



Article

Rescue of *Mycobacterium bovis* DNA Obtained from Cultured Samples during Official Surveillance of Animal TB: Key Steps for Robust Whole Genome Sequence Data Generation

Daniela Pinto ^{1,2} , Gonçalo Themudo ¹ , André C. Pereira ^{1,2} , Ana Botelho ³ and Mónica V. Cunha ^{1,2,*}

¹ Centre for Ecology, Evolution and Environmental Changes (cE3c) & CHANGE—Global Change and Sustainability Institute, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal; dspinto@fc.ul.pt (D.P.); andre.c.pereira94@gmail.com (A.C.P.)

² Biosystems & Integrative Sciences Institute (BioISI), Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

³ National Institute for Agrarian and Veterinary Research (INIAV IP), Av. da República, Quinta do Marquês, 2780-157 Oeiras, Portugal; ana.botelho@iniav.pt

* Correspondence: mscunha@ciencias.ulisboa.pt

Abstract: Epidemiological surveillance of animal tuberculosis (TB) based on whole genome sequencing (WGS) of *Mycobacterium bovis* has recently gained track due to its high resolution to identify infection sources, characterize the pathogen population structure, and facilitate contact tracing. However, the workflow from bacterial isolation to sequence data analysis has several technical challenges that may severely impact the power to understand the epidemiological scenario and inform outbreak response. While trying to use archived DNA from cultured samples obtained during routine official surveillance of animal TB in Portugal, we struggled against three major challenges: the low amount of *M. bovis* DNA obtained from routinely processed animal samples; the lack of purity of *M. bovis* DNA, i.e., high levels of contamination with DNA from other organisms; and the co-occurrence of more than one *M. bovis* strain per sample (within-host mixed infection). The loss of isolated genomes generates missed links in transmission chain reconstruction, hampering the biological and epidemiological interpretation of data as a whole. Upon identification of these challenges, we implemented an integrated solution framework based on whole genome amplification and a dedicated computational pipeline to minimize their effects and recover as many genomes as possible. With the approaches described herein, we were able to recover 62 out of 100 samples that would have otherwise been lost. Based on these results, we discuss adjustments that should be made in official and research laboratories to facilitate the sequential implementation of bacteriological culture, PCR, downstream genomics, and computational-based methods. All of this in a time frame supporting data-driven intervention.

Keywords: animal tuberculosis; *Mycobacterium bovis*; whole genome sequencing; whole genome amplification; computational biology; mixed infection



Citation: Pinto, D.; Themudo, G.; Pereira, A.C.; Botelho, A.; Cunha, M.V. Rescue of *Mycobacterium bovis* DNA Obtained from Cultured Samples during Official Surveillance of Animal TB: Key Steps for Robust Whole Genome Sequence Data Generation. *Int. J. Mol. Sci.* **2024**, *25*, 3869. <https://doi.org/10.3390/ijms25073869>

Academic Editor: Jean-Christophe Marvaud

Received: 13 March 2024

Revised: 28 March 2024

Accepted: 28 March 2024

Published: 30 March 2024



Copyright: © 2024 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

Mycobacterium bovis is the main causative agent of animal tuberculosis (TB) [1]. This infectious disease of cattle causes significant costs to the livestock sector, as applicable European legislation determines infected animals must be slaughtered and affected herds quarantined [2]. This hampers the commercialization of animal-derived products and animal movement. In Portugal, the animal TB eradication program involves passive surveillance at abattoirs and determines as well that livestock (i.e., cattle) must be regularly tested by the single-intradermal cervical comparative test or by interferon-gamma blood testing. However, *M. bovis* infection is not limited to cattle (*Bos taurus*), affecting other livestock species and occurring at the wildlife-livestock interface [3–5]. In the so-called

Portuguese TB epidemiological risk area, where hunting activities are seasonally regular, big game (wild boar, *Sus scrofa*; red deer, *Cervus elaphus*; fallow deer, *Dama dama*) must be preliminarily examined in the field by certified veterinarians. Samples from TB-compatible lesions found in hunted animals or slaughtered cattle are sent to the national reference laboratory, where they are processed by gold standard methods, such as bacteriological culture, followed by molecular identification and differentiation (by PCR) of *Mycobacterium tuberculosis* complex (MTBC) members [6,7].

Methods for culture-based detection of mycobacteria have evolved over the years. Classically, the method includes tissue maceration, decontamination, inoculation into selective solid media, and visual inspection of bacterial colonies. Automatic growth detection systems are now available, speeding up bacterial growth detection and increasing the ability to manage a large number of samples. While based on the classic culture method, the automatic systems rely on an oxygen-sensitive fluorescent dye for bacterial growth detection.

Molecular methods based on the detection of specific genome targets have been implemented over the years, either to classify the MTBC ecovar after mycobacterial isolation or to speed up laboratorial diagnosis via the direct detection of *M. bovis* in animal tissue samples [8,9]. These methods range from more traditional protocols, including restriction fragment length polymorphism analysis of the *gyrB* gene [10] and commercially available reverse hybridization assays, to real-time PCR detection of *IS6110*, *mpb70*, or the Regions of Difference (RD) [8].

Over the last years, epidemiological surveillance programs based on whole genome sequencing (WGS) have been gaining track in low prevalence and endemic settings, as shown by the large number of studies performed in France [11], Mexico [12], Portugal [13], Spain [14], Germany [15], the USA [16], New Zealand [17], the United Kingdom [18], and Uruguay [19]. Genomic surveillance has major benefits compared to traditional molecular genotyping schemes, such as spoligotyping or mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR), due to its superior resolution to identify infection sources, characterize *M. bovis* population structure, trace contacts, and inform biosecurity interventions [12]. In developed countries with ongoing official control and surveillance plans, genomic surveillance is being progressively integrated by reference laboratories into routine analyses. This implies the integration in real-time of culture-dependent and molecular methods with next-generation sequencing. Nonetheless, this implementation may raise methodological challenges.

Here, we share our experience as end-users of biological material (i.e., archived whole cell extracts) obtained from TB official surveillance and routine gold-standard processing (culture, PCR) of tissue samples from animals presumptively infected with *M. bovis* [20]. We used the extracts either from isolated colonies on solid media or from bacterial suspensions grown in automatic detection systems to perform WGS and sequence analyses up to variant [single nucleotide polymorphism (SNP)] calling and de novo assembly. We have found three main issues: (i) routine recovery of a low amount of DNA, not sufficient for WGS application; (ii) high contamination rate with DNA from other organisms, which compromised the amount and quality of the sequence data obtained; and (iii) samples containing more than one *M. bovis* strain (mixed infection), which precluded correct SNP calling. To overcome each one of these issues, we implemented rescue approaches and recovered a large number of samples that would otherwise be unusable for whole genome-based epidemiological surveillance. This has enabled the recovery of samples and the associated data, which is already of key importance, but has also uncovered steps of the upstream procedure before WGS that could be adjusted to minimize the issues found and further facilitate the integration of culture-dependent *M. bovis* detection with whole genome sequence analyses.

2. Results

2.1. Sample Collection and *M. bovis* Isolation

As a result of the implementation of the National Control and Eradication Program for Animal Tuberculosis in Portugal, coordinated by the National Veterinarian Authority (DGAV), the laboratorial confirmation of suspect cases is under the responsibility of the National Reference Laboratory (NRL) for Animal Tuberculosis, INIAV IP. Over the years, the NRL has archived DNA (i.e., frozen whole cell lysates) extracted from *M. bovis* isolates obtained during bacteriological culture of cattle or big game tissues.

For this study, we selected a set of archived DNAs based on: (i) geographical location and collection year of animal samples, selecting those that were collected in the Epidemiological Risk Area of Animal Tuberculosis for big game species, i.e., near the Spanish border in the Central East and South East area of mainland Portugal; (ii) a balanced representation of the main animal hosts of *M. bovis* in Portugal (see below).

We selected 302 samples from the 2002–2021 period, in which the presence of *M. bovis* was confirmed by culture- and PCR-dependent methods (Figure 1). Of those, 102 samples originated from cattle (*Bos taurus*), 100 from wild boar (*Sus scrofa*), and 100 from red deer (*Cervus elaphus*). A total of 219 DNA samples were recovered from bacteria grown in solid medium (2002–2018), while 83 were extracted directly from the BD BACTEC™ MGIT™ 960 automatic detection system liquid cultures (2019–2021). Of the first 219, 44 samples were first described in the study of Reis et al. [13] (Supplementary Table S1).

