Article

Global distribution and evolution of *Mycobacterium bovis* lineages

Cristina Kraemer Zimpel[1,2], José Salvatore L. Patané[3], Aureliano Coelho Proença Guedes[4], Robson F. de Souza[4], Taiana T. Pereira-Silva[1,2], Naila C. Soler Camargo[1,2], Antônio F. de Souza Filho[2], Cássia Y. Ikuta[2], José Soares Ferreira Neto[2], João Carlos Setubal[3,6], Marcos Bryan Heinemann[2], Ana Marcia de Sá Guimarães[1*]

[1]Laboratory of Applied Research in Mycobacteria, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil.
[2]Department of Preventive Veterinary Medicine and Animal Health, School of Veterinary Medicine and Animal Sciences, University of São Paulo, São Paulo, Brazil.
[3]Department of Biochemistry, Institute of Chemistry, University of São Paulo, São Paulo, Brazil.
[4]Laboratory of Protein Structure and Evolution, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil.
[6]Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24061, USA.
*anamarcia@usp.br

Abstract

Mycobacterium bovis is the main causative agent of zoonotic tuberculosis in humans and frequently devastates livestock and wildlife worldwide. Previous studies suggested the existence of genetic groups of *M. bovis* strains based on limited DNA markers (a.k.a. clonal complexes), and the evolution and ecology of this pathogen has been marginally explored at the global level. We have screened over 2,600 publicly available *M. bovis* genomes and newly sequenced two wildlife *M. bovis* strains, gathering 823 genomes from 21 countries and 21 host-species, including humans, to complete a phylogenomic analyses. We propose the existence of three distinct global lineages of *M. bovis* (Lb1, Lb2 and Lb3) underlying the current disease distribution. These lineages are not fully represented by clonal complexes and are dispersed based on geographic location rather than host species. Our data divergence analysis agreed with previous studies reporting independent archeological data of ancient *M. bovis* [(South Siberian infected skeletons at ~2,000 BP (before present)] and indicates that extant *M. bovis* originated during the Roman Empire period, subsequently dispersing across the world with the discovery and settlement of the New World and Oceania, directly influenced by trades among countries.

Keywords: *Mycobacterium bovis,* animal tuberculosis, phylogenetics, *Mycobacterium tuberculosis* Complex, evolution.

Introduction

Tuberculosis (TB) is the leading infectious killer in the world and approximately 10 million new cases are reported annually. In 2017, 1.6 million people died of TB and over 95% of these deaths occurred in low and middle-income countries (World Health Organization 2018). The disease is strongly linked to poverty, with its prevalence following a socioeconomic gradient within and among countries. In addition, there is a significant, and often neglected, contributor to the global disease burden, which is the zoonotic transmission of bovine TB to humans. The WHO (World Health Organization) estimated that 142,000 new cases and 12,500 deaths occurred due to zoonotic TB in 2017 (World Health Organization 2018), numbers that are likely underestimated due to lack of routine surveillance data from most countries (Olea-Popelka et al. 2017). People with zoonotic TB face arduous challenges; the etiologic agent is intrinsically resistant to pyrazinamide, one of the first-line drugs used in TB treatment, and a possible association with extra-pulmonary disease often delays diagnostics and treatment initiation. In addition, bovine TB results in severe economic losses for livestock producers worldwide, respecting no borders and repeatedly affecting animal conservation efforts due to the establishment of wildlife reservoirs or spillover events from cattle to associated animal populations. In order to eradicate TB by 2,030 as part of the United Nations (UN) Sustainable Development Goals, it is imperative that future prevention and control strategies focus on all forms of TB in humans, including its interface with animals.

Human and animal TB are caused by members of the *Mycobacterium tuberculosis* Complex (MTC). The MTC is a clonal bacterial group composed of 12 species or ecotypes with variable virulence and host tropism (Galagan 2014). *Mycobacterium tuberculosis stricto sensu* is the main responsible for the TB numbers and is highly adapted to human hosts. On the other hand, *Mycobacterium bovis*, the causative agent of bovine TB, has a broader host range and is able to infect and cause disease in multiple host species, including humans, with variable populational persistence (Malone and Gordon 2017). MTC

77 members have clonally evolved from a common ancestor with the tuberculous bacteria *Mycobacterium*
78 *canetii*, and alignable regions of MTC genomes are over 99.95% identical, with horizontal gene transfer
79 and large recombination events considered absent. These pathogens have solely evolved through single
80 nucleotide polymorphisms (SNPs) and small deletions of up to 12.7Kb, which translated into a phenotypic
81 array of host tropism and virulence variations (Galagan 2014; Brites et al. 2018).
82 Using whole-genome, SNP-based phylogenetic analyses, *Mycobacterium tuberculosis* strains have
83 been classified into 5 lineages: L1 through L4, and L7. *Mycobacterium africanum*, which is another
84 bacterial species of MTC, accounts for lineages 5 and 6 of the human-adapted MTC strains and is
85 responsible for about 1/3 of TB cases in West Africa (de Jong et al. 2010). Each *M. tuberculosis* lineage
86 is associated with specific global geographical locations and lineage-associated variations in virulence,
87 transmission capacity and in the propensity to acquire drug resistance have been reported (Coscolla and
88 Gagneux 2014). Thus, regional prevalence of specific lineages or sub-lineages have consequences for
89 the epidemiology of TB worldwide. A similar attempt to classify *M. bovis* into different genetic groups was
90 made prior to the large-scale availability of whole-genome sequences and started with the identification of
91 Clonal Complexes (CC). Accordingly, four *M. bovis* CCs have been described (African 1 and 2, European
92 1 and 2), and these are determined based on specific deletions ranging from 806 to 14,094 bp (base
93 pairs), SNPs and spoligotypes. As with *M. tuberculosis* lineages, *M. bovis* CCs appear to have distinct
94 geographical distributions, with African 1 and 2 restricted to Africa, CC European 2 commonly found in
95 the Iberian Peninsula, and CC European 1 distributed globally (Müller et al. 2009; Berg et al. 2011; Smith
96 et al. 2011; Rodriguez-Campos et al. 2012; Smith 2012). Although there are no studies specifically aimed
97 at identifying differences in virulence patterns among *M. bovis* of different CCs, numerous articles report
98 virulence variations among strains of *M. bovis* (Wedlock et al. 1999; Waters et al. 2006; Meikle et al.
99 2011; Wright et al. 2013; Waters et al. 2014; de la Fuente et al. 2015; Vargas-Romero et al. 2016),
100 suggesting a possible link between bacterial genetic polymorphisms and disease development, as
101 observed in *M. tuberculosis*.
102 Since the whole-genome sequence of the first *M. bovis* strain became available in 2003 (Garnier et al.
103 2003), increasing efforts have been made to sequence additional strains and use whole-genome
104 information to tackle bovine and/or wildlife TB transmission within specific outbreaks or countries
105 (Bruning-Fann et al. 2017; Sandoval-Azuara et al. 2017; Ghebremariam et al. 2018; Kohl et al. 2018;
106 Lasserre et al. 2018; Orloski et al. 2018; Price-Carter et al. 2018; Razo et al. 2018). However, no studies
107 to date have comprehensively analyzed *M. bovis* genomes at a global scale. Few studies that have
108 compared transboundary *M. bovis* strains analyzed bacterial isolates obtained from a reduced number of
109 countries (n<9) and included small sample sizes (Dippenaar et al. 2017; Patané et al. 2017; Zimpel et al.
110 2017a; Ghebremariam et al. 2018; Lasserre et al. 2018). Nevertheless, attained results suggest that *M.*
111 *bovis* strains are likely to cluster based on geographical location (Dippenaar et al. 2017; Zimpel et al.
112 2017a; Lasserre et al. 2018). In our previous study, we have also shown that few *M. bovis* genomes do

113 not carry any CC genetic marker (Zimpel et al. 2017a), a phenomenon that was recently observed in *M.*
114 *bovis* isolates from one cattle herd in the USA and from slaughterhouse cattle in Eritrea (Ghebremariam
115 et al. 2018; Orloski et al. 2018). These findings suggest that Clonal Complexes are unlikely to represent
116 the whole diversity of *M. bovis* strains, warranting further evaluation of *M. bovis* molecular lineages
117 (Zimpel et al. 2017a; Lasserre et al. 2018).
118 In this study, we propose and discuss the existence of at least three distinct lineages of *M. bovis* in the
119 world. We have screened over 2,600 publicly available *M. bovis* genomes and newly sequenced two
120 wildlife *M. bovis* strains, gathering 823 *M. bovis* genomes from 21 countries and 21 different host-species,
121 including humans, to complete a phylogenomic analyses. We also used a previously described approach
122 (Bos et al. 2014) to evaluate the evolutionary origin of *M. bovis* strains and lineages and correlated
123 bacterial population dynamics with historical events to gain new insights into the widespread nature of
124 bovine TB worldwide.
125
126 *Results and Discussion*
127 ***Mycobacterium bovis* genomes and host adaptation**
128 After screening ~2,600 publicly available *M. bovis* genomes using pre-determined inclusion criteria
129 (see methods), a total of 823 virulent *M. bovis* genomes from 21 countries and 21 different host species
130 were selected for this study (table 1). This sample constitutes the most comprehensive global *M. bovis*
131 dataset ever analyzed, and includes 724 genomes published previously, and two newly sequenced
132 genomes from this study (supplementary table S1). We used 17,302 nucleotide polymorphic positions
133 detected in these genomes to construct phylogenetic trees inferred with maximum likelihood (ML) and
134 maximum parsimony (MP) methodologies (fig. 1; supplementary files S1 and S2). Both phylogenetic
135 methods employed (MP and ML) generated highly similar trees (supplementary figs. S1 and S2). This
136 observation is supported by the Robinson-Foulds distance between both trees (= 678) compared to all
137 other comparisons among 1,000 random trees with the same number of taxa (for a total of 499,500
138 distances obtained), for which the minimum was 1,632, and the maximum was 1,642 (supplementary fig.
139 S3). Therefore, results are shown for ML only (figs. 1, 2, and 3).
140
141 **Table 1.** *Mycobacterium bovis* genomes selected for analysis.

