Complete Genome Sequencing of *Mycobacterium bovis* SP38 and Comparative Genomics of *Mycobacterium bovis* and *M. tuberculosis* Strains

Cristina Kraemer Zimpel[1,2], Paulo E. Brandão[2], Antônio F. de Souza Filho[2], Robson F. de Souza[3], Cássia Y. Ikuta[2], José Soares Ferreira Neto[2], Naila C. Soler Camargo[1], Marcos Bryan Heinemann[2] and Ana M. S. 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] Laboratory of Protein Structure and Evolution, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

Mycobacterium bovis causes bovine tuberculosis and is the main organism responsible for zoonotic tuberculosis in humans. We performed the sequencing, assembly and annotation of a Brazilian strain of *M. bovis* named SP38, and performed comparative genomics of *M. bovis* genomes deposited in GenBank. *M. bovis* SP38 has a traditional tuberculous mycobacterium genome of 4,347,648 bp, with 65.5% GC, and 4,216 genes. The majority of CDSs (2,805, 69.3%) have predictive function, while 1,206 (30.07%) are hypothetical. For comparative analysis, 31 *M. bovis*, 32 *M. bovis* BCG, and 23 *Mycobacterium tuberculosis* genomes available in GenBank were selected. *M. bovis* RDs (regions of difference) and Clonal Complexes (CC) were identified *in silico*. Genome dynamics of bacterial groups were analyzed by gene orthology and polymorphic sites identification. *M. bovis* polymorphic sites were used to construct a phylogenetic tree. Our RD analyses resulted in the exclusion of three genomes, mistakenly annotated as virulent *M. bovis*. *M. bovis* SP38 along with strain 35 represent the first report of CC European 2 in Brazil, whereas two other *M. bovis* strains failed to be classified within current CC. Results of *M. bovis* orthologous genes analysis suggest a process of genome remodeling through genomic decay and gene duplication. Quantification, pairwise comparisons and distribution analyses of polymorphic sites demonstrate greater genetic variability of *M. tuberculosis* when compared to *M. bovis* and *M. bovis* BCG ($p \leq 0.05$), indicating that currently defined *M. tuberculosis* lineages are more genetically diverse than *M. bovis* CC and animal-adapted MTC (*M. tuberculosis* Complex) species. As expected, polymorphic sites annotation shows that *M. bovis* BCG are subjected to different evolutionary pressures when compared to virulent mycobacteria. Lastly, *M. bovis* phylogeny indicates that polymorphic sites may be used as markers of *M. bovis* lineages in association with CC. Our findings highlight the need to better understand host-pathogen co-evolution in genetically homogeneous and/or diverse host populations, considering the fact that

M. bovis has a broader host range when compared to *M. tuberculosis*. Also, the identification of *M. bovis* genomes not classified within CC indicates that the diversity of *M. bovis* lineages may be larger than previously thought or that current classification should be reviewed.

INTRODUCTION

Tuberculosis is an infectious disease that affects human beings and animals, and is caused by members of the *Mycobacterium tuberculosis* Complex (MTC). The MTC is composed of the human-adapted pathogens *M. tuberculosis* lineages 1 to 4 and *Mycobacterium africanum* lineages 5 and 6, and the animal-adapted species, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium pinnipedii*, *Mycobacterium microti*, *Mycobacterium orygis*, *Mycobacterium suricattae*, *Mycobacterium mungi,* "dassie bacillus" and "chimpanzee bacillus" (Coscolla and Gagneux, 2014; Rodriguez-Campos et al., 2014). These microorganisms have great genomic similarity (>99% of nucleotide identity in alignable regions) and have clonally evolved from a *M. tuberculosis*-like ancestor (Smith et al., 2009; Galagan, 2014). Despite their great genetic relatedness, MTC members are categorized into different species due to species-specific genotypic variations, such as single nucleotide polymorphisms (SNPs) and deletion of regions of difference (RDs), and phenotypic variations related to host adaptability and virulence (Brosch et al., 2002; Galagan, 2014; Rodriguez-Campos et al., 2014).

Nearly 10 million people are affected by tuberculosis worldwide every year. Although the majority of these cases are caused by *M. tuberculosis*, there is an underestimated proportion of zoonotic cases caused by *M. bovis*, the pathogen of bovine tuberculosis (Olea-Popelka et al., 2017). The burden of zoonotic tuberculosis is possibly a reflection of the nations' socioeconomic development and their ability to control the disease in cattle and pasteurize the milk. Unfortunately, bovine tuberculosis is still geographically widespread, especially in developing nations and countries with hard-to-control wildlife reservoirs, and causes significant economic losses for livestock producers (El-Sayed et al., 2016). In Brazil, the epidemiological situation of bovine tuberculosis is well-characterized in 75% of the cattle population; transversal studies showed prevalence of infected herds varying from 0.36 to 9.0% in different States. The disease is more prevalent in high-production dairy farms and the major risk factor for infection is the introduction of animals without tuberculin testing (Dias et al., 2016; Ferreira Neto et al., 2016; Ribeiro et al., 2016). Despite the fact that Brazil is the second world largest producer of bovine meat, little is known about *M. bovis* strains causing disease in the country, especially regarding their transmission capacity, virulence and genetic characteristics. Although few studies have shown variable virulence and transmissibility capabilities among different *M. bovis* strains, these have rarely included genomic analyses of these isolates (Aguilar León et al., 2009; Blanco et al., 2009; Meikle et al., 2011; Andrade et al., 2012; Wright et al.,

2013; Garbaccio et al., 2014). Thus, in contrast to *M. tuberculosis*, the virulence and genomic characteristics of different *M. bovis* lineages are largely unexplored.

Comparative genomics studies among MTC members have been mostly performed aiming to understand *M. tuberculosis* features (Kato-Maeda et al., 2001; Ilina et al., 2013; Coscolla and Gagneux, 2014; Liu et al., 2014; Phelan et al., 2016). *M. bovis* genomes are generally included only for isolated comparison (Cole, 2002; Garcia Pelayo et al., 2009; Joshi et al., 2012; McGuire et al., 2012; Rue-Albrecht et al., 2014; de la Fuente et al., 2015; Periwal et al., 2015). Intra-species comparative genomics studies of *M. bovis* are related to epidemiologic investigations of outbreaks of animal tuberculosis using genome sequencing reads to detect SNPs and infer transmission networks (Biek et al., 2012; Glaser et al., 2016; Crispell et al., 2017). Only two studies using complete and draft genomes of *M. bovis* strains have been published. The first performed sequencing of three *M. bovis* and one *M. caprae* genomes to correlate genomic characteristics with virulence phenotypes (e.g., lesion scores and hosts) (de la Fuente et al., 2015). The second evaluated the phylogenetic relatedness of 38 *M. bovis* genomes, identifying recombination sites, and suggesting that some North-American lineages of *M. bovis* evolved faster than others, which might be related to bacterial maintenance in different hosts causing new outbreaks in the USA (Patané et al., 2017). Thus, a more comprehensive comparative genomic study of *M. bovis* strains using complete and draft genomes is still lacking and should provide important data to understand the biology of this pathogen, especially when compared to *M. tuberculosis*. Thus, the aims of the present study were to sequence, assemble, close, and annotate the genome of a Brazilian strain of *M. bovis* and perform a comparative genomic analysis with other virulent *M. bovis*, *M. bovis* BCG and *M. tuberculosis* genomes available in a public database.

MATERIALS AND METHODS

Isolation of *M. bovis* Strain SP38 and DNA Extraction

The Laboratory of Bacterial Zoonosis of the College of Veterinary Medicine, University of São Paulo, Brazil has been isolating *M. bovis* strains from all over the country since the 1990's. From this bacterial collection, a *M. bovis* strain was randomly selected to be sequenced in this study and called SP38. This strain was originally isolated from a granulomatous lesion collected from a bovine in a slaughterhouse located in the State of São Paulo, Brazil in 2010 (Morato et al., 2016). The lesion was cultured in Stonebrink medium as described previously (Ikuta et al., 2016), following standard procedures

(Centro Panamericano de Zoonosis, 1985), and the isolate was stored at −20°C with no further passages. For the purpose of extracting DNA for this study, this isolate was reactivated in Stonebrink medium. After incubation, a single colony was selected and grown in fresh medium. When growth was detected, a sterile loop was used to collect bacterial mass and DNA was extracted as previously described (van Soolingen et al., 1994; Bemer-Melchior and Drugeon, 1999). All procedures were performed in a Biosafety Level 3+ Laboratory (BSL-3+ Prof. Dr. Klaus Eberhard Stewien) located at the Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Brazil. Tubes containing DNA were properly disinfected, removed from BSL3+ and stored at −20°C until further analysis.

DNA Sequencing, Genome Assembly, and Annotation

Extracted DNA was sent to the Genomics Core Facility of Purdue University, USA for sequencing. Briefly, the quality and concentration of the DNA were measure by Nanodrop 2000C (Thermo Fischer Scientific, Massachusetts, USA) and 0.8% agarose gel using a mass ladder. DNA was sheared into 800 bp (base pairs) fragments using the Hydroshear DNA Shearing Device (Genomic Solutions, Michigan, USA) to create fragments for paired end libraries. Sheared DNA was analyzed using a DNA High Sensitivity Chip in an Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA), resulting in average fragment size of 804 bp. A paired-end genomic library was then constructed using TruSeq DNA PCR-free sample preparation kit (Illumina, California, USA) and 20% of a lane (Illumina v3 chemistry) of Illumina HiSeq2500 was used to sequence it.

De novo assembly of the sequenced reads was performed using ABySS software (Simpson et al., 2009) and CLC Genomics Workbench 8 (QIAGEN, Venlo, Holland). Additionally, reads were mapped to a reference genome (*M. bovis* strain AF2122/97—NC_002945.3) with CLC Genomics Workbench 8 (QIAGEN). ABySS generated contigs were ordered with Projector 2 (van Hijum et al., 2005) using the same reference genome. Contigs showing clear overlaps based on Smith-Waterman alignment, without repetitive sequences, were manually joined. Remaining gaps were closed using complementary information from CLC Genomic Workbench's *de novo* and reference mapping assemblies and/or using PCR followed by Sanger sequencing. Gene prediction and annotation was performed by NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

Comparative Genomic Analyses

As of 2016, 31 virulent *M. bovis* complete or draft genomes were available in GenBank and were selected for this study. These genomes were subjected to MTC species confirmation by manually checking the RD patterns *in silico* (presence or absence of RD1, RD4, RD9, RD12, RD1mic, and RD2seal) (Warren et al., 2006). Following this evaluation, 28 genomes of *M. bovis* were used in this study, which include the *M. bovis* strain SP38 and 3 other complete genomes and 24 drafts. These were from *M. bovis* isolated from cattle, wild boar, sheep and chimpanzee, in Brazil

(10/28), Argentina (7/28), Spain (3/28), South Korea (3/28), United Kingdom (1/28), China (1/28), Uruguay (1/28), France (1/28), and Uganda (1/28) (Supplementary Table 1).

