Tuberculosis 116 (2019) 61–66

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: www.elsevier.com/locate/tube

Subcultured *Mycobacterium tuberculosis* isolates on different growth media are fully representative of bacteria within clinical samples

Charlotte Genestet[a,b,*], Elisabeth Hodille[a,c], Emilie Westeel[a,b], Christophe Ginevra[a,c], Florence Ader[d], Samuel Venner[e], Gérard Lina[a,c], Juliet E. Bryant[a,b], Jean-Luc Berland[a,b,1], Oana Dumitrescu[a,c,1], on behalf of the Lyon TB study group

[a] Centre International de Recherche en Infectiologie (CIRI), INSERM U1111, Université de Lyon, Lyon, France
[b] Fondation Mérieux, Lyon, France
[c] Hospices Civils de Lyon, Institut des Agents Infectieux, Laboratoire de Bactériologie, Lyon, France
[d] Service des Maladies Infectieuses et Tropicales, Hôpital de la Croix Rousse, Hospices Civils de Lyon, Lyon, France
[e] Laboratoire de Biométrie et Biologie Evolutive (LBBE), CNRS UMR 5558, Université Lyon 1, Villeurbanne, France

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Whole genome sequencing
Virulence determinants
Genome-wide diversity

ABSTRACT

Mycobacterium tuberculosis (Mtb) whole genome sequencing (WGS) plays an increasingly important role in tuberculosis diagnosis and research. WGS is typically performed on biobanked isolates obtained by subculture during diagnosis. Genetic variation upon culturing is known to occur in other bacterial species. However, little is understood regarding the impact of different subculture media on genome-wide diversity and variant selection in Mtb. Here we compared WGS derived from direct sequencing of sputa samples to WGS sequences from isolates subcultured on 3 different media. Based on analysis of single nucleotide polymorphisms (SNPs), there was no evidence of variant selection caused by the different culture media used, indicating that subcultured clinical strains can be reliably used to explore genetic determinants of Mtb pathogenesis and epidemiological features.

1. Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (Mtb) complex remains one of the most prevalent and deadly infectious diseases worldwide with an estimated 10 million cases (range, 9.0–11.1 million) and 1.6 million deaths in 2017 [1]. Mtb is thought to have evolved by clonal expansion from a sublineage of *Mycobacterium canettii*-like progenitors that infected early mammals and has since co-evolved in mammalian hosts. Currently circulating Mtb isolates represent successful lineages that have evolved over the last 70 000 years [2]. Mtb infection results in a spectrum of clinical outcomes that ranges from latent asymptomatic infection and/or incipient disease, to pulmonary or extra-pulmonary manifestations of disease, with a wide array of severity. Because the Mtb genome is considered comparatively stable, the diversity of clinical presentations has been historically attributed to host and environmental factors [3]. While many Mtb virulence factors are well described, to date, there are no proven genetic determinants associated with virulence, disease progression, severity, or epidemic dissemination [4].

Although Mtb diversity among isolates worldwide is low (typically differing by $< 0.01\%$ at the genetic level), this low diversity is not the result of a low mutation rates, as evidenced by the frequency of selection of drug resistance mutants that ranges from 10^{-5} to 10^{-8}. Indeed, the high frequency of resistance selection completely eliminates the possibility of mono- or bitherapies during the intensive phase of treatment [4,5]. Genetic changes (SNP variation) have been reported within different stocks of reference strains generated during subculture, for example strain H37Rv, with unknown consequences upon Mtb virulence [6]. Recently, differentially culturable tubercle bacteria were identified in sputa of TB patients, with improved recovery of the bacilli after culturing in liquid versus solid media [7], raising further questions regarding the influence of the growth media on the diversity of subcultured Mtb isolates.