Country	Number of genomes	Host species (# of selected read sets)
Brazil	3	Cattle (1), Bison (1), Llama (1)
Canada	6	Cattle (5), Elk (1)
China	2	Chimpanzee (2)
Eritreia	12	Cattle (12)
Ethiopia	2	Cattle (2)
France	2	Cattle (1), Wild boar (1)
Germany	6	Human (6)

Ghana	3	Human (3)
Great Britain	12	Human (11), Cattle (1)
Italy	2	Human (2)
Mexico	192	Cattle (175), Human (17)
New Zealand	232	Cattle (129), Ferret (38), Pig (7), Cervid (4), Stout (1), Possum (53)
Northern Ireland	88	Cattle (84), Badger (4)
Panama	2	Cattle (2)
Russia	1	Human (1)
South Africa	11	Cattle (1), Lion (2), Kudu (1), Buffalo (7)
Spain	7	Cattle (7)
Switzerland	2	Unknown host*
Tunisia	1	Human (1)
United States	233	Cattle (124), Coyote (1), Human (2), Cervid (64), Cat (7), Raccoon (12), Wild boar (6), Opossum (9), Jaguar (1), Bobcat (1), Bison (1), Elk (5)
Uruguay	4	Cattle (4)
Total	823	

**Mycobacterium bovis* genomes sequenced by Brites et al. (Brites et al. 2018).

In the generated phylogenetic tree, host species are found dispersed among different clades, while most *M. bovis* genomes cluster according to specific geographic locations (fig. 1). This finding suggests that geographic proximity between hosts and their contact rates has played a more important role in determining host range of *M. bovis* than phylogenetic distance among hosts. When considering the over-represented countries USA and New Zealand, the most basal *M. bovis* strains are of bovine origin, suggesting that the introduction of bovine TB in these countries occurred through cattle, spilling over to wildlife in subsequent years. The lack of host clustering supports the hypothesis of *M. bovis* being a generalist member of MTC, able to infect different host species irrespective of the bacterial genetic makeup. Interestingly, all *M. bovis* genomes originating from the African Continent, except for South Africa, appeared as emerging from relatively older nodes of the phylogenetic tree, suggesting the continent is a possible center of dispersion.

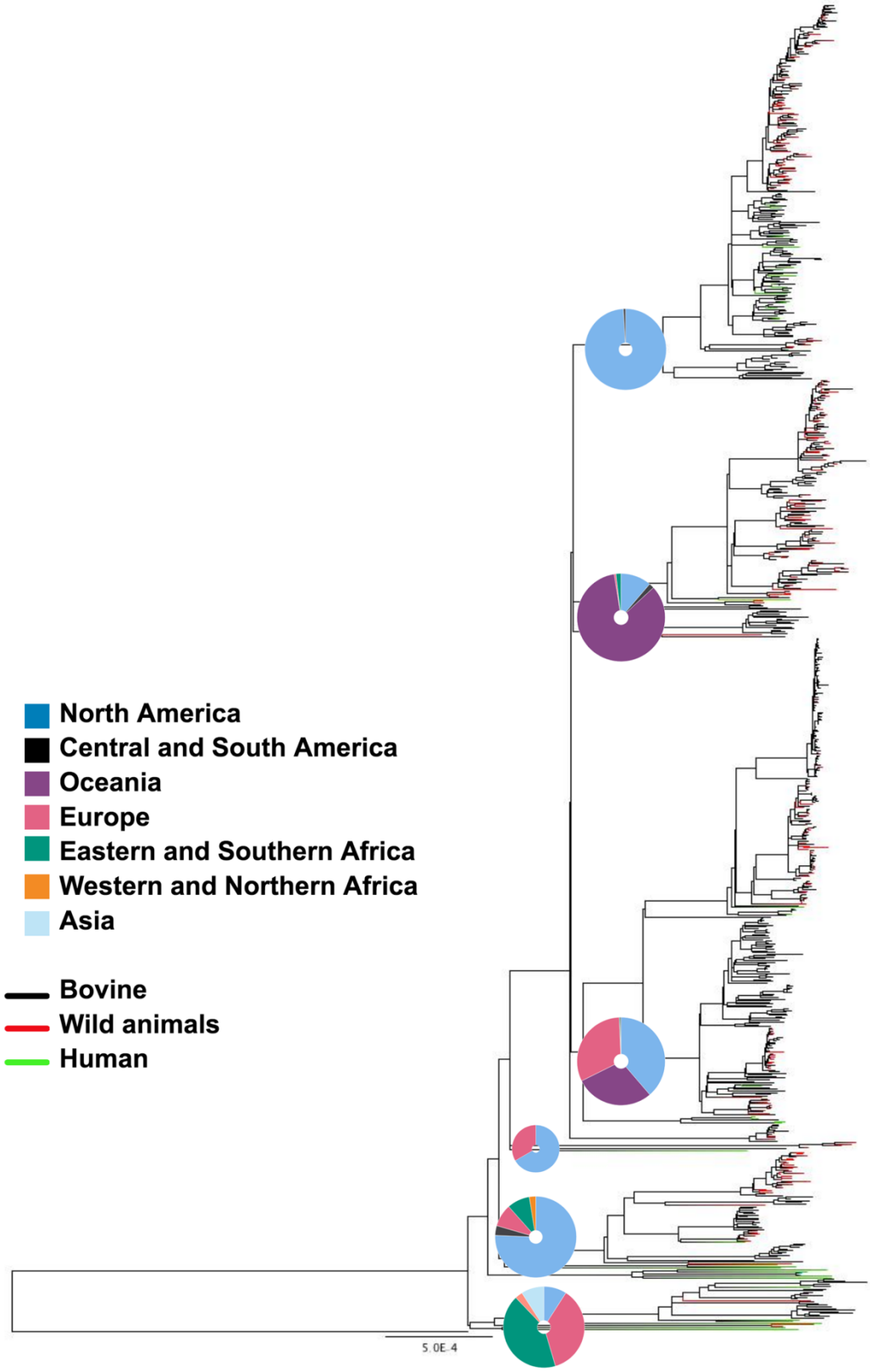


Fig. 1. Host species and geographical distribution of *Mycobacterium bovis* isolates. Maximum likelihood phylogenetic tree based on single nucleotide polymorphisms (SNPs) of 823 *Mycobacterium bovis* genomes, using *Mycobacterium tuberculosis* H37Rv as outgroup. The host species are marked in black tips for bovine, red tips for wild animals, green tips for humans, and in blue tips are two

genomes with unknown host identification. Geographical locations of the isolates are colored in the donut chart (which do not represent ancestral area reconstruction at nodes): blue - North America, black - South and Central America, purple - Oceania, pink - Europe, green - Eastern and Southern Africa, orange - Western and Northern Africa, light blue - Asia. Phylogenetic tree was generated using IQ-Tree with 1,000 bootstrap values and annotated using FigTree v1.4.3 and Mind the Graph. Bootstrap values can be visualized in supplementary fig. S1.

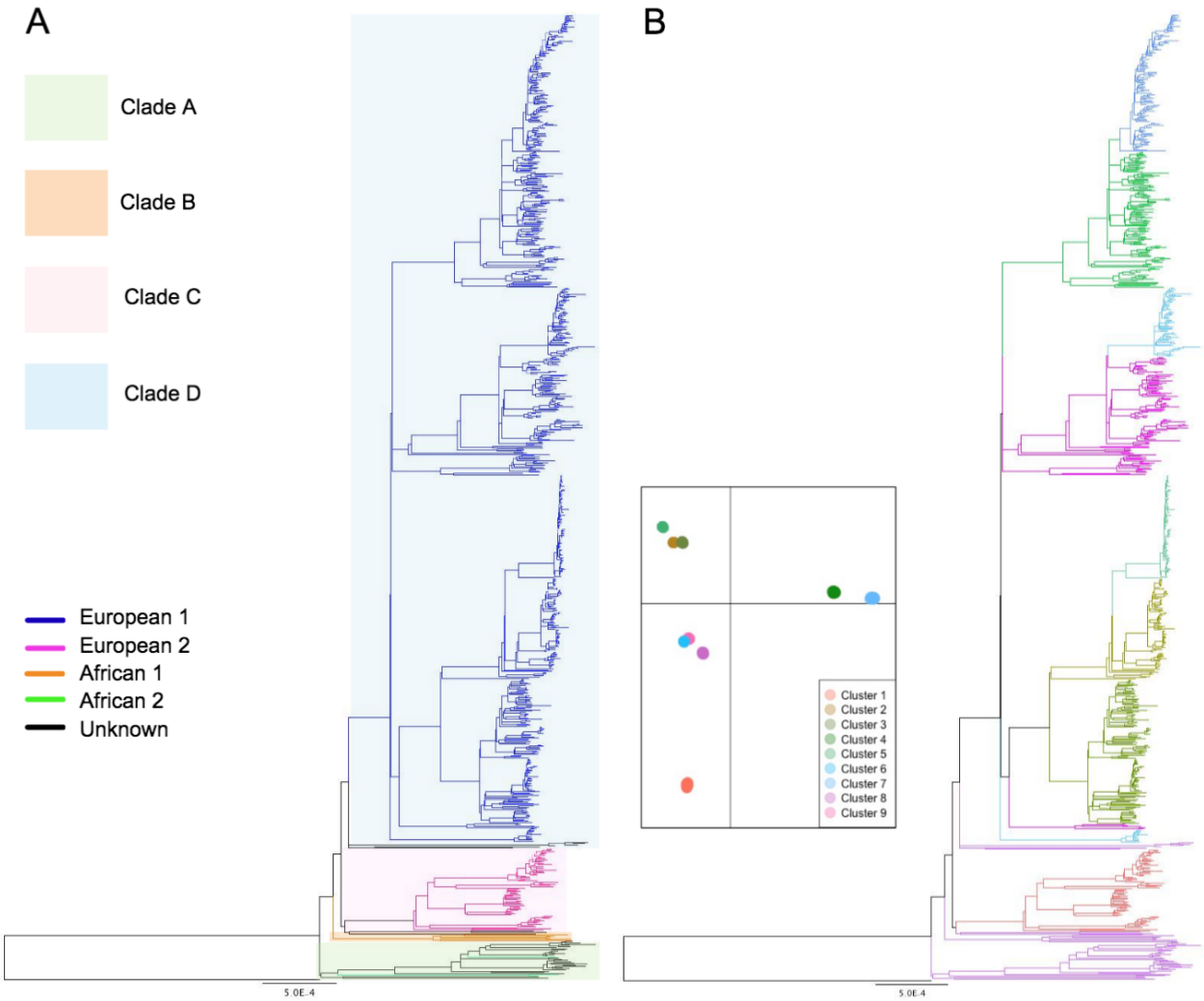
Clonal Complexes do not represent the whole diversity of *M. bovis* genomes

