



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis

Citation for published version:

Trewby, H, Wright, D, Breadon, E, Lycett, S, Mallon, TR, McCormick, C, Johnson, P, Orton, RJ, Allen, A, Galbraith, J, Herzyk, P, Skuce, R, Biek, R & Kao, R 2016, 'Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis' *Epidemics*, vol. 14, pp. 26–35. DOI: 10.1016/j.epidem.2015.08.003

Digital Object Identifier (DOI):

[10.1016/j.epidem.2015.08.003](https://doi.org/10.1016/j.epidem.2015.08.003)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Epidemics

Publisher Rights Statement:

© 2015 The Authors.

Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics



Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis

Hannah Trewby^a, David Wright^b, Eleanor L. Breadon^c, Samantha J. Lycett^d,
Tom R. Mallon^c, Carl McCormick^c, Paul Johnson^a, Richard J. Orton^a, Adrian R. Allen^c,
Julie Galbraith^e, Pawel Herzyk^e, Robin A. Skuce^{b,c}, Roman Biek^a, Rowland R. Kao^{a,*}

^a Boyd Orr Centre for Population and Ecosystem Health, Institute for Biodiversity Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK

^b School of Biological Sciences, Queen's University Belfast, Belfast, Northern Ireland, UK

^c Veterinary Sciences Division, Agri-Food and Biosciences Institute, Stormont, Belfast, Northern Ireland, UK

^d The Roslin Institute, University of Edinburgh, Edinburgh, UK

^e Glasgow Polyomics, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

ARTICLE INFO

Article history:

Received 16 December 2014
Received in revised form 13 July 2015
Accepted 17 August 2015
Available online xxx

Keywords:

Bacterial evolution
Livestock disease
Molecular epidemiology
Mycobacterium bovis
Phylogeography

ABSTRACT

Mycobacterium bovis is the causal agent of bovine tuberculosis, one of the most important diseases currently facing the UK cattle industry. Here, we use high-density whole genome sequencing (WGS) in a defined sub-population of *M. bovis* in 145 cattle across 66 herd breakdowns to gain insights into local spread and persistence. We show that despite low divergence among isolates, WGS can in principle expose contributions of under-sampled host populations to *M. bovis* transmission. However, we demonstrate that in our data such a signal is due to molecular type switching, which had been previously undocumented for *M. bovis*. Isolates from farms with a known history of direct cattle movement between them did not show a statistical signal of higher genetic similarity. Despite an overall signal of genetic isolation by distance, genetic distances also showed no apparent relationship with spatial distance among affected farms over distances <5 km. Using simulations, we find that even over the brief evolutionary timescale covered by our data, Bayesian phylogeographic approaches are feasible. Applying such approaches showed that *M. bovis* dispersal in this system is heterogeneous but slow overall, averaging 2 km/year. These results confirm that widespread application of WGS to *M. bovis* will bring novel and important insights into the dynamics of *M. bovis* spread and persistence, but that the current questions most pertinent to control will be best addressed using approaches that more directly integrate WGS with additional epidemiological data.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The increasing availability of bacterial whole-genome sequence (WGS) data now makes it possible to generate sequence datasets for whole bacterial pathogen populations at high sampling densities. Such comprehensive sequencing has yielded impressive advances in outbreak investigation (Eyre et al., 2013; Harris et al., 2010; Walker et al., 2012), and provided new insights into both spatial dissemination (Gray et al., 2011; Holden et al., 2013) and the complexities of multi-host pathogen systems (Mather et al., 2013; Viana et al., 2014). However, even at the genomic scale

the rates of evolutionary change estimated for bacteria can be substantially lower than those commonly seen in rapidly evolving pathogens such as RNA viruses (Biek et al., 2015; Bryant et al., 2013b). The extent to which slow evolution constrains the type and scale of epidemiological processes that can be resolved for bacterial pathogens, and which analytical approaches are most appropriate to deal with this, remains unclear for many systems.

Mycobacterium bovis is one of a group of closely related bacteria which includes the primary cause of human tuberculosis, *M. tuberculosis*, a pathogen estimated to evolve at a rate of around 0.3–0.5 mutations per genome per year over epidemiological timescales (Bryant et al., 2013b; Walker et al., 2012). *M. bovis* is the causative agent of bovine tuberculosis (bTB), an important disease of cattle and other mammals including man. Herd-to-herd movements of infected cows among farms (Gilbert et al., 2005; Green et al.,

* Corresponding author. Tel.: +44 141 3305770.
E-mail address: rowland.kao@glasgow.ac.uk (R.R. Kao).

<http://dx.doi.org/10.1016/j.epidem.2015.08.003>

1755-4365/© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2008) and infection in Eurasian badger (*Meles meles*) populations (Delahay et al., 2001; Gallagher and Clifton-Hadley, 2000) have both been implicated in the spread of bTB in Britain and Ireland. While much attention has focused on the relative roles of badgers and cattle in the maintenance of bTB, recent studies emphasising the importance of cattle have highlighted the continuing need for a deeper understanding of the role of cattle-based transmission (Brooks-Pollock et al., 2014; Donnelly and Nouvellet, 2013).

Molecular typing of *M. bovis* isolates based on repeated genetic elements has been advocated for some time to aid in the epidemiology and control of bTB (Cousins et al., 1998; Skuce and Neill, 2001), and in Britain and Ireland these typing methods have shown that *M. bovis* molecular types are maintained within well-defined geographic clusters (Skuce et al., 2010; Smith et al., 2006). While such molecular typing has proved useful for identifying local clustering on larger scales, their power to discriminate within-cluster events involved in fine-scale persistence and spread of bTB is limited.

In a previous study, Biek et al. (2012) established the potential of bacterial WGS in investigating the epidemiology of bTB at a local (i.e., farm-to-farm) scale. By sequencing 30 bacterial isolates from a spatially dense cluster of bTB cases within one recently emerged *M. bovis* molecular type (VNTR-10) in Northern Ireland (NI), the study demonstrated (i) close relatedness of bacteria isolated from cattle and badgers, (ii) persistence of bacterial lineages on the same farm, and (iii) that genetic similarity between isolates correlated with geographic distance between sampling locations. The study also showed that, due to slow evolution, even WGS is unlikely to provide sufficient resolution to resolve transmission at the animal-to-animal scale for *M. bovis*, similar to findings in human tuberculosis (Bryant et al., 2013b; Roetzer et al., 2013; Walker et al., 2012), and is more suited to do so at the between-farm scale.

While providing a proof of concept, this previous study was targeted towards a subsample of VNTR-10 infected cattle within a small (approx. 5 km) spatial radius. This spatially restricted sampling precluded a more systematic investigation of processes occurring on a wider, population level scale within the bacterial strain, including the potential identification of under-sampled reservoirs, the rate and mode of spatial spread, and transmission links between bTB breakdowns. A herd breakdown is defined as the period during which movements of cattle out of a herd are restricted due to the detection of bTB in the herd, starting at the detection of one or more infected animals (either through the tuberculin skin test or through abattoir surveillance for bTB lesions), and ending when the herd has undergone two consecutive negative whole-herd tests at least 60 days apart, or a single negative test where the breakdown was not laboratory confirmed.

Here, we extend the analysis of the Biek et al. (2012) study by examining WGS data from all 145 available VNTR-10 isolates in NI since 2003. In NI all cattle herds are tested for bTB on an annual basis, and for over a decade *M. bovis* isolates cultured from test-positive cattle have been extensively typed and stored. These archived samples therefore gave us the opportunity to target a genetically defined sub-population of *M. bovis* (VNTR-10) for high-density sampling with respect to cattle infections, although VNTR-10 infections in any other population would not be accessible through this sampling strategy. To gain insights into the mode and rates of transmission, we used intensive sampling and WGS of *M. bovis* isolates from cattle to address the following questions:

- (1) Does WGS of VNTR-10 isolates from cattle indicate contributions from another host population which is under-sampled under the above sequencing strategy?
- (2) Does the genetic relatedness between sequenced isolates correlate with recorded movements and/or with spatial distance between premises?

- (3) What are the rate and mode of *M. bovis* dispersal across the landscape at the between-breakdown scale as determined by WGS?