The advent of the Next Generation Sequencing (NGS) era has enabled outstanding progress in understanding Mtb global diversity and the evolution of resistance, and Whole Genome Sequencing (WGS) of clinical isolates is becoming increasingly used both for research and clinical practice. Isolates collected are generally stored in various

* Corresponding author. Centre International de Recherche en Infectiologie, Université Lyon 1, ENS de Lyon, 21 Avenue Tony Garnier, 69365 Lyon cedex 07, France.
E-mail address: Charlotte.genestet@gmail.com (C. Genestet).
[1] JLB and OD contributed equally to this work.

https://doi.org/10.1016/j.tube.2019.05.001
Received 7 February 2019; Received in revised form 29 April 2019; Accepted 1 May 2019
1472-9792/

conditions (on slants at room temperature or as frozen stocks), then subcultured on various growth media and finally transferred to the NGS platforms either as live cultures, heat killed cultures, or as isolated DNA. All these pre-analytical steps may either promote changes in the genome, or influence the detection of variants, and may therefore impact the interpretation of WGS results.

Given the very slow turnaround time for Mtb subculturing, WGS obtained directly from the clinical sputum samples has great interest for clinical diagnostics, and to date has focused primarily on resistance mutation detection [8–10]. Few studies have deployed NGS to address whether variants were selected during subculture, or whether subculturing could significantly impact the observed population diversity of obtained isolates, or the relative frequency of minor variants detected. This is a question of utmost importance because, depending on the outcome, it may influence interpretations of sequence data from biobanked Mtb isolates and subsequent inferences regarding genetic determinants of TB physiopathology or Mtb evolution.

In this study, we evaluated the impact of different culture media commonly used in routine diagnosis (MGIT, 7H10 and Coletsos) on the selection of Mtb variants. Residual pellets from smear-positive sputum were analysed and compared by WGS on direct samples and after subculture on media. To study the whole genome and determine minor variants within Mtb populations, and not just the resistance genes, we performed NGS, allowing high coverage of subcultured samples, and sufficient coverage for some directly sequenced sputa; for cases in which coverage was limited, we performed complementary targeted sequencing of specific genome regions.

2. Material and methods

2.1. Sample selection and processing

The smear-positive direct respiratory samples with acid fast bacilli (AFB) scorings of +1 to +4 [11] used in this study were collected from patients with confirmed *M. tuberculosis* infections at the Lyon University Hospital. Residual sputa were collected after completion of the routine diagnostic work flow. All smear-positive samples were digested and decontaminated as previously described by Kent and Kubica [11]. Subsequently 200 µL were inoculated in MGIT PANTA (Polymyxin B, Amphotericin B, Nalidixic acid, Trimethoprim, and Azlocillin) supplemented (Becton Dickinson, Pont-de-Claix, France), 7H10 plate PANTA supplemented and Coletsos (Bio-Rad, Marnes-la-Coquette, France) media. Remaining samples were treated with MolYsis Basic5 kit (Molzym, Germany) to remove human DNA, according to manufacturer instructions, before DNA extraction and Illumina sequencing (Fig. 1 Workflow). In this way paired WGS data were generated from each unique original clinical sputum specimen.

2.2. DNA extraction, targeted PCR and sequencing

Genomic DNA was purified from cleared lysate using a QIAamp DNA mini Kit (Qiagen, Courtaboeuf, France). For targeted PCR, we used Platinum SuperFi DNA polymerase (Life Technologies SAS, Courtaboeuf, France) and primers indicated in Table S1. DNA libraries were prepared with Nextera XT kit (Illumina, San Diego, USA). Samples were sequenced on NextSeq system (Illumina) at the GATC platform (GATC Biotech SARL, Germany), to produce 150 base-pair paired-end reads, obtaining between 2 and 6.9 million read pairs for sequencing on direct samples and between 3.3 and 12 million read pairs for sequencing on subcultured samples.