Also as of 2016, 37 genomes of *M. bovis* BCG were available in GenBank. After excluding drafts with high number of contigs (>321), 32 *M. bovis* BCG genomes (11 complete and 21 drafts) were included in this study (Supplementary Table 2). Additionally, 23 *M. tuberculosis* complete genomes were selected, also from GenBank, representing lineages 1 (Philippines and Indian Ocean), 2 (East Asia), 3 (India and East Africa), and 4 (Europe, America, Africa), and antibiotic sensitive and resistant strains (Supplementary Table 3).

For the phylogenomic analysis, the following MTC genomes were included (in addition to all 28 *M. bovis* and *M. tuberculosis* H37Rv): *M. africanum* GM041182, *M. africanum* MAL010070, *M. microti* 12, *M. caprae* MB2, *M. orygis* 112400015, *M. mungi* BM22813, and *M. suricattae* (Supplementary Table 4). All genomes are available in GenBank with the exception of *M. suricattae*. The reads of *M. suricattae* were obtained from ENA (European Nucleotide Archive).

In Silico Spoligotyping of *M. bovis* Genomes

For the spoligotype identification, the reads of *M. bovis* strain SP38 were analyzed in SpolPred (Coll et al., 2012), and the complete and draft genomes of *M. bovis* were investigated using SpoTyping (Xia et al., 2016). The resulting patterns were submitted to the *M. bovis* Spoligotype Database (www.mbovis. org) for spoligotype pattern identification.

Clonal Complexes of *M. bovis* Genomes

The four Clonal Complexes (Müller et al., 2009; Berg et al., 2011; Smith et al., 2011; Rodriguez-Campos et al., 2012) were investigated among all 28 *M. bovis* genomes. *In silico* evaluation consisted in the evaluation of: RDAf1 presence or absence and deletion of spacer 30 for African 1 using three previously described primers to localize the sequences (Müller et al., 2009); RDAf2 presence or absence and deletion of spacers 3–7 for African 2 using three previously described primers to localize the sequences (Berg et al., 2011); RDEu1 presence or absence and deletion of spoligotype spacer 11 for Clonal Complex European 1 using a pair of previously described primers to localize the sequences (Smith et al., 2011) (Supplementary Table 5); SNP detection in the gene *gua*A at 3,765,573 position according to the reference genome (*M. bovis* strain AF2122/97) and deletion of spoligotype spacer 21 for European 2 (Rodriguez-Campos et al., 2012).

Paralogous Gene Families

Paralogous gene families (PGF) of *M. bovis* strain SP38 were identified using BLASTClust, available online in the MPI bioinformatics Toolkit (Alva et al., 2016). Parameters of 70% coverage and 30% of identity were used. PGF were separated into functional categories according to COG (Cluster of Orthologous Group) (Tatusov et al., 2000).

Gene Orthology and Synteny

For the identification of the pangenome, core and accessory genomes of *M. bovis*, all selected *M. bovis* genomes files in ".gbk" format were uploaded in KBase platform (Arkin et al., 2016). Pangenome analysis was performed with OrthoMCL (Li et al., 2003), selecting "Build Pangenome with OrthoMCL." All CDSs (coding DNA sequences) were categorized according to COGs. When necessary (i.e., proteins identified as singletons or grouped as species or strain-specific orthologous clusters), individual protein sequences were analyzed using BLASTp and/or tBLATn (Altschul et al., 1990) against *M. bovis* and *M. tuberculosis* genomes.

For the identification of groups of orthologous proteins among virulent *M. bovis*, *M. bovis* BCG, and *M. tuberculosis*, all genomes were analyzed using Sybil, a comparative genomics platform provided by the Institute for Genome Sciences, University of Maryland, USA. Sybil and associated algorithms identified protein clusters using reciprocal best BLAST match corrected for paralogs as previously described (Crabtree et al., 2007; Riley et al., 2012). An all-vs-all BLASTp search identified pairs of best-hit in the genomes (*e*-value of 1e-05, 80% identity, 70% coverage). Paralogous genes were then clustered using Jaccard similarity coefficient, with cutoff of 0.6, for each protein. The resulting paralogous proteins and singletons were used to identify orthologous proteins among the genomes (Riley et al., 2012).

Synteny maps were also constructed using Sybil. These maps are characterized by a linear genome alignment arranged by CDSs in vertical bars, and are based on the orthologs identification procedure described above. Synteny plots are built by coloring CDSs of a selected reference genome in a gradient from yellow to blue (left to right). If the query genome shares an ortholog cluster with the reference genome, this cluster is indicated above the sequence using the color that corresponds to the query CDS position in its reference genome. This plot can provide information about gene order conservation, regions without orthology, and rearrangements among genomes (Riley et al., 2012). For this study, only complete genomes were analyzed in synteny maps: four genomes of virulent *M. bovis*, one of *M. bovis* BCG (strain Pasteur) and one of *M. tuberculosis* (strain H37Rv).

Polymorphic Sites

Genomes were clustered into three groups: virulent *M. bovis*, *M. bovis* BCG, and *M. tuberculosis*. Polymorphic sites were detected in each group with kSNP3 (Gardner et al., 2015). Detected mutations were annotated and identified as synonymous or non-synonymous. For the annotation, complete genomes of each group (four complete genomes of *M. bovis*, 11 of *M. bovis* BCG, and 23 of *M. tuberculosis*) were selected as references. Genes with polymorphic sites were then categorized according to COGs (Tatusov et al., 2000).

To determine which mycobacterium species group has the greatest number of polymorphic sites, pairwise analyses were performed in an all-against-all fashion for each group. Results were statistically compiled in GraphPad Prism 6 (GraphPad Software Inc, La Jolla, California, USA), where the three groups

were compared using the non-parametric Kruskal–Wallis test, followed by Dunn test to detect differences between two groups. Results were considered statistically significant when $p \leq 0.05$.

The distribution of core polymorphic sites according to the number of sequenced MTC genomes was also analyzed. For this analyses, the genomes were clustered into four groups: one group containing all selected *M. tuberculosis* genomes; one group containing animal-adapted MTC species (28 genomes of *M. bovis*, and one of each of *M. caprae*, *M. microti*, *M. mungi*, *M. suricattae*, and *M. orygis*); one group with all selected *M. bovis* genomes; and a group with all MTC genomes (all genomes described above and two genomes of *M. africanum* lineages 5 and 6). The convergence of the observed number of segregating sites using subsets of these groups was determined. Briefly, for a given number of genomes, a new multiple sequence alignment of the core polymorphic sites was generated by randomly selecting the same number of rows from the alignment of all core polymorphic sites generated by kSNP3 for each complete dataset. Every combination of genomes was sampled only once (i.e., without replacement) and the number of samples was limited to a maximum of 10,000 samples. The number of segregating sites was then calculated for each resampled alignment by counting all columns that contained at least one sequence with a non-conserved, i.e., divergent, nucleotide.

Phylogenomic Analysis

A phylogenetic tree based on core polymorphic sites was built using all selected *M. bovis* genomes, in addition to one of each MTC species. Polymorphic sites identification was performed with kSNP3, and the file "SNPs_all_matrix.fasta" was selected to construct a multiple sequence alignment in MEGA 7 (Kumar et al., 2016), using MUSCLE algorithm (Edgar, 2004). The resulting alignment was used to construct phylogenetic trees using Neighbor-Joining (NJ) and Maximum-Likelihood (ML) algorithms with bootstrap of 1,000 replications and Jukes Cantor substitution model.

RESULTS AND DISCUSSION

Genome of *M. bovis* Strain SP38: A Traditional Tuberculous Mycobacterium

A total of 29,651,856 reads containing 2,858,321,177 bases resulted from *M. bovis* SP38 genome sequencing (approximate coverage of 657x). First-pass assembly using filtered reads (5,930,382—with 571,709,745 bases) in ABySS software resulted in 51 contigs (4,373,651 bases), with a total of 68 gaps. Following contig ordering with Projector 2 and manual check for overlaps, most gaps were closed. Five remaining gaps were resolved using additional assemblies (*de novo* and "map reads to a reference genome" with CLC Genomics Workbench) and PCR followed by Sanger sequencing. Considering that there are no optimal assembly programs (Wences et al., 2015) and mycobacteria genomes are difficult to close due to repeat regions and high GC content (Kumar and Kaur, 2014), the combination of different assembly algorithms allowed us to obtain a complete genome without gaps.

A singular, circular, and complete genome of *M. bovis* strain SP38 with 4,347,648 bp was obtained, and submitted to GenBank to update the previous draft genome (Guimaraes et al., 2015) (updated accession number: NZ_CP015773.2). The PGAP annotation revealed a traditional tuberculous mycobacterium genome, with high GC content (65.6%) and 4,216 genes, including 154 pseudogenes, 3 rRNA genes (ribosomal RNA), 45 tRNA (transfer RNA), 2 ncRNA (non-coding RNA), 1 tmRNA (transfer-messenger RNA), and 4,011 CDSs. Among the CDSs, most (2,805/4,011; 69.93%) have predicted function, and 1,206 (30.07%) encode for hypothetical proteins. A total of 307 PGF (with 2 to 100 members each), representing 1,049/4,011 (26.79%) CDSs were also found. A great number (318/1,049; 30.31%) of these families encode for hypothetical proteins and/or PE/PPE proteins (proline-glutamic and proline-proline-glutamic, respectively). Similar to *M. tuberculosis* (Voskuil et al., 2004), 3.54% of the CDSs (142/4,011) in *M. bovis* strain SP38 are annotated as PE/PPE proteins. It is believed that these PGFs are important virulence factors, providing immune system evasion through genetic variation (Fishbein et al., 2015; Phelan et al., 2016). Finally, the spoligotype of *M. bovis* strain SP38 was identified as SB0121, the most prevalent spoligotype in Brazil (Rocha, 2013; Carvalho et al., 2016).