2. Materials and methods

2.1. Molecular-typing of *M. bovis* in NI

In NI, *M. bovis* isolates have been stored and typed since the early 2000s using spoligotyping, more recently combined with Variable Number Tandem Repeat (VNTR) typing, to differentiate molecular types (Skuce et al., 2005). Spoligotyping gives a relatively coarse-grained discrimination of the *M. bovis* population, and is based on the presence or absence of multiple spacer oligonucleotides within the direct repeat region of the genome (Kamerbeek et al., 1997). VNTR-typing indexes the number of short nucleotide repeats present at several VNTR loci identified within the mycobacterial genome (Mazars et al., 2001), and provides greater discriminatory power than spoligotyping alone, although with a relatively higher chance of homoplasy (i.e., separate lineages converging on the same molecular type). In NI between 2003 and 2008 one *M. bovis* isolate was VNTR-typed and stored from each herd breakdown for which *M. bovis* was isolated using a panel of 7 VNTR loci optimised for this population of *M. bovis* (Skuce et al., 2010), while from 2009 onwards bacteria were VNTR-typed and stored from all culturable cattle cases, therefore resulting in more intensive sampling in recent years. In addition to these cattle isolates, *M. bovis* has also been typed by spoligotyping and VNTR-typing and archived when isolated from a survey of badgers killed on the roads in NI (Abernethy et al., 2011).

2.2. Bacterial samples and sequencing

Cultures of *M. bovis* were isolated from bovine granulomatous tissue using conventional methods (Skuce et al., 2010). Confirmed isolates were grown on LJ slopes to single colonies, following which single colonies were grown up and DNA was extracted using standard CTAB and solvent extraction protocols (Van Soolingen et al., 2001). A total of 144 VNTR-10 *M. bovis* isolates were included in this study, from 66 herd breakdowns (see earlier definition) occurring in 51 herds between 1996 and 2011. The WGS dataset consisted of the raw reads from 31 VNTR-10 samples originally sequenced in the preceding study (26 cattle and 5 badger isolates; Biek et al., 2012), in addition to 114 VNTR-10 samples (113 cattle and 1 badger isolate) sequenced for the first time in this study. VNTR-10 is located predominantly within the Newtownards area of NI (178/195 VNTR-10 infections recorded between 1996 and 2011 were from the Newtownards district veterinary office), and statistical comparison to other VNTR types confirms that VNTR-10 is generally representative of strains circulating in this area (see supplemental information). Details of the accession numbers for the raw sequencing reads are given in Table S1.

To provide broader evolutionary context we additionally sequenced five samples from VNTR-types thought to be ancestral to VNTR-10, namely four isolates of VNTR-1 (1 cattle isolate (sample A), and 3 badger isolates (samples B–D, Figs. 1 and 2)), and one cattle isolate of VNTR-4. VNTR-1 is thought to be the direct ancestor of VNTR-10 based on the following observations: VNTR-1 and VNTR-10 are separated by a single tandem repeat difference; VNTR-1 has been recorded at a high and approximately stable prevalence in NI since routine VNTR-typing commenced (Skuce et al., 2010, 2005), whereas VNTR-10 has been found in low but increasing numbers suggestive of a newly emerged strain; VNTR-1 is found across a wider spatial range than VNTR-10 (Fig. 2); and a minimum spanning tree of all NI VNTR-types within spoligotype SB0140 shows

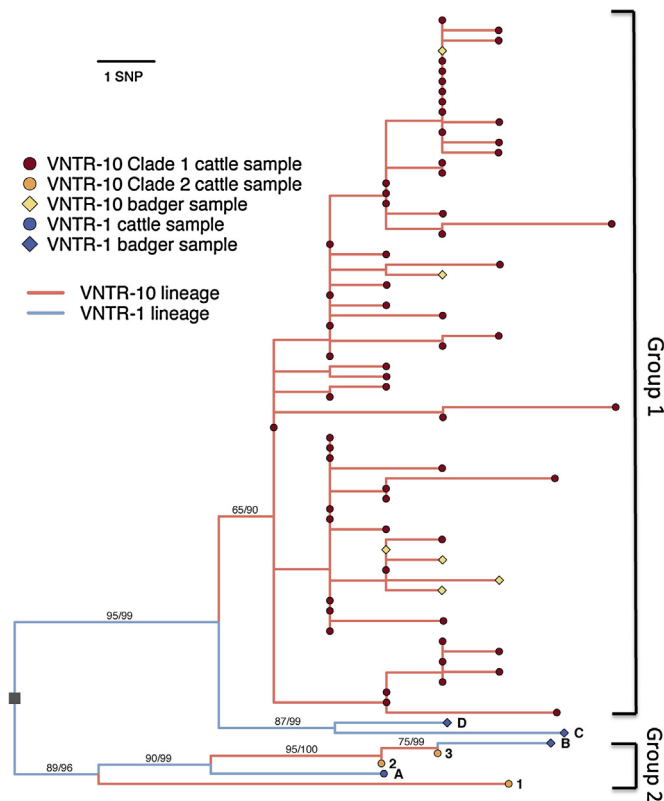


Fig. 1. Maximum Likelihood phylogeny of VNTR-1 and -10 isolates subsampled to one sequence per outbreak and rooted on the VNTR-4 isolate and *M. bovis* reference sequence (Garnier et al., 2003) (not shown; the node used to root the phylogeny is indicated by a grey square). Tip colours give details of the samples: red circles are Group 1 VNTR-10 cattle isolates, orange circles (numbers 1–3) are Group 2 VNTR-10 cattle samples; yellow diamonds are VNTR-10 badger isolates; blue circle (A) is the VNTR-1 cattle isolate, blue diamonds (B–D) are VNTR-1 badger isolates. Branch colours give the likely VNTR-type of each branch, assuming the most recent common ancestor of the group was VNTR-1. Branch labels show the statistical support for selected nodes: the left-hand value indicates percentage bootstrap support from a maximum likelihood phylogeny generated for these isolates, and the right-hand value shows posterior probability of the node in the Bayesian phylogeny generated for these isolates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

that VNTR-1 is basal compared to VNTR-10 (Fig. S1). The VNTR-4 type differs from VNTR-1 by two tandem repeats (Table S2), however it is not possible to determine which of them is ancestral. Sequencing was performed at Glasgow Polyomics at the University of Glasgow using the Illumina Ix platform, with the exception of VNTR-1 samples B–D, sequenced using an Illumina MiSeq.

2.3. Sequence analysis

Full details of the bioinformatics workflow are provided in the Supplementary Information. Briefly, reads were trimmed and mapped to the *M. bovis* reference genome (GenBank accession number BX248333; Garnier et al., 2003) using BWA (Li and Durbin, 2009). Variants were identified using SAMtools (Li et al., 2009) and filtered on base quality, mapping quality, heterozygosity, proportion of samples with high quality calls at each site, clustering of variant loci, and location relative to repeat regions of the genome. The resulting variant sites were concatenated for each isolate, giving the genetic sequences used for downstream analyses.

A maximum likelihood phylogeny was generated in PhyML v3.0 (Guindon and Gascuel, 2003) under the Jukes Cantor model of nucleotide substitution, including the *M. bovis* reference sequence as outgroup, and evaluating statistical support for individual nodes

based on 1000 non-parametric bootstraps. A Bayesian phylogeny was generated in MrBayes (Ronquist and Huelsenbeck, 2003) under the Jukes Cantor model, also including the *M. bovis* reference sequence, and was run for 10^6 MCMC iterations at which point the standard deviation of split frequencies was below 0.01. Raw pairwise single nucleotide polymorphism (SNP) differences between sequenced samples were calculated in MEGA 5 (Tamura et al., 2011), using pairwise deletions in the event of missing data. Due to the increased level of sampling from 2009 onwards mentioned above and the low levels of within-breakdown diversity (see Results), further analyses were restricted to one representative sample per herd breakdown.

2.4. Comparing genetic and epidemiological relationships between breakdowns

In NI, detailed information on the cattle population and movements between herds and bTB test results is recorded by the Department of Agriculture and Rural Development (Houston, 2001). All direct cattle movements between herds with VNTR-10 samples represented in our dataset (66 sequenced breakdowns and 12793 individual cattle movements) were made available, and were combined with the location and date of each sequenced sample using SQL (MySQL v5.5.29; www.mysql.com). All further analyses were conducted in R v3.0.1 (R Core Team and R Development Core Team, 2014) unless otherwise stated.