2.3. Bioinformatic analysis of Illumina data

Reads were mapped with BOWTIE2 to the H37Rv reference genome (Genbank NC000962.2) and variant calling was made with SAMtools mpileup, as previously described [12]. A valid nucleotide variant was called if the position was covered by a depth of at least 10 reads and supported by a minimum threshold rate of 10%. Regions with repetitive or similar sequences were excluded, i.e. PE, PPE, PKS, PPS, ESX. The 16S and 23S regions were also excluded due to possible contamination by oropharyngeal flora on direct samples. On direct samples, the WGS reference genome coverage ranged from 0.03% to 98%, with an average depth of coverage of 0x to 58x; on subculture samples, genome coverage ranged from 98.5% to 99.5%, with an average depth of 178x to 777x. Quality metrics are shown for each sample in Tables S2 and S3.

After pairwise SNP comparison, differential variants were further examed using an additional script that called all nucleotide frequencies at the positions of interest, even for the genome regions not meeting the quality criteria used in previous analyses.

2.4. Nucleotide sequence accession number

All sequence datasets have been deposited in the European Nucleotide Archive under project accession no. PRJEB31112.

2.5. Ethics

This study was in accordance with the ethics committee of the Lyon University hospital, France (declared sample collection: DC-2011-1306).

3. Results

3.1. Evaluation of Illumina sequencing output on sputa and subcultured samples

Paired sputum and subculture samples were sequenced from 14 smear-positive sputa. Fig. 2A shows the average depth of coverage (DoC) across the genome for directly sequenced sputa samples, and the Mtb genome reference coverage. We stratified sputa into 4 groups according to smear positivity: 2 with very high (4 + smear); 8 with high (3 + smear), 2 with medium (2 + smear), and 2 with low (1 + smear) bacterial loads. For 4 of 14 sputa, we obtained sufficient average DoC (≥30x) and reference genome coverage (≥96%) to analyse majority variants (≥50%) within the full genome. For 2 out of 14 sputa, the DoC of 11x and partial reference genome coverage (≥57%) allowed partial genome study of majority variants. Reference genome coverage critically depends on the depth of coverage (Fig. 2A) and a minimum average DoC of 30x is necessary to study whole genomes. We found no strict correlation between smear positivity results and Illumina sequencing output. Low Mtb coverage was reflected in the high number of unmapped reads. We speculate that despite our use of protocols for sputum decontamination and human DNA removal, there may have been residual oropharyngeal flora or human DNA that contributed to the number of unmapped reads [9,10] (Fig. 2B), or that other criteria may have influenced the output data generated directly from sputum samples [13].

Regarding Illumina sequencing output after subculture of samples on MGIT, 7H10 and Coletsos media, we obtained excellent DoC (median from 314x to 316x, Fig. 2C) and reference genome coverage (≥99%, Fig. 2D). For subcultured samples, we observed low numbers of unmapped reads on the Mtb reference genome (median from 1.5% to 1.7%, Fig. 2E). However, 4 out of 14 samples grown on Coletsos medium were not sequenced due to contamination by staphylococci. According to these criteria, we obtained no significant differences in Illumina sequencing output regardless of the subculture media used.

3.2. Concordance of results from sputa and subcultured samples

Firstly, using the data on majority variants (≥50%), we compared pairwise SNP distance between WGS on direct versus subcultured samples, to assess whether the subculture steps induced Mtb genome

