. Salvador LCM, O'Brien DJ, Cosgrove MK, Stuber TP, Schooley A, Crispell J, et al.
576 Disease management at the wildlife-livestock interface: using whole-genome
577 sequencing to study the role of elk in *Mycobacterium bovis* transmission in Michigan,
578 USA. Mol Ecol. 2019; 1–14. doi:10.1111/mec.15061
579 13. Crispell J, Zadoks RN, Harris SR, Paterson B, Collins DM, de-Lisle GW, et al. Using
580 whole genome sequencing to investigate transmission in a multi-host system: bovine
581 tuberculosis in New Zealand. BMC Genomics. 2017;18: 180. doi:10.1186/s12864-
582 017-3569-x
583 14. Anderson DP, Ramsey DSL, Nugent G, Bosson M, Livingstone P, Martin PAJ, et al. A
584 novel approach to assess the probability of disease eradication from a wild-animal
585 reservoir host. Epidemiol Infect. 2013;141: 1509–1521.
586 doi:10.1017/S095026881200310X
587 15. Fitzgerald SD, Kaneene JB. Wildlife Reservoirs of Bovine Tuberculosis Worldwide:
588 Hosts, Pathology, Surveillance, and Control. Vet Pathol. 2013;50: 488–499.

589 doi:10.1177/0300985812467472
590 16. Martin SW, Eves J a, Dolan L a, Hammond RF, Griffin JM, Collins JD, et al. The
591 association between the bovine tuberculosis status of herds in the East Offaly Project
592 Area, and the distance to badger setts, 1988-1993. Prev Vet Med. 1997;31: 113–25.
593 doi:10.1016/S0167-5877(96)01111-7
594 17. Donnelly CA, Woodroffe R, Cox DR, Bourne FJ, Cheeseman CL, Clifton-Hadley RS,
595 et al. Positive and negative effects of widespread badger culling on tuberculosis in
596 cattle. Nature. 2006;439: 843–846. doi:10.1038/nature04454
597 18. Donnelly CA, Woodroffe R, Cox DR, Bourne J, Gettinby G, Le Fevre AM, et al.
598 Impact of localized badger culling on tuberculosis incidence in British cattle. Nature.
599 2003;426: 834–837. doi:10.1038/nature02192
600 19. Olea-Popelka FJ, Flynn O, Costello E, McGrath G, Collins JD, O'Keeffe J, et al.
601 Spatial relationship between *Mycobacterium bovis* strains in cattle and badgers in four
602 areas in Ireland. Prev Vet Med. 2005;71: 57–70. doi:10.1016/j.prevetmed.2005.05.008
603 20. Woodroffe R, Donnelly CA, Cox DR, Gilks P, Jenkins HE, Thomas Johnston W, et al.
604 Bovine tuberculosis in cattle and badgers in localized culling areas. J Wildl Dis.
605 2009;45: 128–143. doi:10.7589/0090-3558-45.1.128
606 21. Biek R, O'Hare A, Wright D, Mallon T, McCormick C, Orton RJ, et al. Whole
607 Genome Sequencing Reveals Local Transmission Patterns of *Mycobacterium bovis* in
608 Sympatric Cattle and Badger Populations. PLoS Pathog. 2012;8.
609 doi:10.1371/journal.ppat.1003008
610 22. Crispell J, Benton CH, Balaz D, De Maio N, Akhmetova A, Allen A, et al. Combining
611 genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium*
612 *bovis* in a multi-host system. Elife. 2019; 1–36.
613 doi:https://doi.org/10.7554/eLife.45833.001

23. Brooks-Pollock E, Roberts GO, Keeling MJ. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature. 2014;511: 228–231. doi:10.1038/nature13529

24. Tildesley MJ, Brand S, Brooks Pollock E, Bradbury N V., Werkman M, Keeling MJ. The role of movement restrictions in limiting the economic impact of livestock infections. Nat Sustain. 2019;2: 834–840. doi:10.1038/s41893-019-0356-5

25. Defra. Bovine TB Eradication Programme for England. 2011.

26. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015. pp. 20140107–20140107. doi:10.1098/rstb.2014.0107

27. Roper T. Badger. Collins; 2010. Available: https://books.google.co.uk/books?id=FZhrPgAACAAJ

28. Judge J, Wilson GJ, Macarthur R, Delahay RJ, McDonald RA. Density and abundance of badger social groups in England and Wales in 2011-2013. Sci Rep. 2014;4: 1–8. doi:10.1038/srep03809

29. Delahay RJ, Walker N, Smith GS, Wilkinson D, Clifton-Hadley RS, Cheeseman CL, et al. Long-term temporal trends and estimated transmission rates for *Mycobacterium bovis* infection in an undisturbed high-density badger (*Meles meles*) population. Epidemiol Infect. 2013;141: 1445–1456. doi:10.1017/S0950268813000721

