# Genomic epidemiology of *Mycobacterium bovis* infection in sympatric badger and cattle populations in Northern Ireland.

Assel Akhmetova[1], Jimena Guerrero[2], Paul McAdam[3], Liliana C.M. Salvador[4], Joseph Crispell[5], John Lavery[6], Eleanor Presho[6], Rowland R. Kao[7], Roman Biek[1], Fraser Menzies[8], Nigel Trimble[8], Roland Harwood[8], P. Theo Pepler[1], Katarina Oravcova[1], Jordon Graham[6], Robin Skuce[6], Louis du Plessis [9], Suzan Thompson[6], Lorraine Wright[6], Andrew Byrne[10], Adrian R. Allen[6].

1 - University of Glasgow, Glasgow, UK.

2 - Centro de Investigacion en Alimentacion y Desarrollo A.C., Hermosillo, Sonora, Mexico

3 – Fios Genomics, Edinburgh, UK.

4 – Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, USA.

5 – Foreign, Commonwealth and Development Office, Glasgow, UK.

6 - Agri-Food and Biosciences Institute, AFBI Stormont, Belfast, UK.

7 – University of Edinburgh, Roslin Institute, Edinburgh, UK

8 – Department of Agriculture, Environment and Rural Affairs (DAERA), Belfast, UK.

9 – Department of Zoology, University of Oxford, Oxford, UK

10 –Department of Agriculture Food and the Marine (DAFM),Dublin, Ireland.

# Abstract

**Background**

Bovine tuberculosis (bTB) is a costly epidemiologically complex, multi-host, endemic disease. Lack of understanding of transmission dynamics may undermine eradication efforts. Pathogen whole genome sequencing improves epidemiological inferences, providing a means to determine the relative importance of inter- and intra- species host transmission for disease persistence. We sequenced an exceptional data set of 619 *Mycobacterium bovis* isolates from badgers and cattle in a 100km$^2$ bTB 'hotspot' in Northern Ireland. Historical molecular subtyping data permitted the targeting of an endemic pathogen lineage, whose long-term persistence provided a unique opportunity to study disease transmission dynamics in unparalleled detail. Additionally, to assess whether badger population genetic structure was associated with the spatial distribution of pathogen genetic diversity, we microsatellite genotyped hair samples from 769 badgers trapped in this area.

**Results**

Graph transmission tree methods and structured coalescent analyses indicated the majority of bacterial diversity was found in the local cattle population. Results pointed to transmission from cattle to badger being more common than badger to cattle. Furthermore, the presence of significant badger population genetic structure in the landscape was not associated with the spatial distribution of *M. bovis* genetic diversity, suggesting that badger-to-badger transmission may not be a key determinant of disease persistence.

**Significance**

Our data were consistent with badgers playing a smaller role in the maintenance of *M. bovis* infection in this study site, compared to cattle. Comparison to other areas suggests that *M. bovis* transmission dynamics are likely to be context dependent, and the role of wildlife difficult to generalise.

## 1. Introduction.

*Mycobacterium bovis* infection in cattle (*Bos taurus*) and badgers (*Meles meles*) is a persistent and costly problem for the farming industries and governments of the United Kingdom and Republic of Ireland (Allen et al. 2018). In Northern Ireland alone, the bovine tuberculosis (bTB) eradication scheme cost £44 million in 2017/2018 (Northern Ireland Audit Office, 2018). The complex epidemiology of the disease is well recognised, with the role of wildlife in transmitting infection to cattle acknowledged as an impediment to eradication (Godfray et al. 2018). A major knowledge gap for this disease has been, until recently, a detailed understanding of inter-host transmission dynamics and their relative importance (Kao et al. 2016).

Multi-host zoonotic infections of slowly-evolving pathogens, such as the members of the *M. tuberculosis* complex (MTBC), present significant challenges to researchers who wish to use molecular epidemiological methods to understand disease transmission dynamics (Biek et al. 2015). Previously, multi-locus variable number of tandem repeats analysis (MLVA) and spoligotyping were used to characterise spatio-temporal patterns in bTB epidemiology (Kamerbeek et al. 1997; Skuce et al. 2010; Milne et al. 2019; Skuce et al. 2020). These methods have demonstrated how *M. bovis* infections typically present as a series of geographically localised micro-epidemics (Skuce et al. 2010; Skuce et al. 2020). However, MLVA and spoligotype loci, whilst extremely useful in defining the home ranges of endemic infections (Trewby, 2016a; Milne et al. 2018), evolve at rates considerably slower than the inter-host transmission rate, thereby limiting their utility for contemporary disease outbreak investigations (Meehan et al. 2018).

Whole Genome Sequencing (WGS) technologies and associated phylogenetic analytical frameworks, have helped to reveal sources of infection and to improve surveillance and control for various pathogens (Harris et al. 2013; Mellman et al. 2011; Walker et al. 2013). These phylodynamic methods have been most effectively applied to fast-evolving viral pathogens, whose mutation rates can, with dense sampling, permit inference of fine scale disease dynamics, over short time intervals (Volz et al. 2009; Biek et al. 2015). While the latter degree of resolution may be unobtainable for slowly evolving bacterial pathogens, recently it has been shown that provided dense sampling is undertaken across a wide temporal window, much can be revealed about inter-host disease transmission dynamics of bTB (Biek et al. 2012; Crispell et al. 2017; Salvador et al. 2019, Crispell et al. 2019; Rossi et al. 2020a and 2020b).

Biek et al (2012) were the first to apply WGS and Bayesian phylogenetics to the *M. bovis* epi-system, focusing on an emerging endemic strain found in the east of Northern Ireland. In this proof of concept study, they demonstrated ongoing transmission between sympatric cattle and badgers at the farm scale. However, due to the small number of samples and the fact that these samples were not strictly contemporaneous, it was difficult to infer which inter-host transmission rates were most important. Subsequent work in New Zealand, applied to *M. bovis* isolates from cattle and brushtail possums (*Trichorus vulpecula*), revealed that the possum population was likely the major reservoir and the primary driver for infection in cattle (Crispell et al. 2017). However, the study also revealed that biases in sampling structure could affect inferences of the relative importance of transmission routes whenever discrete traits analyses were used. More recently, Salvador et al. (2019) used WGS and Bayesian phylogenetic methods to clarify the roles that cattle, white tailed deer (*Odocoileus virginianus*) and elk (*Cervus canadensis*) play in the epidemiology of an endemic *M. bovis* problem in Michigan, USA. Elk and cattle infections were observed to cluster in space, leading to concerns that the wide-ranging behaviour of elk could be a risk for disease spread across the landscape. However, phylogenetic data clarified that deer were the primary source of inter-species transmission in the

region. In the UK, a recent study from Woodchester Park in Gloucestershire has shown that badger and cattle intra-species transmission accounts for the majority of disease dissemination events (Crispell et al. 2019). However, Bayesian structured coalescent methods show that badger-to-cattle transmission occurred up to nine times more commonly than cattle-to-badger transmission in this dataset, further clarifying the role that wildlife can play in maintaining livestock infections in this epi-system (Crispell et al. 2019). In Ireland, phylogenetic methods were recently used to show that in County Wicklow, local deer populations harbour a diversity of *M. bovis* consistent with them being a source of infection for cattle (Crispell et al. 2020).

Pathogen spread across landscapes is recognised to be an inherently spatial process (Biek and Real, 2010), leading to distinct patterns in pathogen genetic structure. Similarly, free ranging wildlife hosts, exhibit partitioning of their own genetic variation across landscapes (Guerrero et al. 2018), for example, isolation by distance (IBD) (Wright, 1943). An appreciation of how these types of landscape-genetic phenomena intersect can help to inform epidemiological investigations in wildlife populations (Biek and Real, 2010). A key question within localised bTB micro-epidemics is whether significant badger population structure is observed over smaller geographic scales in Ireland, and if so, whether it explains any of the partitioning of *M. bovis* genetic variation in the landscape.

In this study we sought to better understand the interspecies transmission dynamics of *M. bovis* and the effect of badger population structure on the spatial partitioning of pathogen diversity. Compared to some previous studies, we undertook a systematic, sympatric sampling of both cattle and badgers. The study was conducted in a small 100 km$^2$ area of Northern Ireland (Figure 1A), using data from a recently completed (2014-2018) wildlife intervention. The 'test and vaccinate or remove' (TVR) selective culling protocol applied to the local badger population (Arnold et al. 2021; Menzies et al. 2021) provided an important opportunity to systematically sample sympatric cattle and badger populations for *M. bovis*, and to apply WGS, phylodynamic and population genetic methods to determine the roles that both cattle and badgers play in the local disease dynamics.

## 2. Methods.

### 2.1 Sampling of cattle and badgers.

The TVR zone (Figure 1A), was chosen because it has a high prevalence of cattle TB (24% confirmed bTB herd prevalence over two years [2011/2012]), and is embedded within County Down, which also has one of the highest average badger densities (3.88 badgers per km$^2$) in Northern Ireland – (Reid et al. 2012).

An initial survey was conducted in 2014 to establish sett locations (DAERA, 2018). Overnight cage trapping was used across all years between the months of May and October 2014-2018. Trapped badgers were anaesthetised, trap-side, hair sampled and tested for *M. bovis* infection using the dual path platform (DPP) serology test based on the StatPak method (Chambers et al. 2009). Tracheal aspirate was taken from all trapped badgers, whether DPP-positive or -negative and sent for bacteriological culture as described below. From 2015-2018, DPP positive badgers were humanely euthanized and DPP-negative badgers were vaccinated using injectable Bacillus Calmette Guerin (BCG), and released (Menzies et al. 2021). Culled badgers underwent post-mortem examination according to a standardised protocol (Courcier et al. 2020) and specified tissues were submitted for bacteriological culture.

All cattle in Northern Ireland are TB tested annually using the standardised (European Council, 1964) single intradermal comparable cervical tuberculin (SICCT) test (Abernethy et al. 2006). Specified tissue

158 from all SICCT positive bTB reactor cattle in the TVR region from 2014-2017 were harvested at the
159 time of slaughter and submitted for bacteriological culture.