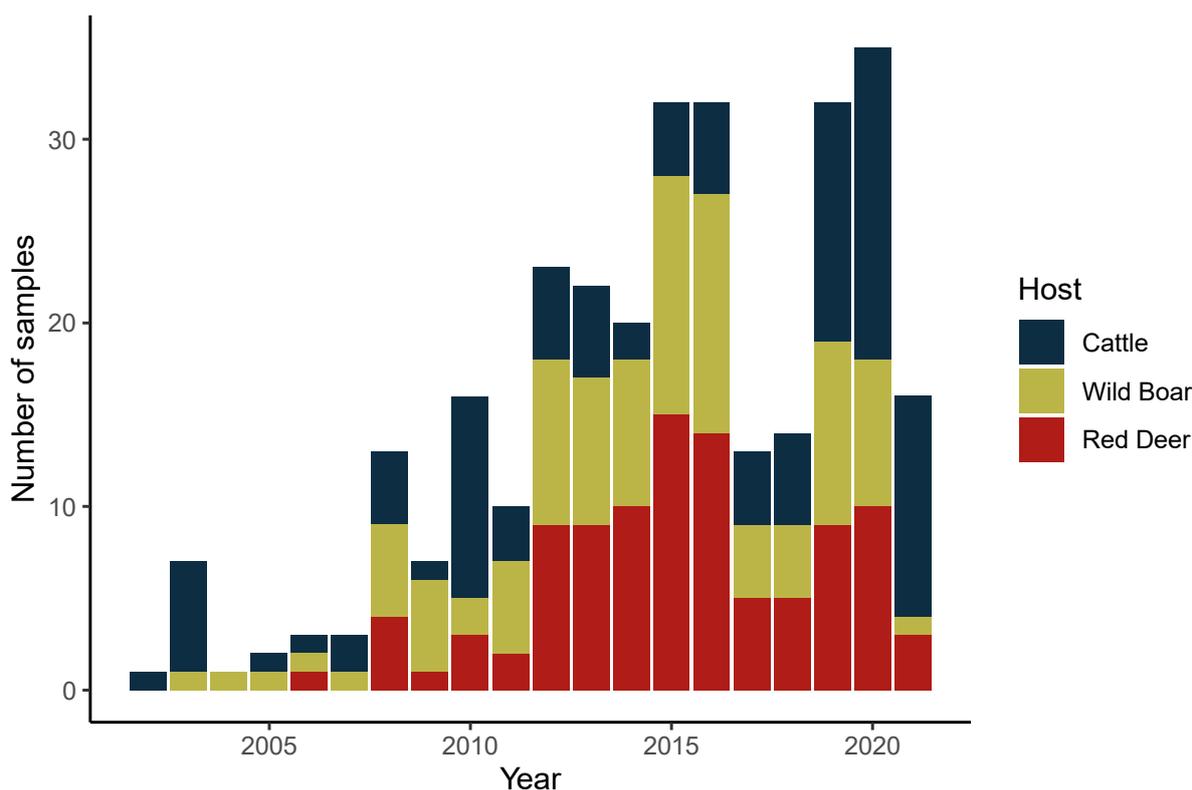


Figure 1. *Mycobacterium bovis* dataset. Distribution of analyzed *M. bovis* samples by year of collection and animal host.

2.2. Recovery of Samples Containing a Low Amount of DNA for WGS Using WGA

With the aim of including these *M. bovis* isolates in our ongoing whole genome-based epidemiological surveillance, we proceeded with the quantification of the DNA present in the archived whole cell (i.e., crude) extracts. We used a highly selective and sensitive method for double-stranded (ds) DNA, which is ideal for quantification in crude extracts. While for the 219 isolates obtained between 2002 and 2018, the amount of DNA was always

enough to proceed with commercial WGS, the same was not true for the most recent samples, obtained between 2019 and 2021. For these, we have detected between 0.1 and 242.9 ng/ μ L of dsDNA in crude extracts, with several samples containing very low DNA concentrations (Figure 2A). This translated into 250 of the 302 samples (82.8%) having enough DNA to allow direct WGS, being 100% of the samples before 2018 and only 37.3% of the more recent samples. For the 250, we proceeded with WGS using paired-end Illumina sequencing technology.

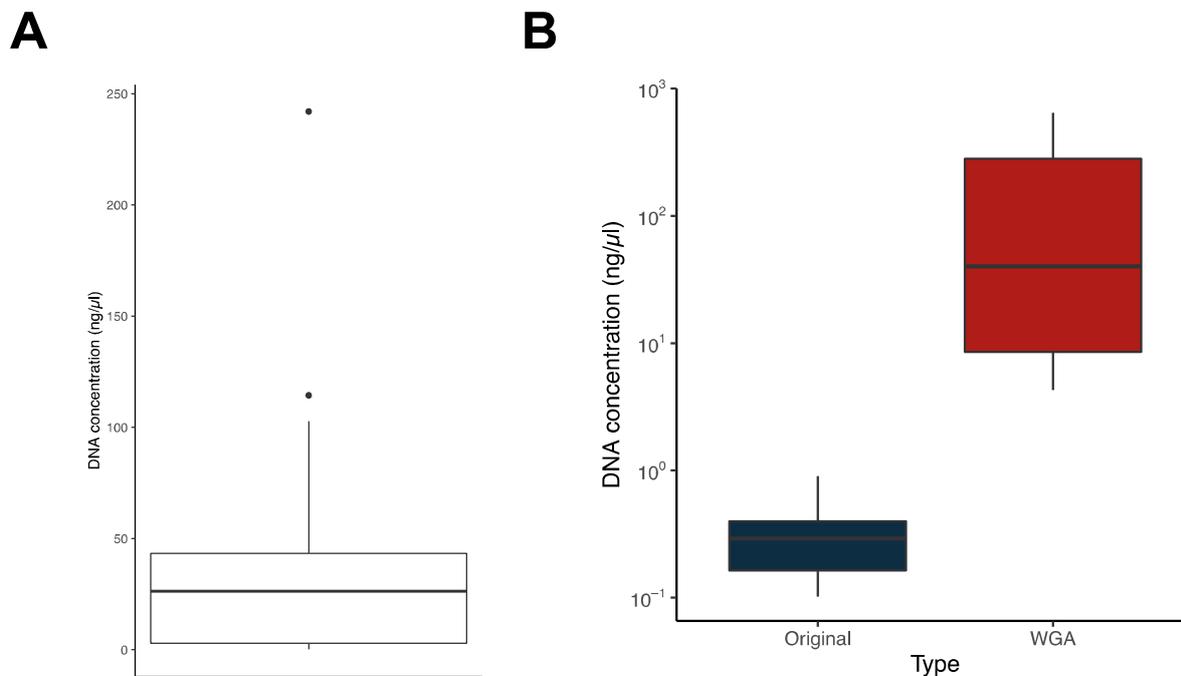


Figure 2. Whole genome amplification (WGA). (A) Box plot showing the distribution of double-stranded (ds) DNA concentration in the whole cell extracts (datasets Batches 2 and 3). (B) Box plots showing the distribution of dsDNA in original samples (in blue) and in the same samples subjected to WGA (in red).

To recover the remaining samples ($n = 52$), which represented 17.2% of the total samples and 62.7% of the samples from the 2019–2021 period, we took advantage of Phi29-dependent whole genome amplification (WGA). As shown in Figure 2B, we have successfully increased the dsDNA concentration in our samples via WGA and hence proceeded with the WGS of these DNA-enriched samples.

2.3. Recovery of Mycobacterial Reads from Heavily Contaminated DNA Samples

From a total of 302, we successfully sequenced 270 samples (89.4% of the initial set), of which 244 were original (i.e., not amplified) and 26 samples were subjected to WGA [[13] and this study]. The remaining either failed the initial DNA quality control coupled to the sequencing service or failed at library construction due to DNA degradation or low amount of library DNA obtained. These 302 samples were sequenced in three batches: Batch 1 is that of Reis et al. [13], Batch 2 contained the remaining samples up to 2018, and Batch 3 contained the samples from 2019 onwards.

For each of the successfully sequenced samples, we checked the quality of the reads with FastQC [21] and plotted the categorical classification by type of sample—i.e., those from Batch 1 [13], those from Batch 2, and those from Batch 3 divided in original and WGA (Figure 3A). The most prominent differences concern the “Per base sequence content”, “Per sequence GC content”, “Sequence Duplication Levels”, “Overrepresented sequences”, and “Sequence Length Distribution”.