A Mantel test was conducted in the R package ecodist (Goslee and Urban, 2007; Lichstein, 2007) to assess correlation between spatial and genetic distance between Group 1 breakdowns, using 10000 permutations to assess significance (see results and Fig. 1 for definition of Group 1). To further confirm this correlation, a multiple regression on distance matrices was also carried out using the R package ecodist (Lichstein, 2007), assessing the correlation between the matrix of pairwise SNP differences and the matrix of geographic distances between Group 1 breakdowns, and using 10000 permutations to assess significance.

For each sequenced Group 1 breakdown, we recorded whether it was linked to one or more other sequenced breakdowns by potentially infectious cattle moving directly into the original breakdown herd (“movement links”). We considered movements occurring within 2, 5, and 10 years prior to the official start of each breakdown. Few direct movement links were identified within the 2- and 5-year windows (15 and 53 links, compared to 102 for the 10-year window). We therefore included all movements identified within the most conservative window of 10 years in our analysis. For each sequenced breakdown, we identified any movement links to that breakdown from any other sequenced breakdown. Where links were present, we recorded the minimum pairwise SNP difference between the linked premises, as the minimum SNP difference is most likely to represent a direct transmission event if one exists between breakdowns.

To assess the effect of short-distance transmission mechanisms in a more targeted manner, we identified pairs of sequenced herds located within 2 km and 5 km of another sequenced herd. This choice was motivated by the observation that the vast majority of badger movements fall within 5 km, although larger distances are occasionally recorded (Byrne et al., 2014; Pope et al., 2006). We note however that transmission over short distances may be driven by various mechanisms and is not necessarily restricted to badgers. For each sequenced breakdown, we identified the presence of other sequenced breakdowns linked through spatial proximity at these distances, and again recorded the minimum number of SNPs separating it from other spatially-linked breakdowns.

A higher number of epidemiological links between breakdowns is likely to result in a lower minimum SNP differences between breakdowns due to the increased number of comparisons, and

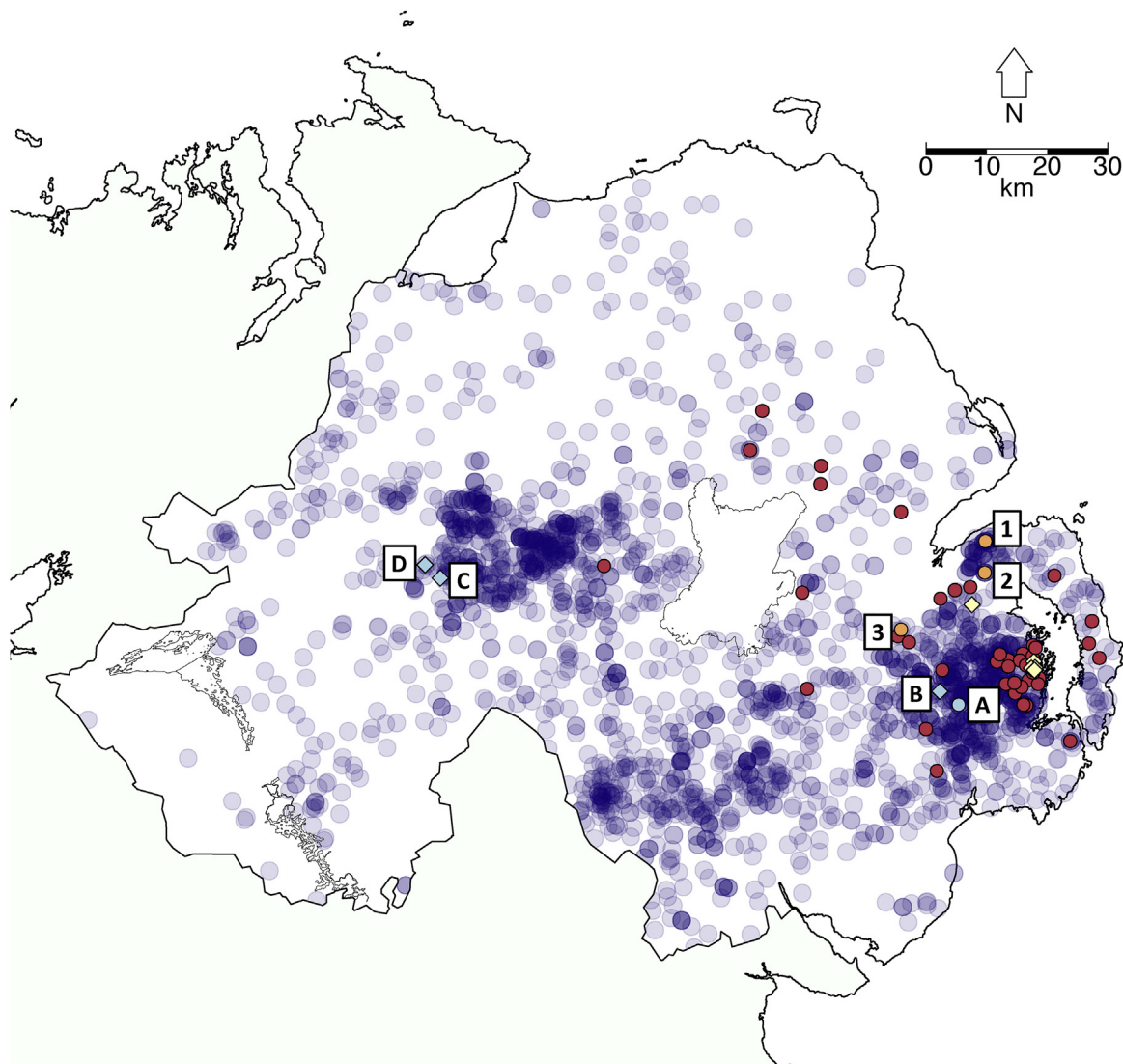


Fig. 2. Map of Northern Ireland showing origins of sequenced samples. Red circles are Group 1 VNTR-10 cattle isolates; orange circles (samples 1-3) are Group 2 VNTR-10 cattle samples, yellow diamonds are VNTR-10 badger isolates, light blue diamonds (badger) and circle (cattle) (samples A–D) are VNTR-1 isolates, and dark blue transparent circles show locations of all other VNTR-1 isolates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

therefore the distributions of minimum SNP differences are not directly comparable between different types of links. To statistically assess the significance of the association between epidemiological links and minimum SNP differences, the distribution of the expected number of SNP differences was simulated for movement and proximity links under a null hypothesis of no association between the presence of a link and genetic relatedness between breakdowns. The number of links to each sequenced breakdown, identified above, was kept but these links were effectively “rewired” by permuting the matrix of linked outbreaks and thus randomising the breakdowns that each sequenced breakdown was linked to, and the minimum SNP difference was again calculated for each breakdown. This was repeated 10,000 times for each type of link (direct recorded movements within 10 years of breakdowns and spatial proximities of 2 km and 5 km) to generate null distributions for comparison to the observed distributions.

To formally compare the null simulations with the observed distribution of minimum SNP differences between linked outbreaks, a goodness of fit test was carried out. We generated 10,000 realisations of a multinomial distribution, with sample size equal to the number of linked breakdowns, and the probability of each category proportional to the expected value of each category (taken

as the mean of the values from the null simulations). A Chi-squared test statistic was then calculated for each of these multinomial distributions. This gave a distribution of simulated test statistics to which the Chi-squared statistic calculated from the observed data was compared. A *p*-value was estimated as the proportion of the null Chi-squared statistics that were greater than the observed Chi-squared value.

2.5. Phylogeographic inference

To quantify the spread of *M. bovis* across the landscape, continuous phylogeographic models were applied using the Bayesian phylogenetic program BEAST v1.7.4 (Drummond et al., 2012; Lemey et al., 2010). This analysis was restricted to the VNTR-10 clade containing the majority of isolates (Group 1, see Results and Fig. 1), using one representative sample per breakdown. A strict Brownian model of spatial diffusion was compared to a relaxed model allowing diffusion rates to vary among branches, with rates drawn from a Cauchy distribution (see Supplemental Information). A relaxed model with branch rates drawn from a gamma distribution was also tested but failed to converge. Models were run for 5×10^8 iterations, assessed for convergence in Tracer, and model

fit evaluated based on log Marginal Likelihood Estimates (MLE) generated by using path-sampling and stepping-stone sampling in BEAST (Baele et al., 2012). Posterior trees for the best fitting model were combined to find and annotate the Maximum Clade Credibility (MCC) tree. Node locations, branch lengths, and branch-specific rates of geographic dispersal were extracted and evaluated for the MCC tree.

