Rev. Sci. Tech. Off. Int. Epiz., 2016, **35** (1), 241-258

Use of genomics to track bovine tuberculosis transmission

R.R. Kao [(1)*], M. Price-Carter [(2)] & S. Robbe-Austerman [(3)]

(1) Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, G61 1QH, Glasgow, United Kingdom
(2) AgResearch, Tennent Drive, Private Bag 11008, Palmerston North 4442, New Zealand
(3) United States Department of Agriculture, Animal and Plant Health Inspection Service, National Veterinary Services Laboratories, Ames, IA, 50010, USA
*Corresponding author: rowland.kao@glasgow.ac.uk

Summary

The control of any infectious disease of livestock is made more difficult by the presence of a wildlife reservoir, as the reservoir is often poorly observed and difficult to manage. This problem is particularly acute for bovine tuberculosis (bTB) because the long duration of infection and low levels of infectiousness make tracing the sources of infection difficult. For over 30 years, the process of contact tracing has been aided by the exploitation of molecular markers in the pathogen, but this has largely only been capable of characterising broad associations between large communities of similar types. However, the recent advent of mass high-throughput 'whole-genome' sequencing (WGS) has revolutionised forensic epidemiology for other diseases, and now it has the potential to do so for bTB. In this review, the authors consider the historical context of WGS use and look at what prior molecular techniques have already achieved. They outline the key approaches to interpreting WGS data and consider both the role of advanced analytical techniques that exploit the evolutionary and epidemiological properties of the system and the problems associated with quantifying the role of hidden reservoirs of disease. Finally, they consider the particular difficulties associated with developing this technology for routine diagnostics and its potential for mass use.

Keywords

Bovine tuberculosis – Contact tracing – Genomics – High-throughput sequencing – Multi-host pathogen – Phylodynamics.

Introduction

Bovine tuberculosis (bTB) is a chronic, zoonotic infectious disease, affecting livestock and a variety of wildlife, which is transmitted via aerosol and various bodily excretions. The causative agent, *Mycobacterium bovis*, is a slow-growing bacterium that is a member of the *M. tuberculosis* complex. While in many countries bTB has been eradicated from livestock, in others eradication efforts are hampered by wildlife reservoirs.

A critical component of bTB control is successful tracing of the sources of infection. Successfully tracing 'who infected whom' can reveal likely points of control and helps predict future directions of spread. In combination with mathematical models, contact tracing data help to predict the outcomes of alternative control methods. Unfortunately, tracing is made difficult by bTB's long incubation period and low transmission probability per contact. Contact tracing has been greatly aided by the development of pathogen molecular markers, which can be used to group infected individuals by the genetic relatedness of pathogen

doi: http://dx.doi.org/10.20506/rst.35.1.2430

samples. Historically, these markers only permitted coarse-grained associations, but high-throughput sequencing technology ('whole-genome sequencing' or WGS) now allows for unprecedented discrimination between possible infection pathways. While the use of WGS for bTB is in its infancy, examples from human TB and other diseases promise considerable opportunities for its use as a forensic tool. In this paper, the authors review the application and interpretation of WGS to bTB, emphasising the difficulties associated with identifying the role of a wildlife reservoir. They also highlight the manner in which a confluence of this technology with sophisticated mathematical and statistical approaches has the potential to produce a paradigm shift in our understanding of bTB transmission and control.

Historical use of molecular data for contact tracing

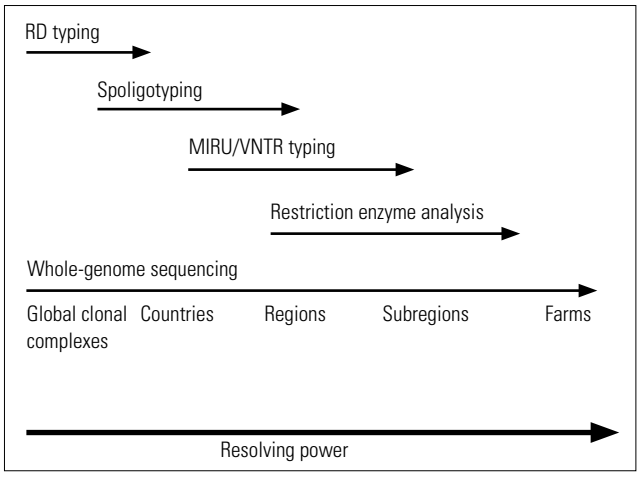
Bovine tuberculosis outbreak investigations have exploited molecular data for over 30 years, with even the earliest molecular methods aiding the identification of the source of herd outbreaks. The ideal molecular epidemiological tool would be inexpensive and easy to use, and provide rapidly generated, reproducible results (within and between laboratories) that are easy to interpret and to document. There are a number of excellent reviews that describe the variety of different molecular typing methods that have been characterised for their ability to distinguish *M. bovis* isolates (1, 2, 3), their roles in characterising the global phylogeography of *M. bovis*, and both animal-to-animal and zoonotic transmission (2, 4, 5, 6, 7, 8, 9, 10, 11). Several techniques that have been used to provide routine information about the source and spread of *M. bovis* infections are compared in Figure 1 for their distinguishing ability and are briefly described below.

Restriction enzyme analysis

The first method for typing *M. bovis* isolates, restriction enzyme analysis (REA), was developed in the early 1980s (12, 13). Genomic DNA is treated separately with three different frequent-cutting enzymes (BstEII, PvuII and BclI) and the resulting restriction fragments electrophoresed so that the restriction patterns can be compared with digests from other isolates. Genomic alterations such as insertions, chromosomal rearrangements, deletions and point mutations lead to observed banding pattern changes. REA is technically challenging, requiring microgram quantities of very pure DNA and, consequently, an additional large-volume culture step. For reproducibility, the DNA must be electrophoresed on very long agarose gels requiring strict temperature control. In addition, despite considerable effort, REA patterns are not easily digitised. REA was the first technology that made use of differences scattered throughout the entire genome for typing *M. bovis* and is reasonably high resolution (Fig. 1). However, these difficulties have restricted the use of this technology to the laboratory in New Zealand (NZ) where it was developed. When last employed for routine typing in 2011, there were over 300 different REA *M. bovis* 'types' in the database of 3,800 NZ isolates of *M. bovis*.

Spoligotyping

Spacer oligotyping or spoligotyping (14) exploits differences detected in the direct repeat (DR) region of the mycobacterial chromosome (15). The DR locus is part of the bacterial immune system which confers resistance to foreign DNA (plasmids and viruses) (16). It contains a series of DRs of 36 base pairs (bp), separated by unique, non-repetitive spacer sequences of 35–41 bp. Variations (typically deletions) are exploited for the typing of *M. bovis* isolates (14). The traditional spoligotyping method involves amplification of the entire DR region of an isolate by polymerase chain reaction (PCR) followed by the detection of specific spacers from *M. tuberculosis* H37Rv and *M. bovis* bacillus Calmette-Guérin (BCG) genomes. Spoligotyping results resemble bar codes that indicate which of 43 standard spacers are present. It is inexpensive, has a high throughput capacity and typing is possible from tissue homogenates (17, 18, 19), thus eliminating the need for additional culture. There is an internationally recognised system for naming spoligotypes (20, 21), with multiple databases that can be used comparatively at the global level (21, 22, 23). However, this method is less resolving than either variable-number tandem repeat (VNTR) typing (see below) or REA, limiting its usefulness in forensic investigations.



RD: regions of difference
VNTR: mycobacterial interspersed repetitive-unit variable-number tandem repeat

Fig. 1
Schematic illustration of the relative resolving power of different typing schemes

Variable-number tandem repeat typing

Segments of the chromosome that contain moderate-sized (50–100 bp) tandem repeats occasionally undergo recombination, which leads to differences in the number of these repeats. The resulting changes in repeat number make them useful for fingerprinting bacterial isolates (24). The entire repeat region is typically PCR-amplified and electrophoresed to determine the region size and thus the number of repeats. Because of the nature of these variations there is a high chance for homoplasy (the same number of repeats in unrelated isolates) at any given VNTR locus; however, comparing isolates at multiple VNTR loci improves both resolution and reliability. Different VNTR loci have a wide array of molecular clocks. Six exact tandem repeat loci were initially characterised (25), and these have been used extensively, especially in the United Kingdom, but 40 loci were later identified when the first *M. tuberculosis* genome was sequenced (26). Discrimination was poor when many of these loci were compared in systemic studies of diverse sets of isolates of *M. bovis* clonal complexes (27, 28, 29), and it varies for different clonal complexes and for descendants of the same complexes in different countries (3). Most routine typing now uses sets of loci that are selected to provide optimal discrimination within regions (28, 29, 30, 31, 32, 33); unfortunately, this prevents systematic global comparisons (34). There is also evidence with *M. tuberculosis* that suggests that VNTR typing may occasionally suffer from laboratory-specific differences (35, 36).

Combining typing methods

Since different typing methods target different regions of the chromosome, combining them can have enhanced epidemiological benefits. This has often been found to be the case for spoligotyping and VNTR (28, 29, 31, 32, 37, 38, 40).