160 To provide extra temporal depth, a selection of historical isolates (n=243) from both hosts in the TVR
161 zone were re-cultured from the AFBI *M. bovis* Northern Ireland wide, strain archive. The temporal
162 window for all isolates runs from 1986 to 2017. All historical badger isolates were derived from road
163 traffic accident (RTA) post-mortems in a surveillance scheme run by the Department of Agriculture,
164 Environment and Rural Affairs in Northern Ireland (DAERA-NI) (Courcier et al. 2017). In addition, an
165 area directly neighbouring the study zone has a distinct lineage of *M. bovis* present, four isolates of
166 which were detected in the study zone. A random sample of historical isolates from this neighbouring
167 area was collected to provide additional phylogenetic context.

168 **2.2 *M. bovis* culture and genomic DNA extraction.**

169 *M. bovis* isolates were initially cultured in the liquid BD BACTEC MGIT system, solid Stonebrinks and
170 Löwenstein-Jensen media and single colonies were selected for sub-culture (Skuce et al. 2005).
171 Isolates were heat-killed in a water bath at 80°C for 30 min. DNA was extracted using standard high
172 salt/cationic detergent cetyl hexadeycl trimethyl ammonium bromide (CTAB) and solvent extraction
173 protocols (Parish and Stoker, 2001; van Soolingen et al. 2002).

174 **2.3 Spoligotyping and MLVA analysis.**

175 *M. bovis* isolates were genotyped by spoligotyping (Kamerbeek et al., 1997) and 8 locus MLVA using
176 previously described methods (Skuce et al., 2010). Authoritative names for spoligotype patterns were
177 obtained from www.mbovis.org (Smith and Upton, 2012). MLVA profiles were named using a
178 laboratory nomenclature (Skuce et al. 2005).

179 **2.4 Genome sequencing and bioinformatic analyses.**

180 Sequencing libraries were prepared using the Illumina Nextera XT method to produce inserts of
181 approximately 500-600bp. One hundred samples were sequenced at AFBI using an Illumina MiSeq
182 platform with Illumina V2 chemistry, producing paired-end reads of 250 bp. A further 100 samples
183 were sequenced at the Glasgow Polyomics facility using an Illumina Miseq producing 2×300 bp paired
184 end reads. All remaining samples were sequenced by Eurofins Scientific using an Illumina HiSeq
185 producing 2 x250 bp paired end reads. For quality assurance (QA) purposes, comparison of sequencing
186 performance across all three sites was undertaken at AFBI. Specifically, we re-sequenced 15 randomly
187 selected isolates from those sent to other institutes and compared these duplicates to the initial
188 sequencing data.

189 Reads for each sample were mapped to the recently updated/annotated (Malone et al. 2017)
190 reference genome for *M. bovis* strain AF2122/97 (GenBank accession LT708304.1) using the mapping-
191 based phylogenomics pipeline, RedDog V1beta.10.3 (Edwards et al. 2016) to identify SNP variants
192 across all isolates. Alignment and mapping were carried out in Bowtie2 v2.2.9 (Langmead et al. 2012)
193 using the sensitive local mapping setting. SNP calling was undertaken using SAMtools and BCFtools (Li
194 et al. 2009) using the consensus caller setting. A minimum depth of 10x was set for SNP calling. The
195 average coverage failure filter, average depth filter and average mapping failure filters were set at
196 98%, 10x, and 75% respectively. Transposable elements, repeat regions and the PE/PPE regions as
197 defined in the Genbank annotation, were excluded from SNP calling using the parseSNP tool in the
198 RedDog pipeline (Edwards et al. 2016).

199 Some isolates were sequenced by Glasgow Polyomics, others were sequenced by Eurofins
200 laboratories. 15 randomly selected isolates were re-sequenced by AFBI. No SNP distance was observed
201 between duplicates and the previously sequenced isolates.

202 **2.5 Badger microsatellite genotyping.**

203 Nuclear DNA extracted from all badger hair samples was genotyped at 14 microsatellite loci using
204 methods previously described (Guerrero et al. 2018). We re-profiled 5% of DNA samples as a quality
205 assurance measure.

206 **2.6 Testing for IBD in badger population.**

207 Using the package 'PopGenReport' (Adamack and Gruber, 2014) in R (R Development Core Team,
208 2020), we constructed a microsatellite distance matrix (Smouse and Peakall method) for all unique
209 badgers captured in the study zone, and for the subset of badgers that produced *M. bovis* cultures
210 from the endemic lineage (see below), to ensure we had sufficient power to detect IBD in sub-
211 populations. For both datasets, we then constructed inter-animal Euclidean distance matrices using
212 the R package, 'Geosphere' (Hijmans et al. 2019). We then performed a Mantel test, with 10,000
213 repetitions, for IBD using the package 'ade4' (Dray and Dufour, 2007) in R. For the larger meta-
214 population, Mantel tests and linear regressions were carried out for each capture year, with an
215 analysis of covariance carried out to compare the slopes of each genetic distance vs euclidean distance
216 relationship.

217 **2.7 Preliminary phylogenetic analyses.**

218 The most appropriate nucleotide substitution model for our phylogenetic analyses of the FASTA
219 alignments of informative SNPs was assessed using the 'modelTest' function of the package
220 'Phangorn' (Schliep, 2011) in R. Specifically, the fit of the General Time Reversible (GTR), Jukes Cantor,
221 and Hasegawa Kishino Yano (HKY) models to the data were assessed. The nucleotide substitution
222 model with the lowest AIC (GTR – AIC: 26860.39) was used to build a maximum likelihood phylogeny
223 using RAxML v8 (Stamatkis, 2014). The rapid bootstrapping search method in RAxML was selected and
224 stopped after 400 replicates. The phylogeny was visualized and assessed in FigTree v1.4.4 (Rambaut,
225 2018) and the ape package (Paradis and Schliep, 2018) in R. Final figures of maximum likelihood trees
226 were produced using ggtree (Yu, 2020). The presence/absence of a temporal signal in the phylogenetic
227 data of the endemic clade was assessed using the program TempEst v1.5.1 (Rambaut *et al.* 2016).
228 After selecting the reference strain (AF2122/97) as the best-fitting, outgroup root isolate, which
229 maximises the temporal signal, the root to tip divergence model was fitted using the residual mean
230 squared method.

231 To test the significance of the temporal signal in our dataset, we randomised the tip dates for the 302
232 chosen isolates from the endemic clade, in ten replicate analyses, as *per* Firth et al (2010). Tip dates
233 were randomised using the 'Tipdatingbeast' package (Rieux and Khatchikian, 2017) in R. The original
234 dataset and each replicate were subjected to a simplified BEAST 2 (Bayesian Evolutionary Analysis by
235 Sampling Trees - Bouckaert et al. 2014), constant population, unstructured coalescent analysis, using
236 a relaxed clock model, the GTR nucleotide substitution model and a chain length of 200,000,000
237 MCMC steps, of which 10% was discarded as burn-in. After checking convergence in Tracer 1.7.1 –
238 (ESS > 200 for all parameters), the median substitution rate for each replicate was compared to those
239 of the non-randomised data set. In addition, we also compared the substitution rates to those inferred
240 from the BASTA and MASCOT analyses (see below) to verify that not taking population structure into
241 account did not result in biased clock rate estimates.

242

243 **2.8 Structured coalescent analyses to determine inter-species *M. bovis* transition rates.**

244 To facilitate the goal of determining the transmission dynamics of bTB in the TVR zone, existing
245 spoligotyping and MLVA data were used to rationally target the historically endemic strain family in
246 this region (Skuce et al. 2010; Skuce et al. 2020). Specifically, we applied two Bayesian structured
247 coalescent methods, using BEAST2 (Bouckaert et al. 2014; Barrido-Sottani et al. 2017). Specifically, we
248 used BASTA (DeMaio et al. 2015) (Bayesian Structured Coalescent Approximation - De Maio et al.
249 2015) and MASCOT (Marginal Approximation of the Structured Coalescent - Müller et al. 2018)) which
250 can estimate both ancestral population structure and transition rates in the presence of biased
251 sampling. To ensure valid comparisons with previously investigated bTB epi-systems in the UK, we
252 used BASTA as *per* Crispell et al (2019) and Rossi et al (2020a and 2020b). In addition, we used MASCOT
253 to estimate the effective population size / diversity of the *M. bovis* population in both cattle and
254 badger hosts.

255 Our study population for estimation of transition rates, was a subset of 302 isolates from the endemic
256 lineage defined below. We randomly selected 248 cattle isolates (one isolate per annum per herd)
257 and 54 isolates from unique badgers. All 302 isolates were collected in the period 1986-2017. We
258 implemented BASTA and MASCOT analyses in BEAST2 to determine inter-species transition rates. We
259 used uncorrelated lognormally distributed relaxed clock models, as these have been observed to best
260 describe the rates of molecular evolution in *M. tuberculosis* and *M. bovis* in previous studies (Crispell
261 et al. 2017; Crispell et al. 2019; Menardo et al. 2019). Each BASTA and MASCOT analysis was run with
262 three replicates. For each BASTA analysis, a chain length of 300,000,000 MCMC steps was set,
263 sampling every 15,000 steps. For each MASCOT analysis, a chain length of 500,000,000 MCMC steps
264 was set, sampling every 50,000 steps. Full details of analyses are supplied in the xml files in
265 Supplementary files. Following the removal of a 10% burn-in, chains were combined using
266 LogCombiner v2.6 (Bouckaert et al. 2014) and analyses were compared based upon the log likelihood
267 scores, model convergence and posterior support of parameters in Tracer v1.7.1 (Rambaut et al.
268 2018). Maximum Clade Credibility (MCC) trees were constructed for combined chains using
269 TreeAnnotator v2.6 (Bouckaert et al. 2014)) using the median ancestor heights criterion. BASTA log
270 and tree output files were used to estimate inter-species transition rates and inter- and intra-species
271 transition counts, using custom R and Java scripts which analysed 10,000 trees from each of the model
272 runs as per the method of Crispell et al (2019) – these scripts are provided as Supplementary files.