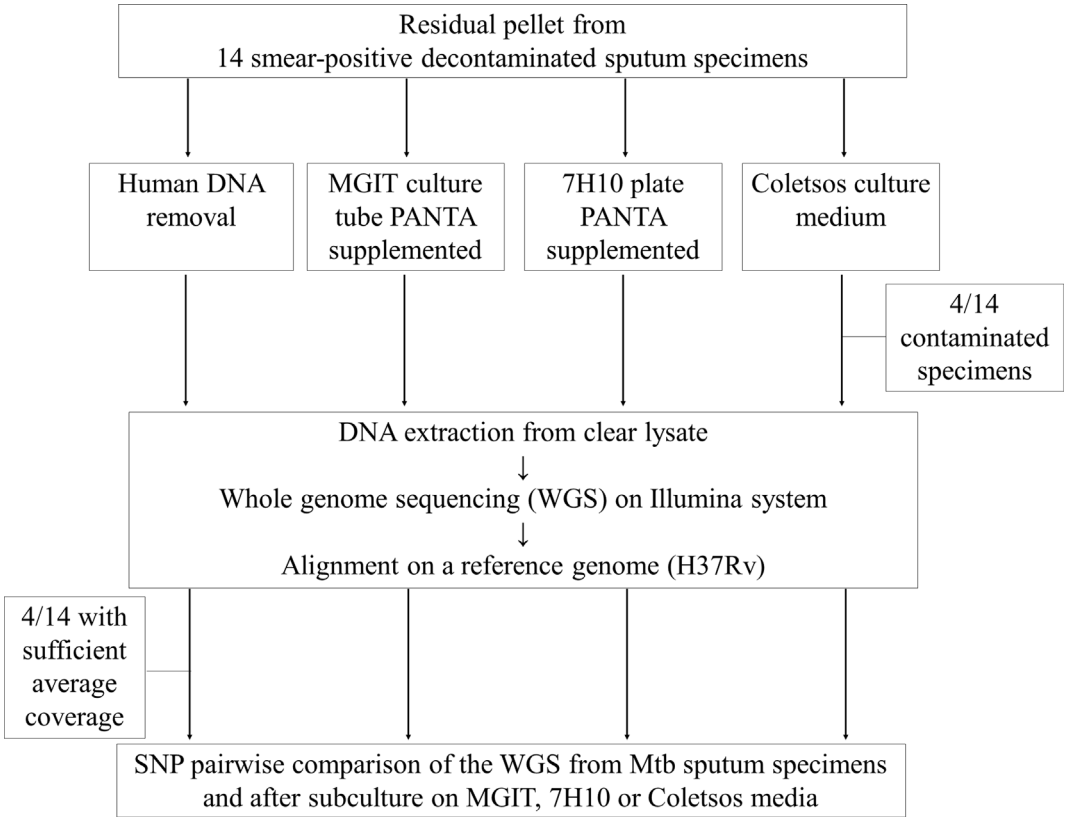

Fig. 1. Flow diagram for preparation of Mtb samples from receipt of smear-positive decontaminated sputum specimens to final data report.

variations. A full comparison of 4 sputa (S1, S2, S4 and S5), and partial comparison of 2 sputa (S8 and S9), and their respective subcultured samples showed no variation between pairs. We also analysed variant frequencies between each individual pair and did not observe any significant differences (Fig. 3).

Secondly, we compared pairwise SNP distances between WGS obtained from cultures grown on MGIT, 7H10 and Coletsos media, to assess whether particular culture media favored Mtb genome variations. WGS generated by subcultures of the same clinical sample on each one of the three types of growth media were compared pairwise (i.e. MGIT versus 7H10, MGIT versus Coltesos and 7H10 versus Coletsos). Given that 4 Coletsos subcultures contaminated by staphylococci were not analysed, a total of 34 pairwise comparisons were conducted. Pairwise comparisons showed no differences in variant frequencies regardless of the media used (Fig. 3).

To further examine the impact of subculture on minor variants that may not have been detected within the WGS datasets, we performed additional NGS sequencing on specific PCR products, targeting the genomic regions where variants had emerged after subculture. In samples 7, 13 and 14 we did not observe unfixed mutations in our settings under any experimental conditions. In samples 1, 2, 3, 4, 5, 8, 9, 10 and 11 we observed majority variants (ranging from 75% to 96%), while minority variants (ranging from 4% to 18%) were observed only in samples 3, 6 and 12. No significant differences were observed between matched pairs of direct and subcultured samples, regardless the culture media used (Fig. 3).

In summary, this study revealed no clear trend of Mtb genome modifications or selection of minor variants after subculture on MGIT, 7H10 or Coletsos media in our setting (Fig. 3B, C and D).