Not All Genomes Available in Genbank Are Virulent *M. bovis*

Three out of the 31 genomes identified in GenBank as *M. bovis* (3/31) did not have the RD pattern that is characteristic of this species. Accordingly, *M. bovis* strain ATCC BAA-935, demonstrated a RD pattern consistent with *M. bovis* BCG; the human isolate from Uganda, *M. bovis* strain B2 7505 (Wanzala et al., 2015), revealed *M. tuberculosis* RD patterns; and *M. bovis* strain MAL010093, also isolated from a human patient in Africa, was characterized as *M. africanum* (Supplementary Table 6). Although *M. bovis* ATCC BAA-935 is labeled in GenBank as AF2122/97, the original genome of strain AF2122/97 (NC_002945.4) has RD patterns consistent with virulent *M. bovis*. Thus, from the 31 selected genomes, 28 were confirmed as virulent *M. bovis* and used in further analysis. These findings warrant caution in species confirmation prior to depositing genome sequences in public databases. In order to avoid errors in comparative genomics analyses, it is essential that bacterial genomes are correctly classified. MTC species can be confirmed prior to genome sequencing through molecular diagnostic (Warren et al., 2006), or after acquiring genome reads using the software RD-analyzer (Faksri et al., 2016). This likely mistaken species definition led Patané et al. (2017) to suggest that *M. bovis* strain ATCC BAA-935, identified as BCG, was a virulent *M. bovis* with significant genetic recombination areas and large sequence polymorphisms (LSP).

The Most Common Spoligotype among Genbank-Deposited *M. bovis* Genomes Is SB0140

The majority (19/28, 67.86%) of the 28 *M. bovis* genomes available in GenBank have the spoligotype SB0140

(Supplementary Table 7). Spoligotype patterns from five *M. bovis* strains failed to be identified using the proposed software. Fortunately, the spoligotype classification of two of these strains (*M. bovis* strains MB1 and MB3) were described in a previous study (de la Fuente et al., 2015). The remaining three (*M. bovis* strains B_322, Bz 31150, and 30) were listed as unknown patterns.

Despite the fact that spoligotype SB0121 is considered the most prevalent in Brazil and in the Iberian Peninsula (Rodriguez-Campos et al., 2013; Zumárraga et al., 2013; Carvalho et al., 2016), only one Brazilian strain (*M. bovis* strain SP38) (1/10) analyzed herein demonstrated this pattern. The other national strains (9/10) were identified as SB0140, which is associated with the United Kingdom and is described as the most prevalent in Argentina (Zumárraga et al., 2013; Garbaccio et al., 2014) due to the importation of British cattle in that country (Zumárraga et al., 2013). In this context, all (7/7) Argentinian strains also demonstrated this pattern. The proximity of Brazil to Argentina and the livestock trade between these countries may explain why SB0140 has high frequency among Brazilian genomes available in GenBank. These results are obviously not representative of an epidemiological survey, but can be taken into consideration in further phylogenomic analysis, as described below.

Clonal Complex European 2 Is Found in Brazil

The Clonal Complexes of *M. bovis* genomes are described below (**Table 1** and Supplementary Table 8). Two genomes failed to be categorized using the current parameters: *M. bovis* strains D_10_02315 and MB3, both isolated from wild boars in France and Spain, respectively. Moreover, the Clonal Complex African 1 was not detected among the studied genomes, probably due to low geographical representativeness of the sample. The lack of Clonal Complex identification in two strains allow us to question if there are other lineages that have yet to be described, or if current complexes' patterns should be reviewed to include more genetically diverse strains.

Mycobacterium bovis strain SP38, sequenced in this study, and the strain 35 of *M. bovis* constitute the first report of Clonal Complex European 2 in Brazil. Contrary to a previous study that reported that only 16% (6/36) of Brazilian *M. bovis* isolates are classified as Clonal Complex European 1 (Smith et al., 2011), this complex was identified as the most common in Brazilian *M. bovis* genomes available in GenBank (8/10). However, this data must be interpreted with caution, as the studied sample is not geographically representative. More studies will allow to confirm the distribution of Clonal Complexes in the Brazilian territory. Nevertheless, it is safe to conclude that Clonal Complexes European 1 and 2 can be found in Brazil.

Genomic Regions with Altered Synteny Are Identified among Virulent *M. bovis* Genomes

As expected, synteny maps of complete *M. bovis* genomes and reference genomes of *M. tuberculosis* and *M. bovis* BCG showed great conservation of overall gene order (**Figure 1**).

TABLE 1 | Identification of Clonal Complexes among the studied *Mycobacterium bovis* genomes.

Genomes	Clonal complex
M. bovis strain AF2122/97	European 1
M. bovis strain SP38	European 2
M. bovis strain 1595	European 1
M. bovis strain 30	European 1
M. bovis strain Bz 31150	African 2
M. bovis strain 04-303	European 1
M. bovis strain 09-1191	European 1
M. bovis strain 05-566	European 1
M. bovis strain 05-567	European 1
M. bovis strain 49-09	European 1
M. bovis strain 32-08	European 1
M. bovis strain 18-08C	European 1
M. bovis strain 35	European 2
M. bovis strain 08-08BF2	European 1
M. bovis strain 09-1193	European 1
M. bovis strain 534	European 1
M. bovis strain 0822-11	European 1
M. bovis strain 61-09	European 1
M. bovis strain 45-08b	European 1
M. bovis strain 09-1192	European 1
M. bovis strain 50	European 1
M. bovis strain W-1171	European 1
M. bovis strain MbURU-001	European 1
M. bovis strain MB4	European 2
M. bovis strain B-3222	European 1
M. bovis strain D_10_02315	?
M. bovis strain MB1	European 2
M. bovis strain MB3	?

?, undetermined Clonal Complex.

Nevertheless, few regions with loss of gene synteny were observed, mostly located between positions 2.2 and 4.3 Mb of the genomes. These may suggest an evolutionary process of gene loss through deletions (LSP, including known RDs highlighted in **Figure 1**) or pseudogenization (absence of orthologous genes due to early gene truncation) among virulent *M. bovis* species when compared to *M. tuberculosis* strain H37Rv (**Figure 1A**) or *M. bovis* AF2122/97 (**Figure 1B**). It was also possible to observe that *M. bovis* strain 30 has a synteny pattern slightly different when compared to other virulent *M. bovis*. More studies must be performed to individually identify these regions and their CDSs. It is possible that this information can be used to infer different lineages of *M. bovis*, as previously described in *M. tuberculosis*, in which different lineages have different RD/deletion profiles and specific gene mutations (Rodriguez-Campos et al., 2014).

Pangenome Analysis of *M. bovis* Revealed an Accessory Genome with High Number of CDSs with Unknown Function

Mycobacterium bovis pangenome is composed of 3,970 groups of orthologous CDSs and 1,329 singletons. Among the groups of orthologous CDSs, 3,149 are part of the core genome (i.e., CDSs presented in all 28 *M. bovis* genomes) and are functionally

categorized in **Table 2**. Approximately half of these core groups are described as having unknown function (55.54%), represented mainly by hypothetical and PE/PPE family proteins. It is also important to highlight groups of CDSs involved in housekeeping functions, such as "transcription" (5.65%), and "lipid transport and metabolism" (5.14%), that contain primordial genes required for bacterium maintenance in the host (Raman et al., 2004) and to survival inside the macrophage (Fisher et al., 2002; Russell et al., 2010), respectively.
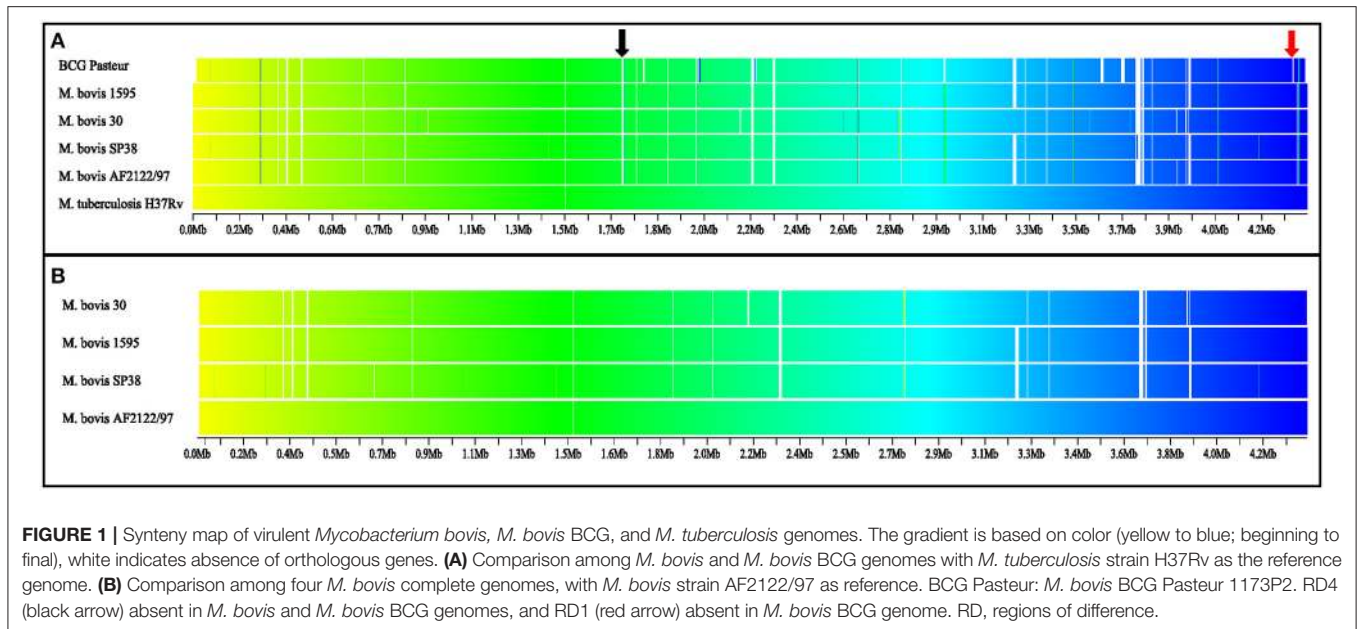
The accessory genome of *M. bovis* [i.e., groups of CDSs present in at least two and up to 27 (~95%) *M. bovis* genomes] is composed of 821 groups of CDSs (**Table 2**). In contrast to the core groups, most of these are categorized as having unknown function (72.96%). Another important function that can be highlighted in addition to transcription and lipid metabolisms is the "mobilome: transposons and prophages" (3.53%), described as mobile genetic elements. These are represented by 29 groups of CDSs, present in 3 to 27 *M. bovis* genomes [the majority (28/29) are present in at least 20 genomes], and characterized as transposases of insertion sequences (IS), phiRV1 phage proteins and resolvases.

