



HAL
open science

Features of *Mycobacterium bovis* Complete Genomes Belonging to 5 Different Lineages

Ciriac Charles, Cyril Conde, Fabien Vorimore, Thierry Cochard, Lorraine Michelet, Maria Laura Boschioli, Franck Biet

► **To cite this version:**

Ciriac Charles, Cyril Conde, Fabien Vorimore, Thierry Cochard, Lorraine Michelet, et al.. Features of *Mycobacterium bovis* Complete Genomes Belonging to 5 Different Lineages. *Microorganisms*, 2023, 11 (1), pp.177. 10.3390/microorganisms11010177 . hal-03952825

HAL Id: hal-03952825

<https://hal.inrae.fr/hal-03952825>

Submitted on 23 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Article

Features of *Mycobacterium bovis* Complete Genomes Belonging to 5 Different Lineages

Ciriac Charles ^{1,2}, Cyril Conde ², Fabien Vorimore ³, Thierry Cochard ², Lorraine Michelet ¹, Maria Laura Boschioli ^{1,*} and Franck Biet ^{2,*}

- ¹ Animal Health Laboratory, National Reference Laboratory for Tuberculosis, Paris-Est University, French Agency for Food, Environmental and Occupational Health and Safety (ANSES), CEDEX, 94701 Maisons-Alfort, France; ciriac.charles11@gmail.com (C.C.); lorraine.michelet@anses.fr (L.M.)
- ² Infectiologie et Santé Publique (ISP), Institut National de Recherche pour L'agriculture, L'alimentation et L'environnement (INRAE), Université de Tours, UMR 1282, 37380 Nouzilly, France; cyril.conde@inrae.fr (C.C.); thierry.cochard@inrae.fr (T.C.)
- ³ Laboratory for Food Safety, Unit of 'Pathogenic *E. coli*' (COLiPATH) & Genomics Platform 'IdentyPath' (IDPA), ANSES, 94701 Maisons-Alfort, France; fabien.vorimore@anses.fr
- * Correspondence: maria-laura.boschioli@anses.fr (M.L.B.); franck.biet@inrae.fr (F.B.); Tel.: +33-149771321 (M.L.B.); +33-247427869 (F.B.)

Abstract: Mammalian tuberculosis (TB) is a zoonotic disease mainly due to *Mycobacterium bovis* (*M. bovis*). A current challenge for its eradication is understanding its transmission within multi-host systems. Improvements in long-read sequencing technologies have made it possible to obtain complete bacterial genomes that provide a comprehensive view of species-specific genomic features. In the context of TB, new genomic references based on complete genomes genetically close to field strains are also essential to perform precise field molecular epidemiological studies. A total of 10 *M. bovis* strains representing each genetic lineage identified in France and in other countries were selected for performing complete assembly of their genomes. Pangenome analysis revealed a "closed" pangenome composed of 3900 core genes and only 96 accessory genes. Whole genomes-based alignment using progressive Mauve showed remarkable conservation of the genomic synteny except that the genomes have a variable number of copies of *IS6110*. Characteristic genomic traits of each lineage were identified through the discovery of specific indels. Altogether, these results provide new genetic features that improve the description of *M. bovis* lineages. The availability of new complete representative genomes of *M. bovis* will be useful to epidemiological studies and better understand the transmission of this clonal-evolving pathogen.

Keywords: *Mycobacterium bovis*; mammalian tuberculosis (TB); complete de novo assembly; transmission; pangenome



Citation: Charles, C.; Conde, C.; Vorimore, F.; Cochard, T.; Michelet, L.; Boschioli, M.L.; Biet, F. Features of *Mycobacterium bovis* Complete Genomes Belonging to 5 Different Lineages. *Microorganisms* **2023**, *11*, 177. <https://doi.org/10.3390/microorganisms11010177>

Academic Editors: Guislaine Réfrégier and Charlotte Genestet

Received: 25 November 2022

Revised: 6 January 2023

Accepted: 7 January 2023

Published: 11 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mammalian tuberculosis (TB) is a zoonotic disease mainly due to *Mycobacterium bovis* (*M. bovis*). Within *M. bovis*, four major clonal complexes were defined by the lack of certain specific regions, single nucleotide polymorphism (SNP), and genetics signatures in the DR region [1–4]. These four groups are the European 1 clonal complex (Eu1) mainly present in the British Islands, and the former British empire colonies [4], the European 2 clonal complex (Eu2), dominant in the Iberian Peninsula [3], the African 1 clonal complex (Af1) present in Mali, Cameroon, Nigeria and Chad [2], and the African 2 clonal complex (Af2) mostly found in East Africa [1].

The generalization of genome sequencing in the last years has made it possible to obtain several thousands of short-reads whole genome sequences of *M. bovis* [5–10]. These data are useful to propose *M. bovis* classification [7,10,11], such as that of Zwyrer et al. proposing to classify this species into eight sublineages named La1.1 to La1.8 [11]. The

M. bovis French diversity has been divided into nine clusters (Cluster A to I, cluster D represents Eu1 and cluster F represents Eu2), which have been defined by specific SNPs, particular signatures in the DR region and in certain VNTR loci [7]. Strains belonging to cluster A, cluster I, and cluster C provoked the majority of outbreaks detected in France in the last 20 years [7,12–14]. New complete genomes belonging to these different groups could help improve the description of *M. bovis* clusters defined previously. Indeed, most sequenced genomes are drafts. These genomes are incomplete and contain indels which can bias genetic structure studies or pangenomic studies [15]. Genome sequencing using long-read technologies now makes it possible to correct these errors and complete the genome at a lower cost [16].

Until recently, only AF2122/97 (NC_002945.4), the complete genome of a Eu1 field strain isolated in the UK, was used as a reference in whole genome SNP (wgSNP) studies [17,18]. Even though this reference genome is well adapted for epidemiological studies where Eu1 strains are common [11,19–23], it could be less fitted for studies in France and other mainland European countries where strains belonging to this clonal complex are not frequent [3]. Recently, a new complete genome, Mb3601, has been published [24]. Mb3601 was obtained by combining short-reads (Illumina) and long-read (PacBio) sequencing technologies. This new genome is specific to one of the most widespread genotypes in France in the last years, SB0120-CO [25]. The study of this complete genome has highlighted the presence of a significant number of IS6110 copies and the presence of several indels in its genome. Its description led to the proposal of a new clonal complex, European 3, to replace Cluster I.

The aim of this work was to obtain new complete genomes that represent *M. bovis* lineages identified in France selected among the main genotypes responsible for TB in the last years to refine our genomic knowledge of *M. bovis* via pangenome analysis and to provide better resolution of the phylogeny needed to study the epidemiology, transmission, and evolution of this clonal pathogen.

2. Materials and Methods

2.1. *Mycobacterium bovis* Strains

A total of 10 strains that cover all *M. bovis* lineages identified in France and representing the main genotypes responsible for TB outbreaks were selected from the National reference laboratory strain collection (Table 1). These strains were grown in Middlebrook 7H9 + ADC enrichment supplement as described previously [7].

2.2. Additional Genomes

To improve pangenomic and SNP studies, 86 genomes representative of the *M. bovis* French diversity obtained by Illumina technology in a previous study [7] and 2 complete genomes (Mb3601 and AF2122/97) were included [18,24].

2.3. DNA Extraction

DNA extraction was performed on 40 mL of stationary phase culture using CTAB and phenol chloroform [7,26,27]. DNA concentration was measured with Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Rodano, MI, Italy) using the “dsDNA BR Assay” kit (Thermo Fisher Scientific). For MinION sequencing, DNA qualities were checked with Nanodrop and DNA integrity was checked with an Agilent 4200 TapeStation. For Illumina sequencing, control quality was performed by Genoscreen (Lille, France) by SybrGreen assay (ThermoFisher scientific) and qualitatively controlled by agarose gel electrophoresis.

Table 1. Information on the 10 *M. bovis* strains selected and sequenced in this study.

