Faculty of Pharmaceutical, Biomedical and Veterinary Sciences

Department of Biomedical Sciences

Mycobacterium tuberculosis complex strain diversity may impact disease presentation, diagnosis and outcome

Hoe verscheidenheid in *Mycobacterium tuberculosis* complex stammen de diagnose, het ziektebeeld en uitkomst van behandeling kan beïnvloeden

Thesis for the degree of

Doctor (PhD) in Biomedical Sciences at the University of Antwerp

by **Chakirath N'Dira SANOUSSI**

Promoters:
Prof. Dr. Leen Rigouts
Prof. Dr. Bouke C. de Jong

Co-promoters:
Prof. Dr. Dissou Affolabi
Dr. Conor Meehan

Antwerp, 2019

Front cover image: shutterstock.com, C. N'Dira Sanoussi (marking on the map for distribution of the lineages of the *Mycobacterium tuberculosis* complex)
Cover design: Anita Muys, Nieuwe Media Dienst, University of Antwerp, Belgium and C. N'Dira Sanoussi
Printing company: PROVO, Gierle, Belgium

Jury

Chairperson:

Prof. Dr. Franck Kooy	University of Antwerp (Belgium)

Promoters:

Prof. Dr. Leen Rigouts	University of Antwerp (Belgium), Institute of Tropical Medicine, Antwerp (Belgium)
Prof. Dr. Bouke C. de Jong	Institute of Tropical Medicine, Antwerp (Belgium)

Co-Promoters:

Prof. Dr. Dissou Affolabi	Laboratoire de Référence des Mycobactéries, Cotonou, National Tuberculosis Program (Benin), University of Abomey-Calavi, Benin
Dr. Conor Meehan	Institute of Tropical Medicine, Antwerp (Belgium), University of Bradford (United Kingdom)

Members:

Prof. Dr. Guy Caljon	University of Antwerp (Belgium)
Prof. Dr Annelies Van Rie	University of Antwerp (Belgium)
Dr Richard Anthony	National Institute for Public Health and the Environment (RIVM), The Netherlands
Dr Emmanuel André	Catholic University of Leuven, KU Leuven (Belgium)
Dr Michel Kaswa	Institut National de Recherche Biomédicale, National Tuberculosis Program, Kinshasa (DR Congo)

Contents

Abbreviations

BCG	Bacilli Calmette Guérin
BeniDiT	BeniDiT study: Benin population Diversity of TB strains and Implications
CPC	Cetylpyridinum chloride
CRISPR	Clustered Regulartory Short Palindromic Repeats
DRS	Drug resistance survey
DST	Drug susceptibility testing
E	Ethambutol, see also EMB
EMB	Ethambutol, see also E
ETOH	Ethanol
H	Isoniazid, see also INH
HIV	Human immune-deficiency virus
IC	Confidence interval
INH	Isoniazid, see also H
ITM	Institute of Tropical Medicine, Antwerp, Belgium
L	Lineage
L1	Lineage 1
L2	Lineage 2
L3	Lineage 3
L4	Lineage 4
L5	Lineage 5 (*M. africanum* West African 1)
L6	Lineage 6 (*M. africanum* West African 2)
L7	Lineage 7
L8	Lineage 8
L0	Lineage 0

L J	Lowenstein Jensen medium
LPA	Line probe assay
LRM	Laboratoire de Référence des Mycobactéries, Cotonou, Bénin
MAF	*Mycobacterium africanum*
MDR	Multi-drug-resistant
MGIT	Mycobacterial Growth Indicator Tube
MIRU-VNTR	Mycobacterial Interspersed Repetitive Unit, Variable Number of Tandem Repeats
MTBC	*Mycobacterium tuberculosis* complex
NALC	N-Acetyl-L-Cysteine
nSNP	Non-synonymous SNP
NTM	Non-tuberculous mycobacteria
OMNI	OMNIgene.SPUTUM
OR	Odds ratio
PacBio	Pacific Biosciences (long read sequencing)
PcbL5Ben	PacBio L5 Benin (complete genome of an L5 isolate from Benin)
PcbL5Gbia	PacBio L5 Gambia (complete genome of an L5 isolate from The Gambia)
PcbL5Nig	PacBio L5 Nigeria (complete genome of an L5 isolate from Nigeria)
PCR	Polymerase chain reaction
pDST	Phenotypic DST (drug-susceptibility testing)
R	Rifampicin, see also RIF
RD	Region of difference
RFLP	Restriction length fragment polymorphism

RIF	Rifampicin, see also R
RR	Rifampicin-resistant
S	Streptomycin, see also STR
S-CPC	Mixtures of a sample with cetylpyridinium chloride
SNP	Single nucleotide polymorphism
STR	Streptomycin, see also S
WHO	World Health Organization

Summary

Tuberculosis (TB) is caused by the bacteria of the *Mycobacterium tuberculosis* complex (MTBC), which comprises 7 human-adapted phylogenetic lineages, including two *M. africanum* lineages that are geographically restricted to West and Central Africa. Lineage 5 (*M. africanum* West-African 1) and Lineage 6 (*M. africanum* West-African 2) together cause up to 40% of TB in West-Africa. While L6 is relatively well studied with regard to host factors, disease presentation and other factors, little is known about L5. Furthermore, there is a suggestion that in some West-African countries the prevalence of L5 and L6 was decreasing. Benin is the country with the highest L5 prevalence worldwide. The sole molecular epidemiology study of TB in Benin was conducted more than 10 years ago, in new patients from one town (Cotonou). This thesis aimed to increase understanding of *M. tuberculosis* West African 1 (Lineage 5) epidemiology and genomic characteristics. We identified numerous novel associations that shed new light on this distinct *M. tuberculosis* complex member. In addition, in the process of conducting the cohort study, we identified technical advances for improved unbiased diagnosis of TB and related molecular epidemiological studies.

We found that L5 is under-represented in positive cultures. This reduced growth of L5 strains could partly be explained by the absence - in the vast majority of L5 strains - of genes implicated in bacterial survival and *in vitro* growth (mainly *Rv1994c*). Besides the identified culture bias, we documented decreased performance of the rapid MPT64-antigen-based lateral flow assay for the identification of L5 as MTBC member in positive cultures, likely due to an L5-wide a non-synonymous SNP (I43N) in the *mpt64* gene.

To enable such large scale multicentric studies/surveys, sputum storage reagents such as cetylpyridinium chloride, OMNIgene.SPUTUM, ethanol (molecular tests only) were compared for up to 28-day storage of sputum in ambient temperature for subsequent mycobacterial culture or molecular TB diagnostic testing. Culture positivity after 8-days storage at ambient temperature of sputum in cetylpyridinium chloride (CPC) or OMNIgene.SPUTUM (OMNI) was comparable to that of fresh sputum. However, after a 28-day storage, the culture positivity significantly decreased in OMNI stored-sputa compared to those stored in CPC. On the other hand, 28-day storage of sputum in CPC, OMNI or ethanol at ambient temperature did not impact short-fragment PCR (GeneXpert®MTB/RIF), including for samples with low smear-positivity grades. However, for long-fragment PCR, ethanol yielded a slightly lower PCR positivity for low smear grades, while the performance of OMNI and CPC was excellent for all smear-positivity grades. We provided an algorithm to help the user in the choice of the storage reagent depending on the samples' bacillary burden, type of molecular test and type of culture combined with molecular test envisaged.

Hence, the algorithm for TB diagnostics testing in *M. africanum*-(especially L5) endemic countries could be improved, favoring direct diagnostics for unbiased results, and expanded identification to ensure identification of all MTBC strains. A general lesson that can be drawn from work in this thesis and by others, is that the performance of diagnostic tests should be validated in a broad variety of settings – including those with geographically limited lineages - ideally before implementation for patient diagnosis/care, or at least as extended post-marketing validation.

In this thesis, the nationwide genetic diversity of the MTBC in Benin was first determined retrospectively using stored culture isolates from previously-treated TB patients. To overcome the culture bias observed, we used direct spoligotyping (followed by "direct PhyloSNP" for uncommon patterns) for lineage determination in the prospective nationwide population structure study of MTBC in new and previously-treated patients in Benin.

In Cotonou, over 10 years (prevalence in 2005-2006 compared to that found in this study), among new patients the L5 prevalence significantly declined by 9.4 % (95% CI: -17.6 to -1.2) as did L1, while the L4 prevalence increased by 16% (95% CI: 7.4 to 24.6), and the L6 prevalence remained similar. In Benin, the L5 and L6 geographical distribution are not driven by the same factors. L5 distribution seems to be explained by decreasing population density and Gbe ethnicity (especially Eastern-Gbe). Conversely L6 distribution is explained by increasing population density and ethnicity (Peulh (Fulani) and Bariba ethnicities), but not with cattle contact. Molecular epidemiology studies investigating the association of lineages with human population density, ethnicity and occupation should be conducted in other *M. africanum* (especially L6) endemic countries.

The suggested lower virulence (immunogenicity, transmissibility) of *M. africanum* may be partly explained in L5 by the absence of genes associated with bacterial survival in macrophages (*Rv1978c)* and during the chronic phase of infection *(Rv1994c),* and genes associated with immune-evasion and virulence (*Rv2074*, vitamin B6: pyridoxine).

The nationwide distribution of lineages differed significantly by patient's treatment history, with *M. africanum* present in 39.2% of new and 26.3% in previously-treated patients (31.1% and 21% respectively for L5 alone). While L5

was less likely to cause TB relapse, our data suggests that it is possibly associated with the acquisition of rifampicin resistance, yet this needs to be further investigated. Furthermore, genomic predictions showed that L5 is possibly also associated with resistance to clofazimine and bedaquiline, important drugs to treat rifampicin-/multidrug-resistant TB. Measures should be implemented to avoid this potential threat to TB treatment in L5-endemic countries.

General molecular epidemiology studies requiring lineage determination could be based on direct spoligotyping, followed by “direct PhyloSNP” analysis for uncommon patterns. Nevertheless, for more advanced molecular epidemiological studies, as well as for future individualized diagnostics, effort should be increased to advance the quality of “direct whole genome sequencing”.

Comparative genomics identified differences in gene content not only between L5 and the currently used *M. tuberculosis* reference genome H37Rv (L4), but also among L5 strains, with up to 32 genes absent in a Nigerian L5 complete genome compared to a Benin and Gambian one. The use of an L5-specific reference genome may help in improved sub-lineage classification and the determination of L5 sub-lineage distribution across West African L5 countries to better understand he L5 transmission, phylogeny and origin.

In conclusion, direct genotyping should be used for MTBC population structure studies, and the algorithm for TB diagnostics testing in *M. africanum*- endemic countries should be improved, favoring direct diagnostics for unbiased results. The high within-lineage gene content variability suggests the pangenome of MTBC may be larger than previously thought, implying a reference-free

genome de novo assembly approach may be preferable over the currently used reference genome H37Rv for genome analyses.

Samenvatting

Tuberculose (tbc) wordt veroorzaakt door de bacteriën van het *Mycobacterium tuberculosis*-complex (MTBC), dat bestaat uit 7 - aan de mens aangepaste- fylogenetische afstammingslijnen, waaronder twee *M. africanum*-lijnen die geografisch beperkt zijn tot West- en Centraal-Afrika. Lijn 5 (L5; *M. africanum* West-African 1) en Lijn 6 (L6; *M. africanum* West-African 2) veroorzaken samen tot 40% van de TB in West-Afrika. Hoewel L6 relatief goed bestudeerd is met betrekking tot gastheerfactoren, ziektepresentatie en andere factoren, is er weinig bekend over L5. Verder wordt gesuggereerd dat in sommige West-Afrikaanse landen de prevalentie van L5 en L6 afnam. Benin is het land met de hoogste L5-prevalentie wereldwijd. De enige studie naar moleculaire epidemiologie van tbc in Benin werd meer dan 10 jaar geleden uitgevoerd, bij nieuwe patiënten afkomstig uit één stad (Cotonou). Het doel van dit proefschrift was om de kennis over *M. africanum* L5 genoomkenmerken en zijn epidemiologie in Benin uit te breiden. Bovendien, leidde de cohortstudie tot het voorstellen van een verbeterde representatieve diagnose van tbc (door technische vooruitgang) en verwante moleculair epidemiologische studies.

We ontdekten dat L5 ondervertegenwoordigd was in positieve kweken. Deze verminderde groei van L5-stammen kon gedeeltelijk worden verklaard door de afwezigheid - in de overgrote meerderheid van de L5 -stammen - van genen die betrokken zijn bij bacteriële overleving en in vitro groei (voornamelijk Rv1994c). Naast de geïdentificeerde kweekbias, documenteerden we ook een verminderde prestaties van de snelle laterale flow-assay (die gebaseerd is op het MPT64-antigeen) voor de identificatie van L5, als lid van het MBTC in positieve kweken, wellicht het gevolg van een L5-brede niet synonieme SNP (I43N) in het *mpt64*-gen.

Om dergelijke grootschalige multicenter studies en surveys mogelijk te maken, vergeleken we de bewaring van sputum tot 28 dagen bij kamertemperatuur in verschillende transportmedia zoals cetylpyridinium chloride, OMNIgene.SPUTUM en ethanol, gevolgd door kweek of moleculair diagnostische testen. Sputa die 8 dagen lang bij omgevingstemperatuur waren bewaard in ofwel cetylpyridiniumchloride (CPC) of in OMNIgene.SPUTUM (OMNI) gaven een vergelijkbare kweekpositiviteit met die van vers sputum. Na een bewaring van 28 dagen nam de kweekpositiviteit echter significant af in OMNI-opgeslagen sputa in vergelijking met deze bewaard in CPC. Anderzijds had 28-dagen opslag van sputum in CPC, OMNI of ethanol bij kamertemperatuur geen invloed op de positiviteit van een PCR met een kort amplificatiefragment (GeneXpert®MTB/RIF), inclusief voor monsters met een zwak-positief microscopie resultaat. Voor de PCR met een lang amplificatiefragment leverde ethanol echter een iets lagere PCR-positiviteit op bij microscopisch zwak-positieve uitstrijkjes, terwijl de prestaties van OMNI en CPC uitstekend waren ongeacht de positiviteitsgraad in microscopie. We hebben een algoritme voorgesteld om de gebruiker te helpen bij de keuze van het transportmedium, afhankelijk van de bacteriële lading van het monster en het beoogde type van moleculaire test al dan niet in combinatie met een bepaalde kweekmethode.

In het algemeen kan het algoritme voor TB-diagnosetesten in *M. africanum*-endemische landen (met name L5) verbeterd worden, met de voorkeur voor niet-selectieve directe diagnostiek en een uitgebreide identificatie zodat alle MTBC-stammen geïdentificeerd kunnen worden. Uit dit proefschrift (en werk van anderen) kan de algemene les getrokken worden dat de werkzaamheid van diagnostische testen uitgebreid gevalideerd moet worden in een breed scala aan geografische regio's, inclusief deze met aanwezigheid van geografisch

beperkte MTBC lijnen. Dit moet idealiter gebeuren vóór implementatie voor diagnose/zorg van de patiënt, of tenminste als uitgebreide post-marketingvalidatie.

Eerst werd de genetische diversiteit van het MTBC in Benin op nationaal niveau retrospectief bepaald aan de hand van bewaarde isolaten van eerder behandelde tuberculosepatiënten.Om de kweekbias te elimineren, gebruikten we directe spoligotypering (met name zonder kweek en gevolgd door “direct PhyloSNP voor ongekende patronen) voor bepaling van de MTBC-lijnen in de prospectieve landelijke populatiestructuurstudie in Benin.

Een vergelijking van de 2005-2006 prevalentiegegevens met de huidige data, toonde aan dat in Cotonou de L5-prevalentie bij nieuwe patiënten gedurende 10 jaar significant afnam met 9,4% (95% CI: -17,6 tot -1,2), evenals L1, terwijl de L4-prevalentie met 16% toenam (95% BI: 7,4 tot 24,6) en de L6-prevalentie ongeveer gelijk bleef. In Benin wordt de geografische verspreiding van L5 en L6 niet door dezelfde factoren bepaald. L5-verspreiding lijkt te worden verklaard door de afnemende bevolkingsdichtheid en Gbe etniciteit (voornamelijk Oost-Gbe). Daarentegen, wordt het voorkomen van L6 verklaard door een toenemende bevolkingsdichtheid en etniciteit (Peulh (Fulani) en Bariba etniciteit),maar niet door contact met vee. Verdere moleculair epidemiologische studies rond de associatie van MTBC-lijnen met menselijke bevolkingsdichtheid, etniciteit en beroepsactiviteit zijn wenselijk in andere *M. africanum*-endemische landen, met name voor L6.

De veronderstelde lagere virulentie (immunogeniciteit, overdraagbaarheid) van *M. africanum* kan deels in L5 worden verklaard door de afwezigheid van genen die geassocieerd zijn met bacteriële overleving in macrofagen (Rv1978c)

of de chronische fase van infectie (Rv1994c), en genen geassocieerd met immuunsysteemontwijking en virulentie (Rv2074, vitamine B6: pyridoxine).

De verspreiding van MTBC-lijnen op nationaal niveau in Benin verschilde aanzienlijk naargelang de tbc voorgeschiedenis van de patiënt, waarbij *M. africanum* aanwezig was in 39,2% van de nieuwe en in 26,3% van de eerder behandelde patiënten (respectievelijk 31,1% en 21% voor L5). Hoewel L5 minder snel TB-terugval veroorzaakte, werd het mogelijk geassocieerd met de verwerving van rifampicine resistentie, maar dit vereist verder onderzoek. Bovendien toonden genomische voorspellingen aan dat L5 mogelijk ook geassocieerd is met resistentie tegen clofazimine en bedaquiline, wat belangrijke geneesmiddelen zijn voor de behandeling van rifampicine-/multidrug-resistente tbc. Er moeten maatregelen worden genomen om deze potentiële bedreiging voor tbc-behandeling in L5-endemische landen te voorkomen.

Algemene moleculair epidemiologische studies die typering van MTBC-lijnen vereisen, zouden kunnen uitgevoerd worden met directe spoligotypering, gevolgd door "directe PhyloSNP" voor identificatie van ongekende patronen. Desalniettemin, zullen er inspanningen geleverd moeten worden om de kwaliteit van de “directe sequentiebepaling van het gehele genoom” te verhogen, zodat meer geavanceerde epidemiologische studies en geïndividualiseerde diagnostiek kunnen plaatsvinden.

Een vergelijkende genoomstudie identificeerde verschillen in het aantal genen, niet alleen tussen L5 en het momenteel gebruikte *M. tuberculosis* referentiegenoom H37Rv (L4), maar ook binnen de L5-lijn, met 33 genen die afwezig waren in het volledige genoom van een Nigeriaans L5 stam vergeleken met een stam uit Benin en Gambia. Het gebruik van een L5-specifiek

referentiegenoom kan helpen bij een verbeterde classificatie binnen de L5-lijn, en de distributie ervan in West-Afrikaanse landen. Bovendien kan hiermee de L5-transmissie, fylogenie en oorsprong beter begrepen worden.

Ter besluit, raden we aan directe genotypering te gebruiken voor het bestuderen van de MTBC populatiestructuur, en het algorithme voor de bacteriologische diagnose van tbc in *M. africanum*-endemische landen te optimaliseren, ten gunste van directe testen om bias te vermijden.

De hoge variabiliteit tussen genen binnen een MTBC-lijn suggereert dat het pangenoom van MTBC groter is dan eerder werd gedacht. Deze bevinding suggereert dat meer accurate genoomanalyses bekomen kunnen worden met een referentievrije de novo-assemblage methode dan genoomanalyses met behulp van het momenteel gebruikte referentiegenoom H37Rv.

Chapter 1

General introduction

Tuberculosis: a major public health problem

Tuberculosis (TB) is an infectious disease which affects the lungs (pulmonary TB) or other organs (extra-pulmonary TB). Pulmonary TB is the commonest form and its principal symptom is cough, which makes it a highly transmissible disease. Worldwide, TB is among the top 10 causes of death, and the first cause of death by a single infectious agent [1]. TB continues to be a major public health problem in many countries [1,2]. In 2017, worldwide, 10 million people developed TB, among whom 1.6 million died [1]. These estimates are higher in resource-limited countries. More than half (64%) of the worldwide TB cases in 2017, were reported from Angola, China, the Democratic Republic of Congo, Ethiopia, India, Indonesia, Kenya, Mozambique, Myanmar, Nigeria, Papua New Guinea, South Africa, Thailand and Zimbabwe [1]. Africa and Asia were the top 2 continents with most cases of TB. Africa had the highest incidence of TB (237 per 100,000 population) which is almost 2-fold the worldwide incidence (133 per 100,000 populations)[1]. Likewise the highest TB mortality rate was reported from Africa: 2.3-fold the worldwide rate in HIV-negative patients (39 vs 17 per 100,000 populations) and 6-fold the worldwide rate in HIV-positive patients (24 vs 4 per 100,000 populations) [1]. Furthermore, drug-resistant TB is an even worse public health threat. Multi-drug resistant TB is defined as resistance to rifampicin (RIF) and isoniazid (INH) the most powerful first-line TB drugs. Rifampicin resistance (RR) is an indicator of multi-drug resistance (98% of RR-TB are MDR-TB) and treated as such. Globally, in 2017, 3.5% of new TB patients and 18% of previously treated patients were MDR/RR. India and China are the top countries with the largest numbers of MDR/RR-TB cases (135000 and 73000), with an incidence of respectively 10 and 5.2 cases per 100,000 populations. This makes Asia the continent with the highest number of MDR/RR-TB patients estimated to 192000 cases whereas the number of

MDR/RR-TB patients in Africa is 90000 cases [1]. Europe is the continent with the highest incidence for MDR/RR-TB (12 per 100,000 populations), followed by South-East Asia and Africa (respectively 9.7 and 8.6 per 100,000 populations). Extensively drug-resistant (XDR)TB is an even more severe form of MDR-TB with additional resistance to fluoroquinolone antibiotics and injectable agents which is also a greater public health threat than non-resistant TB[3]. In 2017 globally, 8.5% of MDR/RR-TB patients were estimated to have XDR-TB (6.2% in 2016 and 6.8% in 2015). In 2015, 10.6% of MDR/RR-TB patients were XDR in Europe (5480 cases), whereas respectively 6.2% and 3% of MDR/RR-TB patients were XDR in South-East Asia (2169 cases) and Africa (483 cases)[1].

Responsible agents and their variable geographical repartition

TB is caused by the bacteria of the *Mycobacterium tuberculosis* complex (MTBC, *M. tuberculosis sensu stricto*, *M. africanum*, *M. bovis*, *M. canettii, M. caprae, M. microti*, *M. pinnipedii*, *M. orygis*)[4]. The human-adapted species are *M. tuberculosis sensu stricto* and *M. africanum*; and consist of 7 phylogenetic lineages (L): L1 (Indo-oceanic), L2 (East-Asian, includes Beijing), L3 (East-African-Indian), L4 (Euro-American), L5 (*M. africanum* West-African 1), L6 (*M. africanum* West-African 2), and L7 (Ethiopia). Thus, *M. africanum* (MAF) consists of two lineages (L5 and L6) which together with L1 are classified as ancestral (or 'ancient') lineages [5]. L7 is classified as an intermediate lineage [6][7] and L2, L3, L4 as modern lineages (Table 1, Fig.1)[5]. As far as their geographic repartition is concerned, L2 and L4 are the most widespread while MAF (L5, L6) and L7 are the most geographically restricted: L5 and L6 are restricted to West-Africa, and L7 is found only in Ethiopia. L1 and L3 are intermediately distributed (Fig.1)[5]. The distribution of L5 and L6 is different

within West-Africa, with L5 mostly found in eastern West Africa while L6 is mostly found in western West Africa [8].

Table 1. MTBC lineages nomenclature

Human-adapted species of the MTBc	**Lineage**	**Abbreviation**	**Other name**	**Ancestral or modern**
Mycobacterium tuberculosis sensu stricto	Lineage 1	L1	Indo-Oceanic	Ancestral
	Lineage 2	L2	East-Asian (includes Beijing)	Modern
	Lineage 3	L3	East-African-Indian	Modern
	Lineage 4	L4	Euro-American	Modern
Mycobacterium africanum (MAF)	Lineage 5	L5 or MAF1 or MAFL5	*M. africanum* West African 1	Ancestral
	Lineage 6	L6 or MAF2 or MAFL6	*M. africanum* West African 2	Ancestral
M. tuberculosis sensu stricto	Lineage 7	L7	Ethiopia	Intermediate

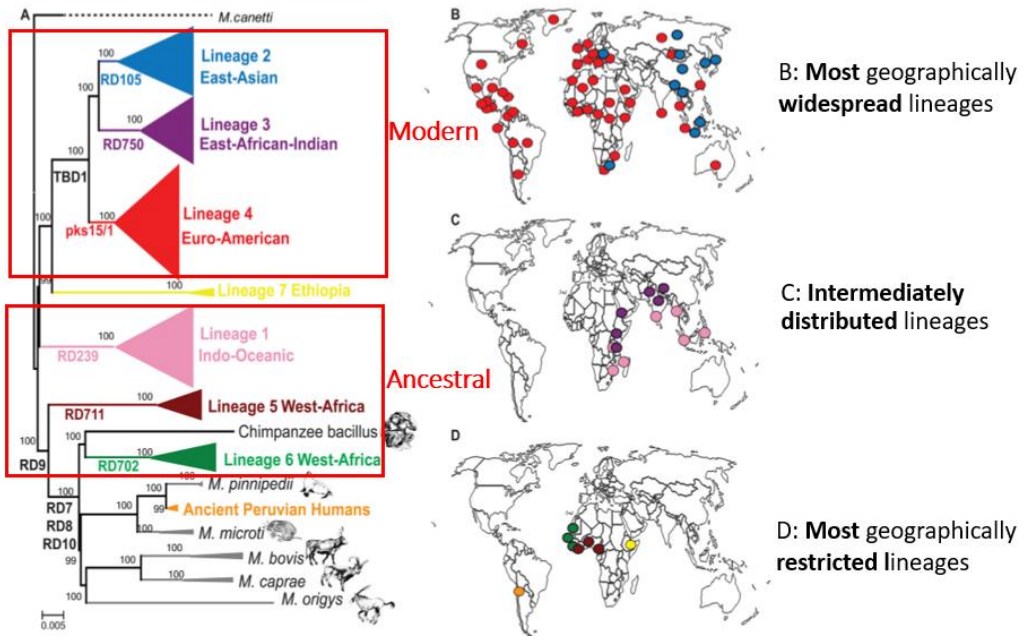

Figure 1. Geographical repartition of the MTBC lineages (adapted from Coscolla et al, 2014[5])

Possible hypotheses for the restricted geographical repartition of *M. africanum*

MAF lineages are responsible for up to 40% of TB in West-Africa [9]. For a long time, it was believed that the MTBC was introduced into Africa by European contact (colonization). But in 2013, it has been shown that, like *homo sapiens*, MTBC originates from Africa and has been coevolving with humans since before humans left the continent 70 thousand years ago [10–13], starting from ancestral lineages (73 to 67 thousand years ago), through the intermediate L7 (64 thousands years ago) to modern lineages (46 to 42 thousand years ago) (Fig.2) [12]. However, a study in 2014, found that TB was brought from Africa to America 6000 years ago, through sea by sea mammals (seals and sea lions),

before European contact [14], contradicting the 70 thousands years ago found in the 2013 study [12].