4. Discussion

Technological advances in routine whole genome sequencing have transformed biomedical research in the last decade. Analysis of Mtb WGS plays an increasingly important role in diagnosis and prediction of drug susceptibility, in outbreak investigations, as well as supporting investigations of host-pathogen interactions, pathogenesis, and global epidemiology [14]. The vast majority of all clinical strains studied by WGS have been isolates biobanked after subculture steps during routine diagnosis. Genetic variation as a result of culturing has been observed in other fastidious bacterial species, such as *Campylobacter jejuni* and *Helicobacter pylori*, often leading to phenotypic changes in virulence [6]. Serial culturing events are known to induce genetic changes in the Mtb genome, however the selective effect of freezing and thawing cycles on cultured populations is inconclusive, and the impact of different subculture media on genome-wide diversity and variant selection of Mtb remains poorly understood. Some studies have investigated concordance between genotypes obtained from matched pairs of sputa and culture, but have either focused on a single culture media or on partial sequencing of particular genome regions [8–10]. Thus, Brown et al. compared directly sequenced sputum and 7H11 subcultures, and found that matched pairs were not significantly different. Moreover, they identified one case with heterogeneity in the gyrA locus, and observed the same proportion of variants in sequences from both sputum and subcultured samples. Doyle and colleagues compared sequences from sputum and MGIT subcultures, and reported variants above a frequency of 50% [9]. In that setting, false SNP differences were observed due to high threshold. Similarly, Votintseva et al. showed no clear trend of higher diversity in directly sequences sputum versus MGIT subcultures [10]. There is some evidence that culture steps may modify the clonal composition of Mtb in mixed artificially spiked sputum samples [15]. However, these results may have been biased by the possible competition occurring among co-culture of different isolates that would not naturally infect the same host [15]. Here we evaluated the impact of MGIT, 7H10 and Coletsos media on the selection of Mtb variants by comparing WGS from directly sequenced sputa to WGS from subcultures produced in 3 different growth media, considering genome-wide diversity and minority variants.

