Veterinary Microbiology 268 (2022) 109416

Contents lists available at ScienceDirect

Veterinary Microbiology

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

Evidence for local and international spread of *Mycobacterium avium* subspecies *paratuberculosis* through whole genome sequencing of isolates from the island of Ireland

Viktor Perets [a], Adrian Allen [b], Joseph Crispell [a,1], Sophie Cassidy [a,2], Aoife O'Connor [c], Damien Farrell [a], John A. Browne [d], Jim O'Mahony [e], Robin Skuce [b], Kevin Kenny [c], Stephen V. Gordon [a,f,g,h,*]

[a] UCD School of Veterinary Medicine, University College Dublin, Dublin 4, Ireland
[b] Agri-Food and Biosciences Institute, AFBI Stormont, Belfast, UK
[c] Central Veterinary Research Laboratory, Department of Agriculture, Food and the Marine, Backweston, Co. Kildare, Ireland
[d] UCD School of Agriculture and Food Science, University College Dublin, Dublin 4, Ireland
[e] Munster Technological University, Department of Biological Sciences, Rossa Avenue, Bishopstown, Cork, Ireland
[f] UCD School of Medicine, University College Dublin, Dublin 4, Ireland
[g] UCD School of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland
[h] UCD Conway Institute, University College Dublin, Dublin 4, Ireland

ARTICLE INFO

Keywords:
Whole genome sequencing
Mycobacterium avium subsp. *paratuberculosis*
Johne's disease
Ireland

ABSTRACT

We describe application of whole genome sequencing (WGS) to a collection of 197 *Mycobacterium avium* subsp *paratuberculosis* (MAP) isolates gathered from 122 cattle herds across 27 counties of the island of Ireland. We compare WGS to MAP diversity quantified using mycobacterial interspersed random unit – variable number tandem repeats (MIRU-VNTR). While MIRU-VNTR showed only two major types, WGS could split the 197 isolates into eight major groups. We also found six isolates corresponding to INMV 13, a novel MIRU-VNTR type for Ireland. Evidence for dispersal of MAP across Ireland via cattle movement could be discerned from the data, with mixed infections present in several herds. Furthermore, comparisons of MAP WGS data from Ireland to data from Great Britain and continental Europe revealed many instances of close genetic similarity and hence evidence for international transmission of infection. BEAST MASCOT structured coalescent analyses, with relaxed and strict molecular clocks, estimated the substitution rate to be 0.10–0.13 SNPs/site/year and disclosed greater transitions per lineage per year from Europe to Ireland, indicating transmission into Ireland. Our work therefore reveals new insight into the seeding of MAP infection across Ireland, highlighting how WGS can inform policy formulation to ultimately control MAP transmission at local, national and international scales.

1. Introduction

Mycobacterium avium subspecies *paratuberculosis* (MAP) is the causative agent of Johne's Disease (JD), a chronic enteritis of ruminants and other animals (Whittington et al., 2019). MAP infection is a threat to animal welfare, production, and trade, with many countries recommending the implementation of control programmes and establishment of a global standard for MAP control (Whittington et al., 2019). The US agricultural economy loses approximately $198 million annually due to JD, while in Ireland an estimated $25 million is lost annually with Irish herd-level true prevalence estimated at 20% (Rasmussen et al., 2021).

Historically, MAP has been genotyped using pulsed field gel electrophoresis (PFGE), restriction fragment length polymorphism (RFLP) typing (de Juan et al., 2005) and mycobacterial interspersed random unit variable number tandem repeats (MIRU-VNTR) (Thibault et al., 2007). Collins originally used RFLP to classify the "Type-S" (sheep) and "Type-C" (cattle) strains; these were later respectively reclassified to "Type I" and "Type II", with an additional "Type III" recognised as a

* Correspondence to: Veterinary Sciences Centre, School of Veterinary Medicine, University College Dublin, Room 037, Dublin 4, Ireland.
E-mail address: stephen.gordon@ucd.ie (S.V. Gordon).
[1] Current address: Office for National Statistics, FCDO Data Hub, East Kilbride G75 8EA, United Kingdom.
[2] Current address: Royal College of Surgeons in Ireland, Dublin 2, Ireland.

https://doi.org/10.1016/j.vetmic.2022.109416
Received 13 July 2021; Received in revised form 14 January 2022; Accepted 1 April 2022
Available online 5 April 2022

subset of "Type-S" (de Juan et al., 2005). The host of isolation did not always conform to the assigned MAP type, and it was recognised that RFLP alone wasn't sufficiently discriminatory to provide a robust basis for molecular epidemiological studies (de Juan et al., 2005). MIRU-VNTR typing was introduced in 2007 for MAP as a cheap, fast, and reproducible genotyping method (Thibault et al., 2007), and it has been widely used (de Kruijf et al., 2017; Stevenson et al., 2009). MIRU-VNTR targets eight repeat regions in the MAP genome, which are PCR-amplified and sized after gel electrophoresis (Thibault et al., 2007).

The application of whole genome sequencing (WGS) to collections of MAP isolates was first reported in studies published by Ahlstrom et al. (2016, 2015) and Bryant et al. (2016). Ahlstrom and colleagues' research compared a collection of VNTR-typed dairy cattle isolates to a subset of those that had been genome sequenced. Their initial WGS analysis identified eight genetically discrete groups of MAP and suggested numerous distinct introductions of MAP into Canada (Ahlstrom et al., 2015). In their VNTR data, Ahlstrom and colleagues observed homoplasy – a situation where isolates shared VNTR profiles but had different ancestry. They concluded that for herd level comparison, VNTR was inadequate to determine diversity as compared to WGS. In a subsequent paper (Ahlstrom et al., 2016), they compared their 2015 MAP WGS dataset to a collection of global isolates. They identified global subtypes of MAP in the Canadian MAP population, supporting their idea of multiple introduction events of MAP into Canada via trade. In 2016, Bryant and colleagues described WGS data for a global collection of MAP isolates (Bryant et al., 2016). They found that MAP has limited genetic diversity, with a substitution rate of less than 0.5 SNPs/genome/year. These authors also showed that WGS was a superior alternative to classical typing methods (IS*1311* RFLP and MIRU-VNTR), and similar to Ahlstrom and colleagues, concluded that VNTR both over- and under-estimated genetic diversity of MAP.

In Ireland, two reports have described the use of VNTR for genotypic analysis of MAP (de Kruijf et al., 2017; Douarre et al., 2011). De Kruijf et al. analysed 114 MAP isolates from 53 herds across 19 counties in Ireland genotyped with 8 locus VNTR, found that 98.2% of these MAP isolates fell within VNTR groups INMV 1 and INMV 2, suggesting low genetic diversity of MAP in Ireland. However, given the known limited resolution of MIRU-VNTR for MAP in comparison to WGS (Ahlstrom et al., 2015; Bryant et al., 2016), a necessary next step was to apply WGS to MAP isolates from Ireland to determine their diversity and provide a basis for the application of WGS for studies of MAP transmission and persistence of infection.

Here we used WGS on a panel of MAP isolates from the island of Ireland (i.e., encompassing both the Republic of Ireland and Northern Ireland). Some of these MAP isolates had previously been VNTR profiled (de Kruijf et al., 2017), while we also generated VNTR and WGS data for a selection of new, previously un-typed isolates. Our aim was to use WGS data to determine the genetic diversity of MAP across the island of Ireland, to evaluate the agreement of WGS and VNTR data, and to compare the Irish WGS data to that of existing British and European WGS data (Bryant et al., 2016). We also attempted to estimate the transitions of MAP between Great Britain and Europe versus Ireland, and determine the rate of substitution for comparison with the previously reported rate of < 0.5 SNPs/genome/year (Bryant et al. 2016).

2. Materials and methods

2.1. Isolate selection

A collection of 197 isolates from MAP culture collection biobanks (held at the Central Veterinary Research Laboratory (CVRL), Co. Kildare, Ireland and Agri-Food and Biosciences Institute (AFBI), Co. Antrim, Northern Ireland) were selected from 122 herds across 27 counties across the island of Ireland to maximise geographic spread and capture broad genetic diversity (see Supplementary Fig. 1 for isolate distribution map). All isolates were cultured from cattle faecal material collected between 2013 and 2017 and 2019. A subset (n = 59) of isolates from the 114 Irish MAP isolates used by de Kruijf et al. (2017) was included which had previously been MIRU-VNTR typed (only 59 of the 114 were used as we could not recover viable cells from frozen culture stocks or stored faecal samples).