In *M. tuberculosis*, there are three main sources of repetitive DNA: duplicated genes and gene families, IS elements, and dispersed non-coding sequences (Cole, 1999). As described above, PE/PPE proteins are important duplicated genes in *M. bovis* strains. In addition, similarly to *M. tuberculosis* H37Rv, 10 IS families were found in *M. bovis* genomes (both in core and accessory genomes), named IS110, IS21, IS256, IS30, IS3, IS5, IS605, IS607, ISL3, and IS1535. The variable copy number of some of these IS in different *M. bovis* strains may indicate possible genetic markers for the epidemiological investigation of bovine tuberculosis.

Two prophage-like elements have been described in *M. tuberculosis* H37Rv genome, named phiRv1 and phiRv2. Their function is still uncertain, but appear to be related to host hypoxia (Fan et al., 2016). All strains of *M. tuberculosis* sequenced before September 2014 have either phiRv1 or phiRv2 (related to RD3 and RD11, respectively) (Fan et al., 2016). As expected (Brosch et al., 2002; Parsons et al., 2002), phiRv1 was found in all but two *M. bovis* genomes (*M. bovis* strains 0822-11 and Bz 31150) and phiRv2, which is generally not identified in this species (Behr et al., 1999; Parsons et al., 2002; Fan et al., 2016), has not been found in any of the analyzed genomes.

Mycobacterium bovis Singleton Analysis Demonstrates a Process of Genomic Decay

OrthoMCL identified 1,329 proteins as being singletons, in other words, present only in one of the 28 *M. bovis* genomes. Of these, 824 (62%) were classified as hypothetical or PE/PPE family proteins and 505 (38%) have known function. Proteins with known function were analyzed using BLASTp and tBLASTn against MTC genomes. Twenty-four proteins were smaller than 30 aa, which is below the threshold allowed in OrthoMCL analysis, and 480/505 (95%) were found to be smaller in size when compared to other *M. bovis* proteins. In addition, there

FIGURE 1 | Synteny map of virulent *Mycobacterium bovis*, *M. bovis* BCG, and *M. tuberculosis* genomes. The gradient is based on color (yellow to blue; beginning to final), white indicates absence of orthologous genes. **(A)** Comparison among *M. bovis* and *M. bovis* BCG genomes with *M. tuberculosis* strain H37Rv as the reference genome. **(B)** Comparison among four *M. bovis* complete genomes, with *M. bovis* strain AF2122/97 as reference. BCG Pasteur: *M. bovis* BCG Pasteur 1173P2. RD4 (black arrow) absent in *M. bovis* and *M. bovis* BCG genomes, and RD1 (red arrow) absent in *M. bovis* BCG genome. RD, regions of difference.

TABLE 2 | *Mycobacterium bovis* orthologous CDSs of the core and accessory genomes categorized by COGs (Cluster of Orthologous Groups).
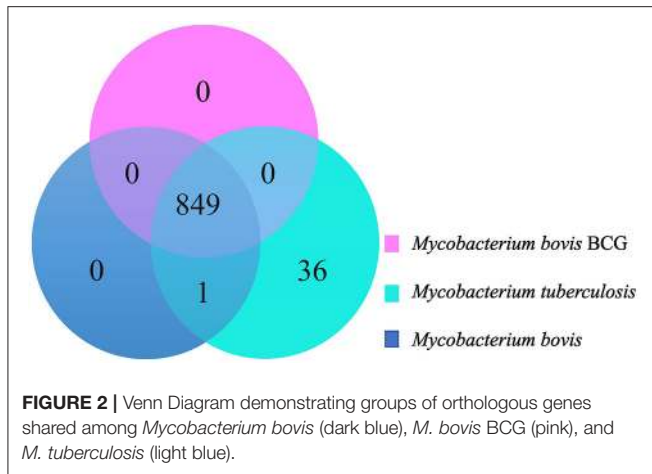
COGs	Core	Accessory
Cell cycle control, cell division, chromosome partitioning	20 (0.63%)	1 (0.12%)
Cell wall/membrane/envelope biogenesis	63 (2.00%)	9 (1.09%)
Cell motility	3 (0.09%)	1 (0.12%)
Post-translational modification, protein turnover, and chaperones	63 (2.00%)	2 (0.24%)
Signal transduction mechanisms	35 (1.11%)	7 (0.85%)
Intracellular trafficking, secretion, and vesicular transport	6 (0.19%)	0 (0.00%)
Defense mechanisms	49 (1.55%)	13 (1.58%)
Mobilome: transposons and prophages	24 (0.76%)	29 (3.53%)
RNA processing and modification	1 (0.03%)	0 (0.00%)
Translation, ribosomal structure and biogenesis	154 (4.89%)	14 (1.70%)
Transcription	178 (5.65%)	23 (2.80%)
Replication, recombination and repair	59 (1.87%)	8 (0.97%)
Energy production and conversion	95 (3.01%)	12 (1.46%)
Amino acid transport and metabolism	113 (3.58%)	15 (1.82%)
Nucleotide transport and metabolism	44 (1.39%)	3 (0.36%)
Carbohydrate transport and metabolism	59 (1.87%)	10 (1.21%)
Coenzyme transport and metabolism	92 (2.92%)	11 (1.34%)
Lipid transport and metabolism	162 (5.14%)	22 (2.68%)
Inorganic ion transport and metabolism	70 (2.22%)	12 (1.46%)
Secondary metabolites biosynthesis, transport, and catabolism	67 (2.12%)	12 (1.46%)
General function prediction only	152 (4.82%)	41 (5.00%)
Function unknown	1,749 (55.54%)	599 (72.96%)
Total of orthologous groups in core genome	3,149	821

One protein can be categorized with more than one COG.

are on average 208 (range: 35 to 363) pseudogenes annotated per genome of *M. bovis*, in a total of 5,620 pseudogenes for 27 genomes (pseudogene information was not available for *M. bovis* strain 30). This corresponds to approximately 5% of

the CDSs in a given *M. bovis* genome. As protein sequences of pseudogenes are not available in genome annotations, these were not included in the OrthoMCL analysis. Thus, these variable annotations may have interfered with the identification

FIGURE 2 | Venn Diagram demonstrating groups of orthologous genes shared among *Mycobacterium bovis* (dark blue), *M. bovis* BCG (pink), and *M. tuberculosis* (light blue).

of singletons. It is also important to highlight that pseudogene annotation might be overestimated in draft genomes; CDSs located in contigs' extremities and without stop codon are annotated as pseudogenes.

From all these 1,329 proteins, only one protein was not identified in BLASTp or BLASTn analyses against *M. bovis* genomes: a membrane protein (341 aa), identified in the genome of *M. bovis* strain Bz 31150 (protein ID: WP_003899748.1), the sole African 2 representative in this study, and with 100% of identity to a probable conserved membrane protein of *M. tuberculosis* (protein ID: Rv3888c). Previous studies have shown that a 2.4 kb region involving membrane, ESX-2 type VII secretion system and ESAT-6 like proteins may or may not be present in *M. bovis* strains (Rauzier et al., 1999; Gey van Pittius et al., 2001). This region has been characterized as ESX-2 locus, one of the five ESAT-6 loci (Bitter et al., 2009; Gröschel et al., 2016). In contrast to many other ESX loci, ESX-2 distribution is restricted to certain mycobacterium species, suggesting that this is a system that arose more recently throughout evolution. Its function, however, remains unknown, as high-density transposon screens have shown that ESX-2 is not required for *in vitro* growth or virulence of *M. tuberculosis* in mice (Gröschel et al., 2016).

The above described characteristics are suggestive of a genomic decay process (gene loss and/or pseudogenization) of *M. bovis*, which is occurring intra-specifically and/or in comparison to other MTC members. The identified singletons are mostly a result of mutations that led to the premature truncation of their genes. Whether or not these are true pseudogenes warrants further investigation. In *M. tuberculosis* strains, pseudogenization has been linked to bacterial host specialization (Bäumler and Fang, 2013; Bolotin and Hershberg, 2015). It is believed that throughout evolution, environmental mycobacteria became specialized host pathogens by losing genes. The impact of this genomic remodeling through genomic decay and PE/PPE family protein duplication on MTC virulence and host adaptability is still unknown.

Groups of Orthologous Proteins among *M. bovis*, *M. bovis* BCG, *M. tuberculosis*: There Are No CDSs Exclusively Present in All *M. bovis* Genomes
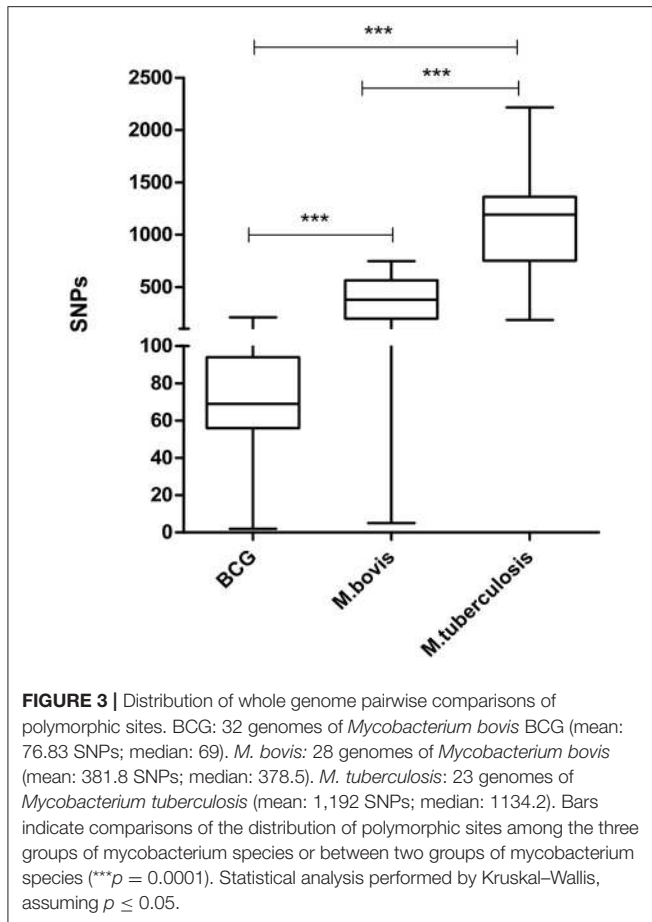
The Venn diagram in **figure 2** illustrates groups of orthologous proteins present in all the studied genomes of *M. bovis*, *M. bovis* BCG, and *M. tuberculosis*. There are no groups of orthologous proteins exclusively present in all 28 *M. bovis* genomes (**Figure 2**). The single group of orthologous proteins present exclusively among *M. bovis* and *M. tuberculosis* genomes and absent in *M. bovis* BCG is annotated as ESAT-6 proteins. These proteins are part of the RD1, an ~9 Kb attenuation-responsible region that codifies 9 CDSs of the ESX-1 locus (PE/PPE protein families and the virulence factor ESAT-6) (Mahairas et al., 1996; Brosch et al., 2002). Moreover, 36 groups of orthologous genes were exclusive of *M. tuberculosis* genomes (**Figure 2**, Supplementary Table 9). A great proportion of these groups (21/26; 58.33%) are related to RD4 and RD7. It is important to highlight the Mce (mammalian cell entry) 3 operon, which is related to the RD7 region and encodes invasin/adhesion proteins that have an important role in *M. tuberculosis* macrophage invasion and survival (Brosch et al., 2002, 2007; Haile et al., 2002; Ahmad et al., 2004). In addition, three groups identified as anti-anti-sigma factor RsfB, ESAT-6, and Cytochrome P450 (Rv3518) were found to be smaller in size (with or without enough amino acid sequence identity) in *M. bovis* genomes, precluding their clustering as orthologs, while other two (Pyridoxine/pyridoxamine 5′-phosphate oxidase and 6-phosphogluconate dehydrogenase Gnd1) were found to be smaller in size in *M. tuberculosis* genomes. These observations demonstrate once more the possibility of a genomic decay process or genome remodeling in *M. bovis* and *M. tuberculosis*.