Mycobacterium bovis clonal complexes

As there is no substantial evidence for horizontal gene transfer in the *M. tuberculosis* complex, the entire species of *M. bovis* is considered to be a single clonal complex, descended from a common ancestor and containing clades that share common ancestors. Within them, large deletions, called regions of difference (RDs), are scattered throughout their genomes and are considered unidirectional phylogenetic markers. The same lineages are evident when *M. bovis* is characterised by spoligotyping, but spoligotyping can often distinguish other *M. bovis* types, in addition to those seen by RD typing (7). Comparisons of the RDs and spoligotypes of isolates from around the world have provided great insight into the regional distribution of *M. bovis* complexes and its causal factors (21, 41) (see Fig. 2). For example, a single clonal complex (European 1) predominates in the United Kingdom and those countries to which it directly or indirectly exported cattle (42). Further discrimination has also been possible with the addition of VNTR methods, showing for example that Great Britain (37), Northern Ireland (32), the Republic

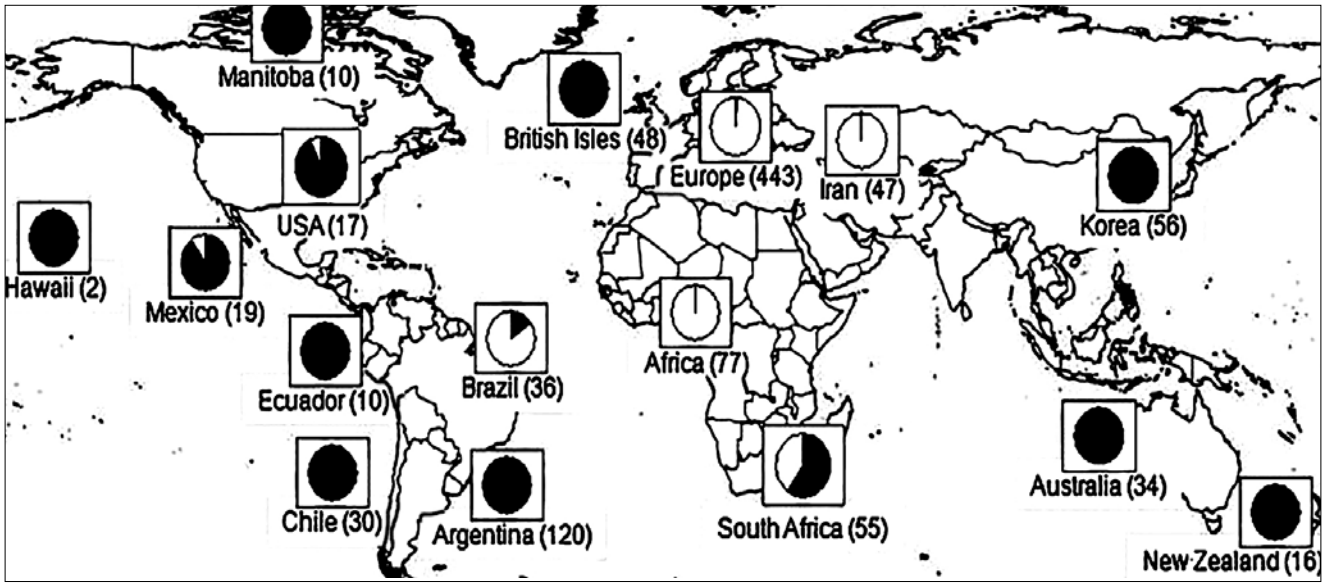


Fig. 2
The global distribution of the European 1 clonal complex of *Mycobacterium bovis* described by Smith and colleagues (42), showing how RD deletion analysis is useful for understanding how *M. bovis* became distributed around the world
The pie charts (in black) show the proportion of isolates that are members of the Eu1 clonal complex. The number of strain deletions typed for RDEu1 in each region are shown in parentheses. The current distribution of the Eu1 clonal complex is consistent with the dispersion of British cattle, such as Herefords, from the United Kingdom to former trading partners

of Ireland (38), Italy (28), Spain (40, 43), the USA (California) (44), Mexico (45) and certain countries in Africa and South America (46, 47, 48, 49, 50, 51) harbour predominant spoligo/VNTR types (see Fig. 3 for an example from Great Britain), aiding epidemiological investigations.

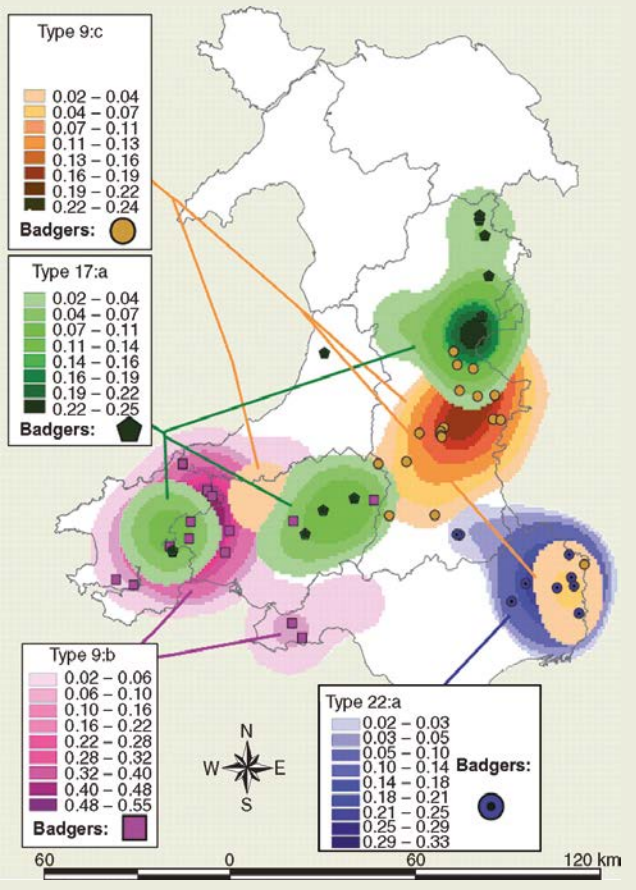


Fig. 3
Resolution with spoligo/ VNTR typing: the geographical distribution of genotypes of *Mycobacterium bovis* isolated from cattle and badgers in Wales
Genotypes isolated from cattle that were examined post mortem and from badgers that were found dead in Wales between 26 October 2005 and 31 May 2006, originally illustrated in Goodchild *et al.* (52). The relative incidence per square kilometre of *M. bovis* genotypes in cattle herds has been kernel-smoothed using a 25-km bandwidth: the darker the colour, the higher the incidence of this type in the indicated region. Locations and genotypes of individual badgers found to have *M. bovis*, shown as coloured symbols, illustrate how the same types are often found in the badger and cattle isolates from the same region. The number in these genotypes refers to the spoligotype and the letter to the subgroup of the spoligotype, e.g. 9:b denotes spoligotype SB0140, VNTR 7-5-5-5*-3-2.1, and 9:c denotes spoligotype SB0140, VNTR 7-5-2-4*-3-3.1

Molecular typing of wildlife reservoirs

In several regions of the world, wildlife reservoirs have greatly hindered efforts to control *M. bovis* (6). This has been the case with the Eurasian badger in Britain and Ireland, wild boar in France, possums in NZ, and deer in Michigan in the USA. Molecular methods have been used with varying degrees of success to establish the role of these reservoirs in *M. bovis* disease dynamics and to guide control measures. An early outcome was the identification of dominant REA 'types' across wildlife and livestock in the same NZ region, often allowing for differentiation of outbreak sources (e.g. wildlife or reinfection from herd movement) (1, 53, 54). REA typing has helped to characterise the role and source of wildlife reservoirs in different regions (55, 56, 57, 58) and the disease dynamics in possum populations (59). It has also aided in tracing the source of infections that result from within-herd transmission and movement of livestock. The role of routine REA typing in NZ bTB control has recently been reviewed (60, 61).

In states of the USA with recent or ongoing *M. bovis* infections, such as Michigan and Minnesota, wildlife, and in particular white-tailed deer, are known to carry *M. bovis* (62, 63). A recent genotyping (spoligotyping and VNTR) survey of isolates from this and other regions in the USA (31) clearly illustrated the geospatial localisation of types, and showed that wild deer and other wildlife harboured the same types as cattle whereas farmed deer harboured an epidemiologically distinct group of *M. bovis* types.

In Britain and Ireland, the Eurasian badger is the main wildlife reservoir of *M. bovis*. In these countries, comparisons of genetic types have provided an indication of the relationship between the prevalence of bTB in cattle and its prevalence in other species, including badgers and other domestic animals (52, 64, 65, 66). Badgers and other wildlife (67, 68) in the same geographical locations have been shown to share the same spoligotype and VNTR type (69). Despite this, the contribution of badgers to cattle bTB has been difficult to quantify (70).

Whole-genome sequencing

The first whole-genome sequence of an *M. bovis* isolate was released in 2003 (71). Genomic DNA was fragmented into uniform-sized pieces that were sufficiently small to be individually sequenced, an approach which remains the basis of current WGS technologies (39). The WGS data files that result from these technologies are large, usually around 250 megabases, and their processing requires specialist bioinformatics expertise. While *de novo* assembly is possible, typically these data are compared or aligned to the sequence of the well-characterised and annotated reference genome AF2122/97 (71), and single nucleotide variations or polymorphisms (SNPs) are compared to those found when other isolates are aligned to the reference. The Illumina technology now widely employed for *M. bovis*

sequencing involves attachment of the fragmented genome to adaptors that facilitate their attachment to the sequencing instrument, followed by amplification to enhance the signal and then sequencing (39). This makes it possible to repeatedly sequence each region of the chromosome (often hundreds of times), resulting in an accurate representation of most of the genome (>4.0×10^6 bases), a vast increase from the 43 spacers in the DR region considered in spoligotyping and the 6–24 VNTR regions considered in VNTR analysis. Although it is not yet possible to sequence directly from tissue homogenates, there are promising signs from work with *M. tuberculosis* that genomic data are obtainable from small volumes of early liquid cultures (72).

Most mutations are detrimental to the bacteria and thus subject to a strong purifying selection. A small minority are beneficial (e.g. facilitate resistance to toxins or drugs) and subject to positive selection. These are a potentially problematic source of homoplasy since, if they occur, they are more likely to take hold in distantly related bacteria. However, this still leaves a substantial proportion of 'nearly neutral' mutations. Although such nearly neutral variation may be selected out given sufficient time (73, 74), over short timescales, these mutations are practical markers of pathogen genealogy. Technical issues, including errors in sequencing and bioinformatics, can result in missed or artefactually added mutations, but the very high genome coverage and the relatively large number of detected differences mean that these types of experimental errors are less influential than similar types of errors acquired with previous typing methods.

How much more does whole-genome sequencing provide?