273 **2.9 Graph methods to assess inter-species transitions.**

274 To reconstruct the transmission tree of the set of 302 *M. bovis* isolates collected from badgers
275 and cattle between 1986 and 2017, we used the graph method *SeqTrack* (Jombart et al.,
276 2011). *SeqTrack* infers the most likely genealogy from genomic and temporal data. The matrix
277 of pairwise genetic distances between *M. bovis* isolates, the isolates' date of collection and
278 associated host information, an estimated *M. bovis* substitution rate, and the sequences total
279 (in number of SNPs) were given as input. *SeqTrack* produced a *M. bovis* genealogy tree, with
280 the number of SNP differences associated to the branch between each ancestor and
281 descendant. The output files were used to compute the distribution of *M. bovis* genetic
282 distances between ancestors and descendants, as well as the percentage of intra and inter
283 species transmission events. To check the effect of *M. bovis* substitution rates on the ancestral
284 history, we ran the analysis with values from previous *M. bovis* genomic studies in NI (0.15
285 and 0.20 substitutions per genome per year from Biek et al. 2012 and Trewby et al. 2016b,

respectively, and 0.35 from this work (Section 3.5). *SeqTrack* was implemented in R using the adegenet package (Jombart and Ahmed, 2011) (scripts provided as Supplementary files).

### 2.10 Assessing effect of badger population structure on *M. bovis* spatial partitioning.

We sought to determine if *M. bovis* inter-isolate SNP distance is associated with pairwise Euclidean distance between trapped badger locations, pairwise difference in time of *M. bovis* isolation and pairwise host genetic distance.For *M. bovis* genome sequences derived from badgers infected with the endemic lineage, we constructed two inter-isolate distance matrices: (i) SNP distance using the R package 'ape' (Paradis and Schliep, 2018); (ii) time of isolation difference using the 'dist' function in R.

Using these *M. bovis* distance matrices and the two already produced to assess IBD in the infected badgers (see above), we performed a multiple regression on distance matrices (MRM) analysis using the R package 'ecodist' (Goslee and Urban, 2007) with 10000 repetitions.

## 3. Results

### 3.1 Sampling of cattle and badgers.

A total of 642 *M. bovis* isolates were used in this study, of which 611 were collected from badgers and cattle in the TVR zone, and 31 from a neighbouring region. Of the 642 isolates, 15 were sequencing QA duplicate controls as discussed in section 2.4. In addition, we re-sequenced the *M. bovis* reference sample, AF2122/97 as an internal control. Of the 642 survey isolates, 399 (282 cattle; 117 badger) were sampled contemporaneously in the TVR zone during the project (2014-2017). A further 242 historical isolates (232 cattle; 10 badger=10) from 1986-2013 were available from archived cultures from the zone and its environs. Cattle isolates across all years were single isolates per animal from 185 herds. Multiple isolates (n=86) were cultured from 24 individual badgers, with single isolates derived from a further 36 badgers. In total, between 1986 and 2017, we collected *M. bovis* isolates from 60 unique badgers and 483 unique cattle (Figure 1B). Full details of sample locations, year of isolation and species of origin are given in Supplementary Data 1.

### 3.2 Spoligotyping and MLVA analysis.

22 MLVA types and 6 spoligotypes were observed in the 642 isolates. From prior analyses (Skuce et al. 2010), the spoligotype and MLVA genotypes could be grouped into eight related 'strain families'. Each is dominated by a probable founder genotype (source of the family name) with related daughter strains varying by spoligotype and / or MLVA polymorphism. Numbers of isolates per strain family are shown in Table 1A. The MLVA 6, spoligotype 263 family (6.263) was considered to be endemic in the region and accounted for most of the observed isolates (Skuce et al. 2010; Skuce et al. 2019). The remaining seven strain families (1.140, 2.142, 3.140, 4.140, 5.140, 19.140 and 20.131) were not likely to be endemic in the TVR cull zone as each has a home range elsewhere in Northern Ireland (Skuce et al. 2010; Trewby, 2016a). Of the 36 isolates from strain family 20.131, 32 were collected from a region neighbouring the TVR zone, 4 badger isolates were found within the zone. Full details of isolate MLVA genotypes and spoligotypes are supplied in Supplementary Data 1.

### 3.3 Sequencing, bioinformatic analyses and Quality Assurance.

The RedDog pipeline was used to process the isolates. 24 (22 cattle and 2 badgers) failed the sequencing QA filters (98% coverage filter for the reference genome) and were excluded. The remaining 618 survey isolates plus AF2122/97 control passed all QA filters. Detailed QA meta-data for all 619 isolates are given in Supplementary Data 2. Forward and reverse reads for all QA passing

327 isolates are deposited at the National Centre for Biotechnology Information (NCBI) Sequence Read
328 Archive (SRA), bio-project XXXXXXXX. Accession numbers for all reads are given in Supplementary Data
329 2.

330 From the 619 isolates with good quality sequence reads, 1562 SNPs passed QA calling rules and were
331 used to conduct phylogenetic analyses. Details of all SNPs passing QA, and their location in the
332 reference sequence are given in Supplementary Data 3.

333 The AF2122/97 control exhibited a 3 SNP distance from the reference sequence (Genbank LT708304.1)
334 likely due to accrual of mutations after culture passages at AFBI.

335 **3.4 Badger microsatellite genotyping and IBD analyses.**

336 769 unique badgers were captured, location recorded, sampled and successfully genotyped between
337 2014 and 2018. Random 5% re-genotyping for QA purposes produced identical allele calls.
338 Microsatellite profiles, capture locations and date of capture can be found in Supplementary Data 4.
339 Summary population genetic statistics for all animals, per year are collated in Supplementary Table
340 S1.

341 Samples from 45 badgers produced positive *M. bovis* cultures. Spoligotyping and MLVA placed them
342 in the major endemic lineage in the study zone. Capture locations for the 45 endemic strain positive
343 badgers are illustrated in Supplementary Figure S1A.

344 Across all years, the badger population exhibited consistent levels of IBD, as indicated by significant
345 Mantel tests ($r$=0.11-0.17, $p<0.05$) (Table 2). The slopes of the relationships between genetic distance
346 and Euclidean distance were very similar. Small, significant differences were observed in the ANCOVA
347 (Supplementary Figure S2 and Table 2), mainly due to the increased slope of the IBD relationships
348 years 2015, 2016, and 2018, consistent with badger genetic differentiation being observed over
349 shorter distances.

350 The 45 *M. bovis* culture positive badgers also exhibited significant IBD (Table 2) similar to that of the
351 larger study population.

352 **3.5 Preliminary phylogenetic analyses.**

353 **3.5.1. All isolates.**

354 The Maximum Likelihood tree constructed in RAxML for all 619 sequenced isolates is shown in Figure
355 2. The phylogeny was rooted using the 20.131 strain family as an out-group, as these isolates are
356 known to derive from an older common ancestor than other extant strains (Allen et al. 2013). Eight
357 major lineages, each with high bootstrap support, were observed in the phylogeny. The eight strain
358 families defined previously by MLVA and spoligotyping were monophyletic and in perfect concordance
359 with the SNP based tree topology. We determined the within-lineage diversity, as defined by total
360 number of SNPs recorded, for each of the eight major lineages observed (Table 1B). Additionally, from
361 distance matrices generated during phylogenetic analyses, pairwise, inter-isolate SNP distance,
362 summary statistics were computed within all eight major lineages (Table 1C).

363 **3.5.2 Endemic lineage – 6.263.**

364 The endemic major lineage of 6.263 presented the best opportunity to investigate *M. bovis*
365 transmission dynamics among cattle and badgers in the TVR zone. 6.263 has been consistently
366 associated with the study area for over two decades in local cattle and badgers, unlike lineages whose
367 home range is elsewhere in Northern Ireland (Skuce et al. 2020). A higher resolution SNP phylogeny
368 of 6.263 is shown in Figure 3, using the subset of 302 isolates described above. Cattle and badger

369 isolates were observed in all sub-lineages of the endemic lineage, with no sub-lineages made up of
370 isolates exclusively from a single host species. Major sub-lineages had good bootstrap support (>90).
371 A bar-plot showing frequency of isolates per host species taken from the endemic clade over the
372 period 1986-2017, is presented in Supplementary Figure S3. A smaller maximum likelihood phylogeny
373 of the 45 endemic clade badgers is presented in Supplementary Figure S1B.

374 Badger isolates were predominantly found in the endemic 6.263 lineage. No badger isolates were
375 found in the 1.140, 2.142, 3.140, 5.140 and 19.140 lineages. Four isolates, from two unique badgers,
376 sampled in the TVR zone, were found in the 20.131 lineage, whilst five isolates from four unique
377 badgers were found in the 4.140 lineage (Supplementary Figures S4 and S5). A full breakdown of
378 number of isolates from each species in each lineage is shown in Table 1A.

379 The n=302 endemic lineage phylogeny analysed in TempEst, rooted against the AF2122/97 reference
380 genome, exhibited a positive correlation between genetic divergence (root to tip distance) and
381 sampling time, with moderate evidence of temporal signal ($R^2$=0.25; p<0.001), and a conservative
382 clock rate of approximately 0.22 substitutions per genome, per year - see Supplementary Figure S6A.
383 All ten tip date randomised replicate datasets run using the simple unstructured coalescent model,
384 exhibited similar substitution rates, all of which were considerably lower than the substitution rate
385 inferred on the non-randomised data set (0.35 substitutions per genome, per year 95% HPD: 0.28-
386 0.43) and exhibited no overlap in 95% highest posterior density (HPD) (Supplementary figure S6B).
387 Furthermore, the BASTA and MASCOT analyses showed similar substitution rate estimates (see below)
388 to the non-randomised data set under a simple unstructured coalescent model. This confirmation of
389 the presence of a temporal signal permitted further analyses using structured coalescent methods
390 within BEAST2 (Firth et al. 2010).