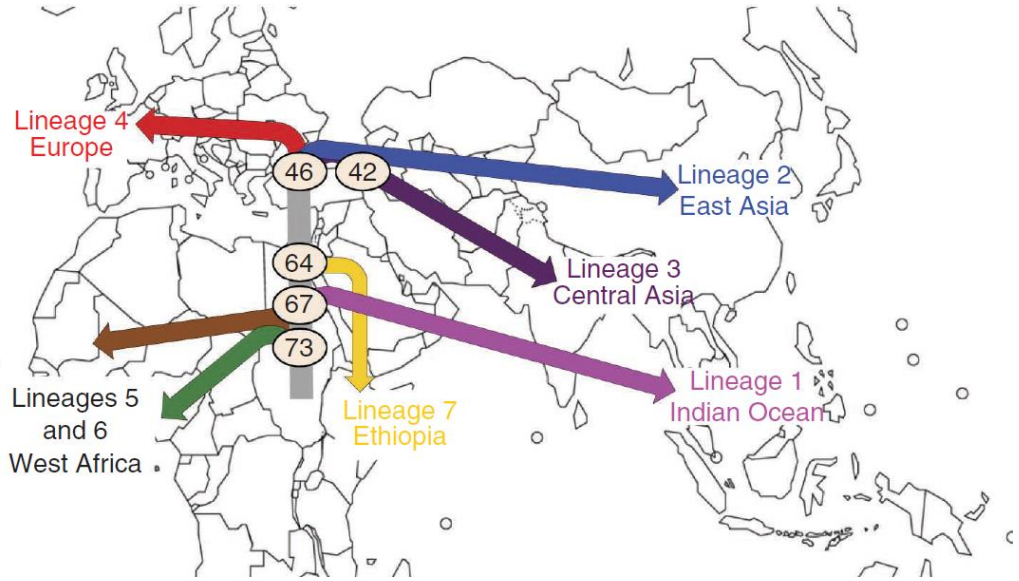

Figure 2. Evolution (spatio-temporal) of MTBC lineages (Adapted from Comas et al, Nat Gen, 2013 [12])

The numbers (73, 67, 64, 46, 42) represent the estimated time (in thousands years ago) when a specific MTBC lineage was found in the specific region on the map. The arrows show how lineages evolved from their region of origin.

With the dating being the contradicting issue, the two studies supported that TB originated from Africa. While L1, L2, L3 and L4 are relatively widespread, the reasons why L5, L6 are restricted to West and Central Africa and L7 to Ethiopa [5] remain unknown. As for other infectious diseases, one of the reasons could be that the reservoir for L5 and L6 is an animal only found in West-Africa. However, there is a suggestion that MAF lineages, like other human-adapted lineages, are obligate human pathogens, and no environmental or substantial animal reservoir has been identified to date [5,12]. In contrast to that

suggestion, L5 and L6 are used successfully to infect animal models for experimental purposes, and are also sporadically isolated from animals, suggesting the possibility of an animal reservoir [15]. Furthermore, a review recently reported that MTBC evolved from Africa, from an environmental organism to an obligate pathogen [16], leaving the possibility of an environmental reservoir in Africa for MAF from which other pathogenic human-adapted lineages evolved. Another proposed hypothesis is that MAF could be adapted to an unchanged environment, i.e. a rural area without the urbanization brought by colonization [11] or low population densities found in rural areas, in contrast to cities[12]. Finally, the geographical restriction could also be due to host genetic factors [17–19]. West-African people may be the only populations susceptible to MAF infection and progression to disease[17], as also supported by a study in the USA where more than 90% of occasionally isolated MAF strains were diagnosed from West-African migrants or people born in USA from parents of African origin [20]. L5 is mostly restricted to the countries around the Guinea gulf (Ghana, Togo, Benin, Nigeria)[8]. A study in Ghana showed that L5 is associated with Ewe ethnicity, which people are also living in countries next to the east of Ghana, as well as people of related ethnic groups (Fon, Aja, Gen (Mina), Goun, Phla-Phera) that are called Gbe languages [21–23]. This association of L5 with Ewe ethnicity could then explain its geographical restriction, and supports the host genetic factors hypothesis [21]. However, more studies are needed for a better understanding of the reasons why MAF is restricted to West-Africa.

Trends in MAF prevalence over-time: is MAF disappearing?

In some countries (Burkina Faso, Cameroon, Guinea-Bissau), recent studies reported a decrease in MAF prevalence [24,25]. The L5 prevalence in Cameroon, which was estimated at 56% (by biochemical/ phenotypic identification) in 1971, decreased to 9% between 1997-1998 (by molecular identification) [26] and has almost disappeared from the country in 2013 [27]. On the other hand, recent (2012) studies reported L5 at 14%-33% in regions of neighboring Nigeria [28,29], while in Ghana, the L5 prevalence was stable over many years [21] and in The Gambia the L6 prevalence did not change over a 7-year period [30]. The question that arises is, first, whether methodological issues explain the apparent disappearance of MAF, and, if real, whether ancestral lineages including MAF are more adapted to low density populations and are disappearing with the increasing density of populations (as hypothesized in the above section on geographical restriction of MAF).

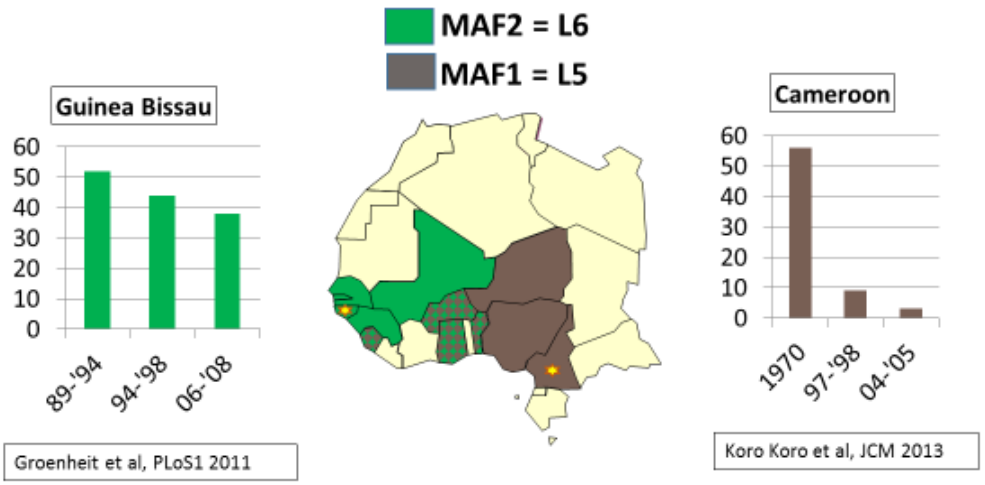

Figure 3. (Trends in) Prevalence of *M. africanum* L5 and L6 in Guinea Bissau and Cameroon (Image courtesy of de Jong B.C.)

Groenheit et al, PLoS One 2011 [25]; Koro Koro et al, JCM 2013 [27]

L5: lineage 5; L6: lineage 6

The two yellow stars/points on the map indicate the geographical location of Guinea Bissau (on the left side of the map) and Cameroon (on the right side of the map).

Implication of strain diversity for TB presentation (clinical characteristics of patients)

Host factors (such as ethnicity, discussed above) could play a role in the presentation of TB. Experiments in animal models showed that modern lineages are more virulent than ancient lineages including MAF, and also in humans there is a suggestion that modern lineages are more virulent than other lineages, while no such association was found in other studies [5]. Indicators of virulence include the ability of the strain to infect a host and cause disease, and impact on disease severity and treatment outcome. Ancestral lineages including MAF have reduced transmissibility compared to modern

lineages [5,20,31]. Compared to L4, L6 was associated with positive HIV status[22], older age of patient, severe malnutrition and worse disease on chest X-ray [30,32]. Furthermore, L6 is associated with a slower progression to disease compared to *M. tuberculosis sensu stricto* [33]. Although, many studies have been conducted on L6 regarding TB presentation, little is known on L5.

Depending on the history of previous TB treatment, TB patients are classified as new patients or previously treated patients. New patients are those who have never been treated before for TB or who have received less than one month of TB treatment. Previously-treated patients -also referred as retreatment patients- are those with a new episode of TB disease and have been previously treated for one month or more for TB. Previously-treated patients are classified in three main categories including 'treatment after failure', 'treatment after relapse', 'treatment after lost to follow-up' (also called default patient) [34]. Relapse patients are those who have previously been treated for TB, were declared cured or treatment completed (without bacteriological proof of cure) at the end of their most recent course of treatment, and have now been diagnosed with a recurrent episode of TB (clinical relapse; either a true relapse with reactivation of the same strain or a new episode of TB caused by reinfection by an exogenous strain). Treatment after failure patients are those who have previously been treated for TB and whose treatment failed at the end of their most recent course of treatment. Finally treatment after loss to follow-up patients (default patients) are those who have previously been treated for TB and have no documentation of cure or treatment completion at the end of their most recent course of treatment [34].

Treatment failure and relapse/reactivation can be caused by drug-resistant MTBC strains, but also by drug-susceptible strains. This could be due, besides other factors, to non-adherence to treatment. Whether the lineage of the MTBC strain is associated with TB recurrence (treatment failure, relapse/reactivation) remains unknown.

L2 (Beijing) is known to spread rapidly [5], most likely to (rapidly) progress to disease [33] and is associated with drug-resistance in some settings [5].

Treatment of TB

Before the 2017 WHO updated recommendation on TB treatment, drug susceptible TB in new patient was treated with a 6-month regimen (Category 1, 1st line treatment) whereas, in previously-treated patients it was treated with a 8-month regimen including streptomycin injection as a difference (Category 2). In 2017, the WHO recommended the use of Category 1 for new and previously-treated patients with proven or probable drug susceptible TB. This regimen is taken daily for 6 months and consists of two months intake of combined tablets containing INH, RIF, pyrazinamide and ethambutol (called intensive phase), followed by four months intake of combined tablets containing RIF and INH (2HRZE/4HR) [35]. MDR-TB and RIF-resistant TB are treated with an MDR-TB regimen, which is either the MDR shorter regimen (9-12 months) or the MDR longer regimen (longer than 9-12 months and usually 18-20 months). The longer MDR regimen can be used as individualized regimen (Fig.4) while the shorter regimen is more standardized (4–6Km–Mfx–Cfz–Pto–Z–E–$H_{high\ dose}$/5Mfx–Cfz–Z–E)[36] and usually last 9 months (4Km–Mfx–Cfz–Pto–Z–E–$H_{high\ dose}$/5Mfx–Cfz–Z–E) . The use of the shorter MDR regimen is

conditional and subjected to some requirements (Fig.5). For patients with INH-resistant but RIF-sensitive strains, the WHO recently recommended the use of RIF, EMB, pyrazinamide-levofloxacin for 6 months (6REZ-Lfx) [37].

Groups & steps	Medicine	
Group A: Include all three medicines	levofloxacin *OR* moxifloxacin	Lfx Mfx
	bedaquiline[2,3]	Bdq
	linezolid[4]	Lzd
Group B: Add one or both medicines	clofazimine	Cfz
	cycloserine *OR* terizidone	Cs Trd
Group C: Add to complete the regimen and when medicines from Groups A and B cannot be used	ethambutol	E
	delamanid[3,5]	Dlm
	pyrazinamide[6]	Z
	imipenem–cilastatin *OR* meropenem[7]	Ipm–Cln Mpm
	amikacin (*OR* streptomycin)[8]	Am (S)
	ethionamide *OR* prothionamide[9]	Eto Pto
	p-aminosalicylic acid[9]	PAS

Figure 4. Grouping of medicines recommended for use in longer MDR-TB regimens[1] (adapted from WHO, 2019 [37])

[1] This table is intended to guide the design of individualized, longer MDR-TB regimens (the composition of the recommended shorter MDR-TB regimen is largely standardized). Medicines in Group C are ranked by decreasing order of usual preference for use subject to other considerations.

[2] Evidence on the safety and effectiveness of bedaquiline use beyond 6 months and below the age of 6 years was insufficient for review. Use of bedaquiline beyond these limits should follow best practices in "off-label" use.

[3] Evidence on the concurrent use of bedaquiline and delamanid was insufficient for review.

[4] Use of linezolid for at least 6 months was shown to increase effectiveness, although toxicity may limit use. The analysis suggested that using linezolid for the whole duration of treatment would optimize its effect (about 70% of patients on linezolid with data received it for more than

6 months and 30% for 18 months or the whole duration). No patient predictors for early cessation of linezolid could be inferred from the IPD (individual patient data) subanalysis.

[5] Evidence on the safety and effectiveness of delamanid beyond 6 months and below the age of 3 years was insufficient for review. Use of delamanid beyond these limits should follow best practices in "off-label" use.

[6] Pyrazinamide is counted as an effective agent only when DST results confirm susceptibility.

[7] Every dose of imipenem–cilastatin and meropenem is administered with clavulanic acid, which is available only in formulations combined with amoxicillin. Amoxicillin–clavulanic acid is not counted as an additional effective TB agent and should not be used without imipenem–cilastatin or meropenem.

[8] Amikacin and streptomycin are to be considered only if DST results confirm susceptibility and high-quality audiometry monitoring for hearing loss can be ensured. Streptomycin is to be considered only if amikacin cannot be used (unavailable or documented resistance) and if DST results confirm susceptibility (resistance to streptomycin is not detectable with second-line molecular LPAs and phenotypic DST is required). Kanamycin and capreomycin are no longer recommended for use in MDR-TB regimens.

[9] These agents showed effectiveness only in regimens without bedaquiline, linezolid, clofazimine or delamanid, and are thus proposed only when other options to compose a regimen are not possible.

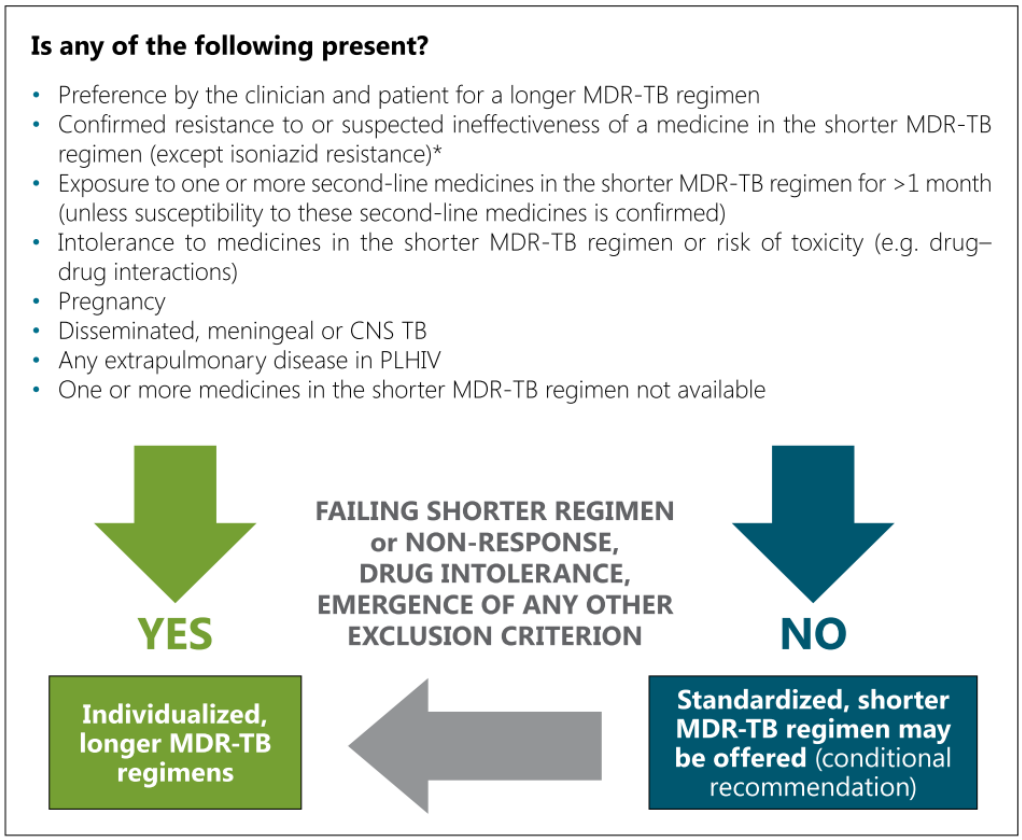

Figure 5. Criteria to decide whether the shorter MDR regimen could be offered (adapted from WHO, 2019 [37])

* Strains from MDR/RR-TB patients should ideally be tested for resistance to fluoroquinolones and other regimen components regardless of the type of MDR-TB treatment regimen offered.

TB and drug-resistant TB diagnosis, and implications of strain diversity for TB diagnosis

TB diagnosis and drug resistance detection

Smear microscopy has been used for decades for TB diagnosis, but is not very sensitive (at least 10^4 AFB/mL sputum to yield positivity [38]) and cannot distinguish MTBC non-tuberculous mycobacteria (NTM). The advent of

GeneXpert MTB/RIF (Cepheid, USA) has improved the sensitivity and specificity of TB diagnosis, yet it is in turn less sensitive than culture (at least 100 AFB/mL) which is the reference standard for bacteriological TB confirmation [1]. The GeneXpert MTB/RIF targets the 81-bp rifampicin-resistance determining region (RDRR) of the *rpoB* gene (5 probes: A, B, C, D, E) using a cartridge in a real-time PCR system which integrates sample purification, DNA amplification and target sequence detection. It detects simultaneously the presence of MTBC and resistance to RIF within 2 hours. RIF resistance is detected by the non-amplification of at least one of the probes or when the delta Ct (cycle threshold) between two probes is at least 4 [39]. The Xpert assays can be directly applied to sputum or on concentrated sediments [39], and was recommended as a rapid molecular diagnostic tool by the World Health Organization (WHO) (in 2010 for adults, in 2013 for children). Hence, some countries are phasing out microscopy as initial tool. Culture as reference standard for the detection of MTBC may be challenged by the new GeneXpert MTB/RIF (Xpert Ultra, Cepheid, USA) which is more sensitive than the classical Xpert (20 CFU/mL versus 130 CFU/mL), by addition of the repetitive elements IS*6110*, IS*1081* as additional targets beyond the *rpoB* gene. Xpert Ultra was recommended by WHO in 2017 for TB diagnosis in all adults and children with symptoms and signs of TB, with a strong recommendation for suspected MDR-TB, HIV-infected persons and TB meningitis [40,41]. The higher sensitivity of Ultra may yield 'trace' positive results [41]. Against culture as reference standard, such results may be considered "false positive", and interpretation of the patient's clinical presentation is necessary to decide whether treatment for TB is indicated, versus close follow-up with additional diagnostic testing. For instance, a patient who was treated for TB in the past may continue to shed detectable MTBC DNA for years after cure [42–44], which is likely prolonged with the more sensitive

Xpert Ultra. For those considerations, for previously-treated patients, it is recommended to repeat a 'trace' result on a second sample (lower specificity for TB diagnosis in patients with previous history of TB; 91.5% versus 95.9%) [40,41].

Other molecular diagnostic tests that can be directly applied on sputum (rapid diagnosis and resistance detection) include line probe assays (LPAs; for first- and second-line drugs), or in-house amplification methods, such as the *rpoB* gene's 1764-bp conventional nested-PCR [45]. The latter is more accurate for RIF resistance detection as it covers a longer fragment of the *rpoB* gene (1764 bp vs 81 bp covered by Xpert and first line LPAs) and detects mutations outside the 81-bp RDRR hotspot region, but requires an additional step of sequencing and sequence analysis after the PCR [45]. Although, molecular diagnostics are rapid, culture-based phenotypic methods that can take up to 12 weeks (culture + pDST) to provide results, are still considered the reference standard for drug-susceptibility testing (DST)[1].

Realization of Xpert MTB/RIF

Xpert MTB/RIF is used for TB diagnosis and also serves as a sputum-based genotypic drug susceptibility testing (detection of resistance to RIF). At least one milliliter of sputum is inactivated with two milliliters of the inactivation reagent provided by the Xpert manufacturer for sputum fluidification and inactivation. After homogenization and incubation at ambient temperature for fifteen minutes, two milliliters of the mixture are added in the Xpert cartridge and the cartridge is loaded in the Xpert machine. After nearly two hours, the results for TB diagnosis and resistance detection for RIF (if TB positive) are obtained (Fig.6)[46]. The obtained Ct value reflects the bacillary load in the sample tested and allows for grading of the positive result, with 'MTB detected

high' reported for Ct values <16, 'MTB detected medium' in case of a Ct value between 16 and 22, 'MTB detected low' with a Ct of 22 to 28, and 'MTB detected very low' for Ct values >28. The qPCR terminates at the 39th cycle.

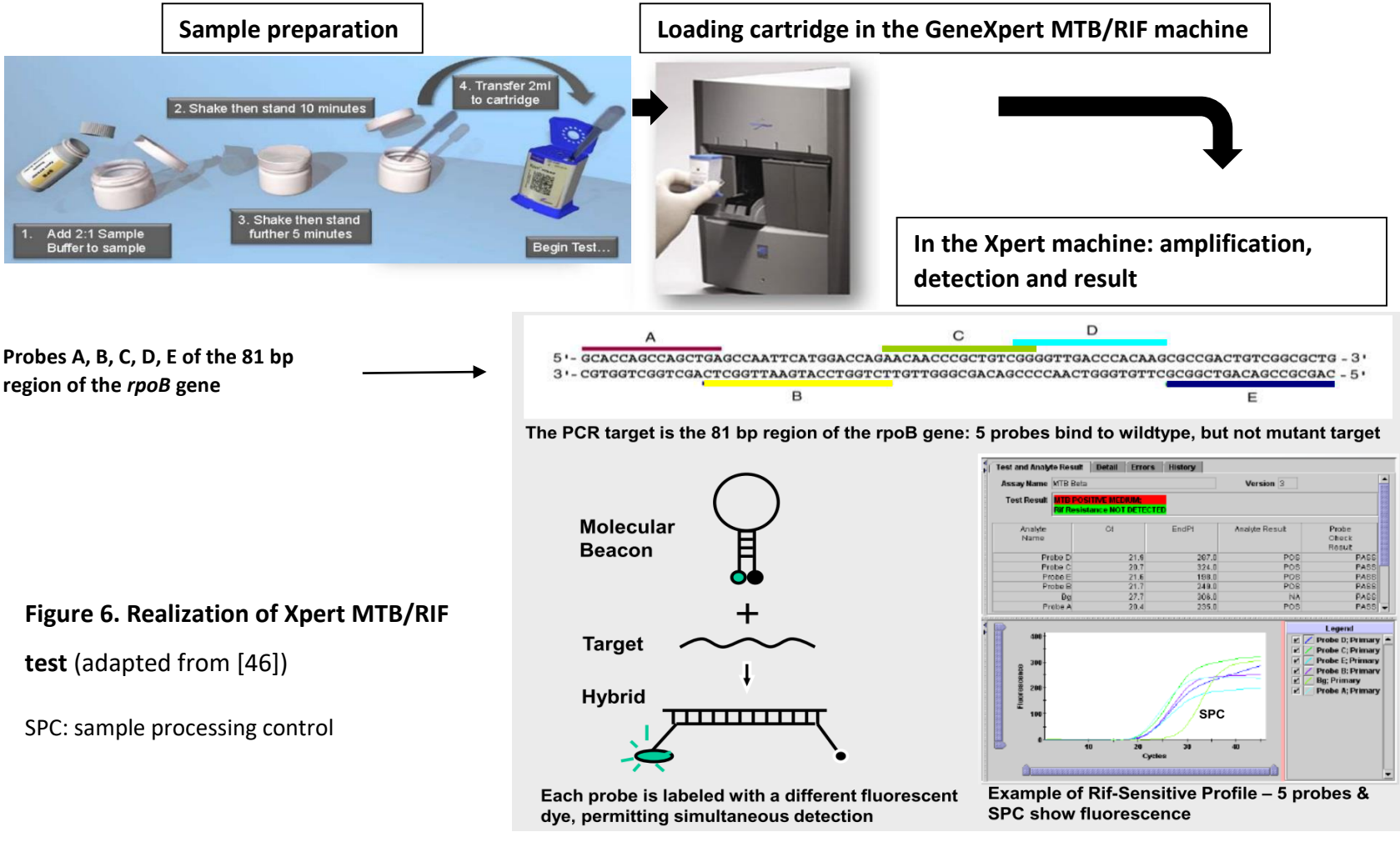

Figure 6. Realization of Xpert MTB/RIF test (adapted from [46])

SPC: sample processing control

Mycobacterial culture

Culture is the reference standard for bacteriological confirmation of TB. Although Xpert is widely used for TB diagnosis, culture is needed for TB diagnosis in Xpert negative specimens, and extra-pulmonary specimens as Xpert is not recommended for such specimens. Furthermore, growth of the MTBC in culture is necessary for phenotypic DST (pDST), MDR/RR-TB patient treatment monitoring [1] and some molecular methods like strain typing [47]. Mycobacterial culture can be performed on solid or liquid media. The solid culture media for mycobacteria culture include the egg-based Löwenstein-Jensen (LJ) medium which is the most used, and Middlebrook 7H10 or 7H11 agar. Liquid media include MGIT (Mycobacterial Growth Indicator Tube) as the most used and Middlebrook 7H9. Incubation of the specimens on the culture medium is generally preceded by a decontamination step to avoid that other bacteria present in the specimens outgrow the mycobacteria, as most mycobacterial species (including MTBC) of mycobacteria grow slower than other commensal bacteria. The decontamination step also enables the liquefaction of the sputum. After inoculation, the LJ medium is incubated at 37°C for up to 8 weeks or even 13 weeks (3 months) -to allow the growth of MAF as recommended by Castets et al [48]- before reporting a negative result.

MGIT culture can be automated (BACTEC 960 MGIT) or manual (BBL MGIT). The MGIT tube, a commercial liquid medium, is more expensive than LJ medium and contains a modified Middlebrook 7H9 broth [49]. In addition to Middlebrook 7H9 liquid media, the MGIT tube contains an oxygen-quenched fluorochrome (tris 4,7-diphenyl-1,10-phenonthroline ruthenium chloride pentahydrate) embedded in silicone at the bottom of the tube. During microbial growth within the tube, the free oxygen is utilized and is replaced

with carbon dioxide. With the reduction of free oxygen, the fluorochrome is no longer inhibited, resulting in fluorescence within the MGIT tube when visualized under UV light. The intensity of fluorescence is directly proportional to the extent of oxygen reduction in the tube. Growth in the MGIT tube resulting into the fluorescence of the tube is detected automatically by the MGIT machine or manually with a UV transilluminator if a conventional incubator is used [49].