The most sensitive PCR-based assays compare isolates at 43 spoligo spacer regions and up to 24 VNTR loci. In contrast, WGS compares changes that occur across the entire genome. Well-defined lineages result (Fig. 4), in which isolates that differ at one or two VNTR loci, differ by tens of SNPs and often share hundreds of SNPs with more distant common ancestors. This extremely high resolving

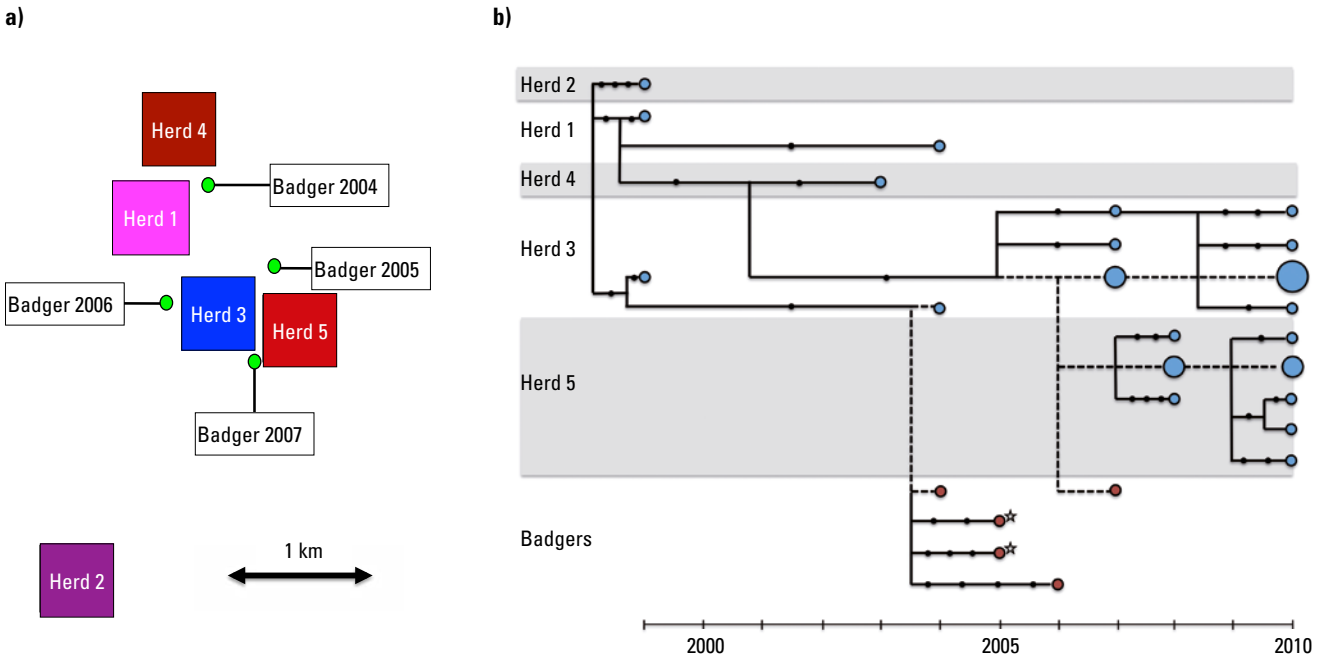


Fig. 4
Resolution with whole-genome sequencing
The work of Biek *et al.* (75), correlating WGS data with the geographical distribution of isolates with the same rare VNTR type from cattle and badgers in Northern Ireland, clearly illustrates the much more detailed relationships that can be defined with WGS data
a) The spatial relationship of the source of cattle and badger isolates that were sampled for this study, and the year in which the samples were acquired
b) Maximum likelihood network of *M. bovis* genomes. Tips in the phylogenetic tree are arranged according to sampling date. Black circles represent single nucleotide polymorphisms (SNPs) separating sequences, dashed lines indicates branches without mutational events. The size of the circle is proportional to the number of isolates sharing the same sequence. In several instances, isolates from badgers are indistinguishable from those from livestock on nearby farms, and numerous SNPs clearly distinguish these isolates from other lineages. Stars indicate the two distinct genotypes that were isolated from different tissues in the same badger in 2005

power has had a transformative effect on investigations into human TB outbreaks. There are several reviews describing the evolution and global spread, transmission and evolution of drug resistance in *M. tuberculosis* (76, 77, 78) that have benefited from this technology. In a seminal study, a social network analysis based on epidemiological tracing was compared with WGS data from 32 isolates of the same VNTR type and identified two separate, concurrent outbreaks of *M. tuberculosis* in crack cocaine houses in British Columbia, Canada (79). The first published use of WGS for *M. bovis* examined genomic differences in a single VNTR type of *M. bovis* that was isolated from cattle and badgers in a small area in Northern Ireland (75). This study showed that spatio-temporal relationships amongst cattle and badger isolates were better predictors of genetic distances than links via cattle movements (Fig. 4). This signature, while still insufficient to identify direction, suggests that larger-scale studies would be able to characterise the spatial-genetic relationship, with interpretation of the multi-host system most likely requiring mechanistic models to maximise the insight gained from the genetic data (75). Similar results were found when WGS was applied to the outbreak in cattle and white-tailed deer in Minnesota (80).

These genomic differences have already provided important additional insights into the origin and spread of *M. bovis* (75, 81). However, the slow and highly variable molecular clock over short time periods (Table I) observed in both *M. tuberculosis* (83) and *M. bovis* (81) limits the inferences that can be made about transmission, with examples of identical types being found from the same farm over a five-year interval, but epidemiologically linked isolates, containing as many as five distinguishing SNPs, being found after only a one-year interval (see Fig. 4). Similar variability has been observed in the SNP lineage of closely linked isolates in the USA (S. Robbe-Austermann, unpublished results) and NZ (M. Price-Carter, unpublished results). Mutation rates may also be reduced during periods of latency compared to active infection (86), causing additional variability that may be greater in some host species than others. This variability creates uncertainty in transmission pathway reconstruction (87), although systematic differences could also be exploited to garner additional insights (Table I).

Analytical approaches to evaluating whole-genome sequencing data

Qualitative comparisons alone have considerable value (88) and are easily augmented by standard statistical analyses that use 'risk factors' such as social network distance to explain the observed genetic distance between bacteria taken from pairs of hosts (75, 82). However, for detailed contact tracing, the difficulties of long incubation period and low infection rates also introduce ambiguities and biases in interpreting WGS data, particularly where there is clustering of potentially infectious contacts (Fig. 5). Different problems exist where many intermediate cases may be absent from the data, although these may be alleviated by the development of temporal markers of infection or reliable indicators of change in the microbial community (89). These uncertainties provide impetus for the use and development of formal analytical approaches to fully exploit WGS data. These integrative approaches broadly fall under the rubric of 'phylodynamics' (90), which covers both evolutionary and epidemiological modelling-based approaches.

Table I
A comparison between estimates of the substitution rates (per genome per year) taken from whole-genome sequencing analyses on the *Mycobacterium* family
The mutation rate found in New Zealand is considerably higher

Citation	Bacteria species	Substitution rate			Host sampled	Country
		Mean	Lower	Upper		
Walker *et al.*, 2013 (82)	*M. tuberculosis*	0.5	0.3	0.7	Human	United Kingdom
Bryant *et al.*, 2013 (83)	*M. tuberculosis*	0.3	N/A	N/A	Human	Netherlands
Roetzer *et al.*, 2013 (84)	*M. tuberculosis*	0.4	0.3	0.7	Human	Germany
Biek *et al.*, 2012 (75)	*M. bovis*	0.15	0.04	0.26	Cattle/badger	United Kingdom
Trewby *et al.*, 2016 (85)	*M. bovis*	0.2	0.1	0.3	Cattle/badger	United Kingdom
Crisp *et al.* (in preparation)	*M. bovis*	0.59	0.3	0.95	Cattle/possum	New Zealand

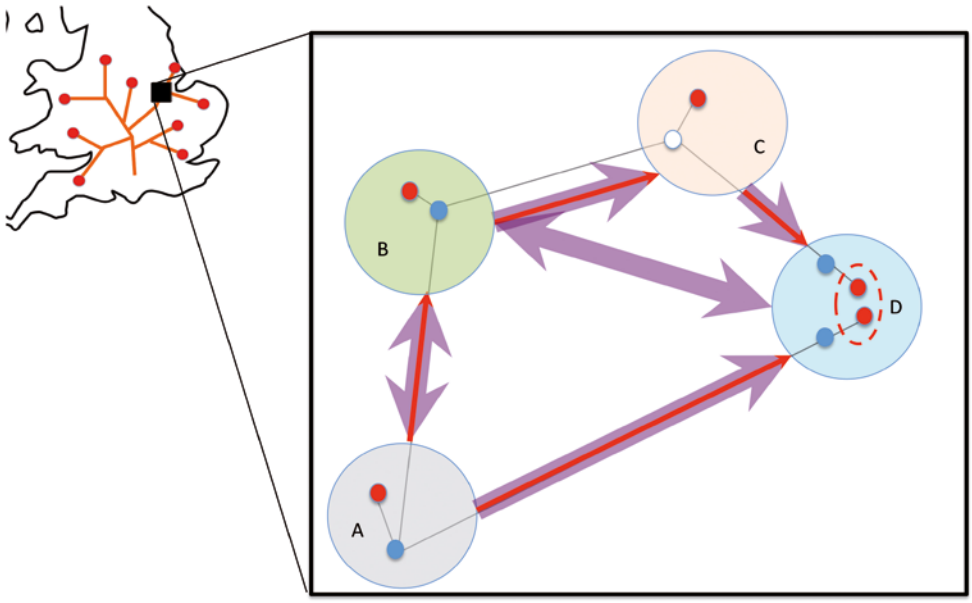


Fig. 5
Illustration of four problems in interpreting whole-genome sequencing data forensically at the micro-scale
1. Due to spatial clustering, there are three possible sources for the ancestors to the sample taken from D
2. All samples taken from individuals (in red) contain potential mutations compared to the transmitted type (in blue)
3. Transmission occurs in one direction (red arrows), but possible pathways are usually bi-directional (purple arrows)
4. The sample in D contains elements of two infections, potentially causing problems for both the bioinformatics and epidemiological interpretation

Evolutionary analyses that directly exploit the molecular clock of mutations over time have been remarkably successful, even when data are relatively scarce (91). Here, the reconstruction of evolutionary pathways depends on the identification of common ancestral traits. At coarse geographical scales, infectious disease processes are either spatially correlated or well described by recordable interactions such as livestock movements (92). The key quantity that underpins these analyses is the estimation of the 'time to most recent common ancestor' (TMRCA) for any two observed strains of pathogen as derived from 'coalescent theory' (93). TMRCAs can then be used to estimate, for example, rates of diffusive spread across landscapes ([94]; Fig. 6a). Although such processes may contain inconsistencies with sampled phylogenetic trees, these can often be neglected at this scale, and neither high sampling densities nor information about the contact relationships are necessary to infer epidemiologically relevant parameters. However, estimates of TMRCAs rely on the key assumption of a well-mixed (or 'panmictic') population. Throughout an epidemic, but especially in the early stages, the underlying pathogen population is expanding, and this influences estimates of the TRMCA (Fig. 7). This influence can be substantial (95); to account for it, current approaches exploit the concept of an 'effective population size' in coalescent-based inference schemes (96), but these still rely on assumptions of panmixia. Where the panmictic assumption is poor, 'structured coalescent' models (97) can be used to divide the population into multiple sub-populations, although implementation of this approach is computationally taxing.