Thirty-eight out of the 823 (4.6%) analyzed *M. bovis* genomes do not have genetic markers of any of the four CCs (Eur1 - European 1, Eur2 - European 2, Afr1 - African 1, Afr2 - African 2). The remaining *M. bovis* isolates were found to cluster according to their CC classification. The vast majority (29/38, 76.32%) *of M. bovis* genomes not classified within CCs appeared dispersed, emerging from relatively ancestral nodes (fig. 2). This finding suggests that isolates clustering in ancient nodes were probably not included in prior studies evaluating CCs, and have thus been poorly studied until now.

Clades A and B

The clade A (fig. 2a, clade A - green box) is composed of 32 *M. bovis* isolates. These are from Africa [Eritrea (*n*=12), Ethiopia (*n*=2), Tunisia (*n*=1)], Europe [Italy (*n*=2), France (*n*=1), Spain (*n*=7), Germany (*n*=1)], Asia [(Russia (*n*=1), China (*n*=2), and the USA (*n*=3). Among these *M. bovis* isolates, only three were characterized with the deletion marker of the CC African 2 (one isolated from wild boar in France, and two from chimpanzees in China; green tips in fig. 2a). The remaining strains (*n*=29) did not show any of the CC specific markers. Clade B (fig. 2a, clade B - orange box) is composed of 6 *M. bovis* isolates, including genomes from Africa [Ghana (*n*=3)] and Europe [Germany (*n*=1), Switzerland (*n*=2)] carrying the deletion marker of the CC African 1 (orange tips in fig. 2a). To date, *M. bovis* strains of the CCs Af1 and Af2 have only been described in Africa (Berg et al. 2011; Firdessa et al. 2013). The finding of *M. bovis* strains of CC Af2 infecting chimpanzees in China may be explained by pathogen importation from the African continent into that country or, less likely, spill-over from infected livestock to captive animals. A thorough characterization of *M. bovis* strains from China in the future could elucidate which CC is more frequent in that country. On the other hand, the presence of both African CCs in countries of Europe, infecting a wild boar (France), humans (Germany) and unknown hosts (Switzerland), is interesting and warrants further investigation into the actual origin of these isolates. Ten out of 43 (23.26%) *M. bovis* isolates from humans are found in clades A and B, four from Africa (Ghana=3; Tunisia=1), four from Europe (Germany=2; Italy=2), one from Russia, and one from the USA. Unfortunately, demographic characteristics of these patients are not described in the related literature (Casali et al. 2014; Walker et al. 2015; Brites et al. 2018; Kandler et al. 2018). With current bovine TB control, the zoonotic transmission of *M. bovis* in Europe and the USA is considered rare. Thus, we speculate that European and North American cases in humans may have arisen from zoonotic transmission acquired in the past that appears

199 in elderly people (Majoor et al. 2011; Davidson et al. 2017) and/or imported human cases of zoonotic TB
200 from countries where bovine TB is highly endemic.
201
202 **Clade C and the CC Eur2 marker**
203 The clade C (fig. 2a, clade C - pink box) is composed of 73 *M. bovis* isolates, including three genomes
204 from Germany without CC markers and 70 genomes from the Americas [Brazil (*n*=3), Canada (*n*=2),
205 Mexico (*n*=17), USA (*n*=40)], Africa [South Africa (*n*=7)], and Europe [Germany (*n*=1)] with the SNP
206 marker of the CC European 2. The CC European 2 has been described as dominant in the Iberian
207 Peninsula and also detected in other countries of Europe and in Brazil (Berg et al. 2011; Zimpel et al.
208 2017a). Its actual worldwide occurrence, however, is unknown.
209



210
211 **Fig. 2. Clonal Complexes and DAPC characterization.** Maximum likelihood phylogenetic trees
212 based on single nucleotide polymorphisms (SNPs) of 823 *Mycobacterium bovis* genomes, using
213 *Mycobacterium tuberculosis* H37Rv as outgroup. **(a)** Phylogenetic tree colored based on clades and
214 clonal complexes. Green box: clade A; orange box: clade B; pink box: clade C; blue box: clade D.
215 Clonal Complexes (tips): Eur1 in blue, Eur2 in pink, Af1 in orange, and Af2 in green. Genomes without

identification of the Clonal Complexes markers are shown in black. **(b)** Phylogenetic tree colored based on Discriminant Analysis of Principal Components (DAPC), revealing the clustering among *Mycobacterium bovis* genomes regarding the first two principal components. Phylogenetic tree was generated using IQ-Tree with 1,000 bootstrap values and annotated using FigTree v1.4.3 and Mind the Graph. Bootstrap of discussed nodes are all ≥ 95% and can be visualized in supplementary fig. S1 and supplementary file S1. Bar shows substitutions per nucleotide.

Clade D and the CC Eur1 marker

The clade D (fig. 2a, clade D - blue box) is composed of 712 *M. bovis* isolates, comprising a small group of genomes without CC markers containing isolates from France (*n*=1), Great Britain (*n*=1), and the USA (*n*=4), and a paraphyletic cluster composed of 706 *M. bovis* isolates originating from the Americas [USA (*n*=186), Mexico (*n*=175), Uruguay (*n*=4), Canada (*n*=4), Panama (*n*=2)], New Zealand (*n*=88), Great Britain (*n*=11), South Africa (*n*=4), and Northern Ireland (*n*=232). All these 706 members have the deletion RDEu1, being classified as CC European 1. Thus, the most widespread CC of *M. bovis* (Smith et al. 2011) forms the most recent evolutionary clade among *M. bovis* strains.

A Proposal for Global Lineages of *M. bovis*

In order to evaluate the genetic relationship among *M. bovis* strains, we used the DAPC (Discriminant Analysis of Principal Components) population genetic algorithm to detect genome clustering. It revealed the existence of nine clusters among *M. bovis* genomes, arranged into four groups (fig. 2b). Interestingly, genomes with more ancestral branching at different clades (C, D, and subclusters of D) share genomic similarities with genomes emerging from the basal node of the phylogenetic tree (clade A and B) (fig. 2b in purple and supplementary fig. S4). This finding indicates the occurrence of successive dispersals followed by clonal expansion in certain geographic locations of the globe throughout the history of *M. bovis*. In other words, specific bacterial clones of the relatively ancestral lineages gave origin to these clades, which then clonally expanded in geographic isolation. This finding agrees with cattle trade and export patterns from Europe and Africa into the New World, Oceania, and Northern Ireland, which eventually led to the introduction of bovine TB in these regions.

Based on the observed tree topology, CC distribution, geography, and DAPC analysis, we propose the existence of at least three major global lineages of *M. bovis*, which we define as Lb1, Lb2 and Lb3 (fig. 3; table 2). Lineage Lb1, emerging from the relatively most basal node, is composed of 38 *M. bovis* isolates, encompassing Clades A and B (fig. 3 - green and orange boxes). The observed geographical origin of extant Lb1 isolates points towards North and East Africa and Europe, although it can also be detected, at a lower frequency, in the United States (3 Lb1 isolates out of 233 *M. bovis* isolates from the USA). Lb2 lineage (fig. 3 - in pink) is composed of the 70 *M. bovis* isolates from Clade C that carry the Eur2 genetic marker. With present data it is not possible to determine whether the remaining genomes of clade C isolated from humans in Germany are also part of Lb2. On the other hand, lineage Lb3 (fig. 3 - in blue) is

253 composed of 706 genomes of Clade D carrying the Eur1 genetic marker. It is not possible to determine
254 whether the remaining six genomes without a CC marker (from France, Great Britain, and the USA) are
255 part of Lb3 either. We believe Eur2 and Eur1 are stable markers for Lb2 and Lb3, respectively.
256
257 **Table 2.** Distribution of *Mycobacterium bovis* lineages in the studied countries.