Name	Mb2487	Mb3602	Mb2269	Mb0820	Mb0531	Mb0486	Mb2377	Mb1101	Mb1855	Mb3114
Accession Number	CP096839	CP096843	CP096840	CP096841	CP096847	CP096848	CP096846	CP096845	CP096844	CP096842
Host species	Cattle	Deer	Cattle	Cattle	Cattle	Cattle	Cattle	Cattle	Cattle	Cattle
Spoligotype ID	SB0999	SB0134	SB0134	SB0840	SB0826	SB0821	SB0853	SB0120	SB0120	SB0120
MLVA profile *	6 4 5 2 8 2 4 7	7 4 5 3 10 4 5 10	6 5 5 3 6 4 5 6	7 5 5 3 8 2 5 s 4	6 7 3 3 10 2 5 s 8	6 5 5 3 11 2 5 s 4	3 6 5 2 9 3 4 6	5 2 3 3 10 3 3 10	5 3 5 3 9 4 5 6	5 5 5 3 11 3 5 4
Cluster	F	C	C	A	A	A	G	I	I	I
Alias	Eu2 CC	SB0134 family	SB0134 family	F4 family	F4 family	F4 family	F9 family	Eu3 CC	Eu3 CC	Eu3 CC
Length (bp)	4,344,516	4,343,218	4,351,057	4,344,564	4,342,977	4,340,629	4,338,946	4,343,846	4,362,894	4,353,147
GC (%)	65.62	65.65	65.64	65.64	65.64	65.65	65.65	65.64	65.64	64.65
CDS	4012	3999	4014	4006	4005	3991	3986	4014	4034	4015
rRNA	3	3	3	3	3	3	3	3	3	3
tRNA	52	52	52	52	52	52	52	52	52	51
tmRNA	1	1	1	1	1	1	1	1	1	1
IS6110 Nb	3	1	3	2	4	3	1 truncated	1	12	1
IS1561 Nb	1	1	1	1	1	1	1	1	1	1
IS1081 Nb	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated	5 + 1 truncated

* MLVA loci: ETR A, ETR B, ETR C, ETR D, QUB 11a, QUB 11b, QUB 26, QUB 3232.

2.4. MinION Sequencing

Each DNA sample was purified with AMPure XP beads (Beckman coulter, Villepinte, France). Samples were adjusted to 2 µg in 50 µL with Qubit (dsDNA BR Assay) quantification and diluted with DNase, RNase-free water. The MinION library was prepared according to Nanopore's protocol "Native barcoding genomic DNA (with EXP-NBD104, EXP-NBD114, and SQK-LSK109)" (Version: NBE_9065_v109_revV_14Aug2019). DNA pool of 324 ng was loaded on an (R9.4.1) flow cell and was sequenced with Oxford Nanopore MinION sequencer for 48h (Table S1).

2.5. Illumina Sequencing

Nextera XT sequencing libraries were generated with the "Nextera XT DNA Library Prep" kit according to the supplier's recommendations, except for the equimolar pool preparation (GenoScreen optimization). Whole genome paired end 2 × 150 bases pairs (bp) sequencing was performed using Illumina MiSeq technology by Genoscreen (Lille, France) (Table S1). To avoid PCR overrepresented fragments during the library preparation, the paired-end FASTQ files were filtered, leaving only one pair of replicated short reads.

2.6. Genome Assembly Method

The quality of sequencing reads was evaluated using FastQC (Version 0.11.9) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 10 August 2021)). Reads were trimmed with Sickle (Version 1.33) (<https://github.com/najoshi/sickle> (accessed on 10 August 2021)) using a quality Phred-score of Q20 [28,29] and Nanoflit (version 1.1.0) [30] with -q 8 and -l 500 options for long-read sequencing. For each genome, Trycycler [31] subsample was used to perform 12 different read sets from initial long reads. These data were assembled with Flye (version 2.8.3-b 1695) [32], Raven (version 1.50) [33], and Unicycler (version v0.4.9b) tools [34]. These tools give different assemblies, and a consensus assembly was obtained for each strain using Trycycler (version v0.5.0) [31]. The consensus assemblies were polished with Medaka (version 1.4.3) (<https://github.com/nanoporetech/medaka> (accessed on 10 August 2021)) to create consensus sequences and variant calls from long-read sequences against the previously assembled genome. Assemblies were corrected using short-reads and Pilon (version 1.24) [35]. Pilon was executed until 2 runs returned no corrections when the reference genome and short-reads were aligned. A circulator (version 1.5.5) [36] was used on the genomes to change the start position to the *dnaA* gene (with -min_id 70 option) (available at GitHub https://github.com/CiriacC/Hybrid_bacterial_genome_assembly (accessed on 4 September 2022)). Assemblies' statistics of new complete genomes were calculated using Quast (version 5.0.2) [37].

2.7. Genome Comparison

IS6110, IS1081, and IS1561 insertion sequences were screened on genomes using Bionumerics (version 7.6.2) software created by Applied Maths NV and available from <http://www.applied-maths.com> (accessed on 10 June 2022). Genomes were aligned with progressiveMauve (version 2.4.0) [38] to determine the genomic structure. The list of indels was obtained using progressiveMauve for each complete genome in comparison to genome reference (Mb3601). In this study, we selected indels of at least 10 bp. Genes involved in these genetic events were inferred in comparison to the reference genome annotation (Mb3601). A comparison of these genetic events and the already known RDs was performed according to the Bespiatykh study [39].

2.8. Pangenomic Analysis

Genomic annotation was carried out with the Prokka (version 1.14.6) [40] tool using prodigal [41] and Mb3601 genbank file to predict ORF (open reading frame).

The pangenomic study was performed with Panaroo (version 1.2.8) [42] with `-merge_p-aralogue` and `-clean-mode strict` options on 12 complete genomes (10 new complete genomes and 2 reference genomes). Visualization of the pangenome, core genome, and new genes accumulation curves were computed using the “gene_presence_absence.Rtab” matrix provided by Panaroo analysis and PanGP software [43] with Distance Guide sample Algorithm, 500 samples size, 100 samples repeat, and 100 amplification coefficients. Accessory genes were aligned with the reference genome (Mb3601) using `blastn` [44] to search for events that could explain their affiliation to the accessory genome.

2.9. Whole Genome SNP Identification and Selection

Genome reads were aligned to Mb3601 using BWA mem and samtools [45,46]. Vcf files were produced using `bcftools mpileup` and the SNP calling was made with `bcftools call` (with `-vm` option) [47]. The product was filtered using `vcffilter` and `-f QUAL > 150 -f DP > 20 -f MQ > 49` options. All SNPs previously detected have been used to list all the variant positions of the panel. This list was used to make a second calling (using `bcftools mpileup` then `bcftools call` with `-m` option) only on these positions in order to have the same information for all strains and facilitate their fusion with `bcftools merge`. SNP annotation and effect prediction was performed using SnpEff and Mb3601 reference genome. The final steps of variant calling were performed with `vcflib vcfsnps`, `vcflib vcffixup`, `vcflib vcfnumalt`, and `vcffilter` (`-f 'NUMALT = 1'` option) [48]. SNPs supported by less than five reads forward and five reads reverse were filtered. Indel and SNPs with ambiguous nucleotide present on at least one strain have also been filtered. SNPs present in PE/PPE family protein and *pks12* were also filtered because of the low confidence and the higher error rate of these regions [48–50].

2.10. Phylogeny Based on SNP

Evolutionary trees were inferred on Mega [51] using the maximum likelihood method (Hasegawa–Kishino–Yano model) based on concatenated and validated SNPs (7023 SNPs for 98 genomes). The trees were drawn to scale, with branch lengths measured in the number of substitutions per site. Trees were midpoint rooted. A phylogenetic tree was visualized using the Interactive Tree of Life [52].

3. Results

3.1. Complete Genomes Features

For each of the 10 genomes, we obtained a complete assembly with 1 circular contig. Genomic characteristics are consistent with previous reference genomes and show great stability in genomic characteristics [18,24]. The genetic structure of the complete genomes has high stability (Figure S1). However, some differences are present especially in length and coding sequences (CDS) number (Table 1). Mb1855 has an addition of 23,948 bp and 48 CDS in comparison to Mb2377.

All genomes have three rRNA and one tmRNA. Almost all genomes have 52 tRNA, one of them presenting a mutation in position 77 (C→T) in Mb3114, which has only 51.

Insertion sequence (IS) analyses showed that all genomes have in the same position one copy of *IS1561* and six copies of *IS1081*, of which one is truncated (Table 1). According to our previous study [53], the number of *IS6110* is variable depending on the genotype. With 12 copies, Mb1855, which belongs to the Eu3 clonal complex, presents the highest *IS6110* (1355 bp) copy number, which is one of the main reasons for its large genome size. In Mb0820, in contrast not only to the other two Cluster A genomes Mb0531 and Mb0486 but also to the rest of the genomes belonging to other clusters, the otherwise ancestral recurrent copy of *IS6110* in the DR locus is absent. In Mb2377 genome, representing Cluster G, there is a large deletion in the DR region which encompasses a portion of *IS6110* including *orfA*. Almost all *IS6110* except for the truncated copy in Mb2377, have a duplication of 2–4 bp in their insertion sites generated during IS transposition [54,55].