Population structures such as those found in explicit contact networks present additional difficulties, but mathematical or simulation models that explicitly represent transmission processes can prove useful in dealing with the problems. These models naturally incorporate epidemiological information (e.g. incubation periods, transmission rates) and can be used to explore assumptions at the population level (implications of contact structure, duration of immunity, changes in population density, etc.). When they directly incorporate epidemiological and genetic data, and are fitted to the available data using modern parameter inference methods (98), they can be powerful tools for understanding epidemiological systems.

Contact structure plays a particularly important role when there are high levels of clustering (Fig. 5). Clustering can be due to spatial proximity (99), or common social or non-spatial contacts (100), such as when two farms

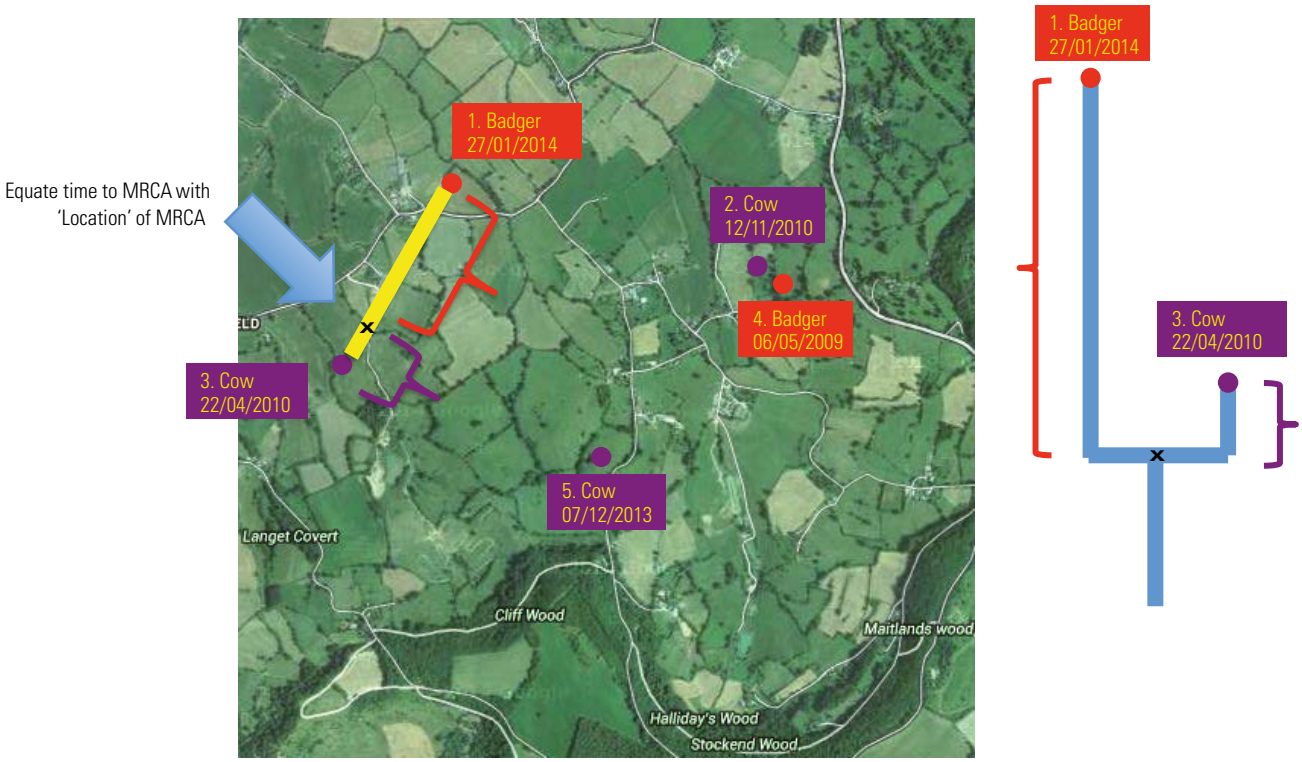


Fig. 6a

Impact of ancestor estimation on estimation of spatial diffusion rates

Equating time to most recent common ancestor (MRCA) to the most likely spatial location of the ancestor forms the basis of phylogeographic estimates of spatial diffusion rates, when this is extended across all samples and to consider all internal nodes

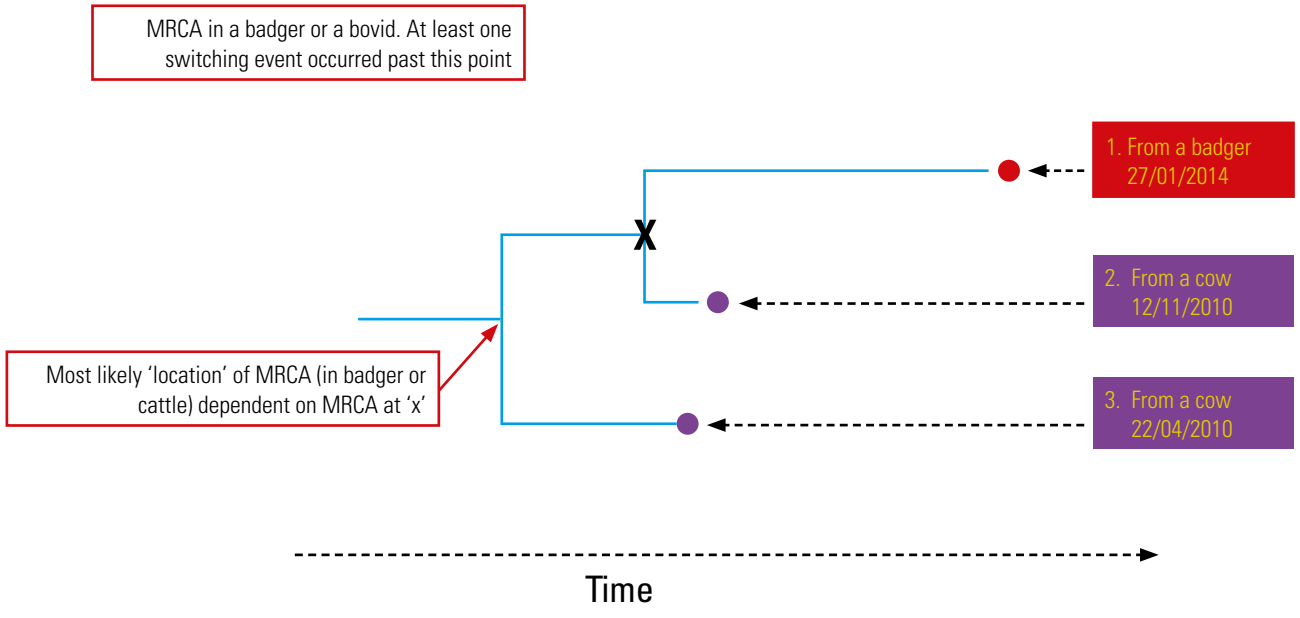


Fig. 6b

Impact of ancestor estimation on estimation of transition rates between species

The time to most recent common ancestor (MRCA) imposes restrictions on the most likely rate at which the bacterium moves between 'discrete traits' (here, species), forming the basis of 'discrete traits analysis' across the whole phylogeny

purchase cattle from a common source, but also trade with each other. Clustering adds ambiguity to the interpretation of contact data, making it difficult to estimate fundamental epidemiological parameters (87). Detailed field investigations can reduce this uncertainty (79), but unless the network of potentially infectious contacts is well characterised, it can be substantially different from the network of actual transmission events (101). Mathematical models have been used to compare spatio-temporal and network data to *M. bovis* WGS data (75), but as yet there are no published reports jointly estimating epidemiological and genetic data, as has been done for other systems (98, 102). These methods are best used where data are dense and accompanied by good-quality epidemiological and demographic data, but can be effective at finely resolved spatio-temporal scales.

Identifying and quantifying the role of reservoir hosts

Wildlife play an important role in the maintenance of bTB in many countries. This presents problems for controlling disease spread, as wildlife are often difficult to manage or even observe. Control options that may work in a single host species may not in another, and if the pathogen can persist in the reservoir on its own, this could render control ineffective.

Estimating the reservoir's importance usually requires quantifying the rates at which the species infect each other. Thus, the problem of clustering is especially relevant, since distinguishing cattle-to-wildlife-to-cattle transmission from direct cattle-to-cattle transmission is critical. The evolutionary history of a multi-host pathogen system can be inferred by mapping discrete traits (i.e. as compartments, as has been done for geographical locations [103, 104]) representing different hosts upon time-scaled trees. This results in: *i)* a posterior set of phylogenetic trees with internal ancestral nodes and branches assigned to particular hosts, and *ii)* the estimated transition rates required to reach that distribution of states (Fig. 6b). The resultant estimates of the size of the pathogen pool per host species as a function of time provides a direct estimate of the rate of cross-species transmission. These approaches are best applied where sampling is relatively unbiased across species and spatio-temporal location, and where mutation rates in the different species are the same, or good prior estimates for the differences exist.