Country	Lb1	Lb2	Lb3.1	Lb3.2
Eritreia	12	-	-	-
Ethiopia	2	-	-	-
Ghana	3	-	-	-
Tunisia	1	-	-	-
France	1	-	-	-
Germany	2	1	-	-
Spain	7	-	-	-
Switzerland	2	-	-	-
Italy	2	-	-	-
China	2	-	-	-
Russia	1	-	-	-
United States	3	40	46	140
Mexico	-	17	67	108
Canada	-	2	3	1
Brazil	-	3	-	-
South Africa	-	7	1	3
Northern Ireland	-	-	88	-
Great Britain	-	-	10	1
New Zealand	-	-	97	135
Panama	-	-	-	2
Uruguay	-	-	1	3
Total *	38	70	313	393

258 Lb - Lineage of *Mycobacterium bovis*. *From 823 studied genomes, 9 were not included in the referred
259 lineages.
260

261 The Lb3 isolates are arranged in two major groups, suggesting the possibility of two sublineages,
262 Lb3.1 and Lb3.2 (table 2). Sublineage Lb3.1 is composed by one cluster with genetic resemblance
263 (based on DAPC) to the ancestral Lb1 containing genomes of New Zealand (*n*=11), Great Britain (*n*=4),
264 Canada (*n*=1), followed by another cluster composed of isolates from Northern Ireland (*n*=88), New
265 Zealand (*n*=86), Mexico (*n*=67), USA (*n*=46), Great Britain (*n*=6), Canada (*n*=2), Uruguay (*n*=1), and
266 South Africa (*n*=1). Sublineage Lb3.2 (table 2) also contains genomes related to ancestral Lb1 (based on
267 DAPC), represented by New Zealand (*n*=135), Mexico (*n*=13), USA (*n*=4), Uruguay (*n*=2), South Africa
268 (*n*=3), Great Britain (*n*=1), Canada (*n*=1), and Panama (*n*=1), and a more derived population,
269 characterized by genomes from the USA (*n*=136), Mexico (*n*=95), Panama (*n*=1), and Uruguay (*n*=1) (fig.
270 3).
271

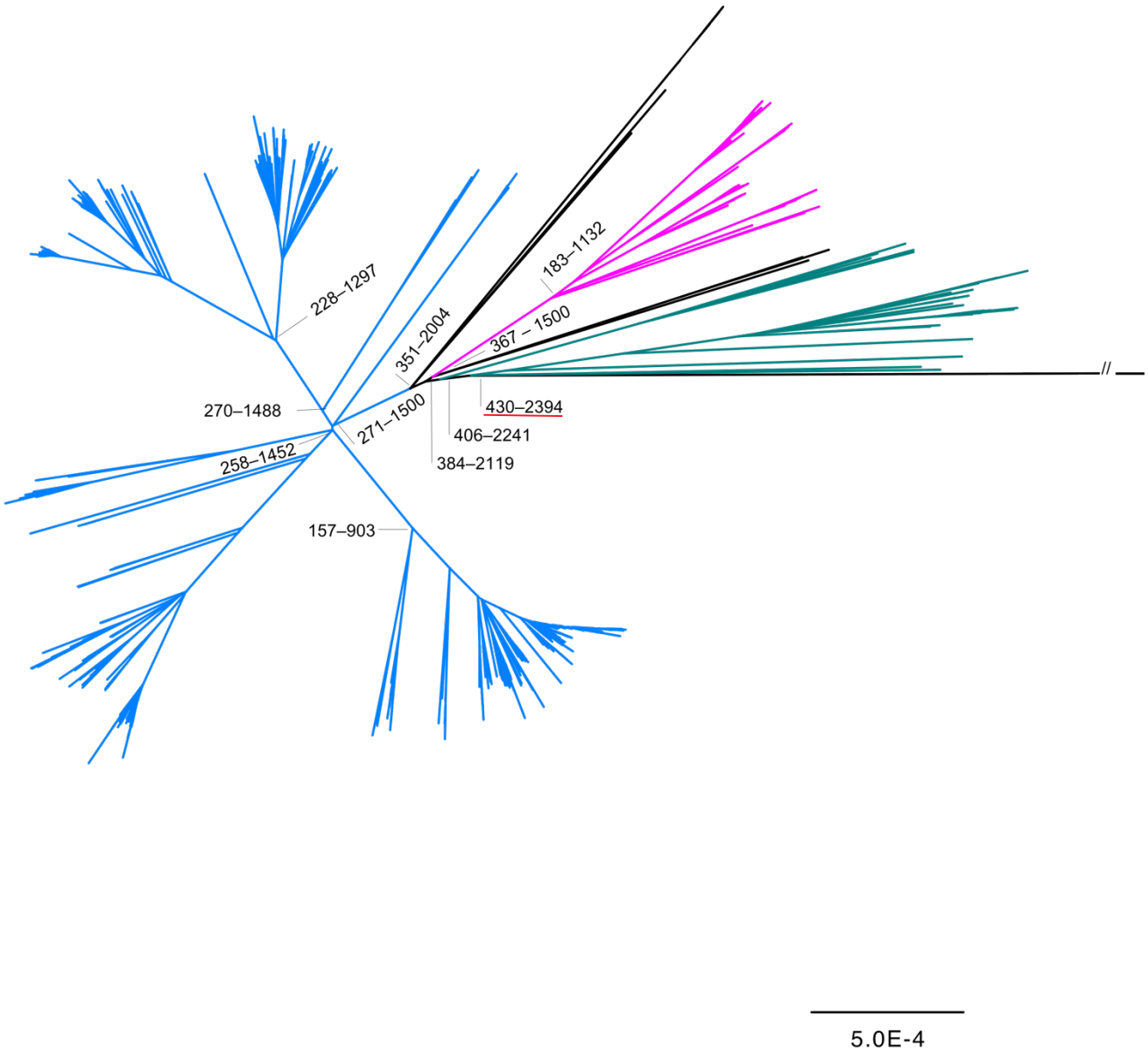


Fig. 3. *Mycobacterium bovis* lineages and evolutionary data. Maximum likelihood phylogenetic tree (radial layout) of 823 *M. bovis* genomes colored based on the proposed Lineages: Lb1 - green; Lb2 - pink; Lb3 - blue. Estimated divergence results in years BP (Before Present) of the main branches. Outgroup: *Mycobacterium tuberculosis* H37Rv. Phylogenetic tree was generated using IQ-Tree with 1,000 bootstrap values and annotated using FigTree v1.4.3. Bootstrap values can be visualized in supplementary fig. S1. Bar shows substitutions per nucleotide.

Spoligotyping patterns are correlated with *M. bovis* lineages

To further support our findings related to the global *M. bovis* lineages, we also evaluated the spoligotype of all isolates. There was an association between spoligotype and the *M. bovis* lineages (table 3). Most of the patterns were specific to the predicted lineages, demonstrating that spoligotyping can be an additional tool to infer these groups.

286 **Table 3.** Spoligotype patterns of *Mycobacterium bovis* strains separated by lineages.

Spoligotype pattern	Lb1	Lb2	Lb3.1	Lb3.2
Unknown	2	-	4	7
SB2467	2	-	-	-
SB0134	3	-	-	-
SB0120	17	-	-	1
SB0948	1	-	-	-
SB1517	2	-	-	-
SB0828	5	-	-	-
SB0288	1	-	-	-
SB0944	4	-	-	-
SB1025	1	-	-	-
SB0121	-	30	-	-
SB0295	-	1	-	-
SB0265	-	28	-	1
SB0339	-	1	-	-
SB1345	-	9	28	-
SB2469	-	1	-	-
SB1031	-	-	11	-
SB0267	-	-	4	-
SB1071	-	-	1	-
SB0140	-	-	186	-
SB0263	-	-	4	-
SB1072	-	-	1	-
SB0673	-	-	60	-
SB0307	-	-	1	-
SB1750	-	-	1	-
SB1751	-	-	1	-
SB1502	-	-	1	-
SB1499	-	-	1	-
SB0986	-	-	2	-
SB1348	-	-	1	-
SB0971	-	-	2	-
SB0272	-	-	1	-
SB0980	-	-	2	-
SB0145	-	-	1	201
SB1054	-	-	-	1
SB0130	-	-	-	79
SB0327	-	-	-	11
SB1216	-	-	-	1
SB1758	-	-	-	1
SB1504	-	-	-	56
SB1446	-	-	-	1
SB2011	-	-	-	3
SB1040	-	-	-	27
SB0162	-	-	-	3
Total	38	70	313	393

287 Lb - lineages. Clade C contains 3 genomes (from Germany) not classified as Lb2: two have unknown
288 spoligotype pattern, and one was classified as SB0871. Clade D contains six genomes that were not
289 grouped into Lb3 (without Eur1 marker): four have spoligotype pattern SB1069, one SB0821, and one
290 SB0134.
291

292 **Evolutionary scenario for the origin of *M. bovis***

293 Recent analyses point towards the evolutionary rate defined by Bos et al. (Bos et al. 2014) as being
294 the most plausible to explain the trajectory of MTC (O'Neill et al. 2017; Menardo et al. 2019). Our
295 posterior Bayesian estimates indicates that the time of the most recent common ancestor (MRCA) for
296 extant *M. bovis* lineages is 430 - 2,394 years BP (Before Present) (fig. 3). This maximum time obtained
297 (2,394 years BP) (fig. 3) agrees with the archeological finding of *M. bovis* (defined by the loss of RD4)
298 causing infection in four human skeletons from the Iron Age in South Siberia, with carbon-dating placing
299 them within the 1,761 - 2,199 years BP range (Taylor et al. 2007). Interestingly, these ancient *M. bovis*
300 strains do not have the deletion marker of CC Eur1 (i.e. RD17), serving as supportive evidence that Lb3
301 is indeed a recently-evolved lineage of *M. bovis*. The oldest description of tubercle ('phymata') lesions in
302 cattle (and also in sheep and pigs) was done by the Greek physician Hippocrates [460-370 BCE (before
303 common era)] (Herzog, Basel 2003), also overlapping our maximum estimate for the MRCA.