391 **3.6 Structured coalescent analyses – BASTA and MASCOT.**

392 The three replicate BASTA and MASCOT chains converged to similar intra-method values across all
393 parameters. Log-combined Data for derived parameters are shown in Table 3. Substitution rates are
394 presented in Supplementary Figure S6B.

395 BASTA indicated a median substitution rate of 0.38 substitutions per genome per year (95% HPD
396 interval 0.27-0.49). MASCOT indicated a similar rate of 0.34 substitutions per genome per year (95%
397 HPD interval: 0.28-0.42).

398 BASTA and MASCOT estimates for median time to the most recent common ancestor (tMRCA) of the
399 endemic lineage were 47.93 years (95% HPD 32.6-85.3) and 37.3 years (95% HPD 31.3-49.4) before
400 the last date of sampling (in 2017), respectively - suggesting an emergence for endemic lineage 6.263
401 in the late 1960s to early 1980s.

402 BASTA estimated the median badger-to-cattle interspecies transition rate to be 0.03 transitions per
403 lineage per year (95% HPD interval: 0.00002-0.2039). In contrast, the cattle-to-badger median rate
404 was approximately three times larger, at 0.098 transitions per lineage per year (95% HPD interval:
405 0.0006-0.154) – see Figure 4A. MASCOT estimated badger-to-cattle transitions to be 0.86 transitions
406 per lineage per year (95% HPD interval: 0.55-1.24), and cattle-to-badger median rate to be
407 approximately three times larger, at 2.73 transitions per lineage per year (95% HPD interval: 0.53-
408 5.83) – see Figure 4B.

409 Time stamped maximum clade credibility (MCC) trees for the BASTA and MASCOT analyses are
410 presented in Supplementary Figures S7 and S8.

411 MASCOT-derived estimates of the *M. bovis* effective population size ($N_e$) were also different between
412 hosts, suggesting that the vast majority of pathogen diversity was found in cattle (median $N_e$ = 465.76;
413 95% HPD: 118.13-1173.25) compared to badgers (median $N_e$ = 1.98: 95% HPD: 0.25-7.52) – see Figure
414 4C.

415 BASTA-derived counts of intra- and inter-species transitions are discussed in the Supplementary
416 Material and presented in Supplementary Figure S9.

417 **3.7 Graph methods – inter-species transitions.**

418 The *SeqTrack* analysis inferred the ancestry of the 302 *M. bovis* isolates by distinguishing
419 (based on date of collection and genetic distances), ancestors from descendants (Figure 5). It
420 identified 68 ancestors, of which 56 were associated with cattle and 12 with badgers. The
421 number of cattle descendants was more than twice the number of badger ones ($\mu$=4.41,
422 $\sigma$=8.44 and $\mu$=1.93, $\sigma$=1.89, respectively). The inferred ancestry displayed very few mutations
423 between ancestors and descendants (maximum of 11 mutations with $\mu$=1.66, $\sigma$=1.95)
424 suggesting direct transmission links between isolates (Supplementary Fig S10). From these,
425 71.8% were within cattle, 13.6% were from cattle to badgers, 10.3% were from badgers to
426 cattle, and 4.3 % were within badgers (Supplementary Fig S11). Running the analysis with
427 different substitution rate values did not influence these results (results not shown).

428 **3.8 Effect of badger population structure on *M. bovis* spatial partitioning.**

429 From the full model in the MRM analysis, modeling *M. bovis* genetic distance (SNP-based) as a function
430 of inter badger genetic distance, inter badger Euclidean distance and inter *M. bovis* time of isolation
431 difference, we observed that only inter-badger Euclidean distance was significantly associated with
432 *M. bovis* genetic distance (p = 0.04). However, the overall fit of the model was non-significant (F-test
433 p >0.05). Badger microsatellite derived genetic relatedness was therefore not associated with *M.*
434 *bovis* SNP-derived genetic differentiation. A full summary of MRM findings is presented in
435 Supplementary Table S2.

436

437 ## 4. Discussion

438 We applied Bayesian phylogenetic methods to investigate transmission dynamics in a systematically
439 sampled, multi-host, endemic disease. Our data shed further light on the intra- and interspecies
440 dynamics of *M. bovis* transmission in an endemic area. They are also useful for informing control
441 policies and comparison to epi-systems in different regions. From the latter, novel insights into drivers
442 of disease transmission dynamics may come to light.

443 Our structured coalescent analyses suggest that the endemic 6.263 lineage has been present in the
444 TVR zone since the 1960s-1980s. The rate of molecular evolution of 0.34 to 0.38 substitutions per
445 genome per year is consistent with previous *M. bovis* phylodynamic studies – see Supplementary
446 Table S3.

447 The presence of bacterial isolates from both cattle and badgers throughout all sub-lineages of the
448 endemic 6.263 lineage indicates likely bi-directional transmission between both species. The
449 structured coalescent BASTA and MASCOT models both suggest that infection in this region is driven
450 primarily by cattle. Point estimates for transition rates from both analyses, and raw transition counts
451 from BASTA do exhibit considerable uncertainty, however, when taken alongside additional data on
452 $N_e$, seqtrack outputs and badger genetic structure (see below), they are consistent with a reduced role

453 for badgers in this epi-system. While BASTA and MASCOT models had similar magnitudes of difference
454 in inter-species transition rates and point to their same relative importance, point estimates of the
455 transition rates were considerably different. This is perhaps not surprising, since while both are
456 approximations of the structured coalescent, they make considerably different assumptions about
457 how the coalescent process affects the shape of the phylogenetic tree from which transition rates are
458 calculated (Müller et al. 2017). Estimates of raw transition counts both between- and within-hosts
459 (see Supplementary material) are broadly congruent and point towards the study area being a
460 primarily cattle-driven, epi-system. An important caveat when using Bayesian structured coalescent
461 methodologies, is that transmission events and transitions are subtly different variables (Crispell et al.
462 2019) – see the Supplementary Materials for some discussion on this and sampling
463 representativeness.

464 Estimates of the effective population size from MASCOT models consistently indicated that most of
465 the pathogen diversity is in cattle. The graph methods were largely in agreement with the
466 reconstructed phylogenies, suggesting that the majority of bacterial ancestors were found in cattle
467 hosts and that after within-cattle transmission events, transmissions from cattle to badgers were the
468 main contributors to this epi-system. The graph methods complemented the phylogenetic analyses
469 by their ability to infer correct genealogies of isolates in densely sampled areas such as this one. This
470 method considers that ancestries can be inferred between sampled isolates within the given dataset
471 following three assumptions: 1) each isolate can only have one ancestor; 2) ancestors must have an
472 earlier collection date than their descendants; and 3) the likelihood of the ancestries can be indicated
473 by the genetic differences among isolates. If any of these assumptions are violated, the method could
474 provide wrong results, however, if this bias exists, it would also affect the graph-based and the
475 phylogenetically analyses. The maximum number of SNP differences between isolates was 11, with a
476 mean number of 1.66, suggesting that most (if not all) transmission events were direct. This result is
477 not surprising since this study was performed in a small area in which an endemic lineage is present.
478 However, it also highlights the quality of data sampling during the study period.

479 These independent lines of evidence suggest that the *M. bovis* strains found in badgers in the area
480 represent a limited sample of the diversity found in the cattle *M. bovis* population, and likely result
481 from recent cattle-to-badger transmission. Indeed, the raw transition count data (Supplementary
482 Table S4 & Supplementary Figure S9) suggest that the number of cattle-to-badger transitions is
483 approximately double that of badger-to-badger transitions, suggesting that spill-over events from
484 cattle are having a greater impact on badger TB status than within-species transmission events.

485 Badger genetic population structure, whilst remaining stable over the intervention period, had no
486 association with how *M. bovis* genetic diversity was spatially distributed. It is possible that this lack of
487 association is due to factors other than reduced badger to badger intraspecies transmission dynamics.
488 The endemic lineage, if it were a relatively recent incursion, may have had little time to establish foci
489 of persistent infection in badgers, and diffuse across the landscape through philopatric contact
490 networks. This lineage has, however, been present in the region for ~40-50 years, providing ample
491 time for establishment to occur. Alternatively, perturbation of the badger population, and associated
492 dispersal arising through the application of culling, even at a small-scale (Bielby et al. 2014), may have
493 served to obscure any association between pathogen and host population structures. The relative
494 stability of the IBD relationship we observe, and the fact that during the culling period of 2015-2018,
495 genetic differentiation occurs over shorter distances than the survey-only year of 2014, are consistent
496 with no perturbation signal. GPS collar data from badgers from this same region have also suggested
497 badger ranging did not increase significantly in cull years compared to survey (O'Hagan et al 2021).
498 Given these observations, and the results of the structured coalescent and graph methods, we

postulate that our badger genetic data further support the hypothesis of a relatively reduced role for badgers in bTB transmission and persistence in this region, compared to cattle.

Our data contrast starkly with the findings of Crispell et al (2019) from the Woodchester Park region of Gloucestershire, GB, that found badger-to-cattle transitions were up to nine times more common than cattle-to-badgers. Without employing detailed, comparative methods, it is difficult to definitively understand why transmission dynamics between the two regions are so divergent. However, it may be due to differences in host density and geography. The Woodchester Park badger population is one of the densest in Europe, with an average of 30-40 badgers per $km^2$ during the period covered by the recent phylodynamic study (Delahay et al. 2013; Crispell et al. 2019). The TVR region's badger population is approximately eight to ten times less dense than this - 3.88 badgers per $km^2$ from a County Down wide survey (Reid et al. 2012), and ~5.6 badgers per $km^2$ as assessed in the TVR study (Menzies et al. 2021), and more comparable to the average badger density observed in the UK Randomised Badger Culling Trial (RBCT - 3.92 badgers per $km^2$ (ISG, 2007)). Conversely, the cattle density in the immediate vicinity of Woodchester Park, in Gloucestershire, is quite variable with estimates ranging from 25-100 cattle per $km^2$ (APHA, 2017). Northern Ireland is recognised as having some of the highest cattle densities in western Europe with recent estimates suggesting an average of 112 cattle per $km^2$ (Allen et al. 2018). The landscapes are also geographically different, with the Irish landscape made up largely of pastureland interspersed with hedgerows that are the primary location for badger setts (Reid et al. 2012), while Woodchester Park consists of prime badger woodland habitat (Delahay et al. 2013).