Quantifying the role of wildlife reservoirs is often complicated by biases in the collection of epidemiological and genetic data. Where mixing between species is either limited or substantial, the role of the under-sampled host can be difficult to identify; moderate mixing creates a much more distinctive phylogenetic signature (105). In contrast,

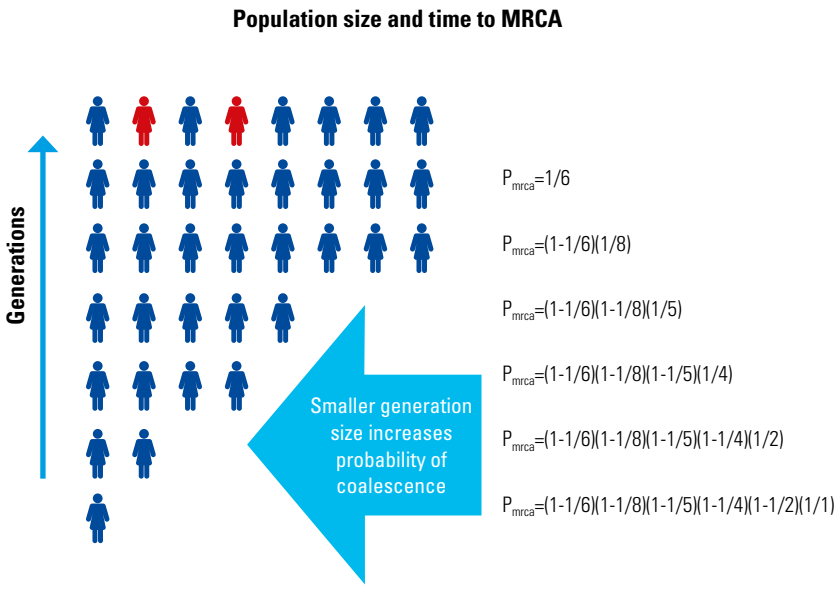


Fig. 7
Impact of population bottlenecks on estimation of the time to most recent common ancestor
The probability of the time to most recent common ancestor (MRCA) occurring in any given generation depends on the number of individuals in that generation. Here, for any two individuals in the current generation (in red), there is a 1/6 probability that their MRCA was in the previous generation. During an epidemic the pathogen population expands, typically shortening estimates of the time to MRCA. Having a single infected individual in a past generation limits the longest possible time to MRCA

the epidemiological signature of the hidden host is much clearer where mixing is high, especially in the case of intervention studies, which would likely have a dramatic effect on disease incidence (106). Thus, the combination of phylogenetic information and epidemiological data again has the potential to provide much deeper insight than either on its own.

Applying whole-genome sequencing to bovine tuberculosis epidemiology on the mass scale

While WGS offers unprecedented potential for inter-laboratory comparison, laboratories and disease programmes face significant challenges integrating and adopting it. Its implementation requires considerable forethought. While sequencing costs have dramatically decreased, it remains an expensive technology, with most of the decrease in costs in the last few years associated with the development of increased throughput, and in the use of data derived from a single sequencing run. For a diagnostic laboratory, batching samples may affect the timeliness of diagnostic results and often the cost advantages are not fully realised. Where throughput is low and the needs limited, the routine deployment of WGS may not be practical, as it requires skilled, well-trained staff and relatively complex, expensive technology. If WGS is reserved for unusual, especially important outbreaks, using external providers may prove a more viable strategy.

Bioinformatics requirements present additional challenges, and meeting these challenges requires a diversity of skill sets and a team approach, incorporating systems administration, database development and management, pipeline development and maintenance, data validation, data visualisation and reporting. In this new paradigm, bioinformatics staff will be equal partners with traditional laboratory staff in providing results. Commercial software packages are often insufficient for this constantly evolving technology and most laboratories will benefit from staff able to use command line programming to manage data and develop analysis pipelines that can capitalise on the latest open source programs. In some cases, larger government laboratories have offered analysis pipelines that can be adopted by laboratories as needed. For example, the United States Department of Agriculture (USDA) Food and Drug Administration has developed the CFSAN SNP Pipeline for salmonella and other foodborne pathogens (github.com/CFSAN-Biostatistics/snp-pipeline) (107), and the USDA National Veterinary Services Laboratories have pipelines available for several animal pathogens, including *Brucella* species, *M. avium* ssp. *paratuberculosis* and the *Mycobacterium tuberculosis* complex (including *M. bovis*) (github.com/USDA-VS).

In principle, WGS data should be easy to compare in laboratories across the world, but there are as yet no uniform standards in place for data processing, and different treatments can have a significant influence on the resulting lineages. Whether communicating to the World Organisation for Animal Health (OIE), to trading partners, or when collaborating with researchers internationally, sharing data and communicating results is enormously beneficial for improving human and animal health. Countries and laboratories must be strongly encouraged to share data and sequences as soon as reasonably possible. As with any genotyping technology, the scope and quality of the database is critical, as this is what will determine how well it can compare new sequences against pre-existing ones. Ideally, there should be an international database that meets predefined sequence quality standards and metadata standards. These are yet to be developed but precedents exist; for example, both the Wellcome Trust and the National Institutes of Health have set standards for developing data-sharing policies for research. This must now be done for sequence data generated from OIE and governmental reference laboratories conducting routine diagnostics on diseases of concern. This is especially important for zoonotic diseases, as it benefits public health researchers and institutions as they leverage these databases for their epidemiological investigations, in the knowledge that they are being maintained by animal health experts with situational intelligence.

Comparing the data generated by older typing methods with the phylogeny generated from WGS will enable us to identify inconsistencies in traditional naming conventions and errors resulting, for example, from homoplasies such as are common with VNTR typing (85, 108). With this new understanding of evolutionary history, these errors can be corrected and historical genotypes named according to how the pathogens evolved, allowing for more accurate communications of disease outbreaks and status reports.

Conclusions

Whole-genome sequencing shares with previous typing methods the need for consistency, high throughput, and low cost; but unlike them it offers many opportunities to improve our understanding of the epidemiology of bTB at multiple scales. However, the resolution of the phylogenetic data is poor and the mutation rate of *M. bovis* highly variable when compared to that of the RNA viruses on which much of the previous success of WGS analysis

in epidemiology is based. Thus, these data at the micro-scale must be interpreted with caution, particularly where wildlife reservoirs are involved. Fortunately, analytical tools that exploit both epidemiological and evolutionary data are available to maximise our ability to interpret these data, but these still need to be proven for bTB. As datasets get both denser and more extensive, the opportunities to integrate data using these analytical techniques will increase and allow for greater epidemiological insight; however, these developments must be supported by concomitant developments in infrastructure and data quality standards.

Acknowledgement

The authors would like to thank Des Collins and Geoff de Lisle for critically reviewing an earlier version of this article.

L'utilisation de la génomique pour retracer la transmission de la tuberculose bovine

R.R. Kao, M. Price-Carter & S. Robbe-Austerman

Résumé

Les maladies infectieuses affectant les animaux d'élevage sont plus difficiles à contrôler lorsqu'il existe un réservoir sauvage, celui-ci étant souvent difficile à observer et à gérer. Ce problème est particulièrement crucial dans le cas de la tuberculose bovine en raison de la durée prolongée de l'infection et des faibles niveaux d'infectiosité qui rendent difficile le traçage des sources d'infection. Pendant plus de 30 ans, le processus de traçage des contacts s'est appuyé sur l'exploitation de marqueurs moléculaires au sein de l'agent pathogène, mais cette technique n'a guère pu aller au-delà d'une caractérisation d'associations générales entre vastes communautés de types similaires. L'avènement récent du séquençage massif à haut débit du génome entier a toutefois révolutionné l'épidémiologie légale appliquée à d'autres maladies, et il en ira bientôt probablement de même pour la tuberculose bovine. Les auteurs de cette synthèse s'intéressent au contexte historique de la mise au point du séquençage du génome entier en relevant ce que les techniques moléculaires antérieures avaient déjà accompli. Ils soulignent les principales méthodes pour interpréter les données générées par le séquençage du génome entier et examinent aussi bien le rôle des techniques analytiques les plus avancées basées sur l'exploitation des propriétés évolutionnistes et épidémiologiques du système que les problèmes qui se posent lorsqu'on cherche à quantifier le rôle joué par les réservoirs inapparents d'une maladie. Enfin, ils exposent les difficultés particulières liées à la mise en œuvre de cette technologie pour des applications diagnostiques de routine ainsi que son potentiel d'utilisation à grande échelle.

Mots-clés

Agent pathogène à spectre d'hôte – Génomique – Phylodynamique – Séquençage à haut débit – Traçage des contacts – Tuberculose bovine.

Uso de la genómica para rastrear la transmisión de la tuberculosis bovina

R.R. Kao, M. Price-Carter & S. Robbe-Austerman

Resumen

La presencia de un reservorio en la fauna salvaje siempre complica la lucha contra las enfermedades infecciosas del ganado, en la medida en que esos reservorios son observados con poca frecuencia y resultan difíciles de gestionar. Este problema cobra especial gravedad en el caso de la tuberculosis bovina, pues la larga duración de la infección y los bajos niveles de infecciosidad hacen difícil localizar el origen de los focos. Durante más de 30 años se han empleado marcadores moleculares del patógeno como método auxiliar en el proceso de localización de los contactos, pero ello casi siempre ha servido únicamente para caracterizar correlaciones más bien laxas entre grandes comunidades de tipos parecidos. En los últimos tiempos, sin embargo, el advenimiento de la secuenciación masiva de alto rendimiento de genomas completos ha revolucionado la epidemiología forense aplicada a otras enfermedades, y ahora puede ocurrir otro tanto con la tuberculosis bovina. Los autores, tras repasar el contexto histórico del uso de la secuenciación de genomas completos, exponen los resultados que hasta la fecha se han podido obtener con las técnicas moleculares anteriores. Asimismo, describen brevemente los principales métodos para interpretar los datos de secuenciación de genomas completos y examinan tanto la función de las técnicas analíticas avanzadas que explotan las propiedades evolutivas y epidemiológicas del sistema como los problemas que surgen para cuantificar la intervención de reservorios ocultos de enfermedad. Por último, exponen las especiales dificultades que plantea el desarrollo de esta tecnología para efectuar diagnósticos sistemáticos y las posibilidades que ofrece para una utilización generalizada.

Palabras clave

Filodinámica – Genómica – Localización de contactos – Patógeno con múltiples anfitriones – Secuenciación de alto rendimiento – Tuberculosis bovina.

References

1. Collins D.M. (1999). – Molecular epidemiology: *Mycobacterium bovis*. *In* Mycobacteria: molecular biology and virulence (C. Ratledge & J. Dale, eds). Blackwell Publishing Ltd, 123–135. doi:10.1002/9781444311433.ch7.

2. Durr P.A., Hewinson R.G. & Clifton-Hadley R.S. (2000). – Molecular epidemiology of bovine tuberculosis. I. *Mycobacterium bovis* genotyping. *Rev. Sci. Tech. Off. Int. Epiz.*, **19** (3), 675–688.

3. Gormley E., Corner L.A.L., Costello E. & Rodriguez-Campos S. (2014). – Bacteriological diagnosis and molecular strain typing of *Mycobacterium bovis* and *Mycobacterium caprae*. *Res. Vet. Sci.*, **97**, S30–S43. doi:10.1016/j.rvsc.2014.04.010.

4. Durr P.A., Clifton-Hadley R.S. & Hewinson R.G. (2000). – Molecular epidemiology of bovine tuberculosis. II. Applications of genotyping. *Rev. Sci. Tech. Off. Int. Epiz.*, **19** (3), 689–701.