304 Considering the upper-end time of 2,394 years BP, the predicted origin of the extant *M. bovis* lineages
305 overlaps with the Ancient Roman period [625-510 BCE to 476 CE (common era)]. Moreover, as most
306 extant Lb1 strains are from Europe and Africa, we speculate that *M. bovis* origin may have had an
307 important contribution from the Roman Empire, possibly due to the Empire's influence on animal
308 husbandry practices in the Mediterranean. The world saw a demographical explosion and a deep
309 connection between different cultures and communities during the Roman Empire (Jongman 2012;
310 Hanson et al. 2017). The economical flourishing was followed by a steep rise in agriculture and livestock
311 capacity to feed the Empire's army and ordinary citizens (Jongman 2012). Archeological evidence show
312 that the domestication of cattle in neolithic times initiated a process of animal body size decrease
313 (Meadow 1988, 1989; Meadow and Zeder 2004). Surprisingly, this trend was reversed during the end of
314 the Iron Age and later intensified during the Roman period, as described by an incremental change in
315 body size of cattle (Trentacoste et al. 2018). These findings imply changes in cattle herding management,
316 including improved feeding practices, probably aimed at increasing production capacity for draft, meat
317 and/or milk (Kron 2002). Reasons for breeding cattle varied considerably throughout the Roman territory
318 and beyond due to culture traditions, diet preferences and climate conditions (Valenzuela-Lamas and
319 Albarella 2017). For instance, it is believed that Romans preferred goat and sheep milk for cheese
320 production, and cattle as a draft animal. Communities from Central and Northern Europe, in contrast,
321 used cattle for dairying, and drank fresh milk in significant amounts (Itan et al. 2009). Nevertheless, by
322 intensifying production systems and animal trade, which was most likely not centered solely in Roman
323 territory, bovine TB may have dispersed significantly.

324 It has been suggested that the demographic explosion and economic rise of the Roman Empire
325 greatly contributed to the spread of human TB worldwide, with naïve human populations from rural areas
326 being exposed to the disease when moving to larger cities (Herzog, Basel 2003; O'Neill et al. 2017).
327 Archeological findings of human skeletons with lesions compatible with TB also suggest a possible shift in

328 bone disease presentation during the Roman Empire (Holloway et al. 2011). In the Mediterranean area,
329 before 600-800 BCE, over 90% of the skeleton lesions were spinal (a.k.a. Pott's Disease) (Holloway et al.
330 2011). From 600 BCE to 1600 CE, the number of lesions localized in other skeletal locations started to
331 increase, reaching over 60% of the lesions, mostly combined with spinal lesions as well (Holloway et al.
332 2011). Diagnosis of TB cases in palaeopathological remains is difficult given the lack of specificity in bone
333 lesions. Nevertheless, this finding suggests an important change in TB presentation that may have been
334 influenced by an increased disease burden as well as by changes in etiological agents and form of
335 transmission with the appearance of *M. bovis* during the Roman Empire. It is thus possible that the
336 zoonotic transmission of TB from cattle to humans through ingestion of contaminated dairy products,
337 meat and close contact may have contributed to the spread of TB in humans during this period.
338 We cannot assume for certain that bovine TB exclusively originated during the Roman Empire. The
339 disease may have been present in cattle before this period, although caused by a different ancestor of the
340 extant *M. bovis* lineages. It is still unclear if the ancestor of MTC was a specific human pathogen or a
341 generalist microorganism able to infect multiple host species (Brites et al. 2018). In the latter case, the
342 intensification of the livestock system and increase in animal and human population density may have
343 selected for pathogens either more adapted to animals (*M. bovis*) or more adapted to human beings (*M.*
344 *tuberculosis*) over evolutionary time. It should be noted that genomes from representative *M. bovis*
345 isolates from Eastern and Western Asia have yet to be sequenced and analyzed. The lack of information
346 from that continent makes it impossible to rule out an Asian origin for *M. bovis* lineages, especially
347 considering the finding of an ancient human infection near present Mongolian border in Russia (Taylor et
348 al. 2007). Alternatively, the well-known commercial trade between the Roman Empire and Asia may also
349 account for this dispersal.
350
351 **Bovine TB after the 16th century**
352 Our dating estimates and evolutionary predictions reveal a complex relationship between spatial
353 dispersal and expansion of *M. bovis*. Bovine TB worldwide distribution is certainly influenced by import
354 and export of cattle breeds over time. Non-European countries imported specialized breed cattle from
355 single sources and then exported to other countries, while others imported animals from multiple
356 locations.
357 Estimates show that the MRCAs of extant *M. bovis* isolates occurring in Northern Ireland date
358 between 54 to 325 ya (supplementary file S3). Archeological data from cattle specimens found in the
359 UK (United Kingdom) indicates that these animals had a shared ancestry with wild British aurochs,
360 being unrelated to the modern European taurine (Orlando 2015). Accordingly, in Northern Ireland,
361 cattle have been the mainstay of farming since the neolithic period (about 6,000 ya). Therefore, despite
362 the ancient presence of cattle in these regions, the introduction of the extant *M. bovis* lineage Lb3 seems
363 to be a more recent event, possibly due to the importation of taurine cattle from other areas of Europe.

364 Most of the USA isolates clustered in Lb3 (186/233), a few in Lb2 (40/233) and only three in Lb1
365 (3/233). A similar trend is observed in Mexico, with the majority of the isolates clustered into Lb3
366 (175/192) and fewer representatives in Lb2 (17/192). Our dating estimates are consistent with bovine TB
367 introduction into these countries during the New World colonization period (supplementary file S3), with
368 Lb2 dating from 137 to 906 ya (USA) and from 41 to 241 ya (Mexico), and Lb3 dating from 71 to 411
369 ya (USA) and 216 to 1151 ya (Mexico). The first cattle to be introduced into the USA came from the
370 Iberian Peninsula in the 16th century. This introduction was followed by cattle importation from Mexico
371 (which also received animals from the Iberian Peninsula) and from England (Martínez et al. 2012). This is
372 consistent with the introduction of Lb2 (CC Eur 2) and Lb3 (CC Eur 1), common lineages in the Iberian
373 Peninsula and the UK, respectively. Isolates from USA and Mexico frequently clustered together,
374 consistent with the close cattle trade relationship between these two countries throughout modern history.
375 Despite the success of the eradication program of bovine TB in the USA, outbreaks in wild animals are
376 frequently reported, such as in white-tailed deer, considered a maintenance host, increasing the risk of
377 disease spread to other animals and humans (Cross et al. 2018). Notwithstanding, prevalence of
378 zoonotic TB in the USA is considered low, ranging from 1.4% to 6% of human TB cases, and having
379 some correlation with Hispanic immigrants due the consumption of contaminated dairy products in
380 their endemic countries (Rodwell et al. 2008; Müller et al. 2013). In Mexico, rates of zoonotic TB are
381 estimated to be around 7%, but in some regions can reach nearly 30% of human TB cases (Müller et
382 al. 2013), suggesting the need for better bovine TB control.
383 At least two main introductions of Lb3 may have happened in New Zealand. The first is estimated to
384 have occurred between 92 and 488 ya (in Lb3.1), while the second from 60 to 333 ya (in Lb3.2)
385 (supplementary file S3). Breeds of cattle were introduced into New Zealand in the 19th century
386 (Livingstone et al. 2015), which is within our time period range and suggests consequent spreading to
387 wildlife. In contrast, *M. bovis* genomes from South Africa appeared clustered into Lb2 and Lb3 (Smith et
388 al. 2011). Despite the small sample size, they are not present in the most ancient lineage, Lb1. In fact,
389 previous genotyping studies have shown that the most common *M. bovis* CC in South Africa is Eur1, i.e.
390 Lb3. Thus, in agreement with previous hypotheses (Michel et al. 2009; Smith et al. 2011), it is likely that
391 bovine TB was introduced into South Africa following European colonization, and subsequently spilled
392 over to naïve wildlife, with destructive consequences (Michel et al. 2006), as exemplified by *M. bovis*
393 genomes from kudus, lions and buffalos analyzed herein.
394
395 *Concluding remarks*
396 We propose the existence of at least three global lineages of *M. bovis*, named Lb1, occurring mostly in
397 Africa and Europe, Lb2 (with CC Eur2 marker), present mainly in the Americas and South Africa, and Lb3
398 (with CC Eur1 marker), dispersed worldwide. Because these lineages tend to cluster based on
399 geographical location rather than host species, it reinforces the idea that bovine TB eradication will only