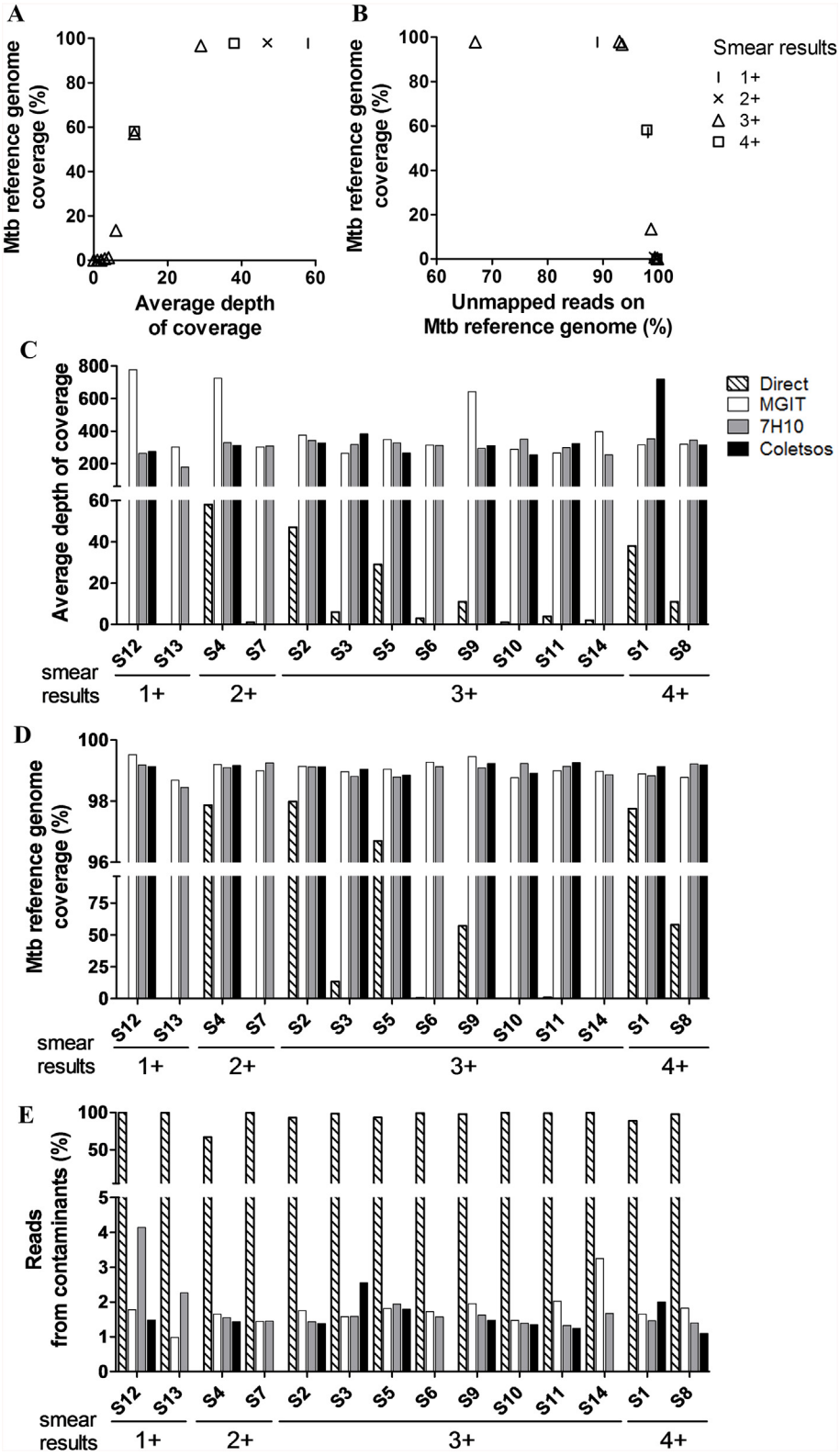

Fig. 2. Illumina sequencing output on sputa and subculture samples. (A) Average depth of coverage versus proportion of the Mtb reference genome coverage obtained for smear-positive samples from sputum, according to the smear results (vertical line, 1 + smear; cross, 2 + smear; triangle, 3 + smear, square, 4 + smear). (B) Proportion of unmapped reads (supposed DNA from contaminant) versus proportion of genome coverage obtained for smear-positive samples from sputum, according to the smear results (cross, 1 + smear result; triangle, 2 + smear result; square, 3 + smear result, circle, 4 + smear result). (C) Average depth of coverage, (D) Mtb genome coverage and (E) proportion of unmapped reads (supposed DNA from contaminant) on direct samples or after subculture on MGIT, 7H10 or Coletsos media, according to sputum smear results.

Due to high abundance of unmapped reads, only 4 out of the 14 sputum samples yielded WGS data that could be exploited for analysis, and 2 sputa yielded sufficient data for partial genome comparisons of majority variants. We did not observe significant differences in variant selection nor in variant frequencies within the paired samples, regardless of the media used for subculture. To further analyse low frequency variants in sputum samples, we performed sequencing on specific PCR amplicons, that targeted the genome regions that harboured variation observed after subculture. This additional analysis did not reveal further Mtb genome modifications or evidence of selection after subculture on MGIT, 7H10 or Coletsos media. Our results were consistent with previous reports, showing low variations in variant frequencies between matched pairs [8,9]. However, due to low average DoC of sputum samples, these experiments did not allow any detection