2.2. Culture of isolates

MAP bacteria were recovered from glycerol stocks, or re-isolated from cattle faecal samples, at CVRL Backweston via growth at 37 °C on Herrold's Egg Yolk Agar slopes with amphotericin B, nalidixic acid and vancomycin (ANV) and Mycobactin J supplements (HEYA; Becton Dickinson, UK). Once colonies were visible, growth was scraped from the HEYA slope and transferred into 10 ml Middlebrook 7H9 (Becton Dickinson, UK) supplemented with 10% (v/v) oleic albumin dextrose catalase (OADC; Becton Dickinson, UK), 0.2% (v/v) glycerol (Sigma, Ireland), 0.05% (v/v) Tween 80 (Sigma, Ireland), 2 μg/ml mycobactin J (Allied Monitor, France) and 0.025% (v/v) polymyxin B, amphotericin B, nalidixic acid, trimethoprim and azlocillin (PANTA). The isolates were grown in a standing incubator at 37 °C for a minimum of eight weeks, until a visible pellet was present in the tube.

2.3. DNA extraction

MAP cells were lysed using a bead beater (MagNA Lyser, Roche) and the DNA was extracted by sodium acetate/ethanol precipitation. Crude MAP DNA extracts were then purified with Ampure XP magnetic purification beads (Beckman Coulter, UK) and resuspended in Tris-EDTA buffer (TE; Fisher Scientific, UK). The step-by-step DNA extraction protocol is available on the GitHub repository of the lead author (https://github.com/ViktorP7/MAP-Publication/blob/master/Protocols/DNA-Extraction-Protocol.md). MAP DNA extracts were assessed for DNA quantity and quality using a Nanodrop spectrophotometer (NDW1000). DNA content was quantified with a Qubit dsDNA BR Assay kit (Invitrogen) following the manufacturer's instructions. DNA extracts were stored at −20 °C pending library preparation and MIRU-VNTR analyses.

2.4. Library preparation and sequencing

The library preparation of each sample used 50 ng of DNA. Library preparations were carried out using Nextera DNA Flex kits (Illumina, USA) as per the manufacturer's protocol with a minor modification of using PCR strip-tubes instead of 96-well plates. Library DNA content was quantified using a Qubit dsDNA HS Assay kit (Invitrogen), and fragment sizes were checked using the DNA chip system on a Bioanalyser 2100 (Agilent) prior to pooling. The standardised pool was adjusted to a recommended concentration (2 nM) before denaturing and sequencing. Sequencing was carried out on a NextSeq500 machine (Illumina, USA) at the Genomics Core, UCD Conway Institute, generating 150 bp paired end reads. Sequence files are available under BioProject ID PRJNA739165.

2.5. Processing of sequencing files and phylogenetic trees

MAP WGS output files were downloaded from the BaseSpace Illumina cloud service. Each pair of.fastq files was examined using FastQC (Andrews, 2010) to determine quality and read trimming parameters. Trimming was performed using Cutadapt (Martin, 2011). Reads were then aligned to the MAP K10 reference genome (Li et al., 2005) using bwa mem (Li and Durbin, 2010). Samtools and bcftools (Li, 2011) were used to process the bwa mem output and call variants, creating variant calling format (VCF) files for each isolate.VCF files were merged and filtered using a read depth of 30, high quality base depth of 4, mapping quality of 30 and allele support of 0.95. For any sites failing initial criteria, a second filter step (termed "rescuing") was performed using

high quality base depth of 5, assuming these sites had passed in other isolates. A SNP alignment FASTA file was then created using the filtered-rescued variants, excluding variants within 10 bp of each other to avoid difficult to sequence or high selection loci (Crispell et al., 2019; Crispell et al., 2017). Using the FASTA file as input, a maximum likelihood phylogeny was constructed using RAxML (Stamatakis, 2014) with n = 100 bootstrap value. Tree visualisation was carried out in R (R Core Team, 2018) using the "ape" (Paradis and Schliep, 2019) package. All additional plots were also generated in R. All R scripts used in this study are available on GitHub (https://github.com/ViktorP7/MAP-Publication).

WGS data from a previously published study of a global MAP collection (Bryant et al., 2016) were included to enable comparison of MAP sequences from the island of Ireland to European and British MAP sequences (the lack of breadth of contemporary data available from outside Europe did not allow robust comparisons of Irish MAP to global MAP strains). These data were downloaded from the European Nucleotide Archive (ENA) using accession PRJEB2204 (Bryant et al., 2016). For this study, only the "C" type isolates were used. For a breakdown of isolate numbers used in analyses, see Table 1.

2.6. MIRU-VNTR typing

Isolates used in this study without existing MIRU-VNTR profiles (n = 138) were typed using the standard eight MAP VNTR loci using primers previously described by Thibault and colleagues (Thibault et al., 2007). Reaction mixtures were as follows: 9.25 μl distilled H_2O, 5 μl 5x Green GoTaq Flexi Buffer (Promega), 2 μl 25 mM $MgCl_2$ (Promega), 0.5 μl 10 mM dNTPs (Promega), 0.25 μl (5 U/μl) GoTaq G2 Flexi DNA polymerase (Promega), 1 μl (100 μM) forward primer, 1 μl (100 μM) reverse primer and 1 μl of DNA (≥ 50 ng/μl) for a total reaction volume of 20 μl. A 1 μl volume of dimethyl sulfoxide (DMSO) and/or 5 μl 5 M betaine were added as outlined in Thibault et al. (2007). PCR was carried out using a SimpliAmp Thermal Cycler (Applied Biosystems) with the following conditions: an initial 5 min incubation at 94 °C, followed by 40 cycles of 30 s at 94 °C, 30 s at 61 °C, 30 s at 72 °C, and a final 7 min cycle at 72 °C. The resulting products were run on a 1.5% agarose gel for 1 h at 100 V to determine the number of repeats present at each of the eight loci.

2.7. Structured coalescent analysis

The assembled dataset provided an opportunity to assess the influence of other European nations on the diversity and population structure of MAP in Ireland. To this end, we undertook the following analyses.

2.7.1. Preliminary phylogenetics and investigating the presence of a temporal signal

A panel of WGS data from 253 MAP isolates with complete information regarding date of isolation and location within Europe was selected (see Table 1). A barplot detailing the frequency of isolates available from both territories is shown in Supplementary Fig. 2.

We used the 'modelTest' function of the package 'Phangorn' (Schliep, 2011), in the R environment (R Core Team, 2018), to assess the most appropriate allele substitution model for our dataset. Specifically, we assessed the fits of the General Time Reversible (GTR), Jukes Cantor, and Hasegawa Kishino Yano (HKY) models to the data. The model with the lowest AIC was then used to build a maximum likelihood phylogeny in RAxML v8 (Stamatakis, 2014), using the rapid bootstrapping search method, stopped after 400 replicates. The phylogeny was visualised in FigTree v1.4.4 (Rambaut et al., 2018).

We used the programme TempEst v1.5.1 (Rambaut et al., 2016) to perform an initial, conservative analysis of root to tip SNP distance vs date of isolation, allowing the programme to select the best fitting root and assess the root to tip divergence model.

To assess the presence or absence of a temporal signal more robustly, we ran a simple coalescent model in BEAST 2 with both relaxed lognormal and strict molecular clocks, assuming constant population size, using the HKY+G allele substitution model and a chain length of 300,000,000 MCMC replicates. Alongside this analysis, we also ran ten replicates of randomised tip dates for the same 253 isolates as per the method of Firth et al. (2010). Tip dates were randomised using the 'Tipdatingbeast' package (Rieux and Khatchikian, 2017) in R. Each tip date randomised dataset was subjected to the same simple coalescent model as described above. All logfiles were inspected in Tracer 1.7.1 (Rambaut et al., 2018) to ensure chain convergence after a 10% burn-in and ESS nos. of 200 or greater for all parameters. Median substitution rates for all models were noted.

2.7.2. MASCOT analysis

To determine inter-regional transition rates between Ireland and the rest of Europe, we used a Bayesian Structured coalescent method, executed in BEAST 2 (Bouckaert et al., 2014). Structured coalescent methods such as those implemented in MASCOT (Marginal Approximation of the Structured Coalescent (Müller et al., 2018)), can estimate ancestral population structure, effective population sizes and transition rates even in the presence of biased sampling (De Maio et al., 2015; Müller et al., 2018, 2017).

We implemented a MASCOT analysis for our 253 isolate dataset as described above. Specifically, we used relaxed log-normal clock models, as these have been observed to best describe the rates of molecular evolution in other slow growing mycobacterial species such as *Mycobacterium bovis* (Crispell et al., 2019, 2017) and *Mycobacterium tuberculosis* (Menardo et al., 2019), but also for MAP (Ahlstrom et al., 2016; Bryant et al., 2016). Also, Biek et al. (2015) have noted before that relaxed clock methods may be more suitable for microbial species, which exhibit among-branch variation in evolutionary rates, a fact shown to be empirically true by Duchêne et al. (2016) (Biek et al., 2015). We also decided to run the MASCOT analyses using a strict clock in case the assumptions on clock model materially affected the outcomes. For both lognormal relaxed clock and strict clock models, we ran 3 replicates of the MASCOT analyses with a chain length of 600,000,000 MCMC replicates, with sampling occurring every 1000 steps. Full details of analyses are supplied in the xml files (Supplementary dataset 1). Following the removal of a 10% burn-in, chains were combined using LogCombiner v2.6 (Bouckaert et al., 2014) and analyses were compared based upon the log likelihood scores, model convergence and posterior support of parameters in TRACER v1.7.1 (Rambaut et al., 2018). ESS nos. of 200 or greater for all parameters were ensured. A Maximum Clade Credibility (MCC) tree was constructed for combined chains using TreeAnnotator v2.6 (Bouckaert et al., 2014) using the median ancestor heights criterion.