Selected *M. tuberculosis* Strains Showed a Greater Number of Polymorphic Sites When Compared to Virulent *M. bovis* and *M. bovis* BCG

A total of 3,448 polymorphic sites were detected among virulent *M. bovis*, 1,088 among *M. bovis* BCG, and 8,335 among *M. tuberculosis* genomes. Pairwise analysis of each group indicates that *M. tuberculosis* has the highest number of polymorphic sites, followed by virulent *M. bovis*, and *M. bovis* BCG ($p \leq 0.05$) (**Figure 3**). Even though *M. bovis* is considered a generalist species (i.e., it is able to infect multiple hosts), this mycobacterium demonstrated lower genetic variability when compared to *M. tuberculosis*. One possible explanation is that the *M. tuberculosis* sample analyzed herein is more geographically representative of the world strains when compared to *M. bovis*. However, it is also possible that *M. tuberculosis* strains present higher genetic variability due to prolonged co-evolution with human populations of various genetic backgrounds (Gagneux, 2012) or for being an evolutionarily older lineage of MTC.

As expected, the number of polymorphic sites was substantially lower in *M. bovis* BCG when compared to virulent *M. bovis* and *M. tuberculosis* (**Figure 3**). *M. bovis* BCG

FIGURE 3 | Distribution of whole genome pairwise comparisons of polymorphic sites. BCG: 32 genomes of *Mycobacterium bovis* BCG (mean: 76.83 SNPs; median: 69). *M. bovis:* 28 genomes of *Mycobacterium bovis* (mean: 381.8 SNPs; median: 378.5). *M. tuberculosis:* 23 genomes of *Mycobacterium tuberculosis* (mean: 1,192 SNPs; median: 1134.2). Bars indicate comparisons of the distribution of polymorphic sites among the three groups of mycobacterium species or between two groups of mycobacterium species (***p = 0.0001). Statistical analysis performed by Kruskal–Wallis, assuming $p \leq 0.05$.

strains are solely subjected to artificial *in vitro* selection and originated from a single isolate. Nevertheless, the presence of variability in BCG strains should not be neglected, as they can reflect in different vaccine success rates (Garcia Pelayo et al., 2009).

M. tuberculosis Lineages Are More Genetically Diverse Than *M. bovis* Clonal Complexes

Genomes of *M. tuberculosis* analyzed herein represent all four lineages of this species. The sole attempt to define *M. bovis* strains into lineages comes from the recent definition of Clonal Complexes. Selected *M. bovis* genomes represent 3 out of the 4 Clonal Complexes currently described (African 1 is the only missing complex). The evaluation of polymorphic sites as a function of the number of genomes being analyzed indicates that *M. tuberculosis* lineages are more diverse than *M. bovis* Clonal Complexes and possibly animal-adapted MTC species (*M. bovis, M. caprae, M. microti, M. mungi, M. suricattae,* and *M. orygis*) (**Figure 4**). This discrepancy as well as the definition of *M. bovis* lineages should be explored in future studies involving a more geographically representative *M. bovis* sample. And finally, it is clear that the diversity of MTC will continue to grow as additional genomes are
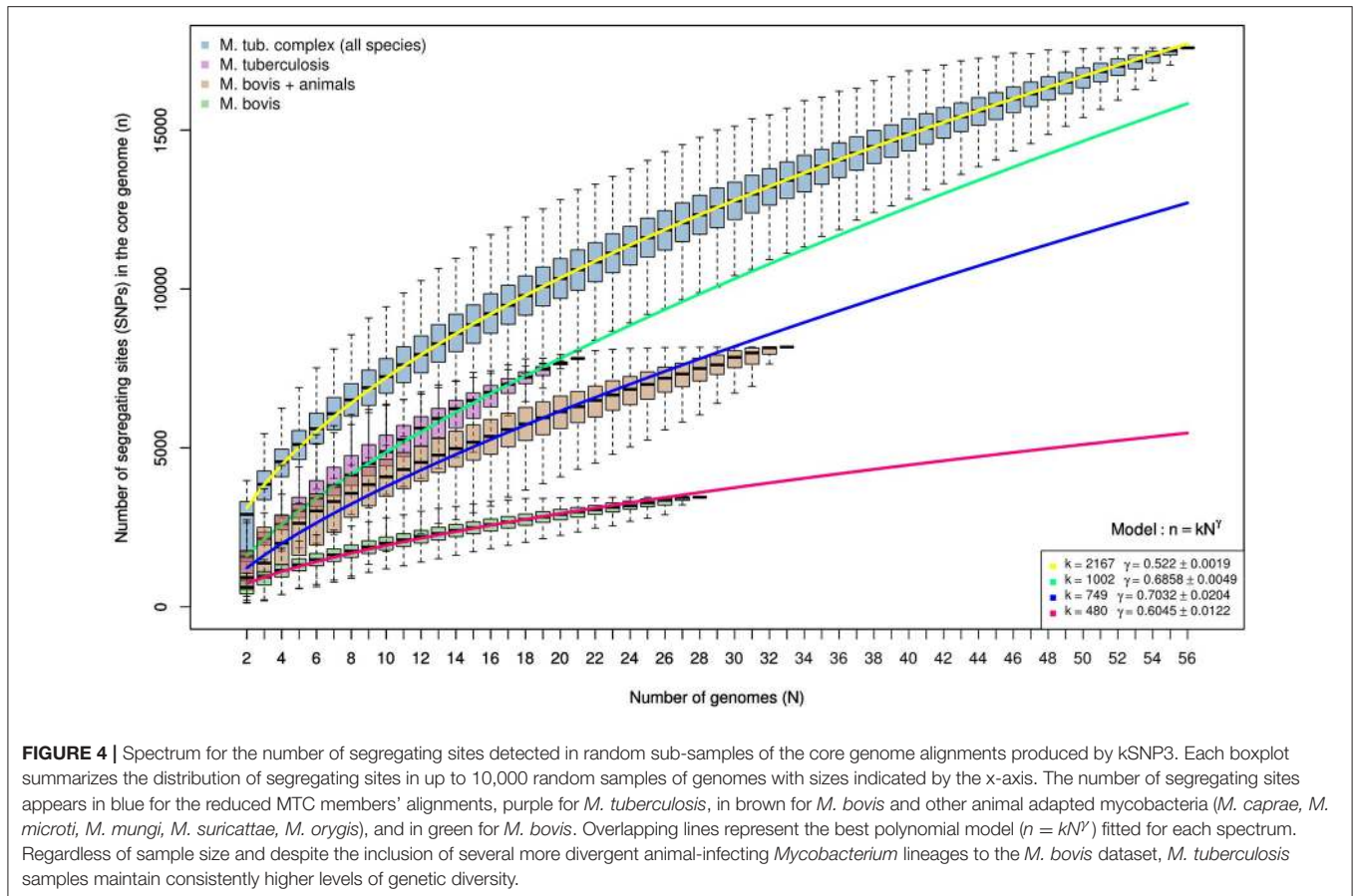
sequenced, having a direct impact on the fight against these pathogens.

Polymorphic Sites Annotation: *M. bovis* BCG Is Subjected to Distinct Evolutionary Pressure When Compared to Virulent Mycobacteria

Polymorphic sites annotation of *M. bovis*, *M. bovis* BCG, and *M. tuberculosis* are shown in Supplementary Tables 10–12. The proportion of synonymous (40.48–51.10%) vs. nonsynonymous (48.90–59.52%) mutations was similar in all three groups. Of the 3,448 polymorphic sites identified in virulent *M. bovis* genomes, 2,804 (81.32%) are located within CDSs, 342 (9.92%) were not in annotated genome, 24 (0.70%) in non-protein coding regions, and 278 (8.06%) in unannotated regions. Similarly, of the 8,335 polymorphic sites identified in *M. tuberculosis* genomes, most of the mutations were located within CDSs (8,293, 99.50%). Only two (0.02%) polymorphic sites were not in annotated genomes, 10 (0.12%) in non-protein coding regions, and 30 (0.36%) in unannotated regions. On the other hand, from the 1,088 polymorphic sites described in *M. bovis* BCGs genomes, 757 (69.58%) of the mutations were identify within CDSs, 178 (16.36%) not in annotated genome, 52 (4.78%) in non-protein coding regions, and 101 (9.28%) in unannotated regions. The higher number of mutations in non-protein coding regions in *M. bovis* BCG when compared to virulent *M. bovis* and *M. tuberculosis* may be a result of the *in vitro* selective pressure that *M. bovis* BCG strains undergo. As intergenic regions may harbor regulatory regions, it is likely that mutations in such sites may lead to alterations in gene expression in these attenuated BCG strains.