As for the “Per base sequence content” category, most of the samples from Batches 1 and 2 raised an Alert or Fail (95.5% and 93.1%, respectively), indicating a non-uniform distribution of bases along the sequence. This is most likely related to the sequencing library construction kits used, as they differ in each case, and it is known that some kits generate this imbalance in base distribution. As for the “Per sequence GC content” category, 34.4% of samples fail in this criterion. This classification occurs because of the presence of sequences with very distinct GC content, which is an indication that the DNA of several different organisms was present in the samples. This does not seem to be a consequence of WGA, as the proportion of samples that fail this criterion is similar between the original and WGA samples in the same period (2019 to 2021; 69.2% in the original and 64.5% in the WGA samples). As for the “Sequence Duplication Levels” category, we observed a high number of “Fail” occurrences in the Batch 2 dataset (84.5%), which is most likely related to the specificities of the sequencing, as the library kit used (TruSeq kit) is known to generate high levels of sequence duplication. The results of the FastP analysis make this aspect of increased duplication also evident, with much higher duplication rates observed in the samples of Batch 2 and those subjected to WGA (Figure 3B). Additionally, an increase in the levels of sequence duplication in the WGA samples when compared with the original samples of the same period was registered, in agreement with previous reports [22]. As for the “Overrepresented Sequences” category, we again observe a high level of “Alert” classification in Batch 2 dataset, again due to the specificities of the sequencing. Finally, as for the “Sequence Length Distribution”, a higher number of alerts come in the Batch 1 dataset, reflecting a distinct sequencing technology that generated longer reads (250 bp vs. 150 bp in the other two datasets) [13].

As mentioned above, FastQC results suggest that more than one genome (and hence, organism) was present in several samples. The taxonomical classification of the reads made this particularly evident. Especially in the most recent samples (original and WGA), a high percentage of reads were not classified by the Kraken2 Standard library, and the percentage of reads classified as belonging to organisms of the genus *Mycobacterium* was less than 50% in over half of the samples (Figure 3C). Looking at each sample individually (Figure 3D), it becomes evident that there are reads classified as belonging to other organisms and high percentages of reads that could not be classified in the most recent samples. The genus of the most frequent contaminants in all samples was *Homo* (human DNA), most likely from cross-contamination in sample manipulation, and other known animal pathogens or regular members of the animal’s microbiota (i.e., *Clostridium*, *Serratia*, *Bacillus*, *Paenibacillus*, *Rhodococcus*, *Staphylococcus*, *Rummeliibacillus*, and *Mycolicibacter*) [23–29]. One aspect that stands out is that the percentage of contamination with Gram-negative pathogens is higher in the WGA samples (average of 6.0%) than in the original (non-amplified) samples (average of 0.3%) from 2019 to 2021. Since each sample is different from one another, we cannot exclude that this result might, by chance, be related to these particular samples, but the hypothesis that WGA has generated a bias towards these bacteria cannot be excluded.

To overcome this issue, we have filtered the mycobacterial reads before proceeding with the alignment to the *M. bovis* AF2122/97 reference genome. This filtering step was performed whenever the percentage of *Mycobacterium* reads in the sample was below 70%. As shown in Figure 3E, the obtained values of average depth of coverage were higher in the Batch 2 dataset and on the WGA samples, especially when compared to the original samples (Batch 3). Conversely, the coverage of reference genome was lower in the WGA sample set than in the original one (Figure 3E). This is most likely due to the low percentage of mycobacterial reads in several of the WGA samples (Figure 3D) along with the higher duplication rates in these samples (Figure 3B).

To include these samples in our ongoing whole genome-based epidemiological surveillance, we applied quality thresholds to our alignments to allow for robust SNP calling: at least 90% coverage of the reference genome and at least 30X depth of coverage. Based on these criteria, 201 samples (74.4% of the sequenced samples) were deemed suitable for inclusion in our ongoing epidemiological analysis (Figure 3F).

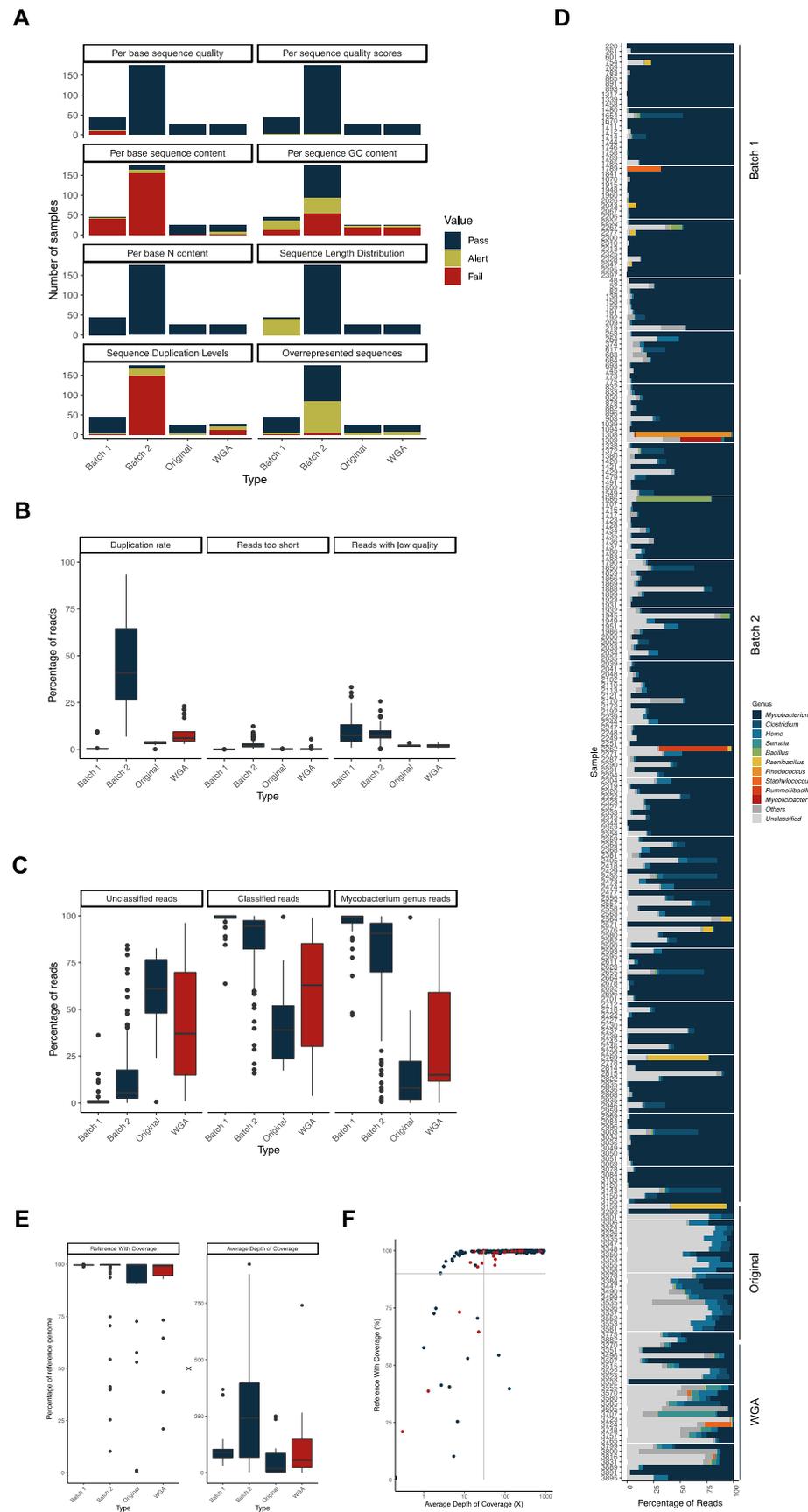


Figure 3. (A) Comparison of the sequencing read quality parameters assessed by FastQC [21], divided by datasets Reis et al. [13] (Batch 1), Batch 2, and Batch 3 (original and WGA). FastQC classified the

samples into three qualitative categories: Pass (blue), Alert (yellow), and Fail (red). (B) Box plots of the distribution of the parameters analyzed by FastP [30] in the samples from the same dataset as panel A (WGA samples in red). The determined values of duplication rate, reads too short, and reads with low quality are percentages of the total of reads. (C) Box plots of the distribution of the Kraken2 [31] determined percentages of unclassified, classified, and mycobacterial reads in each sample. The samples were grouped by dataset, as in panel A (WGA samples in red). (D) Percentage of reads per sample classified in each of the genera (color coded) indicated on the right. The “Unclassified” category comprises all the reads Kraken2 [31] was not able to classify, and the “Other” category comprises the total of reads classified to other genera or not classified at the genus level. Represented are the ten most abundant genera in the full sample set. Datasets are indicated on the right. (E) Box plot of the distribution of the vSNP determined parameters “Average Depth of Coverage” and “Reference With Coverage” (WGA samples in red). (F) Distribution of the samples—original (in blue) and WGA (in red)—by the percentage of the reference with coverage and the average depth of coverage. The gray lines indicate the adopted quality thresholds of 90% reference with coverage and 30X average depth of coverage. All datasets are included here, and those of Reis et al. [13] (Batch 1) and Batch 2 were colored blue because they have not been subjected to WGA. Note that other non-tuberculous mycobacteria can be isolated from TB-compatible lesions [32] and consequently generate *Mycobacterium* genus reads. Although our starting material (archived DNA obtained from cultured samples) had already been tested by molecular methods for the presence of *M. bovis* DNA (see Section 4 for details), we assessed the presence of mycobacterial reads other than from *M. bovis* by analyzing the reads that did not align with its reference genome. We have observed that the percentage of unaligned reads was very low, both in the samples in which the reads were taxonomically filtered due to high levels of contamination (i.e., above 30%) or those that were not (Supplementary Figure S1A). As expected, the percentage of mycobacterial reads in those unaligned samples was higher in the filtered samples (Supplementary Figure S1B), but they were assigned at very low percentages to several other mycobacterial species (Supplementary Figure S1C), sometimes even with only one read being assigned to a given species. Altogether, these data do not support contamination of our samples with other non-tuberculous mycobacteria.