Table 1
Summary of MAP isolates used in each analysis.

Analysis	Irish Isolates from this study	Irish Isolates from Bryant et al. (2016)	Isolates from GB and continental Europe from Bryant et al. (2016)	Total
VNTR & WGS (from this paper)	197	0	0	197
Combined Phylogeny of our data with Bryant et al. (2016) dataset	197	8	82	287
BEAST/MASCOT using all isolates for which date of isolation available	197	8	48	253

3. Results

3.1. MIRU-VNTR typing and WGS

MIRU-VNTR profiles for all 197 isolates used in this study were either obtained from existing data (de Kruijf et al., 2017) or generated using standard protocols (Thibault et al., 2007). Five different VNTR patterns were observed. Overall, 95.4% of isolates fell into either INMV 1 or 2; 90/197 (45.7%) isolates were assigned INMV 1 (4–2–3–3–2–2–2–8 repeats at the 8 loci analysed), and 98/197 (49.7%) were assigned INMV 2 (3–2–3–3–2–2–2–8 repeat profile). Two isolates were reported as INMV 3 (3–2–3–3–2–2–1–8 repeat profile), and one was assigned to INMV 116 (4–1–3–3–2–2–2–8 repeat profile). Six isolates were INMV 13 (2–2–3–3–2–2–2–8 repeats repeat profile), which to our knowledge, is the first reported occurrence of this INMV type in Ireland.

Of the 31 herds represented by > 1 isolate, 16 herds contained isolates with mixed INMV types. One INMV 13 isolate was present in herd 'Cork 4', which had one other isolate with INMV 1, the remaining five INMV 13 isolates were identified in herd 'Down Downpatrick Herd C' (DoC), which also contained two type INMV 2 isolates and one INMV 1 isolate. The second INMV 3 isolate was identified in herd 'Westmeath 1', which also contained four INMV 2 isolates. The remaining mixed INMV herds consisted of types 1 and 2.

To compare the VNTR data to WGS, a phylogenetic tree was built with the tip labels coloured based on the corresponding INMV for that isolate (Fig. 1). Based on the WGS data in the tree, eight distinct MAP groups were identified, the maximum difference between isolates was

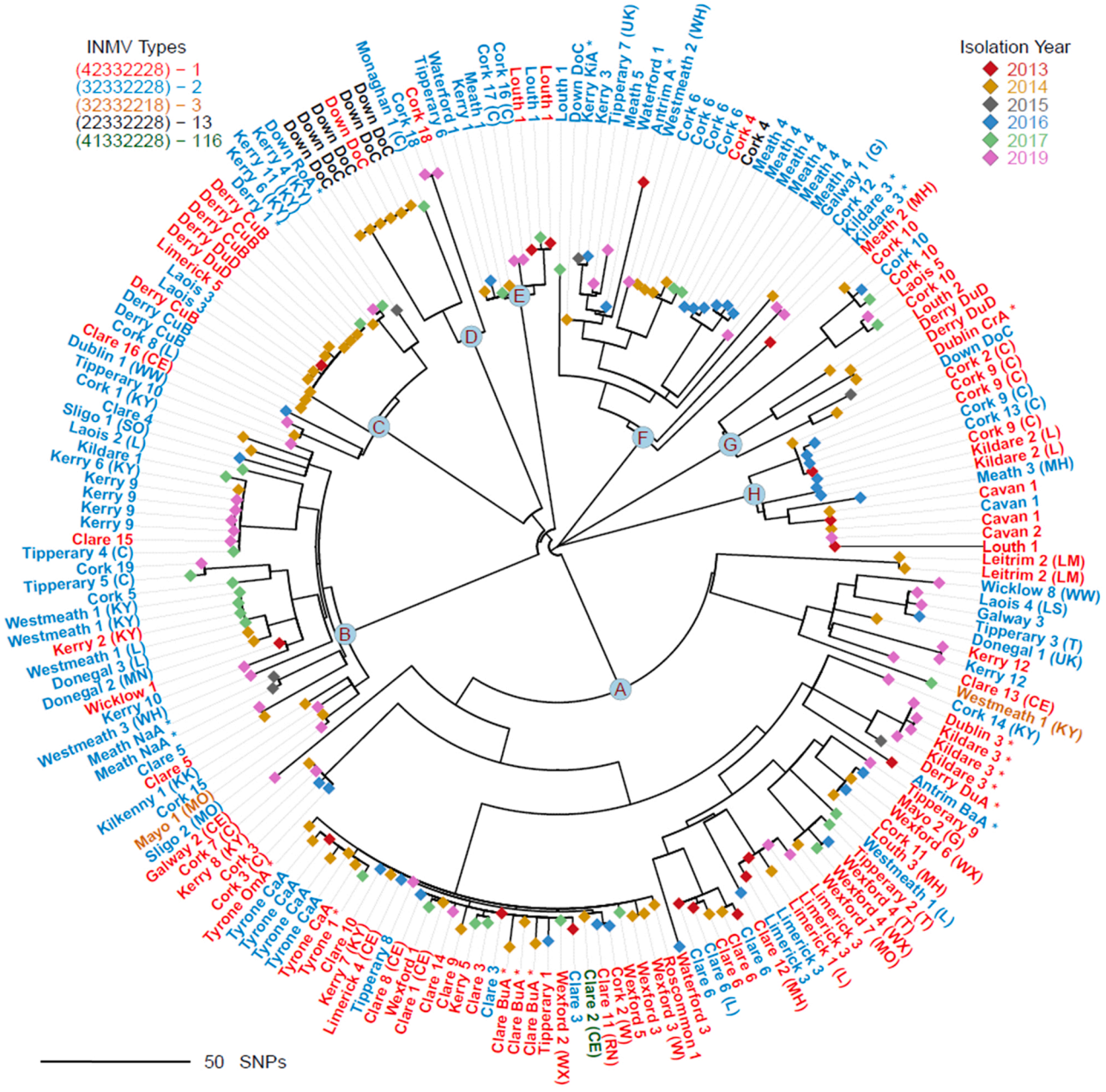

Fig. 1. Phylogenetic tree of MAP isolates from the island of Ireland. Isolates are divided into groups A-H based on genetic proximity. Tip labels are colour coded by INMV type and indicate year, county, and herd of isolation, as well as birth county of animal in brackets (if different from isolation). Tip labels marked with an asterisk (*) indicate unknown/unavailable birth county data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

277 SNPs. Each of the eight distinct groups contained isolates with a mixture of VNTR profiles. INMVs 1 and 2 were present across all groups, with SNP distances ranging between 0 and 277 SNPs for isolates within INMVs 1 (median 213 SNPs) and between 0 and 274 SNPs within INMV 2 (median 204 SNPs). The INMV 13 isolates were present in groups D and F in two herds ('Cork 4' and 'Down DoC'). The SNP distances for INMV 13 isolates ranged from 0 to 222 SNPs (median 1 SNP); four of the INMV 13 'Down DoC' isolates were genetically identical, with the other isolate being 1 SNP away, thus the SNP distance maximum arose from the difference between the 'Down DoC' isolates and the 'Cork 4' isolate. The INMV 116 isolate was within 2 SNPs of several INMV 1 and 2 isolates from different herds. The two isolates with INMV 3 were present in groups A & B, differing from each other by 227 SNPs and were 72 and 3 SNPs, respectively, away from their closest isolates (both INMV 2).

3.2. Spatial overview

The county of isolation data (see Supplementary Table 1) was included in the tip labels of the phylogenetic tree to enable spatial interpretation of the WGS data. As with the VNTR data, none of the eight main groups of Irish MAP isolates showed geographic clustering. Isolates within each group were found in several different counties in Ireland. The group with the most isolates, group A (indicated in Fig. 1), contained 77 isolates recovered from 21 counties across the island of Ireland. The next largest group, B, had 33 isolates representing 13 counties. The remaining six groups all contained isolates from at least two different counties across the island (see Fig. 1).

At the level of the individual county of isolation, the most isolates were recovered from county Cork (n = 32), and these isolates were present in all strain groups. SNP distances for isolates recovered from Cork ranged from 0 to 269 SNPs, with a median of 211 SNPs. Isolates recovered from Clare (n = 25) were spread out across groups A - C. The SNP range for Clare isolates was 0–248 SNPs, and the median was 83 SNPs. Overall, 24 counties were represented by > 1 isolate. Of these 24, only isolates taken from Cavan, Leitrim and Wexford were found exclusively in one group.