Interestingly, the transmission dynamics we describe here are very similar to those reported by Rossi et al (2020b) in a recently introduced incursion of bTB into the low-risk area of Cumbria in northern GB. Specifically, they observed an initial phase of primarily cattle-to-cattle transmission, spilling over into local badgers. Whilst exact estimates of GB wide badger density are not available (Rossi et al. 2020b), it is likely that density in Cumbria is considerably lower than that observed in Woodchester Park, and comparable to that observed in our study area - perhaps in the order of 0.3 to 2.5 badgers per $km^2$ (Judge et al. 2017). It is interesting that the transition counts between hosts are very similar given that the Cumbria outbreak has not persisted as long as the endemic situation in the TVR area.

Host density is a major driver of *M. bovis* persistence (Allen et al. 2018) and it is conceivable that the relative densities of cattle and badgers in different regions, alongside other factors, may affect transmission dynamics. The observed disparity between epi-systems suggests that there may be no simple bTB transmission paradigm on which to base all interventions. Such heterogeneity in regional disease epidemiology may well call for a more heterogeneous approach in the application of disease eradication schemes.

An important consideration from this study is that while our data are supportive of badgers playing a lesser role in intra-species bTB transmission dynamics in this region, this may not be capturing the full impact of badger-to-cattle transmission, which seeds new infection into herds. Subsequent within herd 'amplification' by cattle-to-cattle transmission may mean the initial seeding event has greater impact than that standalone event, results in an outsized contribution to disease spread, as has been postulated before from RBCT data. While the badger-to-cattle contribution was estimated at 5.7%, this was modelled to amplify to ~52% (bootstrap 95% CI: 9.1 – 100%), although confidence intervals were very wide (Donnelly and Nouvellet, 2013). Owing to the low rate of molecular evolution observed in *M. bovis*, the phylogenetic methods we employ lack the resolution over shorter outbreak time scales to identify such amplification events and inform on their impact.

Any study of a multi-host infectious disease system in the field will have biases that compromise the representativeness of the sample set assembled. However, we believe the systematic sampling undertaken here, and analytical approaches taken, largely mitigate these issues. See Supplementary materials for a full discussion.

A separate issue with our findings is that they come from a badger population undergoing selective culling and vaccination, both interventions that are likely to affect interspecies disease transmission dynamics (Buddle et al. 2018; ISG, 2007). However, any study seeking to harvest systematically sampled, culturable *M. bovis* from wildlife, would have involved disturbance and culling of badgers for post-mortem and pathogen isolation. The application of vaccination is admittedly a different matter however, and without the necessary non vaccinated control population in which to study transmission dynamics, we are unable to determine the likely impact of the 5 years of vaccination.

## Conclusions

We describe how in a small region of Northern Ireland, cattle-associated transmission appears to drive bTB disease dynamics. Of the inter-species transitions estimated, cattle-to-badger transitions were three times more common than badger-to-cattle. We contrast this to the recently published work of Crispell et al. (2019) from Woodchester Park, in which starkly different transmission dynamics were observed. There may be regional heterogeneity in the epidemiology of bTB. Further work in other regions of the UK and Ireland is required to assess just how heterogeneous disease dynamics may be. If substantial heterogeneity is observed, it may be advisable for different regions to adopt bespoke eradication schemes tailored to the prevailing host dynamics in their areas, leading to superior control outcomes.

## Acknowledgments

Assel Akhmetova is supported by a Bolashak International Scholarship. This work was funded by the Department of Agriculture, Environment and Rural Affairs for Northern Ireland (DAERA-NI) through its 'Evidence and Innovation' programme – project no. 15/3/07. Additional funding was provided by the UK's Biotechnology and Biological Sciences Research Council (BBSRC) – grant numbers BB/P0105598 and BB/M01262X. The funders had no role in study design, analysis or decision to publish. The authors wish to recognise Shane Collins and Carl McCormick and their team for supervising all field work and AFBI Disease Surveillance and Bacteriology teams who provided excellent post-mortem examinations and mycobacteriology support. Thanks to Dr Dez Delahay of the Animal and Plant Health Agency (APHA) in GB for advice on surveying local badger populations and trapping. Thanks to Dr Nicola Müller for her advice on the use of MASCOT and to Dr Nicola DeMaio for his advice on preparing XML files for BASTA.

## Research Ethics

All badger field work was carried out under licences issued by the Northern Ireland Environment Agency. All scientific procedures performed on badgers were conducted according to the guidelines of the Animals Scientific Procedures Act (ASPA - Licence 2767) overseen by the Department of Health for Northern Ireland.

## Conflict of interest

The authors declare no conflict of interest in the production of this work.

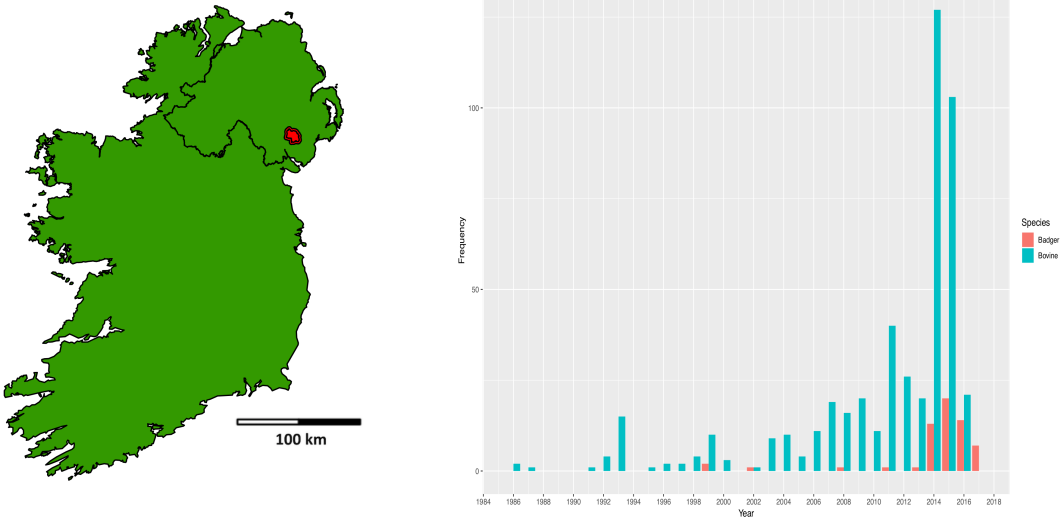
## Supplementary Files

R scripts and BEAST xml run files used in the performance of this work are curated at the following GitHub repository:

https://github.com/AdrianAllen1977/Genome-epidemiology-of-Mycobacterium-bovis-infection-in-contemporaneous-sympatric-badger-and-cattle

612 Figures

1A 1B




613
614 **Figure 1 –** A: Map of Ireland with location of 100 km$^2$ TVR cull zone highlighted in red. B: Frequency
615 bar-plot of *Mycobacterium bovis* isolates sampled annually, from unique badgers and cattle, in the
616 TVR zone prior to (1986-2013) and during the intervention (2014-2017).