We have also evaluated how our sequencing metrics of quality and contamination varied with the origin of the sample, as those of wildlife are collected in the field and may be more prone to contamination. However, no significant differences were observed between samples of different host origins in either of the evaluated metrics (Supplementary Figure S2), which indicates that the results and conclusions we obtained can be generalized to all sample origins (i.e., host species).

2.4. Genome Separation in Cases of Mixed Infections

While aligning the filtered reads to the reference genome with vSNP, multiple *M. bovis* isolates were detected in five samples, i.e., five samples in which defining SNPs had ambiguous calls. To recover samples 219, 1790, 1923, 2110, and 2244, SplitStrains was used [33]. The mixed *M. bovis* isolates were separated into ten strains, with the major strains occurring in proportions between 0.62 and 0.95 (Table 1). Of these ten separated samples, seven passed the quality thresholds (Table 1) and were included in the whole genome epidemiological surveillance dataset. Hence, this approach allowed us to recover 70% of otherwise unusable samples.

Table 1. Mixed isolates separation. Analysis was performed with SplitStrains [33] and vSNP.

Sample	Strain	Proportion	Reference with Coverage (%)	Average Depth of Coverage (X)
219	Major	0.93	84.69	133.1
	Minor	0.07	92.38	133.5
1790	Major	0.72	82.81	8.5
	Minor	0.28	65.71	4.3

Table 1. Cont.

Sample	Strain	Proportion	Reference with Coverage (%)	Average Depth of Coverage (X)
1923	Major	0.94	99.82	81.8
	Minor	0.06	99.24	76.5
2110	Major	0.62	98.38	60.0
	Minor	0.38	91.11	34.9
2244	Major	0.95	99.92	96.5
	Minor	0.05	99.34	91.0

2.5. Recovery of Fragmented Assemblies

Not all relevant analyses (e.g., comparative and functional genomics to further characterize the population regarding pangenome composition or the distribution of antibiotic resistance and virulence determinants) can be performed under a map-to-reference approach, and, in such cases, it becomes necessary to use the sequencing reads to generate a genome assembly. Hence, we generated genome assemblies of the 213 samples that passed the map-to-reference sequencing depth and genome coverage quality thresholds (30X and 90%, respectively). We assembled the reads with Unicycler [34] and analyzed the quality of the resulting assemblies with Quast [35] (Supplementary Table S2). We looked at the data the same way as in the map-to-reference data and saw that the quality of the assemblies was much lower in the samples that were subjected to WGA ($n = 12$) (Figure 4). In fact, not only was the number of contigs generated higher (Figure 4A), but also the number of bases covered in the assembly was lower (Figure 4B). In addition, the assembler was unable to generate large contigs (Figure 4C), and the percentage of reference genome covered by the assemblies was much lower (Figure 4E) when compared with the samples not subjected to WGA (i.e., those of Batches 1 and 2, and the original samples of Batch 3).

We also noticed that the assemblies of the samples subjected to WGA had a much lower GC content (Figure 4D), which led us to hypothesize that the low quality of the assemblies was due to the WGA bias towards low GC content sequences, which would generate a bias in the amplification, reducing the quality of the assemblies. However, such a hypothesis would not be in agreement with the high genome coverage obtained with the map-to-reference approach (see above) and was also not in agreement with what was seen by plotting the reference genome GC content versus the contig coverage (see Figure 4F for the full extent of the genome and Figure 4G for an insert into a 16 kb region) or previous descriptions [36]. Despite refuting our hypothesis, this analysis highlighted the lack of contigs in areas where the read coverage was very high.

To further understand this aspect, we analyzed 25 sequenced samples that were previously subjected to WGA and 26 randomly chosen samples selected from among those that were not, concerning map-to-reference mapping quality, depth of coverage, and SNP density (Supplementary Figure S3). This analysis made particularly evident the uneven coverage throughout the entire *M. bovis* chromosome, while no obvious differences were observed regarding mapping quality or SNP density, with both sample types generating very similar results. This uneven sequencing depth has already been noted in samples subjected to WGA (as were those of single cell sequencing [36]) and prompted us to use an assembler specifically developed to deal with such samples—i.e., IDBA-UD [37]. By using this assembler and posterior contig fusion, we were able to generate four additional quality assemblies (Supplementary Table S3) and rescue 36.4% of the WGA assemblies.

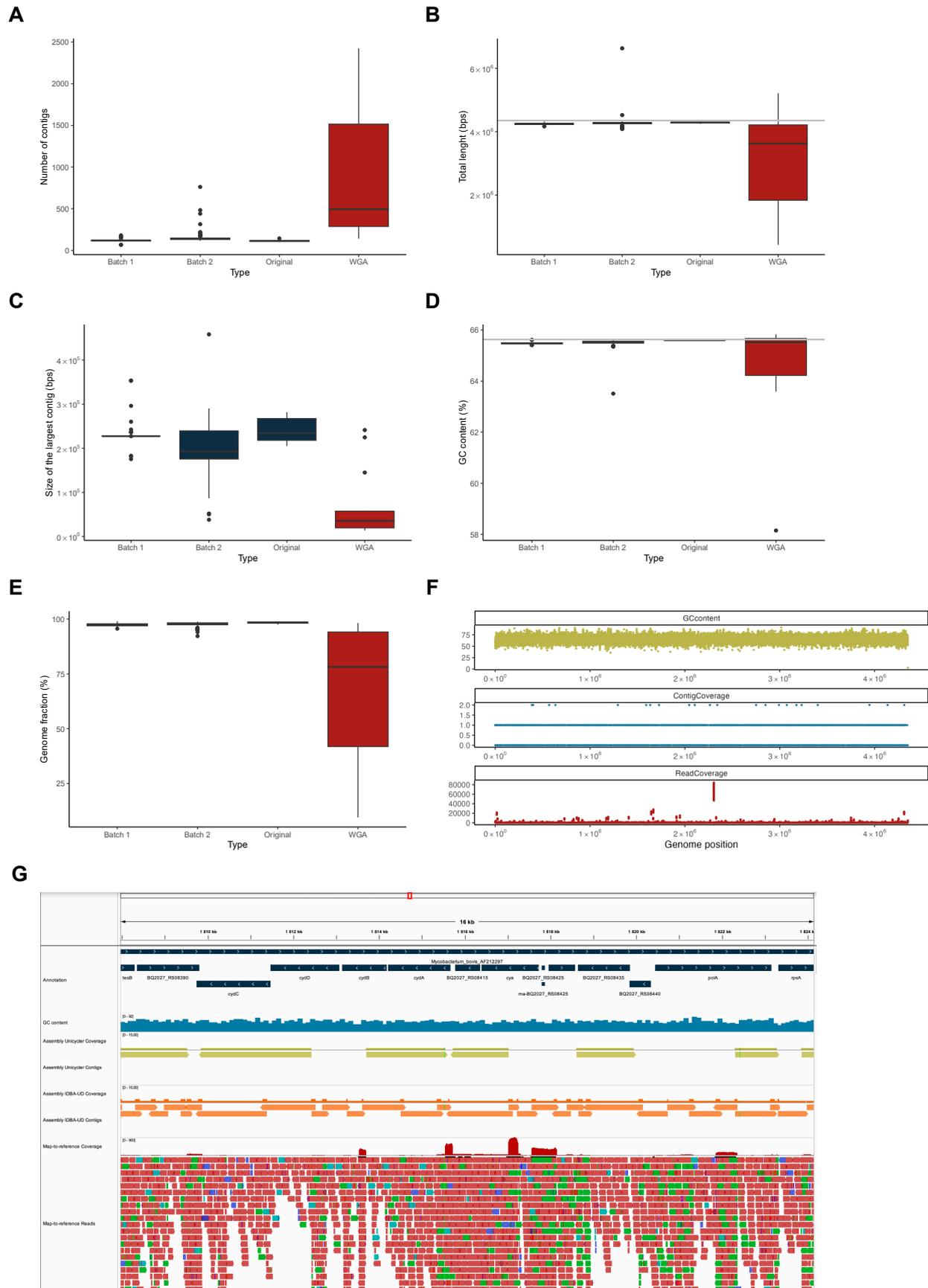


Figure 4. De novo assembly. (A–E) Box plots of the distributions of the number of contigs (A), total length of the sum of all contigs (B), size of the largest contig (C), GC content (D), and genome fraction

(E) of the de novo assemblies generated by Unicycler and assessed by Quast, divided by the original datasets of Reis et al. [13] (Batch 1), Batch 2, and Batch 3 (original and WGA). (F) Variation of GC content (calculated in 100 bp windows), contig coverage, and read coverage in the reference genome of *M. bovis* AF2122/97 of sample 3270. (G) Snippet of a genome region showing the genome annotation (dark blue), the GC content (blue), the coverage by contigs generated by Unicycler (green) and IDBA-UD (orange), and the coverage of the reads (red) in sample 3270.