Together with the presence of IS6110 variable copy numbers, genome size differences can also be explained by the presence of deletions or insertions (indels).

3.2. Pangenome Analysis and Gene Content Variation

Pangenome analysis on 12 complete genomes (10 described here, plus AF2122/97 and Mb3601) showed 3996 ortholog clusters and confirmed the high clonality of this species as regards the high core genome (98%, 3900 core genes) (Figure 1). The analysis showed 78 shell genes and 18 cloud genes (Table S2).

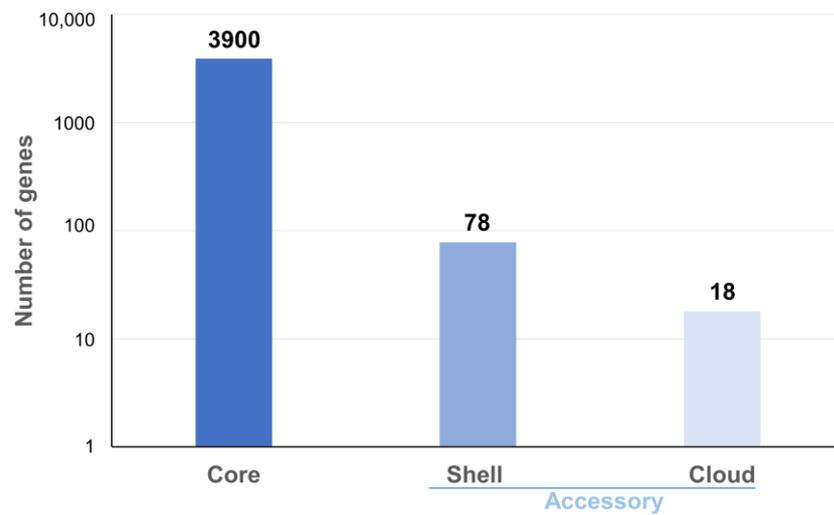


Figure 1. Pan-genomic histogram of 12 complete genomes of *M. bovis*. The figure shows the core and accessory genes proportion in the genome’s panel.

Cloud genes are found in seven genomes that belong to three different clusters (A, C and I) (Figure 2). Cloud genes of Mb0486 and Mb0531 correspond to PE PPE genes and one hypothetical protein for Mb0486 (Table S2). The cloud genes of Mb1101, Mb3114, and Mb3602 are annotated as hypothetical proteins. The cloud genes of Mb1855 and Mb3601 are due to IS6110 insertion in the CDS except for *folp* found in Mb3601 (Table S2).

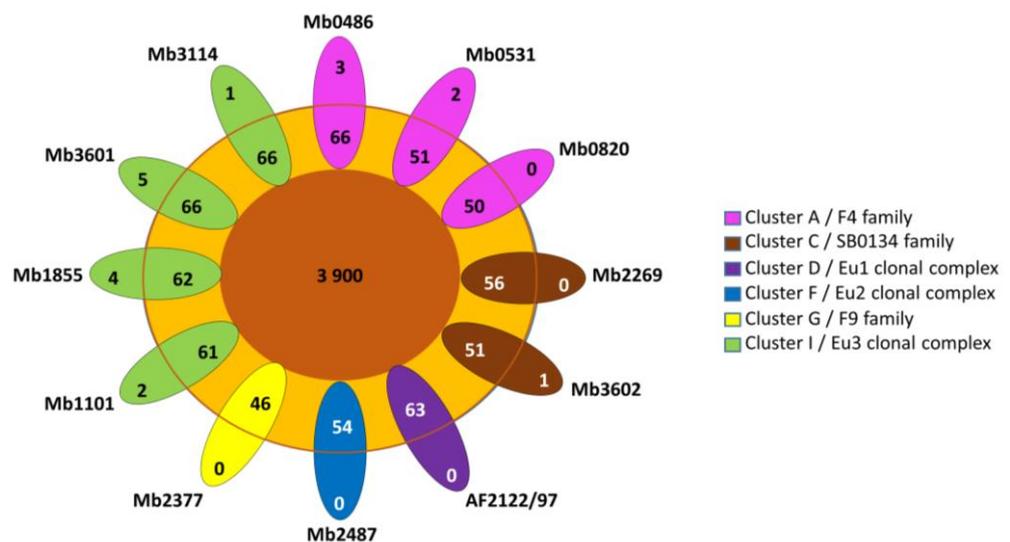


Figure 2. The gene distribution of pangenome. Flower plot showing in the center, genes present in all strains (core-genes), genes present in some strains (shell genes) in the annulus, and strain-specific genes of the 12 *M. bovis* complete genomes in the petals (cloud genes). Genomes are grouped in 6 previously described clusters [7].

The low number of accessory (shell and cloud) genes is consistent with the alpha diversity of 1.11 which highlights a closed pangenome (Figure 3). In addition, a more detailed examination of accessory genes shows that some of them are present but pseudogenized (Table S2). Indeed, 14 accessory genes have been listed due to an *IS6110* insertion which interrupts CDS. For example, the *rpfD_2* orthologous gene present in Mb3601 is due to *IS6110* insertion, in *rpfD*. Fourteen other accessory genes were implied in the putative PhiRv1 phage protein (RD3) [39]. RD3 is present in the three complete genomes of Cluster A, Mb3602 (Cluster C), and Mb2377 (Cluster G). Two of the ortholog clusters present in RD3 are absent in Mb1101 but this genome has the other 12.

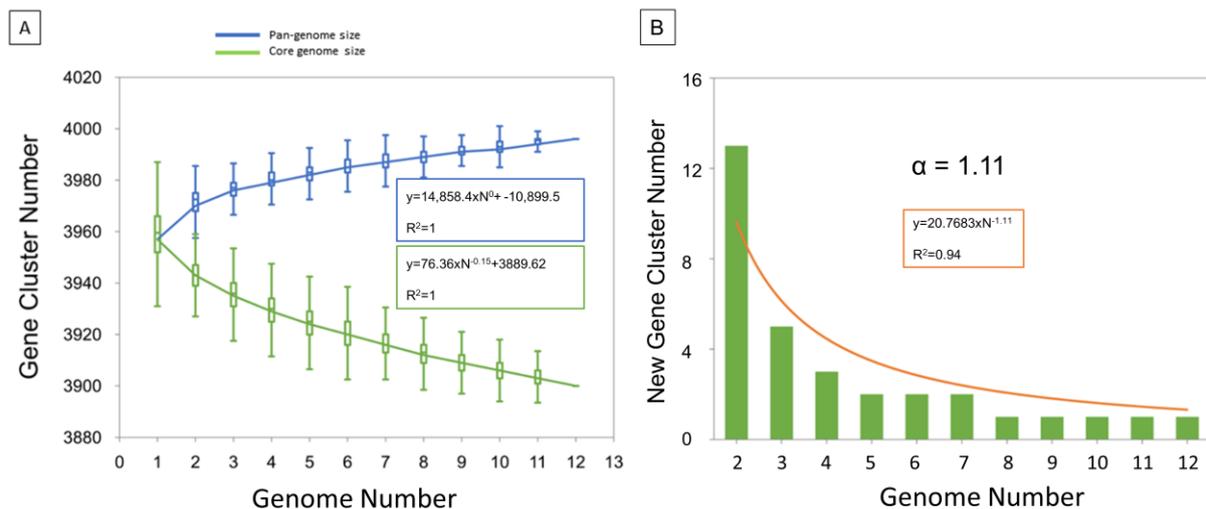


Figure 3. Pan-genome profile calculated with PanGP tool. (A) The figure shows two gene cluster accumulation curves for pangenome (blue) and core genome (green). (B) Evolution of new gene cluster numbers over genome number. The trend line (in orange) defines the curve equation and the alpha diversity.

The other 16 accessory genes concern PE and PPE genes. These genes are known to be highly polymorphic and are often excluded from analyses [56]. They were excluded from our wgSNP study but not from the indel analysis. The region with the most indels found in our study, located at position 3,890,000 bp in Mb3601 genome, encompasses PE PGRS genes (*pe_pgrs* 59, *pe_pgrs* 54, *pe_pgrs* 56 and *pe_pgrs* 53). Other regions presenting numerous indels are CRISPR-Cas (position 3,090,000 bp in Mb3601) or a region including PPE genes (position 2,165,000 bp in Mb3601). This result shows that the indel distribution is not random (Table S3).