617

618

619

620

621

622

623

624

625

626

627

628

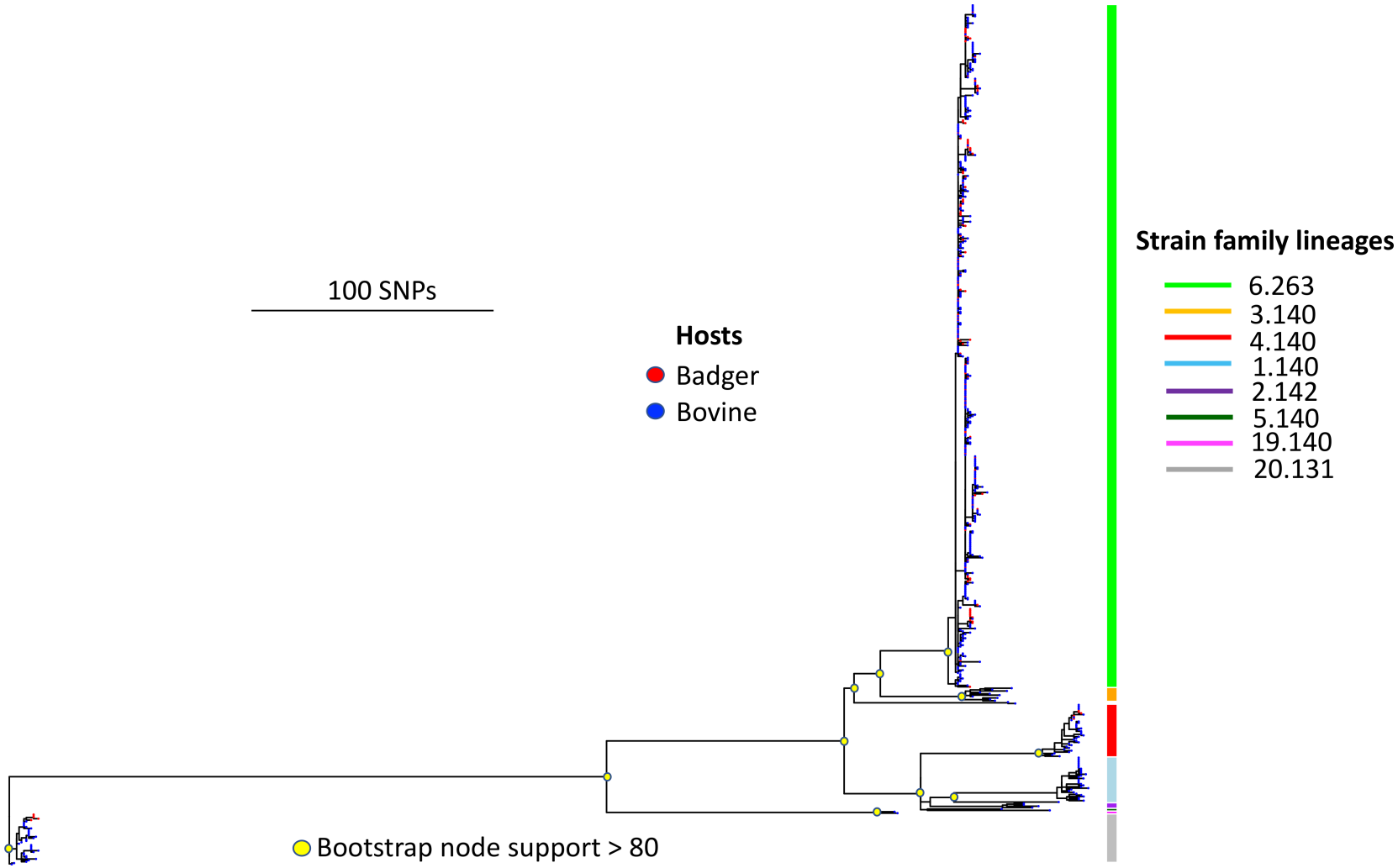
629

630

631

632

633




634

635 **Figure 2** – 1562 SNP Maximum Likelihood phylogeny of all 619 isolates that passed sequencing QA.

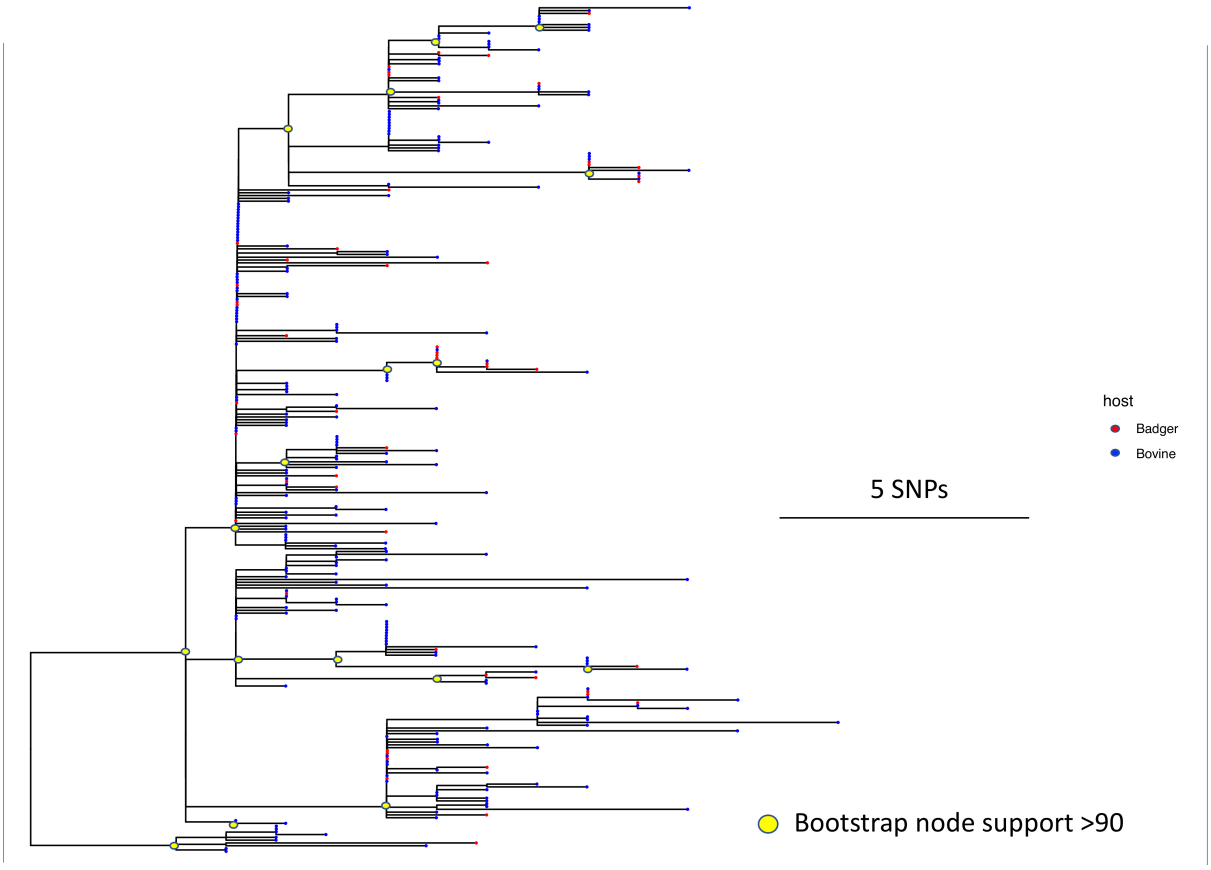



**Figure 3** – 377 SNP Maximum Likelihood phylogeny of 302 isolates from the 6.263 endemic lineage in the TVR zone. Tree rooted to AF2122/97 reference, but reference removed for aid of visualisation.

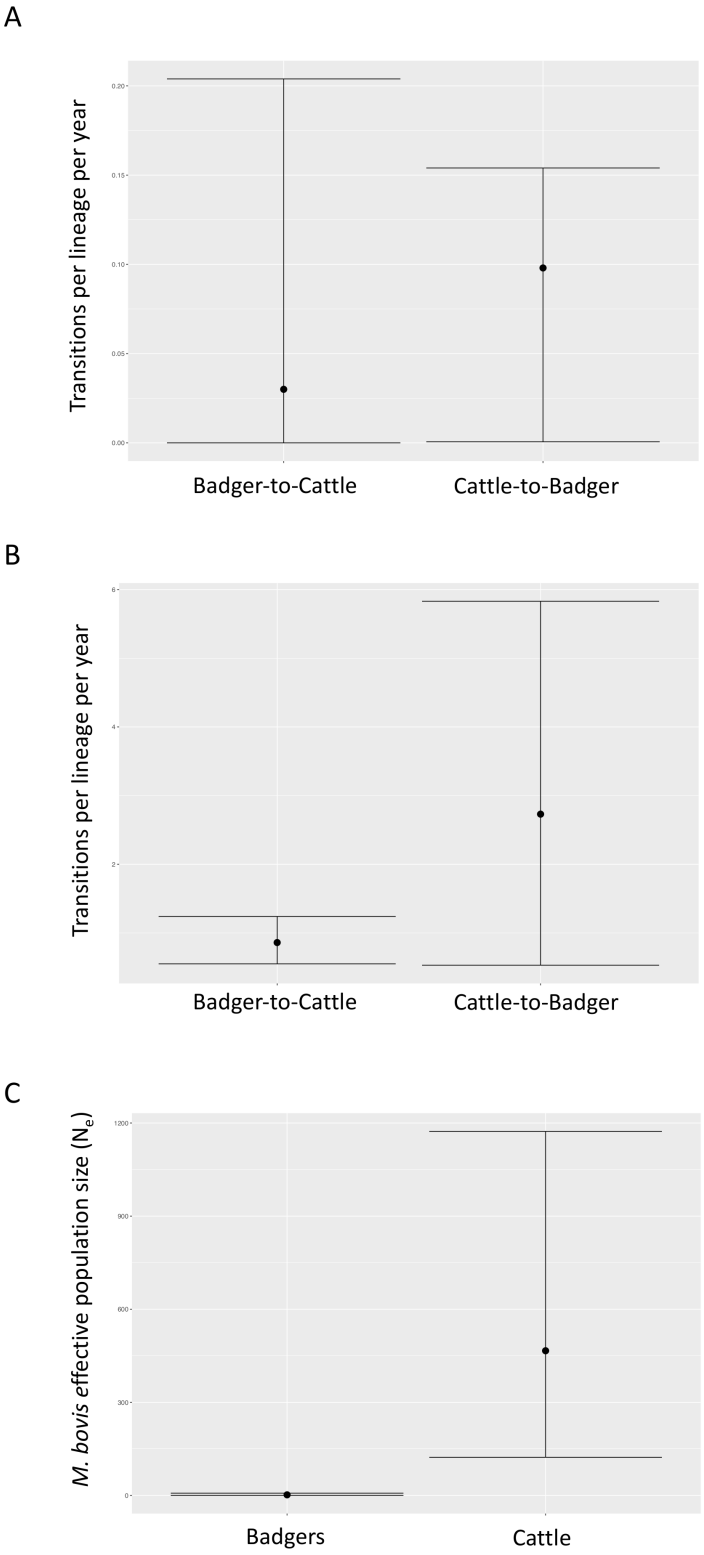



**Figure 4** – Logcombined interspecies pathogen transition rates as assessed by A: BASTA and B: MASCOT structured coalescent models. Transition rates are expressed as median transitions per lineage per year and 95% highest posterior density (HPD). C: Logcombined MASCOT inferred *M. bovis* effective population sizes in badgers and cattle and 95% highest posterior density (HPD).