3. Discussion

In this work, we explored different strategies to rescue DNA samples obtained by routine approaches performed by national reference laboratories for animal TB. Obtaining DNA in sufficient quantity and quality for WGS can be challenging when interconnecting the routine culture- and PCR-based schemes with WGS-based methods. These culture- and PCR-based methods are primarily used to confirm *M. bovis* infection in suspect animals (i.e., reactors or abattoirs) and thus implement biosecurity measures. In Portugal, as in many countries with animal TB, WGS-based epidemiological surveillance is not standard practice at NRLs. Consequently, animal tissue samples are not routinely processed for high-quality DNA extraction, which would facilitate subsequent WGS. Our study aimed to pinpoint the main challenges hindering the integration of WGS-based epidemiology into routine NRL procedures optimized for other purposes and suggest strategies to enable this integration. It is therefore important to have available a set of alternative approaches to recovering samples that suffer from the most commonly found limitations, e.g., low amounts of DNA, high rates of contamination, and mixed infections (i.e., infections with different strains of the same pathogen). Why is it so important to recover as many *M. bovis* samples as possible? Because the loss of an isolate's genome represents the loss of a link in the reconstruction of transmission chains, hampering the interpretation of the biological and epidemiological data as a whole.

Following routine procedures, the amount of DNA recovered from the samples processed in the automatic growth detection system was low, hindering direct WGS in 17.8% of total samples (which corresponded to 62.7% of the samples recovered directly from the automatic growth detection system cultures) (Supplementary Table S4). To overcome this issue, we applied WGA and successfully amplified all samples, with an increase in the amount of DNA from 9- to 2025-fold. This approach enabled the recovery of all samples for the sequencing workflow, i.e., we obtained at least 100 ng of total DNA for all 56 samples.

Another recurrent issue was the contamination with DNA from organisms other than *M. bovis*, particularly evident in the most recent samples (i.e., from sample number 3270 onwards) (Supplementary Table S4). The decontamination is made with 4% sodium hydroxide, which kills more sensitive cells but does not necessarily eliminate the DNA present in the sample because it might still be protected by cellular structures (e.g., endospores), or even if free, it may suffer denaturation but not complete degradation [38]. The culturing step should increase the amount of mycobacterial DNA by allowing mycobacterial multiplication, but if the cultures are processed immediately after being positive for growth (as assessed by fluorescence, which is a very sensitive indicator), that enrichment might not be sufficient. One methodological change that might minimize this issue would consist of extending the incubation time of the culture tubes to allow additional growth of mycobacteria.

The bacterial genera that appear in these samples, all of which have been associated with animal infections or the animal's regular microbiota [23–29], suggest that the contamination source is related to the sampled animals and not to posterior contamination introduced by handling the samples. Of course, if such organisms are present in a high amount in the sampled tissues, and since the decontamination procedure does not eliminate the DNA (as discussed above), they will generate WGS reads.

Another hypothesis that could explain this observation is that the organisms we find are resistant to the decontamination process and the antibiotics present in the growth medium. The antibiotics included in the PANTA supplement target fungi (amphotericin),

Gram-negative bacteria (polymyxin B, nalidixic acid, trimethoprim, and azlocillin), and a small group of Gram-positive bacteria (trimethoprim; *Staphylococcus* and *Streptococcus*). Hence, if the contamination we found would have come from the growth of the microorganisms in the MGIT tube, they would have to resist the decontamination process along with being resistant to the antibiotics present in the medium. *Clostridium* spp., the most common contaminants, are strict anaerobes, which makes the growth in the MGIT tube unlikely, but their spores could be resistant to decontamination, hence their DNA detection. The same is true for other spore forming organisms that were also detected (i.e., *Bacillus* spp., *Paenibacillus* spp., and *Rummellibacillus* spp.) or other actinobacteria (i.e., *Rhodococcus* spp. and *Mycolicibacter* spp.), whose more robust cell envelope might make them more resistant to the decontamination protocol. Finally, the carbon sources available in the Middlebrook medium (supplemented with OADC) are not primary carbon sources readily used by non-mycobacterial organisms. Altogether, the growth of contaminants seems unlikely in these conditions.

We have attempted to overcome this high contamination issue by filtering out the mycobacterial WGS reads. In almost a quarter of all sequenced samples, the fraction of mycobacterial reads was low, and the alignment to the reference genome did not have sufficient quality to be included in the subsequent whole-genome based epidemiological analysis. One possible way to overcome this issue is by increasing the sequencing depth (i.e., the number of reads generated with WGS) as a way to increase the number of mycobacterial reads. Another approach might rely on whole genome enrichment, which was already successfully applied to enrich *Mycobacterium tuberculosis* in sputum samples [39].

Another challenge was the presence of more than one *M. bovis* strain in the same sample, as defined by the presence of ambiguous defining SNPs. This phenomenon might occur if two *M. bovis* strains independently infect the same animal host or if clonal evolution occurs inside the host during the typical TB chronic infection phenotype [40]. Strain separation, although effective, leads to an expected reduction in the percentage of genome coverage and average sequencing depth, given that the reads containing the SNPs are separated between the two resulting strains—i.e., major and minor strains. Despite this additional limitation, seven out of the ten resulting strains could be rescued and included in the ongoing genomic epidemiology studies.

While generating de novo genome assemblies, we encountered a limitation related to WGA, i.e., in the samples subjected to this amplification process, the assemblies had very low continuity. Our analysis correlated the areas in which a contig could not be generated to regions of extremely high read coverage. Given that most of the de novo assemblers identify regions with above-average coverage as repetitive regions and break the contigs on those regions for not being able to unequivocally assign a read to a given repeat region, we were prompted to explore assemblers especially designed to generate de novo assemblies from data with uneven coverage. Our approach allowed us to recover four out of the 11 WGA samples that failed in the regular de novo assembly, but this must be taken into account when considering the use of WGA to rescue samples with low amounts of DNA, as it is a strategy that performs better when the data is to be used for map-to-reference approaches than for de novo assemblies.

A subset of the samples analyzed in this study had been previously used in published studies. Specifically, the map-to-reference alignments were used for phylogenomic analysis of the *M. bovis* population circulating between wildlife and livestock in Portugal [13], while de novo assemblies were used to support pangenome analysis [41]. The remaining samples (83.7%) have been incorporated into our ongoing studies encompassing both phylogenomics and phylodynamics (relying on the generated map-to-reference alignments), as well as the expansion of the pangenome analysis (using the de novo assemblies). These recovery strategies have not only prevented the loss of several genomes but have also permitted the coherent integration of these genomes into the *M. bovis* population phylogenetic trees for Portugal. This integration occurred without raising any suspicions of potential artifacts introduced by our recovery methods. In other words, these genomes were distributed

across various branches of the phylogenetic trees without any noticeable clustering [42]. This outcome was entirely expected due to the lack of any disparity in SNP density across genomes subjected to whole genome amplification (WGA) compared to those that were not (Supplementary Figure S3). Regarding the strains recuperated through read splitting, their positioning in the tree indicated that they are unrelated, suggesting instances of mixed infections rather than evolution within a single host [42]. Consequently, when these recovered genomes are applied in subsequent analyses, their behavior aligns with that of genomes that did not require the implementation of recovery strategies. This reinforces the validity of the recovery strategies employed in this study and sustains the inclusion of the resulting genomes in downstream epidemiological and evolutionary analyses.