The polymorphic site categorization according to COGs showed that *M. bovis* and *M. tuberculosis* have similar functional mutation patterns (Supplementary Tables 10, 11). In addition to CDSs categorized as having unknown and general functions, CDSs involved in "secondary metabolites biosynthesis, transport, and catabolism," "lipid transport and metabolism," "energy production and conversion," and "amino acid transport and metabolism" represent a great proportion of the mutations. These shared patterns can be due to the intra-host natural selection pressure *M. tuberculosis* and *M. bovis* undergo in their lifecycle. Nevertheless, it is not possible to ignore that phenotype differences between both species may be due the presence of specific SNPs in certain genes (Filliol et al., 2006; Bigi et al., 2016, 2017).

In contrast to virulent mycobacteria, a significant proportion of polymorphic sites of *M. bovis* BCG were characterized as being in CDSs involved with "energy production and conversion," followed by "cell cycle control, cell division, chromosome partitioning," and "signal transduction mechanisms." These characteristics reflect the *in vitro* selective pressure *M. bovis* BCG undergoes during its growth for vaccine production, contrasting with the intra-host environment encountered by virulent mycobacteria. Inside macrophages, virulent mycobacteria usually

FIGURE 4 | Spectrum for the number of segregating sites detected in random sub-samples of the core genome alignments produced by kSNP3. Each boxplot summarizes the distribution of segregating sites in up to 10,000 random samples of genomes with sizes indicated by the x-axis. The number of segregating sites appears in blue for the reduced MTC members' alignments, purple for *M. tuberculosis*, in brown for *M. bovis* and other animal adapted mycobacteria (*M. caprae, M. microti, M. mungi, M. suricattae, M. orygis*), and in green for *M. bovis*. Overlapping lines represent the best polynomial model ($n = kN^{\gamma}$) fitted for each spectrum. Regardless of sample size and despite the inclusion of several more divergent animal-infecting *Mycobacterium* lineages to the *M. bovis* dataset, *M. tuberculosis* samples maintain consistently higher levels of genetic diversity.

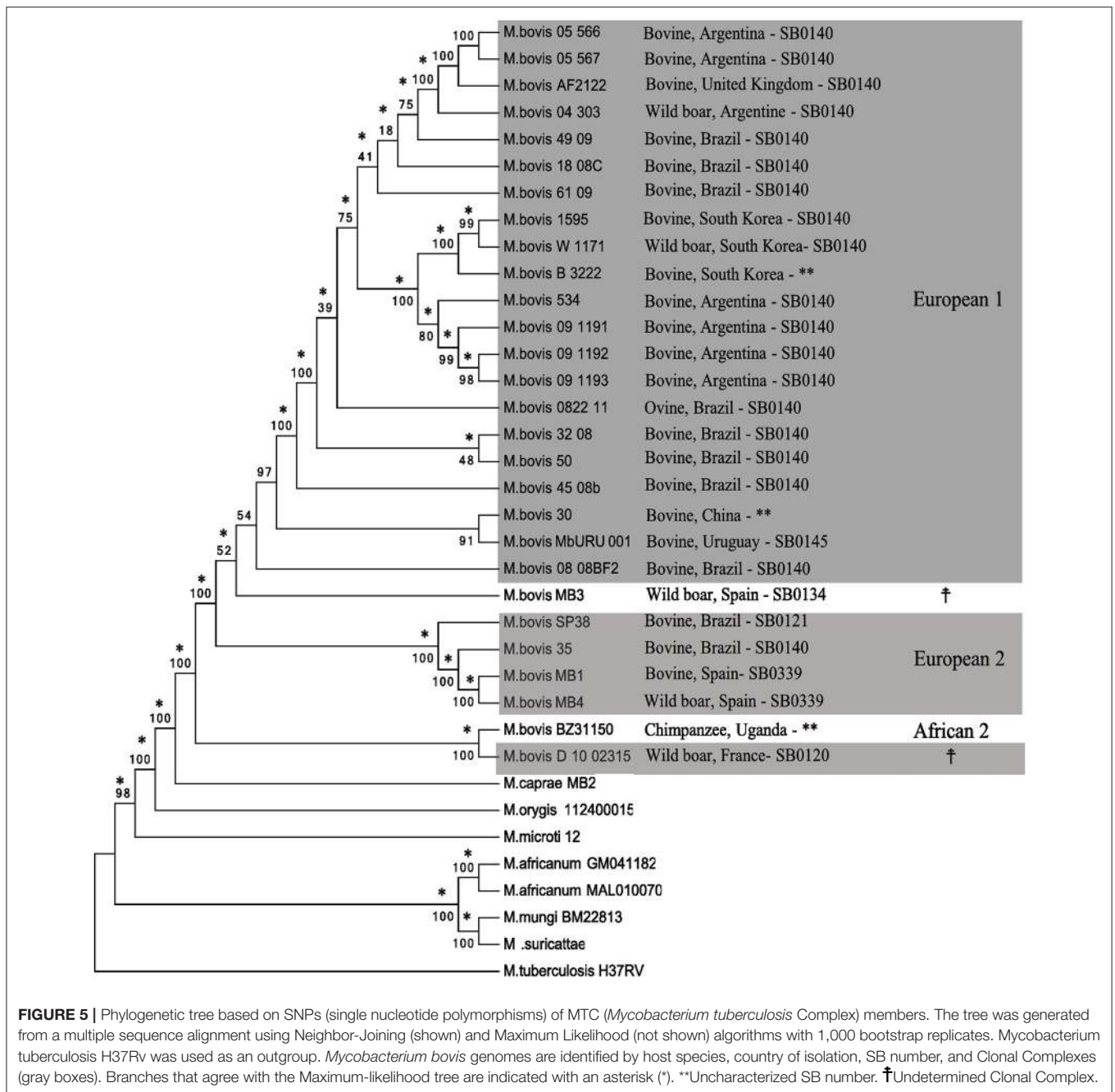rely on lipids and cholesterol for growth (Pandey and Sassetti, 2008; Lee et al., 2013).

Phylogenomics: *M. bovis* Lineages Form Clades According to Their Clonal Complexes

The phylogenetic tree based on polymorphic sites supports the classification of *M. bovis* into clades according to their Clonal Complexes, and not according to the host in which they were isolated from (**Figure 5**). The identification and analyses of SNPs in MTC genomes is an important tool used for their phylogenetic reconstruction and discrimination among different lineages, as described for *M. tuberculosis,* and as a marker to distinguish MTC species (Gutacker et al., 2002; Namouchi et al., 2012; Coll et al., 2014; Coscolla and Gagneux, 2014; Phelan et al., 2016). There are few studies regarding the comparative genomics of *M. bovis* (Garcia Pelayo et al., 2009; Joshi et al., 2012; de la Fuente et al., 2015; Patané et al., 2017) and, to our knowledge, this is the first study involving SNP phylogenomics of *M. bovis* genomes deposited in GenBank, complete or draft, in association with Clonal Complexes. Even though representatives of the Clonal Complex African 1 were not present in our sample, these results support the use of SNPs as a possible phylogenetic marker of *M. bovis* in association with Clonal Complexes,

which can be explored to distinguish this species into different lineages.

Final Considerations

This is the first study to sequence and close a complete genome of a Brazilian strain of *M. bovis*. The *M. bovis* strain SP38 genome demonstrated traditional characteristics of a tuberculous mycobacterium genome, and was categorized as Clonal Complex European 2, the first report of this Clonal Complex in Brazil. Comparative genomics of *M. bovis* SP38 and other *M. bovis* genomes deposited in GenBank demonstrated an active process of intra-species genomic decay. The remodeling of mycobacterium genomes through gene loss (deletion or pseudogenization) and duplication (particularly of PE/PPE family genes) must be comparatively explored among MTC species regarding their impact on host adaption and virulence. Also, the identification of *M. bovis* genomes that fail to be classified within current Clonal Complexes indicates that the diversity of *M. bovis* may be larger than previously thought or that current classification needs to be reviewed. The detected smaller genetic diversity of *M. bovis* when compared to *M. tuberculosis* genomes contrasts with the generalist behavior of this species. In fact, *M. tuberculosis* lineages appeared more genetically diverse than animal-adapted MTC species. This finding highlights the need to better understand host-pathogen

FIGURE 5 | Phylogenetic tree based on SNPs (single nucleotide polymorphisms) of MTC (*Mycobacterium tuberculosis* Complex) members. The tree was generated from a multiple sequence alignment using Neighbor-Joining (shown) and Maximum Likelihood (not shown) algorithms with 1,000 bootstrap replicates. Mycobacterium tuberculosis H37Rv was used as an outgroup. *Mycobacterium bovis* genomes are identified by host species, country of isolation, SB number, and Clonal Complexes (gray boxes). Branches that agree with the Maximum-likelihood tree are indicated with an asterisk (*). **Uncharacterized SB number. †Undetermined Clonal Complex.

co-evolution in the context of genetically homogeneous and/or diverse host populations. And finally, the phylogeny based on SNPs can be used as a complementary tool for *M. bovis* Clonal Complex identification and explored as a gold-standard for lineages classification of *M. bovis* strains, as performed for *M. tuberculosis*.

AUTHORS CONTRIBUTIONS

CZ, PB, AdS, RdS, CI, NC and AG: conceived and designed experiments; CZ, RdS, NC and AG: analyzed the data; PB, JF, MH

and AG: contributed reagents, materials, analysis tools; CZ, RdS and AG, wrote the paper; CZ, PB, AdS, RdS, CI, JF, NC, MH and AG: approved the paper.

FUNDING

had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb. 2017.02389/full#supplementary-material

REFERENCES

Aguilar León, D., Zumárraga, M. J., Jiménez Oropeza, R., Gioffré, A. K., Bernardelli, A., Orozco Estévez, H., et al. (2009). *Mycobacterium bovis* with different genotypes and from different hosts induce dissimilar immunopathological lesions in a mouse model of tuberculosis. *Clin. Exp. Immunol.* 157, 139–147. doi: 10.1111/j.1365-2249.2009.03923.x

Ahmad, S., El-Shazly, S., Mustafa, A. S., and Al-Attiyah, R. (2004). Mammalian cell-entry proteins encoded by the mce3 operon of *Mycobacterium tuberculosis* are expressed during natural infection in humans. *Scand. J. Immunol.* 60, 382–391. doi: 10.1111/j.0300-9475.2004.01490.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Alva, V., Nam, S.-Z., Söding, J., and Lupas, A. N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* 44, W410–W415. doi: 10.1093/nar/gkw348

Andrade, M. R., Amaral, E. P., Ribeiro, S. C., Almeida, F. M., Peres, T. V., Lanes, V., et al. (2012). Pathogenic *Mycobacterium bovis* strains differ in their ability to modulate the proinflammatory activation phenotype of macrophages. *BMC Microbiol.* 12:166. doi: 10.1186/1471-2180-12-166

Arkin, A. P., Stevens, R. L., Cottingham, R. W., Maslov, S., Henry, C. S., Dehal, P., et al. (2016). The DOE systems biology knowledgebase (KBase). *bioRxiv.* doi: 10.1101/096354

Bäumler, A., and Fang, F. C. (2013). Host specificity of bacterial pathogens. *Cold Spring Harb. Perspect. Med.* 3:a010041. doi: 10.1101/cshperspect.a010041

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., et al. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284, 1520–1523. doi: 10.1126/science.284.5419.1520

Bemer-Melchior, P., and Drugeon, H. B. (1999). Inactivation of *Mycobacterium tuberculosis* for DNA typing analysis. *J. Clin. Microbiol.* 37, 2350–2351.