The comparison of complete genomes against Mb3601 as reference genome shows 72 indels for Mb2487, 56 for Mb1101, 54 for Mb1855, 34 for Mb3114, 74 for Mb0820, 88 for Mb0531, 83 for Mb0486, 58 for Mb2377, 77 for Mb2269, and 69 for Mb3602 (Table S3).

Some large indels (more than 2 kb) were identified in the complete genomes (Tables 2 and S3).

3.3. Contribution of the Complete Genome to *M. bovis* Lineages Definition

Obtaining complete genomes was an opportunity to revisit the population structure of French *M. bovis* strains by looking at the topology of the SNP-based phylogenetic tree and identifying genetic traits that could complete the new nomenclature covering the main *M. bovis* phylogenetic groups [11]. 7023 SNPs were found among the 98 genomes (12 complete and 86 draft genomes) (Table S4). The majority (87.7%) of SNPs were present in CDS and 12.3% are intergenic. The analysis showed that 31.4% are synonymous variants and 56.3% are non-synonymous variants. The phylogenetic distribution of SNPs, shown in Figure 4, discriminated *M. bovis* genomes into 10 clusters well resolved by at least 200 SNP. This population structure is congruent with previous studies [7,53]. Indeed, the heatmap

clearly highlights lineages based on the absolute SNP distance between strains, which supports a very clear separation between the lineage La1.2 (clusters G+H+I) and lineage La1.8.2 (cluster A+B+C). However, lineages 1.7 or 1.8 are not clearly identified in this figure.

Table 2. Details of large indels affecting the genomes.

Nomenclature	Length (in bp)	Number of Locus Tags Associated	Genome
Indel-Mb0531-33	2384	5	Mb0531
Indel-Mb0486-6	3148	4	Mb0486
Indel-Mb0486-11	3634	4	Mb0486
Indel-Mb3602-33	2150	2	Mb3602
Indel-Mb2269-1	2122	4	Mb2269
Indel-Mb2269-24	2368	6	Mb2269
Indel-Mb2487-5	2691	3	Mb2487
Indel-Mb2487-36	2387	6	Mb2487
Indel-Mb2487-50/RDBovis	2409	3	Mb2487
Indel-Mb2377-27	5539	6	Mb2377
Indel-Mb1101-1	2966	2	Mb1101
Indel-Mb1101-8	4384	2	Mb1101
Indel-Mb1101-21	1160	1	Mb1101
Indel-Mb1855-26	1730	1	Mb1855
Indel-Mb1855-29	3058	3	Mb1855

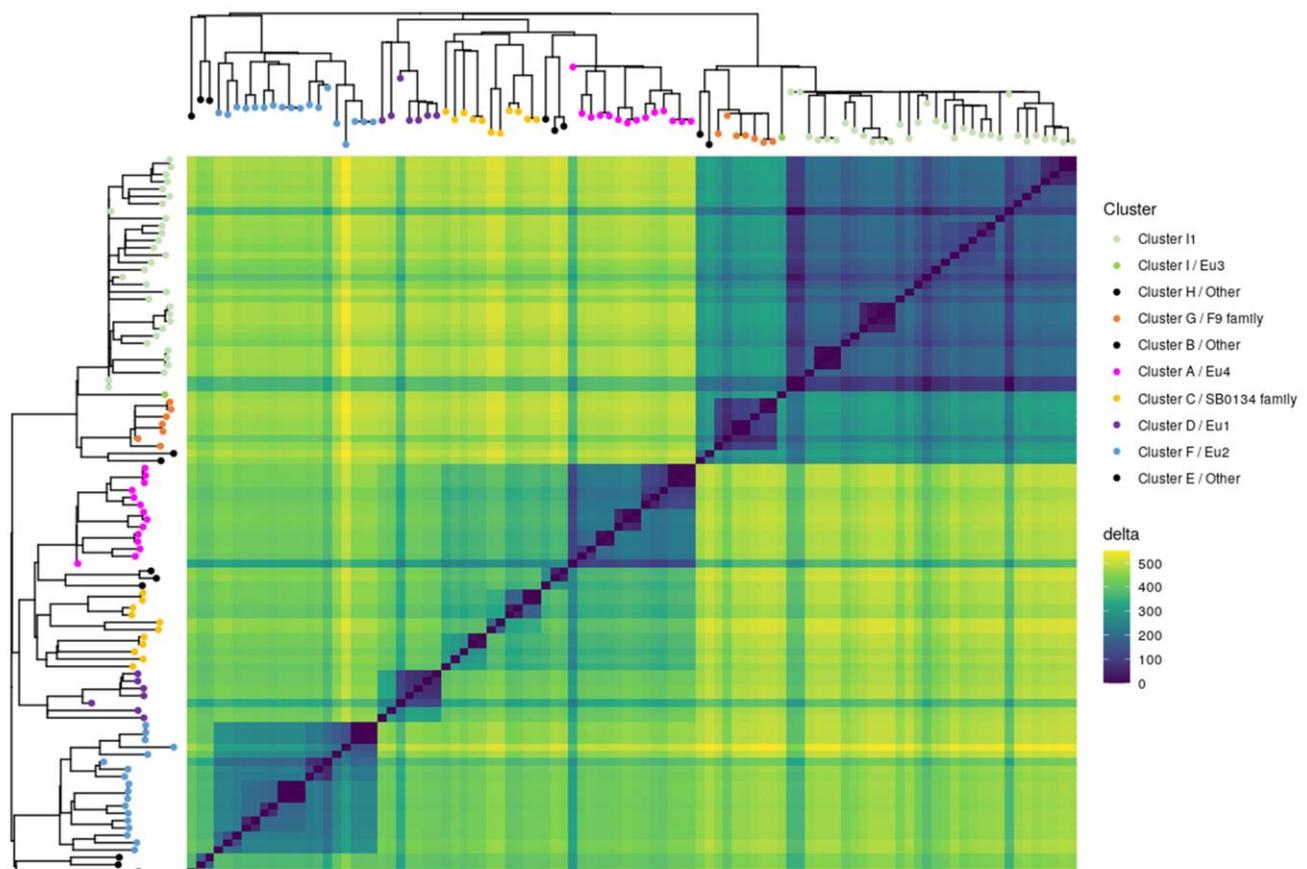


Figure 4. *M. bovis* isolates separated into clusters. The heatmap illustrates pairwise SNP distance between genomes belonging to each cluster. Both axes have a maximum-likelihood SNP-based tree inferred on 98 genomes with leaf colored according to cluster defined in this study. Trees were midpoint rooted. The SNP difference key is shown on the right.

All groups described in this study have specific SNPs (Figure 5). Some groups have few specific SNP such as Cluster C, and Eu3. Other groups have more than 60 specific SNPs such as Eu2, Cluster A, or Cluster G.

<i>M. bovis</i> group	Complete genome	Specific SNPs	Specific indels	Other specific genetic characteristic
La1.8 + La1.7		108	8	
La1.7		6		
La1.7.X / unk4		5		
La1.7.1 / Eu2	Mb2487	77		missing spoligotype spacer 21
La1.8		12	3	
La1.8.1 / Eu1	AF2122/97	47		RDEu1
La1.8.2 / unk7		27	4	
Cluster C / SB0134 family	Mb3602 / Mb2269	4	1	missing spoligotype spacers 4 and 5
Cluster B / Other		47		
Cluster A / F4 family	Mb0820 / Mb0531 / Mb0486	66	8	missing spoligotype spacer 33 and QUB26 truncated
La1.2 / unk2		108	8	
Cluster G / F9 family	Mb2377	83		missing spoligotype spacers 1 to 17/ IS6110 truncated
Cluster H / Other		46		
Cluster I / Eu3	Mb1101	2		
Cluster I1	Mb3601 / Mb3114 / Mb1855	53	3	

Figure 5. List of specific genetic events of the different *M. bovis* groups (Table S4). The colors of the *M. bovis* groups are in accordance with the previously described clusters and lineage Hauer et al. 2019 [7].

Indels were also examined on the 98 genomes and the specificity of an indel for an *M. bovis* group was determined when it was identified in all genomes of this group (Figure 5).