In conclusion, when adding to the routine culture and PCR-based detection of *M. bovis* a series of procedures leading to the obtention of a complete genome sequence, one might expect additional challenges. These challenges take form in the contamination of the sample, the low amount of available DNA, and the potential presence of multiple *M. bovis* strains in the same sample (Supplementary Table S4). Despite being undesirable, these issues can be overcome. Here we showed one way to overcome these challenges, but others could certainly be explored with the necessary adjustments to the particularities of the samples and resources available. We were able to recover 100% of the samples with a low amount of total DNA through WGA, 60.8% of the contaminated samples through read filtering, and 70% of the mixed strains through strain separation. Overall, these approaches allowed the recovery of 62 samples that would have been otherwise unusable for epidemiological purposes.

Thus, the direct coupling of genomics to *M. bovis* standard diagnosis is possible, provided that specific laboratory and computational strategies are applied to circumvent the limitations imposed. Modification of the diagnostic protocol and a dedicated pipeline integrating computational tools to tackle different data challenges minimize biological data losses, optimize available material resources, and greatly facilitate the coupling of diagnosis to genomic surveillance. In fact, the results shown here also argue in favor of the inclusion of an additional step of culture of *M. bovis* in solid medium after the automatic detection whenever WGS is the final goal. The advantage of including this step, present in the conventional method but not in the automatic one, is particularly obvious in the amount and purity of the DNA that can be retrieved. This would facilitate the downstream computational analysis and the likelihood of obtaining useful sequencing data for genomic epidemiology purposes.

Surveillance of animal TB has many diagnostic challenges, and the culture of *M. bovis* remains time consuming. In the future, WGS applied to DNA directly extracted from tissue samples could be a very useful tool, shortening the overall time for diagnosis, improving surveillance, and enabling the timely implementation of intervention measures. However, many challenges for its full optimization and routine implementation exist for the moment.

Here, we have applied these rescue strategies to our particular problem of recovering *M. bovis* genomes to answer animal TB-related epidemiological questions. However, the value of the implementation of these rescue strategies is not limited to our particular model system and has a broader application. In fact, they can be applied to any other sample type in which similar problems might arise, as is the case of blood cultures for the diagnosis of bloodstream infections, where the percentage of reads belonging to the pathogen is low [43], impairing the capability of predicting antibiotherapy susceptibility from WGS data [44], amongst other limitations [45].

4. Materials and Methods

4.1. Sample Collection and Processing

Tissue samples obtained between 2002 and 2021 from hunted wild boar and red deer and slaughtered cattle, positive in the intradermal tuberculin test and/or exhibiting suspect TB-lesions during veterinarian inspection, were submitted to the National Reference Laboratory for animal TB (INIAV, I.P.) for analysis.

Sample processing was performed according to the OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals [20]. Before 2018, samples were processed by inoculation in Löwenstein-Jensen solid medium and simultaneously with a BD BACTEC™ 9000 automatic detection system (Becton, Dickinson and Company, Franklin Lakes, NJ, USA), and from 2019 onwards with a BD BACTEC™ MGIT™ 960 system.

Collected tissues were homogenized with a pestle and mortar, followed by decontamination with 4% sodium hydroxide for 30 min (15 min from 2016 onwards) to inactivate other contaminating bacteria. After neutralization with hydrogen chloride and centrifugation of the resulting suspension, the sediment was inoculated into Löwenstein–Jensen with pyruvate slants. The inoculated tubes were then incubated at 37 °C for up to 42 days, while continuously monitoring the medium for the presence of bacterial colonies. Colonies from positive cultures were visually inspected for the characteristic morphology of *M. bovis*. Some of those isolates from Batch 2 ($n = 29$) were subcultured in Middlebrook 7H9 (BD Difco, Franklin Lakes, NJ, USA) and passaged in Middlebrook 7H10 (BD Difco, Franklin Lakes, NJ, USA). Both media were supplemented with 5% sodium pyruvate and 10% ADS (50 g albumin, 20 g glucose, and 8.5 g sodium chloride in 1 L of water). Incubation was at 37 °C in a level 3 biosecurity facility.

In parallel, and according to the BD BACTEC™ manufacturer instructions, collected tissues were homogenized and decontaminated as described above and incubated in Middlebrook 7H9 broth supplemented with OADC and PANTA (Polymyxin B, Amphotericin B, Nalidixic Acid, Trimethoprim, and Azlocillin) and an oxygen-sensitive fluorescent compound. The inoculated tubes were incubated at 37 °C for up to 42 days while continuously monitoring the fluorescence signal representative of bacterial growth. Once the fluorescence crossed the threshold, the sample was deemed positive and readily processed.

Cell biomass (one colony) grown on solid medium or 1 mL of liquid culture suspension were collected, and the cells were recovered by centrifugation. The cell pellet was then washed with 1 mL of phosphate buffered saline (PBS), resuspended in 200 µL of PBS, and heat-inactivated at 99 °C in a water bath for 30 min. The suspension was cleared by centrifugation, and the supernatant (i.e., the crude whole cell extract) was used for downstream analysis. Until use, these were archived at –20 °C.

4.2. *Mycobacterium bovis* Identification

The presence of *M. bovis* in the whole cell extracts was confirmed by one of three methods: (i) by restriction fragment length polymorphism analysis of the PCR-amplified *gyrB* gene after hydrolysis with *RsaI* and *SacII* as described before [10]; (ii) by reverse hybridization assays of INNO-LiPA Mycobacteria (Innogenetics, Zwijnaarde, Ghent, Belgium) or GenoType *Mycobacterium* (Hain diagnostics, Nehren, Tübingen, Germany), following the manufacturer's instructions; (iii) by real-time PCR detection of Regions of Difference (RD) 1, 4, and 9 as described elsewhere [8].

4.3. DNA Concentration Determination

DNA quantification was performed using the Qubit dsDNA HS assay kit or the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. This kit is highly selective for double-stranded (ds) DNA and, hence, ideal for the quantification of dsDNA in crude extracts.

4.4. Whole Genome Amplification (WGA)

WGA was performed using the GenomiPhi™ V2 DNA Amplification Kit (Cytiva, Marlborough, MA, USA) according to the manufacturer's instructions with a starting dsDNA amount of up to 10 ng.

4.5. Whole Genome Sequencing (WGS) and Data Analysis

The WGS of the samples described in Reis et al. [13] (Batch 1) was performed by MiSeq and HiSeq Illumina technology (paired-end 250 pb) at the United States Department of

Agriculture (USDA, Ames, IA, USA) with the Nextera XT DNA Library Prep Kit from Illumina. WGS of the samples in Batch 2 was commercially conducted by STABVida (Caparica, Portugal) using the Illumina NovaSeq platform with the TrueSeq Library Prep kit (paired-end 150 bp). WGS of the samples from Batch 3 was commercially performed by Eurofins Genomics (Konstanz, Germany) using the Genome Sequencer Illumina NovaSeq platform (Illumina, San Diego, CA, USA) (paired-end 150 bp reads).

The new raw data (Supplementary Table S1) have been deposited on a public domain server at the NCBI SRA database, under BioProject accession number PRJNA946560.

Upon receiving the read files, a quality control check was performed with a local installation of FastQC (version 0.11.9) [21]. For read filtering, trimming, and sequence overrepresentation analysis, FastP (version 0.23.2) [30], hosted on the Galaxy server [46], was used with “overrepresented sequence analysis”. The cleaned-up reads were submitted to Kraken2 (version 2.1.1) [31], also hosted at Galaxy server [46], for read taxonomical classification using its Standard database. KrakenTools (<https://github.com/jenniferlu717/KrakenTools> (accessed on 18 June 2021)) were then used to extract reads classified as belonging to the *Mycobacterium* taxon (NCBI tax ID 1763) every time the *Mycobacterium* read content in the sample was below 70%. Mycobacterial reads were aligned to the genome of *M. bovis* AF2122/97 (NC_002945.4) using the vSNP pipeline (<https://github.com/USDA-VS/vSNP> (accessed on 21 February 2022)). Quality control of the alignments was performed with QualiMap BamQC (version 2.2.2c+galaxy1) [47,48], hosted at Galaxy Europe with 4350 bins, and SNP density per kilobase was determined with VCFtools (version 0.1.16) [49].

4.6. Strain Separation

Whenever appropriate, and upon the alert from vSNP of mixed reads, strain separation was performed with SplitStrains [33], starting from the vSNP generated alignment to the reference BAM file. The reference was changed to *M. bovis* AF2122/97 (NC_002945.4). The same parameters were used for all the strains: -s, specifying the start position on the genome, was set to 1; -e, specifying the end position on the genome, was set to 4,349,100; -fd was set to 0 to force SplitStrains to not ignore sites based on the depth of coverage (as vSNP also does not); -u was set to 99 to ignore proportions only above this value; -l was set to 1, to ignore proportions only below this value; -m was set to 5, to exclude reads below 5 map quality; -q was set to 5, to exclude reads below 5 quality; -ft was set to 0 so as to not exclude any position from the analysis.