5. Skuce R.A. & Neill S.D. (2001). – Molecular epidemiology of *Mycobacterium bovis*: exploiting molecular data. *Tuberculosis (Edinb.)*, **81** (1–2), 169–175. doi:10.1054/tube.2000.0270.

6. De Lisle G.W., Bengis R.G., Schmitt S.M. & O'Brien D.J. (2002). – Tuberculosis in free-ranging wildlife: detection, diagnosis and management. *In* Infectious diseases of wildlife: detection, diagnosis and management (Part Two) (R.G. Bengis, ed.). *Rev. Sci. Tech. Off. Int. Epiz.*, **21** (2), 317–334.

7. Smith N.H., Gordon S.V., de la Rua-Domenech R., Clifton-Hadley R.S. & Hewinson R.G. (2006). – Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.*, **4** (9), 670–681. doi:10.1038/nrmicro1472.

8. Collins D.M. (2011). – Advances in molecular diagnostics for *Mycobacterium bovis*. *Vet. Microbiol.*, **151** (1–2), 2–7. doi:10.1016/j.vetmic.2011.02.019.

9. Ramos D.F., Tavares L., da Silva P.E. & Dellagostin O.A. (2014). – Molecular typing of *Mycobacterium bovis* isolates: a review. *Braz. J. Microbiol.*, **45** (2), 365–372. doi:10.1590/S1517-83822014005000045.

10. Rodriguez-Campos S., Smith N.H., Bonjotti M.B. & Aranaz A. (2014). – Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics and legislation of bovine tuberculosis. *Res. Vet. Sci.*, **97**, S5–S19. doi:10.1016/j.rvsc.2014.02.009.

11. Pérez-Lago L., Navarro Y. & García-de-Viedma D. (2014). – Current knowledge and pending challenges in zoonosis caused by *Mycobacterium bovis*: a review. *Res. Vet. Sci.*, **97**, S94–S100. doi:10.1016/j.rvsc.2013.11.008.

12. Collins D.M. & de Lisle G.W. (1984). – DNA restriction endonuclease analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG. *J. Gen. Microbiol.*, **130** (4), 1019–1021. doi:10.1099/00221287-130-4-1019.

13. Collins D.M. & de Lisle G.W. (1985). – DNA restriction endonuclease analysis of *Mycobacterium bovis* and other members of the tuberculosis complex. *J. Clin. Microbiol.*, **21** (4), 562–564.

14. 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** (4), 907–914.

15. Hermans P.W., van Soolingen D., Bik E.M., de Haas P.E., Dale J.W. & van Embden J.D. (1991). – Insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.*, **59** (8), 2695–2705.

16. Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A. & Horvath P. (2007). – CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315** (5819), 1709–1712. doi:10.1126/science.1138140.

17. Roring S., Hughes M.S., Beck L.A., Skuce R.A. & Neill S.D. (1998). – Rapid diagnosis and strain differentiation of *Mycobacterium bovis* in radiometric culture by spoligotyping. *Vet. Microbiol.*, **61** (1–2), 71–80. doi:10.1016/S0378-1135(98)00167-9.

18. Roring S., Hughes M.S., Skuce R.A. & Neill S.D. (2000). – Simultaneous detection and strain differentiation of *Mycobacterium bovis* directly from bovine tissue specimens by spoligotyping. *Vet. Microbiol.*, **74** (3), 227–236. doi:10.1016/S0378-1135(00)00189-9.

19. Bidovec-Stojkovič U., Seme K., Žolnir-Dovč M. & Supply P. (2014). – Prospective genotyping of *Mycobacterium tuberculosis* from fresh clinical samples. *PLoS One*, **9** (10), e109547. doi:10.1371/journal.pone.0109547.

20. Cousins D.V., Skuce R.A., Kazwala R.R. & van Embden J.D. (1998). – Towards a standardized approach to DNA fingerprinting of *Mycobacterium bovis*. International Union Against Tuberculosis and Lung Disease, Tuberculosis in Animals Subsection. *Int. J. Tuberc. Lung Dis.*, **2** (6), 471–478.

21. Smith N.H. (2012). – The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. *Infect. Genet. Evol.*, **12** (4), 857–865. doi:10.1016/j.meegid.2011.09.007.

22. Rodriguez-Campos S., González S., de Juan L., Romero B., Bezos J., Casal C., Álvarez J., Fernández-de-Mera I.G., Castellanos E., Mateos A., Sáez-Llorente J.L., Domínguez L., Aranaz A. & Spanish Network on Surveillance Monitoring of Animal Tuberculosis (2012). – A database for animal tuberculosis (mycoDB.es) within the context of the Spanish national programme for eradication of bovine tuberculosis. *Infect. Genet. Evol.*, **12** (4), 877–882. doi:10.1016/j.meegid.2011.10.008.

23. Weniger T., Krawczyk J., Supply P., Niemann S. & Harmsen D. (2010). – MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res.*, **38** (Suppl. 2), W326–W331. doi:10.1093/nar/gkq351.

24. Versalovic J., Koeuth T. & Lupski J.R. (1991). – Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.*, **19** (24), 6823–6831. doi:10.1093/nar/19.24.6823.

25. Frothingham R. & Meeker-O'Connell W.A. (1998). – Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology*, **144** (5), 1189–1196. doi:10.1099/00221287-144-5-1189.

26. Le Flèche P., Fabre M., Denoeud F., Koeck J.L. & Vergnaud G. (2002). – High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiol.*, **2**, 37. doi:10.1186/1471-2180-2-37.

27. Allix C., Walravens K., Saegerman C., Godfroid J., Supply P. & Fauville-Dufaux M. (2006). – Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of *Mycobacterium bovis* and comparison of the method with IS*6110* restriction fragment length polymorphism analysis and spoligotyping. *J. Clin. Microbiol.*, **44** (6), 1951–1962. doi:10.1128/JCM.01775-05.

28. Boniotti M.B., Goria M., Loda D., Garrone A., Benedetto A., Mondo A., Tisato E., Zanoni M., Zoppi S., Dondo A., Tagliabue S., Bonora S., Zanardi G. & Pacciarini M.L. (2009). – Molecular typing of *Mycobacterium bovis* strains isolated in Italy from 2000 to 2006 and evaluation of variable-number tandem repeats for geographically optimized genotyping. *J. Clin. Microbiol.*, **47** (3), 636–644. doi:10.1128/JCM.01192-08.

29. Martinez L.R., Harris B., Black W.C. 4th, Meyer R.M., Brennan P.J., Vissa V.D. & Jones R.L. (2008). – Genotyping North American animal *Mycobacterium bovis* isolates using multilocus variable number tandem repeat analysis. *J. Vet. Diagn. Invest.*, **20** (6), 707–715. doi:10.1177/104063870802000601.

30. Price-Carter M., Rooker S. & Collins D.M. (2011). – Comparison of 45 variable number tandem repeat (VNTR) and two direct repeat (DR) assays to restriction endonuclease analysis for typing isolates of *Mycobacterium bovis*. *Vet. Microbiol.*, **150** (1–2), 107–114. doi:10.1016/j.vetmic.2011.01.012.

31. Tsao K., Robbe-Austerman S., Miller R.S., Portacci K., Grear D.A. & Webb C. (2014). – Sources of bovine tuberculosis in the United States. *Infect. Genet. Evol.*, **28**, 137–143. doi:10.1016/j.meegid.2014.09.025.

32. 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** (17), 501–504. doi:10.1136/vr.157.17.501.

33. Rodriguez-Campos S., Aranaz A., de Juan L., Sáez-Llorente J.L., Romero B., Bezos J., Jiménez A., Mateos A. & Domínguez L. (2011). – Limitations of spoligotyping and variable-number tandem-repeat typing for molecular tracing of *Mycobacterium bovis* in a high-diversity setting. *J. Clin. Microbiol.*, **49** (9), 3361–3364. doi:10.1128/JCM.00301-11.

34. Allix-Beguec C., Harmsen D., Weniger T., Supply P. & Niemann S. (2008). – Evaluation and strategy for use of MIRU-VNTR*plus*, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *J. Clin. Microbiol.*, **46** (8), 2692–2699. doi:10.1128/JCM.00540-08.

35. De Beer J.L., Kremer K., Ködmön C., Supply P., van Soolingen D. & Global Network for the Molecular Surveillance of Tuberculosis 2009 (2012). – First worldwide proficiency study on variable-number tandem-repeat typing of *Mycobacterium tuberculosis* complex strains. *J. Clin. Microbiol.*, **50** (3), 662–669. doi:10.1128/JCM.00607-11.

36. De Beer J.L., Ködmön C., van Ingen J., Supply P., van Soolingen D. & Global Network for Molecular Surveillance of Tuberculosis 2010 (2014). – Second worldwide proficiency study on variable number of tandem repeats typing of *Mycobacterium tuberculosis* complex. *Int. J. Tuberc. Lung Dis.*, **18** (5), 594–600. doi:10.5588/ijtld.13.0531.

37. Smith N.H., Dale J., Inwald J., Palmer S., Gordon S.V., Hewinson R.G. & Smith J.M. (2003). – The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc. Natl Acad. Sci. USA*, **100** (25), 15271–15275. doi:10.1073/pnas.2036554100.

38. McLernon J., Costello E., Flynn O., Madigan G. & Ryan F. (2010). – Evaluation of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat analysis and spoligotyping for genotyping of *Mycobacterium bovis* isolates and a comparison with restriction fragment length polymorphism typing. *J. Clin. Microbiol.*, **48** (12), 4541–4545. doi:10.1128/JCM.01175-10.

39. Illumina (2015). – An introduction to next-generation sequencing. Available at: www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf. (accessed on 7 July 2015).

40. Rodriguez-Campos S., Navarro Y., Romero B., de Juan L., Bezos J., Mateos A., Golby P., Smith N.H., Hewinson G.R., Domínguez L., García-de-Viedma D. & Aranaz A. (2013). – Splitting of a prevalent *Mycobacterium bovis* spoligotype by variable-number tandem-repeat typing reveals high heterogeneity in an evolving clonal group. *J. Clin. Microbiol.*, **51** (11), 3658–3665. doi:10.1128/JCM.01271-13.

41. Brosch R., Gordon S.V., Marmiesse M., Brodin P., Buchrieser C., Eiglmeier K., Garnier T., Gutierrez C., Hewinson G., Kremer K., Parsons L.M., Pym A.S., Samper S., van Soolingen D. & Cole S.T. (2002). – A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl Acad. Sci. USA*, **99** (6), 3684–3689. doi:10.1073/pnas.052548299.