30. Pollock JM, Neill SD. *Mycobacterium bovis* infection and tuberculosis in cattle. Vet J. 2002;163: 115–127. doi:10.1053/tvjl.2001.0655

31. Cassidy JP. The pathogenesis and pathology of bovine tuberculosis with insights from studies of tuberculosis in humans and laboratory animal models. Veterinary Microbiology. 2006. pp. 151–161. doi:10.1016/j.vetmic.2005.11.031

32. Nuñez-Garcia J, Downs SH, Parry JE, Abernethy DA, Broughan JM, Cameron AR, et

al. Meta-analyses of the sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland. Prev Vet Med. 2018;153: 94–107. doi:10.1016/j.prevetmed.2017.02.017

33. Drewe JA, Tomlinson AJ, Walker NJ, Delahay RJ. Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. PLoS One. 2010;5. doi:10.1371/journal.pone.0011196

34. Perrin LD, Harris KA, Reynolds M, Lawes JR, Frost S, Brouwer A, et al. Bovine TB infection status in cattle in Great Britain in 2017. Vet Rec. 2019;184: 371–378. doi:10.1136/vr.l1321

35. Defra. An update on TB surveillance in wildlife. 2019 [cited 26 May 2020] p. 6. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/830810/surveillance-wildlife-2018.pdf

36. Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: How whole-genome sequencing and big data are transforming epidemiology. Trends Microbiol. 2014;22: 282–291. doi:10.1016/j.tim.2014.02.011

37. Defra. TB surveillance in badgers during year 1 badger control operations in eastern Cumbria, Low Risk Area (2018). 2019 [cited 31 Oct 2020]. Available: https://www.gov.uk/government/publications/bovine-tb-surveillance-in-wildlife-in-england/tb-surveillance-in-badgers-during-year-1-badger-control-operations-in-eastern-cumbria-low-risk-area-2018

38. DAERA-NI. The test and vaccinate or remove (TVR) wildlife intervention research project - year 2 report. www.dardni.gov.uk/ test-and-vaccinate-or-remove.htm. 2015. Available: https://www.daera-ni.gov.uk/articles/test-and-vaccinate-or-remove-tvr-wildlife-intervention-research

664 39. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim
665 Ecol. 2008;77: 802–813. doi:10.1111/j.1365-2656.2008.01390.x
666 40. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, et al. Predictive
667 analysis across spatial scales links zoonotic malaria to deforestation. Proc R Soc B
668 Biol Sci. 2019;286. doi:10.1098/rspb.2018.2351
669 41. Rossi G, Crispell J, Balaz D, Lycett SJ, Delahay RJ, Kao RR. Identifying likely
670 transmission pairs with pathogen sequence data using Kolmogorov Forward
671 Equations; an application to *M.bovis* in cattle and badgers. BiorXiv. 2020; 1–32.
672 doi:10.1101/2020.06.11.146894
673 42. De Maio N, Wu CH, O'Reilly KM, Wilson D. New Routes to Phylogeography: A
674 Bayesian Structured Coalescent Approximation. PLoS Genet. 2015;11: 1–22.
675 doi:10.1371/journal.pgen.1005421
676 43. Green DM, Kiss IZ, Mitchell AP, Kao RR. Estimates for local and movement-based
677 transmission of bovine tuberculosis in British cattle. Proc R Soc B Biol Sci. 2008;275:
678 1001–1005. doi:10.1098/rspb.2007.1601
679 44. Gilbert M, Mitchell A, Bourn D, Mawdsley J, Clifton-Hadley R, Wint W. Cattle
680 movements and bovine tuberculosis in Great Britain. Nature. 2005;435: 491–496.
681 doi:10.1038/nature03548
682 45. Atkins PJ, Robinson PA. Bovine tuberculosis and badgers in Britain: Relevance of the
683 past. Epidemiol Infect. 2013;141: 1437–1444. doi:10.1017/S095026881200297X
684 46. Salvador LCM, Deason M, Enright J, Bessell PR, Kao RR. Risk-based strategies for
685 surveillance of tuberculosis infection in cattle for low-risk areas in England and
686 Scotland. Epidemiol Infect. 2018;146: 107–118. doi:10.1017/S0950268817001935
687 47. Brooks-Pollock E, Keeling M. Herd size and bovine tuberculosis persistence in cattle
688 farms in Great Britain. Prev Vet Med. 2009;92: 360–365.

doi:10.1016/j.prevetmed.2009.08.022

48. Silk MJ, Drewe JA, Delahay RJ, Weber N, Steward LC, Wilson-Aggarwal J, et al. Quantifying direct and indirect contacts for the potential transmission of infection between species using a multilayer contact network. Behaviour. 2018. doi:10.1163/1568539X-00003493