Identification of MTBC

Identification of the MTBC, and its distinction from NTM, is crucial for pDST, as NTM can be intrinsically resistant to anti-TB drugs and be mistaken for MDR-TB. Conventional essential identification tests include growth on paranitro-benzoic acid (PNB) medium and the heat (68°C) labile catalase test (Fig. 7)[50]. Mpt64 antigen testing is an easy and rapid immuno-chromatography alternative test that has been shown to be highly sensitive and specific for the identification of MTBC, although only few studies have used strains isolated from solid medium [51,52]. However, a recent study in The Gambia reported that the test has a low sensitivity for identifying L6 as a member of the MTBC [53]. Furthermore, it is suggested [54] that the presence of a non-synonymous single nucleotide polymorphism (nSNP) in the *mpt64* gene of L5 strains could have consequences for the identification of the MTBC with the Mpt64 antigen test in countries where L5 is endemic. Four MPT64 tests exist and include Capilia TB-Neo (Tauns Laboratories, Numazu, Japan), TB Ag MPT64 Rapid Test (SD Bioline, Kyonggi-do, South Korea), TBcID (Becton Dickinson Microbiology Systems, Sparks, USA) and TBCheck MPT64 (Hain Lifescience). The three first are recommended by the WHO.

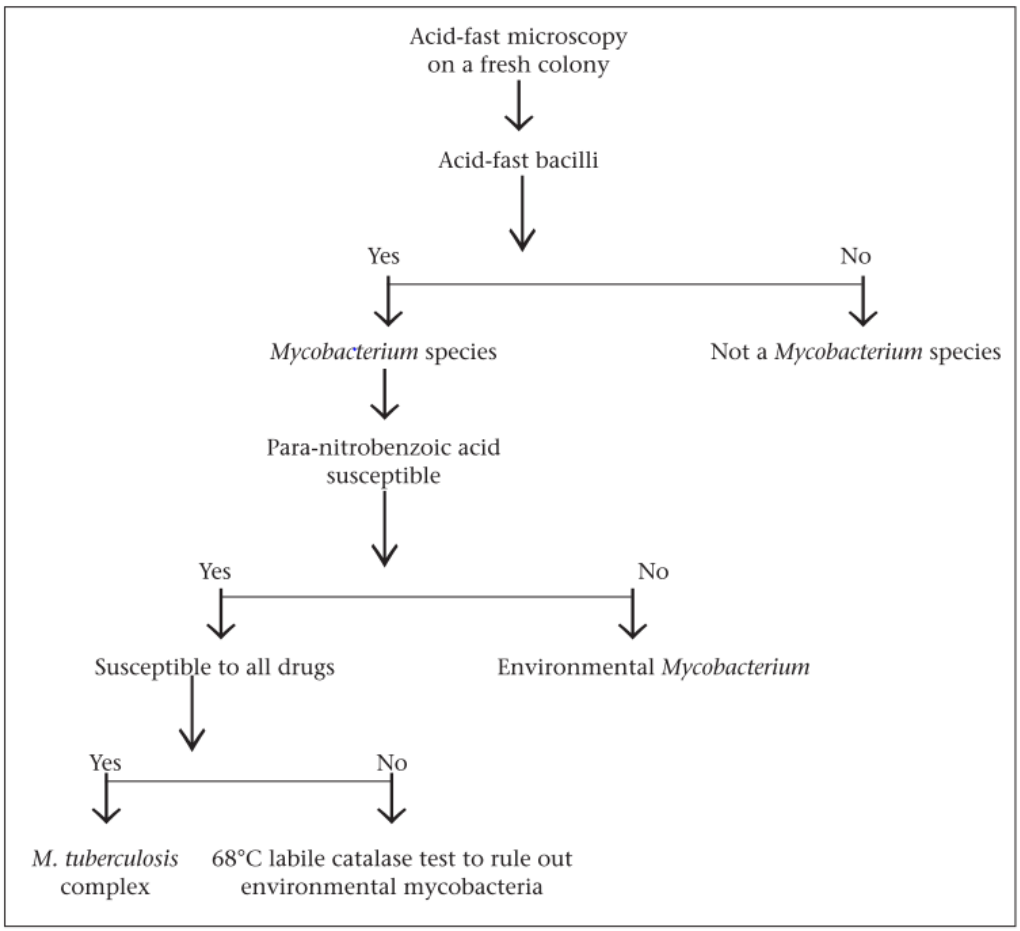

Figure 7. Conventional phenotypic identification of the MTBC (Rieder et al, 2007 [50]). Environmental mycobacteria: include NTM that are opportunistic pathogens.

The conventional PNB identification tests the ability of the bacteria to grow in the presence of PNB. Identification using PNB is usually simultaneously performed along with pDST, even after confirming presence of MTBC by the MPT64 assay, to exclude co-existence of MTBC and NTM in the isolate, which could still confound the resistance profile. For the identification on LJ medium, a slant containing PNB (at a finale concentration of 500µ µg/mL in the LJ medium) is inoculated with approximately 1000 bacilli corresponding to 10 µL

of the 10^{-2} dilution (of the bacterial suspension MacFarland 1) used for pDST to inoculate control and drug containing slants. After inoculation, the LJ slants are incubated for up to 6 weeks before reporting no growth on the PNB containing slant. Bacilli of the MTBC do not grow in the presence of PNB, while most NTM do [50].

The heat (68°C) labile catalase test is also used to differentiate bacilli of the MTBC from NTM (environmental mycobacteria). NTM produce a catalase that is stable when heated (68°C) or not, while most MTBC produce a catalase that is only stable at room temperature and labile when heated, except some isoniazid-resistant MTBC strains that do no produce catalase due to mutations in the *katG* gene (catalase-negative strains). For each isolate, two tubes of bacterial suspension (2-3 loopfuls of colonies suspended into 0.5 mL of phosphate buffered saline) are made of which one is incubated in a water bath at 68°C and the other at room temperature. After twenty minutes of incubation, the 68°C tube is cooled down to room temperature. Then 0.5mL of freshly prepared Tween-peroxide mixture (mixture 1:1 of 30% hydrogen peroxide and 10% Tween 80) is added to each of the two tubes. The production of air bubbles indicates the stability/presence of catalase, and can be recorded as negative if twenty minutes has passed without formation of the bubbles after the Tween-peroxide mixture is added. For MTBC isolates the room temperature tube is catalase positive while the 68°C-heated tube is catalase negative. For NTM, both tubes are catalase positive [50].

The MPT64 antigen rapid identification test is a lateral flow immuno-chromatographic test based on the detection of the MPT64 antigen produced by MTBC strains in positive cultures. This antigen is not produced by NTM. A strip-like device/casette is used for the test. The MPT64 test device consists of

a well, one test band and one control band. In the well, anti-MPT64 conjugated with colloidal gold particles were immobilized, for (sample's) MPT64 antigen capture and detection in a sandwich type assay. The test band contains immobilized 'mouse monoclonal anti-MPT64' (an antibody which recognizes another epitope of the MPT64 antigen). In the control band, 'goat anti-mouse immunoglobulin G' is immobilized. From a solid culture, a few (3-4) colonies of the isolate to test are suspended in 200 µL of the extraction buffer provided by the test manufacturer or the condensation fluid (if any) in the LJ slant is used. If the test is done from a liquid culture, the culture suspension can be directly used without dilution with the extraction buffer provided. Then 100 µL of the suspension (extraction buffer based or condensation fluid or liquid culture) is loaded into the well of the MPT64 test device. The bacterial suspension flows by capillarity to the test and control bands. As the bacterial suspension applied flows laterally through the membrane, the antibody-colloidal gold conjugate (in the sample well) binds to the MPT64 antigen in the sample, and forms a complex that flows further and bind to the mouse monoclonal anti-MPT64 on the solid phase in the test band, producing red to purple color line if a reaction occurs. In the absence of MPT64, there is no line in the test band region. After fifteen minutes of incubation (room temperature) the MPT64 test is reported positive if a colour (red to purple) signal line appears in the control and the test bands, whereas the test is reported negative if a colour signal line only appears in the control band and not in the test band (Fig.8) [55].

The GenoType Mycobacterium CM/AS (Hain Lifescience, Germany) test is a LPA used for the genotypic identification of mycobacteria species from positive culture. After a PCR, the amplicons are hybridized on a strip on which probes were immobilized by the manufacturer. Two kind of strips can be used: the strip for GenoType Mycobacterium CM (common mycobacteria) identifies common

mycobacteria species and classifies the other as a group of species for further identification of individual species using the GenoType Mycobacterim AS (additional species) strip. Based on the banding pattern of the strip after hybridization, the isolate tested can be classified as belonging to a specific mycobacteria species or a specific group of mycobacteria species (GenoType CM, Fig.9)[56].

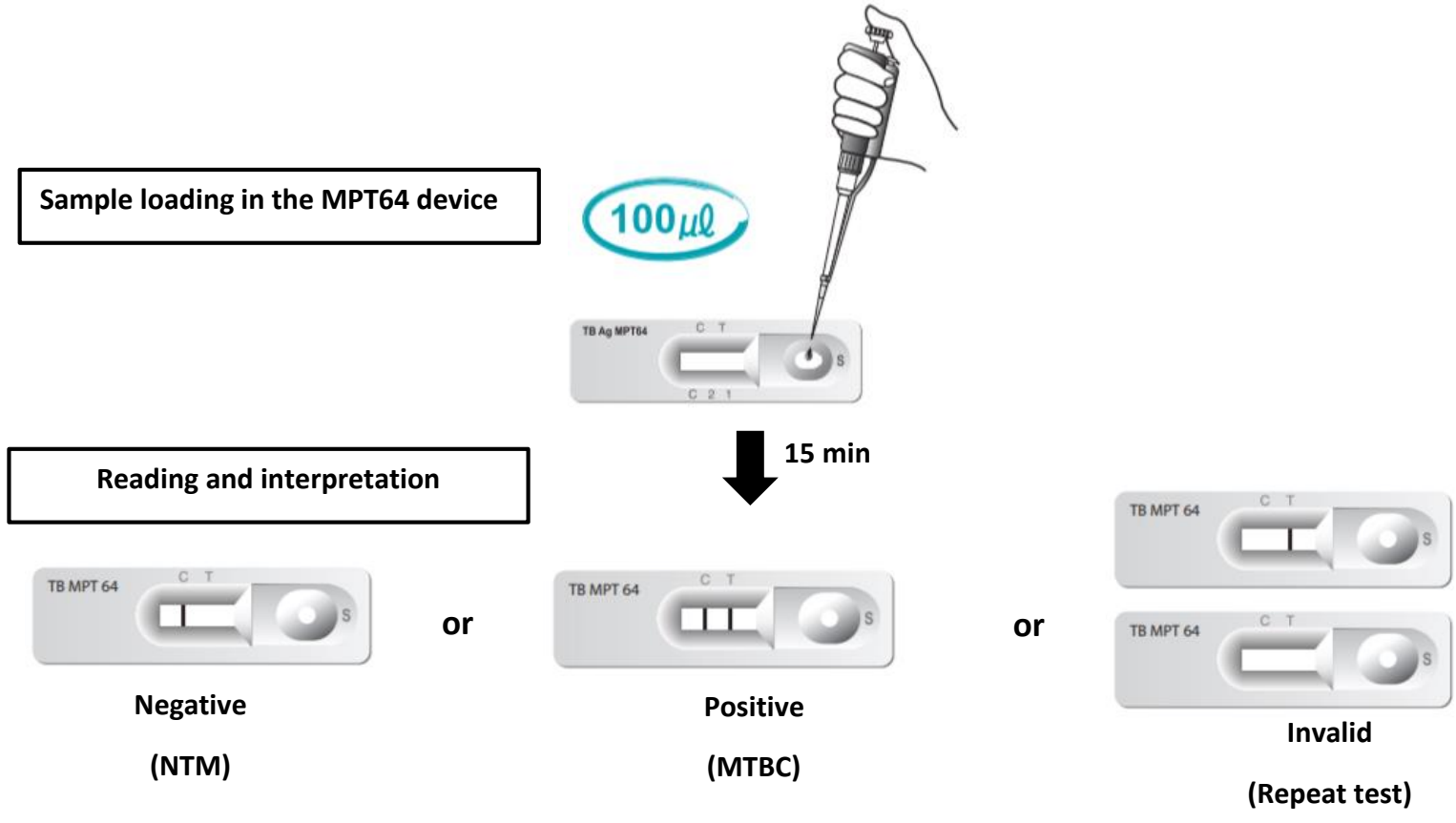

Figure 8. Realization, reading and interpretation of the MPT64 test (adapted from [55]). C: control, T: test, S: sample well

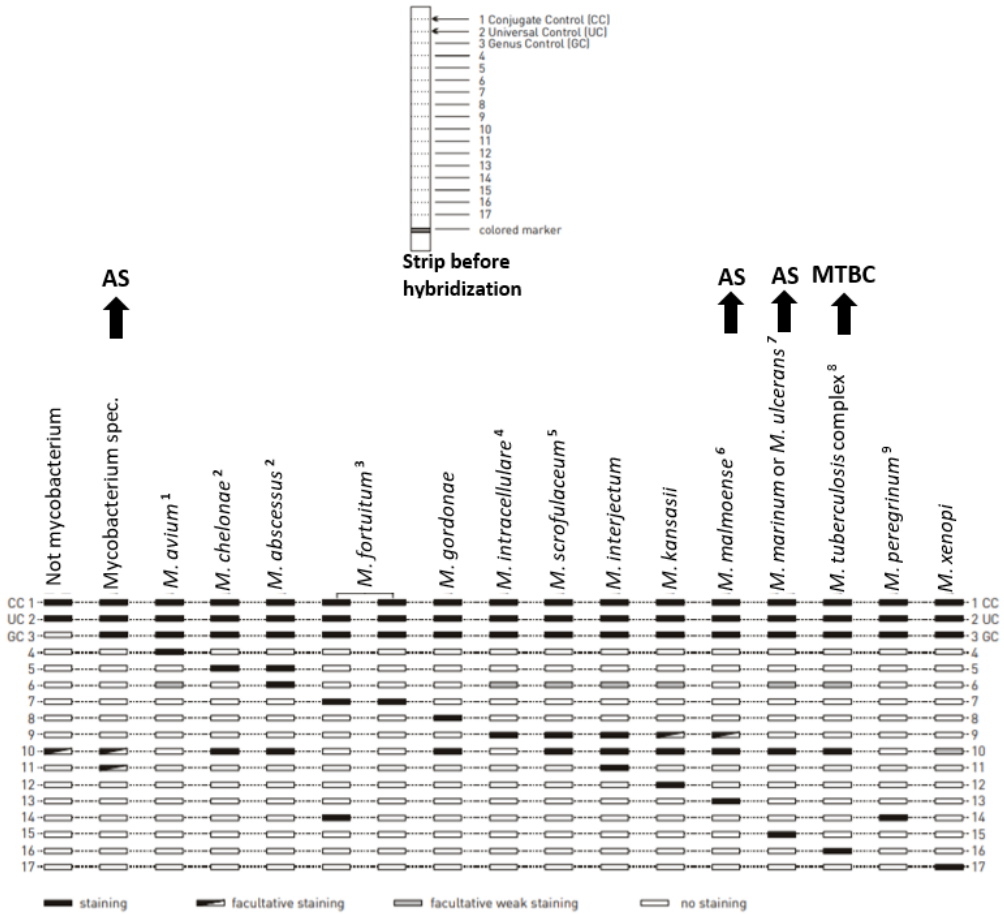

Figure 9. Genotype CM/AS strips before and after hybridization & interpretation (adapted from [56]**)**

AS: Species may possibly be identified with the GenoType Mycobacterium AS kit (strip).

MTBC: For further differentiation use the GenoType MTBC kit.

1 Does not include other species of the *M. avium* complex.

2 *M. immunogenum* (belongs to the *M. abscessus*/*M. chelonae* group) shows the same banding pattern as *M. chelonae* or *M. abscessus*. In case the quality and/or quantity of the extracted DNA does not allow an efficient amplification, the amplicon hybridizing both to the Genus Control and to band 6 may have been supplanted due to competition of the single reactions during amplification. In this case, *M. abscessus* shows the banding pattern identifying *M.*

chelonae. However, as long as the specifications given in these instructions for use are observed and the DNA polymerase used for performance evaluation is applied, this does not occur.

[3] Due to sequence variations within the species two different *M. fortuitum* banding patterns do occur. *M. mageritense* shows the *M. fortuitum* banding pattern as depicted in the right column.

[4] *M. chimaera* shows the same banding pattern as *M. intracellulare*.

[5] *M. paraffinicum* and *M. parascrofulaceum* show the same banding pattern as *M. scrofulaceum*.

[6] *M. haemophilum*, *M. palustre*, and *M. nebraskense* show the same banding pattern as *M. malmoense*. *M. haemophilum*/*M. nebraskense* can be identified with the GenoType Mycobacterium AS kit.

[7] *M. ulcerans* can be identified with the GenoType Mycobacterium AS kit.

[8] If band 15 has also stained positive, additional detection methods must be applied.

[9] *M. alvei* and *M. septicum* show the same banding pattern as *M. peregrinum*.

Culture-based phenotypic drug susceptibility testing (pDST)

pDST can be done on various culture media including the solid LJ medium and liquid MGIT media which are the most used, apart from Middlebrook 7H9 broth, or 7H10 and 7H11 agar. Among pDST techniques, the proportion method on LJ or pDST in MGIT medium (liquid) are the most used worldwide [34,50]. The critical concentration of 1st line and 2nd line TB drugs tested for DST on various media are presented respectively in Fig.10a and Fig.10b [57].

For the proportion method, the critical proportion to differentiate susceptible from resistant strains, is 1% for most anti-TB drugs (10% for pyrazinamide, which can only be tested in adapted MGIT medium). LJ medium slants are incorporated with the anti-TB drugs at their critical concentration. When less bacteria grow (number of colony) on the slant incorporated with drug compared to the 1/100 diluted slant without drug (<1% growth), the strain is considered susceptible to that drug. However, when 1% or more of bacteria grow on the drug-incorporated slant, the strain is reported resistant to that drug. After inoculation LJ slants are incubated at 37°C for 4-6 weeks, so six weeks before reporting that a strain is drug sensitive [50,57].

pDST in MGIT, is more rapid than testing on solid medium with primary cultures being positive after a mean of 10 days, and a maximum incubation time of 13 days for subsequent susceptibility testing [57]. For pDST in MGIT, the principle is similar to what is described above for culture in MGIT medium, except that drug are added to the MGIT tubes for pDST. For each isolate, MGIT tube containing anti-TB drugs and control MGIT tube (without drug) are inoculated with the bacterial suspension. Growth is checked during the incubation automatically (if the MGIT machine is used) or manually with a UV transilluminator (if a conventional incubator is used). If there is growth in the

control tube and the tube containing drug, the isolated is reported resistant to anti-TB drug, and sensitive if there is no growth in the tube containing drug.

Medicine	Abbreviation	Critical concentrations (µg/ml) for DST by medium			
		Löwenstein–Jensen[a]	Middlebrook 7H10[a]	Middlebrook 7H11[a]	BACTEC MGIT liquid culture[a]
Rifampicin	RIF	40.0	1.0	1.0	1.0[b]
Isoniazid[c]	INH	0.2	0.2	0.2	0.1
Ethambutol[d]	EMB	2.0	5.0	7.5	5.0
Pyrazinamide[e]	PZA	-	-	-	100

Figure 10a. Critical concentrations for 1st line TB drugs DST on various media (WHO, 2018 [57])

[a] The use of the indirect proportion method is recommended. Other methods using solid media (such as the resistance ratio or absolute concentration) have not been adequately validated for anti-TB agents.

[b] The detection of rifampicin resistance using the BACTEC MGIT 960 system has limitations and cannot detect clinically significant resistance in certain isolates. The detection of resistance conferring mutations in the entire *rpoB* gene using DNA sequencing may be the most reliable method for the detection of rifampicin resistance.

[c] Patients with MTBC isolates that are resistant at the critical concentration may be effectively treated with high dose isoniazid.

[d] All phenotypic DST methods for ethambutol produce inconsistent results. DST is not recommended. [e]BACTEC MGIT 960 liquid culture method is the only WHO-recommended method for PZA susceptibility testing, though even this testing is reportedly associated with a high rate of false-positive resistance results. Careful inoculum preparation is essential for performing PZA testing reliably. The detection of resistance conferring mutations in the pncA gene using DNA sequencing may be the most reliable method for the detection of pyrazinamide resistance although there is emerging evidence of non-pncA mutational resistance to PZA.

Group	Medicine	Abbreviation	Critical concentrations (µg/ml) for DST by medium			
			Löwen-stein Jensen[1]	Middle-brook 7H10[1]	Middle-brook 7H11[1]	BACTEC MGIT liquid culture[1]
Group A	Levofloxacin (CC)	LFX[2,3]	**2.0**	1.0	-	1.0
	Moxifloxacin (CC)	MFX[2,3]	**1.0**	0.5	0.5	0.25
	Moxifloxacin (CB)[4]			2.0	-	1.0
	Bedaquiline[5]	BDQ	-	-	**0.25**	**1.0**
	Linezolid[6]	LZD	-	1.0	1.0	1.0
Group B	Clofazimine	CFZ	-	-	-	**1.0**
	Cycloserine TerizidoneTerizidone	CS TZD	- -	- -	- -	- -
Group C	Ethambutol[7]	E	2.0	5.0	7.5	5.0
	Delamanid[8]	DLM	-	-	**0.016**	**0.06**
	Pyrazinamide[9]	PZA	-	-	-	100.0
	Imipenem-cilastatin Meropenem	IMP/CLN MPM	- -	- -	- -	- -
	Amikacin[10] (Or Streptomycin)	AMK (S)	30.0 4.0	2.0 2.0	- 2.0	1.0 1.0
	Ethionamide Prothionamide	ETO PTO	40.0 40.0	5.0 -	10.0 -	5.0 2.5
	Para-aminosalicylic acid	PAS	-	-	-	-

Figure 10b. Critical concentrations for 2nd line TB drugs DST on various media

(Interim critical concentration are highlighted in red) (WHO, 2018 [57])

[1] The use of the indirect proportion method is recommended. Other methods using solid media (such as the resistance ratio or absolute concentration) have not been adequately validated for anti-TB agents.

[2] Testing of ofloxacin is not recommended as it is no longer used to treat DR-TB and laboratories should transition to testing the specific fluoroquinolones (levofloxacin and moxifloxacin) used in treatment regimens.

[3] Levofloxacin and moxifloxacin interim CCs for LJ established despite very limited data.

[4] Clinical breakpoint concentration (CB) for 7H10 and MGIT apply to high-dose moxifloxacin (i.e. 800 mg daily).

[5] No evidence is available on safety and effectiveness of BDQ beyond six months; individual patients who require prolonged use of BDQ will need to be managed according to 'off-label' best practices.

[6] Optimal duration of use of LZD is not established. Use for at least 6 months was shown to be highly effective, although toxicity may limit use for extended periods of time.

[7] DST not reliable and reproducible. DST is not recommended.

[8] The position of delamanid will be re-assessed once individual patient data by Otsuka are available for review. No evidence is available on effectiveness and safety of DLM beyond six months; individual patients who require prolonged use of DLM will need to be managed according to 'off-label' best practices.

[9] Pyrazinamide is only counted as an effective agent when DST results confirm susceptibility in a quality-assured laboratory. Its use with BDQ may be synergistic.

[10] Amikacin and streptomycin are only to be considered if DST results confirm susceptibility and high-quality audiology monitoring for hearing loss can be ensured. Streptomycin is to be considered only if amikacin cannot be used and if DST results confirm susceptibility. Streptomycin resistance is not detectable with 2nd line molecular line probe assays

Specimen storage/transport

Within countries, TB diagnostic services are usually shared between peripheral and/or intermediate TB clinic laboratories and a central reference laboratory. Not all TB diagnostics are available in peripheral laboratories, and samples are shipped to remote intermediate or reference laboratories for further tests such as culture, DST and molecular analysis (e.g. resistance detection, strain typing). This shipment of specimens to remote laboratories where specific platforms are available, is also useful for nationwide systematic drug resistance surveillance, periodic drug resistance surveys, and representative epidemiological/transmission studies. Logistical constraints can delay the time between sample collection and processing, which can impact specimen quality and result in unreliable test results. To avoid specimen deterioration, sputum storage reagents such as cetylpyridinium chloride (CPC), OMNIgene.SPUTUM (DNA Genotek, Canada), or ethanol (ETOH) can be used, although these have not been compared for long-term storage before subsequent culture or subsequent molecular diagnostic testing.

Typing of MTBC strain

The lineage of MTBC strains can be determined using spoligotype (CRISPR, Clustered Regulartory Short Palindromic Repeats) analysis [58,59], MIRU-VNTR (Mycobacterial Interspersed Repetitive Unit, Variable Number of Tandem Repeats)[60], IS*6110*-RFLP (Restriction Fragment Length Polymorphism) [61], RD (regions of difference) identification [62], whole genome sequencing (WGS), or lineage-specific SNP (single nucleotide polymorphism) analysis [59,62,63]. These typing techniques require culture positivity as they are usually performed

on positive cultures [47]. Among those methods, WGS has the highest resolution [47,62,64,65], yet is still relatively expensive and with limited availability in laboratories in TB endemic countries. The genotyping method of choice depends partially on the purpose. For transmission studies a high resolution (discriminatory power) may be needed [64,65] while a relatively lower resolution genotyping method like spoligotyping is sufficient for lineage assignment in most cases [47,66,67]. Furthermore, spoligotyping is cheaper and easily implementable in laboratories in developing countries, with the option of in-house preparation of membranes. In addition, it enables the detection of *M. africanum* lineages (L5 and L6) that have respectively more than five and five or less copies of the insertion sequence IS6110 [68] Although MIRU-VNTR has a higher resolution than spoligotyping, its discriminatory power varies by strain lineage [62]. A comparison of the main typing methods is presented in Table 2.

Spoligotyping

It tests the presence or absence of spacers (DNA sequence between sequence repeats) of the direct repeat (DR) region (Fig.11). There are 43 types of spacers detected by the spoligotyping analysis [58]. After a PCR of the DR region, the amplicons are hybridized on a membrane labelled with probes representing the 43 spacers (Fig.12). When a specific spacer is present in the strain, its hybridization to the membrane is indicated as a dark spot, while the absence of a spacer leaves the spot white (blank). The absence of specific spacers defines the MTBC lineage (Fig.13). The spoligotyping signature of lineages is presented in Fig. 13. The lineage can be assigned by uploading the spoligo-pattern of a strain in databases such as TBlineage database (http://tbinsight.cs.rpi.edu/run_tb_lineage.html), MIRU-VNTRplus (miru-

vntrplus.org), TBminer (https://info-demo.lirmm.fr/tbminer/). The SIT (spoligotype international type) can be found by uploading the spoligo-pattern in the SITVITWEB database (http://www.pasteur-guadeloupe.fr:8081/SITVIT2/).