Upon read separation, reads were extracted from the original BAM alignment files with the SplitStrains package script rmreads.py. Each read set was once again aligned to the reference genome with the vSNP pipeline.

4.7. De Novo Genome Assembly

Genomes were de novo assembled with the Unicycler pipeline v.0.5.0, available at <https://github.com/rrwick/Unicycler> (accessed on 1 July 2022) [34]. For short read only data, Unicycler assembles genomes de novo, optimizing the use of SPAdes [50] by evaluating which k-mer size produces the best assembly. We applied a conservative bridging mode to avoid misassembly, and the k-mer size was searched and selected between 20 and 95% of the read length. Contigs smaller than 300 bp were removed, considering the reads’ average size and following SPAdes guidelines [51].

IDBA-UD assembler [37], hosted on the Galaxy Europe server, was used to generate the assemblies from the samples subjected to WGA. The assembler was run with a minimum k value of 5, a maximum k value of 315, and an increment of k-mer of 10 in each iteration. The seed k-mer size for alignment was set to 15, and the minimum size of contig considered was 500 bps. Minimap2 [52] was used for mapping the contigs to the reference genome and Samtools [53] for contig fusion.

To assess the quality of the de novo assemblies, we ran the Quast pipeline (<http://quast.sourceforge.net/quast.html> (accessed on 1 July 2022)), which remaps contigs to the

M. bovis AF2122/97 reference genome and calculates a number of metrics, such as genome fraction covered [35].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25073869/s1> [21,30,54].

Author Contributions: Study design, D.P. and M.V.C.; provision of DNA samples, A.B.; computational analysis, D.P., A.C.P. and G.T.; data analysis, D.P., A.C.P., G.T. and M.V.C.; data visualization, D.P.; writing, D.P. and M.V.C. with contributions from A.C.P. and G.T. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge funding from Fundação para a Ciência e a Tecnologia, IP (FCT)/MCTES through national funds (PIDDAC) and co-funding by the European Regional Development Fund (FEDER) of the European Union, through the Lisbon Regional Operational Program and the Competitiveness and Internationalization Operational Program for Portugal 2020 or other programs that may succeed it in the scope of project “Colossus: Control Of tubercuLOsis at the wildlife/livestock interface uSing innovative natUre-based Solutions” (references PTDC/CVT- CVT/29783/2017, LISBOA-01-0145-FEDER-029783, POCI-01-0145- FEDER-029783). Strategic funding from FCT to cE3c and BioISI Research Units (DOI:10.54499/UIDB/00329/2020 and DOI:10.54499/UIDB/04046/2020) and to the associate lab CHANGE (LA/P/0121/2020) are gratefully acknowledged. ACP was funded by FCT (SFRH/BD/136557/2018). MVC acknowledges institutional support from FCT (DOI: 10.54499/CEECINST/00032/2018/CP1523/CT0005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sequence data included in this work are deposited under Bioproject accession numbers PRJNA682618 and PRJNA946560 at a public domain server in the National Centre for Biotechnology Information (NCBI) SRA database.

Acknowledgments: We thank Rogério Tenreiro and Rute Fonseca for technical suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Pereira, A.C.; Reis, A.C.; Ramos, B.; Cunha, M.V. Animal Tuberculosis: Impact of Disease Heterogeneity in Transmission, Diagnosis and Control. *Transbound. Emerg. Dis.* **2020**, *67*, 1828–1846. [[CrossRef](#)] [[PubMed](#)]
2. Food, E.; Authority, S. The European Union One Health 2018 Zoonoses Report. *EFSA J.* **2019**, *17*. [[CrossRef](#)]
3. Cunha, M.V.; Matos, F.; Canto, A.; Albuquerque, T.; Alberto, J.R.; Aranha, J.M.; Vieira-Pinto, M.; Botelho, A. Implications and Challenges of Tuberculosis in Wildlife Ungulates in Portugal: A Molecular Epidemiology Perspective. *Res. Vet. Sci.* **2012**, *92*, 225–235. [[CrossRef](#)] [[PubMed](#)]
4. Ramos, B.; Pereira, A.C.; Reis, A.C.; Cunha, M.V. Estimates of the Global and Continental Burden of Animal Tuberculosis in Key Livestock Species Worldwide: A Meta-Analysis Study. *One Health* **2020**, *10*, 100169. [[CrossRef](#)] [[PubMed](#)]
5. Reis, A.C.; Ramos, B.; Pereira, A.C.; Cunha, M.V. The Hard Numbers of Tuberculosis Epidemiology in Wildlife: A Meta-Regression and Systematic Review. *Transbound. Emerg. Dis.* **2021**, *68*, 3257–3276. [[CrossRef](#)] [[PubMed](#)]
6. DGAV Plano de Controlo e Erradicação de Tuberculose Em Caça Maior 2011. Available online: https://www.dgav.pt/wp-content/uploads/2021/03/Plano_controlo_erradicacao_Tuberculose_Caca_Maior.pdf (accessed on 17 July 2023).
7. DGAV Programme for the Eradication of Bovine Tuberculosis, Bovine Brucellosis or Sheep and Goat Brucellosis 2019. Available online: https://www.dgav.pt/wp-content/uploads/2021/01/PT_BB-2018_ref12178web.pdf (accessed on 17 July 2023).
8. Costa, P.; Amaro, A.; Ferreira, A.S.; Machado, D.; Albuquerque, T.; Couto, I.; Botelho, A.; Viveiros, M.; Inácio, J. Rapid Identification of Veterinary-Relevant Mycobacterium Tuberculosis Complex Species Using 16S rDNA, IS6110 and Regions of Difference-Targeted Dual-Labelled Hydrolysis Probes. *J. Microbiol. Methods* **2014**, *107*, 13–22. [[CrossRef](#)]
9. Lorente-Leal, V.; Liandris, E.; Castellanos, E.; Bezos, J.; Domínguez, L.; de Juan, L.; Romero, B. Validation of a Real-Time PCR for the Detection of Mycobacterium Tuberculosis Complex Members in Bovine Tissue Samples. *Front. Vet. Sci.* **2019**, *6*. [[CrossRef](#)] [[PubMed](#)]
10. Abass, N.A.; Suleiman, K.M.; El Jalii, I.M. Differentiation of Clinical Mycobacterium Tuberculosis Complex Isolates by Their GyrB Polymorphism. *Indian J. Med. Microbiol.* **2010**, *28*, 26–29. [[CrossRef](#)] [[PubMed](#)]