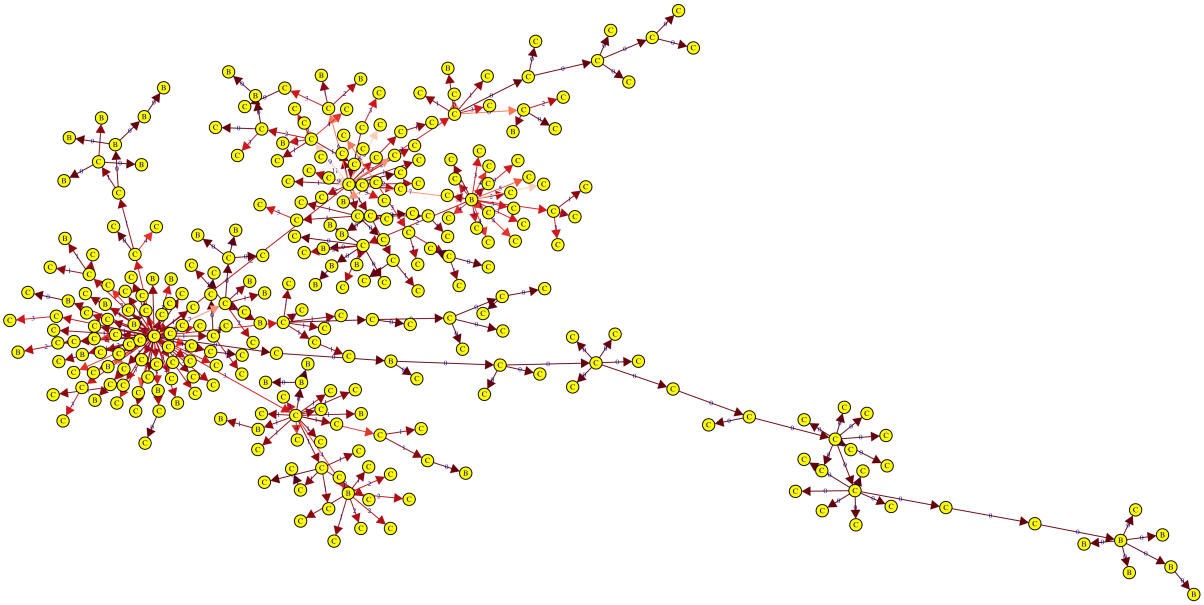


**Figure 5** - SeqTrack transmission tree of 302 *M. bovis* isolates collected from cattle and badgers. Nodes shown in yellow labelled as hosts (B for badgers, C for cattle), arrows delineate transmission direction (dark red = no/few mutations, light red = more mutations) with numbers of mutations between isolates shown in blue.

681 ## Tables

| | MLVA Strain families | 1.140 | 2.142 | 3.140 | 4.140 | 5.140 | 6.263 | 19.140 | 20.131 |
|---|---|---|---|---|---|---|---|---|---|
| A | No. of isolates | 33 | 4 | 10 | 38 | 2 | 478 | 2 | 36 |
| | No. cattle isolates | 33 | 4 | 10 | 33 | 2 | 367 | 2 | 32 |
| | No. badger isolates | 0 | 0 | 0 | 5 | 0 | 111 | 0 | 4 |
| B | No. SNPs in clade | 186 | 38 | 84 | 92 | 82 | 377 | 13 | 53 |
| C | Min | 0.0 | 8.0 | 8.0 | 0.0 | N/A | 0.0 | N/A | 0.0 |
| | 1st Quartile | 10.0 | 2.0 | 16.0 | 8.0 | N/A | 5.0 | N/A | 6.0 |
| | Median | 15.0 | 20.5 | 21.0 | 12.0 | N/A | 7.0 | N/A | 8.0 |
| | Mean | 17.6 | 19.5 | 21.6 | 11.7 | N/A | 7.6 | N/A | 8.0 |
| | 3rd Quartile | 19.0 | 21.0 | 27.0 | 15.0 | N/A | 10.0 | N/A | 11.0 |
| | Max | 96.0 | 27.0 | 34.0 | 28.0 | 79.0 | 25.0 | 13.0 | 18.0 |
| | St. Dev | 20.0 | 6.2 | 7.3 | 6.3 | N/A | 4.0 | N/A | 3.8 |

682 **Table 1** – A. number of isolates per strain family sampled in the study area with breakdown of
683 number of isolates per host species. B. number of SNPs detected in strain family clades, and C.
684 pairwise SNP distance statistics for each of the eight major lineages of *M. bovis* found in the TVR
685 zone.

686

687

688

689

690

691

692

693

694

695

696

697

| Cohort genotyped | No. animals | Mantel test Pearson coefficient r | Mantel p value | Linear model beta | Linear model p value | $R^2$ | 1 unit differentiation per x km distance |
|---|---|---|---|---|---|---|---|
| **2014** | 273 | 0.11 | 0.0001 | $1.6x10^{-4}$ | $<2x10^{-16}$ | 0.012 | 6.25 km |
| **2015** | 152 | 0.16 | 0.0001 | $2.1x10^{-4}$ | $<2x10^{-16*}$ | 0.024 | 4.74 km |
| **2016** | 97 | 0.17 | 0.0001 | $2.6x10^{-4}$ | $<2x10^{-16**}$ | 0.030 | 3.84 km |
| **2017** | 113 | 0.11 | 0.0001 | $1.5x10^{-4}$ | $<2x10^{-16}$ | 0.011 | 6.67 km |
| **2018** | 134 | 0.13 | 0.0004 | $1.7x10^{-4}$ | $<2x10^{-16***}$ | 0.017 | 5.88 km |
| **TB +ve** | 45 | 0.16 | 0.0024 | $3.1x10^{-4}$ | $1.1x10^{-7}$ | 0.030 | 3.22 km |

**Table 2** – Badger meta population isolation by distance (IBD) relationship for all sampling years. * =2015 significantly difference than slopes for years 2014, 2016 and 2017. ** = 2016 slope significantly different than slopes for years 2014 and 2015. ***=2018 slope significantly different than slopes for years 2014 and 2017.

| | Substitution rate | tMRCA | Badger *M. bovis* Ne | Cattle *M. bovis* Ne | Cattle to badger transition rate | Badger to cattle transition rate |
|---|---|---|---|---|---|---|
| **BASTA** | 0.38 (0.27-0.49) | 47.93 (32.64-85.34) | N/A | N/A | 0.098 (0.0006-0.154) | 0.030 (0.000002-0.2039) |
| **MASCOT** | 0.34 (0.28-0.42) | 37.30 (31.28-49.39) | 1.97 (0.26-7.39) | 466.19 (122.94-1172.92) | 2.73 (0.53-5.83) | 0.86 (0.55-1.24) |

**Table 3** – Summary statistics for log-combined triplicate BASTA and MASCOT models. All estimates are median and 95% highest posterior density (HPD). Substitution rate is expressed as sites per genome, per year. Transition rates are expressed as transitions per lineage per year. Time to most recent common ancestor (tMRCA) for the endemic 6.263 clade is expressed in years prior to the last sampling date, 2017.

743 ## References


744 Abernethy DA, Denny GO, Menzies FD, McGuckian P, Honhold N, Roberts AR (2006) The Northern
745 Ireland programme for the control and eradication of *Mycobacterium bovis*. Vet Microbiol 112(2-4):
746 231-237.

747 Adamack AT, Gruber B (2014) PopGenReport : simplifying basic population genetic analyses in R.
748 Methods Ecol Evol 5(4): 384-387.

749 Allen AR, Dale J, McCormick C, Mallon TR, Costello E, Gordon SV et al (2013) The phylogeny and
750 population structure of *Mycobacterium bovis* in the British Isles. Infect Genet Evol 20: 8-15.

751 Allen AR, Skuce RA, Byrne AW (2018) Bovine Tuberculosis in Britain and Ireland - A Perfect Storm?
752 the Confluence of Potential Ecological and Epidemiological Impediments to Controlling a Chronic
753 Infectious Disease. Front Vet Sci 5: 109.

754 APHA (2017) Livestock Demographic Data Group: Cattle population report. Animal and Plant Health
755 Agency, United Kingdom. http://apha.defra.gov.uk/documents/surveillance/diseases/lddg-pop-
756 report-cattle1117.pdf

757 Arnold ME, Courcier EA, Stringer LA, McCormick CM, Pascual-Linaza AV, Collins SF et al (2021) A
758 Bayesian analysis of a Test and Vaccinate or Remove study to control bovine tuberculosis in badgers
759 (*Meles meles*). PLoS One 16(1): e0246141.

760 Barido-Sottani J, Bošková V, Plessis LD, Kühnert D, Magnus C, Mitov V et al (2018) Taming the BEAST-
761 A Community Teaching Material Resource for BEAST 2. Syst Biol 67(1): 170-174.

762 Biek R, O'Hare A, Wright D, Mallon T, McCormick C, Orton RJ et al (2012) Whole Genome Sequencing
763 Reveals Local Transmission Patterns of *Mycobacterium bovis* in Sympatric Cattle and Badger
764 Populations. PLoS Pathog 8(11): e1003008.

765 Biek R, Pybus OG, Lloyd-Smith JO, Didelot X (2015) Measurably evolving pathogens in the genomic
766 era. Trends Ecol Evol 30(6): 306-313.

767 Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. Mol Ecol
768 19(17): 3515-3531.

769 Bielby J, Donnelly CA, Pope LC, Burke T, Woodroffe R (2014) Badger responses to small-scale culling
770 may compromise targeted control of bovine tuberculosis. Proc Nat Acad Sci USA 111(25): 9193-9198.

771 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D et al (2014) BEAST 2: a software platform
772 for Bayesian evolutionary analysis. PLoS Comp Biol 10(4): e1003537.

773 Buddle BM, Vordermeier HM, Chambers MA, de Klerk-Lorist LM (2018) Efficacy and Safety of BCG
774 Vaccine for Control of Tuberculosis in Domestic Livestock and Wildlife. Front Vet Sci 5: 259.

775 Chambers MA, Waterhouse S, Lyashchenko K, Delahay R, Sayers R, Hewinson RG (2009)
776 Performance of TB immunodiagnostic tests in Eurasian badgers (*Meles meles*) of different ages and
777 the influence of duration of infection on serological sensitivity. BMC Vet Res 5: 42.

778 Courcier EA, Menzies FD, Strain SAJ, Skuce RA, Robinson PA, Patterson IAP et al (2018) Monitoring
779 *Mycobacterium bovis* in Eurasian badgers (*Meles meles*) killed by vehicles in Northern Ireland
780 between 1998 and 2011. Vet Rec 182(9): 259.