400 be attained once the disease is controlled in wildlife and vice-versa. The observed lack of host specificity
401 supports the hypothesis of *M. bovis* being a generalist member of MTC. Nonetheless, *M. bovis* ability to
402 be transmitted among cattle is the main reason why this pathogen has spread geographically over
403 evolutionary time, because of animal trade. The lack of stable CC markers in the most ancient lineage,
404 Lb1, indicates that these pathogens have been the least studied, and that there is an urgent need for
405 additional evaluations of *M. bovis* isolates from Asia and Africa.
406 Our results delineate independent evolutionary trajectories of bacterial subpopulations (i.e. lineages) of
407 *M. bovis* underlying the current disease distribution. Whether or not these events are associated with
408 further specialization of *M. bovis* to the bovine species and breeds or increased/decreased virulence of
409 this pathogen in domesticated and/or wildlife have yet to be determined. Lineages of *M. tuberculosis* are
410 known for their virulence variations, warranting further similar studies regarding *M. bovis* lineages.
411 Finally, our dating estimates for the MCRA of extant *M. bovis* lineages agreed with previous studies
412 based on independent archeological data (South Siberian skeletons at ~2,000 BP). The results suggest
413 that dispersion and expansion of the species may have increased during the Ancient Roman period,
414 stimulated by the demographic expansion and rising need for cattle breeding. This may have contributed
415 to an increased number of TB cases in humans, albeit of zoonotic origin. Subsequently, the history of
416 economic trade, especially involving animals, may explain the continuous spread of bovine TB worldwide.
417 By understanding the evolutionary origin and genomic diversification of *M. bovis*, we expect that the
418 results presented herein will help pave the way to avoid future outbreaks of the disease in cattle, wildlife
419 and humans.
420
421 *Materials and Methods*
422 **Genome sequencing of Brazilian *M. bovis* genomes**
423 Two Brazilian *M. bovis* isolates obtained from an European bison (*Bison bonasus*) (Zimpel et al.
424 2017b), and a llama (*Llama glama*) (provided by the Laboratory of Bacterial Zoonosis of the College of
425 Veterinary Medicine, University of São Paulo, Brazil) were reactivated in Stonebrink medium and a
426 single colony was sub-cultured for DNA extraction using a previously described protocol (Zimpel et al.
427 2017a). DNA quality was evaluated using Nanodrop 2000c (Thermo Scientific, MA, USA) and Agilent
428 2100 High Sensitivity Chip DNA Bioanalyzer (Agilent Technologies, CA, USA). All procedures involving
429 live tuberculous mycobacteria were performed in a Biosafety Level 3+ Laboratory (BSL-3+ Prof. Dr. Klaus
430 Eberhard Stewien) located at the Department of Microbiology, Institute of Biomedical Sciences, University
431 of São Paulo, Brazil.
432 Paired-end genomic libraries were constructed using TruSeq DNA PCR-free sample preparation kit
433 (Illumina, CA, USA), and Illumina HiSeq2500 (Illumina v3 chemistry) was used to sequence the
434 genomic library (100 bp). These procedures were performed at the Central Laboratory of High
435 Performance Technologies in Life Sciences (LaCTAD), State University of Campinas (UNICAMP),

436 São Paulo, Brazil. Illumina sequencing reads were deposited in the Sequence Read Archive (SRA)
437 from the National Center for Biotechnology Information (NCBI) (accession numbers: SRR7693912 and
438 SRR7693877).
439
440 **Selection of *M. bovis* genomes**
441 As of September 2018, there were approximately 2,600 sequence read sets of *M. bovis* genomes
442 deposited in SRA, NCBI. Genomes from *M. bovis* were initially selected if they: (i) were sequenced using
443 Illumina technology [i.e., genomes sequenced using MinION (Oxford Nanopore Technologies, Oxford,
444 United Kingdom) and PacBio (Pacific Biosciences, California, USA) were excluded]; (ii) had a sequencing
445 coverage greater than 30X; (iii) presented known sample location; (iv) were virulent *M. bovis* strains, i.e.
446 not named as *M. bovis* BCG, *M. bovis* AN5, *M. bovis* Ravenel.
447 From all initially selected genomes, four countries were over-represented (n>100) in the dataset
448 (supplementary table S2) and were subsequently subsampled. Briefly, as to obtain host-species
449 representation, first, all *M. bovis* genomes obtained from these countries that were isolated from non-
450 bovine hosts were selected (except for cervids from the USA). From the remaining read sets from cattle
451 and USA cervids, *M. bovis* genomes were randomly subsampled (supplementary table S2) using the
452 online sample size calculator (http://www.raosoft.com/samplesize.html), choosing an error of 5%,
453 confidence level of 95%, and response distribution of 50% (as to maximize the sample size). A random
454 number generator (https://www.randomizer.org) was used to randomly select genomes from these read
455 sets.
456 We also selected 28 additional *M. bovis* genomes that were sequenced or identified after September
457 2018 (Brites et al. 2018). These isolates were from Germany (*n*=7), Ghana (*n*=5), Malawi (*n*=1), Republic
458 of Congo (*n*=3), Russia (*n*=2), Switzerland (*n*=2), and United Kingdom (*n*=6). The same inclusion criteria
459 described above was followed in the selection of these genomes. In the end, a total of 949 virulent *M.*
460 *bovis* genomes were initially selected using the above described methods and criteria (supplementary
461 table S3).
462
463 **Sequencing quality criteria**
464 FASTQ files of all 949 initially selected *M. bovis* genomes were downloaded from SRA, NCBI and
465 trimmed using Trimmomatic (Sliding window: 5:20) (Bolger et al. 2014). To guarantee the selection of
466 genomes with good sequence quality, the following quality criteria were considered after trimming: (i)
467 coverage ≥15x, (ii) median read length of 70 bp, (ii) absence of low quality sequences and anomalous
468 GC content as determined using FastQC, (iv) mapping coverage against reference genome *M. bovis*
469 AF2122/97 (NC_002945.4) greater than 95% as indicated by using Burrows-Wheeler Aligner (Li and
470 Durbin 2010).
471

472 **Mapping and variant calling of reads**

473 From the quality-approved *M. bovis* read sets, duplicated reads were removed using Picard v 2.18.23
474 (https://github.com/broadinstitute/picard). Reads were mapped against the reference genome *M. bovis*
475 AF2122/97 (NC_002945.4) using Burrows-Wheeler Aligner (Li and Durbin 2010). SNPs were called using
476 Samtools v 1.9 mpileup (Li 2011), and VarScan v2.4.3 mpileup2cns (Koboldt et al. 2012), using
477 parameters of read depth of 7, minimum map quality and minimum base quality of 20. Generated vcf files
478 were annotated using snpEff (Cingolani et al. 2012) based on the same reference genome, and
479 manipulated using awk programming language to remove INDELs (insertion and deletions) and SNPs
480 located in PE/PPE, transposases, integrases, maturase, phage and repetitive family 13E12 genes.
481 Mixed-strain infection was defined when 15% of total SNPs was composed by heterogeneous SNPs.
482 Read sets originating from mixed-strain cultures were excluded from downstream analyses.
483

484 **Phylogenetic Reconstruction**

485 A customized script in Python language was used to build a matrix of the polymorphic positions
486 identified in all *M. bovis* genomes. IQ-Tree (Nguyen et al. 2015) was used to perform a maximum
487 likelihood (ML) phylogenetic reconstruction, using 1,000 bootstrap inferences. Additionally, a
488 phylogenetic tree was constructed based on maximum parsimony (MP) with standard bootstrap (1,000
489 replicates) using MPBoot (Hoang et al. 2018). Both phylogenies were rooted using *M. tuberculosis*
490 H37Rv and topology was annotated using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) and
491 Mind the Graph (https://mindthegraph.com). In order to compare if the ML and MP trees were significantly
492 similar, the Robinson-Foulds distance (Robinson and Foulds 1981) between them was obtained, then
493 tested for significance of this estimated distance by generating 1,000 random trees in IQTree with the
494 same number of taxa used in the analysis. Finally, the original distance to the distribution of distances
495 among all random trees (for a total of 499,500 comparisons) was compared, and results were plotted in a
496 histogram done in R software.
497

498 **Spoligotyping and Clonal Complexes**

499 Spoligotypes of all selected genomes were investigated using SpoTyping (Xia et al. 2016). Identified
500 genetic spacers were processed in the *Mycobacterium bovis* Spoligotype Database (www.mbovis.org) to
501 retrieve spoligotype pattern and SB number. Intact regions of RdEu1 (European 1 - Eur1, RD17, 1,207
502 bp), SNP in the *guaA* gene (European 2 - Eur2), RDAf1 (African 1 - Afr1, 350 bp) and RDAf2 (African 2 -
503 Afr2, 451 bp) (Müller et al. 2009; Berg et al. 2011; Smith et al. 2011; Rodriguez-Campos et al. 2012) were
504 searched against all genomes using the BLASTn tool available in SRA, NCBI. Genomes with “no
505 significant similarity found” as result in BLASTn when using intact regions of CC markers were mapped
506 against the *Mycobacterium tuberculosis* H37Rv (NC_000962.3), using CLC Genomics Workbench 11
507 (QIAGEN, Venlo, Netherlands), to confirm the presence of RDs/SNPs.

508

509 **Population Structure**

510 To investigate the molecular clustering pattern of *M. bovis* genomes, the R package ADEGENET 2.1.1
511 (Jombart et al. 2010) was used to perform a Discriminant Analysis of Principal Components (DAPC),
512 resulting in the identification of clusters, using the Bayesian Information Criterion (BIC) for assessing the
513 best supported model, and therefore the number and nature of clusters. The software ggtree (Yu et al.
514 2017) was then used to generate an illustrated phylogenetic tree according to DAPC results.

515

516 **Dating divergences**

517 Molecular dating of divergences was performed using BEAST v2.5.2 (Bouckaert et al. 2014) with a
518 coalescent approach (extended bayesian skyline plot). The age of the root was set using the Bos et al.
519 (Bos et al. 2014) calibration, which is based on ancient human skeletons infected with *M. pinnipedii.* The
520 minimum age was set to 800 BP, after reasonable carbon-dating skeleton data, and the soft upper
521 maximum set as ~7,000 BP (a conservative upper bound based on Bos et al. (Bos et al. 2014) posterior
522 estimates; this soft bound allows older dates in our analyses, if these are likely to occur in the posterior
523 distribution of dates). Tip-dating was implemented with collection dates from known sampling times, plus
524 a conservative Uniform [0, 100] BP for the unknown dates.