11. Duault, H.; Michelet, L.; Boschirolì, M.-L.; Durand, B.; Canini, L. A Bayesian Evolutionary Model towards Understanding Wildlife Contribution to F4-Family Mycobacterium Bovis Transmission in the South-West of France. *Vet. Res.* **2022**, *53*, 28. [[CrossRef](#)] [[PubMed](#)]
12. Ortiz, A.P.; Perea, C.; Davalos, E.; Velázquez, E.F.; González, K.S.; Camacho, E.R.; García Latorre, E.A.; 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. Vet. Sci.* **2021**, *8*. [[CrossRef](#)]
13. Reis, A.C.; Salvador, L.C.M.; Robbe-Austerman, S.; Tenreiro, R.; Botelho, A.; Albuquerque, T.; Cunha, M.V. Whole Genome Sequencing Refines Knowledge on the Population Structure of Mycobacterium Bovis from a Multi-Host Tuberculosis System. *Microorganisms* **2021**, *9*, 1585. [[CrossRef](#)]
14. Perea, C.; Ciaravino, G.; Stuber, T.; Thacker, T.C.; Robbe-Austerman, S.; Allepuz, A.; de Val, B.P. Whole-Genome SNP Analysis Identifies Putative Mycobacterium Bovis Transmission Clusters in Livestock and Wildlife in Catalonia, Spain. *Microorganisms* **2021**, *9*, 1629. [[CrossRef](#)]
15. 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](#)]
16. Orloski, K.; Robbe-Austerman, S.; Stuber, T.; Hench, B.; Schoenbaum, M. Whole Genome Sequencing of Mycobacterium Bovis Isolated From Livestock in the United States, 1989–2018. *Front. Vet. Sci.* **2018**, *5*, 00253. [[CrossRef](#)]
17. 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](#)]
18. Trewby, H.; Wright, D.; Breadon, E.L.; Lycett, S.J.; Mallon, T.R.; McCormick, C.; Johnson, P.; Orton, R.J.; Allen, A.R.; Galbraith, J.; et al. Use of Bacterial Whole-Genome Sequencing to Investigate Local Persistence and Spread in Bovine Tuberculosis. *Epidemics* **2016**, *14*, 26–35. [[CrossRef](#)]
19. Lasserre, M.; Berná, L.; Greif, G.; Díaz-Viraqué, F.; Iraola, G.; Naya, H.; Castro-Ramos, M.; Juambeltz, A.; Robello, C. Whole-Genome Sequences of Mycobacterium Bovis Strain MbURU-001, Isolated from Fresh Bovine Infected Samples. *Genome Announc.* **2015**, *3*, e01237-15. [[CrossRef](#)]
20. World Organization for Animal Health. *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*, 6th ed.; World Organization for Animal Health: Geneva, Switzerland, 2008.
21. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 1 January 2020).
22. Pinard, R.; de Winter, A.; Sarkis, G.J.; Gerstein, M.B.; Tartaro, K.R.; Plant, R.N.; Egholm, M.; Rothberg, J.M.; Leamon, J.H. Assessment of Whole Genome Amplification-Induced Bias through High-Throughput, Massively Parallel Whole Genome Sequencing. *BMC Genom.* **2006**, *7*, 216. [[CrossRef](#)]
23. Otter, A.; Uzal, F.A. Clostridial Diseases in Farm Animals: 1. Enterotoxaemias and Other Alimentary Tract Infections. *In Pract.* **2020**, *42*, 219–232. [[CrossRef](#)]
24. Keck, N.; Dunie-merigot, A.; Dazas, M.; Hirschaud, E.; Laurence, S.; Gervais, B.; Madec, J.Y.; Haenni, M. Long-Lasting Nosocomial Persistence of Chlorhexidine-Resistant *Serratia Marcescens* in a Veterinary Hospital. *Vet. Microbiol.* **2020**, *245*, 108686. [[CrossRef](#)]
25. Logan, N.A. Bacillus Species of Medical and Veterinary Importance. *J. Med. Microbiol.* **1988**, *25*, 157–165. [[CrossRef](#)]
26. Rampacci, E.; Sforza, M.; Dentini, A.; Di Matteo, I.; Lidano, P.; Capucci, C.; Passamonti, F. Paenibacillus Amyolyticus Osteomyelitis in a Poodle Dog: Case Report and Literature Review. *J. Vet. Diagn. Investig.* **2022**, *34*, 703–708. [[CrossRef](#)]
27. Vázquez-Boland, J.A.; Giguère, S.; Hapeshi, A.; MacArthur, I.; Anastasi, E.; Valero-Rello, A. Rhodococcus Equi: The Many Facets of a Pathogenic Actinomycete. *Vet. Microbiol.* **2013**, *167*, 9–33. [[CrossRef](#)]
28. Derakhshani, H.; Plaizier, J.C.; De Buck, J.; Barkema, H.W.; Khafipour, E. Composition and Co-Occurrence Patterns of the Microbiota of Different Niches of the Bovine Mammary Gland: Potential Associations with Mastitis Susceptibility, Udder Inflammation, and Teat-End Hyperkeratosis. *Anim. Microbiome* **2020**, *2*, 11. [[CrossRef](#)]
29. Supré, K.; Roupie, V.; Ribbens, S.; Stevens, M.; Boyen, F.; Roels, S. Short Communication: Mycolicibacterium Smegmatis, Basonym Mycobacterium Smegmatis, Causing Pyogranulomatous Mastitis and Its Cross-Reactivity in Bovine (Para)Tuberculosis Testing. *J. Dairy Sci.* **2019**, *102*, 8405–8409. [[CrossRef](#)]
30. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)]
31. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [[CrossRef](#)]
32. Tingan, T.K.; Mensah, G.I.; Agyekum, E.B.; Amanor, I.B.; Addo, S.O.; Ayamdoo, Y.I.; Duah, M.S.; Mosi, L.; Addo, K.K. Non-Tuberculous Mycobacteria, Not Mycobacterium Bovis, Are a Significant Cause of TB-like Lesions Observed in Slaughtered Cattle in Ghana. *IJID Reg.* **2022**, *3*, 8–14. [[CrossRef](#)]
33. Gabbassov, E.; Moreno-Molina, M.; Comas, I.; Libbrecht, M.; Chindelevitch, L. SplitStrains, a Tool to Identify and Separate Mixed Mycobacterium Tuberculosis Infections from WGS Data. *Microb. Genom.* **2021**, *7*, 000607. [[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. Mikheenko, A.; Prjibelski, A.; Saveliev, V.; Antipov, D.; Gurevich, A. Versatile Genome Assembly Evaluation with QUAST-LG. *Bioinformatics* **2018**, *34*, i142–i150. [[CrossRef](#)]
36. Rodrigue, S.; Malmstrom, R.R.; Berlin, A.M.; Birren, B.W.; Henn, M.R.; Chisholm, S.W. Whole Genome Amplification and de Novo Assembly of Single Bacterial Cells. *PLoS ONE* **2009**, *4*, e0006864. [[CrossRef](#)]
37. Peng, Y.; Leung, H.C.M.; Yiu, S.M.; Chin, F.Y.L. IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth. *Bioinformatics* **2012**, *28*, 1420–1428. [[CrossRef](#)]
38. Wang, X.; Lim, H.J.; Son, A. Characterization of Denaturation and Renaturation of DNA for DNA Hybridization. *Environ. Health Toxicol.* **2014**, *29*, e2014007. [[CrossRef](#)]
39. Soundararajan, L.; Kambli, P.; Priyadarshini, S.; Let, B.; Murugan, S.; Irvatham, C.; Tornheim, J.A.; Rodrigues, C.; Gupta, R.; Ramprasad, V.L. Whole Genome Enrichment Approach for Rapid Detection of Mycobacterium Tuberculosis and Drug Resistance-Associated Mutations from Direct Sputum Sequencing. *Tuberculosis* **2020**, *121*, 101915. [[CrossRef](#)]
40. Thacker, T.C.; Palmer, M.V.; Robbe-Austerman, S.; Stuber, T.P.; Waters, W.R. Anatomical Distribution of Mycobacterium Bovis Genotypes in Experimentally Infected White-Tailed Deer. *Vet. Microbiol.* **2015**, *180*, 75–81. [[CrossRef](#)]
41. Reis, A.C.; Cunha, M.V. The Open Pan-Genome Architecture and Virulence Landscape of Mycobacterium Bovis. *Microb. Genomics* **2021**, *7*, 000664. [[CrossRef](#)]
42. Pereira, A.C.; Reis, A.C.; Cunha, M.V. Genomic Epidemiology Sheds Light on the Emergence and Spread of Mycobacterium Bovis Eu2 Clonal Complex in Portugal. *Emerg. Microbes Infect.* **2023**, *12*, 2253340. [[CrossRef](#)]
43. Grumaz, S.; Stevens, P.; Grumaz, C.; Decker, S.O.; Weigand, M.A.; Hofer, S.; Brenner, T.; von Haeseler, A.; Sohn, K. Next-Generation Sequencing Diagnostics of Bacteremia in Septic Patients. *Genome Med.* **2016**, *8*, 73. [[CrossRef](#)]
44. Anson, L.W.; Chau, K.; Sanderson, N.; Hoosdally, S.; Bradley, P.; Iqbal, Z.; Phan, H.; Foster, D.; Oakley, S.; Morgan, M.; et al. DNA Extraction from Primary Liquid Blood Cultures for Bloodstream Infection Diagnosis Using Whole Genome Sequencing. *J. Med. Microbiol.* **2018**, *67*, 347–357. [[CrossRef](#)]
45. Pekar, N.; Couto, N.; Sinha, B.; Rossen, J.W. Diagnosis of Bloodstream Infections from Positive Blood Cultures and Directly from Blood Samples: Recent Developments in Molecular Approaches. *Clin. Microbiol. Infect.* **2018**, *24*, 944–955. [[CrossRef](#)] [[PubMed](#)]
46. Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Čech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B.A.; et al. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [[CrossRef](#)] [[PubMed](#)]
47. García-Alcalde, F.; Okonechnikov, K.; Carbonell, J.; Cruz, L.M.; Götz, S.; Tarazona, S.; Dopazo, J.; Meyer, T.F.; Conesa, A. Qualimap: Evaluating next-Generation Sequencing Alignment Data. *Bioinformatics* **2012**, *28*, 2678–2679. [[CrossRef](#)] [[PubMed](#)]
48. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* **2016**, *32*, 292–294. [[CrossRef](#)] [[PubMed](#)]
49. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
50. Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinform.* **2020**, *70*, e102. [[CrossRef](#)] [[PubMed](#)]
51. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
52. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)] [[PubMed](#)]
53. 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](#)]
54. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments. *Genome Biol.* **2014**, *15*, R46. [[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.