For the 31 individual herds within each county represented by > 1 isolate, only herds 'Clare 6', 'Cork 2', 'Cork 10', Derry 'Dungiven Herd D (DuD)', Down 'Downpatrick Herd C' (DoC), 'Kerry 6', 'Kildare 3', 'Louth 1', 'Waterford 1' and 'Westmeath 1' presented more than 10 SNPs intra-herd isolate variation. The remaining herds had little (<10 SNPs) to no genetic difference between their isolates.

3.3. Suspected inter-herd movement

In addition to having the county of isolation for each isolate, the animal birth county was also available. As shown in Fig. 1, there were 31 isolates recovered from animals born in a different county (shown in brackets) from where their infection was disclosed and MAP isolated. In addition to this, 38 isolates were from animals born in a different herd within the same county. The number of moves each animal made after birth is indicated in Supplementary Table 1. For animals with > 1 move, intermediate locations/destinations were not available. This allowed us to explore evidence for suspected inter-herd movement of MAP.

In group B, a total of ten isolates were recovered from cattle with different counties of birth to their disclosure counties. Taking an example of five isolates from within group B separated by large geographic distances (up to ~300 km) - three isolates from herd 'Westmeath 1' were highly similar (within 0–2 SNPs) to isolates from herds 'Donegal 3' and 'Kerry 2'. While the counties of isolation were situated far apart (Kerry to Donegal, ~300 km; Kerry to Westmeath, ~200 km; Westmeath to Donegal, ~150 km), birth county data shows that two of the 'Westmeath 1' animals originated in Kerry, while the other 'Westmeath 1' animal and 'Donegal 3' animal originated in Limerick (Kerry and Limerick border each other).

Group A contained fifteen isolates from animals originating in differing birth counties. One of these animals was from the aforementioned 'Westmeath 1' herd but was born in Limerick (17–7446) – 237 SNPs separated this animal's isolate from the other MAP isolates from 'Westmeath 1' in group B (Figs. 1 and 2). Another 'Westmeath 1' isolate was found in group A, but the animal was born in Kerry (17–7457; Fig. 2). This 'Westmeath 1' isolate was 239 SNPs away from the group B 'Westmeath 1' isolates (Figs. 1 and 2). There were 160 SNPs between the group A 'Westmeath 1' isolates, and they were separated by 39 and 72 SNPs from their closest respective neighbours (Figs. 1 and 2).

All isolates from 'Westmeath 1' (with exception of 17–7457), as well as the other closely related isolates shown in Fig. 1, were recorded to have moved to at least one other unknown location before arriving at their herd of MAP isolation. As the intermediate destinations were unknown, they were not shown in Fig. 2.

Cattle movements were seen in other herds throughout the phylogeny. In group A, 31 closely related isolates (9 SNPs median difference) within the group (as seen in Fig. 1) from 8 counties showed animal movements from Roscommon, Clare and Waterford. Other group A examples included movements from Galway and Meath into Mayo and Louth respectively, which were within 5 SNPs of each other, but also within 5 SNPs of isolates from Cork and Wexford. Another move occurred from Limerick into 'Clare 6' – this isolate was within 10 SNPs of four other 'Clare 6' isolates.

3.4. Comparison to Continental Europe and Great Britain MAP WGS data

MAP WGS data from the island of Ireland were compared to an existing European subset of a published MAP WGS study (Bryant et al., 2016). For this comparison, the latter European WGS data, spanning 12 European countries, were included in the phylogeny for Ireland. As shown in Fig. 3, MAP isolates from Ireland were interspersed amongst the Continental European isolates – Continental European isolates were seen across all previous defined groups A-H. Furthermore, several Continental European isolates fell outside the defined groups, indicating that there are additional strains of MAP present on the continent not seen in our dataset. Isolates from Great Britain (GB) were present in the majority of groups, with the exception of groups B and H, all the GB isolates were located within the defined MAP groupings. Given the biased nature of our sampling, we elected to use a MASCOT approach to conduct an appropriate comparison of our data with that of GB & Continental Europe.

3.5. Preliminary phylogenetics and investigation for the presence of temporal signal

Model test revealed the best fitting nucleotide substitution model was the HKY plus gamma rate heterogeneity model (AIC 22444.67). The maximum likelihood phylogeny imported into TempEst revealed a positive, significant but weak root to tip distance regression (Fig. 4A: $p < 0.001$, R^2 0.05).

For the models run using the relaxed normal clock, the logcombined MASCOT substitution rate of 0.13 substitutions per genome per year was observed to exhibit substantial overlap in its 95% HPD interval (0.07–0.21) with the simple, non-structured, non-randomised tip date, coalescent model substitution rate of 0.21 substitutions per genome per year (95% HPD interval 0.12–0.30). The latter is suggestive that both substitution rates are similar (Fig. 4B). However, both the MASCOT and simple coalescent model substitution rates were considerably larger than those of the ten randomised tip date, simple coalescent models (Fig. 4B), and exhibited no overlap with their 95% HPD intervals.

Additionally, the strict clock, logcombined MASCOT substitution rate, 0.10 substitutions per genome per year (95% HPD interval 0.06–0.14), and the simple, non-structured, non-randomised tip coalescent model substitution rate of 0.10 substitutions per genome per year (95% HPD 0.05–0.14) were very similar and exhibited substantial HPD overlap (Supplementary Fig. 3). As with the lognormal findings, both

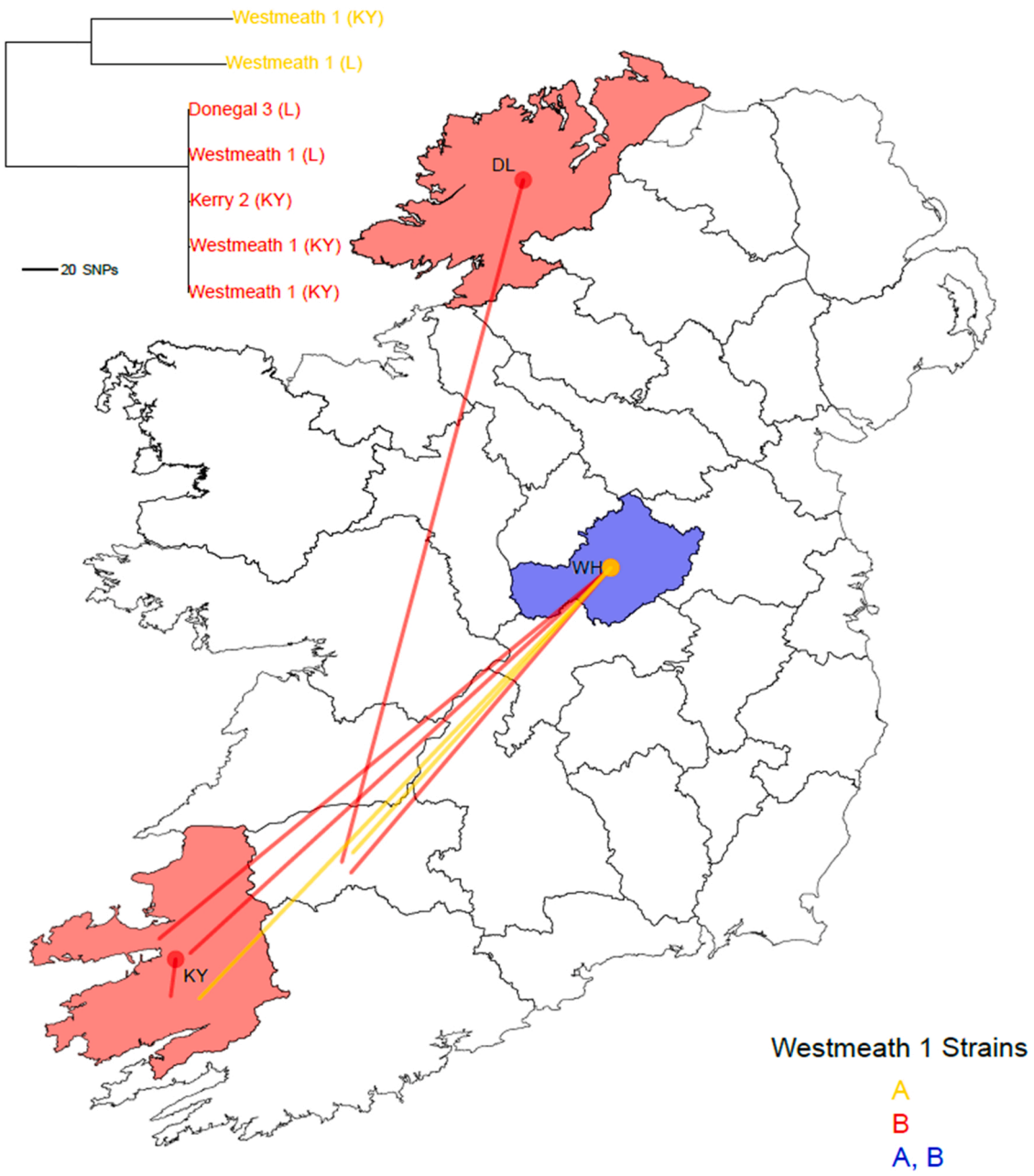

Fig. 2. Movement of MAP strains across Ireland facilitated by cattle movement. The map focuses on the example herd of Westmeath 1 (WH). Strain groups A (gold), and B (red) are indicated; WH contains both groups. Birth county is indicated in brackets e.g. (KY) in phylogeny. Arrows show movement of the animals (from which these strains were isolated) from their respective birth county to the location where their MAP infection was disclosed. It should be noted that exact herd locations within a county are not shown to comply with GDPR regulations around anonymity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rates were larger than those observed for the ten randomised tip date simple coalescent models run with a strict clock (Supplementary Fig. 3), and again exhibited no overlap in 95% HPD intervals.