49. Kinsley AC, Rossi G, Silk MJ, VanderWaal K. Multilayer and Multiplex Networks: An Introduction to Their Use in Veterinary Epidemiology. Frontiers in Veterinary Science. 2020. p. 596. doi:10.3389/fvets.2020.00596

50. Hamilton L, Evans N, Allcock J. "i don't go to Meetings": Understanding farmer perspectives on bovine TB and biosecurity training. Vet Rec. 2019;184: 1–8. doi:10.1136/vr.104995

51. Campbell EL, Byrne AW, Menzies FD, McBride KR, McCormick CM, Scantlebury M, et al. Interspecific visitation of cattle and badgers to fomites: A transmission risk for bovine tuberculosis? Ecol Evol. 2019;9: 8479–8489. doi:10.1002/ece3.5282

52. Defra. Bovine TB Information Note 03 / 16 Reducing TB risks from the sale of cattle from 4-yearly testing herds in England. 2016. Available: https://www.gov.uk/government/publications/bovine-tb-information-note-0116-post-movement-testing

53. Campbell F, Strang C, Ferguson N, Cori A, Jombart T. When are pathogen genome sequences informative of transmission events? PLoS Pathog. 2018;14: 1–21. doi:10.1371/journal.ppat.1006885

54. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018. Available: https://www.r-project.org/

55. Csárdi G, Nepusz T. The igraph software package for complex network research. J Comput Appl. 2014;Complex Sy: 9. doi:10.3724/SP.J.1087.2009.02191

56. Kuhn M, Weston S, Keefer C, Engelhardt A, Cooper T, Mayer Z, et al. Classification and Regression Training. 2016. p. 198. Available: https://cran.r-project.org/package=caret

57. Stollenwerk N, Jansen VAA. Meningitis, pathogenicity near criticality: The epidemiology of meningococcal disease as a model for accidental pathogens. J Theor Biol. 2003;222: 347–359. doi:10.1016/S0022-5193(03)00041-9

58. Sharkey KJ. Deterministic epidemiological models at the individual level. J Math Biol. 2008;57: 311–331. doi:10.1007/s00285-008-0161-7

59. Keeling MJ, Ross J V. On methods for studying stochastic disease dynamics. J R Soc Interface. 2008;5: 171–181. doi:10.1098/rsif.2007.1106

60. Rossi G, De Leo GA, Pongolini S, Natalini S, Vincenzi S, Bolzoni L. Epidemiological modelling for the assessment of bovine tuberculosis surveillance in the dairy farm network in Emilia-Romagna (Italy). Epidemics. 2015;11: 62–70. doi:10.1016/j.epidem.2015.02.007

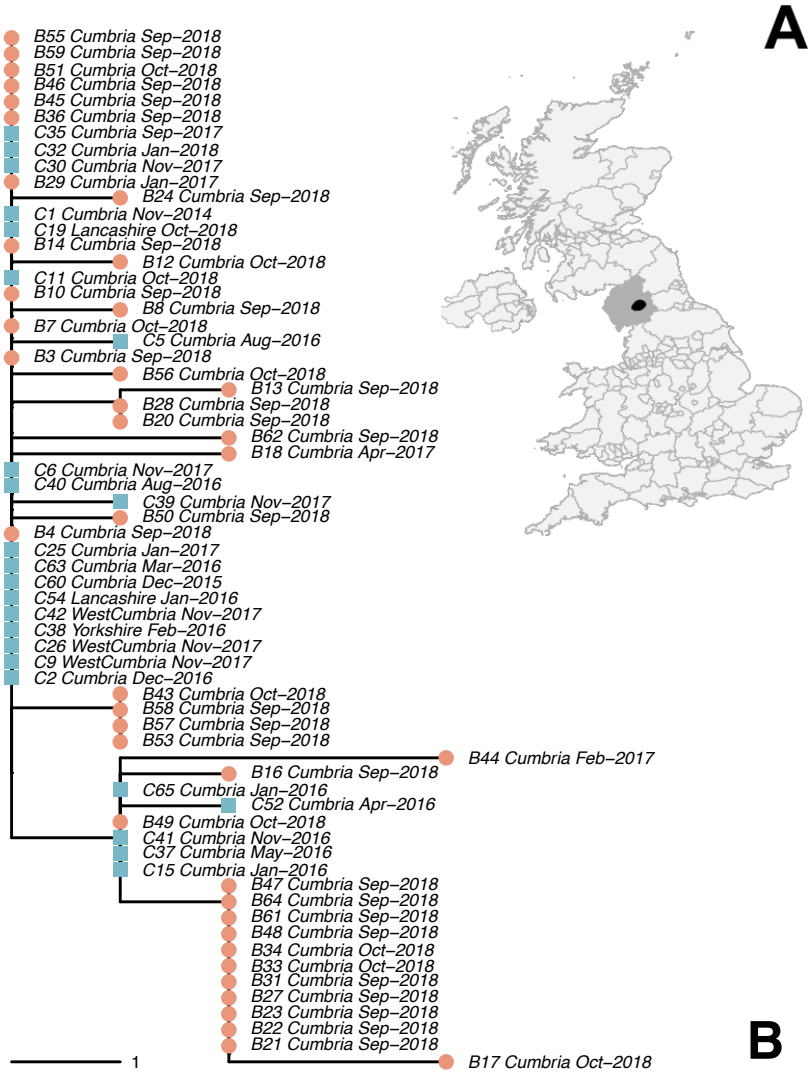
61. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 2014;10: 1–6. doi:10.1371/journal.pcbi.1003537