42. Smith N.H., Berg S., Dale J., Allen A., Rodriguez S., Romero B., Matos F., Ghebremichael S., Karoui C., Donati C., da Conceição Machado A., Mucavele C., Kazwala R.R., Hilty M., Cadmus S., Ngandolo B.N., Habtamu M., Oloya J., Muller A., Milian-Suazo F., Andrievskaia O., Projahn M., Barandiarán S., Macías A., Müller B., Zanini M.S., Ikuta C.Y., Rodriguez C.A., Pinheiro S.R., Figueroa A., Cho S.N., Mosavari N., Chuang P.C., Jou R., Zinsstag J., van Soolingen D., Costello E., Aseffa A., Proaño-Perez F., Portaels F., Rigouts L., Cataldi A.A., Collins D.M., Boschiroli M.L., Hewinson R.G., Ferreira Neto J.S., Surujballi O., Tadyon K., Botelho A., Zárraga A.M., Buller N., Skuce R., Michel A., Aranaz A., Gordon S.V., Jeon B.Y., Källenius G., Niemann S., Boniotti M.B., van Helden P.D., Harris B., Zumárraga M.J. & Kremer K. (2011). – European 1: a globally important clonal complex of *Mycobacterium bovis*. *Infect. Genet. Evol.*, **11** (6), 1340–1351. doi:10.1016/j.meegid.2011.04.027.

43. Parra A., Larrasa J., García A., Alonso J.M. & de Mendoza J.H. (2005). – Molecular epidemiology of bovine tuberculosis in wild animals in Spain: a first approach to risk factor analysis. *Vet. Microbiol.*, **110** (3–4), 293–300. doi:10.1016/j.vetmic.2005.07.013.

44. McCluskey B., Lombard J., Strunk S., Nelson D., Robbe-Austerman S., Naugle A. & Edmondson A. (2014). – *Mycobacterium bovis* in California dairies: a case series of 2002–2013 outbreaks. *Prev. Vet. Med.*, **115** (3–4), 205–216. doi:10.1016/j.prevetmed.2014.04.010.

45. Gutiérrez Reyes J.A., García Casanova L., Romero Torres C., Sosa Gallegos S.L., Cantó Alarcón G.J., Mercado Pezzat M., Pizano Martínez O., Estrada Chávez C. & Milián Suazo F. (2012). – Population structure of *Mycobacterium bovis* isolates from cattle in Mexico. *Prev. Vet. Med.*, **106** (1), 1–8. doi:10.1016/j.prevetmed.2012.05.008.

46. Cadmus S.I., Gordon S.V., Hewinson R.G. & Smith N.H. (2011). – Exploring the use of molecular epidemiology to track bovine tuberculosis in Nigeria: an overview from 2002 to 2004. *Vet. Microbiol.*, **151** (1–2), 133–138. doi:10.1016/j.vetmic.2011.02.036.

47. Muller B., Hilty M., Berg S., García-Pelayo M.C., Dale J., Boschiroli M.L., Cadmus S., Ngandolo B.N., Godreuil S., Diguimbaye-Djaibé C., Kazwala R., Bonfoh B., Njanpop-Lafourcade B.M., Sahraoui N., Guetarni D., Aseffa A., Mekonnen M.H., Razanamparany V.R., Ramarokoto H., Djønne B., Oloya J., Machado A., Mucavele C., Skjerve E., Portaels F., Rigouts L., Michel A., Müller A., Källenius G., van Helden P.D., Hewinson R.G., Zinsstag J., Gordon S.V. & Smith N.H. (2009). – African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. *J. Bacteriol.*, **191** (6), 1951–1960. doi:10.1128/JB.01590-08.

48. Müller B., Steiner B., Bonfoh B., Fané A., Smith N.H. & Zinsstag J. (2008). – Molecular characterisation of *Mycobacterium bovis* isolated from cattle slaughtered at the Bamako abattoir in Mali. *BMC Vet. Res.*, **4**, 26. doi:10.1186/1746-6148-4-26.

49. Parreiras P.M., Andrade G.I., Figueiredo do Nascimento T., Oelemann M.C., Gomes H.M., Alencar A.P., Assis R.A., Mota P.M., Pereira M.A., Lobato F.C., Lage A.P. & Suffys P.N. (2012). – Spoligotyping and variable number tandem repeat analysis of *Mycobacterium bovis* isolates from cattle in Brazil. *Mem. Inst. Oswaldo Cruz*, **107** (1), 64–73. doi:10.1590/S0074-02762012000100009.

50. Figueiredo E.E., Ramos D.F., Medeiros L., Silvestre F.G., Lilenbaum W., Silva J.T., Paschoalin V.M. & Dellagostin O.A. (2012). – Multiple strains of *Mycobacterium bovis* revealed by molecular typing in a herd of cattle. *Vet. J.*, **193** (1), 296–298. doi:10.1016/j.tvjl.2011.11.006.

51. Zumarrága M.J., Arriaga C., Barandiaran S., Cobos-Marín L., de Waard J., Estrada-García I., Figueiredo T., Figueroa A., Giménez F., Gomes H.M., Gonzalez-y-Merchand J.A., Macías A., Milián-Suazo F., Rodríguez C.A., Santillán M.A., Suffys P.N., Trangoni M.D., Zárraga A.M. & Cataldi A. (2013). – Understanding the relationship between *Mycobacterium bovis* spoligotypes from cattle in Latin American countries. *Res. Vet. Sci.*, **94** (1), 9–21. doi:10.1016/j.rvsc.2012.07.012.

52. Goodchild A.V., Watkins G.H., Sayers A.R., Jones J.R. & Clifton-Hadley R.S. (2012). – Geographical associations between the genotype of bovine tuberculosis in found dead badgers and in cattle herds. *Vet. Rec.*, **170** (10), 259–263. doi:10.1136/vr.100193.

53. Collins D.M., Gabric D.M. & de Lisle G.W. (1988). – Typing of *Mycobacterium bovis* isolates from cattle and other animals in the same locality. *N.Z. Vet. J.*, **36** (1), 45–46. doi:10.1080/00480169.1988.35476.

54. Collins D.M., Erasmuson S.K., Stephens D.M., Yates G.F. & de Lisle G.W. (1993). – DNA fingerprinting of *Mycobacterium bovis* strains by restriction fragment analysis and hybridization with insertion elements IS*1081* and IS*6110*. *J. Clin. Microbiol.*, **31** (5), 1143–1147.

55. Collins D.M., de Lisle G.W. & Gabric D.M. (1986). – Geographic distribution of restriction types of *Mycobacterium bovis* isolates from brush-tailed possums (*Trichosurus vulpecula*) in New Zealand. *J. Hyg. (Lond.)*, **96** (3), 431–438. doi:10.1017/S0022172400066201.

56. De Lisle G.W., Yates G.F., Collins D.M., MacKenzie R.W., Crews K.B. & Walker R. (1995). – A study of bovine tuberculosis in domestic animals and wildlife in the MacKenzie Basin and surrounding areas using DNA fingerprinting. *N.Z. Vet. J.*, **43** (7), 266–271. doi:10.1080/00480169./1995.35905.

57. Collins D.M. (1999). – DNA typing of *Mycobacterium bovis* strains from the Castlepoint area of the Wairarapa. *N.Z. Vet. J.*, **47** (6), 207–209. doi:10.1080/00480169.1999.36145.

58. De Lisle G.W., Kawakami R.P., Yates G.F. & Collins D.M. (2008). – Isolation of *Mycobacterium bovis* and other mycobacterial species from ferrets and stoats. *Vet. Microbiol.*, **132** (3–4), 402–407. doi:10.1016/j.vetmic.2008.05.022.

59. Corner L.A., Stevenson M.A., Collins D.M. & Morris R.S. (2003). – The re-emergence of *Mycobacterium bovis* infection in brushtail possums (*Trichosurus vulpecula*) after localised possum eradication. *N.Z. Vet. J.*, **51** (2), 73–80. doi:10.1080/00480169.2003.36343.

60. Buddle B.M., de Lisle G.W., Griffin J.F. & Hutchings S.A. (2015). – Epidemiology, diagnostics, and management of tuberculosis in domestic cattle and deer in New Zealand in the face of a wildlife reservoir. *N.Z. Vet. J.*, **63** (Suppl. 1), 19–27. doi:10.1080/00480169.2014.929518.

61. Livingstone P.G., Hancox N., Nugent G. & de Lisle G.W. (2015). – Toward eradication: the effect of *Mycobacterium bovis* infection in wildlife on the evolution and future direction of bovine tuberculosis management in New Zealand. *N.Z. Vet. J.*, **63** (Suppl. 1), 1–18. doi:10.1080/00480169.2014.971082.

62. O'Brien D.J., Schmitt S.M., Fitzgerald S.D., Berry D.E. & Hickling G.J. (2006). – Managing the wildlife reservoir of *Mycobacterium bovis*: the Michigan, USA, experience. *Vet. Microbiol.*, **112** (2–4), 313–323. doi:10.1016/j.vetmic.2005.11.014.

63. Carstensen M., O'Brien D.J. & Schmitt S.M. (2011). – Public acceptance as a determinant of management strategies for bovine tuberculosis in free-ranging US wildlife. *Vet. Microbiol.*, **151** (1–2), 200–204. doi:10.1016/j.vetmic.2011.02.046.

64. Shrikrishna D., de la Rua-Domenech R., Smith N.H., Colloff A. & Coutts I. (2009). – Human and canine pulmonary *Mycobacterium bovis* infection in the same household: re-emergence of an old zoonotic threat? *Thorax*, **64** (1), 89–91. doi:10.1136/thx.2008.106302.

65. Bailey S.S., Crawshaw T.R., Smith N.H. & Palgrave C.J. (2013). – *Mycobacterium bovis* infection in domestic pigs in Great Britain. *Vet. J.*, **198** (2), 391–397. doi:10.1016/j.tvjl.2013.08.035.

66. Brooks-Pollock E., Roberts G.O. & Keeling M.J. (2014). – A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature*, **511** (7508), 228–231. doi:10.1038/nature13529.

67. Broughan J.M., Harris K.A., Brouwer A., Downs S.H., Goodchild A.V., Upton P.A. & Smith N.H. (2015). – Bovine TB infection status in cattle in Great Britain in 2013. *Vet. Rec.*, **176** (13), 326–330. doi:10.1136/vr.h1145.