525

526 **Acknowledgements**

527 The authors are in debt to Gisele Oliveira de Souza and Carolina Bertelli de Souza Ferreira from the
528 University of São Paulo, São Paulo, Brazil for invaluable technical assistance. We also thank the
529 LaCTAD, UNICAMP, Campinas, Brazil for aiding in the genome sequencing of *M. bovis* isolates, and
530 CEFAP, University of São Paulo, for computer core services. We are in debt with Dr Paulo Eduardo
531 Brandão for continuous mentoring support throughout this study. Fellowships for CKZ, JSLP, ACPG,
532 TTPS, NCSC, AFSF, CYI are provided by CNPq (Conselho Nacional de Pesquisa Científica), Ministry of
533 Science, Brazil (134266/2017-0; and 140003/2019-3), CAPES (Coordenação de Aperfeiçoamento de
534 Pessoal de Nível Superior), Ministry of Education, Brazil (88887.283881/2018-00; 1650900; and
535 1539669) and São Paulo Research Foundation (FAPESP) (2017/04617-3; and 2017/20147-7). This
536 study was financed in part by CAPES (Finance Code 001). Main research funding was made available
537 through Morris Animal Foundation (grant number D17ZO-307).

538

539 **Author contributions**

540 CKZ performed and/or participated in all experiments, analyzed and interpreted the data, and wrote the
541 manuscript; JSLP performed dating divergence experiments, analyzed and interpreted the resulting
542 data; ACPG, TTPS and NCSC aided in bioinformatics analyses; CYI and AFSF contributed with
543 biological data; JSFN, JCS, MBH helped design and analyze the data; AMSG designed and coordinated

544 the study, and wrote the manuscript. All co-authors contributed to data discussion and structuring of the
545 manuscript.
546
547 **Competing interests**
548 The authors declare no competing financial interests.
549
550 *References*

551 Berg S, Garcia-Pelayo MC, Müller B, Hailu E, Asiimwe B, Kremer K, Dale J, Boniotti MB, Rodriguez S,
552 Hilty M, et al. 2011. African 2, a clonal complex of *Mycobacterium bovis* epidemiologically
553 important in East Africa. *J. Bacteriol.* 193:670–678.
554 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
555 *Bioinformatics* 30:2114–2120.
556 Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR,
557 Schuenemann VJ, et al. 2014. Pre-Columbian mycobacterial genomes reveal seals as a source
558 of New World human tuberculosis. *Nature* 514:494–497.
559 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond
560 AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.*
561 10:e1003537.
562 Brites D, Loiseau C, Menardo F, Borrell S, Boniotti MB, Warren R, Dippenaar A, Parsons SDC, Beisel
563 C, Behr MA, et al. 2018. A New Phylogenetic Framework for the Animal-Adapted *Mycobacterium*
564 *tuberculosis* Complex. *Front. Microbiol.* 9:2820.
565 Bruning-Fann C, Robbe-Austerman S, Kaneene J, Thomsen B, Jr JT, Ray J, Smith R, Fitzgerald S,
566 Bolin S, O'Brien D, et al. 2017. Use of whole-genome sequencing and evaluation of the apparent
567 sensitivity and specificity of antemortem tuberculosis tests in the investigation of an unusual
568 outbreak of Mycobacterium bovis infection in a Michigan dairy herd. *J. Am. Vet. Med. Assoc*.
569 251:206–216.
570 Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant
571 J, Parkhill J, Nejentsev S, et al. 2014. Evolution and transmission of drug-resistant tuberculosis in
572 a Russian population. *Nat. Genet.* 46:279–286.
573 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A
574 program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*
575 *(Austin)*. 6:80–92.
576 Coscolla M, Gagneux S. 2014. Consequences of genomic diversity in Mycobacterium tuberculosis.
577 *Semin. Immunol.* 26:431–444.
578 Cross M, Heeren A, Cornicelli LJ, Fulton DC. 2018. Bovine Tuberculosis Management in Northwest
579 Minnesota and Implications of the Risk Information Seeking and Processing (RISP) Model for

Wildlife Disease Management. *Front. Vet. Sci.* 5:190.

Davidson JA, Loutet MG, O'Connor C, Kearns C, Smith RMM, Lalor MK, Thomas HL, Abubakar I, Zenner D. 2017. Epidemiology of *Mycobacterium bovis* Disease in Humans in England, Wales, and Northern Ireland, 2002-2014. *Emerg. Infect. Dis.* 23:377–386.

Dippenaar A, Parsons SDC, Miller MA, Hlokwe T, Gey van Pittius NC, Adroub SA, Abdallah AM, Pain A, Warren RM, Michel AL, et al. 2017. Progenitor strain introduction of Mycobacterium bovis at the wildlife-livestock interface can lead to clonal expansion of the disease in a single ecosystem. *Infect. Genet. Evol.* 51:235–238.

de la Fuente J, Díez-Delgado I, Contreras M, Vicente J, Cabezas-Cruz A, Tobes R, Manrique M, López V, Romero B, Bezos J, et al. 2015. Comparative Genomics of Field Isolates of *Mycobacterium bovis* and *M. caprae* Provides Evidence for Possible Correlates with Bacterial Viability and Virulence. *PLoS Negl. Trop. Dis.* 9:1–22.

Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, Gadisa E, Kiros T, Habtamu M, Hussein J, et al. 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg. Infect. Dis.* 19:460–463.

Galagan JE. 2014. Genomic insights into tuberculosis. *Nat. Rev. Genet.* 15:307–320.

Garnier T, Eiglmeier K, Camus J, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, et al. 2003. The Complete Genome Sequencing of Mycobacterium bovis. *Proc. Natl. Acad. Sci.* 100:7887–82.

Ghebremariam MK, Hlokwe T, Rutten VPMG, Allepuz A, Cadmus S, Muwonge A, Robbe-Austerman S, Michel AL. 2018. Genetic profiling of Mycobacterium bovis strains from slaughtered cattle in Eritrea. *PLoS Negl. Trop. Dis.* 12:e0006406.

Hanson JW, Ortman SG, Lobo J. 2017. Urbanism and the division of labour in the Roman Empire. *J. R. Soc. Interface* 14.

Herzog, Basel H. 2003. History of Tuberculosis. *Respiration* 65:5–15.

Hoang DT, Vinh LS, Flouri T, Stamatakis A, von Haeseler A, Minh BQ. 2018. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol. Biol.* 18:11.

Holloway KL, Henneberg RJ, de Barros Lopes M, Henneberg M. 2011. Evolution of human tuberculosis: A systematic review and meta-analysis of paleopathological evidence. *HOMO- J. Comp. Hum. Biol.* 62:402–458.

Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG. 2009. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5:17–19.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94.

de Jong BC, Antonio M, Gagneux S. 2010. Mycobacterium africanum-review of an important cause of

human tuberculosis in West Africa. *PLoS Negl. Trop. Dis.* 4.

Jongman WM. 2012. Re-constructing the roman economy. In: The Cambridge History of Capitalism Volume I: The Rise of Capitalism: From Ancient Origins to 1848. Larry Neal & Jeffrey G. Williamson, editors. Cambridge University Press: Cambridge pp. 75–100.

Kandler JL, Mercante AD, Dalton TL, Ezewudo MN, Cowan LS, Burns SP, Metchock B, van der Walt M, Brand J, Tupasi T, et al. 2018. Validation of Novel *Mycobacterium tuberculosis* Isoniazid Resistance Mutations Not Detectable by Common Molecular Tests. *Antimicrob. Agents Chemother*. 62.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22:568–576.

Kohl TA, Utpatel C, Niemann S, Moser I. 2018. Mycobacterium bovis persistence in two different captive wild animal populations in Germany: A longitudinal molecular epidemiological study revealing pathogen transmission by whole-genome sequencing. *J. Clin. Microbiol.* 56:1–9.

Kron G. 2002. Archaeozoology and the Productivity of Roman Livestock Farming. *Münstersche Beiträge zur Antiken Handel.* 21:53–73.

Lasserre M, Fresia P, Greif G, Iraola G, Castro-Ramos M, Juambeltz A, Nuñez Á, Naya H, Robello C, Berná L. 2018. Whole genome sequencing of the monomorphic pathogen Mycobacterium bovis reveals local differentiation of cattle clinical isolates. *BMC Genomics* 19:1–14.

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.

Livingstone PG, Hancox N, Nugent G, de Lisle GW. 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 1:4–18.

Majoor CJ, Magis-Escurra C, van Ingen J, Boeree MJ, van Soolingen D. 2011. Epidemiology of *Mycobacterium bovis* Disease in Humans, the Netherlands, 1993–2007. *Emerg. Infect. Dis.* 17:457–463.

Malone KM, Gordon S V. 2017. Mycobacterium tuberculosis Complex Members Adapted to Wild and Domestic Animals. In: Advances in experimental medicine and biology. Gagneux, S, editor. Springer, Cham pp. 135–154.

Martínez AM, Gama LT, Cañón J, Ginja C, Delgado J V., Dunner S, Landi V, Martín-Burriel I, Penedo MCT, Rodellar C, et al. 2012. Genetic Footprints of Iberian Cattle in America 500 Years after the Arrival of Columbus. *PLoS One* 7:e49066.