These data, from two distinct clock models, serve as confirmation of a low, but robust temporal signal in the dataset, and permitted further analyses using structured coalescent outputs as discussed below.

3.6. MASCOT analysis

The three replicate MASCOT models converged on similar values across all parameters, for both the relaxed log normal and strict clock variants. Logcombined median values for substitution rate, effective population size (N_e), time to most recent common ancestor (tMRCA) and inter regional transition rates are presented in Table 2 below. The relaxed clock MASCOT substitution rate is also illustrated in Fig. 4B alongside tip randomisation derived rates as described above. Inter-regional transition rates and regional MAP effective population sizes are illustrated in Fig. 4C and D. The strict clock MASCOT substitution rate is shown in Supplementary Fig. 3, while both inter-regional transition rates and regional MAP effective population sizes can be found in Supplementary Fig. 4A and B.

The relaxed clock MASCOT model estimated a median substitution rate of 0.13 substitutions per genome per year (95% HPD interval 0.07–0.21), and an estimated tMRCA of 1288.37 years before the final sampling date of 2019 (95% HPD interval 392.44–2771.11). Inter-regional median transition rates between Ireland and the rest of

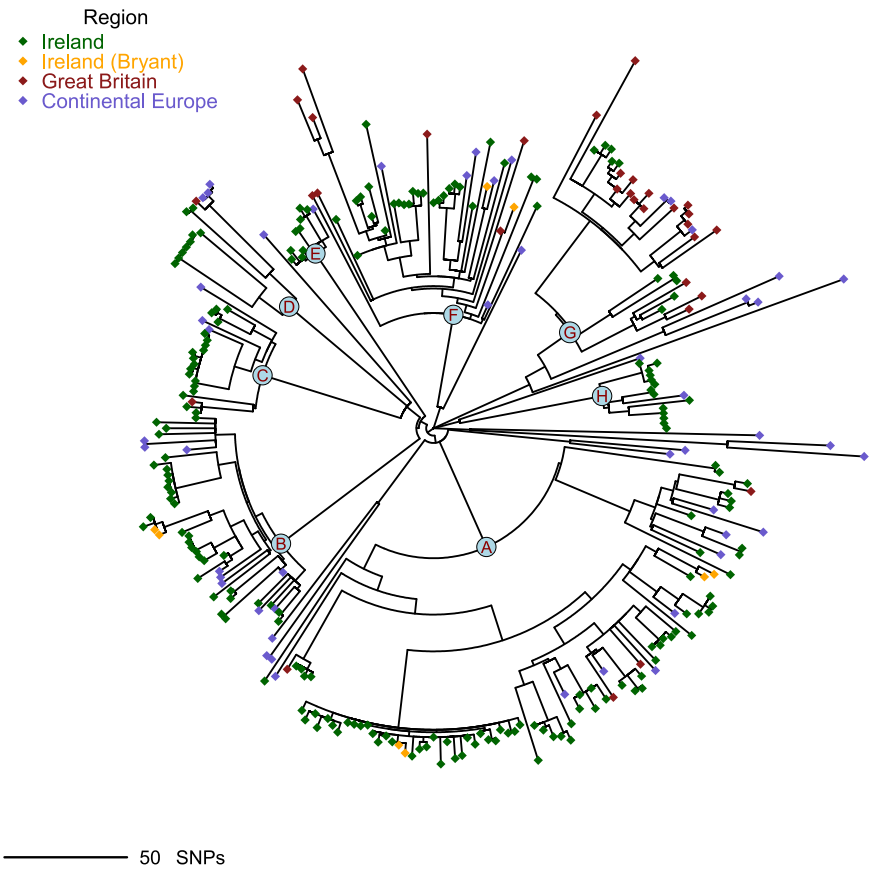

Fig. 3. Phylogenetic tree combining MAP WGS data from Ireland, Great Britain and Continental European. Data from this study was integrated with a wider European and British MAP WGS subset from Bryant et al. (2016) that includes 8 Irish isolates. Tree tips are colour coded based on their origin from Ireland (green, this study; yellow, Bryant et al. (2016) study; Great Britain (red), or Continental Europe (purple), with the latter 2 groups also from Bryant et al. (2016). Groups A-H described in this study are indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Europe were discordant, with Europe to Ireland transitions of 2.2 transitions per lineage per year (95% HPD interval 0.79–4.81), observed to be approximately 27.5 times larger than Ireland to Europe transitions of 0.08 transitions per lineage per year (95% HPD interval 0.01–0.18). MASCOT also estimates most of the pathogen's diversity is found in the non-Irish European countries (N_e 1428.11 95% HPD interval 668.09–2700.85) compared to Ireland (N_e 5.49 95% HPD interval 1.67–11.04).

The strict clock MASCOT model estimated a median substitution rate of 0.10 substitutions per genome per year (95% HPD interval 0.06–0.14) and an estimated tMRCA of 880.60 years before final sampling date of 2019 (95% HPD interval 518.94–1290.77). Inter-regional median transition rates between both territories, were again strongly discordant. The observed Europe to Ireland transition rate of 0.93 transitions per lineage per year (95% HPD interval 0.42–1.70) was 15.5 times larger than the Ireland to Europe transition rate of 0.06 transitions per lineage per year (95% HPD interval 0.03–0.10). As per the lognormal relaxed clock model, the strict clock model also indicated the majority of MAP diversity was found in European territories (N_e 5104.17 95% HPD interval 2115.13– 10,135.86) versus Ireland (N_e 11.77 95% HPD interval 4.63–22.06).

The MCC tree presented in Fig. 5 illustrates that most Irish isolates share an ancestor with isolates from other European territories within the last 50–100 years.

4. Discussion

Here we describe genetic and genomic analysis of 197 MAP geographically dispersed isolates from across the island of Ireland, with phylogenetic analysis revealing eight distinct groups (A-H). As with previous work, VNTR typing of our isolates showed that the majority (95.4%) were of INMV types 1 and 2 (de Kruijf et al., 2017), with rarer INMV types 3, 13 and 116. This is the first reported case of INMV 13 in Ireland; all other INMV 13 isolates to date in Europe have originated from isolates acquired in the UK, Sweden, and Spain (Thibault et al., 2007; Stevenson et al., 2009). WGS groups B, C, E & F were predominantly INMV 2, and groups A & G were predominantly INMV 1, suggesting that VNTR does recapitulate a degree of the underlying genetic relationships. However, the resolution of VNTR is limited, as evidenced here by groups that showed the same VNTRs being separated by large genetic distances (up to 277 SNPs) defined by WGS. These groups also contained isolates of different VNTRs that were genetically close to one another, including multiple cases of INMV 1 and 2 isolates showing < 5 SNPs difference. Of the rarer INMV types, the INMV 3 isolates were 3 and 72 SNPs away from closest isolates, while INMV 116 was within 5 SNPs of both INMV 1 and 2 isolates, and INMV 13 isolates were identical to INMV 1 isolates. Furthermore, while a standard eight loci are examined in MAP MIRU-VNTR, the INMV types identified here only differed at three of these loci. INMVs 1 & 2, which respectively comprised 45.7% and 49.7% of Irish isolates in our study, as well as INMV 13, differ by only one (or two in the case of INMV 1 vs 13) repeats at the first locus. As such, our work echoes previous findings that VNTR both under- and over-estimates true genetic relatedness (Ahlstrom et al., 2015; Bryant et al., 2016).