Table 2. Comparison of strain typing methods

Characteristics	Spoligotyping	IS6110-RFLP	MIRU-VNTR (24 loci)	SNP (WGS)
Hunter Gaston Index (HGI) of discriminatory ability (in Stockholm)	0.9701 [69]	0.9975 [69]	0.9985 [69]	Highest [5,47,59,62, 70]
Turnaround time	+	++++ (Time-consuming [69])	+++ (Lower than RFLP [69])	++
Complexity of the method	+ (technically simple)[69]	++++ (labor-intensive [69])	+++	+++
Ease of interpretation	+++	No (Format of result difficult to exchange or compare between different laboratory [69])	+++	++++
Cost	+	+++	++	++++
Easily implementable in low resources settings	Yes	No	No	No
Need of skilled personnel	Yes +	Yes ++++	Yes +++	Yes +++
Ability to discriminate between strains with less than 6 copies of IS6110	Yes [71,72]	No [69,71]	Yes [71]	Yes
Culture isolate	Yes	Yes	Yes	Yes
Uncultured specimen	Sometimes used (performance not evaluated) [71,73]	Yes	Possible (more likely successful for high grade microcopy 3+ and 2+ specimens) [74,75]	no

Other		reproducibility problem in isolates presenting large copy number of IS6110 [76]	less discriminatory power for Beijing isolates [69]	
		needs high amount of genomic DNA [69]		needs high amount of genomic DNA [69]

HGI for Spoligotyping + MIRU-VNTR-24= 0.9986, HGI for Spoligotyping + IS6110-RFLP = 0.9988, HGI for Spoligo+MIRU-VNTR 24 +RFLP = 0.999

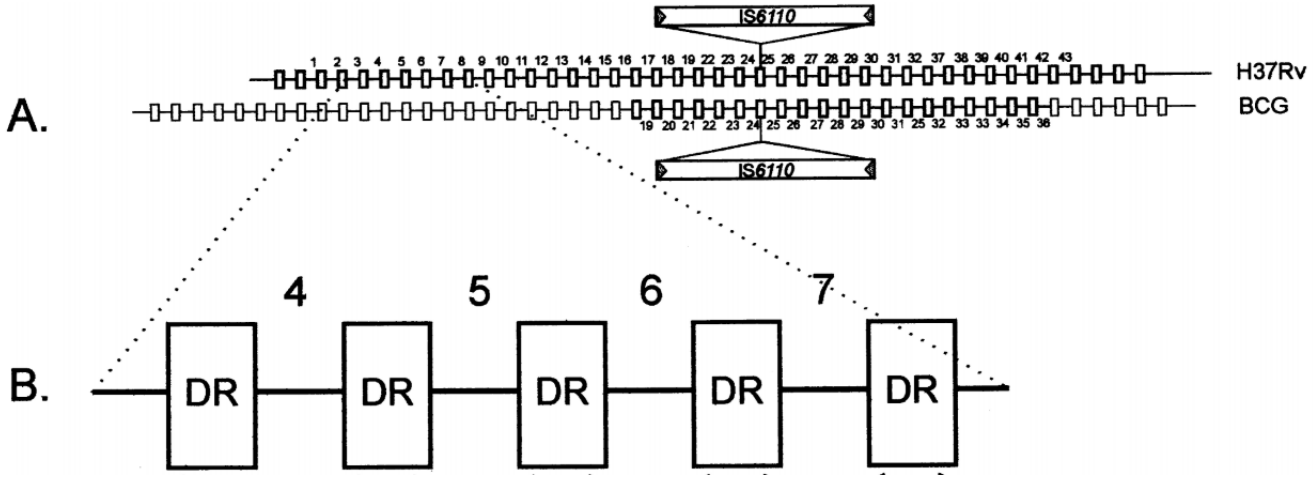

Figure 11. Target of the spoligotyping analysis: spacers in the direct repeat (DR) region of the MTBC (adapted from Kamerbeek et al, JCM 1997 [58]).

(A) Structure of the DR locus in the mycobacterial genome. The chromosomes of M. tuberculosis H37Rv and M. bovis BCG contain 48 and 41 DR repeats, respectively (depicted as rectangles), which are interspersed with unique spacers varying in length from 35 to 41 bp. (B) Example of the spacers 4-7 separated by DR repeats.

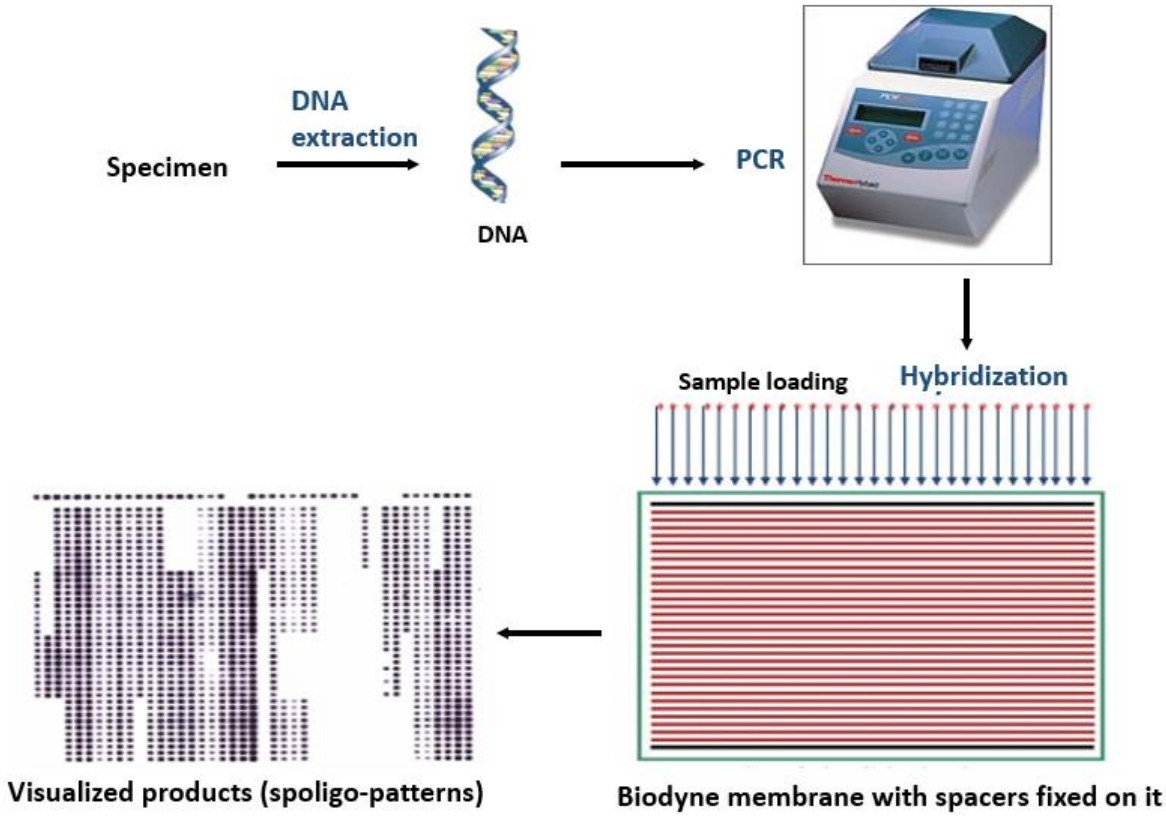

Figure 12. Realization of spoligotyping analysis (adapted from figure courtesy of Rigouts L.) On the biodyne membrane each red line indicates a specific spacer of the 43 spacers analyzed. In practice there are in total 43 spacers fixed on the membrane (so 43 such red lines).

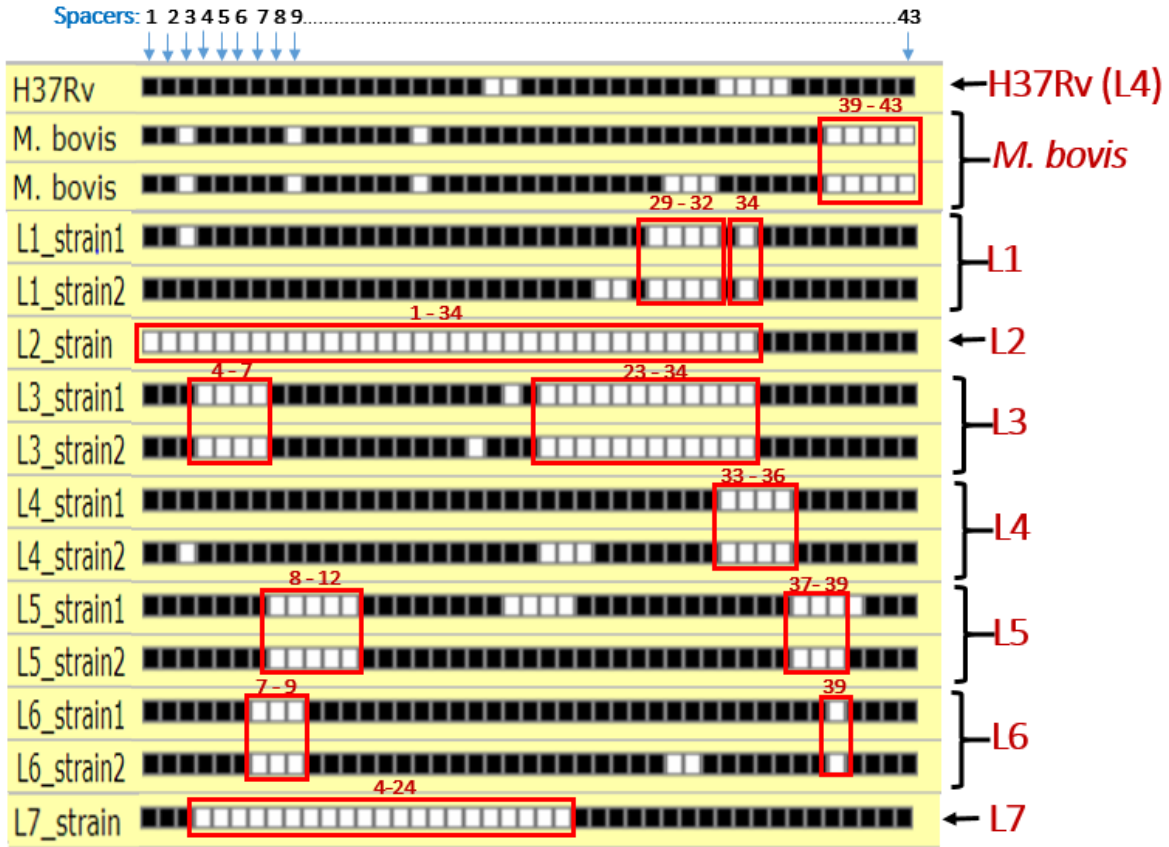

Figure 13. Spoligotyping signature for the different MTBC lineages (source: Sanoussi C. N.)

Mixed infections

Mixed (or polyclonal) TB infection occurs when more than one clone of TB bacilli is found in a patient specimen at a single point in time. Current methods used for diagnosis and typing have low sensitivity to detected mixed infections, and when they do, only a small proportion of the actual cases of mixed infection are detected [77]. However, genotyping applied directly on uncultured specimens alone or combined with isolate-based genotyping, is more likely to detect mixed infections [75,77,78]. In patients with mixed TB infection, a specific clone of TB bacilli may not be present in the collected sputum specimen, as different clones may be present in different cavities or parts of lungs. Also when the minority clone is present, it can be lost during specimen transport/storage [75], decontamination or culture. Hence, clonal ratio's of a specimen can change after culture, leaving only one clone/strain which grows in culture and is detected after culture. Mixed infections may have an impact on accurate diagnosis, effective treatment of individuals and control of TB[77]. For example, it was reported that in a mixed infection of an MDR strain with a drug-susceptible strain, the MDR strain persisted and grew during the treatment with 1st line drugs. Then, a subsequent treatment (of the MDR strain) was done with 2nd line drugs [77,79].

MTBC diversity in the era of genome sequencing/analysis

Genome sequencing is widely used for lineage determination/ strain classification, drug resistance detection, transmission studies, comparative genomics, understanding strain biology/characteristics, and beyond [70]. Genome sequencing (next-generation sequencing, Sanger sequencing) allows

in-depth and extended analyses that are not possible with most genotypic or phenotypic tests. For example, it allows drug resistance detection for one (targeted sequencing) or many drugs at once (WGS), can detect mixed infections and can be used to resolve discrepancies (evidence of mutation). However, this rapidly advancing field lacks a sufficient evidence base on which to correlate specific mutations with phenotypic resistance and clinical impact. While for 1st line drugs, genotypic prediction of drug susceptibility was found to be correlated with phenotypic susceptibility, this has not been fully established yet for second line drugs or new and repurposed drugs [80]. The resistance association with mutation, including information on the confidence and likelihood is planned to be made available in a database [81].

Since decades, in clinical laboratories, the H37Rv isolate is used as the reference isolate for TB laboratory analyses [82], and was the first MTBC genome to be fully sequenced (1998)[83][84]. In this era of genome sequencing/analysis, the completed H37Rv genome continues to be used as reference genome for MTBC genome analysis [70], despite possible differences between the different lineages [5,70,85,86].

TB in Benin

Benin is a West-African country with 11,186,785 inhabitants over 114,763 km^2. Each year, over 4000 cases of TB occur in Benin, with an incidence of 39 cases per 100,000 populations.

In total, there are 78 TB clinics (each including a laboratory) in Benin, including five intermediate-level and one central-level TB clinic, associated with the

national reference laboratory for tuberculosis in Cotonou (Laboratoire de Référence des Mycobactéries, LRM, Cotonou, Benin).

Microscopy or GeneXpert MTB/RIF are used for the bacteriological diagnosis of TB in TB clinics. Xpert MTB/RIF is being implemented nationwide in Benin to replace microscopy for TB diagnosis (and RIF resistance detection). So far Xpert is implemented and used in 31 of the 78 TB clinics in Benin. Xpert (two machines of four modules each) was first implemented in LRM in 2012 and used for RIF resistance detection in all previously-treated patients in Benin (after their samples are shipped from the TB clinics), diagnosis of TB in children and in microscopy-negative HIV-positive patients. In 2016, Xpert was implemented in the five other intermediary TB clinics in Benin. In January 2018, LRM started to use Xpert in replacement of microscopy for TB diagnosis in all patients including new patients (using two Xpert machine of 16 modules each). It is since then, that Xpert is being implemented in peripheral TB clinics nationwide, and used for the diagnosis of TB in new patients as well. Culture, pDST and other phenotypic and molecular TB diagnostic tests like LPAs are only available in the LRM, Cotonou. There is a system of sample referral from TB clinics to the LRM for continuous drug resistance surveillance and further laboratory analysis as needed. When a previously-treated TB patient is diagnosed in a 7TB clinic, his/her sputum sample is shipped to the LRM (within 3-4 days) using the package/courier delivery service of public transport. The LRM staff is informed beforehand so that the package is picked up at the public transport station upon arrival in Cotonou.

The first study of the molecular epidemiology of TB in Benin was conducted in 2005-2006, on patients recruited in only one city (Cotonou), and included only new TB patients. The prevalence of MTBC lineages was 7.7% for L1, 10.3% for

L2, 0% for L3, 42.3% for L4, 31% for L5 and 6% for L6, amounting to 37% for MAF (L5 + L6) [9,87]. An outbreak caused by a streptomycin-resistant L2 Beijing strain was also detected [88].

A review in 2010 [15], showed that Benin had the highest prevalence of L5 among West-African countries [8,9,15], based on the 2005-2006 molecular epidemiology study in Cotonou.

No *M. bovis* strain was identified from patients during that study. Another study on bovine TB in cattle was conducted in 2007 in Cotonou. From 1480 cattle clinically screened, 124 were suspected of TB. Among these, 100 were sampled. Specimens from 68 animals yielded a positive culture, with 46 of them successfully subcultured. For the culture, LJ medium slants along with slants of L.J supplemented with pyruvate (as preferred by *M. bovis* [89]) were inoculated. Phenotypic identification (PNB and catalase) showed that all the isolates were NTMs. So, no MTBC - including *M. bovis* - was isolated from cattle in Cotonou.

Regarding drug resistance, Benin is a low burden country for drug resistant TB. In addition to the continuous surveillance of TB drug resistance, periodic drug resistance surveys (DRS) are conducted. From the first DRS conducted in Benin in 1999, RIF-resistance was detected among 0.3% of new TB patients and 12% of previously-treated patients. During the 2010 DRS, RIF-resistance was detected among 13.3% of previously-treated TB patients, and 0.7% of new TB patients. No XDR was detected. In 2013, RIF-resistance was detected in 14% of previously-treated patients (continuous surveillance). To date, in Benin, DRS has been based on pDST, hence depending on culture positivity. Since the advent of Xpert MTB/RIF, which overcomes the necessity of positive culture for the detection of resistance to RIF since it is applied directly on sputa, no DRS

has been conducted. The use of Xpert could enable an accurate estimation of the current national trends of RIF-resistance in new and previously-treated TB patients.

In Benin, before 2018, the 6-month TB treatment regimen (2HRZE/4HR) was used for new patients with drug susceptible TB, while the 8-month regimen was used for the treatment of previously-treated patients with drug susceptible TB (as recommended by WHO at that time). After WHO has updated TB treatment regimens for the treatment of drug susceptible TB in 2017, the 6-month regimen (2HRZE/HR) has been used since 2018 for the treatment of both new and previously-treated TB patients with drug susceptible TB. For the treatment of MDR and RIF resistant TB, the shorter MDR regimen is used for 9 months (Fig. 14).

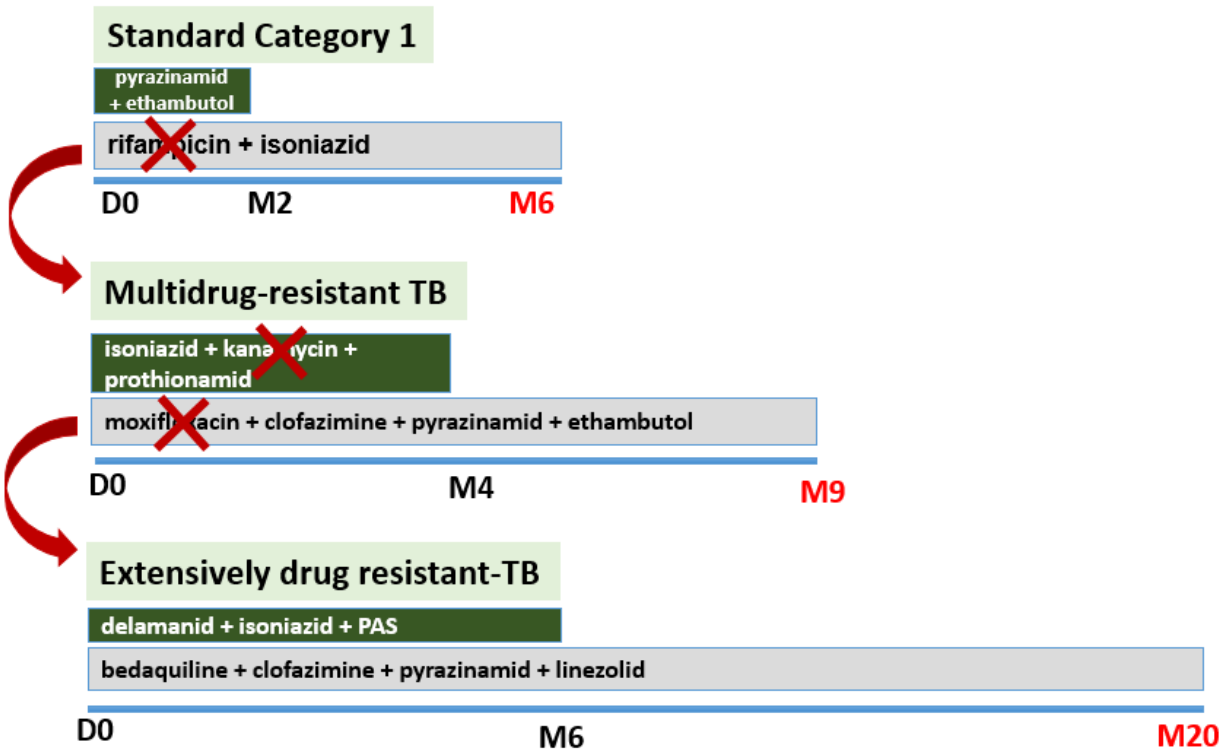

Figure 14. Drug regimens used for the treatment of drug susceptible, and drug-resistant (MDR, RIF-resistant) TB in Benin (adapted from image courtesy of Rigouts L.). M: month (M2: month 2); D0: day 0 (day of treatment initiation); X: if the strain is resistant to the drug (s) use the regimen indicated with the red arrow below.

Research questions addressed in this thesis

MAF is restricted to West Africa for reasons that remain unclear, although some hypotheses have been proposed as detailed earlier, including an association with ethnicity in West-Africa and adaption to low host population density [12,21]. Many studies were conducted about the virulence of L6 [5,15,22,32,33], in contrast to L5 which remains largely unstudied. A nationwide study on the molecular epidemiology of TB has to date not been done in Benin, and the one study done in 2005-2006 included only new patients from a single city [87]. As L5 is common in Benin (highest prevalence among West-African countries) [9,15], a nationwide study is needed to check whether L5 is equally distributed in the country, and whether its prevalence is decreasing as suggested in other countries.

Regarding diagnostics, growth in culture followed by distinction of MTBC from NTM, is needed for pDST. Most molecular epidemiological studies conducted to date worldwide performed lineage determination on positive cultures. It is unknown whether growth in culture is lineage-dependent. The rapid MPT64 antigen identification test reported to have low sensitivity for L6 detection [53], targets a protein for which the gene is suggested to be mutated in L5 [54]. For (nationwide) surveys/studies or further routine diagnostic analyses, a good preservation of the samples shipped to reference laboratories is crucial for good culture positivity yield and/or reliable molecular tests results, also for relatively long storage duration (logistic constraints).

Finally, in this era of genome sequencing, comparative genomics may allow an improved understanding of the genetic difference between L5 and others MTBC members.

The principal aim of this thesis was to improve the understanding of the implications of the diversity of the MTBC, on TB presentation, diagnosis, and outcome. Specifically, we aimed:

- To establish the population structure in a countrywide cohort of enhanced surveillance of new and previously treated patients in Benin
- To compare the characteristics of ancestral lineages such as L5 to modern lineages in terms of:
    - their prevalence in new patients versus previously-treated patients
    - their prevalence in rural versus urban (low versus high population density) patient residence areas
- To determine the clinical implications of TB due to L5 in Benin by comparing patient characteristics (age, sex, ethnic group, HIV status, BCG vaccine scar) and outcome of treatment between MAFL5 and other lineages
- To assess the implications of strain differences on TB diagnostics (culture, identification, sputum preservation for culture and/or molecular TB diagnostic testing)
- To compare genetic differences between L5 and the current reference H37Rv (L4) and other MTBC to identify candidate genes for phenotypic differences.

References

[1] WHO. Global tuberculosis report 2018. 2018.

[2] Enarson DA. Tuberculosis as a Global Public Health Problem. Mycobact. TB, vol. 2, Basel: KARGER; 2002, p. 1–6. doi:10.1159/000066885.

[3] Guidelines for surveillance of drug resistance in tuberculosis 5 th Edition. 2015.

[4] Leao S, Martin A, Mejia GI, Portaels F, Leao SC, Martin A, Mejia GI, Palomino JC, Robledo J, Telles MAS P-TF. Practical handbook for the phenotypic and genotypic identification of mycobacteria. 2004:77–126. doi:10.1007/BF01045823.

[5] Coscolla M, Gagneux S. Consequences of genomic diversity in mycobacterium tuberculosis. Semin Immunol 2014;26:431–44. doi:10.1016/j.smim.2014.09.012.

[6] Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial lineages causing pulmonary and extrapulmonary Tuberculosis, Ethiopia. Emerg Infect Dis 2013;19:460–3. doi:10.3201/eid1903.120256.

[7] Brosch R, Gordon S V, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci 2002;99:3684–9. doi:10.1073/pnas.052548299.

[8] Gehre F, Kumar S, Kendall L, Ejo M, Secka O, Ofori-Anyinam B, et al. A Mycobacterial Perspective on Tuberculosis in West Africa: Significant Geographical Variation of M. africanum and Other M. tuberculosis Complex Lineages. PLoS Negl Trop Dis 2016;10. doi:10.1371/journal.pntd.0004408.

[9] Gehre F, Antonio M, Faïhun F, Odoun M, Uwizeye C, de Rijk P, et al. The First Phylogeographic Population Structure and Analysis of Transmission Dynamics of M. africanum West African 1- Combining Molecular Data from Benin, Nigeria and Sierra Leone. PLoS One 2013;8:8–13. doi:10.1371/journal.pone.0077000.

[10] Wikipedia. Early human migrations n.d. https://en.wikipedia.org/wiki/Early_human_migrations.

[11] Comas I, Hailu E, Kiros T, Bekele S, Mekonnen W, Gumi B, et al. Population Genomics of Mycobacterium tuberculosis in Ethiopia Contradicts the Virgin Soil Hypothesis for Human Tuberculosis in Sub-Saharan Africa. Curr Biol 2015;25. doi:10.1016/j.cub.2015.10.061.

[12] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet 2013;45. doi:10.1038/ng.2744.

[13] Gagneux S. Host-pathogen coevolution in human tuberculosis n.d.

doi:10.1098/rstb.2011.0316.

[14] Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 2014;514:494–7. doi:10.1038/nature13591.

[15] De Jong BC, Antonio M, Gagneux S. Mycobacterium africanum-Review of an Important Cause of Human Tuberculosis in West Africa 2010. doi:10.1371/journal.pntd.0000744.

[16] Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol 2018;16. doi:10.1038/nrmicro.2018.8.

[17] van Tong H, Velavan TP, Thye T, Meyer CG. Human genetic factors in tuberculosis: an update. Trop Med Int Heal 2017;22:1063–71. doi:10.1111/tmi.12923.

[18] Gagneux S, Deriemer K, Van T, Kato-Maeda M, De Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. 2006.

[19] Salie M, van der Merwe L, Möller M, Daya M, van der Spuy GD, van Helden PD, et al. Associations between human leukocyte antigen class I variants and the Mycobacterium tuberculosis subtypes causing disease. J Infect Dis 2014;209:216–23. doi:10.1093/infdis/jit443.

[20] Sharma A, Bloss E, Heilig CM, Click ES. Tuberculosis Caused by Mycobacterium africanum, United States, 2004-2013. Emerg Infect Dis 2016;22:396–403. doi:10.3201/eid2203.151505.

[21] Asante-Poku A, Yeboah-Manu D, Otchere ID, Aboagye SY, Stucki D, Hattendorf J, et al. Mycobacterium africanum Is Associated with Patient Ethnicity in Ghana. PLoS Negl Trop Dis 2015. doi:10.1371/journal.pntd.0003370.