3.3.1. Cluster A/F4 Family

Three complete genomes were obtained for Cluster A: Mb0486, Mb0820, and Mb0531. This cluster is described by 66 specific SNPs and 8 specific indels (Table S4). In comparison to Mb3601, the deletions involve *metk* (MBS3601_RS07300) and a *leuA* (MBS3601_RS19090) partial deletion. However, *leuA* was also partially deleted in the different and larger indel of Mb2487 (cluster F). Two recurrent IS6110 insertions sites were found in the three complete genomes. Other genomic characteristics of this cluster are the absence of spacer 33 in the DR region, RD3, and the truncated repetition of QUB26.

3.3.2. Cluster C/SB0134 Family

This cluster is composed of two subgroups and is described by few SNP and only one deletion of 514 bp (Figures 4 and 5). This group does not present spacers 4 and 5 in their spoligotypes. Mb3602 and Mb2269 are present in each of these subgroups.

3.3.3. Cluster F/Eu2

Mb2487 is representative of this clonal complex, which is defined by 77 SNPs, including that in *guaA* described originally [3] and a lack of spacer 21 in their spoligotypes.

3.3.4. Lineages 1.7 and 1.8

A deletion of 2409 bp (Indel-Mb0486-49, Indel-Mb0531-56, Indel-Mb0820-44, Indel-Mb3602-44, Indel-Mb2269-47, and Indel-Mb2487-50) which corresponds to RDBovis [39] is common to genomes of cluster A, C, and F and allow to define the lineage La1.7 + La1.8. This lineage is also defined by 108 SNPs and an insertion of more than 2000 bp (Indel-Mb2487-64, Indel-Mb2269-63, Indel-Mb3602-58, Indel-Mb0820-65, Indel-Mb0531-76, and Indel-Mb0486-49). However, the insertion present in Mb2487 is the largest compared to those in the other complete genomes. This region contains PPE genes. These two indels are also present in AF2122/97, as shown in a previous study comparing this genome to Mb3601.

3.3.5. Cluster G/F9 Family

Mb2377 is representative of cluster G, which is defined by 83 SNPs. As mentioned before, this cluster is characterized by the truncated *IS6110* in the DR region and the lack of spacer 1 to 17. These specific genetic characteristics are due to a large indel (Indel-Mb2377-27).

3.3.6. Cluster I/Eu3

In addition to Mb3601, three complete genomes were obtained for this cluster, which is the most represented among the strains studied in France [12,14]. The Eu3 clonal complex is only defined by two SNPs. Indeed, Mb1101 is close to BCG vaccine strains and is separated from other Eu3 strains (Figure 6).

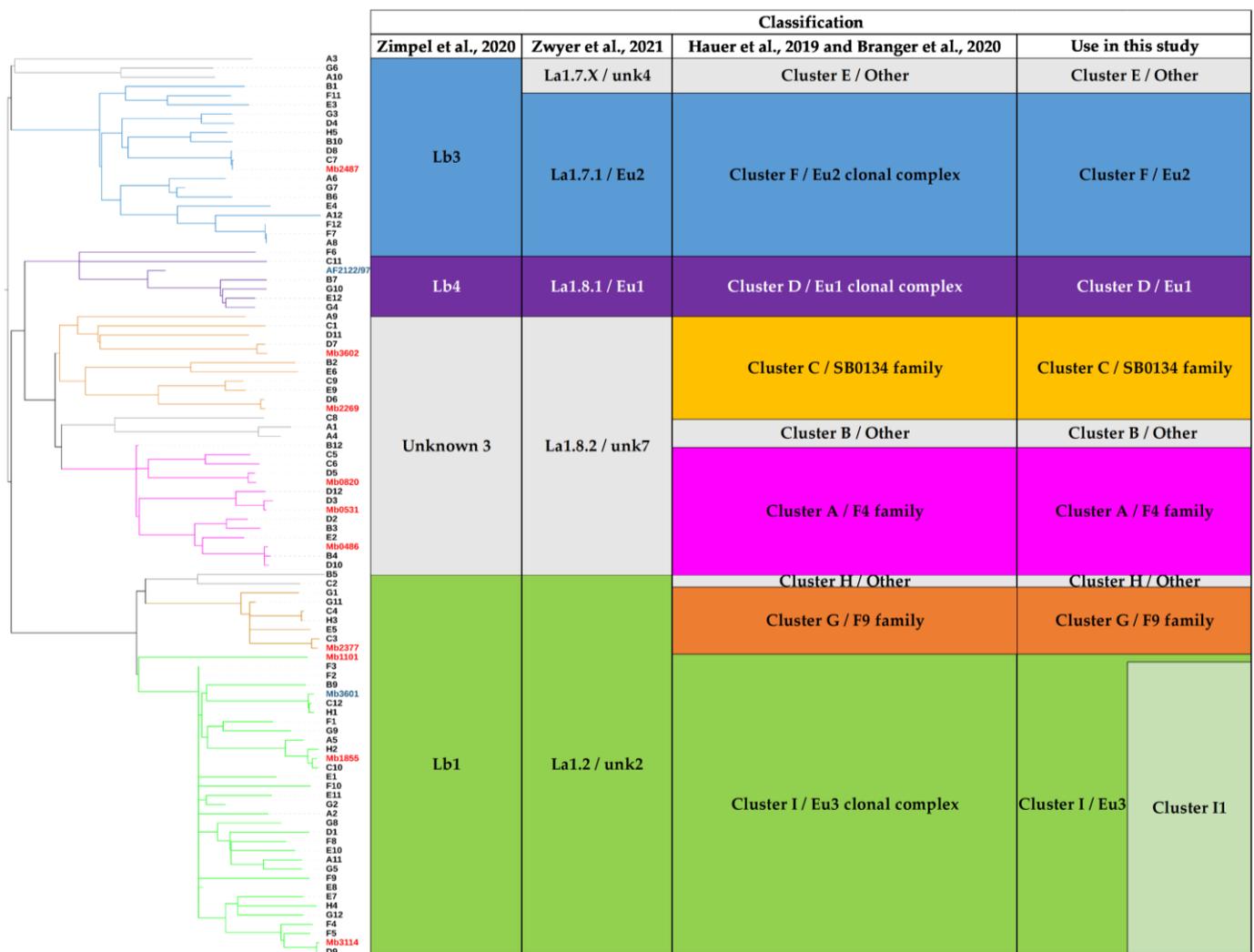


Figure 6. Phylogenetic tree of 98 *M. bovis* genomes. The two previous reference genomes (Mb3601 and AF2122/97) are marked in blue. The 10 new complete genomes are indicated in red. The phylogenetic tree is based on 7023 whole genome SNPs. The strains are grouped according to the previous classification Hauer et al. 2019, Zwyer et al. 2021 and Guimares et al. 2020 [7,11,16]. The colors of the *M. bovis* groups are in accordance with the previously described clusters and lineages Hauer et al. 2019, Zimpel et al. 2020 and Zwyer et al. 2021 [7,10,11].

We propose to define Cluster I1 which corresponds to Cluster I strains by removing vaccine strains and Mb1101. Mb3114, Mb3601, and Mb1855 are present in this cluster, which is defined by 53 SNPs and 3 indels. One of these indels, a deletion of 622 bp, includes *VapB46* (Tables S3 and S4).

4. Discussion

Using Illumina and MinION sequencing technologies, 10 complete genomes of *M. bovis* were obtained and represented—with AF2122/97 and Mb3601—the main *M. bovis* clusters described previously [7]. Our analysis showed highly similar genomic features and a conserved synteny within these new 10 complete genomes. Pangenome established with the ten complete genomes and the two other available complete genomes [18,24] confirmed a closed pangenome with a core gene representing 98% of total genes in agreement with previous studies [4,15,57]. However, a recent study using the “Get-homologues pipeline” and draft genomes showed an open pangenome and a larger accessory genome in comparison to our study [58,59]. This difference can be explained by the short-read sequences data, which lead to the increase of the accessory genome [15]. To overcome this problem, Panaroo can be used to clean up annotation errors due to fragmented assemblies or misassembly [42]. Indeed, Panaroo produces superior ortholog clusters, which induce a reduction in the accessory genome estimate size and an increase in the core genome. The pangenomic analysis of 12 complete genomes highlighted an alpha diversity of 1.11 consistent with a closed pangenome [60]. In addition, the presence or absence of certain ortholog clusters in genomes is due to gene pseudogenization. Our analysis showed that the size of the core genome decreases more rapidly than the increase in the pangenome size corroborating that evolution of the MTBC complex members genomes, as recently demonstrated for *M. bovis* [61], occurs by gene loss or pseudogenized instead of gene gain. This event could explain the pathogen’s host specialization as shown in *M. tuberculosis* [62,63].