We attempted to sample isolates from as broad a geographical spread of the island of Ireland as possible to maximise the potential genetic diversity captured. However, given the much greater density of cattle in southern Irish counties, it was difficult to obtain an equal representation of the whole island. Furthermore, of the 122 herds in the study, only 31 were represented by > 1 isolate, with only 10 herds containing isolates that differed by > 10 SNPs from each other. Within these constraints, no clear signal of spatial clustering of the WGS groups in our data was apparent. Groups A-H contained isolates recovered from diverse geographical locations around Ireland, and individual counties

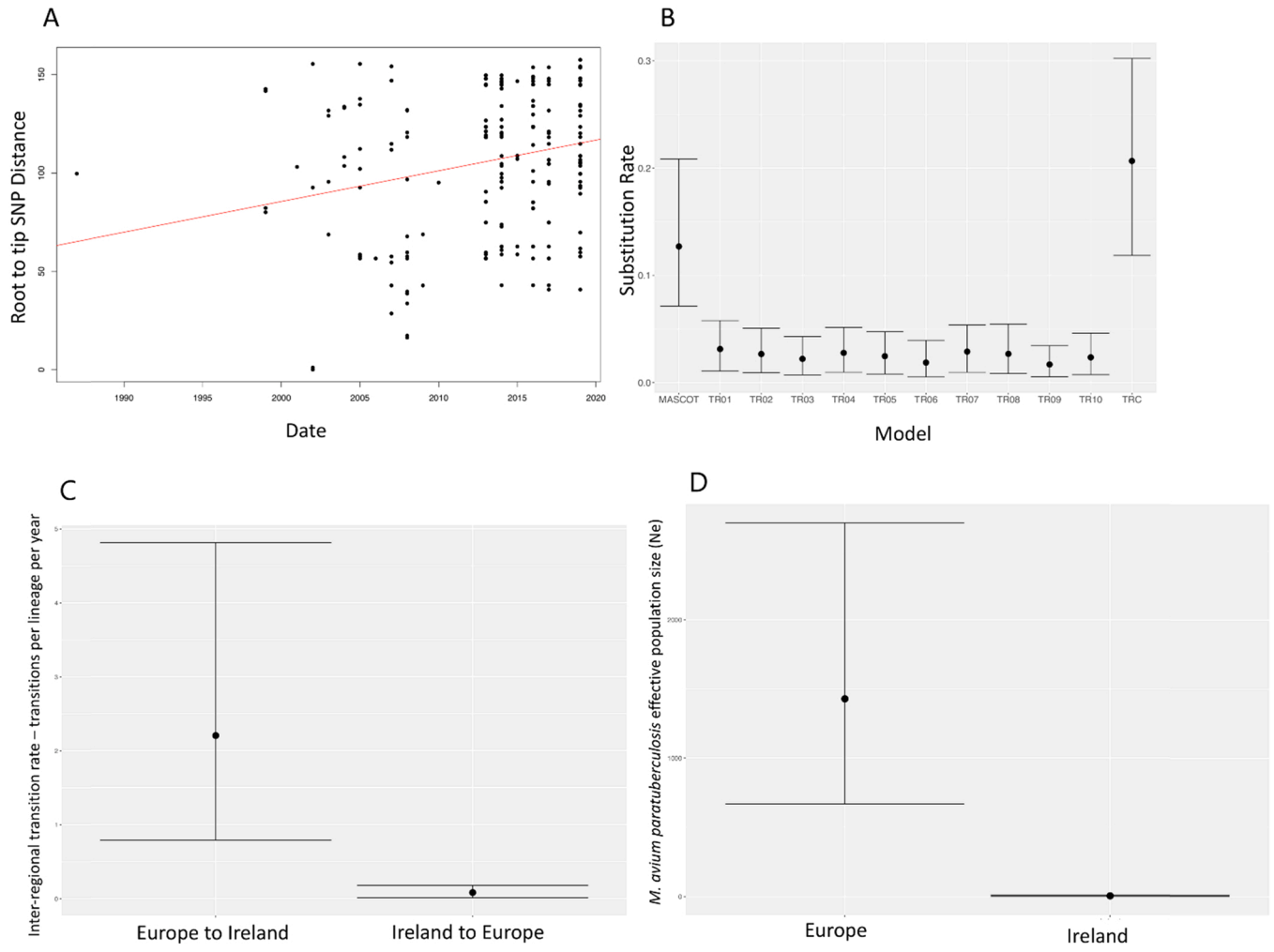

Fig. 4. A: TempEst root to tip distance for n = 253 *MAP* isolates (R^2 0.05, $p < 0.001$). B: Median substitution rates and 95% HPD for various models run in MASCOT using a relaxed clock model. Logcombined MASCOT rate, tip date randomisation runs (TR01-TR10) with simple, non-structured coalescent model in BEAST2; Tip randomisation control (TRC), non-randomised tip dates run with simple non-structured coalescent model in BEAST2. C: Logcombined median inter-regional pathogen transition rates as assessed by the MASCOT structured coalescent model using a relaxed clock model. Transition rates are expressed as transitions per lineage per year and 95% highest posterior density (HPD). D: Logcombined median MASCOT inferred pathogen effective population sizes (N_e) in Ireland and the rest of Europe and 95% highest posterior density (HPD).

Table 2
Summary statistics for log-combined triplicate MASCOT models.

	Substitution rate	tMRCA	Ireland N_e	Europe N_e	Ireland to Europe transition rate	Europe to Ireland transition rate
Relaxed lognormal clock model	0.13 (0.07–0.21)	1288.37 (392.44–2771.11)	5.49 (1.67–11.04)	1428.11 (668.09–2700.85)	0.08 (0.01–0.18)	2.20 (0.79–4.81)
Strict clock model	0.10 (0.06–0.14)	880.60 (518.94–1290.77)	11.77 (4.63–22.06)	5140.17 (2115.13–10,135.86)	0.06 (0.03–0.10)	0.93 (0.42–1.70)

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 253 isolates is expressed in years prior to the last sampling date, 2019.

contained multiple strains within the respective herds studied. This mixing reflects the well documented mobility of the Irish cattle population, with for example 1.3 million cattle movement events in Ireland in 2016 (McGrath et al., 2018).

There were 31 recorded cases in the data where the animal from which MAP was isolated were disclosed in a different county to that of their birth county. There were cases where animals changed locations more than once, but only the final destinations of these animals were available. These animal movements could be used to identify situations where geographically distant herds had cattle infected with the same MAP strains and hence disclose movement of infection from a shared source. For example, in the Westmeath 1 herd (Fig. 2), all MAP isolates recovered from cattle in this herd had been born in southern Irish counties (Kerry and Limerick). Isolates from two distinct WGS groups (A & B) were present in Westmeath 1, with identical Group B strains also present in Donegal and Kerry. The Donegal animal was likewise born in Limerick. The shared origin of all animals infected with related strains suggests that these animals were infected in their birth herd, or at a second unknown location shared with the other infected animals (such as another herd or market). It is likely that the spread of MAP strains