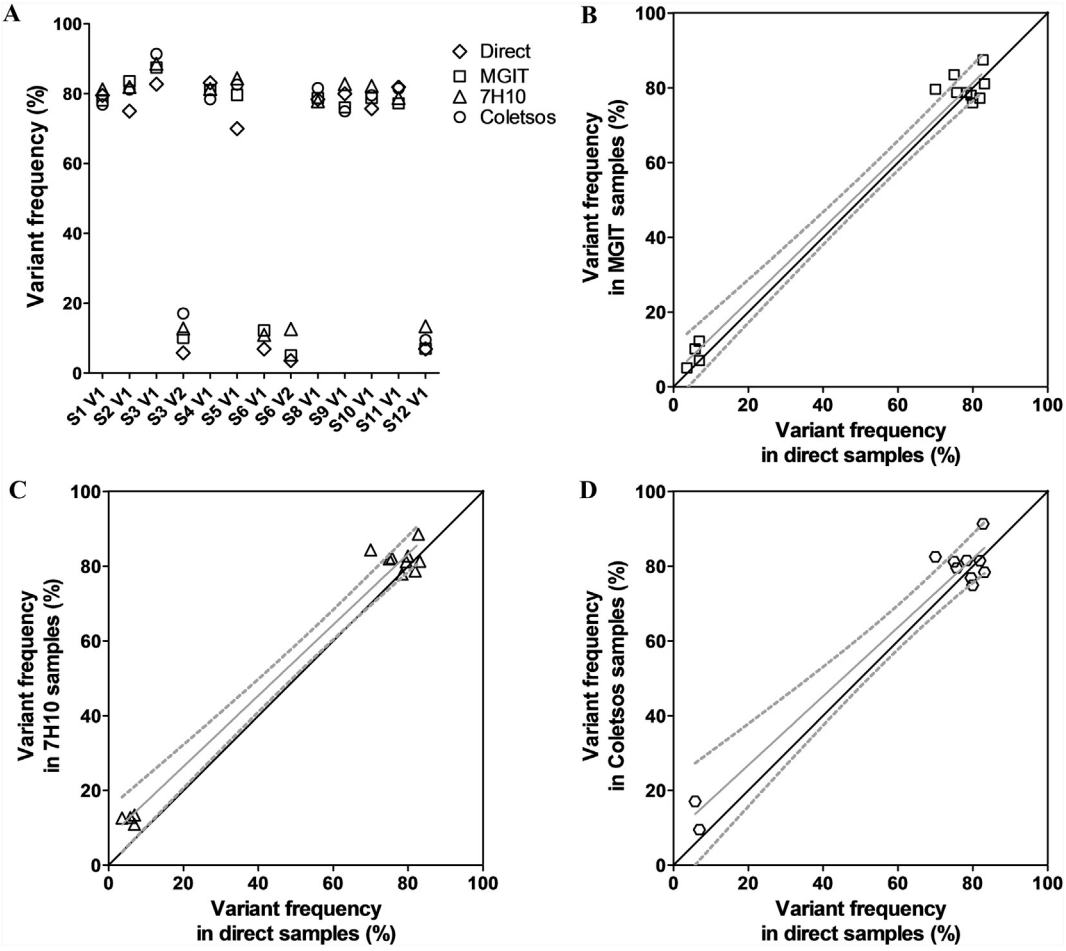

Fig. 3. Frequencies of variants observed on direct and subculture samples. (A) Frequencies of variants (V1 and V2) observed after sequencing of specific PCR products on direct samples (S3, S6 S8, S9, S10, S11, S12; diamond) or after WGS on direct samples (diamond) and after subculture on MGIT (square), 7H10 (triangle) or Coletsos (circle). (B–D) Correlation of variant frequencies between matched pairs of direct (x axis) and subculture samples (y axis) on MGIT (B), 7H10 (C) or Coletsos (D). Solid grey lines indicate linear regression curves and dashed lines the 95 confidence bands.

of non-cultivable minority variants potentially present in the sputa.