Courcier EA, Pascual-Linaza AV, Arnold ME, McCormick CM, Corbett DM, O'Hagan MJH et al (2020) Evaluating the application of the dual path platform VetTB test for badgers (*Meles meles*) in the test and vaccinate or remove (TVR) wildlife research intervention project in Northern Ireland. Res Vet Sci 130: 170-178.

Crispell J, Benton CH, Balaz D, De Maio N, Ahkmetova A, Allen A et al (2019) Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. Elife 8.

Crispell J, Cassidy S, Kenny K, McGrath G, Warde S, Cameron H et al (2020) *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. Microb Genom 6(8).

Crispell J, Zadoks RN, Harris SR, Paterson B, Collins DM, de-Lisle GW et al (2017) Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. BMC Genomics 18(1): 180.

DAERA. (2018) Test and vaccinate or remove (TVR) wildlife intervention research. https://www.daera-ni.gov.uk/articles/test-and-vaccinate-or-remove-tvr-wildlife-intervention-research

De Maio N, Wu CH, O'Reilly KM, Wilson D (2015) New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. PLoS Genet 11(8): e1005421.

Delahay RJ, Walker N, Smith GC, Wilkinson D, Clifton-Hadley RS, Cheeseman CL et al (2013) Long-term temporal trends and estimated transmission rates for *Mycobacterium bovis* infection in an undisturbed high-density badger (*Meles meles*) population. Epidemiol Infect 141(7): 1445-1456.

Donnelly CA, Nouvellet P (2013) The contribution of badgers to confirmed tuberculosis in cattle in high-incidence areas in England. PLoS Currents Outbreaks 5.

Dray S, Dufour A (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. J Stat Software 22: 1-20.

Edwards D, Pope B, Holt K (2016) RedDog mapping based phylogenomics pipeline. https://github.com/katholt/RedDog

European Council. (1964) "Council Directive 64/432/EEC". Off J Eur Comm.

Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A (2010) Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol 27(9): 2038-2051.

Godfray C, Donnelly C, Hewinson RG, Winter M, Wood J (2018) Bovine TB strategy review. DEFRA https://www.gov.uk/government/publications/a-strategy-for-achieving-bovine-tuberculosis-free-status-for-england-2018-review

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. J Stat Software 22(7): 1-19.

Guerrero J, Byrne AW, Lavery J, Presho E, Kelly G, Courcier EA et al (2018) The population and landscape genetics of the European badger (*Meles meles*) in Ireland. Ecol Evol 8(20): 10233-10246.

Harris SR, Cartwright EJ, Torok ME, Holden MT, Brown NM, Ogilvy-Stuart AL et al (2013) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13(2): 130-136.

Hijmans RJ, Williams, E., Vennes C. (2019) Geosphere: Spherical trigonometry for geographic applications. https://cran.r-project.org/web/packages/geosphere/index.html

ISG - Independent Scientific Group on Cattle TB (2007) DEFRA London. http://www.bovinetb.info/docs/final_report.pdf

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21): 3070-3071.

Jombart T, Eggo RM, Dodd PJ, Balloux F (2011) Reconstructing disease outbreaks from genetic data: a graph approach. Heredity 106(2): 383-390.

Judge J, Wilson GJ, Macarthur R, McDonald RA, Delahay RJ (2017) Abundance of badgers (*Meles meles*) in England and Wales. Sci Reports 7(1): 276.

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S et al (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35(4): 907-914.

Kao RR, Price-Carter M, Robbe-Austerman S (2016) Use of genomics to track bovine tuberculosis transmission. Revue Sci Tech 35(1): 241-258.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9(4): 357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16): 2078-2079.

Malone KM, Farrell D, Stuber TP, Schubert OT, Aebersold R, Robbe-Austerman S et al (2017) Updated Reference Genome Sequence and Annotation of *Mycobacterium bovis* AF2122/97. Genome Announc 5(14).

Meehan CJ, Moris P, Kohl TA, Pečerska J, Akter S, Merker M et al (2018) The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology. EBioMedicine 37: 410-416.

Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A et al (2011) Prospective Genomic Characterization of the German Enterohemorrhagic Escherichia coli O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. PLoS ONE 6(7): e22751.

Menardo F, Duchêne S, Brites D, Gagneux S (2019) The molecular clock of Mycobacterium tuberculosis. PLoS Pathog 15(9): e1008067.

Menzies FD, McCormick, C.M., O'Hagan, M.J.H., Collins, S.F., McEwan, J., McGeown, C.F.et al (2021) Test and vaccinate or remove: methodology and preliminary results from a badger intervention research project. Vet Rec - Accepted.

Milne GM, Graham J, Allen A, Lahuerta-Marin A, McCormick C, Presho E et al (2019) Spatiotemporal analysis of prolonged and recurrent bovine tuberculosis breakdowns in Northern Irish cattle herds reveals a new infection hotspot. Spat Spatiotemporal Epidemiol 28: 33-42.

Milne MG, Graham J, Allen A, McCormick C, Presho E, Skuce R et al (2019) Variation in *Mycobacterium bovis* genetic richness suggests that inwards cattle movements are a more important source of infection in beef herds than in dairy herds. BMC Microbiol 19(1): 154.

Müller NF, Rasmussen D, Stadler T (2018) MASCOT: parameter and state inference under the marginal structured coalescent approximation. Bioinformatics 34(22): 3843-3848.

Müller NF, Rasmussen DA, Stadler T (2017) The Structured Coalescent and Its Approximations. Mol Biol Evol 34(11): 2970-2981.

Northern Ireland Audit Office. (2018) Bovine Tuberculosis Report. http://www.niauditoffice.gov.uk

O'Hagan M, Gordon, A., McCormick, C., Collins, S., Trimble, N., McGeown, C. et al (2021) The effect of ranging behaviour after selective removal of bovine tuberculosis test positive badgers (*Meles meles*) using a test and vaccinate or remove intervention in Northern Ireland. In review.

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3): 526-528.

Parish T, Stoker, N.G. (2001) Mycobacterium tuberculosis protocols. Humana Press.

Rambaut A (2018) Figtree v1.4.4. https://github.com/rambaut/figtree/releases

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 67(5): 901-904.

Rambaut A, Lam TT, Max Carvalho L, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus evol 2(1): vew007.

Reid N, Etherington TR, Wilson GJ, Montgomery WI, McDonald RA (2012) Monitoring and population estimation of the European badger *Meles meles* in Northern Ireland. Wildlife Biol 18(1): 46-57.

Rieux A, Khatchikian CE (2017) tipdatingbeast: an r package to assist the implementation of phylogenetic tip-dating tests using beast. Mol Ecol Res 17(4): 608-613.

Rossi G, Crispell J, Balaz D, Lycett SJ, Benton CH, Delahay RJ et al (2020a) Identifying likely transmissions in *Mycobacterium bovis* infected populations of cattle and badgers using the Kolmogorov Forward Equations. Sci Reports 10(1): 21980.

Rossi G, Crispell, J., Brough, T., Lycett, S.J., White, P.C.L., Allen, A. et al (2020b) Phylodynamic analysis of an emergent *Mycobacterium bovis* outbreak in an area with no previously known wildlife infections. bioRxiv. 2020.11.12.379297; doi: https://doi.org/10.1101/2020.11.12.379297

Salvador LCM, O'Brien DJ, Cosgrove MK, Stuber TP, Schooley AM, Crispell J et al (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. Mol Ecol 28(9): 2192-2205.

Schliep KP (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27(4): 592-593.

Skuce R, Breadon E, Allen A, Milne G, McCormick C, Hughes C et al (2020) Longitudinal dynamics of herd-level *Mycobacterium bovis* MLVA type surveillance in cattle in Northern Ireland 2003-2016. Infect Genet Evol 79: 104131.

Skuce RA, Mallon TR, McCormick CM, McBride SH, Clarke G, Thompson A et al (2010) *Mycobacterium bovis* genotypes in Northern Ireland: herd level surveillance (2003 to 2008). Vet Rec 167(18): 684-689.

896 Skuce RA, McDowell SW, Mallon TR, Luke B, Breadon EL, Lagan PL et al (2005).Discrimination of
897 isolates of *Mycobacterium bovis* in Northern Ireland on the basis of variable numbers of tandem
898 repeats (VNTRs). Vet Rec 157(17): 501-504.

899 Smith NH, Upton P (2012) Naming spoligotype patterns for the RD9-deleted lineage of the
900 Mycobacterium tuberculosis complex; www.Mbovis.org. Infect Genet Evol 12(4): 873-876.

901 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
902 phylogenies. Bioinformatics 30(9): 1312-1313.

903 R Development Core Team. (2020) R Foundation for Statistical Computing,Vienna, Austria.

904 Trewby H (2016a) The genetic and spatial epidemiology of bovine tuberculosis in the UK: from
905 molecular typing to bacterial whole genome sequencing. PhD thesis, University of Glasgow.

906 Trewby H, Wright D, Breadon EL, Lycett SJ, Mallon TR, McCormick C et al (2016b) Use of bacterial
907 whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis.
908 Epidemics 14: 26-35.

909 van Soolingen D, de Haas PE, Kremer K (2001) Restriction fragment length polymorphism typing of
910 mycobacteria. Methods Mol Med 54: 165-203.

911 Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ (2014) Efficient Bayesian inference
912 under the structured coalescent. Bioinformatics 30(16): 2272-2279.

913 Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SDW (2009) Phylodynamics of
914 Infectious Disease Epidemics. Genetics 183(4): 1421-1430.

915 Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ et al (2013) Whole-genome
916 sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study.
917 Lancet Infect Dis 13(2): 137-146.

918 Wright S (1943) Isolation by Distance. Genetics 28(2): 114-138.

919 Yu G. (2020) ggtree: Elegant Graphics for Phylogenetic Tree Visualization and Annotation.
920 https://guangchuangyu.github.io/ggtree-book/short-introduction-to-r.html

921