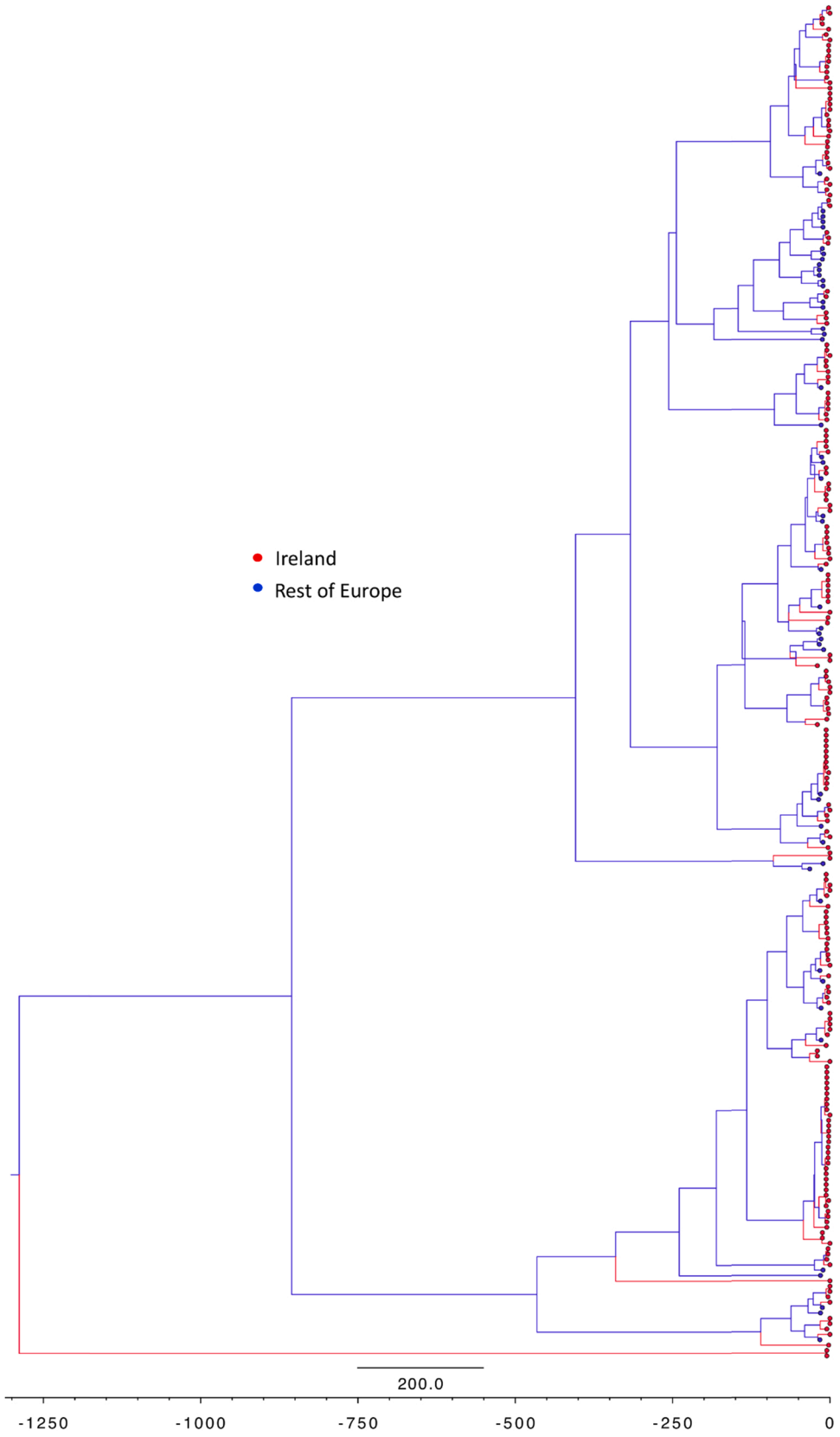

Fig. 5. Time stamped Maximum Clade Credibility (MCC) tree for log combined MASCOT models. X axis scale is no. of years from root date, which was estimated to be median 1288 years before last sampling date of 2019. Coloured lines indicate ancestry: red, Ireland; blue, Rest of Europe. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

across Ireland is facilitated by the insidious nature of MAP disease progression, whereby animals are infected at, or shortly after, birth and only develop JD at a later date, facilitating dissemination of infection. The high degree of cattle movement in Ireland (McGrath et al., 2018) and insidious nature of infection hence provides the opportunity for geographical dispersal of MAP.

While the regression co-efficient for the TempEst analysis was low, this has been noted before in root to tip regressions of this slow growing pathogen and has not been a barrier to subsequent application of coalescent based approaches, provided the robustness of clock like behaviour can be demonstrated (Bryant et al., 2016). We applied Bayesian phylogenetic approaches to better understand the phylogeography of MAP in Ireland relative to other European nations. We found low, but robust evolutionary rates for both relaxed and strict clock models, whose magnitude and 95% HPD interval was consistent with that described before in independent studies (Bryant et al., 2016). This rate, being lower than that estimated for *Mycobacterium bovis*, another common veterinary mycobacterial pathogen (Akhmetova et al., 2021; Crispell et al., 2019, 2017; Salvador et al., 2019), is consistent with the in vitro doubling time of MAP which is slower (22–26 h) than that of the *Mycobacterium tuberculosis* complex (16 h) (Bannantine et al., 2003; Beste et al., 2009).

A consistent finding across both the strict and relaxed clock MASCOT models was that MAP N_e in European territories other than Ireland, was larger than that observed within Ireland. Admittedly, the estimates of Ne in wider Europe are imprecise, with wide 95% HPD intervals, which make it difficult to quantify exactly how much larger the population of MAP outside Ireland is compared to that within Ireland. However, the lack of overlap in N_e 95% HPD for both clock models between wider European and Irish estimates is indicative that there is a meaningful difference in diversity between both territories. Our data are supportive therefore, of the extant Irish population being a subset of the wider diversity seen across Europe. This subset has likely arisen because of movement of continental cattle into Ireland from Europe in the 20th century. Our coalescent data, from both the relaxed and strict clock models suggest this is the case for most Irish samples studied here, which share a recent common ancestor with extant MAP isolates from the rest of Europe within the last 50–100 years. The latter is consistent with the historical record that saw large scale movement of European continental cattle into Ireland after the formation of the European Economic Community after 1957, and further movement with the development of the EU single market in 1992 (Walsh, 2017). An apparent exception to this finding can be observed on the MCC tree (Fig. 5), in which isolates 14–6195 and 14–6194, both from Ireland, exhibit a long branch length back to the base of the phylogeny. However, the latter may be an artefact of the associated lineage both isolates occur in, which is not well sampled in regions outside of Ireland. Had non-Irish European isolates from this lineage been available, we may have seen similar branching towards the tips and recent shared ancestry between territories. Interestingly, both isolates were derived from a Continental European breed, the French Montbeliarde, whose presence in Ireland is relatively recent (Walsh, 2017).

Recorded incidence of JD increased in Ireland in the 1990s. Ninety-two JD cases were reported in Ireland between 1932 and 1992, the majority of which were in imported cattle (Douarre et al., 2010). Between 1995 and 2002, 232 cases were reported across 106 herds (Barrett et al., 2006). While the increase in 1995–2002 could be attributed to better detection and improved veterinary practises, this period also coincides with the opening of the EU single market that saw increased cattle importation from Europe into Ireland, with 85,000 animals arriving between 1992 and 2004 (Barrett et al., 2006; Douarre et al., 2010). Hence while different MAP strains may have been initially brought to Ireland in the last 50–100 years, it is possible the cattle imports of the 1992–2004 period amplified MAP introgression and hence the increase in JD cases. On the other hand, given the volume of live cattle exports (for example, a total of 250,788 in 2020) from Ireland to destinations such as the UK, Spain, Netherlands and Italy (Bord Bia, 2021), it is possible that transmission is now also occurring from Ireland via trade.

The MASCOT derived diversity indices / effective population sizes for Irish and non-Irish European populations and inter-regional transition rates are strongly discordant, with lack of overlap in HPD ranges suggesting the divergences observed are significant. A potential criticism of our structured coalescent findings is that they are based on a data set that exhibits both geographical and temporal biases. Our sampling across Ireland (n = 205) is systematic, taking in at least 27 of Ireland's 32 counties (see Supplementary Fig. 1). Across Britain and the wider western European continent, our sampling (n = 48 from 8 jurisdictions) was not systematic, relying on archived genomic data derived from the Bryant et al. (2016) study. In addition, the Irish samples were collected over the period 1999–2019, but with a particular focus on 2013–2019 (197 of 287 isolates), and hence occupied a temporal range distinct from those from the rest of Europe (2001–2010); see Supplementary Fig. 2. These biases were unavoidable as the scope of this project did not extend to systematically sampling across Europe. We believe however, that the broad inferences we have made from our analyses are robust.

Discrete traits-based methods have been used to infer phylogeographic parameters in the past; however biased sampling, such as that we describe here, can skew outcomes and lead to incorrect inferences being made given that they assume the shape of a phylogeny is not influenced by the migration process (Müller et al., 2017). Conversely, structured coalescent methods can infer ancestral population structure and transition rates between phylogeographic states in the presence of biased sampling (De Maio et al., 2015; Müller et al., 2018, 2017). These methods achieve this by explicitly taking account of the demography of isolates from a given state or territory over time. The underlying diversity within each territory is then used alongside phylogenetic and temporal data to infer transitions from one state to another. Discrepant levels of genetic diversity have been noted before to be indicative of a clade's origin within the more genetically diverse geographical region (Morin et al., 2004). Provided a sufficiently representative sampling across states has been undertaken, the inferred population structure and transition rate estimates are robust.

Our systematic sampling (n = 205) across all of Ireland enhanced the chances of capturing a cohort of isolates which are representative of island-wide pathogen diversity. In contrast, we have only 48 isolates from the rest of Europe. Despite this comparative deficit, and resulting potential to underestimate pathogen diversity, the isolates available still suggest diversity across the rest of Europe is substantially larger than that seen in Ireland. It is likely that a more systematic, representative sampling of continental Europe and Britain would have revealed greater pathogen diversity. The latter may have helped to refine our estimates of N_e within, and transition rates between, territories. However, given that structured coalescent methods control for sampling biases, it is unlikely such additional sampling would have greatly changed the broad findings of our analyses. Ireland's MAP population likely represents a subset of the pathogen genetic diversity seen in neighbouring European countries. This bottlenecked population was likely founded by largely unidirectional movement of infected cattle into Ireland, from multiple European source locations, in the recent past.

5. Conclusion

In summary, WGS reveals that MAP isolates from the island of Ireland are more diverse than previous MIRU-VNTR data would suggest. Distinct strains of MAP appear across different counties of Ireland; integration of WGS with county of birth for each animal revealed evidence for dispersal of MAP strains across Ireland. Comparison to European MAP WGS data suggested transmission of the pathogen from Europe to Ireland. Coalescent analyses indicate that MAP was likely introduced from Europe into Ireland via cattle trade over the last 50–100 years. Our study highlights the potential of WGS to support studies on

the epidemiology and transmission dynamics of MAP infection at local, national, and international scales.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge funding from the Department of Agriculture, Food and Marine awards 15/S/651 ('NexusMAP') and 2019R404 ('BTBGenIE'). We thank Alison Murphy and the UCD Conway Genomics Core for their support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.vetmic.2022.109416.