Berg, S., Garcia-Pelayo, M. C., Müller, B., Hailu, E., Asiimwe, B., Kremer, K., et al. (2011). African 2, a clonal complex of *Mycobacterium bovis* epidemiologically important in East Africa. *J. Bacteriol.* 193, 670–678. doi: 10.1128/JB.00750-10

Biek, R., O'Hare, A., Wright, D., Mallon, T., McCormick, C., Orton, R. J., et al. (2012). Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathog.* 8:e1003008. doi: 10.1371/journal.ppat.1003008

Bigi, M. M., Blanco, F. C., Araújo, F. R., Thacker, T. C., Zumárraga, M. J., Cataldi, A. A., et al. (2016). Polymorphisms of 20 regulatory proteins between *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *Microbiol. Immunol.* 60, 552–560. doi: 10.1111/1348-0421.12402

Bigi, M. M., Lopez, B., Blanco, F. C., Sasiain, M. D. C., De la Barrera, S., Marti, M. A., et al. (2017). Single nucleotide polymorphisms may explain the contrasting phenotypes of two variants of a multidrug-resistant *Mycobacterium tuberculosis* strain. *Tuberculosis* 103, 28–36. doi: 10.1016/j.tube.2016.12.007

Bitter, W., Houben, E. N. G., Bottai, D., Brodin, P., Brown, E. J., Cox, J. S., et al. (2009). Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog.* 5:e1000507. doi: 10.1371/journal.ppat.1000507

Blanco, F. C., Nunez-García, J., García-Pelayo, C., Soria, M., Bianco, M. V., Zumárraga, M., et al. (2009). Differential transcriptome profiles of attenuated and hypervirulent strains of *Mycobacterium bovis*. *Microbes Infect.* 11, 956–963. doi: 10.1016/j.micinf.2009.06.006

Bolotin, E., and Hershberg, R. (2015). Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. *Genome Biol. Evol.* 7, 2173–2187. doi: 10.1093/gbe/evv135

Brosch, R., Gordon, S. V., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., et al. (2007). Genome plasticity of BCG and impact on vaccine efficacy. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5596–5601. doi: 10.1073/pnas.0700869104

Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., et al. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3684–3689. doi: 10.1073/pnas.052548299

Carvalho, R. C. T., Vasconcellos, S. E. G., Issa, Mde A., Soares Filho, P. M., Mota, P. M., Araújo, F. R., et al. (2016). Molecular typing of *Mycobacterium bovis* from cattle reared in midwest Brazil. *PLoS ONE* 11:e0162459. doi: 10.1371/journal.pone.0162459

Centro Panamericano de Zoonosis (1985). *Manual de normas y procedimientos tecnicos para la bacteriologia de la tuberculosis.* Buenos Aires: Organizacion Panamericana de Ia Salud.

Cole, S. T. (1999). Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett.* 452, 7–10. doi: 10.1016/S0014-5793(99)00536-0

Cole, S. T. (2002). Comparative and functional genomics of the *Mycobacterium tuberculosis* complex a. *Microbiology* 148, 2919–2928. doi: 10.1099/00221287-148-10-2919

Coll, F., Mallard, K., Preston, M. D., Bentley, S., Parkhill, J., McNerney, R., et al. (2012). SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. *Bioinformatics* 28, 2991–2993. doi: 10.1093/bioinformatics/bts544

Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., et al. (2014). A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* 5:4812. doi: 10.1038/ncomms5812

Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* 26, 431–444. doi: 10.1016/j.smim.2014.09.012

Crabtree, J., Angiuoli, S. V., Wortman, J. R., and White, O. R. (2007). Sybil: methods and software for multiple genome comparison and visualization. *Methods Mol. Biol.* 408, 93–108. doi: 10.1007/978-1-59745-547-3_6

Crispell, J., Zadoks, R. N., Harris, S. R., Paterson, B., Collins, D. M., De-Lisle, G. W., et al. (2017). Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. *BMC Genomics* 18:180. doi: 10.1186/s12864-017-3569-x

de la Fuente, J., Díez-Delgado, I., Contreras, M., Vicente, J., Cabezas-Cruz, A., Tobes, R., 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. doi: 10.1371/journal.pntd.0004232

Dias, R. A., Ulloa-Stanojlovic, F. M., Belchior, A. P. C., Ferreira, R. D. S., Gonçalves, R. C., de Aguiar, R. S. C. B., et al. (2016). Prevalence and risk factors for bovine tuberculosis in the State of São Paulo, Brazil. *Semin. Ciências Agrárias* 37:3673. doi: 10.5433/1679-0359.2016v37n5Supl2p3673

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

El-Sayed, A., El-Shannat, S., Kamel, M., Castañeda-Vazquez, M. A., and Castañeda-Vazquez, H. (2016). Molecular epidemiology of *Mycobacterium bovis* in humans and cattle. *Zoonoses Pub. Heal.* 63, 251–264. doi: 10.1111/zph.12242

Faksri, K., Xia, E., Tan, J. H., Teo, Y.-Y., and Ong, R. T.-H. (2016). *In silico* region of difference (RD) analysis of *Mycobacterium tuberculosis* complex from sequence reads using RD-Analyzer. *BMC Genomics* 17:847. doi: 10.1186/s12864-016-3213-1

Fan, X., Abd Alla, A. A. E., and Xie, J. (2016). Distribution and function of prophage phiRv1 and phiRv2 among *Mycobacterium tuberculosis* complex. *J. Biomol. Struct. Dyn.* 34, 233–238. doi: 10.1080/07391102.2015.1022602

Ferreira Neto, J. S., da Silveira, G. B., Rosa, B. M., Gonçalves, V. S. P., Grisi-Filho, J. H. H., Amaku, M., et al. (2016). Analysis of 15 years of the national program for the control and eradication of animal brucellosis and tuberculosis, Brazil. *Semin. Ciências Agrárias* 37:3385. doi: 10.5433/1679-0359.2016v37n5Supl2p3385

Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., Bobadilla del Valle, M., et al. (2006). Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP s. *J. Bacteriol.* 188, 759–772. doi: 10.1128/JB.188.2.759-772.2006

Fishbein, S., van Wyk, N., Warren, R. M., and Sampson, S. L. (2015). Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity. *Mol. Microbiol.* 96, 901–916. doi: 10.1111/mmi.12981

Fisher, M. A., Plikaytis, B. B., and Shinnick, T. M. (2002). Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J. Bacteriol.* 184, 4025–4032. doi: 10.1128/JB.184.14.4025-4032.2002

Gagneux, S. (2012). Host-pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 850–859. doi: 10.1098/rstb.2011.0316

Galagan, J. E. (2014). Genomic insights into tuberculosis. *Nat. Rev. Genet.* 15, 307–320. doi: 10.1038/nrg3664

Garbaccio, S., Macias, A., Shimizu, E., Paolicchi, F., Pezzone, N., Magnano, G., et al. (2014). Association between spoligotype-VNTR types and virulence of *Mycobacterium bovis* in cattle. *Virulence* 5, 297–302. doi: 10.4161/viru.27193

Garcia Pelayo, M. C., Uplekar, S., Keniry, A., Mendoza Lopez, P., Garnier, T., Nunez Garcia, J., et al. (2009). A comprehensive survey of single nucleotide polymorphisms (SNPs) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent M. bovis strains and M. *Infect. Immun.* 77, 2230–2238. doi: 10.1128/IAI.01099-08

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878. doi: 10.1093/bioinformatics/btv271

Gey van Pittius, N. C., Gamieldien, J., Hide, W., Brown, G. D., Siezen, R. J., and Beyers, A. D. (2001). The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol.* 2:research0044.1. doi: 10.1186/gb-2001-2-10-research0044

Glaser, L., Carstensen, M., Shaw, S., Robbe-Austerman, S., Wunschmann, A., Grear, D., et al. (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:e0145735. doi: 10.1371/journal.pone.0145735

Gröschel, M. I., Sayes, F., Simeone, R., Majlessi, L., and Brosch, R. (2016). ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat. Rev. Microbiol.* 14, 677–691. doi: 10.1038/nrmicro.2016.131

Guimaraes, A. M., Zimpel, C. K., Ikuta, C. Y., do Nascimento, N. C., Dos Santos, A. P., Messick, J. B., et al. (2015). Draft genome sequence of *Mycobacterium bovis* strain SP38, a pathogenic bacterium isolated from a Bovine in Brazil. *Genome Announc.* 3:e00511-15. doi: 10.1128/genomeA.00511-15

Gutacker, M. M., Smoot, J. C., Migliaccio, C. A. L., Ricklefs, S. M., Hua, S., Cousins, D. V., et al. (2002). Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* 162, 1533–1543.