[22] Asante-Poku A, Otchere ID, Osei-Wusu S, Sarpong E, Baddoo A, Forson A, et al. Molecular epidemiology of Mycobacterium africanum in Ghana. BMC Infect Dis 2016;16:385. doi:10.1186/s12879-016-1725-6.

[23] Kluge Angela. A sociolinguistic survey of the Gbe language communities of Benin and Togo: Gbe language family overview | SIL International. SIL Electron Surv Reports 2011. https://www.sil.org/resources/publications/entry/41567 (accessed May 21, 2019).

[24] Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, et al. First Molecular Epidemiology Study of Mycobacterium tuberculosis in Burkina Faso. J Clin Microbiol 2007;45:921–7. doi:10.1128/JCM.01918-06.

[25] Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, et al. The Guinea-Bissau family of Mycobacterium tuberculosis complex revisited. PLoS One 2011;6:e18601. doi:10.1371/journal.pone.0018601.

[26] Niobe-eyangoh SN, Kuaban C, Cunin P, Thonnon J, Sola C, Rastogi N, et al. Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon. J Cinical Microbiol 2003;41:2547–53. doi:10.1128/JCM.41.6.2547.

[27] Koro FK, Simo YK, Piam FF, Noeske J, Gutierrez C, Kuaban C, et al. Population dynamics of tuberculous bacilli in cameroon as assessed by spoligotyping. J Clin Microbiol 2013;51:299–302. doi:10.1128/JCM.01196-12.

[28] Lawson L, Zhang J, Gomgnimbou MK, Abdurrahman ST, Le Moullec S, Mohamed F, et al. A molecular epidemiological and genetic diversity study of tuberculosis in Ibadan, Nnewi and Abuja, Nigeria. PLoS One 2012;7:e38409. doi:10.1371/journal.pone.0038409.

[29] Thumamo BP, Asuquo AE, Abia-Bassey LN, Lawson L, Hill V, Zozio T, et al. Molecular epidemiology and genetic diversity of Mycobacterium tuberculosis complex in the Cross River State, Nigeria. Infect Genet Evol 2012;12:671–7. doi:10.1016/j.meegid.2011.08.011.

[30] de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M and AR. Differences between TB cases infected with M. africanum, West-African type 2, relative to Euro-American M. tuberculosis- an update. FEMS Immunol Med Microbiol 2010;58:102–5. doi:10.1038/ng.840.Principles.

[31] Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, Darko Otchere I, et al. Reduced transmission of Mycobacterium africanum compared to Mycobacterium tuberculosis in urban West Africa. Int J Infect Dis 2018;73:30–42. doi:10.1016/j.ijid.2018.05.014.

[32] de Jong BC, Hill PC, Brookes RH, Otu JK, Peterson KL, Small PM, et al. Mycobacterium africanum: a new opportunistic pathogen in HIV infection? AIDS 2005;19:1714–5. doi:10.1097/01.aids.0000185991.54595.41.

[33] Jong BC de, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by M. tuberculosis lineage in The Gambia. J Infect Dis 2008;198:1037. doi:10.1086/591504.

[34] WHO. Guidelines for surveillance of drug resistance in tuberculosis 5 th Edition. 2015.

[35] TREATMENT OF TUBERCULOSIS Guidelines for treatment of drug-susceptible tuberculosis and patient care. 2017.

[36] WHO. THE SHORTER MDR-TB REGIMEN FEATURES OF THE SHORTER MDR-TB REGIMEN REGIMEN COMPOSITION 4-6 Km-Mfx-Pto-Cfz-Z-H high-dose-E / 5 Mfx-Cfz-Z-E. 2016.

[37] WHO. WHO consolidated guidelines on drug-resistant tuberculosis treatment. 2019.

[38] GLI. Mycobacteriology Laboratory Manual (Global Laboratory Initiative advancing TB diagnosis: A publication of the Global Laboratory Initiative a Working Group of the Stop TB Partnership). 2014.

[39] Cepheid. Xpert MTB/RIF package insert. REF GXMTB/RIF-US-10. 2015.

[40] WHO | Next-generation Xpert® MTB/RIF Ultra assay recommended by WHO. WHO 2017.

[41] Xpert ultra and MTB/RIF same price. n.d.

[42] Kang HK, Jeong B-H, Lee H, Park HY, Jeon K, Huh HJ, et al. Clinical significance of smear positivity for acid-fast bacilli after ≥5 months of treatment in patients with drug-susceptible pulmonary tuberculosis 2016. doi:10.1097/MD.0000000000004540.

[43] Al-Moamary MS, Black W, Bessuille E, Elwood RK, Vedal S. The Significance of the Persistent Presence of Acid-fast Bacilli in Sputum Smears in Pulmonary Tuberculosis. Chest 1999;116:726–31. doi:10.1378/chest.116.3.726.

[44] Boyles TH, Hughes J, Cox V, Burton R, Meintjes G, Mendelson M. False-positive Xpert® MTB/RIF assays in previously treated patients: need for caution in interpreting results. Int J Tuberc Lung Dis 2014;18:876–8. doi:10.5588/ijtld.13.0853.

[45] Rigouts L, Nolasco O, De Rijk P, Nduwamahoro E, Van Deun A, Ramsay A, et al. Newly developed primers for comprehensive amplification of the rpoB gene and detection of rifampin resistance in Mycobacterium tuberculosis. J Clin Microbiol 2007;45:252–4. doi:10.1128/JCM.01489-06.

[46] Perkins M, Find C. Xpert MTB/RIF Automated molecular detection of TB and MDR screening in peripheral laboratories. n.d.

[47] Comas I. Genomic epidemiology of tuberculosis. Adv. Exp. Med. Biol., vol. 1019, 2017. doi:10.1007/978-3-319-64371-7_4.

[48] Castets M. [Mycobacterium africanum (author's transl)]. Med Trop (Mars) 1979;39:145–8.

[49] Siddiqi SH, Rüsch-Gerdes S. For BACTEC™ MGIT 960™ TB System (Also applicable for Manual MGIT) Mycobacteria Growth Indicator Tube (MGIT) Culture and Drug Susceptibility Demonstration Projects. 2006.

[50] Rieder HL, Van A, Kai D, Kam M, Kim SJ, Chonde TM, et al. Priorities for Tuberculosis Bacteriology Services in Low-Income Countries Second edition 2007 International Union Against Tuberculosis and Lung Disease. 2007.

[51] Yin X, Zheng L, Lin L, Hu Y, Zheng F, Hu Y, et al. Commercial MPT64-based tests for rapid identification of Mycobacterium tuberculosis complex : A meta-analysis. J Infect 2013;67:369–77. doi:10.1016/j.jinf.2013.06.009.

[52] Hopprich R, Shephard L, Taing B, Kralj S, Smith A, Lumb R. Evaluation of (SD) MPT64 antigen rapid test, for fast and accurate identification of Mycobacterium tuberculosis complex. Pathol - J RCPA 2012;44.

[53] Ofori-Anyinam B, Kanuteh F, Agbla SC, Adetifa I, Okoi C, Dolganov G, et al. Impact of the Mycobaterium africanum West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased Sensitivity of Rapid Identification Tests in The Gambia. PLoS Negl Trop Dis 2016;10:1–12. doi:10.1371/journal.pntd.0004801.

[54] Gagneux S, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. vol. 7. 2007.

[55] Diagnostics Standard Inc. SD BIOLINE TB Ag MPT 64 Rapid: one step identification of Mycobacterium tuberculosis complex by RAPID test. 2018.

[56] Hain Lifescience. GenoType Mycobacterium CM. 2014. doi:IFU-299-22.

[57] WHO. Technical manual for drug susceptibility testing of medicines used in the treatment of tuberculosis. 2018.

[58] Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14. doi:10.1016/S0305-4403(02)00239-X.

[59] Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One 2009;4:e7815. doi:10.1371/journal.pone.0007815.

[60] Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis † Downloaded from. J Clin Microbiol 2006;44:4498–510. doi:10.1128/JCM.01392-06.

[61] Thierry D. S6110, an IS-like element of Mycobacterium tuberculosis complex. vol. 18. n.d.

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

[63] Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R. High Resolution Discrimination of Clinical Mycobacterium tuberculosis Complex Strains Based

on Single Nucleotide Polymorphisms. PLoS One 2012;7:39855. doi:10.1371/journal.pone.0039855.

[64] Meehan CJ, Moris P, Kohl TA, Pečerska J, Akter S, Merker M, et al. The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology. EBioMedicine 2018;37:410–6. doi:10.1016/j.ebiom.2018.10.013.

[65] Wyllie DH, Davidson JA, Grace Smith E, Rathod P, Crook DW, Peto TEA, et al. A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying Mycobacterium tuberculosis Transmission: A Prospective Observational Cohort Study. EBioMedicine 2018;34:122–30. doi:10.1016/j.ebiom.2018.07.019.

[66] De Jong BC, Antonio M, Awine T, Ogungbemi K, De Jong YP, Gagneux S, et al. Use of spoligotyping and large sequence polymorphisms to study the population structure of the Mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in the gambia??? J Clin Microbiol 2009;47:994–1001. doi:10.1128/JCM.01216-08.

[67] Stucki D, Gagneux S. Single nucleotide polymorphisms in Mycobacterium tuberculosis and the need for a curated database. Tuberculosis (Edinb) 2013;93:30–9. doi:10.1016/j.tube.2012.11.002.

[68] Viana-Niero C, Gutierrez C, Sola C, Filliol I, Boulahbal F, Vincent VR, et al. Genetic Diversity of Mycobacterium africanum Clinical Isolates Based on IS6110-Restriction Fragment Length Polymorphism Analysis, Spoligotyping, and Variable Number of Tandem DNA Repeats Downloaded from. J Clin Microbiol 2001;39:57–65. doi:10.1128/JCM.39.1.57-65.2001.

[69] Jonsson J, Hoffner S, Berggren I, Bruchfeld J, Ghebremichael S. Comparison between RFLP and MIRU-VNTR Genotyping of Mycobacterium tuberculosis Strains Isolated in Stockholm. PLoS One 2014;9:95159. doi:10.1371/journal.pone.0095159.

[70] Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. Nat Rev Microbiol 2019. doi:10.1038/s41579-019-0214-5.

[71] Rozo-Anaya JC, Ribón W. Molecular tools for Mycobacterium tuberculosis genotyping. vol. 12. 2010.

[72] Bauer J, Andersen ÅB, Kremer K, Mio¨rner H, Mio¨rner M. Usefulness of Spoligotyping To Discriminate IS6110 Low-Copy-Number Mycobacterium tuberculosis Complex Strains Cultured in Denmark. vol. 37. 1999.

[73] Cafrune P, Possuelo L. Prospective study applying spoligotyping directly to DNA from sputum samples of patients suspected of having tuberculosis. Can J …

2009;900:895–900. doi:10.1139/W09-033.

[74] Bidovec-Stojkovič U, Seme K, ˇ Olnir-Dovč Z, Supply M. Prospective Genotyping of Mycobacterium tuberculosis from Fresh Clinical Samples. PLoS One 2014;9:109547. doi:10.1371/journal.pone.0109547.

[75] Kargarpour Kamakoli M, Hadifar S, Khanipour S, Farmanfarmaei G, Fateh A, Siadat SD, et al. Comparison of MIRU-VNTR genotyping between old and fresh clinical samples in tuberculosis. Infect Dis (Auckl) 2019:1–9. doi:10.1080/23744235.2019.1638963.

[76] Gutacker MM, Smoot JC, Migliaccio CAL, Ricklefs SM, Hua S, Cousins D V, et al. Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 2002;162:1533–43.

[77] Cohen T, van Helden PD, Wilson D, Colijn C, McLaughlin MM, Abubakar I, et al. Mixed-strain Mycobacterium tuberculosis infections and the implications for tuberculosis treatment and control. Clin Microbiol Rev 2012;25:708–19. doi:10.1128/CMR.00021-12.

[78] Kamakoli MK, Khanipour S, Hadifar S, Ghajavand H, Farmanfarmaei G, Fateh A, et al. Challenge in direct Spoligotyping of Mycobacterium tuberculosis: a problematic issue in the region with high prevalence of polyclonal infections. BMC Res Notes 2018;11:486. doi:10.1186/s13104-018-3579-z.

[79] Van Rie A, Victor TC, Richardson M, Johnson R, Van Der Spuy GD, Murray EJ, et al. Reinfection and Mixed Infection Cause Changing Mycobacterium tuberculosis Drug-Resistance Patterns n.d. doi:10.1164/rccm.200503-449OC.

[80] CRYPTIC consortium. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med 2018;379:1403–15. doi:10.1056/NEJMoa1800474.

[81] Ezewudo M, Borens A, Chiner-Oms Á, Miotto P, Chindelevitch L, Starks AM, et al. Integrating standardized whole genome sequence analysis with a global Mycobacterium tuberculosis antibiotic resistance knowledgebase 2018. doi:10.1038/s41598-018-33731-1.

[82] Bifani P, Moghazeh S, Shopsin B, Driscoll J, Ravikovitch A, Kreiswirth BN. Molecular characterization of Mycobacterium tuberculosis H37Rv/Ra variants: distinguishing the mycobacterial laboratory strain. J Clin Microbiol 2000;38:3200–4.

[83] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998;393:537–44. doi:10.1038/31159.

[84] Camus J-C, Pryor MJ, Me C, Cole ST. Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. vol. 24. 2019.

[85] Periwal V, Patowary A, Karuthedath Vellarikkal S, Gupta A, Singh M, Mittal A, et al. Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of Mycobacterium tuberculosis Pangenome 2015. doi:10.1371/journal.pone.0122979.

[86] Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, et al. Comparing Genomes within the Species Mycobacterium tuberculosis 2001. doi:10.1101/gr166401.

[87] Affolabi D, Anyo G, Faïhun F, Sanoussi N, Shamputa IC, Rigouts L, et al. First molecular epidemiological study of tuberculosis in Benin. Int J Tuberc Lung Dis 2009;13.

[88] Affolabi D, Faïhun F, Sanoussi N, Anyo G, Shamputa IC, Rigouts L, et al. Possible outbreak of streptomycin-resistant Mycobacterium tuberculosis Beijing in Benin. Emerg Infect Dis 2009;15. doi:10.3201/eid1507.080697.

[89] Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, et al. The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: Implications for in vivo growth. Mol Microbiol 2005;56:163–74. doi:10.1111/j.1365-2958.2005.04524.x.

[90] WHO. WHO treatment guidelines for isoniazid-resistant tuberculosis. 2018. doi:WHO/CDS/TB/2018.7.

[91] Werngren J, Alm E, Mansjö M. Non-pncA Gene-Mutated but Pyrazinamide-Resistant Mycobacterium tuberculosis: Why Is That? 2017. doi:10.1128/JCM.02532-16.

PART 1

Technical advances

Chapter 2

Performance of OMNIgene.SPUTUM (DNA Genotek) and cetylpyridinium chloride for sputum storage prior to mycobacterial culture

Dissou Affolabi[1,2 #], **N'Dira Sanoussi**[2,3], Adelaide, Sossou[1], Tom Nys[3], Ousman Bodi[2], Marius Esse[2], Sabine Houeto[2], Faridath Massou[1,2], Bouke C. de Jong[3], Leen Rigouts[3]

[1]Faculté des Sciences de la Santé, Université d'Abomey-Calavi, Cotonou, Bénin; [2]Laboratoire de Référence des Mycobactéries, Cotonou, Bénin, [3]Institute of Tropical Medicine, Antwerp, Belgium

J Med Microbiol. 2018 May 2. doi: 10.1099/jmm.0.000745

https://www.ncbi.nlm.nih.gov/pubmed/29717969

Conceived and designed the research: DA, **NS**, BCdJ, LR

Performed the experiments: AS, **NS**, TN, OB

Data analysis: DA, ME, AS, **NS**

Supervision: DA, BCdJ, LR

Writing original draft: DA

Writing, review and editing: DA, **NS**, AS, TN, OB, ME, SH, FM, BCdJ, LR

Contributed reagents, materials, analysis tools: DA, BCdJ, LR

Abstract

Purpose. The aim was to assess the performance of both cetylpyridinium chloride (CPC) and OMNIgene.SPUTUM (OMNI) reagents for maintenance of *Mycobacterium tuberculosis* viability in sputum prior to recovery by culture.

Methodology. Using 312 sputa, we evaluated the performance of the two reagents using culture on Löwenstein-Jensen medium after sputum storage in CPC or OMNI for up to 28 days. In addition, the viability of *M. tuberculosis* isolates stored in both reagents was assessed.

Results. Contamination rates were not statistically different between freshly processed samples and those stored in CPC, whereas the contamination rate for OMNI was significantly lower than that of fresh sputa (p=0.026 for 8 days and p=0.002 for 28 days of storage). Culture positivity for fresh sputa (81.7 %) was similar to that of samples stored in CPC regardless of the storage time (89.8 % for CPC-8, and 73.0 % for CPC-28). For OMNI-preserved samples, culture positivity was similar after 8 days of storage (84.2 %), yet decreased significantly after 28 days (42.7 %; p<0.0001) compared to fresh sputa, CPC-8, CPC-28 and OMNI-8. There was a significant loss of viability of H37Rv strain when stored in OMNI at room temperature beyond 8 days compared to CPC, but storage at 37°C decreased recovery from both CPC- and OMNI-stored suspensions.

Conclusion. Culture from sputum stored for 8 days at room temperature in OMNI or CPC gave comparable culture positivity rates to culture from fresh sputum, but at 28 days of storage, the performance of OMNI decreased significantly compared to CPC.

Keywords: Mycobacterium, OMNIgene.SPUTUM, cetylpyridinium chloride, culture

Introduction

Among available tests, culture is the reference method for both laboratory confirmation and follow-up of tuberculosis (TB). Although rarely performed for this purpose for all presumptive TB cases in developing countries, culture is still the test of choice for follow-up of drug-resistant TB treatment even in low resource and TB endemic countries. It is also a prerequisite for phenotypic drug susceptibility testing to first- and second-line anti-TB drugs [1].

However, culture requires expensive and sophisticated infrastructure and highly qualified personnel; thus in low resource countries, only few laboratories are able to perform this test [2]. A sample referral system from TB clinics to the culture laboratory is therefore essential in most endemic countries. Depending on the size of the country, geographical and logistical constraints, samples can take several weeks to reach the culture laboratory. Such a delay can lead to the increased likelihood of overgrowth of other organisms and false negative results [3 - 5].

For long-term storage of sputum samples prior to culture, cetylpyridinium chloride (CPC) has been widely used with good results, but CPC-stored samples may require an additional decontamination step before inoculation to reach acceptably low culture contamination rates [6 - 8]. CPC is also used to ship cultured isolates, to lower the risk of contamination. Recently, a new storage reagent has been developed by DNA Genotek (Ottawa, Canada): OMNIgene.SPUTUM (OMNI) [9]. OMNI is a non-toxic and highly stable reagent that liquefies and decontaminates sputum samples at the point-of-collection or in the laboratory. Like CPC, this reagent allows for specimen storage at room temperature; yet has the advantage that no further decontamination is necessary, thus simplifying its use and enhancing its suitability for laboratories

in developing countries. Furthermore, OMNI is said to preserve samples for up to 8 days before culture, but few data are available on samples preserved for more than 8 days [10 - 12].

In this study, we aimed to evaluate the ability to recover *Mycobacterium* spp. by culture on Löwenstein Jensen (LJ) medium after sputum storage in CPC or OMNI for up to 28 days, compared to culture of fresh sputum. In addition, the viability of *Mycobacterium tuberculosis* isolates stored in CPC or OMNI was assessed using the H37Rv reference strain.

Materials and methods

Sample collection. Sputa were collected from microscopy-positive new TB cases attending the National Hospital for Tuberculosis and Pulmonary Diseases in Cotonou, Benin, from June to October 2016. For each patient, three sputa were collected in two days: first spot - early morning - second spot. After microscopic examination of sputa, an equal volume of CPC or OMNI was added to one of the three samples (approximately 4 mL). Mixtures of a sample with 1 % CPC-2 % NaCl (S-CPC) or with OMNI (S-OMNI) were stored for 8 or 28 days at room temperature, which was recorded daily. The third sample was not stored but decontaminated directly within 24 hours of collection (fresh sputum). In order to ensure a homogeneous distribution of sputa between different methods, the same type of sputum (first spot, early morning or second spot) was used for a given method during two weeks before changing to another method. Then, every two weeks, the type of sputum for a given method (sequence for assignment) changed and so on until the sample size was reached (Table 1).

Table 1. Sequential assignment of sputa per method

	Sequence for assignment to a given transport/decontamination method		
	1	**2**	**3**
Fresh sputum	First spot	Early morning	Second spot
CPC	Second spot	Second spot	First spot
OMNI	Early morning	First spot	Early morning

CPC : cetylpyridinium chloride; OMNI: OMNIgene.SPUTUM

Processing of fresh sputa. Fresh sputa were decontaminated within 24 hours of collection with the Petroff method as previously described [8]. Briefly, samples were added to an equal volume of 4 % NaOH, vortexed and then mixed by shaking at room temperature for 15 min, followed by neutralization with 1N HCl, and centrifugation at 3000 x g for 20 min at 4°C.

Processing of S-CPC. After 8 days of storage, samples in CPC were vortexed and split into two equal parts: one part was processed at day 8 (CPC-8) and the second stored again until day 28 before processing (CPC-28). For processing of S-CPC, the mixture was centrifuged at 3000 x g for 20 min without refrigeration, the supernatant (CPC) was discarded and the pellet suspended in a volume of phosphate buffered saline equal to the processed sputum volume. The resulting suspension was processed in the same way as described above for

fresh sputa, yet with only 5 min instead of 15 min NaOH incubation and without refrigeration of the centrifuge [8].

Processing of S-OMNI. After 8 days of storage, samples in OMNI were vortexed and split into two equal parts: one part was processed at day 8 (OMNI-8) and the second part stored again until day 28 and processed (OMNI-28). For processing of S-OMNI, the mixture was centrifuged at 3000 x g for 20 min at 4°C [10].

Culture of decontaminated samples. After centrifugation, pellets were dissolved in 2 ml phosphate buffered saline for fresh sputa or 1 ml for S-CPC or S-OMNI to prevent over-dilution of OMNI or CPC specimens compared to fresh sputa, as S-CPC and S-OMNI were each split into two. Each suspension was inoculated onto three LJ tubes (two tubes supplemented with 0.75 % glycerol and one with 0.5 % pyruvate). Media were incubated at 37°C and read weekly until 56 days. Results from positive cultures were graded as previously described from scanty to 3+ [6].

The overview of sputum specimen processing is summarized in Figure 1.

Transport/decontamination method on strains. At the Institute of Tropical Medicine (ITM), Antwerp, Belgium, *M. tuberculosis* H37Rv bacterial suspensions were prepared in triplicate from freshly grown bacteria on LJ medium at a concentration of McFarland #1 in sterile distilled water. Each suspension was divided into three parts, and equal volume of OMNI, CPC or sterile distilled water were added, and split into two aliquots each after careful mixing. The first aliquot was stored at room temperature and the second at 37°C. From each aliquot, 2.5 ml was processed on day 0, day 8, day 28 and day 56. OMNI samples were centrifuged for 20 min at 3500 x g, the pellets resuspended in 800 µl phosphate buffered saline and inoculated on LJ medium

and in supplemented MGIT (Mycobacterial Growth Indicator Tubes, Becton Dickinson, USA) for further incubation at 37°C (standard incubator or MGIT 960 respectively). Reading of LJ media was performed weekly until 56 days (Figure 2).

Data analysis. Contamination as well as positivity rates for processed sputa were calculated for each method. Chi-square and Fisher exact tests were used to compare proportions, with a p-value < 0.05 considered significant.

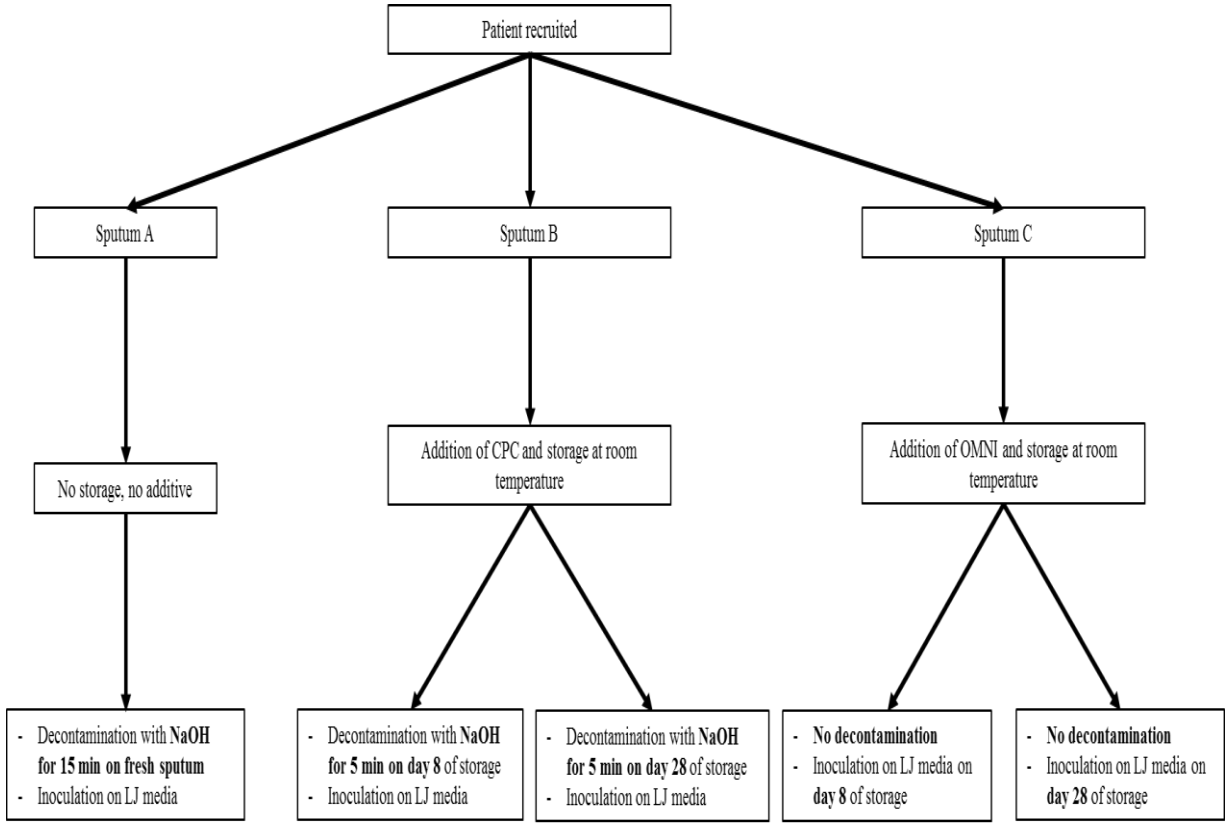

Figure 1. Overview of sputum specimen processing

CPC: cetylpyridiniumm chloride; **LJ:** Löwenstein-Jensen; **OMNI:** OMNIgene sputum**. A, B and C:** types of sputa depending on the sequence described in table 1.