In conclusion, we report the absence of evidence for selective pressure on minority variant populations following subculture in the 3 most commonly used media for Mtb culture. These findings indicate that biobanked clinical strains, that have not undergone multiple culture passages or freezing and thawing events, are likely to be representative of Mtb populations present in original patient sputa, and are therefore relevant for use in pinpointing putative genetic determinants of Mtb pathogenesis and epidemiological features. Our study also revealed the frequent presence of minor variants within the diversity of Mtb clinical isolates. Though the relevance of minor variants has been proved for loci linked to TB-drug resistance [16], future studies are needed to better understand the involvement of Mtb minor variants in TB pathogenesis and transmission.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by the LABEX ECOFECT (ANR-11-LABX-0048) of Université de Lyon, within the program “Investissements d'Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tube.2019.05.001.

References

[1] WHO G. World health organization Global Tuberculosis Report 2018
[2] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic co-expansion of Mycobacterium tuberculosis with modern humans. Nat Genet 2013;45:1176–82. https://doi.org/10.1038/ng.2744.
[3] Gagneux S, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 2007;7:328–37. https://doi.org/10.1016/S1473-3099(07)70108-1.
[4] Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol 2018;16:202–13. https://doi.org/10.1038/nrmicro.2018.8.
[5] Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug resistant tuberculosis. Nat Genet 2013;45:784–90. https://doi.org/10.1038/ng.2656.
[6] Ioerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, Fortune S, et al. Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. J Bacteriol 2010;192:3645–53. https://doi.org/10.1128/JB.00166-10.
[7] Dhillon J, Fourie PB, Mitchison DA. Persister populations of Mycobacterium tuberculosis in sputum that grow in liquid but not on solid culture media. J Antimicrob Chemother 2014;69:437–40. https://doi.org/10.1093/jac/dkt357.
[8] Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZM, et al. Rapid whole-genome sequencing of Mycobacterium tuberculosis isolates directly

from clinical samples. J Clin Microbiol 2015;53:2230–7. https://doi.org/10.1128/JCM.00486-15.

[9] Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, et al. Direct whole-genome sequencing of sputum accurately identifies drug-resistant *Mycobacterium tuberculosis* faster than MGIT culture sequencing. J Clin Microbiol 2018;56. https://doi.org/10.1128/JCM.00666-18.

[10] Votintseva AA, Bradley P, Pankhurst L, del Ojo Elias C, Loose M, Nilgiriwala K, et al. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol 2017;55:1285–98. https://doi.org/10.1128/JCM.02483-16.

[11] Handbook on tuberculosis laboratory diagnostic methods in the European Union - updated 2018. European Centre for Disease Prevention and Control; 2018http://ecdc.europa.eu/en/publications-data/handbook-tuberculosis-laboratory-diagnostic-methods-european-union-updated-2018, Accessed date: 14 November 2018.

[12] Genestet C, Tatai C, Berland J-L, Claude J-B, Westeel E, Hodille E, et al. Prospective whole-genome sequencing in tuberculosis outbreak investigation, France, 2017–2018. Emerg Infect Dis 2019:25. https://doi.org/10.3201/eid2503.181124.

[13] Murray PR, Washington JA. Microscopic and bacteriologic analysis of expectorated sputum. Mayo Clin Proc 1975;50:339–44.

[14] Cabibbe AM, Walker TM, Niemann S, Cirillo DM. Whole genome sequencing of Mycobacterium tuberculosis. Eur Respir J 2018;52:1801163. https://doi.org/10.1183/13993003.01163-2018.

[15] Martín A, Herranz M, Ruiz Serrano MJ, Bouza E, García de Viedma D. The clonal composition of Mycobacterium tuberculosis in clinical specimens could be modified by culture. Tuberculosis 2010;90:201–7. https://doi.org/10.1016/j.tube.2010.03.012.

[16] Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, et al. Detection of low-level mixed-population drug resistance in Mycobacterium tuberculosis using high fidelity amplicon sequencing. PLoS One 2015;10:e0126626https://doi.org/10.1371/journal.pone.0126626.