Among genomic features, we observed variations in genomes size between the 10 complete genomes. This observation was explained by a variable number of copies of IS6110 according to the genomes and the indel content. Indeed, the 12 copies of IS6110 in Mb1855 represent an addition of 14,905 bp in comparison to genomes with only one copy of IS6110. The complete genomes allowed us to confirm the presence of multiple copies of IS6110 in certain *M. bovis* genotypes according to our previous study [24,53]. The transposition of this genetic element can play an important role in bacterial evolution by interrupting or leading to the overexpression of genes [53,64–66]. Indeed, some genetic changes such as gene deletion or gene pseudogenization that could affect the core genome, can be attributed to IS6110. Multiple examples are present in literature and show the deletion of some genes like *cas* genes in the CRISPR-Cas locus [65]. In our study, one of these examples is present in Cluster G strains with the absence of *Cas* genes and the first 17 spoliotype spacers. However, except for this example, all IS6110 have a duplication of 2–4 bp in their insertion sites which in the nine other complete genomes shows the absence of IS recombination events between two IS6110.

Indels can also explain length differences among genomes. Some large deletions are identified in this study as Indel-Mb2377-27 of 5539 bp in Mb2377, Indel-Mb2487-64 of 5166 bp in Mb2487, RDBovis of 2409 bp present in genomes of La1.7 and La1.8 or RD3 of 9253 bp present in Cluster A and several other genomes [39]. This last indel corresponds to prophage phiRv1 which seems to have a role in host hypoxia [61,67,68]. However, Mb1101 has a specific deletion pattern in this region that involves two ortholog clusters instead of 14 in RD3. In addition, our results showed that indel positions are not random. Many indels are present in the CRISPR-Cas region [65] but the most polymorphic region is that containing PE and PPE genes. This high frequency of deletions and insertions in these regions is in agreement with the previous *M. bovis* complete genome publication [24]. Further studies on these indels are needed to better understand their role in bacterial evolution.

In this study, the selected *M. bovis* strains to obtain complete genomes, represent the main genotypes responsible for TB outbreaks in France and are also representative of *M. bovis* genotypes found in other countries. Indeed, Mb2487 belongs to the lineage 1.7.1, formerly described as Eu2 clonal complex [3,8,11]. Four complete genomes belong to lineage La 1.2, 3 of which belong to the Eu3 clonal group (in addition to Mb3601). Mb1855 is representative of highly prevalent strains in France with several copies of IS6110. Mb3114 is representative of a common genotype in Italy with only one copy of IS6110 [69]. Five

genomes belong to lineage 1.8.2. This lineage had previously been separated in the Hauer study into Cluster A, B, and C [7]. Harmonization of the nomenclature used to describe *M. bovis* lineages may facilitate comparisons of WGS studies. Specific indels and SNPs were described for complete genomes or *M. bovis* lineages. Some of these genetic events such as *guaA* and other 68 SNPs specific to the Eu2 strains [10], already described in the literature, were confirmed in this study. Nevertheless, the number of specific SNPs found for the previously described clusters was larger than what was found in a recent study [11]. These differences can be explained by the smaller number of strains used in our study. This result shows the importance of using a panel of strains as exhaustive as possible to describe specific events of the *M. bovis* lineage. Some indels were found to be specific to *M. bovis* lineages, others appear to be specific to certain genomes. They will need to be investigated in larger panels of strains to determine if they are the signature of groups or subgroups of *M. bovis*.

TB cattle outbreaks in France are present in specific regions where *M. bovis* circulates in wild and domestic communities of hosts [13,25] where the transmission links between infected animals remain difficult to establish as *M. bovis* strains share spoligotype and a multilocus variable number of tandem repeats analysis (MLVA) identical profiles [70–72]. WGS-SNP can be used to refine these studies but requires adapted reference genomes to the field strains. Mb3601 and other representative complete genomes could be used to improve epidemiological studies for the surveillance of TB and contact tracing between infected animals [16]. The new complete genomes described in this study are closer to field strains than AF2122/97, the genome used as a reference until now, which will allow better epidemiological surveillance of the disease based on WGS data.

5. Conclusions

Ten new *M. bovis* complete genomes were obtained in this study. These new complete genomes cover the *M. bovis* French diversity but are also representative of *M. bovis* lineages present in other countries. These genomes allow us to better describe *M. bovis* lineages. A comparison of these complete genomes confirmed that the global genome organization of *M. bovis* is very stable and shows a closed pangenome. The search for indels and SNPs made it possible to specify certain genomic traits and the absence of certain genes characterizing each cluster described in this article.

These complete genomes, adapted to *M. bovis* clusters, will be useful to better understand TB transmission dynamics in multi-host systems and therefore to implement more effective control measures.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms11010177/s1>, Figure S1: Alignments of the 12 *M. bovis* complete genomes. Table S1: Sequencing metric of the 10 new genomes and obtained with fastqc. A: Metric provide to Illumina metric. B: metric provide to MinION metric. Table S2: Pangenomic study performed on 12 *M. bovis* complete genomes. The table indicates the genes accessory. "1" shows the presence of CDS and "0" his absence. Table S3: Indels between the ten new complete genomes and Mb3601 using progressiveMauve. Table S4: WgSNP analysis performed on 98 *M. bovis*.

Author Contributions: Conceptualization, M.L.B., F.B. and L.M.; Experimental work, C.C. (Ciriac Charles), C.C. (Cyril Conde), T.C., F.V. and M.L.B. Software, C.C. (Ciriac Charles), C.C. (Cyril Conde) and L.M.; Formal Analysis, C.C. (Ciriac Charles), C.C. (Cyril Conde), M.L.B. and L.M.; Writing—Original Draft Preparation, C.C. (Ciriac Charles), M.L.B., L.M. and F.B.; Writing—Review and Editing, C.C. (Ciriac Charles), C.C. (Cyril Conde), M.L.B., L.M. and F.B. All authors have read and agreed to the published version of the manuscript.

Funding: CCh is a doctoral fellow financed by the ANSES and the One Health European Joint Programme (OHEJP), Grant number 773830.