Given that the molecular clock rate of *M. bovis* and other closely related mycobacteria has been shown to be slow and variable (Biek et al., 2012; Bryant et al., 2013b), it was uncertain whether these data would contain enough genetic signal to accommodate phylogeographic analyses. To test this, we simulated a homogeneous spatial diffusion process along the MCC phylogeny generated above, guided by empirical rates, generating a set of spatial coordinates for sampled sequences under a set rate of spatial diffusion along the existing phylogeny. We then evaluated whether phylogeographic analysis in BEAST, using the settings described above, using the simulated coordinates and observed sequences and sampling dates as input, could recover the originally specified diffusion rate for each of 100 simulations.

3. Results

3.1. VNTR-10 isolates

Among the intensively sampled VNTR-10 isolates, genetic divergence was low; averaging 6.4 SNPs (range 0–19) over the whole group of 145 sequenced VNTR-10 isolates. Despite limited divergence, the group contained 39 shared polymorphisms resulting in a non-star-like phylogenetic structure (Fig. 1, showing one sample per herd breakdown).

Regressing genetic distance from the root of the phylogeny against sampling date revealed a moderate positive correlation ($R^2 = 0.32$), indicative of a molecular clock signal within these data (Firth et al., 2010).

Average diversity for sequenced isolates from within the same herd breakdowns (including two breakdowns which were polyphyletic) was low, with mean 0.69 SNPs and range 0–4 SNPs. This was considerably lower than the average minimum SNP differences between different VNTR-10 breakdowns (mean 4.73 SNPs, range 0–17 SNPs). Multiple samples per breakdown were only available for 19 of the 66 VNTR-10 breakdowns. Given the low level of within-breakdown divergence and the fact that multiple samples per breakdown were only available after 2008 (see Materials and methods), we chose to use one representative sequence per herd breakdown for further analysis in order to focus on the dynamics of *M. bovis* spread at the between-breakdown scale. The full phylogeny, including all samples described here, is shown in Fig. S2.

Both Bayesian and Maximum Likelihood phylogenies indicated that the majority of VNTR-10 sequences fall into a single main group, hereafter referred to as Group 1, including 126 samples from 63 breakdowns, and including the six badger samples (Fig. 1). Nineteen other VNTR-10 samples, belonging to three individual breakdowns (represented by samples 1, 2, and 3, Fig. 1) were not positioned within Group 1 and instead clustered with two of the VNTR-1 samples (samples A and B, Fig. 1, Group 2). Both Group 1 and Group 2 isolates were defined by multiple unique SNPs, however statistical support for the nodes defining these two lineages was not consistently high (bootstraps of 65 and 89 respectively, and posterior probabilities of 90 and 96, Fig. 1).

Despite the much larger number of Group 1 samples, these two sets of VNTR-10 samples were similar in terms of maximum genetic divergence among isolates (Group 1 samples: 15 SNPs; Group 2 samples: 13 SNPs). Within Group 1, the average pairwise

distance was low (mean: 4.4 SNPs), many isolates were genetically indistinguishable, and the majority of mutational steps were represented by one or more sequenced isolates. These observations are indicative of a population that has been comprehensively sampled, consistent with our expectations based on sampling all VNTR-10 outbreaks detected over the study period. In contrast, sequences from the Group 2 VNTR-10 breakdowns were non-identical, more divergent from each other (mean: 8.5 SNPs) and many of the intermediate mutational steps were not represented in the sequenced samples. This suggests that comprehensive sampling of the lineage represented by these isolates had not been achieved.

To evaluate whether these differences might simply be due to sampling, 1000 random subsamples of three sequences (matching the number of Group 2 VNTR-10 breakdowns) were generated from Group 1 isolates. The observed pairwise differences between Group 2 VNTR-10 isolates (1, 12, and 13 SNP differences) were at the upper end of the simulated distribution of pairwise distances among subsamples (mean 4.16 SNPs, range 0–12 SNPs; Fig. S3). This suggests that Group 2 isolates were sampled from a bacterial population that is at least as large as, or even larger than Group 1 (assuming no significant differences in evolutionary rates between the two lineages), but that a lower proportion of this population has been sampled and sequenced compared to Group 1, despite the consistently high sampling effort for VNTR-10 outbreaks in cattle.

Given that the Group 2 isolates appeared to come from a population with different characteristics compared to Group 1, and given that these comprised only three sampled VNTR-10 breakdowns, we restricted all further analyses to Group 1 samples unless otherwise indicated.

3.2. VNTR-types 4 and 1

As anticipated, the single VNTR-4 isolate was genetically highly distinct from the VNTR-10 group, differing by an average of 89 SNPs (Fig. S2). In contrast, VNTR-1 isolates failed to form a separate clade, as would have been expected for a separate VNTR-type, and instead were found to be nested within the VNTR-10 group. The two VNTR-1 isolates originating in the same geographical area as the majority of the VNTR-10 samples (samples A and B, Fig. 2) were found to cluster with VNTR-10 Group 2 samples (Fig. 1), while those originating from outside the VNTR-10 range formed a sister group to VNTR-10 Group 1 (Fig. 1, samples C and D, see also 'Badger isolates' below).

3.3. Badger isolates

As discussed by Biek et al. (2012) using a subset of the data described here, VNTR-10 *M. bovis* isolates from badgers and cattle were highly similar genetically, with a minimum distance of 0–3 SNPs to the most closely related cattle isolate (Fig. 1), suggestive of recent transmission links between badger and cattle. While the current study included only one additional VNTR-10 badger isolate, our high density sampling of VNTR-10 cattle infections also showed that the cattle isolates most closely related to those from badgers were all found within very close spatial proximity (<1.5 km) to the locations of these badger isolates.

Of the three newly sequenced VNTR-1 badger isolates, two samples (C and D) originated from an area outside the distributional range of VNTR-10, and approximately 100 km from the area where the majority of VNTR-10 isolates were located (Fig. 2). These two isolates were closest to VNTR-10 Group 1 sequences, separated by a minimum genetic distance of 5 SNPs (Fig. 1). As described above, the other VNTR-1 badger isolate (sample B) originated from the same area as the majority of the VNTR-10 samples (Fig. 2), and clustered with Group 2 VNTR-10 isolates (Fig. 1)

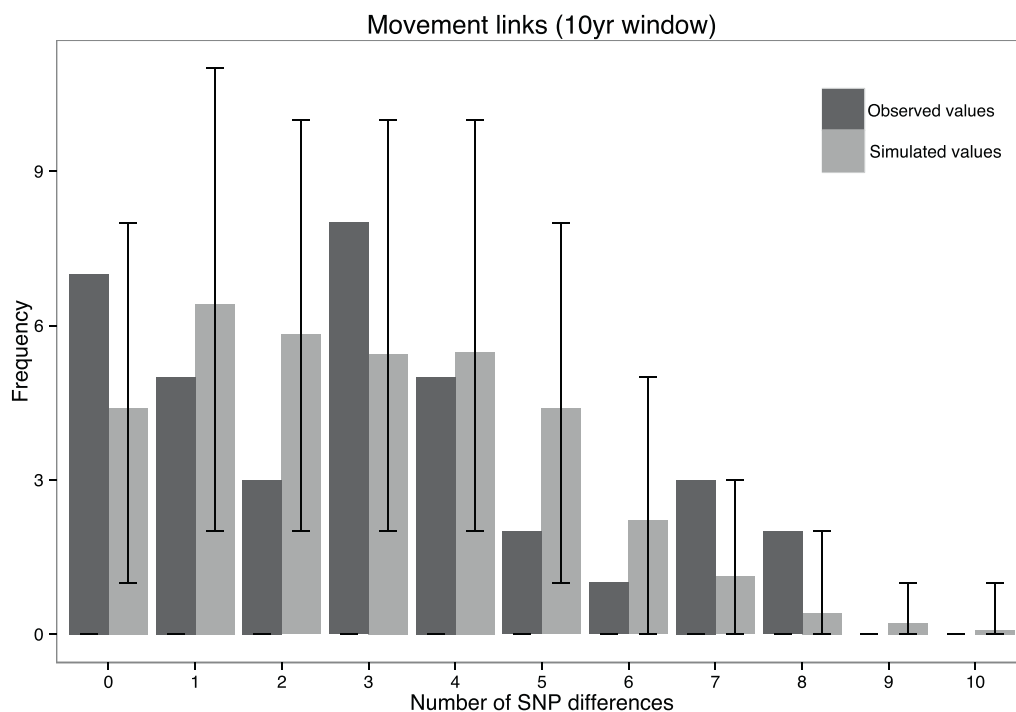


Fig. 3. Observed number of SNP differences between outbreaks linked by movements of cattle within a 10-year timeframe (dark grey), and expected SNP differences from 10^4 simulations of the null hypothesis of no association between presence of a link and genetic similarity (light grey). Bars show the intervals containing 95% of the results from the null simulations.