Meadow R, Zeder M. 2004. Approaches to Faunal Analysis in the Middle East. Cambridge, MA:

652 Peabody Museum of Archaeology and Ethnology, Harvard University
653 Meadow RH. 1988. Archeology: The Archaeology of Animals. Simon J. M. Davis. *Am. Anthropol.*
654 90:697.
655 Meadow RH. 1989. Osteological evidence for process of animal domestication. In: The Walking larder:
656 patterns of domestication, pastoralism, and predation. Unwin Hyman: London pp. 80–90.
657 Meikle V, Bianco M V, Blanco FC, Gioffré A, Garbaccio S, Vagnoni L, Di Rienzo J, Canal A, Bigi F,
658 Cataldi A. 2011. Evaluation of pathogenesis caused in cattle and guinea pig by a *Mycobacterium*
659 *bovis* strain isolated from wild boar. *BMC Vet. Res*. 7:37.
660 Menardo F, Duchene S, Brites D, Gagneux S. 2019. The molecular clock of Mycobacterium
661 tuberculosis. bioRxiv. Available from: https://www.biorxiv.org/content/10.1101/532390v2
662 Michel AL, Bengis RG, Keet DF, Hofmeyr M, Klerk LM de, Cross PC, Jolles AE, Cooper D, Whyte IJ,
663 Buss P, et al. 2006. Wildlife tuberculosis in South African conservation areas: Implications and
664 challenges. *Vet. Microbiol.* 112:91–100.
665 Michel AL, Coetzee ML, Keet DF, Maré L, Warren R, Cooper D, Bengis RG, Kremer K, van Helden P.
666 2009. Molecular epidemiology of Mycobacterium bovis isolates from free-ranging wildlife in South
667 African game reserves. *Vet. Microbiol.* 133:335–343.
668 Müller B, Dürr S, Alonso S, Hattendorf J, Laisse CJM, Parsons SDC, Helden PD Van, Zinsstag J.
669 2013. Zoonotic *Mycobacterium bovis*-induced Tuberculosis in Humans. *Emerg. Infect. Dis.*
670 19:899–908.
671 Müller B, Hilty M, Berg S, Garcia-Pelayo MC, Dale J, Boschiroli ML, Cadmus S, Ngandolo BNR,
672 Godreuil S, Diguimbaye-Djaibé C, et al. 2009. African 1, an epidemiologically important clonal
673 complex of Mycobacterium bovis dominant in Mali, Nigeria, Cameroon, and Chad. *J. Bacteriol.*
674 191:1951–1960.
675 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic
676 Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32:268–274.
677 O'Neill M, Kitchen A, Zarley A, Aylward W, Eldholm V, Pepperell C. 2017. Lineage specific histories of
678 Mycobacterium tuberculosis dispersal in Africa and Eurasia. bioRxiv. Available from:
679 https://www.biorxiv.org/content/10.1101/210161v1
680 Olea-Popelka F, Muwonge A, Perera A, Dean AS, Mumford E, Erlacher-Vindel E, Forcella S, Silk BJ,
681 Ditiu L, El Idrissi A, et al. 2017. Zoonotic tuberculosis in human beings caused by *Mycobacterium*
682 *bovis* - a call for action. *Lancet Infect. Dis*. 17:e21–e25.
683 Orlando L. 2015. The first aurochs genome reveals the breeding history of British and European cattle.
684 *Genome Biol.* 16:225.
685 Orloski K, Robbe-austerman S, Stuber T, Hench B, Schoenbaum M. 2018. Whole genome
686 sequencing of Mycobacterium bovis isolated from livestock in the United States, 1989-2018.
687 *Front. Vet. Sci.* 5:253:1–23.

688 Patané JSL, Martins J, Castelão AB, Nishibe C, Montera L, Bigi F, Zumárraga MJ, Cataldi AA, Junior
689 AF, Roxo E, et al. 2017. Patterns and Processes of Mycobacterium bovis Evolution Revealed by
690 Phylogenomic Analyses. *Genome Biol. Evol.* 9:521–535.
691 Price-Carter M, Brauning R, de-Lisle GW, Livingstone P, Neill M, Sinclair J, Paterson B, Atkinson G,
692 Knowles G, Crews K, et al. 2018. Whole Genome Sequencing for Determining the Source of
693 Mycobacterium bovis Infections in Livestock Herds and Wildlife in New Zealand. *Front. Vet. Sci.*
694 5:1–13.
695 Razo CAP, Hernández ER, Ponce SIR, Suazo FM, Robbe-Austerman S, Stuber T, Alarcón GJC.
696 2018. Research article molecular epidemiology of cattle tuberculosis in mexico through whole-
697 genome sequencing and spoligotyping. *PLoS One* 13:1–14.
698 Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131–147.
699 Rodriguez-Campos S, Schürch AC, Dale J, Lohan AJ, Cunha M V., Botelho A, Cruz K De, Boschiroli
700 ML, Boniotti MB, Pacciarini M, et al. 2012. European 2 – A clonal complex of *Mycobacterium*
701 *bovis* dominant in the Iberian Peninsula. *Infect. Genet. Evol.* 12:866–872.
702 Rodwell TC, Moore M, Moser KS, Brodine SK, Strathdee SA. 2008. Tuberculosis from Mycobacterium
703 bovis in binational communities, United States. *Emerg. Infect. Dis.* 14:909–916.
704 Sandoval-Azuara SE, Muñiz-Salazar R, Perea-Jacobo R, Robbe-Austerman S, Perera-Ortiz A, López-
705 Valencia G, Bravo DM, Sanchez-Flores A, Miranda-Guzmán D, Flores-López CA, et al. 2017.
706 Whole genome sequencing of Mycobacterium bovis to obtain molecular fingerprints in human
707 and cattle isolates from Baja California, Mexico. *Int. J. Infect. Dis.* 63:48–56.
708 Smith NH. 2012. The global distribution and phylogeography of Mycobacterium bovis clonal
709 complexes. *Infect. Genet. Evol.* 12:857–865.
710 Smith NH, Berg S, Dale J, Allen A, Rodriguez S, Romero B, Matos F, Ghebremichael S, Karoui C,
711 Donati C, et al. 2011. European 1: A globally important clonal complex of *Mycobacterium bovis*.
712 *Infect. Genet. Evol.* 11:1340–1351.
713 Taylor GM, Murphy E, Hopkins R, Rutland P, Chistov Y. 2007. First report of *Mycobacterium bovis*
714 DNA in human remains from the Iron Age. *Microbiology* 153:1243–1249.
715 Trentacoste A, Nieto-Espinet A, Valenzuela-Lamas S. 2018. Pre-Roman improvements to agricultural
716 production: Evidence from livestock husbandry in late prehistoric Italy. *PLoS One* 13:e0208109.
717 Valenzuela-Lamas S, Albarella U. 2017. Animal Husbandry across the Western Roman Empire:
718 Changes and Continuities. *Eur. J. Archaeol.* 20:402–415.
719 Vargas-Romero F, Mendoza-Hernández G, Suárez-Güemes F, Hernández-Pando R, Castañón-
720 Arreola M. 2016. Secretome profiling of highly virulent *Mycobacterium bovis* 04-303 strain reveals
721 higher abundance of virulence-associated proteins. *Microb. Pathog.* 100:305–311.
722 Walker TM, Kohl TA, Omar S V, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus
723 KE, Wilson DJ, et al. 2015. Whole-genome sequencing for prediction of *Mycobacterium*

tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet. Infect. Dis.* 15:1193–1202.

Waters WR, Palmer M V, Thacker TC, Bannantine JP, Vordermeier HM, Hewinson RG, Greenwald R, Esfandiari J, McNair J, Pollock JM, et al. 2006. Early antibody responses to experimental *Mycobacterium bovis* infection of cattle. *Clin. Vaccine Immunol.* 13:648–654.

Waters WR, Thacker TC, Nelson JT, DiCarlo DM, Maggioli MF, Greenwald R, Esfandiari J, Lyashchenko KP, Palmer MV. 2014. Virulence of Two Strains of Mycobacterium bovis in Cattle Following Aerosol Infection. *J. Comp. Pathol.* 151:410–419.

Wedlock DN, Aldwell FE, Collins DM, de Lisle GW, Wilson T, Buddle BM. 1999. Immune responses induced in cattle by virulent and attenuated *Mycobacterium bovis* strains: correlation of delayed-type hypersensitivity with ability of strains to grow in macrophages. *Infect. Immun.* 67:2172–2177.

World Health Organization. 2018. WHO Tuberculosis Report 2018. Geneva: World Health Organization

Wright DM, Allen AR, Mallon TR, McDowell SWJ, Bishop SC, Glass EJ, Bermingham ML, Woolliams JA, Skuce RA. 2013. Field-Isolated Genotypes of *Mycobacterium bovis* Vary in Virulence and Influence Case Pathology but Do Not Affect Outbreak Size. *PLoS One* 8:e74503.

Xia E, Teo Y-Y, Ong RT-H. 2016. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. *Genome Med.* 8:19.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8:28–36.

Zimpel CK, Brandão PE, de Souza Filho AF, de Souza RF, Ikuta CY, Ferreira Neto JS, Camargo NCS, Heinemann MB, Guimarães AMS. 2017a. Complete Genome Sequencing of *Mycobacterium bovis* SP38 and Comparative Genomics of *Mycobacterium bovis* and *M. tuberculosis* Strains. *Front. Microbiol.* 8:2389.

Zimpel CK, Brum JS, de Souza Filho AF, Biondo AW, Perotta JH, Dib CC, Bonat M, Neto JSF, Brandão PE, Heinemann MB, et al. 2017b. *Mycobacterium bovis* in a European bison (*Bison bonasus*) raises concerns about tuberculosis in Brazilian captive wildlife populations: a case report. BMC Res. Notes 10:91.