Data Availability Statement: The raw data are deposited in a public domain server at the NCBI SRA database, under BioProject accession number PRJNA832544.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Berg, S.; Garcia-Pelayo, M.C.; Müller, B.; Hailu, E.; Asiimwe, B.; Kremer, K.; Dale, J.; Boniotti, M.B.; Rodriguez, S.; Hilty, M.; et al. African 2, a Clonal Complex of *Mycobacterium bovis* Epidemiologically Important in East Africa. *J. Bacteriol.* **2011**, *193*, 670–678. [[CrossRef](#)]
2. Muller, B.; Hilty, M.; Berg, S.; Garcia-Pelayo, M.C.; Dale, J.; Boschioli, M.L.; Cadmus, S.; Ngandolo, B.N.R.; Godreuil, S.; Diguimbaye-Djaibé, C.; et al. African 1, an Epidemiologically Important Clonal Complex of *Mycobacterium bovis* Dominant in Mali, Nigeria, Cameroon, and Chad. *J. Bacteriol.* **2009**, *191*, 1951–1960. [[CrossRef](#)]
3. Rodriguez-Campos, S.; Schürch, A.C.; Dale, J.; Lohan, A.J.; Cunha, M.V.; Botelho, A.; De Cruz, K.; Boschioli, M.L.; Boniotti, M.B.; Pacciarini, M.; et al. European 2—A clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. *Infect. Genet. Evol.* **2012**, *12*, 866–872. [[CrossRef](#)]
4. Smith, N.H.; Berg, S.; Dale, J.; Allen, A.; Rodriguez, S.; Romero, B.; Matos, F.; Ghebremichael, S.; Karoui, C.; Donati, C.; et al. European 1: A globally important clonal complex of *Mycobacterium bovis*. *Infect. Genet. Evol.* **2011**, *11*, 1340–1351. [[CrossRef](#)]
5. Almaw, G.; Mekonnen, G.A.; Mihret, A.; Aseffa, A.; Taye, H.; Conlan, A.J.K.; Gumi, B.; Zewude, A.; Aliy, A.; Tamiru, M.; et al. Population structure and transmission of *Mycobacterium bovis* in Ethiopia. *Microb. Genom.* **2021**, *7*, 000539. [[CrossRef](#)]
6. da Conceição, M.L.; Conceição, E.C.; Furlaneto, I.P.; Da Silva, S.P.; Guimarães, A.E.D.S.; Gomes, P.; Boschioli, M.L.; Michelet, L.; Kohl, T.A.; Kranzer, K.; et al. Phylogenomic Perspective on a Unique *Mycobacterium bovis* Clade Dominating Bovine Tuberculosis Infections among Cattle and Buffalos in Northern Brazil. *Sci. Rep.* **2020**, *10*, 1747. [[CrossRef](#)]
7. Hauer, A.; Michelet, L.; Cochard, T.; Branger, M.; Nunez, J.; Boschioli, M.-L.; Biet, F.; Hauer, A.; Michelet, L.; Cochard, T.; et al. Accurate Phylogenetic Relationships among *Mycobacterium bovis* Strains Circulating in France Based on Whole Genome Sequencing and Single Nucleotide Polymorphism Analysis. *Front. Microbiol.* **2019**, *10*, 955. [[CrossRef](#)]
8. Kohl, T.A.; Kranzer, K.; Andres, S.; Wirth, T.; Niemann, S.; Moser, I. Population Structure of *Mycobacterium bovis* in Germany: A Long-Term Study Using Whole-Genome Sequencing Combined with Conventional Molecular Typing Methods. *J. Clin. Microbiol.* **2020**, *58*, e01573-20. [[CrossRef](#)]
9. Loiseau, C.; Menardo, F.; Aseffa, A.; Hailu, E.; Gumi, B.; Ameni, G.; Berg, S.; Rigouts, L.; Robbe-Austerman, S.; Zinsstag, J.; et al. An African origin for *Mycobacterium bovis*. *Evol. Med. Public Health* **2020**, *2020*, 49–59. [[CrossRef](#)]
10. Zimpel, C.K.; Patané, J.S.L.; Guedes, A.C.P.; de Souza, R.F.; Silva-Pereira, T.T.; Camargo, N.C.S.; Filho, A.F.D.S.; Ikuta, C.Y.; Neto, J.S.F.; Setubal, J.C.; et al. Global Distribution and Evolution of *Mycobacterium bovis* Lineages. *Front. Microbiol.* **2020**, *11*, 843. [[CrossRef](#)]
11. Zwyrer, M.; Çavusoglu, C.; Ghielmetti, G.; Pacciarini, M.L.; Scaltriti, E.; Van Soelingen, D.; Dötsch, A.; Reinhard, M.; Gagneux, S.; Brites, D. A new nomenclature for the livestock-associated *Mycobacterium tuberculosis* complex based on phylogenomics. *Open Res. Eur.* **2021**, *1*, 100. [[CrossRef](#)]
12. Boschioli, M.; Michelet, L.; Hauer, A.; De Cruz, K.; Courcou, A.; Hénault, S.; Palisson, A.; Karoui, C.; Biet, F.; Zanella, G. Tuberculose bovine en France: Cartographie des souches de *Mycobacterium bovis* entre 2000–2013. *Bull. Épidémiologique* **2015**, *70*, 2–8.
13. Hauer, A.; De Cruz, K.; Cochard, T.; Godreuil, S.; Karoui, C.; Henault, S.; Bulach, T.; Bañuls, A.-L.; Biet, F.; Boschioli, M. Genetic evolution of *Mycobacterium bovis* causing tuberculosis in livestock and wildlife in France since. *PLoS ONE* **2015**, *10*, e0117103. [[CrossRef](#)]
14. Michelet, L.; Durand, B.; Boschioli, M.-L. Tuberculose bovine: Bilan génotypique de *M. bovis* à l'origine des foyers bovins entre 2015 et 2017 en France Métropolitaine. *Bull. Epidemiol.* **2020**, *91*, 13.
15. Ceres, K.M.; Stanhope, M.J.; Gröhn, Y.T. A critical evaluation of *Mycobacterium bovis* pangenomics, with reference to its utility in outbreak investigation. *Microb. Genom.* **2022**, *8*, 000839. [[CrossRef](#)]
16. Guimarães, A.M.S.; Zimpel, C.K. *Mycobacterium bovis*: From Genotyping to Genome Sequencing. *Microorganisms* **2020**, *8*, 667. [[CrossRef](#)]
17. Farrell, D.; Crispell, J.; Gordon, S.V. Updated functional annotation of the *Mycobacterium bovis* AF2122/97 reference genome. *Access Microbiol.* **2020**, *2*, e000129. [[CrossRef](#)]
18. Garnier, T.; Eiglmeier, K.; Camus, J.-C.; Medina, N.; Mansoor, H.; Pryor, M.; Duthoy, S.; Grondin, S.; Lacroix, C.; Monsempe, C.; et al. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7877–7882. [[CrossRef](#)]
19. Crispell, J.; Cassidy, S.; Kenny, K.; McGrath, G.; Warde, S.; Cameron, H.; Rossi, G.; MacWhite, T.; White, P.C.L.; Lycett, S.; et al. *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. *Microb. Genom.* **2020**, *6*, e000388. [[CrossRef](#)]
20. Crispell, J.; Zadoks, R.N.; Harris, S.R.; Paterson, B.; Collins, D.M.; De-Lisle, G.W.; Livingstone, P.; Neill, M.A.; Biek, R.; Lycett, S.J.; et al. Using whole genome sequencing to investigate transmission in a multi-host system: Bovine tuberculosis in New Zealand. *BMC Genom.* **2017**, *18*, 180. [[CrossRef](#)]
21. Ortiz, A.P.; Perea, C.; Davalos, E.; Velázquez, E.F.; González, K.S.; Camacho, E.R.; Latorre, E.A.G.; Lara, C.S.; Salazar, R.M.; Bravo, D.M.; et al. Whole Genome Sequencing Links *Mycobacterium bovis* from Cattle, Cheese and Humans in Baja California, Mexico. *Front. Veter.-Sci.* **2021**, *8*, 674307. [[CrossRef](#)]

22. Price-Carter, M.; Brauning, R.; De Lisle, G.W.; Livingstone, P.; Neill, M.; Sinclair, J.; Paterson, B.; Atkinson, G.; Knowles, G.; Crews, K.; et al. Whole Genome Sequencing for Determining the Source of *Mycobacterium bovis* Infections in Livestock Herds and Wildlife in New Zealand. *Front. Veter-Sci.* **2018**, *5*, 272. [[CrossRef](#)]
23. Salvador, L.C.M.; O'Brien, D.J.; Cosgrove, M.K.; Stuber, T.P.; Schooley, A.M.; Crispell, J.; Church, S.V.; Gröhn, Y.T.; Robbe-Auerman, S.; Kao, R.R. Disease management at the wildlife-livestock interface: Using whole-genome sequencing to study the role of elk in *Mycobacterium bovis* transmission in Michigan, USA. *Mol. Ecol.* **2019**, *28*, 2192–2205. [[CrossRef](#)]
24. Branger, M.; Loux, V.; Cocharde, T.; Boschioli, M.L.; Biet, F.; Michelet, L. The complete genome sequence of *Mycobacterium bovis* Mb3601, a SB0120 spoligotype strain representative of a new clonal group. *Infect. Genet. Evol.* **2020**, *82*, 104309. [[CrossRef](#)]
25. Delavenne, C.; Pandofi, F.; Girard, S.; Réveillaud, É.; Boschioli, M.-L.; Dommergues, L.; Garapin, F.; Keck, N.; Martin, F.; Moussu, M.; et al. Tuberculose bovine: Bilan et évolution de la situation épidémiologique entre 2015 et 2017 en France Métropolitaine. *Bull. Epidemiol. Infect.* **2020**, *91*, 12. [[CrossRef](#)]
26. Biek, R.; O'Hare, A.; Wright, D.; Mallon, T.; McCormick, C.; Orton, R.J.; McDowell, S.; Trewby, H.; Skuce, R.A.; Kao, R.R. Whole Genome Sequencing Reveals Local Transmission Patterns of *Mycobacterium bovis* in Sympatric Cattle and Badger Populations. *PLOS Pathog.* **2012**, *8*, e1003008. [[CrossRef](#)]
27. Imai, T.; Ohta, K.; Kigawa, H.; Kanoh, H.; Taniguchi, T.; Tobari, J. Preparation of High-Molecular-Weight DNA: Application to Mycobacterial Cells. *Anal. Biochem.* **1994**, *222*, 479–482. [[CrossRef](#)]
28. Wingett, S.W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **2018**, *7*, 1338. [[CrossRef](#)]
29. Xia, E.; Teo, Y.-Y.; Ong, R.T.-H. SpoTyping: Fast and accurate in silico Mycobacterium spoligotyping from sequence reads. *Genome Med.* **2016**, *8*, 1–9. [[CrossRef](#)]
30. De Coster, W.; D'Hert, S.; Schultz, D.T.; Cruys, M.; Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **2018**, *34*, 2666–2669. [[CrossRef](#)]
31. Wick, R.R.; Judd, L.M.; Cerdeira, L.T.; Hawkey, J.; Méric, G.; Ben Vezina, B.; Wyres, K.L.; Holt, K.E. Tricycler: Consensus long-read assemblies for bacterial genomes. *Genome Biol.* **2021**, *22*, 266. [[CrossRef](#)]
32. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [[CrossRef](#)]
33. Vaser, R.; Šikić, M. Raven: A de novo genome assembler for long reads. *bioRxiv* **2021**. [[CrossRef](#)]
34. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [[CrossRef](#)]
35. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
36. Hunt, M.; De Silva, N.; Otto, T.D.; Parkhill, J.; Keane, J.A.; Harris, S.R. Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **2015**, *16*, 294. [[CrossRef](#)]
37. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
38. Darling, A.E.; Tritt, A.; Eisen, J.A.; Facciotti, M.T. Mauve Assembly Metrics. *Bioinformatics* **2011**, *27*, 2756–2757. [[CrossRef](#)]
39. Bespyatykh, D.; Bespyatykh, J.; Mokrousov, I.; Shitikov, E. A Comprehensive Map of *Mycobacterium tuberculosis* Complex Regions of Difference. *MSphere* **2021**, *6*, e0053521. [[CrossRef](#)]
40. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
41. Hyatt, D.; Chen, G.-L.; Locascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [[CrossRef](#)]
42. Tonkin-Hill, G.; MacAlasdair, N.; Ruis, C.; Weimann, A.; Horesh, G.; Lees, J.A.; Gladstone, R.A.; Lo, S.; Beaudoin, C.; Floto, R.A.; et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **2020**, *21*, 180. [[CrossRef](#)] [[PubMed](#)]
43. Zhao, Y.; Jia, X.; Yang, J.; Ling, Y.; Zhang, Z.; Yu, J.; Wu, J.; Xiao, J. PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* **2014**, *30*, 1297–1299. [[CrossRef](#)]
44. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
45. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
46. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
47. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [[CrossRef](#)]
48. Marin, M.; Vargas, R.; Harris, M.; Jeffrey, B.; Epperson, L.E.; Durbin, D.; Strong, M.; Salfinger, M.; Iqbal, Z.; Akhundova, I.; et al. Genomic sequence characteristics and the empiric accuracy of short-read sequencing. *bioRxiv* **2021**. [[CrossRef](#)]