3.4. Molecular clock rate

Bayesian evolutionary analysis revealed an evolutionary rate of 0.2 substitutions per genome per year (95% HPD 0.1–0.3), estimating the time of the most recent common ancestor of the VNTR-10 group as a whole as 1974 (1954–1989), and the time of the most recent common ancestor of Group 1 at 1988 (1979–1995) and Group 2 at 1984 (1968–1998). Our evolutionary rate estimate is lower than some estimates for *M. tuberculosis* in humans, (95% CIs: 0.3–0.7 substitutions genome⁻¹ year⁻¹ (Roetzer et al., 2013; Walker et al., 2012), whereas other studies reported similar rates (0.13–0.41; Bryant et al., 2013b). While differences in pathogen life history, such as disease latency (Colangeli et al., 2014) or non-replicating persistence of bacteria in the environment (Courtenay et al., 2006; Maddock, 1933; Young et al., 2005), might contribute to rate variation among different mycobacteria, further studies will be required to verify this.

3.5. Comparing genetic and epidemiological relationships between breakdowns

A Mantel test showed a significant, though weak, association between the genetic and spatial distances of Group 1 breakdowns ($p=0.014$, Spearman's rank coefficient=0.20; Fig. S4), as did multiple regression on matrices ($p=0.034$, $R^2=0.031$).

The minimum SNP differences observed between pairs of sequenced Group 1 breakdowns linked by direct recorded cattle movements (Fig. 3, mean 1.29 SNP differences, range 0–11 SNPs) and by spatial proximity of 5 km (Fig. 4B, mean 0.49 SNPs, range 0–4 SNPs) were not significantly different from expectations under the null hypothesis of no association (p -values of 0.350 and 0.338 respectively), whereas the association showed borderline significance for spatial proximities of 2 km (Fig. 4A: mean 0.57 SNPs, range 0–11 SNPs, $p=0.048$). These analyses were also conducted on a subset of the Group 1 data comprising isolates occurring from 2009 onwards (i.e., after sampling intensity was increased to include

all infected cattle in a bTB breakdown), to check for an effect of temporal differences in sampling intensity, and these also gave non-significant results (Fig. S5).

Four of the Group 1 VNTR-10 breakdowns from which isolates were sequenced for this study showed no apparent epidemiological links to any other breakdown; neither through direct recorded movements nor spatial proximity up to 5 km (Fig. S6). One of the Group 2 VNTR-10 breakdowns had no apparent epidemiological links to other VNTR-10 isolates, while the other two Group 2 breakdowns did show epidemiological links, but are only comparatively distantly related to the linked breakdowns (>10 SNP differences).

Data from repeat breakdowns of VNTR-10 within the same herd were available for 11 herds and 27 breakdowns. For nine of these breakdowns, the later breakdown was caused by an isolate closely related to the preceding breakdown (less than 3 SNP differences), indicating the possibility of local persistence even after a cattle herd has been declared free of bTB.

3.6. Application of phylogeographic tools

Comparison of different phylogeographic models of bacterial dispersal showed statistical support for a heterogeneous model of spatial diffusion allowing different rates of spread among phylogenetic branches (log MLEs: –1518 for the relaxed model and –1605 for the homogeneous model). This model further provides information for each phylogenetic branch about the estimated distance travelled over the time period represented by its length. Based on the MCC tree for Group 1 (terminal branches only), the estimated mean diffusion rate was comparatively low at 2 km/year, but with higher rates up to 30 km/year seen rarely (Fig. S7). The majority of branches underlying these rates involve distances of <5 km travelled over less than 5 years (Fig. S8).

In 95 out of 100 simulations of homogeneous spatial diffusion along the time-stamped VNTR-10 phylogeny, the originally-specified diffusion rate could be recovered in BEAST, in that the originally specified value fell within the estimated 95% HPD

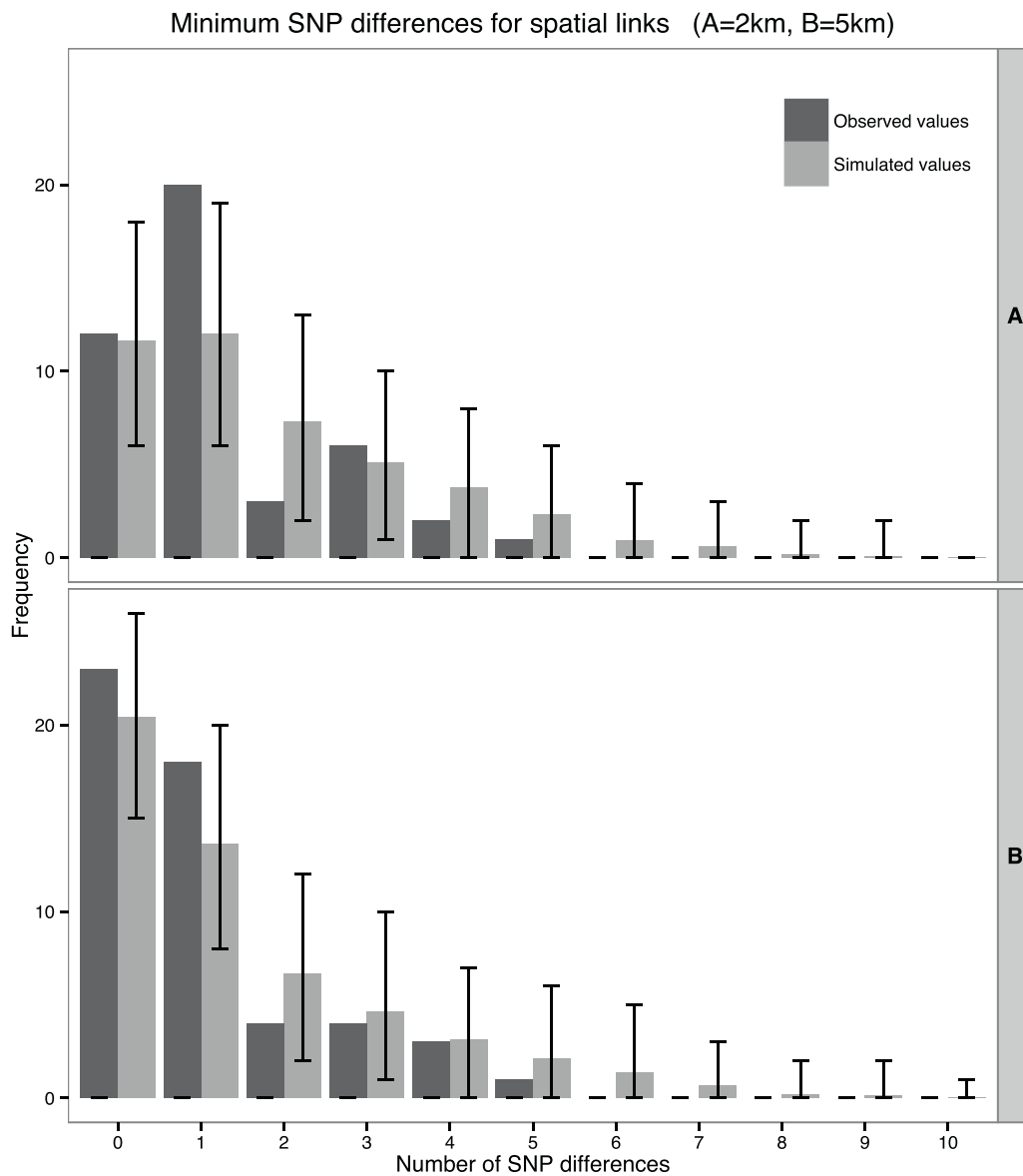


Fig. 4. Observed number of SNP differences between outbreaks linked by spatial proximity (dark grey), and expected SNP differences from 10^4 simulations of the null hypothesis of no association between presence of a link and genetic similarity (light grey). Bars give the intervals containing 95% of the results from the null simulations. A. shows results for spatial proximity of 2 km and B. shows results for 5 km.