References

Ahlstrom, C., Barkema, H.W., Stevenson, K., Zadoks, R.N., Biek, R., Kao, R., Trewby, H., Haupstein, D., Kelton, D.F., Fecteau, G., Labrecque, O., Keefe, G.P., McKenna, S.L., Tahlan, K., De Buck, J., 2016. Genome-wide diversity and phylogeography of Mycobacterium avium subsp. paratuberculosis in Canadian dairy cattle. PLoS One 11, e0149017.

Ahlstrom, C., Barkema, H.W., Stevenson, K., Zadoks, R.N., Biek, R., Kao, R., Trewby, H., Haupstein, D., Kelton, D.F., Fecteau, G., Labrecque, O., Keefe, G.P., McKenna, S.L.B., 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.

Akhmetova, A., Guerrero, J., McAdam, P., Salvador, L.C.M., Crispell, J., Lavery, J., Presho, E., Kao, R.R., Biek, R., Menzies, F., Trimble, N., Harwood, R., Pepler, P.T., Oravcova, K., Graham, J., Skuce, R., du Plessis, L., Thompson, S., Wright, L., Byrne, A., Allen, A.R., 2021. Genomic epidemiology of *Mycobacterium bovis* infection in sympatric badger and cattle populations in Northern Ireland. bioRxiv, 2021.2003.2012.435101.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

Bannantine, J.P., Zhang, Q., Li, L.L., Kapur, V., 2003. Genomic homogeneity between Mycobacterium avium subsp. avium and Mycobacterium avium subsp. paratuberculosis belies their divergent growth rates. BMC Microbiol. 3, 10.

Barrett, D.J., Good, M., Hayes, M., More, S.J., 2006. The economic impact of Johne's disease in an Irish dairy herd: a case study. Ir. Vet. J. 59, 282–+.

Beste, D.J., Espasa, M., Bonde, B., Kierzek, A.M., Stewart, G.R., McFadden, J., 2009. The genetic requirements for fast and slow growth in mycobacteria. PLoS One 4, e5349.

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.

Bord Bia, 2021. Irish Live Cattle Exports. ⟨https://www.bordbia.ie/farmers-growers/prices-markets/cattle-trade-prices/live-cattle-exports/⟩.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Bryant, J.M., Thibault, V.C., Smith, D.G., McLuckie, J., Heron, I., Sevilla, I.A., Biet, F., Harris, S.R., Maskell, D.J., Bentley, S.D., Parkhill, J., Stevenson, K., 2016. Phylogenomic exploration of the relationships between strains of Mycobacterium avium subspecies paratuberculosis. BMC Genom. 17, 79.

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

Crispell, J., Zadoks, R.N., Harris, S.R., Paterson, B., Collins, D.M., de-Lisle, G.W., Livingstone, P., Neill, M.A., Biek, R., Lycett, S.J., Kao, R.R., Price-Carter, M., 2017. Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. BMC Genom. 18, 180.

de Juan, L., Mateos, A., Dominguez, L., Sharp, J.M., Stevenson, K., 2005. Genetic diversity of Mycobacterium avium subspecies paratuberculosis isolates from goats detected by pulsed-field gel electrophoresis. Vet. Microbiol. 106, 249–257.

de Kruijf, M., Lesniak, O.N., Yearsley, D., Ramovic, E., Coffey, A., O'Mahony, J., 2017. Low genetic diversity of bovine Mycobacterium avium subspecies paratuberculosis isolates detected by MIRU-VNTR genotyping. Vet. Microbiol. 203, 280–285.

De Maio, N., Wu, C.H., O'Reilly, K.M., Wilson, D., 2015. New routes to phylogeography: a Bayesian structured coalescent approximation. PLoS Genet. 11, e1005421.

Douarre, P.E., Cashman, W., Buckley, J., Coffey, A., O'Mahony, J., 2011. Molecular characterization of Mycobacterium avium subsp. paratuberculosis using multi-locus short sequence repeat (MLSSR) and mycobacterial interspersed repetitive units-variable number tandem repeat (MIRU-VNTR) typing methods. Vet. Microbiol. 149, 482–487.

Douarre, P.E., Cashman, W., Buckley, J., Coffey, A., O'Mahony, J.M., 2010. Isolation and detection of Mycobacterium avium subsp. paratuberculosis (MAP) from cattle in Ireland using both traditional culture and molecular based methods. Gut Pathog. 2, 11.

Duchêne, S., Holt, K.E., Weill, F.X., Le Hello, S., Hawkey, J., Edwards, D.J., Fourment, M., Holmes, E.C., 2016. Genome-scale rates of evolutionary change in bacteria. Microb. Genom. 2, e000094.

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.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.

Li, L., Bannantine, J.P., Zhang, Q., Amonsin, A., May, B.J., Alt, D., Banerji, N., Kanjilal, S., Kapur, V., 2005. The complete genome sequence of Mycobacterium avium subspecies paratuberculosis. Proc. Natl. Acad. Sci. USA 102, 12344–12349.

Martin, M., 2011. Cutadapt Removes Adapter Sequences from High-throughput Sequencing Reads, 17, p. 3.

McGrath, G., Tratalos, J.A., More, S.J., 2018. A visual representation of cattle movement in Ireland during 2016. Ir. Vet. J. 71, 18.

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

Morin, P., Luikart, G., Wayne, R., Morin, P.A., Luikart, G., Wayne, R.K., 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216.

Müller, N.F., Rasmussen, D., Stadler, T., 2018. MASCOT: parameter and state inference under the marginal structured coalescent approximation. Bioinformatics 3843–3848.

Müller, N.F., Rasmussen, D.A., Stadler, T., 2017. The structured coalescent and its approximations. Mol. Biol. Evol. 34, 2970–2981.

Paradis, E., Schliep, K., 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.

R Core Team, 2018. R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in bayesian phylogenetics using tracer 1.7. Syst. Biol. 67, 901–904.

Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2 vew007.

Rasmussen, P., Barkema, H.W., Mason, S., Beaulieu, E., Hall, D.C., 2021. Economic losses due to Johne's disease (paratuberculosis) in dairy cattle. J. Dairy Sci.

Rieux, A., Khatchikian, C.E., 2017. tipdatingbeast: an r package to assist the implementation of phylogenetic tip-dating tests using beast. Mol. Ecol. Resour. 17, 608–613.

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

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stevenson, K., Alvarez, J., Bakker, D., Biet, F., de Juan, L., Denham, S., Dimareli, Z., Dohmann, K., Gerlach, G.F., Heron, I., Kopecna, M., May, L., Pavlik, I., Sharp, J.M., Thibault, V.C., Willemsen, P., Zadoks, R.N., Greig, A., 2009. Occurrence of Mycobacterium avium subspecies paratuberculosis across host species and European countries with evidence for transmission between wildlife and domestic ruminants. BMC Microbiol. 9, 212.

Thibault, V.C., Grayon, M., Boschiroli, M.L., Hubbans, C., Overduin, P., Stevenson, K., Gutierrez, M.C., Supply, P., Biet, F., 2007. New variable-number tandem-repeat markers for typing Mycobacterium avium subsp. paratuberculosis and M. avium strains: comparison with IS900 and IS1245 restriction fragment length polymorphism typing. J. Clin. Microbiol. 45, 2404–2410.

Walsh, G., 2017. Cattle Breeds in Ireland: A History. Borie Press, Wexford, Ireland. ISBN 978-1-9998711-0-9.

Whittington, R., Donat, K., Weber, M.F., Kelton, D., Nielsen, S.S., Eisenberg, S., Arrigoni, N., Juste, R., Saez, J.L., Dhand, N., Santi, A., Michel, A., Barkema, H., Kralik, P., Kostoulas, P., Citer, L., Griffin, F., Barwell, R., Moreira, M.A.S., Slana, I., Koehler, H., Singh, S.V., Yoo, H.S., Chavez-Gris, G., Goodridge, A., Ocepek, M., Garrido, J., Stevenson, K., Collins, M., Alonso, B., Cirone, K., Paolicchi, F., Gavey, L., Rahman, M.T., de Marchin, E., Van Praet, W., Bauman, C., Fecteau, G., McKenna, S., Salgado, M., Fernandez-Silva, J., Dziedzinska, R., Echeverria, G., Seppanen, J., Thibault, V., Fridriksdottir, V., Derakhshandeh, A., Haghkhah, M., Ruocco, F., Kawaji, S., Momotani, E., Heuer, C., Norton, S., Cadmus, S., Agdestein, A., Kampen, A., Szteyn, J., Frossling, J., Schwan, E., Caldow, G., Strain, S., Carter, M., Wells, S., Munyeme, M., Wolf, R., Gurung, R., Verdugo, C., Fourichon, C., Yamamoto, T., Thapaliya, S., Di Labio, E., Ekgatat, M., Gil, A., Alesandre, A.N., Piaggio, J., Suanes, A., de Waard, J.H., 2019. Control of paratuberculosis: who, why and how. A review of 48 countries. BMC Vet. Res. 198.