49. Lorente-Leal, V.; Farrell, D.; Romero, B.; Álvarez, J.; de Juan, L.; Gordon, S.V. Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization. *Front. Veter-Sci.* **2021**, *8*, 780018. [[CrossRef](#)]
50. Meehan, C.J.; Goig, G.; Kohl, T.A.; Verboven, L.; Dippenaar, A.; Ezewudo, M.; Farhat, M.R.; Guthrie, J.L.; Laukens, K.; Miotto, P.; et al. Whole genome sequencing of *Mycobacterium tuberculosis*: Current standards and open issues. *Nat. Rev. Genet.* **2019**, *17*, 533–545. [[CrossRef](#)]
51. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
52. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)] [[PubMed](#)]
53. Charles, C.; Conde, C.; Biet, F.; Boschioli, M.L.; Michelet, L. IS6110 Copy Number in Multi-Host *Mycobacterium bovis* Strains Circulating in Bovine Tuberculosis Endemic French Regions. *Front. Microbiol.* **2022**, *13*, 891902. [[CrossRef](#)] [[PubMed](#)]
54. Dale, J.W. Mobile genetic elements in mycobacteria. *Eur. Respir. J. Suppl.* **1995**, *20*, 633s–648s.
55. Mendiola, M.; Martin, C.; Otal, I.; Gicquel, B. Analysis of the regions responsible for IS6110 RFLP in a single *Mycobacterium tuberculosis* strain. *Res. Microbiol.* **1992**, *143*, 767–772. [[CrossRef](#)]
56. Gupta, A.; Alland, D. Reversible gene silencing through frameshift indels and frameshift scars provide adaptive plasticity for *Mycobacterium tuberculosis*. *Nat. Commun.* **2021**, *12*, 4702. [[CrossRef](#)]
57. Galagan, J.E. Genomic insights into tuberculosis. *Nat. Rev. Genet.* **2014**, *15*, 307–320. [[CrossRef](#)]
58. Contreras-Moreira, B.; Vinuesa, P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Appl. Environ. Microbiol.* **2013**, *79*, 7696–7701. [[CrossRef](#)]
59. Reis, A.C.; Cunha, M.V. The open pan-genome architecture and virulence landscape of *Mycobacterium bovis*. *Microb. Genom.* **2021**, *7*, 000664. [[CrossRef](#)]
60. Richard, G. Eukaryotic Pangenomes. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020. [[CrossRef](#)]
61. Zimpel, C.K.; Brandão, P.E.; de Souza Filho, A.F.; de Souza, R.F.; Ikuta, C.Y.; Ferreira Neto, J.S.; Camargo, N.C.S.; Heinemann, M.B.; Guimarães, A.M.S. Complete Genome Sequencing of *Mycobacterium bovis* SP38 and Comparative Genomics of *Mycobacterium bovis* and *M. tuberculosis* Strains. *Front. Microbiol.* **2017**, *8*, 2389. [[CrossRef](#)]
62. Baumler, A.; Fang, F.C. Host Specificity of Bacterial Pathogens. *Cold Spring Harb. Perspect. Med.* **2013**, *3*, a010041. [[CrossRef](#)]
63. Bolotin, E.; Hershberg, R. Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. *Genome Biol. Evol.* **2015**, *7*, 2173–2187. [[CrossRef](#)]
64. Gonzalo-Asensio, J.; Pérez, I.; Aguilo, N.; Uranga, S.; Picó, A.; Lampreave, C.; Cebollada, A.; Otal, I.; Samper, S.; Martín, C. New insights into the transposition mechanisms of IS6110 and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. *PLOS Genet.* **2018**, *14*, e1007282. [[CrossRef](#)]
65. Refrégier, G.; Sola, C.; Guyeux, C. Unexpected diversity of CRISPR unveils some evolutionary patterns of repeated sequences in *Mycobacterium tuberculosis*. *BMC Genom.* **2020**, *21*, 1–12. [[CrossRef](#)]
66. Soto, C.Y.; Menéndez, M.C.; Pérez, E.; Samper, S.; Gómez, A.B.; García, M.J.; Martín, C. IS6110 Mediates Increased Transcription of the *phoP* Virulence Gene in a Multidrug-Resistant Clinical Isolate Responsible for Tuberculosis Outbreaks. *J. Clin. Microbiol.* **2004**, *42*, 212–219. [[CrossRef](#)]
67. Fan, X.; Alla, A.A.E.A.; Xie, J. Distribution and function of prophage ϕ Rv1 and ϕ Rv2 among *Mycobacterium tuberculosis* complex. *J. Biomol. Struct. Dyn.* **2015**, *34*, 233–238. [[CrossRef](#)]
68. Mahairas, G.G.; Sabo, P.J.; Hickey, M.J.; Singh, D.C.; Stover, C. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J. Bacteriol.* **1996**, *178*, 1274–1282. [[CrossRef](#)]
69. Lombardi, G.; Botti, I.; Pacciarini, M.L.; Boniotti, M.B.; Roncarati, G.; Monte, P.D. Five-year surveillance of human tuberculosis caused by *Mycobacterium bovis* in Bologna, Italy: An underestimated problem. *Epidemiology Infect.* **2017**, *145*, 3035–3039. [[CrossRef](#)]
70. Driscoll, J.R. Spoligotyping for Molecular Epidemiology of the *Mycobacterium tuberculosis* Complex. *Methods Mol. Biol.* **2009**, *551*, 117–128. [[CrossRef](#)]
71. Kamerbeek, J.; Schouls, L.; Kolk, A.; van Agterveld, M.; van Soolingen, D.; Kuijper, S.; Bunschoten, A.; Molhuizen, H.; Shaw, R.; Goyal, M.; et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **1997**, *35*, 907–914. [[CrossRef](#)]
72. Supply, P.; Allix, C.; Lesjean, S.; Cardoso-Oelemann, M.; Rüsche-Gerdes, S.; Willery, E.; Savine, E.; de Haas, P.; van Deutekom, H.; Roring, S.; et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **2006**, *44*, 4498–4510. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.