(Fig. S9). This suggests that, despite low levels of genetic divergence, our data contain sufficient phylogenetic signal for meaningful phylogeographic inference.

4. Discussion

Applying WGS to an intensively sampled molecular type of *M. bovis* in cattle allowed us to explore the potential of this approach to assess the role of cattle movements or spatial proximity in transmission, and to quantify bacterial dispersal across the landscape. Our findings not only demonstrate the potential of WGS as a tool for epidemiological investigation of bTB, but also clearly expose certain limitations.

4.1. Differential sampling intensity between VNTR-10 clades and switching of VNTR-type

As expected for a well-sampled population of slowly evolving bacteria, the majority of VNTR-10 isolates (specifically: Group 1)

were genetically highly similar, often identical, and included most of the recent ancestral sequence types that can be inferred from the VNTR-10 phylogeny. In contrast, we found a small number of VNTR-10 samples (Group 2, Fig. 1) that were phylogenetically distinct from the Group 1 samples and showed higher pairwise genetic diversity. Sampling and sequencing effort was even across all VNTR-10-typed isolates from cattle, so the finding of a rare group with more divergent isolates was surprising and suggests that our sampling of Group 2 isolates was less complete compared to Group 1. Observing a much smaller proportion of the overall bacterial diversity might have suggested that this lineage was maintained in a host population largely missed by our sampling, such as a non-cattle reservoir host. However, the sequence data presented here from the closely related VNTR-1 strain indicates that the apparent under-sampling in the Group 2 was caused by switching of VNTR phenotype within this lineage. The placement of VNTR-1 isolates C and D implies that the emergence of VNTR-10 from VNTR-1 occurred independently for the Group 1 clade, whereas the relationship between VNTR-1 samples A and B, and VNTR-10 samples in

Group 2 (samples 1–3) suggests that in Group 2 bacteria the VNTR-type has switched VNTR-type multiple times between VNTRs-1 and -10 (Fig. 1). Because our sampling strategy was reliant on VNTR-typing, and focussed on VNTR-10, lineages that had changed to a different type were almost certain to be missed. Similar evidence for type switching from whole genome data has also recently emerged for human TB (Bryant et al., 2013a; Walker et al., 2012) and in other mycobacteria (Ahlstrom et al., 2015).

Given that we have sequenced so few VNTR-1 isolates, the finding that all of them cluster within the VNTR-10 samples is significant, however it is difficult to say whether the Group 2 lineage is genuinely more prone to VNTR-switching than Group 1. Although studies have found only limited phenotypic differences between different molecular types of *M. bovis* (Wright et al., 2013a,b), it is possible that a difference between the two groups of VNTR-10 isolates, in terms of their propensity for VNTR-switching, may be related to functional genetic differences between them. Two out of the four SNPs differentiating the clades show non-synonymous changes in annotated regions, one in the *Sth A* gene and one in a region coding for an unknown hypothetical protein. Additionally, the MV2163B locus which differentiates VNTRs-1 and -10 (Table S2) is known to occur within the open reading frame of PPE gene Rv1917c, one of a family of proteins thought to play a role in antigenic variation (Skuce et al., 2002). Therefore a difference in VNTR-switching between clades could also be a response to selective pressures acting on this VNTR locus. However, MV2163B does not show any greater diversity of tandem repeat variation than other NI VNTR-loci, and if anything appears slightly less prone to variation than the others (Skuce et al., 2005).

4.2. Comparing genetic and epidemiological relationships between breakdowns

An analysis of the association between close spatial proximity among breakdowns (2 km and 5 km) and genetic similarity gave only marginally significant results ($p=0.048$) for distances of under 2 km, and showed no significance for distances of under 5 km (Fig. 4). A similar evaluation of the correlation between direct recorded movements of cattle between breakdowns and genetic similarity also showed no significant associations (Fig. 3).

In contrast to findings for more rapidly evolving bacteria (Eyre et al., 2013), our results suggest that the use of pairwise distances to track transmission and to assess the relative roles of potential transmission mechanisms lacks power for *M. bovis*. The lack of power seen here is likely to be due to the low level of genetic signal in this slowly evolving pathogen as well as the variable, often prolonged, duration of infection within an animal. As a consequence, the genetic distance between breakdowns will be difficult to predict even for breakdowns linked by direct transmission of infection. Adding to this genetic uncertainty, over distances of 2 and 5 km the discrepancy between the registered location of a herd and the actual location of the cattle will be brought to the fore (especially in NI where use of rented pasture is common and level of farm fragmentation is high (Abernethy et al., 2006)), and recorded movements assessed here do not include indirect movements between herds which may also play a role in transmission.

On a broader scale however, the significant, though weak, correlation between genetic and spatial distances within Group 1 breakdowns demonstrated by the Mantel test and multiple regression on matrices indicates that spatially localised mechanisms are likely involved in the transmission of this lineage. This is consistent with earlier findings (Biek et al., 2012), as well as the large-scale patterns of spatial expansion of the bTB-endemic areas in the south-west of Britain (Brunton et al., 2015).

4.3. Application of phylogeographic tools

Despite the limitations described above, phylogenetic data for *M. bovis* can provide insight into the pattern and process of spatial spread. Encouragingly, even taking into account the estimation uncertainties and the low evolutionary rate discussed above, our simulations demonstrate that, genome-wide variation of *M. bovis* contains sufficient information to support meaningful Bayesian phylogeographic analyses over the temporal and spatial scales covered by our data. Based on such an approach, a heterogeneous model of spatial spread fitted the WGS data significantly better than a model assuming a homogeneous diffusion process, implying that the VNTR-10 group has spread across the landscape at a variable rate. Such a pattern might indicate that transmission is underpinned by multiple mechanisms, each associated with a different diffusion process and rate, or it may suggest that transmission is largely driven by a single mechanism involving a variable rate of spatial spread (for example, human-mediated movements of cattle between herds). The low mean spatial diffusion rate of VNTR-10 of 2 km/year fits the observation of strong spatial clustering characteristic of *M. bovis* in the UK, seen over different scales and typing methods (Skuce et al., 2010; Smith et al., 2006).

4.4. Implications for bTB management

Our finding that homoplasy due to VNTR-switching may be common in bTB is significant from an applied point of view since it could confound the epidemiological distinctions made between closely related VNTR-types. However, VNTR-typing in NI is used in conjunction with spoligotyping, and the VNTR loci used have been chosen for optimal discrimination within the NI *M. bovis* population (Skuce et al., 2005), both of which are expected to reduce the impact of VNTR-type homoplasies (Reyes et al., 2012). Of the ten most common VNTR-types in NI (accounting for 85% of all VNTR-typed isolates (Skuce et al., 2010)), only one pair is separated by less than two VNTR tandem repeat differences while sharing the same spoligotype. VNTR-switching is therefore expected to have limited impact on the routine application of VNTR-typing for bTB in NI. Additionally, the VNTR-typing of all culture positive animals in a breakdown (currently standard in NI) will facilitate early detection of VNTR-switching events should they occur.

This study also suggests that it may be possible in principle to use WGS to identify under-sampled populations in *M. bovis*, in this case due to switching of VNTR-type between VNTR-10 and the closely related VNTR-1. However, whether WGS will provide sufficient resolution to characterise the involvement of under-sampled wildlife reservoirs for *M. bovis* is unclear, and inferences will also be affected by the rate at which transmission occurs between the two host populations (Kao et al., 2014).

Despite the exceptional quality of epidemiological data available for *M. bovis* in NI and the high intensity of sampling, we found that four Group 1 VNTR-10 breakdowns in this study showed no links to other VNTR-10 breakdowns, neither through direct recorded movements within 10 years of the breakdown nor through spatial proximity of 5 km or less. Although in Group 2 samples such a finding is likely due to the under-sampling of this lineage due to VNTR-switching, the presence of “unlinked” breakdowns in Group 1 is more surprising, suggesting that the epidemiological links assessed here do not cover all the routes through which infection spreads.

We suggest due to its slow evolutionary rate, some limitations will always be inherent in the application of WGS to *M. bovis* epidemiology and accordingly care must be taken in interpreting results: certain analyses will always remain problematic, for example, unambiguous determination of the underlying transmission tree, “who infected whom” (Didelot et al., 2014; Köser et al., 2012).

However, we show here that other approaches, such as Bayesian phylogeographic techniques to explore the spatial spread of disease, appear feasible for this pathogen.