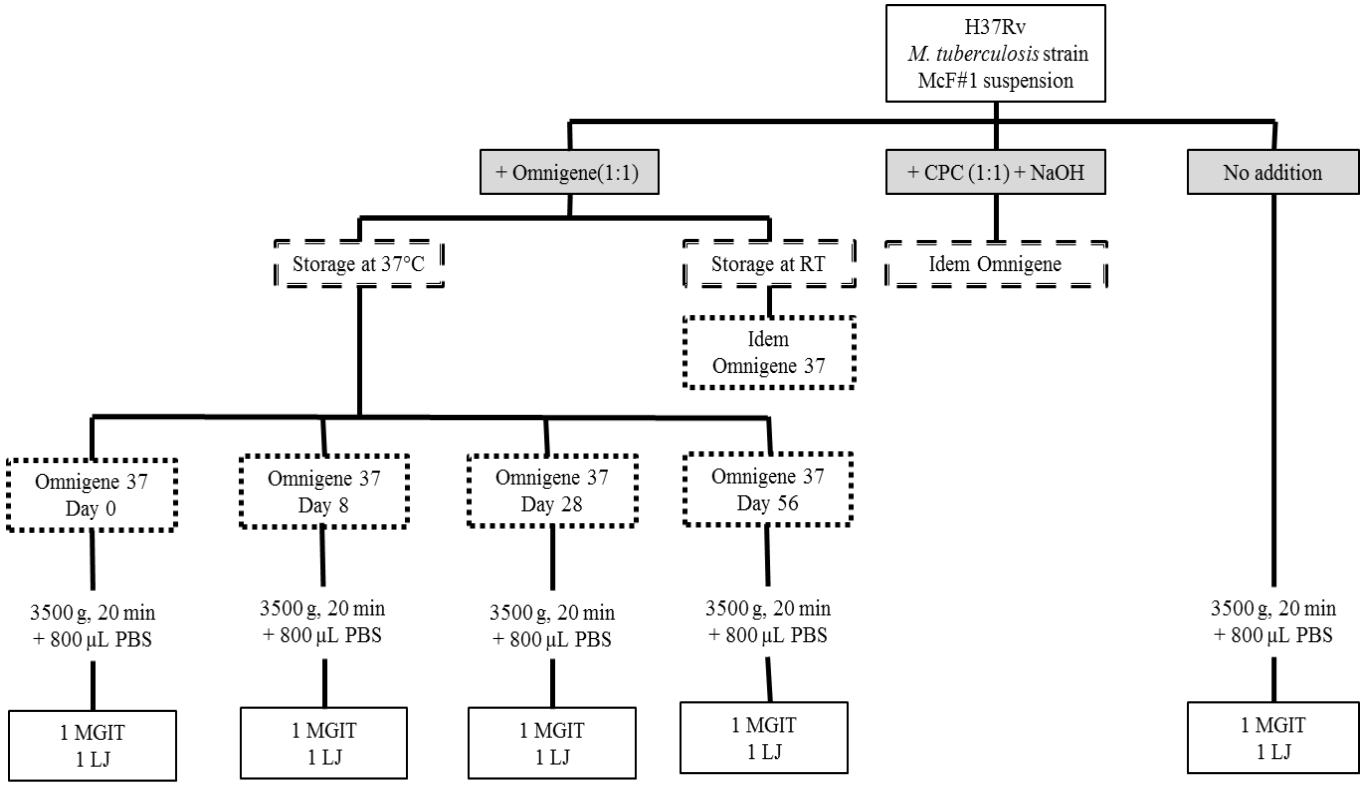

Figure 2. Schematic overview of processing of H37Rv bacterial suspensions

CPC = cetylpyridinium chloride; LJ: Löwenstein Jensen ; McF: McFarland; MGIT : Mycobacterial Growth Indicator Tube ; OMNI = OMNIgene.SPUTUM; PBS: Phosphate Buffered Saline; RT = room temperature; SDW = sterile distilled water; 37 = 37°C

Results

During the study period in Cotonou, the recorded room temperature ranged from 25.5°C to 29.4ºC with a mean of 27.8ºC. In total, 312 sputa were collected from 104 patients. These samples were equally distributed among methods regarding sputum sampling type (p = 0.999) and bacterial load (p=0.425) as shown in Supplementary Tables 1 and 2.

Regarding culture contamination, there was no statistical difference between freshly processed samples and those stored in CPC (8 or 28 days), whereas the contamination rate for OMNI was statistically lower than that of fresh sputa (p=0.026 for 8 days and p=0.002 for 28 days of storage) (Table 2).

Culture positivity for fresh sputa reached 81.7 %, which is similar to samples stored in CPC regardless of the storage time (89.8 % for CPC-8, and 73.0 % for CPC-28). Direct comparison between 8 days and 28 days of storage showed a significant decrease for CPC-28 (p=0.002). Regarding OMNI, culture positivity was similar after 8-days of storage (84.2 %), yet decreased significantly after 28 days (42.7 %; p<0.0001 compared to fresh sputa, CPC-8, CPC-28 and OMNI-8). (Table 3).

Regarding the number of colonies observed on LJ tubes, we found that culture on fresh sputa yielded more highly positive cultures (++/ +++) (65.8 %) than those obtained with CPC 8 (47.7 %), CPC 28 (26.0 %), OMNI 8 (43.5 %) and OMNI 28 (6.8 %). Furthermore, the proportion of highly positive cultures with CPC 8 was similar to that with OMNI 8, but colonies were fewer for CPC 28 and OMNI 28. At day 28, CPC treated sputa grew significantly more “highly positive” than OMNI treated sputa (Table 4).

Table 2. Contaminated cultures by sample type

	Culture		Total n (%)	p
	Contaminated n (%)	Not contaminated n (%)		
Fresh sputa*	11(10.6)	93 (89.4)	104 (100)	-
CPC 8	6(5.8)	98 (94.2)	104 (100)	0.206
CPC 28	4(3.8)	100 (96.2)	104 (100)	0.060
OMNI 8	3(2.9)	101 (97.1)	104 (100)	**0.026**
OMNI 28	1(1)	103 (99)	104 (100)	**0.002**
CPC 8*	6(5.8)	98 (94.2)	104 (100)	-
CPC 28	4(3.8)	100 (96.2)	104 (100)	0.516
OMNI 8	3(2.9)	101 (97.1)	104 (100)	0.308
OMNI 28	1(1)	103 (99)	104 (100)	0.054
CPC 28*	4(3.8)	100 (96.2)	104 (100)	-
OMNI 8	3(2.9)	101 (97.1)	104 (100)	0.722
OMNI 28	1(1)	103 (99)	104 (100)	0.368
OMNI 8*	3(2.9)	101 (97.1)	104 (100)	-
OMNI 28	1(1)	103 (99)	104 (100)	0.371

Samples were processed within 24 hours of collection without any storage reagent (fresh sputa) or after 8 days' storage in CPC (CPC 8), OMNIgene.SPUTUM (OMNI 8), 28 days' storage in CPC (CPC 28) or OMNIgene.SPUTUM (OMNI 28). * Method used as reference method for comparison with methods just below.

Table 3. Culture positivity for non-contaminated cultures

	Culture		Total n (%)	p
	Positive n (%)	Negative n (%)		
Fresh sputa*	76 (81.7)	17 (18.3)	93 (100.0)	-
CPC 8	88 (89.8)	10 (10.2)	98 (100.0)	0.109
CPC 28	73 (73.0)	27 (27.0)	100 (100.0)	0.149
OMNI 8	85 (84.2)	16 (15.8)	101(100.0)	0.651
OMNI 28	44 (42.7)	59 (57.3)	103 (100.0)	**<0.0001**
CPC 8*	88 (89.8)	10 (10.2)	98 (100.0)	-
CPC 28	73 (73.0)	27 (27.0)	100 (100.0)	**0.002**
OMNI 8	85 (84.2)	16 (15.8)	101(100.0)	0.238
OMNI 28	44 (42.7)	59 (57.3)	103 (100.0)	**<0.0001**
CPC 28*	73 (73.0)	27 (27.0)	100 (100.0)	-
OMNI 8	85 (84.2)	16 (15.8)	101(100.0)	0.053
OMNI 28	44 (42.7)	59 (57.3)	103 (100.0)	**<0.0001**
OMNI 8*	85 (84.2)	16 (15.8)	101(100.0)	-
OMNI 28	44 (42.7)	59 (57.3)	103 (100.0)	**<0.0001**

Samples were processed within 24 h of collection without any storage reagent (fresh sputa) or after 8 days' storage in CPC (CPC 8), OMNIgene.SPUTUM (OMNI 8), 28 days' storage in CPC (CPC 28) or OMNIgene.SPUTUM (OMNI 28).

** Method used as reference method for comparison with methods just below.*

Table 4. Culture positivity grade

	Culture positivity scale		Total	p
	scanty/ + n (%)	++/ +++ n (%)		
Fresh sputa *	26 (34.2)	50 (65.8)	76 (100.0)	-
CPC 8	46 (52.3)	42 (47.7)	88 (100.0)	**0.010**
CPC 28	54 (74.0)	19 (26.0)	73 (100.0)	**<0.001**
OMNI 8	48 (56.5)	37 (43.5)	85 (100.0)	**0.002**
OMNI 28	41 (93.2)	3 (6.8)	44 (100.0)	**<0.001**
CPC 8*	46 (52.3)	42 (47.7)	88 (100.0)	-
CCP 28	54 (74.0)	19 (26.0)	73 (100.0)	**0.004**
OMNI 8	48 (56.5)	37 (43.5)	85 (100.0)	0.579
OMNI 28	41 (93.2)	3 (6.8)	45 (100.0)	**<0.001**
CPC 28*	54 (74.0)	19 (26.0)	73 (100.0)	-
OMNI 8	48 (56.5)	37 (43.5)	85 (100.0)	0.579
OMNI 28	41 (93.2)	3 (6.8)	46 (100.0)	**<0.001**
OMNI 8*	48 (56.5)	37 (43.5)	85 (100.0)	-
OMNI 28	41 (93.2)	3 (6.8)	47 (100.0)	**<0.001**

Samples were processed within 24 h of collection without any storage reagent (fresh sputa) or after 8 days' storage in CPC (CPC 8), OMNIgene.SPUTUM (OMNI 8), 28 days' storage in CPC (CPC 28) or OMNIgene.SPUTUM (OMNI 28).

** Method used as reference method for comparison with methods just below.*

Assessment of storage conditions directly on *M. tuberculosis* H37Rv strain in suspension revealed a significant loss of viability when stored in OMNI at room temperature beyond 8 days, compared to baseline growth without storage as well as compared to CPC at room temperature, and this both in MGIT and on LJ medium (Figure 3-5). Storage at 37°C affected both CPC- and OMNI-stored suspensions, with absence of growth beyond 8 days for both culture systems. CPC suspensions stored at room temperature still yielded growth after 28 and 56 days of storage, albeit with reduced positivity as expressed by lower mean number of colonies on LJ, and both lower mean growth units and higher time to positivity in MGIT (Figure 3-5).

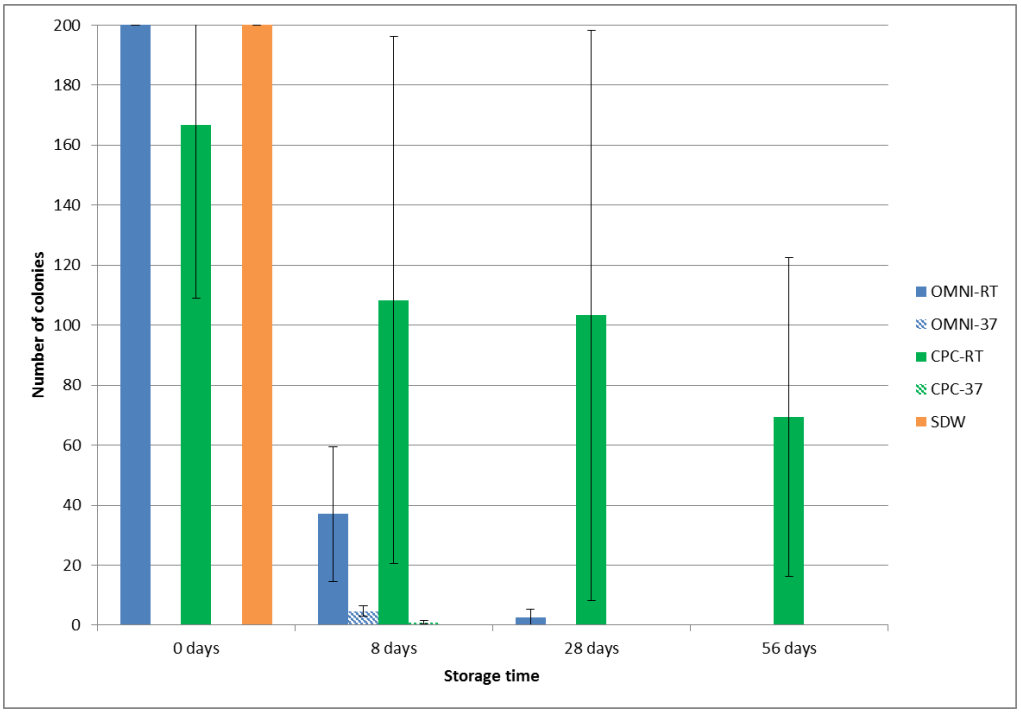

Figure 3. Observed number of colonies on Löwenstein-Jensen medium after 8 weeks of incubation for H37Rv bacterial suspensions stored in OMNI, CPC or sterile distilled water prior to inoculation

RT = room temperature; SDW = sterile distilled water; CPC = cetylpyridinium chloride; OMNI = OMNIgene.SPUTUM; 37 = 37°C. Confluent growth was recorded as 200 colonies.

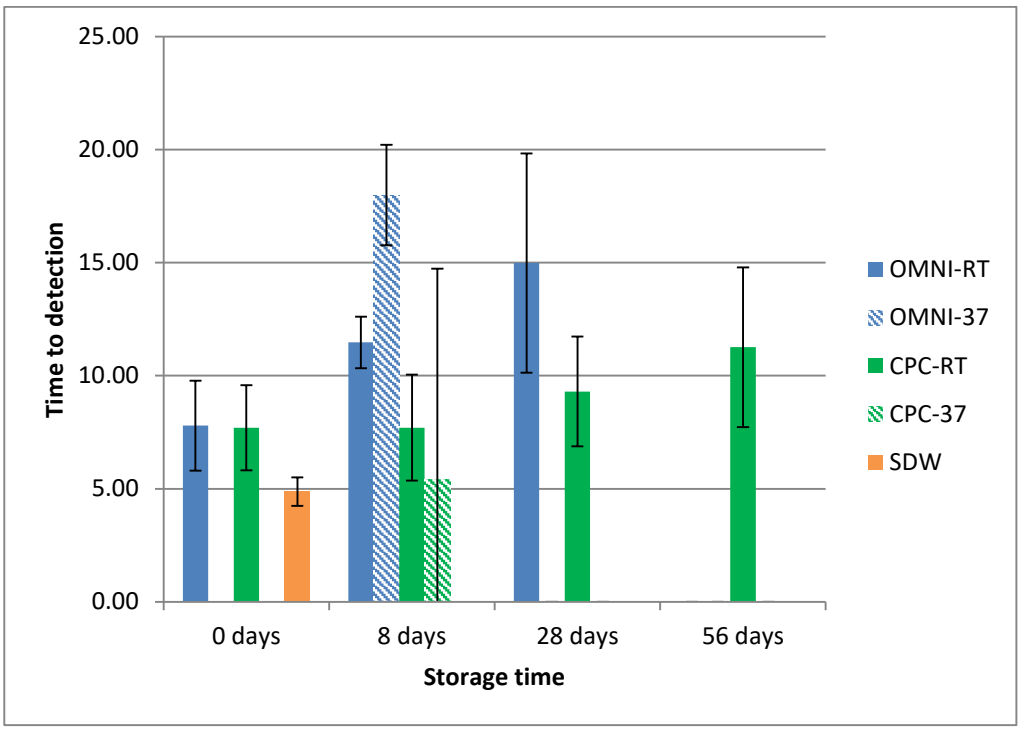

Figure 4. Mean time to detection in MGIT960 for H37Rv bacterial suspensions stored in OMNI, CPC or sterile distilled water prior to inoculation

RT = room temperature; SDW = sterile distilled water; CPC = cetylpyridinium chloride; OMNI = OMNIgene.SPUTUM; 37 = 37°C.

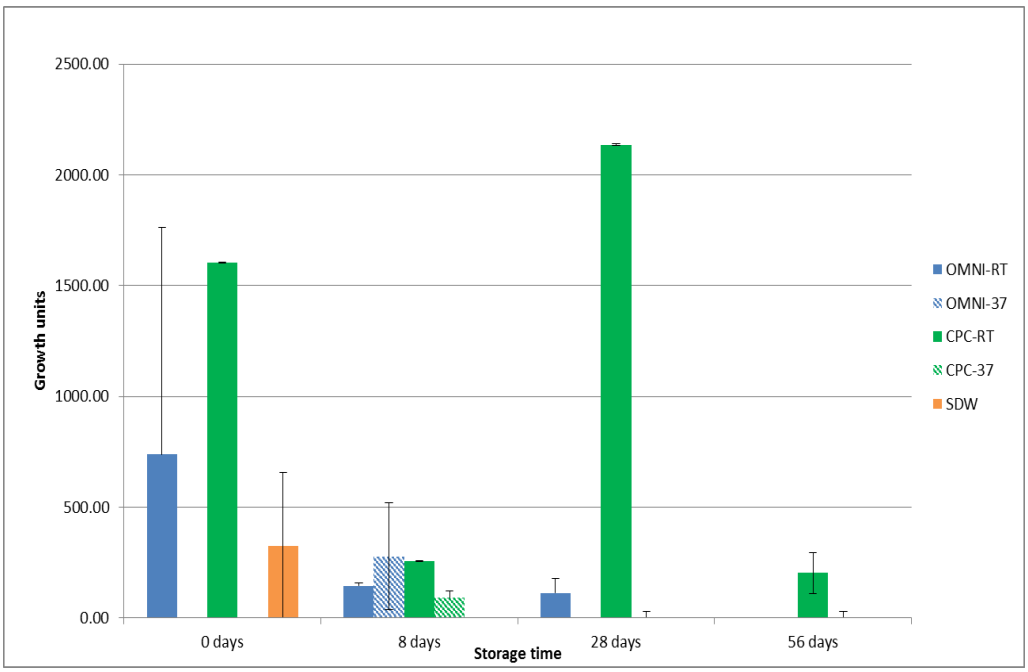

Figure 5. Mean growth units observed in MGIT960 for H37Rv bacterial suspensions stored in OMNI, CPC or sterile distilled water prior to inoculation

RT = room temperature; SDW = sterile distilled water; CPC = cetylpyridinium chloride; OMNI = OMNIgene.SPUTUM; 37 = 37°C.

Discussion

In this study, OMNI performed equally well as CPC at day 8 of sputum storage relative to freshly processed samples, both in terms of culture contamination and positivity rates, which matches with previous findings [10; 11]. In these previous studies, NALC-NaOH decontamination was used as comparative method rather than the Petroff method (NaOH) used in the present study, and one of the previous studies included culture in MGIT, with similar results, showing that OMNI was suitable for preserving sputa up to 8 days [11].

Recently, Tagliani et al. compared culture in MGIT of fresh sputa with MGIT culture of sput astored in OMNI from 4 to 22 days [12]. The study showed a significant delay in time to culture positivity for samples stored for more than two weeks in OMNI. In addition, excluding culture contaminated samples, 9 were positive by culture from fresh sputa but negative after storage in OMNI, while only one was positive after storage in OMNI but negative with fresh sputa, suggesting that culture recovery after long-term storage in OMNI may not be optimal for MGIT culture either. Like Tagliani et al, we assessed the performances of OMNI beyond 8 days of storage, exceeding the manufacturer's recommendations [10], as sputum shipments regularly exceed 8 days in settings where access to mycobacterial laboratories is limited. Our study found OMNI unsuitable for sputum storage for 28 days, as culture sensitivity declined to under 50 %, unlike storage in CPC (Figure 3).

Many low resource and TB endemic countries face important geographical and logistical constraints leading to prolonged delays from sample collection to culture, even up to several weeks or months as mentioned by Abeygunawardena et al in Sri Lanka [3], Das et al in India [4] and Narasimooloo et al in South Africa [5]. In those conditions, OMNI will not be helpful for

culture-based diagnostics, unlike CPC that retained culture sensitivity around 73/104 (70.2 %) of microscopy positive sputa at 28 days compared to 44/104 (42.3 %) sputa for OMNI.

CPC has been used extensively for long-term storage of sputa prior to culture on solid media with relatively good results [3 - 5]. Nevertheless, its misuse in operational conditions can lead to decrease of its performance, since CPC crystalizes at 4°C. Hence, neither the solution itself, nor CPC-specimens should be stored in the refrigerator, and a refrigerated centrifuge should be avoided for CPC-preserved specimens [8]. In addition, CPC has shipping restrictions, and its use involves additional costs and effort of decontamination. Lastly, previous studies reported interference of CPC with MGIT media making CPC-stored samples not suitable for culture in MGIT, although our findings on stored isolates suggest that MGIT culture after CPC storage yielded stronger growth than after OMNI storage [10].

The main advantage of OMNI compared to CPC is that OMNI does not require an additional decontamination step prior to culture. Thus, it can be easily used in less experienced-laboratories. This step is indeed critical and if not well controlled, it can lead to both false negative culture results and increased contamination rates. In contrast, a mild decontamination is recommended in case of long-term CPC storage [13]. In our study, decontamination with 4 % NaOH for 5 min for CPC-preserved samples, gave relatively good results, albeit with decreased positivity rate for CPC-28.

The number of observed colonies in primary culture was significantly higher for fresh sputa compared to those stored in CPC or OMNI, indicating a loss of viability for any delay in specimen processing, especially in microscopy scanty positive specimens (data in Supplementary Tables 3, 4 and 5). This observation

is supported by our experiment on H37Rv bacterial suspensions with decreased growth units and observed colonies in MGIT and on LJ respectively, in addition to increased time to positivity.

In addition to primary isolation for the diagnosis of TB, culture is recommended also in endemic countries for follow-up of patients treated for multidrug resistant TB strains, in order to detect treatment failure. Delay in processing culture will cause delay in diagnosing treatment failure and thus promote spread of resistant strains within the community. Therefore, efforts should be made by National Tuberculosis Programs in low resource and endemic countries to reduce time from collection to culture as much as possible by setting up an affordable, reliable, sustainable and rapid sample referral system to reduce time from sample collection to culture, despite geographical and logistical constraints in these countries.

This study was performed in a laboratory where air conditioning was functional, and the room temperature recorded during the study period ranged from 25.5°C to 29.4ºC. Performance of both OMNI and CPC on sputum preservation in settings with higher temperature might be different from what was found in this study, as suggested by our findings on preservation of bacterial colonies. Furthermore, the laboratory is a reference laboratory with recommendations from the manufacturer on sample processing strictly followed. In an operational context with less qualified personnel the execution of procedures and resulting performance of storage reagents may be different. Mycobacteria isolated during this study were not identified; thus we were not able to assess performance of both storage reagents depending on mycobacteria species.

In conclusion, culture from sputum stored for 8 days in OMNI or CPC gave comparable culture positivity rates as culture from fresh sputum. Storage

beyond 8 days decreased the performance of OMNI significantly, while CPC showed a lesser decrease in sensitivity at 28 days of storage.

Funding information. Leen Rigouts and Bouke de Jong were supported by European Research Council-INTERRUPTB starting grant (N° 311725); Dissou Affolabi and N'Dira Sanoussi by DGD, Belgium. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication

Conflicts of interest. DNA Genotek kindly provided part of the OMNIgene SPUTUM reagents used in this study but had no role in the design, execution, or analysis of data of the study.

Ethical statement. This study was a sub-study within the 'BeniDiT study' approved by the National ethical committee of Benin (N°030 of 15th/12/2015).

References

1. **Gilpin C, Korobitsyn A, Weyer K.** Current tools available for the diagnosis of drug-resistant tuberculosis. *Ther Adv Infect Dis* 2016; 3:145-151.
2. **World Health Organization.** *Laboratory services in tuberculosis control. Part III*. Culture. Geneva, Switzerland: WHO; 1998. http://apps.who.int/iris/bitstream/10665/65942/3/WHO_TB_98.258_(part3).pdf
3. **Abeygunawardena SC, Sharath BN, Van den Bergh R, Naik B, Pallewatte N, Masaima MNN.** Management of previously treated tuberculosis patients in Kalutara district, Sri Lanka: how are we faring? *Public Health Action* 2014; 4:105-109.
4. **Das D, Dwibedi B, Kar SK.** Transportation of sputum samples in cetylpyridinium chloride for drug resistance studies from remote areas of Odisha, India. *J Infect Dev Ctries* 2014; 8:1059-1062.
5. **Narasimooloo R, Ross A.** Delay in commencing treatment for MDR TB at a specialised TB treatment centre in KwaZulu-Natal. *S Afr Med J* 2012;102:360-362
6. **Smithwick RW, Stratigos CB, David HL.** Use of cetylpyridinium chloride and sodium chloride for the decontamination of sputum specimens that are transported to the laboratory for the isolation of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1975; 1: 411-413.
7. **Selvakumar N, Kumar V, Gopi PG, Venkataramu KV, Datta M, Paramasivan CN, Prabhakar R.** Isolation of tubercle bacilli from sputum samples of patients in the field studies by the cetylpyridinium chloride-sodium chloride & sodium hydroxide methods. *Ind J Med Res* 1995; 102: 149-151.
8. **Rieder HL, Van Deun A, Kam KM, Kim SJ, Chonde TM, Trébucq A, Urbanczik R.** *Priorities for tuberculosis bacteriology services in low-income countries*. Second edition. Paris: International Union Against Tuberculosis and Lung Disease; 2007. https://tbrieder.org/publications/books_english/red_book.pdf
9. **Kelly-Cirino CD, Curry PS, Marola JL, Helstrom NK, Salfinger M.** Novel multi-day sputum transport reagent works with routine tuberculosis tests and eliminates need for cold chain: preliminary study of compatibility with the Xpert MTB/RIF assay. *Diagn Microbiol Infect Dis* 2016; 86: 273–276.
10. **Maharjan B, Shrestha B, Weirich A, Stewart A, Kelly-Cirino CD.** A novel sputum transport solution eliminates cold chain and supports routine tuberculosis testing in Nepal. *J Epidemiol Glob Health* 2016; 6:257-265.
11. **Kelly-Cirino CD, Musisi E, Byanyima P, Kaswabuli S, Andama A, Sessolo A, Sanyu I, Zawedde J, Curry PS, Huang L.** Investigation of OMNIgene.SPUTUM performance in delayed tuberculosis testing by smear, culture, and Xpert MTB/RIF assays in Uganda. *J Epidemiol Glob Health* 2017; 7:103-109.