Tables

Category	Covariates	Description	All samples model	Cattle model	Badger model	Inter-species model
Temporal	SampleTimeDist	Time distance (in days) between the isolates sample dates	YES	YES	YES	YES
Spatial	SampleGeoDist	Geographical Euclidean distance between the isolates sampling locations	YES	YES	YES	YES
Spatial	ParcelsDist	Minimum geographical Euclidean distance between the land parcels where isolates have been sampled	YES	YES	YES	YES
Group	SameSpecies	Isolates from same species animals (YES/NO)	YES	NO	NO	NO
Group	SameGroup	Isolates from same herd or sett animals (YES/NO)	YES	YES	YES	NO
Group	Sizes	Size coefficient (squared root of product) of the isolates herd/social groups	YES	YES	YES	YES
Network	ShortestPathSNet	Shortest path length between isolates nodes in single-species network	YES	YES	YES	NO
Network	InfNodesInPathSNet	Number of infected nodes in shortest path between isolates in single-species network	YES	YES	YES	NO
Network	DegreeSNet	Degree coefficient (squared root of product) of isolates nodes in single-species network	YES	YES	YES	NO
Network	SameCommSNet	Isolates' nodes in same community in single-species network (YES/NO)	YES	YES	YES	NO
Network	ShortestPathLPNet	Shortest path length between isolates nodes in spatial (land parcels) network	YES	YES	YES	YES
Network	InfParcInPathLPNet	Number of infected nodes in shortest path between isolates in spatial network	YES	YES	YES	YES
Network	DegreeLPNet	Degree coefficient (squared root of product) of isolates nodes in spatial network	YES	YES	YES	YES
Network	SameCommLPNet	Isolates' nodes in same community in spatial network (YES/NO)	YES	YES	YES	YES
Network	ShortestPathML	Shortest path length between isolates nodes in multi-layer network	YES	YES	YES	YES
Network	InfNodesInPathML	Number of infected nodes in shortest path between isolates in multi-layer network	YES	YES	YES	YES
Network	DegreeML	Degree coefficient (squared root of product) of isolates nodes in multi-layer network	YES	YES	YES	YES
Network	SameCommML	Isolates' nodes in same community in multi-layer network (YES/NO)	YES	YES	YES	YES

735 **Table 1. Epidemiological covariates tested in the BRT algorithm.** List of epidemiological
736 covariates tested in the boosted regression trees (BRT) algorithm against the isolates genetic
737 distance, calculated in number of SNVs. The four columns indicate in which model the
738 covariate has been used.
739

740 **Figures**



741

742 **Figure 1. East Cumbria outbreak area and phylogenetic tree.** A: East Cumbria outbreak

743 area (black) location in Cumbria (dark grey). B: *M. bovis* genomes phylogenetic tree

744 (distance calculated in Single Nucleotide Variants, SNVs). Red dots represent badgers, and

745 blue square cattle, and the label report the code assigned to each individual, sampling

746 location and sample date.