5. Conclusion

Despite a rate of evolution amongst the lowest recorded to date among bacteria, the genomic data presented show a substantial improvement in genetic resolution over previous methods of genetic typing. While WGS data have considerable potential to enhance both our in-depth understanding of bTB epidemiology as well as routine bTB surveillance, the slow evolutionary rate of *M. bovis* does impose a limit to this potential, as has been noted in human tuberculosis (e.g., Didelot et al., 2014; Köser et al., 2012). For the future, we suggest that continued advances in mathematical models integrating epidemiological and genetic information will allow a more confident resolution of the factors involved in the spread of bovine tuberculosis, giving a better understanding of the interplay between epidemiological and genetic factors for this important and troubling pathogen.

Funding

The Wellcome Senior Research Fellowship acknowledged as a funder has a Wellcome Reference 081696/Z/06/Z.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.epidem.2015.08.003.

References

- Abernethy, D.A., Denny, G.O., Menzies, F.D., McGuckian, P., Honhold, N., Roberts, A.R., 2006. The Northern Ireland programme for the control and eradication of *Mycobacterium bovis*. *Vet. Microbiol.* 112, 231–237, <http://dx.doi.org/10.1016/j.vetmic.2005.11.023>.
- Abernethy, D.A., Walton, E., Menzies, F., Courcier, E., Robinson, P., 2011. *Mycobacterium bovis* surveillance in European badgers (*Meles meles*) killed by vehicles in Northern Ireland: an epidemiological evaluation. In: *International Conference on Animal Health Surveillance. Association pour l'Étude de l'Épidémiologie des Maladies Animales*, Lyon, France, pp. 216–218.
- Ahlstrom, C.A., Barkema, H.W., Stevenson, K., Zadoks, R.N., Biek, R., Kao, R., Trewby, H., Hendrick, S., Hauptstein, D., Kelton, D., Fecteau, G., Labrecque, O., Keefe, G., McKenna, S., De Buck, J., 2015. Limitations of variable number of tandem repeat typing identified through whole genome sequencing of *Mycobacterium avium* subsp. paratuberculosis on a national and herd level. *BMC Genom.* 16, 161, <http://dx.doi.org/10.1186/s12864-015-1387-6>.
- Baële, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* 29, 2157–2167, <http://dx.doi.org/10.1093/molbev/mss084>.
- Biek, R., O'Hare, A., Wright, D., Mallon, T., McCormick, C., Orton, R.J., McDowell, S., Trewby, H., Skuce, R.A., Kao, R.R., 2012. Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathog.* 8, e1003008, <http://dx.doi.org/10.1371/journal.ppat.1003008>.
- Biek, R., Pybus, O.G., Lloyd-Smith, J.O., Didelot, X., 2015. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* 30, 306–313, <http://dx.doi.org/10.1016/j.tree.2015.03.009>.
- Brooks-Pollock, E., Roberts, G.O., Keeling, M.J., 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature* 511, 228–231, <http://dx.doi.org/10.1038/nature13529>.
- Brunton, L.A., Nicholson, R., Ashton, A., Alexander, N., Wint, W., Enticott, G., Ward, K., Broughan, J.M., Goodchild, A.V., 2015. A novel approach to mapping and calculating the rate of spread of endemic bovine tuberculosis in England and Wales. *Spat. Spatiotemporal. Epidemiol.* 13, 41–50, <http://dx.doi.org/10.1016/j.sste.2015.04.002>.
- Bryant, J.M., Harris, S.R., Parkhill, J., Dawson, R., Diacon, A.H., van Helden, P., Pym, A., Mahayiddin, A.A., Chuchottaworn, C., Sanne, I.M., Louw, C., Boeree, M.J., Hoelscher, M., McHugh, T.D., Bateson, A.L.C., Hunt, R.D., Mwaigwisya, S., Wright, L., Gillespie, S.H., Bentley, S.D., 2013a. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir. Med.* 1, 786–792, [http://dx.doi.org/10.1016/S2213-2600\(13\)70231-5](http://dx.doi.org/10.1016/S2213-2600(13)70231-5).
- Bryant, J.M., Schürch, A.C., van Deutekom, H., Harris, S.R., de Beer, J.L., de Jager, V., Kremer, K., van Hijum, S.A.F.T., Siezen, R.J., Borgdorff, M., Bentley, S.D., Parkhill, J., van Soolingen, D., 2013b. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect. Dis.* 13, 110, <http://dx.doi.org/10.1186/1471-2334-13-110>.
- Byrne, A.W., Quinn, J.L., O'Keefe, J.J., Green, S., Sleeman, D.P., Martin, S.W., Davenport, J., 2014. Large-scale movements in European badgers: has the tail of the movement kernel been underestimated? *J. Anim. Ecol.* 83, 991–1001, <http://dx.doi.org/10.1111/1365-2656.12197>.
- Colangeli, R., Arcus, V.L., Cursons, R.T., Ruthe, A., Karalus, N., Coley, K., Manning, S.D., Kim, S., Marchiano, E., Alland, D., 2014. Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. *PLoS One* 9, e91024, <http://dx.doi.org/10.1371/journal.pone.0091024>.
- Courtenay, O., Reilly, L.A., Sweeney, F.P., Hibberd, V., Bryan, S., Ul-Hassan, A., Newman, C., Macdonald, D.W., Delahay, R.J., Wilson, G.J., Wellington, E.M.H., Ul-Hassan, A., Macdonald, D.W., 2006. Is *Mycobacterium bovis* in the environment important for the persistence of bovine tuberculosis? *Biol. Lett.* 2, 460–462, <http://dx.doi.org/10.1098/rsbl.2006.0468>.
- Cousins, D.V., Skuce, R.A., Kazwala, R.R., van Embden, J.D.A., 1998. Towards a standardised approach to DNA fingerprinting of *Mycobacterium bovis*. *Int. J. Tuberc. Lung Dis.* 2, 471–478.
- Delahay, R.J., Cheeseman, C.L., Clifton-Hadley, R.S., 2001. Wildlife disease reservoirs: the epidemiology of *Mycobacterium bovis* infection in the European badger (*Meles meles*) and other British mammals. *Tuberculosis* 81, 43–49, <http://dx.doi.org/10.1054/tube.2000.0266>.
- Didelot, X., Gardy, J., Colijn, C., 2014. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol. Biol. Evol.* 31, 1869–1879, <http://dx.doi.org/10.1093/molbev/msu121>.
- Donnelly, C.A., Nouvellet, P., 2013. The contribution of badgers to confirmed tuberculosis in cattle in high-incidence areas in England. *PLoS Curr. Outbreaks* 1, 1–15, <http://dx.doi.org/10.1371/currents.outbreaks.097a904d3f3619db2fe78d24bc776098>. Abstract.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973, <http://dx.doi.org/10.1093/molbev/mss075>.
- Eyre, D.W., Cule, M.L., Wilson, D.J., Griffiths, D., Vaughan, A., O'Connor, L., Ip, C.L.C., Golubchik, T., Batty, E.M., Finney, J.M., Wyllie, D.H., Didelot, X., Piazza, P., Bowden, R., Dingle, K.E., Harding, R.M., Crook, D.W., Wilcox, M.H., Peto, T.E.A., Walker, A.S., 2013. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N. Engl. J. Med.* 369, 1195–1205, <http://dx.doi.org/10.1056/NEJMoa1216064>.
- Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., Rambaut, A., 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* 27, 2038–2051, <http://dx.doi.org/10.1093/molbev/msq088>.
- Gallagher, J., Clifton-Hadley, R.S., 2000. Tuberculosis in badgers—a review of the disease and its significance for other animals. *Res. Vet. Sci.* 69, 203–217, <http://dx.doi.org/10.1053/rvsc.2000.0422>.
- Garnier, T., Eiglmeier, K., Camus, J.C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V., Hewinson, R.G., 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci.* 100, 7877–7882, <http://dx.doi.org/10.1126/science.1182395>.
- Gilbert, M., Mitchell, A., Bourn, D., Mawdesley, J., Clifton-Hadley, R., Wint, W., 2005. Cattle movement and bovine tuberculosis in Great Britain. *Nature* 435, 491–496, <http://dx.doi.org/10.1038/nature03548>.
- Goslee, S.C., Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* 22, 1–19.
- Gray, R.R., Tatem, A.J., Johnson, J.A., Alekseyenko, A.V., Pybus, O.G., Suchard, M.A., Salemi, M., 2011. Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a Bayesian framework. *Mol. Biol. Evol.* 28, 1593–1603, <http://dx.doi.org/10.1093/molbev/msq319>.
- Green, D.M., Kiss, I.Z., Mitchell, A.P., Kao, R.R., 2008. Estimates for local and movement-based transmission of bovine tuberculosis in British cattle. *Proc. R. Soc. B* 275, 1001–1005, <http://dx.doi.org/10.1098/rspb.2007.1601>.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704, <http://dx.doi.org/10.1080/10635150390235520>.
- Harris, S.R., Feil, E.J., Holden, M.T.G.J., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J.A., Edgeworth, J.D., de Lencastre, H., Parkhill, J., Peacock, S.J., Bentley, S.D., Traverses, A., 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327, 469–474, <http://dx.doi.org/10.1126/science.1182395>.
- Holden, M.T.G., Hsu, L., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Zemlickova, J., Coombs, G., Kearns, A.M., Hill, R.L., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G.F., McAdam, P.R., Templeton, K.E., McCann, A., Zhou, Z., Castillo-Ramirez, S., Feil, E.J., Hudson, L.O., Enright, M.C., Balloux, F., Aanensen, D.M., Spratt, B.G., Fitzgerald, J.R., Parkhill, J., Achtman, M., Bentley, S.D., Nubel, U., 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* 23, 653–664, <http://dx.doi.org/10.1101/gr.147710.112>. Freely.
- Houston, R., 2001. A computerised database system for bovine traceability. *Rev. Sci. Tech. Int. Off. Epizoot.* 20, 652–661.

- Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* 35, 907–914.
- Kao, R.R., Haydon, D.T., Lycett, S.J., Murcia, P.R., 2014. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol.* 22, 282–291, <http://dx.doi.org/10.1016/j.tim.2014.02.011>.
- Köser, C.U., Ellington, M.J., Cartwright, E.J.P., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T.G., Dougan, G., Bentley, S.D., Parkhill, J., Peacock, S.J., 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 8, e1002824, <http://dx.doi.org/10.1371/journal.ppat.1002824>.
- Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* 27, 1877–1885, <http://dx.doi.org/10.1093/molbev/msq067>.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760, <http://dx.doi.org/10.1093/bioinformatics/btp324>.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAM tools. *Bioinformatics* 25, 2078–2079, <http://dx.doi.org/10.1093/bioinformatics/btp352>.
- Lichstein, J.W., 2007. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol.* 188, 117–131, <http://dx.doi.org/10.1007/s11258-006-9126-3>.
- Maddock, E.C.G., 1933. Studies on the survival time of the bovine tubercle bacillus in soil, soil and dung, in dung and on grass, with experiments on the preliminary treatment of infected organic matter and the cultivation of the organism. *J. Hyg. (Lond.)* 33, 103–117.
- Mather, A.E., Reid, S.W.J., Maskell, D.J., Parkhill, J., Fookes, M.C., Harris, S.R., Brown, D.J., Coia, J.E., Mulvey, M.R., Gilmour, M.W., Petrovska, L., de Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T.R., Suchard, M.A., Lemey, P., Mellor, D.J., Haydon, D.T., Thomson, N.R., 2013. Distinguishable epidemics of multidrug-resistant *Salmonella typhimurium* DT104 in different hosts. *Science* 341, 1514–1517, <http://dx.doi.org/10.1126/science.1240578>.
- Mazars, E., Lesjean, S., Banuls, A.L., Gilbert, M., Vincent, V., Gicquel, B., Tibaïre, M., Locht, C., Supply, P., 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci.* 98, 1901–1906, <http://dx.doi.org/10.1073/pnas.98.4.1901>.
- Pope, L.C., Domingo-Roura, X., Erven, K., Burke, T., 2006. Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. *Mol. Ecol.* 15, 371–386, <http://dx.doi.org/10.1111/j.1365-294X.2005.02815.x>.
- R Core Team and R Development Core Team, 2014. R: A Language and Environment for Statistical Computing.
- Reyes, J.F., Chan, C.H.S., Tanaka, M.M., 2012. Impact of homoplasmy on variable numbers of tandem repeats and spoligotypes in *Mycobacterium tuberculosis*. *Infect. Genet. Evol.* 12, 811–818, <http://dx.doi.org/10.1016/j.meegid.2011.05.018>.
- Roetzer, A., Diel, R., Kohl, T.A., Rückert, C., Nübel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rüschi-Gerdes, S., Supply, P., Kalinowski, J., Niemann, S., 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med.* 10, e1001387, <http://dx.doi.org/10.1371/journal.pmed.1001387>.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574, <http://dx.doi.org/10.1093/bioinformatics/btg180>.
- Skuce, R.A., Mallon, T.R., McCormick, C.M., McBride, S.H., Clarke, G., Thompson, A., Couzens, C., Gordon, A.W., McDowell, S.W.J., 2010. *Mycobacterium bovis* genotypes in Northern Ireland: herd-level surveillance (2003–2008). *Vet. Rec.* 167, 684–689, <http://dx.doi.org/10.1136/vr.c5108>.
- Skuce, R.A., McCorry, T.P., McCarrroll, J.F., Roring, S.M.M., Scott, A.N., Brittain, D., Hughes, S.L., Hewinson, R.G., Neill, S.D., 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* 148, 519–528.
- Skuce, R.A., McDowell, S.W., Mallon, T.R., Luke, B., Breadon, E.L., Lagan, P.L., McCormick, C.M., McBride, S.H., Pollock, J.M., 2005. Discrimination of isolates of *Mycobacterium bovis* in Northern Ireland on the basis of variable numbers of tandem repeats (VNTRs). *Vet. Rec.* 157, 501–504.
- Skuce, R.A., Neill, S.D., 2001. Molecular epidemiology of *Mycobacterium bovis*: exploiting molecular data. *Tuberculosis* 81, 169–175, <http://dx.doi.org/10.1054/tube.2000.0270>.
- Smith, N.H., Gordon, S.V., de la Rúa-Domenech, R., Clifton-Hadley, R.S., Hewinson, R.G., 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.* 4, 670–681, <http://dx.doi.org/10.1038/nrmicro1472>.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739, <http://dx.doi.org/10.1093/molbev/msr121>.
- Van Soolingen, D., de Haas, P.E.W., Kremer, K., 2001. Restriction fragment length polymorphism typing of *Mycobacteria*. *Methods Mol. Med.* 54, 165–203, <http://dx.doi.org/10.1385/1-59259-147-7:165>.
- Viana, M., Mancy, R., Biek, R., Cleaveland, S., Cross, P.C., Lloyd-Smith, J.O., Haydon, D.T., 2014. Assembling evidence for identifying reservoirs of infection. *Trends Ecol. Evol.*, 1–10, <http://dx.doi.org/10.1016/j.tree.2014.03.002>.
- Walker, T.M., Ip, C.L.C., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W., Parkhill, J., Harris, D., Walker, A.S., Bowden, R., Monk, P., Smith, E.G., Peto, T.E.A., 2012. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.*, 3099, [http://dx.doi.org/10.1016/S1473-3099\(12\)70277-3](http://dx.doi.org/10.1016/S1473-3099(12)70277-3).
- Wright, D.M., Allen, A.R., Mallon, T.R., McDowell, S.W.J., Bishop, S.C., Glass, E.J., Bermingham, M.L., Woolliams, J.A., Skuce, R.A., 2013a. Field-isolated genotypes of *Mycobacterium bovis* vary in virulence and influence case pathology but do not affect outbreak size. *PLoS One* 8, e74503, <http://dx.doi.org/10.1371/journal.pone.0074503>.
- Wright, D.M., Allen, A.R., Mallon, T.R., McDowell, S.W.J., Bishop, S.C., Glass, E.J., Bermingham, M.L., Woolliams, J.A., Skuce, R.A., 2013b. Detectability of bovine TB using the tuberculin skin test does not vary significantly according to pathogen genotype within Northern Ireland. *Infect. Genet. Evol.* 19, 15–22, <http://dx.doi.org/10.1016/j.meegid.2013.05.011>.
- Young, J.S., Gormley, E., Wellington, E.M.H., Elizabeth, M.H., Wellington, E.M.H., 2005. Molecular detection of *Mycobacterium bovis* and *Mycobacterium bovis* BCG (Pasteur) in soil. *Appl. Environ. Microbiol.* 71, 1946–1952, <http://dx.doi.org/10.1128/AEM.71.4.1946>.