12. **Tagliani E, Alagna R, Tafaj S, Hafizi H, Cirillo DM.** Evaluation of *Mycobacterium tuberculosis* viability in OMNIgene-SPUTUM reagent upon multi-day transport at ambient temperature. *BMC Infect Dis* 2017; 17:663.
13. **Sankar MM, Kumar P, Munawwar A, Singh J, Parashar D, Singh S.** Recovery of *Mycobacterium tuberculosis* from sputum treated with cetyl pyridinium chloride. *J Clin Microbiol* 2009; 47:4189-4190.

Abbreviations.

CPC	Cetylpyridinium chloride
DGD	Directorate-general Development Cooperation and Humanitarian Aid
ITM	Institute of Tropical Medicine
LJ	Löwenstein Jensen
MGIT	Mycobacterial Growth Indicator Tubes
NALC	N-Acetyl-L-Cysteine
OMNI	OMNIgene.SPUTUM
S-CPC	Mixtures of a sample with cetylpyridinium chloride
S-OMNI	Mixtures of a sample with OMNIgene.SPUTUM
TB	Tuberculosis

Supplementary Table 1. Distribution of sputum sampling type per method

	Type of sputum			
	First Spot n *(%)*	Early morning n *(%)*	Second Spot n *(%)*	p
Fresh sputa	35 *(33.7)*	34 *(32.6)*	35 *(33.7)*	
CPC	34 *(32.6)*	35 *(33.7)*	35*(33.7)*	0.999
OMNI	35 *(33.7)*	35 *(33.7)*	34 *(32.6)*	
Total	**104** ***(100.0)***	**104** ***(100.0)***	**104** ***(100.0)***	

CPC : cetylpyridinium chloride; OMNI: OMNIgene.SPUTUM

Supplementary Table 2. Stratification of smear microscopy results per method

	Microscopy results				Total(%)	p
	scanty	+	++	+++		
	n *(%)*	n *(%)*	n *(%)*	n *(%)*		
Fresh sputa	24 *(23.1)*	29 *(27.9)*	44 *(42.3)*	7 *(6.7)*	104 *(100.0)*	0.425
CPC	20 *(19.2)*	28 *(26.9)*	42 *(40.4)*	14 *(13.5)*	104 *(100.0)*	
OMNI	20 *(19.2*)	20 *(19.2)*	52 *(50.0)*	12 *(11.5)*	104 *(100.0)*	

CPC : cetylpyridinium chloride; OMNI: OMNIgene.SPUTUM

Chapter 3

Storage of sputum in cetylpyridinium chloride, OMNIgeneSPUTUM and ethanol is compatible with molecular tuberculosis diagnostic testing

C. N'Dira Sanoussi[1,2,3#], Bouke C. de Jong[2], Dissou Affolabi[1], Conor J. Meehan[2], Mathieu Odoun[1], Leen Rigouts[2,3]

[1]Laboratoire de Référence des Mycobactéries, Cotonou, Benin

[2]Unit of Mycobacteriology, Institute of Tropical Medicine, Antwerp, Belgium; [3]Department of Biomedical Sciences, University of Antwerp

Corresponding author: C. N'Dira Sanoussi

J Clin Microbiol. 2019 Jun 25;57(7). pii: e00275-19. doi: 10.1128/JCM.00275-19 ;

https://www.ncbi.nlm.nih.gov/pubmed/31092592

Conceived and designed the research: **CNS**, BCdJ, DA, LR

Performed the experiments: **CNS**, MO

Data analysis: **CNS**, BCdJ, LR

Supervision: BCdJ, DA, LR

Writing original draft: **CNS**

Writing, review and editing: **CNS**, BCdJ, DA, CJM, MO, LR

Contributed reagents, materials, analysis tools: BCdJ, DA, LR

ABSTRACT

We compared cetylpyridinium chloride (CPC), ethanol (ETOH) and OMNIgene.SPUTUM (OMNI) for 28-day storage of sputum at ambient temperature before molecular tuberculosis diagnostics.

Three sputa were collected from 133 smear-positive TB patients (399 sputa). Each patient's sputum was stored with either CPC, ETOH or OMNI for 28 days at ambient temperature, with subsequent *rpoB* amplification, targeting a short fragment (81-bp, GeneXpert MTB/RIF (Xpert)) or a long fragment (1764-bp, in-house nested-PCR). For 36 patients, Xpert was also performed at baseline on all 108 fresh sputa.

After the 28-day storage (D28), Xpert positivity did not significantly differ between storage methods. In contrast, higher positivity for *rpoB* nested-PCR was obtained with OMNI (125, 94%), relative to ETOH (114, 85.7%, p=0.001). Smears with scanty AFB had lower *rpoB*-PCR positivity for ETOH storage (10, 41.7%) relative to CPC (16, 66.7%; difference=25%, 95%CI (3.5-46.5), p=0.031) or OMNI (16, 69.6%; difference=26.1%, 95%CI (3.8-48.4), p=0.031) with no difference between CPC and OMNI. Post-storage Ct-values significantly decreased compared to pre-storage with ETOH (D=-1.1, 95%CI (-1.6 to -0.6), p=0.0001), but not with CPC (p=0.915) or OMNI (p=0.33). For one patient's ETOH- and CPC-stored specimens with a Ct<10, Xpert tested rifampicin false-resistant at D28, which was resolved by repeating Xpert on a 1/100 diluted specimen.

In conclusion, 28-day storage of sputum in OMNI, CPC or ETOH at ambient temperature does not impact short fragment PCR (Xpert), including for low smear-grades. However, for long fragment PCR, ETOH yielded a lower PCR

positivity for low smear-grades, while the performance of OMNI and CPC was excellent for all smear-grades.

Keywords: sputum, storage, cetylpyridinium-chloride, ethanol, OMNIgene.Sputum, molecular tests, short/long-fragment PCR, Xpert, AFB-scanty, isolate

INTRODUCTION

Molecular tuberculosis (TB) diagnostics are more sensitive than microscopy and faster than culture. In addition to detection of TB, molecular analyses can provide the drug resistance profile of the affecting *Mycobacterium tuberculosis* complex (MTBC) strain(s), and strain typing for diversity studies. In low-income countries, molecular analyses may not be available in peripheral laboratories, and specimens need to be shipped to remote intermediate or central reference laboratories. Logistic constraints can delay the time between specimen collection and analysis, potentially negatively impacting the results.

Sample deterioration could be overcome by the use of sputum preservation reagents, such as ethanol (ETOH)(1) and cetylpyridinium chloride (CPC), already widely used, or a proprietary buffer such as OMNIgene.SPUTUM (OMNI, DNA Genotek, Ottawa, Canada) (2, 3). ETOH preservation allows for molecular analyses only, as it inactivates TB bacilli(1), while CPC and OMNI preservation can precede both molecular analyses and culture (2, 3). ETOH preserved samples no longer constitute a biosafety risk, hence not requiring a specific containment laboratory level, and can be shipped with minimal restrictions. ETOH is less expensive than CPC, which in turn is cheaper than OMNI.

No limit has been specified for specimen preservation in ETOH and CPC, while for OMNI storage for a maximum of 8 days is specified by the manufacturer. The time between specimen collection and analysis can vary widely depending on shipping logistics and distance, the urgency of the molecular analysis to be performed and whether specimens are assembled for batch processing. A recent systematic review highlighted the lack of evidence on the performance of commercial storage reagents to preserve sputum samples, especially paucibacillary samples, which is of great diagnostic importance(4, 5).

Molecular methods for the detection of rifampicin (RIF), for example, vary by platform and target length, ranging from the 81 bp rifampicin resistance determining region (RRDR) target of the *rpoB* gene covered in GeneXpert MTB/RIF (Xpert, Cepheid, Sunnyvale, California, USA)(6) to the 1764 bp target of the *rpoB* gene that also covers resistance conferring mutations positioned outside the RRDR (7).

We compared ETOH, CPC and OMNI for one-month (28-day) storage of smear-positive sputa at ambient temperature, with subsequent automated Xpert and conventional gel-based *rpoB* nested-PCR, including paucibacillary specimens.

MATERIALS AND METHODS

Ethics considerations

This evaluation was embedded in the BeniDiT study that was approved by the national ethics committee of Benin, and the ethics committees of the Institute of Tropical Medicine (ITM) of Antwerp, and the University of Antwerp in Belgium. The study was registered on ClinicalTrials.gov (NCT02744469). Before inclusion, all the study participants provided written informed consent. Participants' specimens were pseudonymized before laboratory analyses.

Study design, participants and specimens

The main study design and specimen workflow is summarized in Fig. 1. In total 399 sputa were collected from 133 consecutive new smear-positive TB patients who were prospectively recruited among patients registered for TB screening in the National University Hospital for TB and Lung Diseases in Cotonou, Benin (Centre National Universitaire Hospitalier de Pneumo-phtisiologie, CNHU-PPC)

during a three-month period. The sputa were collected before patients started anti-TB treatment. Laboratory analyses were performed in the Supranational Reference Laboratory for mycobacteria in Cotonou, Benin (Laboratoire de Référence des Mycobactéries, LRM) located in the CNHU-PPC.

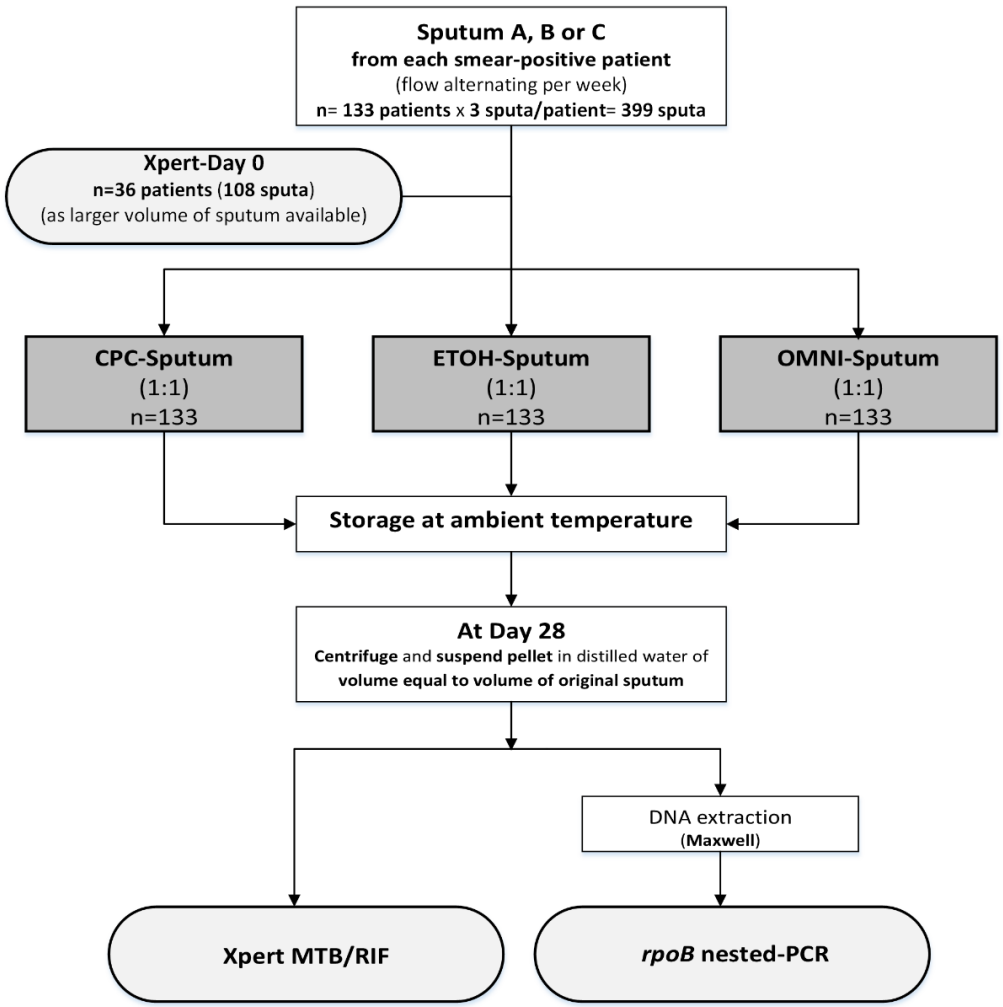

Fig. 1. Flow diagram for specimens and methods

In total, three pre-treatment sputa (spot day 1, morning and spot day 2) were collected from each participating patient. TB confirmation, per national guidelines (8), was based on smear microscopy on unprocessed sputa using the WHO/IUATLD AFB grading scale for fluorescence microscopy (9). Each sputum sample was stored with either an equal volume of ETOH (final concentration 50%), CPC1% (final 0.5%) or OMNI (final 50%)[1] [3]. The 1% CPC solution consists of dissolving the necessary amount of CPC in a 2% sodium chloride solution. The sequence of the sputum samples (spot day 1, morning and spot day 2) assigned to a specific storage reagent alternated weekly. The sputum mixtures were vortexed and stored at ambient temperature for 28 days, after which they were centrifuged at 3800 g (3) for 20 minutes at 28°C, the supernatant was discarded and the pellet was re-suspended in sterile distilled water (of equal volume to the initial unprocessed sputum). At least one milliliter of the suspension was used for Xpert and 200µL extracted for *rpoB* nested-PCR. For 36 patients (108 sputa) who produced large volume samples (at least 2.5 ml per sputum container), the Xpert analysis was also done at baseline before storage of the sputum in the storage reagent. The ambient temperature of the storage-room was recorded every day, no temperature control measures such as air-conditioning were in place.

Prior to the main study, at the Institute of Tropical Medicine in Antwerp (ITM), the reference *M. tuberculosis* isolate *H37Rv* (dilution 10^{-3} McFarland#1) was also stored in ETOH, CPC and OMNI, in triplicate, for 37 days at ambient temperature and 37°C and tested with Xpert to check whether the detectable bacillary load would be affected. As the ITM laboratory is air-conditioned (20-25°C) and Belgium has a moderate climate, the standard set temperature of the incubator (37°C) was used to mimic higher temperature conditions encountered in many high endemic countries. To mimic delayed inter-

continental shipments – potentially facing custom clearance issues – storage time was extended to 37 days.

Molecular analyses

Xpert MTB/RIF is an automated system with DNA extraction and amplification integrated for the real-time qPCR of five probes (A, B, C, D, E) covering the RRDR of the *rpoB* gene. Resistance to RIF is detected through absence or delayed reaction of one or more probes (6). The assay was used as per the manufacturer instructions (6). For the fresh sputum (when baseline Xpert was done before storage) or for the suspension after storage, one volume of the specimen (suspension/sputum) was mixed with 2 volumes of the Xpert 'sample reagent' as recommended by the manufacturer for sputum, to assure the comparability of fresh sputum and respective suspension after storage. Post-storage Ct-values were compared across storage methods, and to Ct-values of fresh sputum when available. To ensure valid comparisons, specimens and Xpert' sample reagent' volumes were measured with graduated pipets.

In-house *rpoB* nested-PCR. Sediments from CPC- and OMNI-stored specimens were inactivated at 95°C for 10 min (10). After a digestion of 200 µL of the suspension with proteinase K, DNA was extracted using the semi-automated Maxwell 16 tissue DNA purification kit AS1030 with a Maxwell machine (model 4.9, Promega), and eluted in 300 µL Maxwell elution buffer as previously described (11, 12). *rpoB* nested-PCR was performed as previously described (7). Positive- and negative extraction as well as amplification controls were included. The positive extraction control was a sediment from a sputum known as MTBC PCR positive. The positive amplification controls consisted of DNA

extracts from *H37Rv* suspensions (10^{-2}, 10^{-3}, 10^{-4} of McFarland#1) and DNA extracts (10^{-2}, 10^{-4} µg/µL) from a MTBC PCR positive sputum. The negative extraction control consisted of molecular grade water processed as other sputa specimens in each DNA extraction run. The negative amplification control consisted of molecular grade water (both 1st PCR run to nested run & nested run separately) processed as a specimen DNA extract along with DNA extracts from sputum. All specimens (ETOH, CPC, OMNI) from a patient were processed in the same extraction run or PCR run. For the conventional PCR, specimens were coded so that the person reading the gel was blinded to the patient's identification (ID) or storage method assigned. To check for possible PCR inhibition, the *rpoB* nested-PCR was repeated for specimens with a negative result using one tenth dilution of the original DNA extract.

Statistical analyses

Statistical data analysis was performed using Stata 12.0 (StataCorp, USA). We used the Mc Nemar Chi-squared test to compare paired categorical data across ETOH, CPC, OMNI; and paired t-test to compare Ct-values. The difference (Diff) in proportion or mean difference was calculated with a 95% confidence interval (CI) and p-value, which was considered significant below 0.05.

RESULTS

The sputa stored in the respective reagents had similar AFB-positivity grades (Table S1). In the room where sputum mixtures with storage reagent were stored, the temperature ranged from 26 to 31.9°C, with an average of 28.5°C and a median of 28.4°C. All included negative controls remained negative by nested-PCR.

After the 28 days of storage (D28), the presence of MTBC was confirmed by Xpert in 132 (99.2%) patients. Xpert yielded a positive result for most sputa (98.5%, 99.2%, 99.2% respectively for ETOH, CPC, OMNI), and the Xpert positivity across storage methods did not significantly differ (Table 1). After stratification by AFB smear-grading, all but three AFB-scanty sputa (2 stored with ETOH, 1 stored with CPC) and a 1+ AFB-positive sputum (stored with OMNI) were Xpert positive (Table S2). However, also for scanty specimens, positivity was not significantly different across storage methods for Xpert (Table S2).

In contrast, higher positivity for *rpoB* nested-PCR was obtained with OMNI storage (94%), which differed significantly between OMNI and ETOH (p=0.001, Table 1). Most negative nested PCRs occurred among AFB-scanty sputa (14/24 for ETOH, 9/25 for CPC, 7/24 for OMNI, Table 2). Lower *rpoB* PCR positivity among AFB-scanty sputa was found with ETOH storage relative to either CPC (D=25%, 95%CI (3.5-46.5), p=0.031) or OMNI (D=26.1%, 95%CI (3.8-48.4), p=0.031), whereas CPC and OMNI storage yields were similar (p=1, Table S3, Table 2). For 1+ AFB-positive sputa, the differences between ETOH, CPC and OMNI were not significant (Table S3).

In total, sputa from two patients tested RIF-resistant in Xpert, with probes A or E absent. Hence, we compared Ct-values of probe B and D to assess Ct-value variability across storage methods. The mean Ct-value for probe D after 28-day storage was 17 for ETOH, 17.9 for CPC and 18.1 for OMNI. The probe D Ct-values after ETOH storage were significantly lower (more MTBC DNA amplified) than those observed either after CPC storage (CPC vs ETOH, p<0.00001), or after OMNI storage (OMNI vs ETOH, p<0.00001), but did not significantly differ between OMNI vs CPC storage (p=0.51, Table 3). Likewise, for probe B, Ct-

values were significantly lower for ETOH than CPC or OMNI storage, but no difference was observed for OMNI vs CPC (data not shown).

Table 1. Positivity of PCR by storage method

	Xpert MTB/RIF at D28				*rpoB* PCR at D28			
	Positive		**Diff (95%CI)**		**Positive**		**Diff (95%CI)**	
Storage	**n (%)**	**Total**		**p**	**n(%)**	**Total**		**p**
ETOH-28	131[#] (98.5)	133	-	-	114[‡] (85.7)	133	-	-
CPC-28	132 (99.2)	133	0.8 (-1.5 to 3)[A]	1	121[‡] (91)	133	5.3 (-1.1 to 11.7) [A]	0.119
OMNI-28	132[†] (99.2)	133	0.8 (-1.5 to 3) [B]	1	125[‡] (**94**)	133	**8.3 (2.8 to 13.7)** [B]	**0.001**
OMNI-28 vs CPC-28			0 (-0.8 to 0.8) [C]	1			3 (-1.9 to 7.9) [C]	0.289
Total	395 (99)	399			360 (90.2)	399		

p-values were calculated using Exact Mc Nemar test, comparing (A) CPC-28 versus ETOH-28, (B) OMNI-28 versus ETHO-28 and (C) OMNI-28 versus CPC-28. Diff: difference

Includes 2 errors that became positive (high) after testing 1/10 dilution of the sediment.

† Includes 1 invalid that became positive (very low) after testing 1/10 dilution of the sediment.

‡ For all PCR –negative specimens, including within a patient set of specimens with discrepant PCR result, the *rpoB* PCR was repeated on a 1/10 dilution of the DNA extract. All remained negative, but one stored in OMNI became positive.

Table 2. Positivity of *rpoB* PCR by sputum AFB-grade

	***rpoB* PCR**					
	ETOH		**CPC**		**OMNI**	
Sputum AFB-grade‡#	**Positive n (%)**	**Total**	**Positive n (%)**	**Total**	**Positive n(%)**	**Total**
+++pos	13 (100)	13	16 (100)	16	15 (100)	15
++pos	63 (96.9)	65	63 (98.4)	64	68 (100)	68
+pos	28 (90.3)	31	26 (92.9)	28	25 (96.2)	26
Scanty*	10 (41.7)	24	16 (**64**)	25	17 (**70.8**)	24
Total	114	133	121	133	125	133

‡UIATLD/WHO scale for fluorescent microscopy (9)

Original (fresh unprocessed) sputum

*Among smear-scanty specimens, *rpoB* PCR positivity was significantly different for CPC vs ETOH (Mc Nemar exact, p=0.031) and OMNI vs ETOH (p=0.031), but similar for OMNI vs CPC (p=1) (See Table S3 in the supplemental material).

Table 3. Change in Xpert Ct value for probe D after storage (D28) across storage method

Xpert D28_Ct_probeD					
Storage method	**Total (per storage method)**	**Mean Ct (95%CI)**	**$Diff_{Ct}$ (95%CI)**	**$p^{\#}$ (Diff=0)**	**$p^{\#}$ (Ha: Diff>0)**
ETOH-28	131*	**17.0 (16.3 to 17)**	-		
CPC-28	131*	17.9 (17.2 to 18)	**0.9(0.5 to 1.3)**	**<0.00001**	**<0.00001**
ETOH-28	130*	**17.0 (16.4 to 17.7)**	**-**		
OMNI-28	130*	18.1 (17.4 to 18.8)	**1.1(0.6 to 1.6)**	**<0.00001**	**<0.00001**
CPC-28	131*	18 (17.3 to 18.7)	-		
OMNI-28	131*	18.2 (17.5 to 18.8)	0.2(-0.3 to 0.7)	0.51	0.253

#p-values were calculated using paired t-test.
* Only paired sputa (ie sputa from the same patient) with positive Xpert (ie Ct ≠ 0) for the 2 compared storage methods were included.
ETOH-28: Storage in ETOH for 28 days
CPC-28: Storage in CPC for 28 days
OMNI-28: Storage in OMNI for 28 days

As a sensitivity analysis, for 36 patients (108 sputa) who had Xpert also done on fresh sputum, we compared Ct-values for three fresh unprocessed (D0) and three post-storage (D28) sputa across storage methods. No significant difference in D0 Ct-value was observed between the sputa assigned to either ETOH, CPC or OMNI storage (mean D0_Ct_probe D=18.6 for ETOH versus 18.4 for CPC versus 18.1 for OMNI, non-significant for all comparisons), showing that groups of sputa assigned to each of the three storage methods were comparable in terms of fresh bacillary load before mixing with storage reagents. However, overall D28 Ct-values compared across storage methods were significantly lower for ETOH relative to CPC or OMNI, yet not different for OMNI vs CPC (mean D28_Ct_probe D=17.5 for ETOH versus 18.3 for CPC versus 18.8 for OMNI; with $p_{(CPC=ETOH)}$ = 0.015; $p_{(OMNI=ETOH)}$ = 0.005; $p_{(OMNI=CPC)}$ =0.31). This confirmed the findings on all sputa that ETOH preservation, unlike CPC or OMNI, lowers Xpert Ct-values. Despite the fact that the quantification of the bacterial load by sputum AFB-microscopy is less precise than by Xpert Ct-value, ETOH storage also yielded more Xpert “high” bacillary load among AFB weakly-positive (1+ positive and scanty) sputa (25.5% vs 11.3% for CPC and 6% for OMNI, Table S2).

Likewise, a direct comparison of Ct-values from baseline and stored sputum from the same patient ($Ct_{stored(D28)}$-$Ct_{fresh(D0)}$) significantly decreased (more detectable DNA) after storage in ETOH (-1.1, 95%CI (-1.6 to -0.6), p=0.0001 for probe D), whereas it did not differ after storage in CPC (p=0.915, probe D) or OMNI (p=0.33, probe D) (Table 4 for probe D, Table S4 for probe B).

For one patient, all three baseline sputa (D0) were RIF sensitive by Xpert. At D28, ETOH stored specimen turned RIF resistant, while CPC and OMNI stored sputa remained RIF sensitive. After repeat testing of the three D28 specimens

by Xpert, RIF-resistance was confirmed for the ETOH specimen (probe A and E missing, Ct=8.7 for probe D) and also the CPC specimen tested RIF-resistant (only probe A missing, Ct=8.7 for probe D), while the OMNI remained RIF-sensitive (Ct=11.5 for probe D). As the Xpert Ct-values were low, we repeated Xpert on a 1/100 dilution of the D28 suspensions, and all became RIF-sensitive. RIF sensitivity was further confirmed by testing the undiluted suspensions with Xpert-Ultra (Cepheid, Sunnyvale, California, USA) and DNA sequencing of the *rpoB* nested-PCR amplicons (wild type *rpoB*).

Regarding the *H37Rv* bacterial suspensions, the Xpert positivity was the same ("medium") for CPC and OMNI at either ambient temperature or 37°C, yet for ETOH results were temperature dependent, remaining "medium" at ambient temperature but decreasing ("low") at 37°C in two replicates.

Table 4. Change in Ct-value for Xpert probe D from storage at D0 to processing at D28

		Total (per storage method)	Mean $Ct_{probe\ D}$ (95%CI)	Diff Ct_{D28-D0} (95%CI)	p# (Ha: Diff<0)	p# (Diff=0)
ETOH	D0	36	18.6 (17.3 to 19.9)	-		
	D28	36	**17.5 (16.1 to 19)**	**-1.1 (-1.6 to -0.6)**	**0.0001**	**0.0002**
CPC	D0	36	18.4 (17 to 19.8)	-		
	D28	36	18.3 (16.8 to 19.9)	-0.03 (-0.7 to 0.5)	0.457	0.915
OMNI	D0	35*	18.1 (16.8 to 19.4)	-		
	D28	35*	18.3 (17 to 19.7)	0.2 (-0.3 to 0.7)	0.835	0.330

#p-values were calculated using paired t-test.

*One sputum excluded from that Ct difference analysis on OMNI as the AFB-positive sputum (of the 3 AFB-scanty sputa of a patient) randomly assigned to OMNI had a negative D0 Xpert.