68. Barnett J.E., Booth P., Brewer J.I., Chanter J., Cooper T., Crawshaw T., Davison N.J., Greenwood A., Riley P., Smith N.H. & Wessels M. (2013). – *Mycobacterium bovis* infection in a grey seal pup (*Halichoerus grypus*). *Vet. Rec.*, **173** (7), 168. doi:10.1136/vr.101480.

69. Harris K.A., Downs S.H., Goodchild A.V., Broughan J.M., Upton P.A. & Smith N.H. (2014). – Bovine TB infection status in cattle in Great Britain in 2012. *Vet. Rec.*, **174** (24), 600–604. doi:10.1136/vr.g3644.

70. Godfray H.C., Donnelly C.A., Kao R.R., Macdonald D.W., McDonald R.A., Petrokofsky G., Wood J.L., Woodroffe R., Young D.B. & McLean A.R. (2013). – A restatement of the natural science evidence base relevant to the control of bovine tuberculosis in Great Britain. *Proc. Biol. Sci.*, **280** (1768), 20131634. doi:10.1098/rspb.2013.1634.

71. Garnier T., Eiglmeier K., Camus J.C., Medina N., Mansoor H., Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe 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. USA*, **100** (13), 7877–7882. doi:10.1073/pnas.1130426100.

72. Votintseva A.A., Pankhurst L.J., Anson L.W., Morgan M.R., Gascoyne-Binzi D., Walker T.M., Phuong Quan T., Wyllie D.H., del Ojo Elias C., Wilcox M., Walker A.S., Peto T.E. & Crook D.W. (2015). – Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. *J. Clin. Microbiol.*, **53** (4), 1137–1143. doi:10.1128/JCM.03073-14.

73. Morelli M.J., Wright C.F., Knowles N.J., Juleff N., Paton D.J., King D.P. & Haydon D.T. (2013). – Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. *Vet. Res.*, **44** (1), 12. doi:10.1186/1297-9716-44-12.

74. Nelson M.I. & Holmes E.C. (2007). – The evolution of epidemic influenza. *Nat. Rev. Genet.*, **8** (3), 196–205. doi:10.1038/nrg2053.

75. 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** (11), e1003008. doi:10.1371/journal.ppat.1003008.

76. Ford C., Yusim K., Ioerger T., Feng S., Chase M., Greene M., Korber B. & Fortune S. (2012). – Mycobacterium tuberculosis: heterogeneity revealed through whole genome sequencing. *Tuberculosis (Edinb.)*, **92** (3), 194–201. doi:10.1016/j.tube.2011.11.003.

77. Wlodarska M., Johnston J.C., Gardy J.L. & Tang P. (2015). – A microbiological revolution meets an ancient disease: improving the management of tuberculosis with genomics. *Clin. Microbiol. Rev.*, **28** (2), 523–539. doi:10.1128/CMR.00124-14.

78. Niemann S. & Supply P. (2014). – Diversity and evolution of *Mycobacterium tuberculosis*: moving to whole-genome-based approaches. *Cold Spring Harb. Symp. Quant. Biol.*, **4** (12), a021188. doi:10.1101/cshperspect.a021188.

79. Gardy J.L., Johnston J.C., Ho Sui S.J., Cook V.J., Shah L., Brodkin E., Rempel S., Moore R., Zhao Y., Holt R., Varhol R., Birol I., Lem M., Sharma M.K., Elwood K., Jones S.J., Brinkman F.S., Brunham R.C. & Tang P. (2011). – Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.*, **364** (8), 730–739. doi:10.1056/NEJMoa1003176.

80. Glaser L., Carstensen M., Shaw S., Robbe-Austerman S., Wunschmann A., Grear D., Stuber T. & Thomsen B. (2016). – Descriptive epidemiology and whole genome sequencing analysis for an outbreak of bovine tuberculosis in beef cattle and white-tailed deer in northwestern Minnesota. *PLoS ONE*, **11** (1), e0145735. doi:10.1371/journal.pone.0145735.

81. Robbe-Austerman S. (2014). – Interpreting whole genome sequencing results. United States Department of Agriculture, Washington, DC. Available at: www.screencast.com/t/Ew4ejXYjs (accessed on 26 February 2016).

82. Walker T.M., 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. (2013). – Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.*, **13** (2), 137–146. doi:10.1016/S1473-3099(12)70277-3.

83. 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., Siezen R.J., Borgdorff M., Bentley S.D., Parkhill J. & van Soolingen D. (2013). – Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect. Dis.*, **13**, 110. doi:10.1186/1471-2334-13-110.

84. Roetzer A., Diel R., Kohl T.A., Rückert C., Nübel U., Blom J., Wirth T., Jaenicke S., Schuback S., Rüsch-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** (2), e1001387. doi:10.1371/journal.pmed.1001387. E-pub.: 12 February 2013.

85. Trewby H., Wright D., Breadon E.L., Lycett S.J., Mallon T.R., McCormick C., Johnson P., Orton R.J., Allen A.R., Galbraith J., Herzyk P., Skuce R.A., Biek R. & Kao R.R. (2016). – Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. *Epidemics* (in press). E-pub.: 19 October 2015. doi:10.1016/j.epidem.2015.08.003.

86. 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** (3), e91024. doi:10.1371/journal.pone.0091024.

87. 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** (5), 282–291. doi:10.1016/j.tim.2014.02.011.

88. Schürcha A.C., Kremer K., Kiers A., Daviena O., Boeree M.J., Siezen R.J., Smith N.H. & van Soolingen D. (2010). – The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. *Infect. Genet. Evol.*, **10** (1), 108–114. doi:10.1016/j.meegid.2009.10.002.

89. Zaas A.K., Burke T., Chen M., McClain M., Nicholson B., Veldman T., Tsalik E.L., Fowler V., Rivers E.P., Otero R., Kingsmore S.F., Voora D., Lucas J., Hero A.O., Carin L., Woods C.W. & Ginsburg G.S. (2013). – A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Sci. Transl. Med.*, **5** (203), 203ra126. doi:10.1126/scitranslmed.3006280.

90. Grenfell B.T., Pybus O.G., Gog J.R., Wood J.L., Daly J.M., Mumford J.A. & Holmes E.C. (2004). – Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303** (5656), 327–332. doi:10.1126/science.1090727.

91. Biek R., Pybus O.G., Lloyd-Smith J.O. & Didelot X. (2015). – Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.,* **30** (6), 306–313. doi:10.1016/j.tree.2015.03.009.

92. Faria N.R., Suchards M.A., Rambaut A. & Lemey P. (2011). – Toward a quantitative understanding of viral phylogeography. *Curr. Opin. Virol.*, **1** (5), 423–429. doi:10.1016/j.coviro.2011.10.003.

93. Rosenberg N.A. & Nordborg M. (2002). – Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.*, **3** (5), 380–390. doi:10.1038/nrg795.

94. Pybus O.G., Suchard M.A., Lemey P., Bernardin F.J., Rambaut A., Crawford F.W., Gray R.R., Arinaminpathy N., Stramer S.L., Busch M.P. & Delwart E.L. (2012). – Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc. Natl Acad. Sci. USA*, **109** (37), 15066–15071. doi:10.1073/pnas.1206598109.

95. Koelle K. & Rasmussen D.A. (2012). – Rates of coalescence for common epidemiological models at equilibrium. *J. Roy. Soc. Interface*, **9** (70), 997–1007. doi:10.1098/rsif.2011.0495.

96. Minin V.N., Bloomquist E.W. & Suchard M.A. (2008). – Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.*, **25** (7), 1459–1471. doi:10.1093/molbev/msn090.

97. Vaughan T.G., Kühnert D., Popinga A., Welch D. & Drummond A.J. (2014). – Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, **30** (16), 2272–2279. doi:10.1093/bioinformatics/btu201.

98. Morelli M.J., Thébaud G., Chadœuf J., King D.P., Haydon D.T. & Soubeyrand S. (2012). – A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. *PLoS Comput. Biol.*, **8** (11), e1002768. doi:10.1371/journal.pcbi.1002768.

99. Keeling M.J. (1999). – The effects of local spatial structure on epidemiological invasions. *Proc. Roy. Soc. Lond., B, Biol. Sci*, **266** (1421), 859–867. doi:10.1098/rspb.1999.0716.

100. Kiss I.Z., Green D.M. & Kao R.R. (2005). – Disease contact tracing in random and clustered networks. *Proc. Biol. Sci.*, **272** (1570), 1407–1414. doi:10.1098/rspb.2005.3092.

101. Kao R.R., Green D.M., Johnson J. & Kiss I.Z. (2007). – Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *J. Roy. Soc. Interface*, **4** (16), 907–916. doi:10.1098/rsif.2007.1129.

102. Jombart T., Cori A., Didelot X., Cauchemez S., Fraser C. & Ferguson N. (2014). – Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput. Biol.*, **10** (1), e1003457. doi:10.1371/journal.pcbi.1003457.

103. Lemey P., Rambaut A., Drummond A.J. & Suchard M.A. (2009). – Bayesian phylogeography finds its roots. *PLoS Comput. Biol.*, **5** (9), e1000520. doi:10.1371/journal.pcbi.1000520.

104. Rabaa M.A., Ty Hang V.T., Wills B., Farrar J., Simmons C.P. & Holmes E.C. (2010). – Phylogeography of recently emerged DENV-2 in southern Viet Nam. *PLoS Negl. Trop. Dis.*, **4** (7), e766. doi:10.1371/journal.pntd.0000766.

105. Mather A.E., Reid S.W., 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** (6153), 1514–1517. doi:10.1126/science.1240578.

106. Woodroffe R., Donnelly C.A., Cox D.R., Gilks P., Jenkins H.E., Johnston W.T., Le Fèvre A.M., Bourne F.J., Cheeseman C.L., Clifton-Hadley R.S., Gettinby G., Hewinson R.G., McInerney J.P., Mitchell A.P., Morrison W.I. & Watkins G.H. (2009). – Bovine tuberculosis in cattle and badgers in localized culling areas. *J. Wildl. Dis.*, **45** (1), 128–143. doi:10.7589/0090-3558-45.1.128.

107. Davis S., Pettengill J., Luo Y., Payne J., Shpuntoff A., Rand H. & Strain E. (2015). – CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. *Peer J. Comput. Sci.*, **1**, e20. doi:10.7717/peerj-cs.20.

108. 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. & 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 Genomics*, **16**, 161. doi:10.1186/s12864-015-1387-6.