747

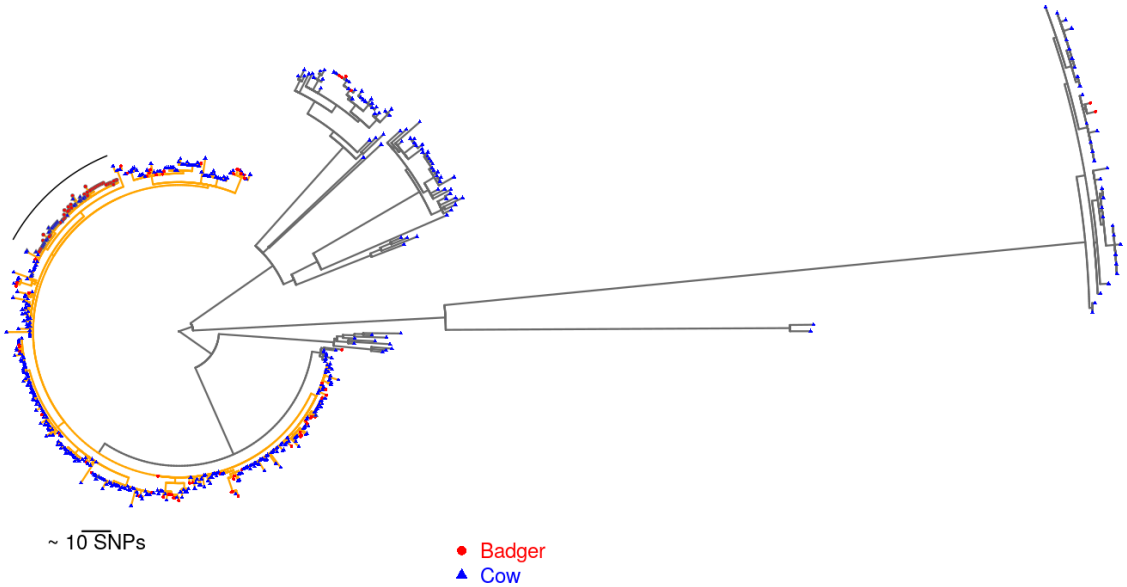


Figure 2. East Cumbria and TVR *M. bovis* phylogenies. A maximum likelihood phylogeny of the *M. bovis* genomes sourced from Cumbria and the Test, Vaccinate, or Release (TVR) area in Northern Ireland. The tree is rooted with the *M. bovis* reference genome AF2122/97. The *M. bovis* genomes sourced from infected cattle and badgers in Cumbria are highlighted with a black semi-circle at the top left. The branches of the clade containing the *M. bovis* genomes sourced from Cumbria, and those from the TVR area that are most similar is highlighted in orange.

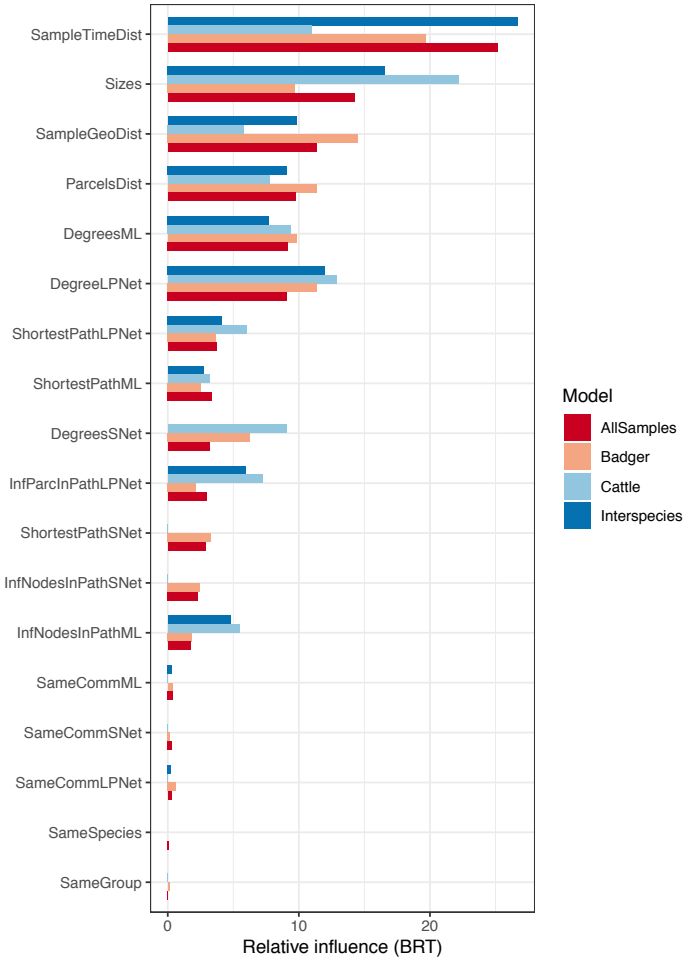


Figure 3. Covariate influence in the BRT models. The relative influence of the 18 epidemiological covariates calculated by the Boosted Regression Trees (BRT) model. Bars colours correspond to the four models run with different sub-samples of the dataset: dark-red for the full model account all samples, light red for the badger-to-badger model, light-blue for the cattle-to-cattle only model, and dark-blue for the interspecies (badger-to-cattle) model.