Haile, Y., Caugant, D. A., Bjune, G., and Wiker, H. G. (2002). *Mycobacterium tuberculosis* mammalian cell entry operon (*mce*) homologs in *Mycobacterium*

other than tuberculosis (MOTT). *FEMS Immunol. Med. Microbiol.* 33, 125–132. doi: 10.1111/j.1574-695X.2002.tb00581.x

Ikuta, C. Y., Morato, F., Souza, G. O., de, Heinemann, M. B., Amaku, M., Ferreira Neto, J. S., et al. (2016). Influence of the incubation conditions on culture media to optimize primary isolation of *Mycobacterium bovis*. *Semin. Ciências Agrárias* 37:3693. doi: 10.5433/1679-0359.2016v37n5Supl2p3693

Ilina, E. N., Shitikov, E. A., Ikryannikova, L. N., Alekseev, D. G., Kamashev, D. E., Malakhova, M. V., et al. (2013). Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia. *PLoS ONE* 8:e56577. doi: 10.1371/journal.pone.0056577

Joshi, D., Harris, N. B., Waters, R., Thacker, T., Mathema, B., Krieswirth, B., et al. (2012). Single nucleotide polymorphisms in the *Mycobacterium bovis* genome resolve phylogenetic relationships. *J. Clin. Microbiol.* 50, 3853–3861. doi: 10.1128/JCM.01499-12

Kato-Maeda, M., Rhee, J. T., Gingeras, T. R., Salamon, H., Drenkow, J., Smittipat, N., et al. (2001). Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* 11, 547–554. doi: 10.1101/gr.166401

Kumar, A., and Kaur, J. (2014). Primer based approach for PCR amplification of high GC content gene: *Mycobacterium* gene as a model. *Mol. Biol. Int.* 2014:937308. doi: 10.1155/2014/937308

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lee, W., VanderVen, B. C., Fahey, R. J., and Russell, D. G. (2013). Intracellular *Mycobacterium tuberculosis* exploits host-derived fatty acids to limit metabolic stress. *J. Biol. Chem.* 288, 6788–6800. doi: 10.1074/jbc.M112.445056

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liu, F., Hu, Y., Wang, Q., Li, H., Gao, G. F., Liu, C., et al. (2014). Comparative genomic analysis of *Mycobacterium tuberculosis* clinical isolates. *BMC Genomics* 15:469. doi: 10.1186/1471-2164-15-469

Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C., and Stover, C. K. (1996). Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J. Bacteriol.* 178, 1274–1282. doi: 10.1128/jb.178.5.1274-1282.1996

McGuire, A., Weiner, B., Park, S., Wapinski, I., Raman, S., Dolganov, G., et al. (2012). Comparative analysis of *Mycobacterium* and related actinomycetes yields insight into the evolution of *Mycobacterium tuberculosis* pathogenesis. *BMC Genomics* 13:120. doi: 10.1186/1471-2164-13-120

Meikle, V., Bianco, M. V., Blanco, F. C., Gioffré, A., Garbaccio, S., Vagnoni, L., et al. (2011). Evaluation of pathogenesis caused in cattle and guinea pig by a *Mycobacterium bovis* strain isolated from wild boar. *BMC Vet. Res.* 7:37. doi: 10.1186/1746-6148-7-37

Morato, F., Ikuta, C. Y., Gonçales, A. P., Souza, G. O., de, Amaku, M., Cortez, A., et al. (2016). Evaluation of optimal storage time of tuberculous lesions stored in sodium borate. *Semin. Ciências Agrárias* 37:3685. doi: 10.5433/1679-0359.2016v37n5Supl2p3685

Müller, B., Hilty, M., Berg, S., Garcia-Pelayo, M. C., Dale, J., Boschiroli, M. L., et al. (2009). African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. *J. Bacteriol.* 191, 1951–1960. doi: 10.1128/JB.01590-08

Namouchi, A., Didelot, X., Schöck, U., Gicquel, B., and Rocha, E. P. C. (2012). After the bottleneck: genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. *Genome Res.* 22, 721–734. doi: 10.1101/gr.129544.111

Olea-Popelka, F., Muwonge, A., Perera, A., Dean, A. S., Mumford, E., Erlacher-Vindel, E., et al. (2017). Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis* - a call for action. *Lancet Infect. Dis.* 17, e21–e25. doi: 10.1016/S1473-3099(16)30139-6

Pandey, A. K., and Sassetti, C. M. (2008). Mycobacterial persistence requires the utilization of host cholesterol. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4376–4380. doi: 10.1073/pnas.0711159105

Parsons, L. M., Brosch, R., Cole, S. T., Somoskövi, A., Loder, A., Bretzel, G., et al. (2002). Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. *J. Clin. Microbiol.* 40, 2339–2345. doi: 10.1128/JCM.40.7.2339-2345.2002

Patané, J. S. L., Martins, J., Castelão, A. B., Nishibe, C., Montera, L., Bigi, F., et al. (2017). Patterns and processes of *Mycobacterium bovis* evolution revealed by phylogenomic analyses. *Genome Biol. Evol.* 9, 521–535. doi: 10.1093/gbe/evx022

Periwal, V., Patowary, A., Vellarikkal, S. K., Gupta, A., Singh, M., Mittal, A., et al. (2015). Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome. *PLoS ONE* 10:e0122979. doi: 10.1371/journal.pone.0122979

Phelan, J. E., Coll, F., Bergval, I., Anthony, R. M., Warren, R., Sampson, S. L., et al. (2016). Recombination in pe/ppe genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. *BMC Genomics* 17:151. doi: 10.1186/s12864-016-2467-y

Raman, S., Hazra, R., Dascher, C. C., and Husson, R. N. (2004). Transcription regulation by the *Mycobacterium tuberculosis* alternative sigma factor SigD and its role in virulence. *J. Bacteriol.* 186, 6605–6616. doi: 10.1128/JB.186.19.6605-6616.2004

Rauzier, J., Gormley, E., Gutierrez, M. C., Kassa-Kelembho, E., Sandall, L. J., Dupont, C., et al. (1999). A novel polymorphic genetic locus in members of the Mycobacterium tuberculosis complex. *Microbiology* 145, 1695–1701. doi: 10.1099/13500872-145-7-1695

Ribeiro, L. A., Gonçalves, V. S. P., Francisco, P. F. C., de Alencar Mota, A. L. A., Nascimento, G. T. do, Licurgo, J. B., et al. (2016). Epidemiological status of bovine tuberculosis in the Federal District of Brazil. *Semin. Ciências Agrárias* 37:3561. doi: 10.5433/1679-0359.2016v37n5Supl2p3561

Riley, D. R., Angiuoli, S. V., Crabtree, J., Dunning Hotopp, J. C., and Tettelin, H. (2012). Using Sybil for interactive comparative genomics of microbes on the web. *Bioinformatics* 28, 160–166. doi: 10.1093/bioinformatics/btr652

Rocha, V. C. F. (2013). *Avaliação do* Spoligotyping, MIRU-VNTR e Multispacer Sequence Typing na discriminação de isolados autóctones de *Mycobacterium bovis*.

Rodriguez-Campos, S., Navarro, Y., Romero, B., de Juan, L., Bezos, J., Mateos, A., et al. (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, 3658–3665. doi: 10.1128/JCM.01271-13

Rodriguez-Campos, S., Schürch, A. C., Dale, J., Lohan, A. J., Cunha, M. V., Botelho, A., et al. (2012). European 2 – A clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. *Infect. Genet. Evol.* 12, 866–872. doi: 10.1016/j.meegid.2011.09.004

Rodriguez-Campos, S., Smith, N. H., Boniotti, M. B., and Aranaz, A. (2014). Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics and legislation of bovine tuberculosis. *Res. Vet. Sci.* 97 (Suppl. 1), S5–S19. doi: 10.1016/j.rvsc.2014.02.009

Rue-Albrecht, K., Magee, D. A., Killick, K. E., Nalpas, N. C., Gordon, S. V., and MacHugh, D. E. (2014). Comparative functional genomics and the bovine macrophage response to strains of the mycobacterium genus. *Front. Immunol.* 5:536. doi: 10.3389/fimmu.2014.00536

Russell, D. G., VanderVen, B. C., Lee, W., Abramovitch, R. B., Kim, M., Homolka, S., et al. (2010). *Mycobacterium tuberculosis* wears what it eats. *Cell Host Microbe* 8, 68–76. doi: 10.1016/j.chom.2010.06.002

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19, 1117–1123. doi: 10.1101/gr.089532.108

Smith, N. H., Berg, S., Dale, J., Allen, A., Rodriguez, S., Romero, B., et al. (2011). European 1: a globally important clonal complex of *Mycobacterium bovis*. *Infect. Genet. Evol.* 11, 1340–1351. doi: 10.1016/j.meegid.2011.04.027

Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R., and Gordon, S. V. (2009). Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* 7, 537–544. doi: 10.1038/nrmicro2165

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

van Hijum, S. A. F. T., Zomer, A. L., Kuipers, O. P., and Kok, J. (2005). Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. *Nucleic Acids Res.* 33, W560–W566. doi: 10.1093/nar/gki356

van Soolingen, D., de Haas, P. E., Hermans, P. W., and van Embden, J. D. (1994). DNA fingerprinting of *Mycobacterium tuberculosis*. *Methods Enzymol.* 235, 196–205. doi: 10.1016/0076-6879(94)35141-4

Voskuil, M. I., Schnappinger, D., Rutherford, R., Liu, Y., and Schoolnik, G. K. (2004). Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis* 84, 256–262. doi: 10.1016/j.tube.2003.12.014

Wanzala, S. I., Nakavuma, J., Travis, D. A., Kia, P., Ogwang, S., and Sreevatsan, S. (2015). Draft genome sequences of *Mycobacterium bovis* BZ 31150 and *Mycobacterium bovis* B2 7505, pathogenic bacteria isolated from archived captive animal bronchial washes and human sputum samples in Uganda. *Genome Announc.* 3, e01102–e01115. doi: 10.1128/genomeA.01102-15

Warren, R. M., Gey Van Pittius, N. C., Barnard, M., Hesseling, A., Engelke, E., De Kock, M., et al. (2006). Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference. *Int. J. Tuberc. Lung Dis.* 10, 818–822.

Wences, A. H., Schatz, M. C., Hernandez Wences, A., and Schatz, M. C. (2015). Metassembler: merging and optimizing *de novo* genome assemblies. *Genome Biol.* 16:207. doi: 10.1186/s13059-015-0764-4

Wright, D. M., Allen, A. R., Mallon, T. R., McDowell, S. W. J., Bishop, S. C., Glass, E. J., et al. (2013). Field-isolated genotypes of *Mycobacterium bovis* vary in virulence and influence case pathology but do not affect outbreak size. *PLoS ONE* 8:e74503. doi: 10.1371/journal.pone.0074503

Xia, E., Teo, Y.-Y., and Ong, R. T.-H. (2016). SpoTyping: fast and accurate *in silico Mycobacterium* spoligotyping from sequence reads. *Genome Med.* 8:19. doi: 10.1186/s13073-016-0270-7

Zumárraga, M. J., Arriaga, C., Barandiaran, S., Cobos-Marín, L., de Waard, J., Estrada-Garcia, I., et al. (2013). Understanding the relationship between *Mycobacterium bovis* spoligotypes from cattle in Latin American Countries. *Res. Vet. Sci.* 94, 9–21. doi: 10.1016/j.rvsc.2012.07.012