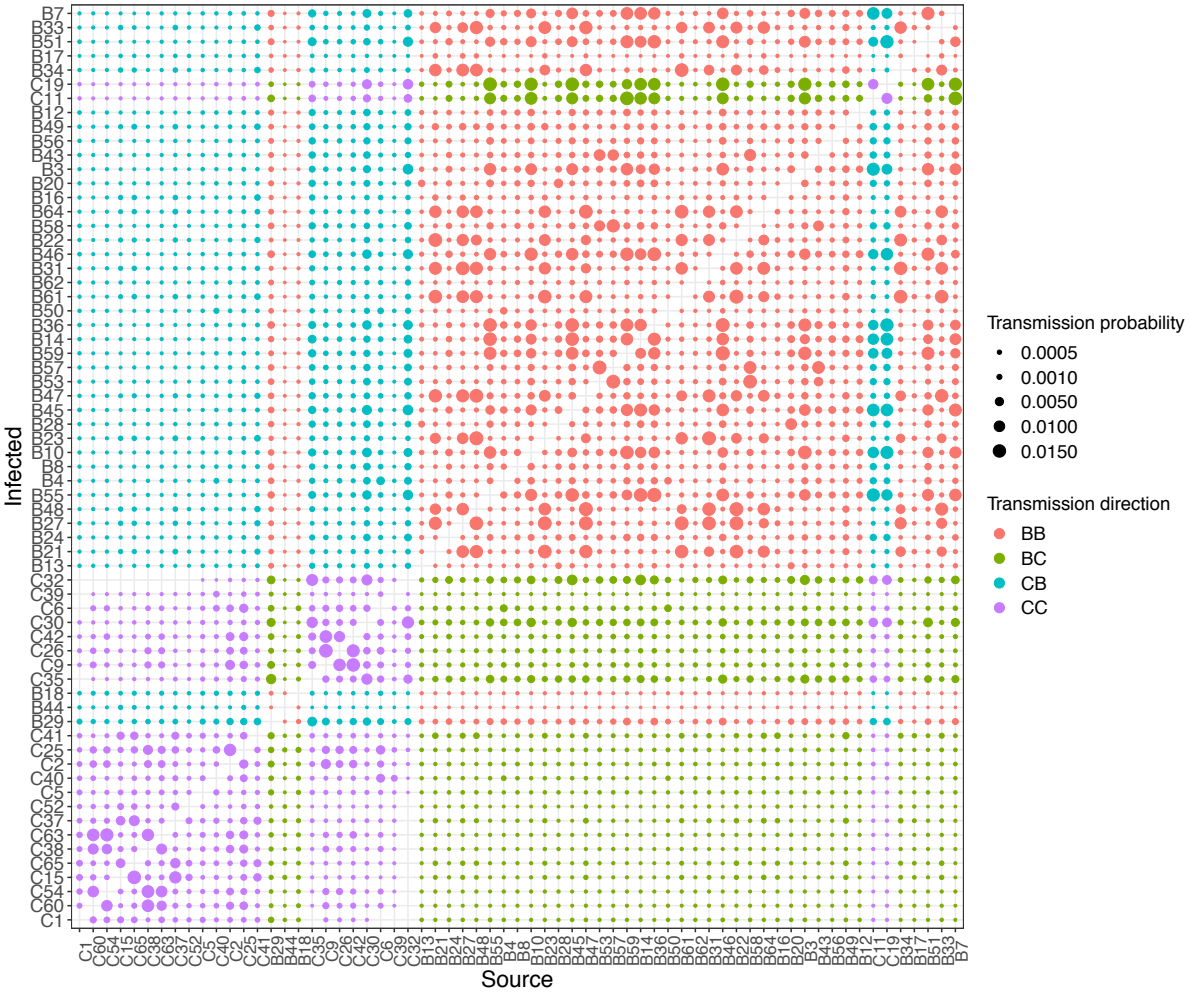


Figure 4. Transmission matrix. Pairwise transmission probabilities between infected animals in the East Cumbria outbreak. Animals are reported in *x* (source animal) and *y* (infected animal) axes in the order they have been sampled, and they are labelled from one to 65 and the species name (B for badgers and C for Cattle). Different colours correspond to transmission directions (red: badger-to-badger, green: badger-to-cow, light-blue: cattle-to-badger, and magenta: cattle-to-cattle).

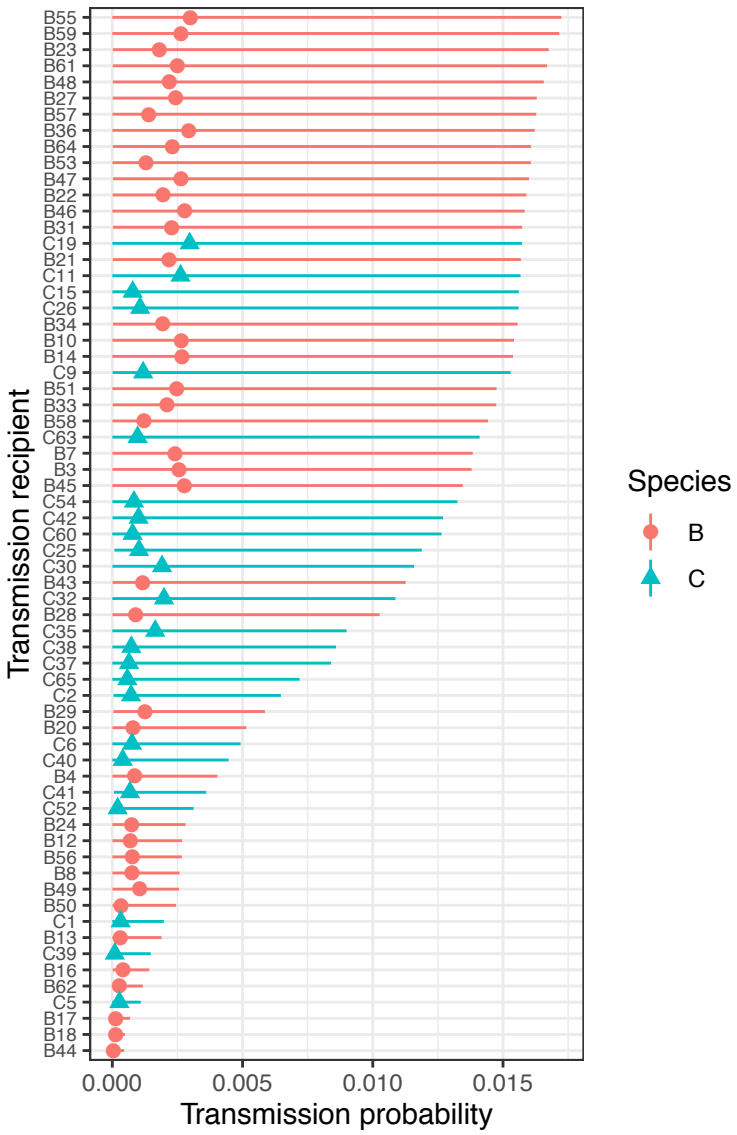


Figure 5. Transmission probability to sequenced animals. Average (dots/triangles) and range (line) of infection probabilities (x axis) from all sampled infection sources to each individual animal (y axis). Red lines/dots correspond to badgers and blue lines/triangles to cattle.

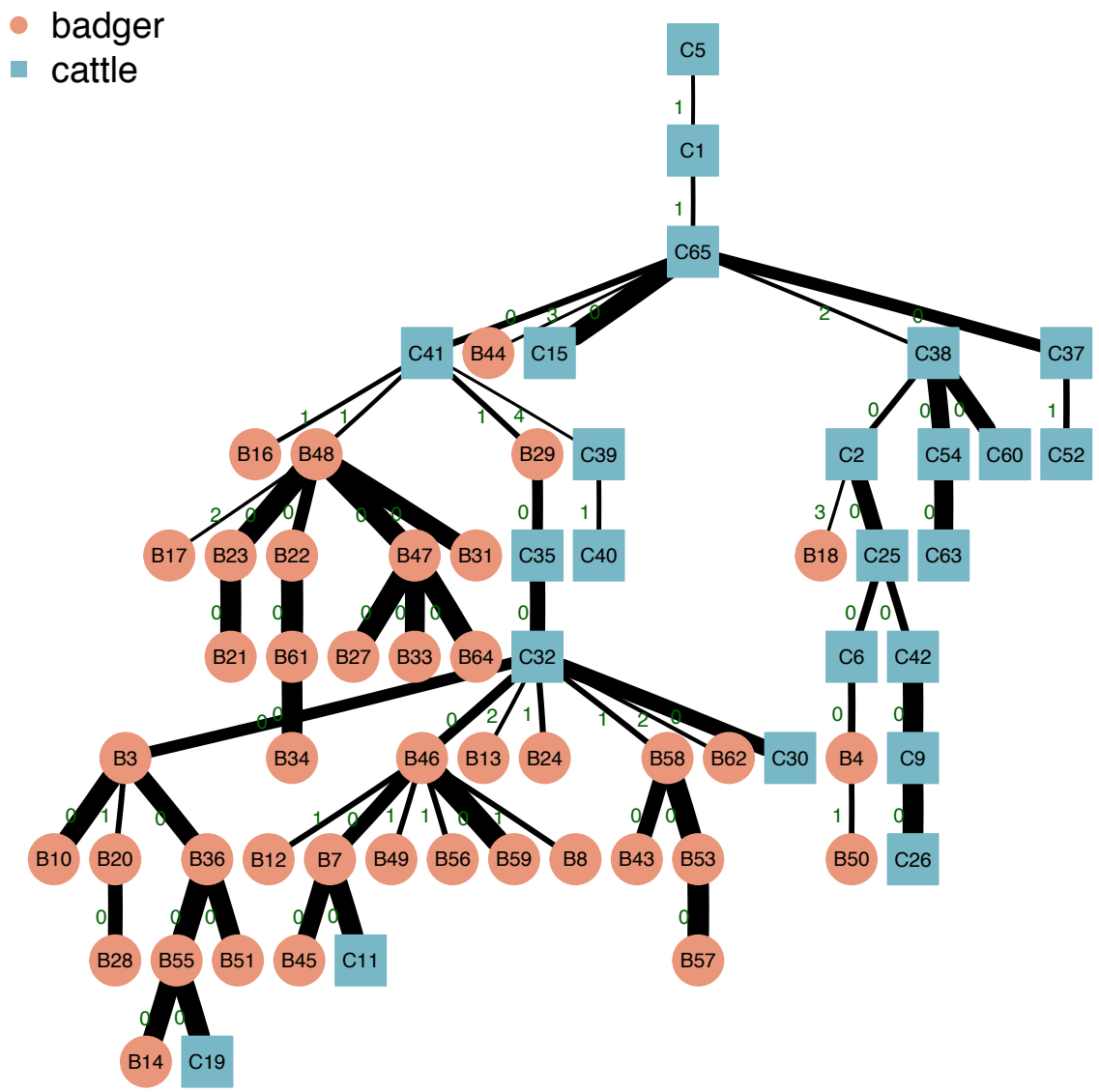


Figure 6. East Cumbria *M. bovis* outbreak transmission tree. Most likely transmission for the sampled infectious individuals in the East Cumbria outbreak (transmissions from top to bottom). Red circles correspond to badgers, and blue squares to cattle. Edge thickness is proportional to the pair transmission probability and the edge label (dark green) indicates the SNV distance between the two individuals.

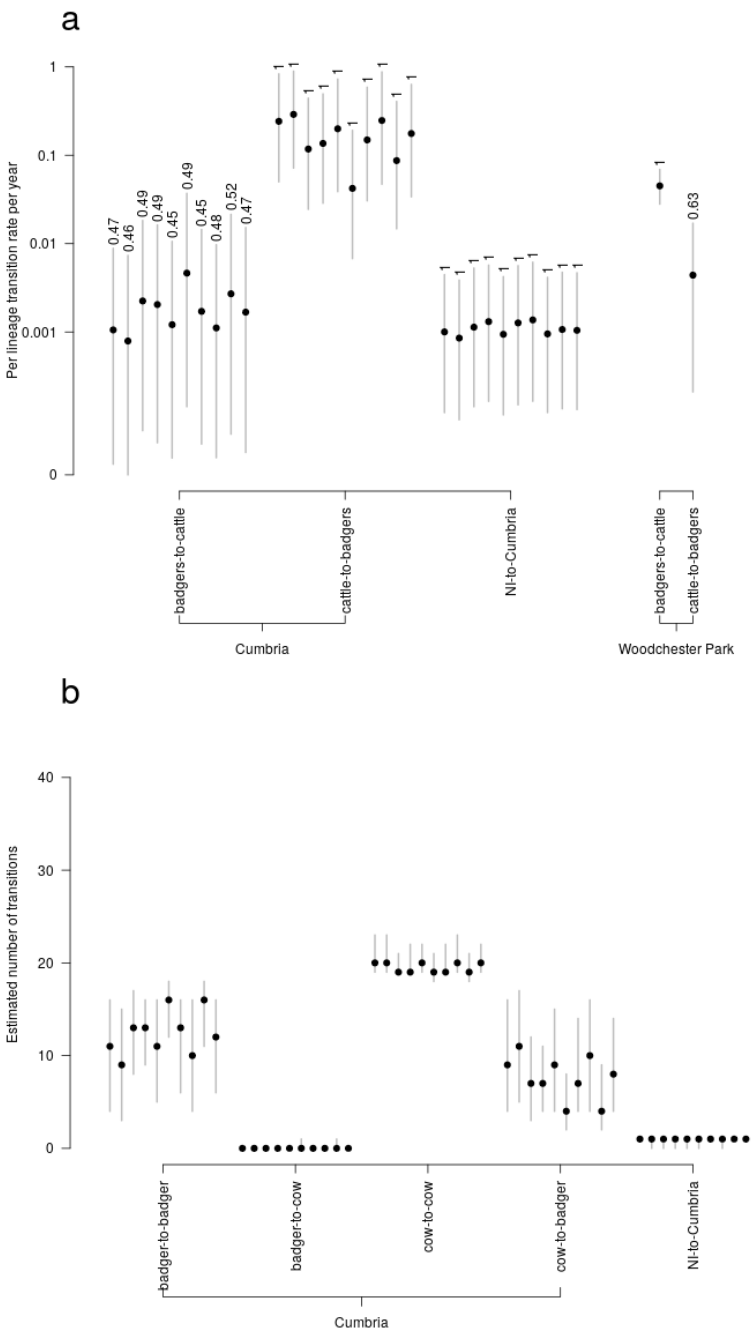


Figure 7. Transmission rates and number of transmissions estimated with BASTA. a) Inter-species and Northern Ireland to Cumbria *M. bovis* transmission rates (y axes, log scale) estimated using BASTA, based on analyses that used the genomic data and compared with the results for Woodchester Park [22]; b) estimates from BASTA of the number of transmission events (y axes) between the sampled cattle and badgers.