D0: day 0 (baseline before storage with a storage reagent); D28: after 28-day storage with a storage reagent

DISCUSSION

ETOH, CPC and OMNI storage for 28 days did not yield significant differences in the proportion that tested Xpert positive, whereas successful amplification of a large *rpoB* target was significantly more likely after OMNI storage (94%) than after ETOH storage (85.7%), without significant difference with CPC (91%). Of concern, one sample tested false RIF resistant after 28-day storage in ETOH and CPC, likely related to the lowered Ct-value after storage, below the range for optimal detection of rifampicin resistance. Thus, for all sputa – whether fresh or stored - in the context of Ct values ≤ 10, a RIF resistant result should be confirmed by dilution (1/100) of the remaining sample-or a new sample if no left-over. Studies similar to this, could be conducted on Xpert Ultra as this becomes the standard Xpert MTB/RIF testing (13).

Other authors comparing shorter storage durations (2-21 days) also found comparable Xpert positivity after storage in CPC (98.9% vs 99.2% in our study) or OMNI (97.9% (14) and 95% (15) vs 99.2% in our study).

The yields of a long *rpoB* fragment from AFB-scanty sputa was significantly higher for OMNI and CPC storage than for ETOH storage. Contrary, Xpert Ct-values after the 28-day storage were significantly lower in ETOH stored sputa than in CPC or OMNI, despite comparable Ct-values in fresh sputa at baseline. Similarly, Asandem et al. found that OMNI yielded less Xpert 'high' bacillary load (after 7 days storage) versus freshly decontaminated sediments (decrease of 6%)(14). The differences observed for ETOH storage, with lower yield for long-fragment PCR yet higher bacterial loads detected in Xpert compared with CPC and OMNI, is unlikely to be explained by a lower compatibility of ETOH with Maxwell DNA extraction, as ETOH rather enhances/facilitates DNA extraction (as also shown by Rabodoarivelo et al(16)) by destroying the protein structure

of the bacterial cell surface (17). This was confirmed by the lowered Ct at D28 for EtOH samples in our study. Furthermore, the storage reagent mixed with the sputum was discarded after the post-storage centrifugation, before the pellet was re-suspended in sterile distilled water, thus limiting interaction between storage and extraction reagents. Also such interaction between ETOH and proteinase K in lysing buffer (pH 7.5) used prior to our modified Maxwell extraction is unlikely (11, 18). In addition, ETOH is a component of the Maxwell extraction cartridge (security datasheet (12)). The more plausible explanation is the fragmentation/degradation of the DNA, as ETOH creates a disordered state in the bacterial DNA by cross-linking the bases so that the DNA strands can no longer properly separate, interfering with replication and transcription (17). Therefore, the initial denaturation step of the PCR process could cause fragmentation of the bacterial DNA, which no longer properly separates, as result of the ETOH effect. This could affect the recovery/amplification of the bacterial DNA, which is more perceivable in a long fragment target (1764 bp in *rpoB*-PCR versus 81 bp in Xpert), especially in paucibacillary samples. Likewise, NaOH induced changes in the bacterial DNA sequence of paucibacillary samples after prolonged storage in Xpert 'sample reagent' risk causing false resistant results (19). Furthermore, the impact of ETOH storage could be temperature dependent, as evidenced by data of the *H37Rv* isolate with decreased positivity at 37°C. This needs to be further investigated as we only tested one strain, albeit in triplicate, and our sputum samples were not exposed to temperatures above 32°C.

Storage of sputum in 50% (used in this study) or 70% ETOH final (used in some laboratories) renders the TB bacilli non-viable after 1 hour of storage ((1); unpublished data from our laboratory (ITM)). PCR results for a 123 bp fragment were similar for both ETOH concentrations after one day of storage(1). The long

fragment *rpoB* PCR after 1-hour storage of 4 triplicates of isolate-spiked sputa yielded 12 positive results in 50% ETOH vs 11 in 70% ETOH; and after 14-days storage 10 vs 11 positives respectively (unpublished data from ITM). In other unpublished data from ITM, analysing 943 sputa after long term (20-218 days) storage in ~70% ETOH at non controlled 'ambient' temperature in Bangladesh, showed a similar positivity in *rpoB* nested-PCR for 3+, 2+ and 1+ AFB-grade sputa (96%, 95%, 90% versus 100%, 96.9%, 90.3% in the current study). In contrast, positivity among AFB-scanty sputa was lower in the present study (41.7%) than in the unpublished data (77%). Van Deun et al found that sputa stored in 70% ETOH for a period ranging from 2 to 7 years yielded an overall positivity of 94% in the long fragment *rpoB* nested-PCR vs 85.7% for 50% ETOH in this study (20).

A strength of this study is the paired design, with comparable bacterial loads by AFB smear-positivity grading and Xpert Ct-values between all groups. As a limitation, ambient temperature varies, potentially affecting the generalizability of our findings to hotter climates, nor did we test long term storage without additives, which may be challenging because of potential overgrowth by other organisms. Smear-negative sputa from TB patients were not included in this study, with likely further decreased yield of *rpoB* amplification in all storage solutions. Moreover, we did not test compatibility of different storage solutions with other DNA extraction methods beyond Maxwell. The use of basic extraction methods consisting of boiling sputum (such as the Chelex method) resulted in good yield after storage of sputum in ETOH (16). We centrifuged and discarded the storage reagent from sputa before testing (as recommended for OMNI(3)), to avoid eventual reagent interaction, which is unlikely for ETOH and unknown for CPC and OMNI. This may not be feasible in Xpert centres without centrifuge. We centrifuged stored

sputa at 3800g, the upper range (3000-3800g) recommended for OMNI by its manufacturer (3), and compatible with other TB analyses (≥3000 g recommended (9)). At 3000 g but not 3800 g we noted that the OMNI pellet tended to slip and be discarded with supernatant after centrifugation.

As summarized in Fig. 2, in conclusion, the overall performance of ETOH, CPC and OMNI for 28-day storage of sputum at ambient temperature is excellent, especially for subsequent Xpert analysis and long-fragment PCR of non-scanty AFB-positive samples. It is advisable to use ETOH for subsequent short-fragment PCR (Xpert) when culture is not needed, and use either CPC or OMNI when culture is needed. For paucibacillary sputa, OMNI or CPC will have better yield of subsequent long-fragment PCR. When culture is needed, the choice between OMNI and CPC, which are equally performant for either subsequent short-fragment PCR (Xpert) or long-fragment (conventional) PCR, should be based on their cost (OMNI is ~57x more expensive than CPC (~1.15 versus ~0.02 US$/ml of sputum))(5), their performance for culture after 28-day storage (CPC followed by a short decontamination, better than OMNI for culture on Lowenstein-Jensen medium (21)). The choice should be also based on the culture medium to be used. There is a reduced/delayed growth in MGIT after OMNI (2, 22, 23). Likewise direct inoculation of CPC-preserved is not compatible with MGIT (24) with a reduced/delayed positivity(25); however washing off the CPC prior to inoculation in MGIT (2, 21, 25) increases the culture positivity rate (25), a procedure not yet common in routine practice. One should also take into account, whether post-storage fluorescence microscopy-based tests will be realized (CPC not compatible with fluorescence microscopy(2)), the constraints regarding import or transport of dangerous

products (CPC), availability of a centrifuge adjustable to ambient temperature (CPC crystalizes at cold temperatures(9)), and the availability of a centrifuge reaching 3800 g (needed for OMNI). CPC and OMNI are likely more stable than ETOH for the storage of specimens/cultures at high ambient temperature (≤37°C) and possibly beyond.

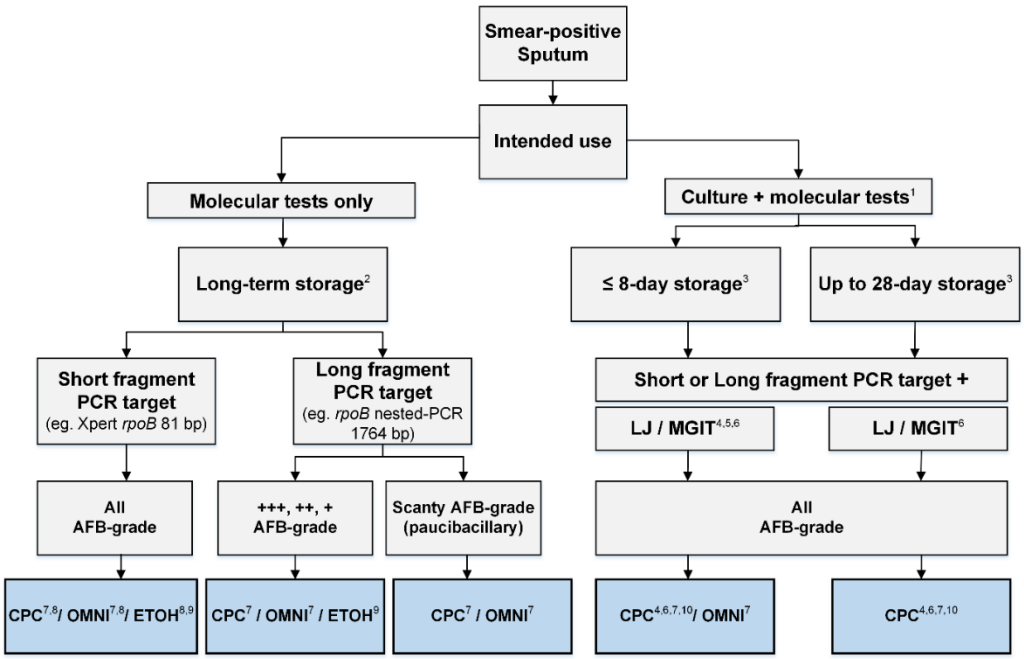

Fig. 2. Algorithm for the choice of the suitable reagent for the storage of sputum for TB diagnostics

[1] Molecular tests on the stored sputum (direct molecular tests)
[2] From up to 28 days (current study) to 2-7 years (20)
[3] Previously showed by Affolabi et *al* (16)
[4] **Extra washing required** before inoculation on culture medium (2, 21, 25)
[5] **Reduced/delayed growth after OMNI storage** (2, 22, 23)
[6] CPC is not compatible with direct MGIT inoculation (24), with reduced/delayed positivity(25); however washing off the CPC prior to inoculation in MGIT (2, 21, 25) increases the culture positivity rate (25).
[7] **Cost OMNI > CPC > ETOH**. **Shipping restriction** for specimens stored with CPC and OMNI: IATA Cat B
[8] When Xpert Ct values < 10, a RIF-resistant result should be confirmed on a dilution (1/100) of the remaining sample, or a new sample if no left over remains.
[9] **Not optimal** with **higher temperature** (≥37°C tested in this study). ETOH has **shipping restrictions**: dangerous good if total volume > 100 ml
[10] Short decontamination required before inoculation on culture medium (21)
LJ: Löwenstein Jensen medium

MGIT: Mycobacterial growth indicator tube (manual or automated)

LJ/MGIT: For LJ or MGIT
CPC/OMNI/ETOH: Use either CPC, OMNI or ETOH
CPC/OMNI: Use either CPC or OMNI
LJ: Lowënstein Jensen medium, **MGIT:** Mycobacterial growth indicator tube

ACKNOWLEDGMENT

We wish to thank Maren Diels, Wim Mulders (Unit of Mycobacteriology of the ITM, Antwerp) for their assistance in performing supplementary analyses to resolve discrepancies; Mourad Gumusboga (Unit of Mycobacteriology of the ITM, Antwerp) for providing unpublished data on 70% ETOH, and Tom Nys who did his bachelor thesis research on the storage of the reference isolate *H37Rv* in various temperatures.

FUNDING

This work was supported by funds from the Directorate General for Development (DGD), Belgium (FA4) to CNS, BCdJ, DA and LR, and the European Research Council-INTERRUPTB starting grant nr.311725 to BCdJ, CJM and LR. The funders had no role in study design, data collection and interpretation or the decision to submit the work for publication.

SUPLLEMENTAL MATERIALS (See after the references section)

REFERENCES

1. Williams DL, Gillis Tp Fau - Dupree WG, Dupree WG. 1995. Ethanol fixation of sputum sediments for DNA-based detection of Mycobacterium tuberculosis. J Clin Microbiol 1995 Jun;33(6):1558-61.
2. GLI. GLI guide to TB specimen referral systems and integrated networks.
3. DNA-Genotek. 2017. OMNIgene.SPUTUM Reagent for optimizing samples.
4. Reeve BWP, McFall SM, Song R, Warren R, Steingart KR, Theron G. 2018. Commercial products to preserve specimens for tuberculosis diagnosis: a systematic review. Int J Tuberc Lung Dis 22:741-753.
5. WHO. 2017. Technical Expert Group Meeting Report: Commercial products for preserving clinical specimens for the diagnosis of tuberculosis.
6. Cepheid. 2009. Xpert MTB/RIF: two hour detection of MTB and resistance to rifampicin.
7. Rigouts L, Nolasco O, de Rijk P, Nduwamahoro E, Van Deun A, Ramsay A, Arevalo J, Portaels F. 2007. Newly developed primers for comprehensive amplification of the rpoB gene and detection of rifampin resistance in Mycobacterium tuberculosis. J Clin Microbiol 45:252-4.
8. WHO. 2015. Implementing tuberculosis diagnostics: policy framework. ISBN: 978 92 4 150861 2, WHO/HTM/TB/2015.11,
9. Rieder H L VA, Kam K M , Kim S J, Chonde T M , Trébucq A , Urbanczik R. Second edition 2007. Priorities for Tuberculosis Bacteriology Services in Low-Income Countries, Second edition
   ed. International Union Against Tuberculosis and Lung Disease (The Union)
10. Bemer-Melchior P HBD. 1999. Inactivation of Mycobacterium tuberculosis for DNA Typing Analysis. J Clin Microbiol July 1999:p. 2350–2351.
11. Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. 2017. Genotypic characterization directly applied to sputum improves the detection of Mycobacterium africanum West African 1, under-represented in positive cultures. PLoS Negl Trop Dis 11:e0005900.
12. Promega. 2018. Maxwell 16 DNA Purification Kits Technical Manual. https://worldwide.promega.com/resources/protocols/technical-manuals/0/maxwell-16-dna-purification-kits-protocol/. Accessed
13. WHO. 2017. Next-generation Xpert MTB/RIF Ultra assay recommended by WHO.
14. Asandem DA, Asante-Poku A, Asare P, Aboagye SY, Stephen OW, Danso E, Klevor PM, Hayibor KM, Yeboah-Manu D. 2018. OMNIgene SPUTUM: A good transport and decontaminating reagent for tuberculosis testing. Int J Mycobacteriol 7:222-227.
15. Maharjan B, Kelly-Cirino CD, Weirich A, Curry PS, Hoffman H, Avsar K, Shrestha B. 2016. Evaluation of OMNIgene(R)*SPUTUM-stabilised sputum for long-term transport and Xpert(R) MTB/RIF testing in Nepal. Int J Tuberc Lung Dis 20:1661-1667.

16. Rabodoarivelo M, Brandao A, Cergole Novella M, C. Bombonatte A, Imperiale B, Rakotosamimanana N, Morcillo N, Rasolofo V, Palomino J, Martin A. 2018. Detection of multidrug-resistant tuberculosis from stored DNA Samples: A multicenter study. International Journal of Mycobacteriology 7:40-44.
17. Yoo JH. 2018. Review of Disinfection and Sterilization - Back to the Basics. Infect Chemother 50:101-109.
18. Tomar R, Dubey VK, Jagannadham MV. 2009. Effect of alkyl alcohols on partially unfolded state of proteinase K: Differential stability of alpha-helix and beta-sheet rich regions of the enzyme. Biochimie 91:951-60.
19. Banada PP, Sivasubramani SK, Blakemore R, Boehme C, Perkins MD, Fennelly K, Alland D. 2010. Containment of Bioaerosol Infection Risk by the Xpert MTB/RIF Assay and Its Applicability to Point-of-Care Settings. Journal of Clinical Microbiology 48:3551.
20. Van Deun A, Aung KJ, Hossain A, de Rijk P, Gumusboga M, Rigouts L, de Jong BC. Disputed rpoB mutations can frequently cause important rifampicin resistance among new tuberculosis patients.
21. Affolabi D, Sanoussi N, Sossou A, Nys T, Bodi O, Esse M, Houeto S, Massou F, de Jong BC, Rigouts L. 2018. Performance of OMNIgene*SPUTUM (DNA Genotek) and cetylpyridinium chloride for sputum storage prior to mycobacterial culture. J Med Microbiol doi:10.1099/jmm.0.000745.
22. Zallet J, Olaru ID, Witt AK, Vock P, Kalsdorf B, Andres S, Hillemann D, Kranzer K. 2018. Evaluation of OMNIgene((R))*SPUTUM reagent for mycobacterial culture. Int J Tuberc Lung Dis 22:945-949.
23. Tagliani E, Alagna R, Tafaj S, Hafizi H, Cirillo DM. 2017. Evaluation of Mycobacterium tuberculosis viability in OMNIgene-SPUTUM reagent upon multi-day transport at ambient temperature. BMC Infect Dis 17:663.
24. FIND. 2006. MGIT procedure manual.
25. Sankar MM, Kumar P, Munawwar A, Singh J, Parashar D, Singh S. 2009. Recovery of Mycobacterium tuberculosis from sputum treated with cetyl pyridinium chloride. Journal of clinical microbiology 47:4189-4190.

Table S1. AFB-grading of sputa by storage method

Sputa# assigned to	AFB-grade (auramine)‡#											
	+++ Pos	Diff (95%CI)	P	++ Pos	Diff (95%CI)	p	+ Pos	Diff (95%CI)	p	Scanty	Diff (95%CI)	p
ETOH	13 (9.8)	-		65 (48.9)	-		31 (23.3)	-		24 (18.0)	-	
CPC	16 (12.0)	2.3(-1 to 5.5)[A]	0.25	64 (48.1)	-0.7(-5.9 to 4.4)[A]	1	28 (21.1)	-2.3(-6.9 to 2.4)[A]	0.453	25 (18.8)	0.8(-1.5 to 3)[A]	1
OMNI	15 (11.3)	1.5(-2.2 to 5.2)[B]	0.625	68 (51.1)	2.3(-2.9 to 7.4)[B]	0.51	26 (19.5)	-3.8(-9.8 to 2.3)[B]	0.267	24 (18.0)	0(-2.8 to 2.8)[B]	1
OMNI vs CPC		-0.7(-5.4 to 3.9)[C]	1		3(-3.2 to 9.2)[C]	0.424		-1.5(-6.9 to 3.9)[C]	0.754		-0.8(-3 to 1.5)[C]	1
Total	44			197			85			73		

‡UIATLD/WHO scale for fluorescent microscopy (1)

Original (fresh unprocessed) sputum; Diff: difference

p-values were calculated using Exact Mc Nemar's Chi2 test, comparing (A) sputa assigned to (to be stored with) CPC versus those assigned to ETOH, (B) those assigned to OMNI versus ETOH and (C) those assigned to OMNI versus CPC

Table S2. Positivity and bacilli load of Xpert by sputum AFB-grade

Sputum AFB-grade‡##	Xpert																	
	ETOH						CPC						OMNI					
	H n(%)	M n(%)	L n(%)	VL n(%)	All pos n(%)	Total n	H n(%)	M n(%)	L n(%)	VL n(%)	All pos n(%)	Total n	H n(%)	M n(%)	L n(%)	VL n(%)	All pos n(%)	Total n
+++pos	13 (100)	0	0	0	**13** (100)	13	15 (93.8)	1	0	0	**16** (100)	16	15 (100)	0	0	0	**15** (100)	15
++pos	51 (**78.5**)	14 (21.5)	0	0	**65** (100)	65	38 (59.4)	26 (40.6)	0	0	**64** (100)	64	40 (58.8)	27 (39.7)	0	1	**68** (100)	68
+pos	11 (**35.5**)	18 (58.1)	2 (6.5)	0	**31** (100)	31	5 (17.9)	21 (75)	2 (7.1)	0	**28** (100)	28	3 (11.5)	19 (73.1)	3 (**11.5**)	0	**25** (96.2)	26
Scanty*	3 (**12.5**)	9 (37.5)	10 (41.7)	0	**22** (91.7)	24	1 (**4**)	9 (36)	14 (**56**)	0	**24** (96)	25	0 (**0**)	12 (**50**)	11 (45.8)	1	**24** (100)	24
Total	78	41	12	0	131	**133**	59	57	16	0	132	**133**	58	58	14	3	132	**133**
pos+ and scanty	14 (**25.5**)	25 (45.5)	12 (21.8)	0	53 (96.4)	55	6 (11.3)	30 (56.6)	16 (30.2)	0	52 (98.1)	63	3 (6)	31 (62)	14 (28)	1 (2)	49 (98)	50

H: high (Ct<16); **M**: medium (Ct 16-22); **L**: low (Ct 22 - 28); **VL**: very low (Ct >28); pos: positive

‡UIATLD/WHO scale for fluorescent microscopy (1); **#** Original (fresh unprocessed) sputum

***Among smear-scanty specimens, Xpert positivity was similar (Mc Nemar exact, p=1 for CPC vs ETOH, p=1 for OMNI vs ETOH, p=1 for OMNI vs CPC

Table S3. Pairwise comparison of storage methods for *rpoB* PCR positivity among lower grade AFB positive sputa (scanty and AFB-positive +)

	CPC-28 vs ETOH-28					OMNI-28 vs ETOH-28					OMNI vs CPC			
AFB scanty														
	ETOH +	**ETOH -**	**Total**	**Diff(95%CI) p***		**ETOH +**	**ETOH -**	**Total**	**Diff(95%CI) p***		**CPC +**	**CPC -**	**Total**	**Diff(95%CI) p***
CPC +	10	**6**	16	**25 (3.5 to 46.5)**	**OMNI +**	10	**6**	16	**26.1 (3.8 to 48.4)**	**OMNI +**	14	**3**	17	4.2 (-18.2 to 26.5)
CPC -	0	8	8	**0.031**	**OMNI -**	0	7	7	**0.031**	**OMNI -**	**2**	5	7	1
Total	10	14	24#		Total	10	13	23#		Total	16	8	24#	
AFB positive +														
	ETOH +	**ETOH -**	**Total**	**Diff(95%CI) p***		**ETOH +**	**ETOH -**	**Total**	**Diff(95%CI) p***		**CPC +**	**CPC -**	**Total**	**Diff(95%CI) p***
CPC +	21	**3**	24	3.8 (-16.8 to 24.5)	**OMNI +**	21	**2**	23	8.7 (7.2 to 24.6)	**OMNI +**	20	**2**	22	9.1 (-7.5 to 25.6)
CPC -	2	0	2	1	**OMNI -**	0	0	0	0.5	**OMNI -**	**0**	0	0	1
Total	23	3	26#		Total	21	2	23#		Total	20	2	22#	

Specimens (from a same patient) and with the same AFB smear grading (scanty or positive+ accordingly) for the (two) storage methods compared were included in the analysis.

*p-values were calculated with the Mc Nemar Exact test. **Diff**: difference

Table S4. Change in Ct-value for Xpert probe B from storage at D0 to processing at D28

		Total (per storage method)	Mean Ct_{probeB} (95%CI)	Diff $Ct_{D28 - D0}$ (95%CI)	p[#]
ETOH	D0	36	19.8 (18.6 to 21)	-	
	D28	36	**18.2 (16.9 to 19.6)**	**-1.5 (-2.1 to -0.9)**	**<0.00001**
CPC	D0	36	19.4 (18.1 to 20.7)	-	
	D28	36	19 (17.4 to 20.5)	-0.4 (-1.1 to 0.2)	0.247
OMNI	D0	35*	19.2 (17.9 to 20.5)	-	
	D28	35*	19.1 (17.8 to 20.4)	-0.1 (-0.7 to 0.4)	0.646

#p-values were calculated using paired t-test.

*****One sputum excluded from that Ct difference analysis on OMNI as the AFB-positive sputum (of the 3 AFB-scanty sputa of a patient) randomly assigned to OMNI had a negative D0 Xpert.

REFERENCES FOR SUPPLEMETAL MATERIALS

1. Rieder H L VA, Kam K M , Kim S J, Chonde T M , Trébucq A , Urbanczik R. Second edition 2007. Priorities for Tuberculosis Bacteriology Services in Low-Income Countries, Second edition
ed. International Union Against Tuberculosis and Lung Disease (The Union)

Chapter 4

Low sensitivity of MPT64 identification test to detect Lineage 5 of the *M. tuberculosis complex*

C. N'Dira Sanoussi [1, 2 #], Bouke C. de Jong [1], Mathieu Odoun [2], Karamatou Arekpa [2,3], Moulikatou Ali Ligali [2], Bodi Ousman [2], Simon Harris [4], Ofori-Anyinam Boatema [5], Dorothy Yeboah-Manu [6], Isaac Darko Otchere [6], Adwoa Asante-Poku [6], Severin Anagonou [2], Sébastien Gagneux [7], Mireia Coscolla [8], Leen Rigouts [1, 9], Dissou Affolabi [2]

[1] Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium; [2] Laboratoire de Référence des Mycobactéries, Cotonou, Benin; [3] Génie de Biologie Humaine, Ecole Polytechnique d'Abomey-Calavi, Université d'Abomey-Calavi; [4]Wellcome Trust Sanger Institute, United Kingdom; [5]Vaccine and Immunity Theme, Medical Research Council (MRC) Unit, Serrekunda, The Gambia; [6]Noguchi Memorial Institute for Medical Research, Legon, Accra, Ghana; [7]Swiss Tropical and Public Health Institute, Basel, Switzerland; [8] University of Valencia, Spain; [9]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium; # Corresponding author: C. N'Dira Sanoussi

J Med Microbiol. 2018 Dec;67(12):1718-1727. doi: 10.1099/jmm.0.000846.

https://www.ncbi.nlm.nih.gov/pubmed/30388066

Conceived and designed the research: **CNS**, BCdJ, LR, DA

Performed the experiments: **CNS**, MO, KA, MA, BO

Data analysis: **CNS**, BCdJ, LR;

Supervision: BCdJ, LR, DA

Writing original draft: **CNS**;

Writing, review and editing: **CNS**, BCdJ, MO, KA, MAL, BO, SH, OAB, DHM, IDO, AAP, SA, SG, MC, LR, DA;

Contributed reagents, materials, analysis tools: BCdJ, SH, DHM, IDO, AAP, SG, MC, LR, DA

Abstract

Objectives: Differentiation of the *Mycobacterium tuberculosis* complex (MTBc) from non-tuberculous mycobacteria (NTM) is important for tuberculosis diagnosis and is a prerequisite for reliable phenotypic drug-resistance testing. We evaluated the performance of the rapid MPT64-antigen identification test for the detection of *M. africanum* Lineage 5 (MAF L5).

Methods: Smear-positive tuberculosis patients' sputa were included prospectively. Culture was performed on Löwenstein-Jensen medium, and when positive, the MPT64 test and classical identification tests of para-nitro-benzoic acid susceptibility and a heat-labile catalase (PNB/catalase) were performed. MPT64 test was repeated 14 days after an initially negative first testing. Direct spoligotyping was performed for MTBc lineage determination.

Results: In total 333 isolates were tested for all methods, including 322 (96.7%) pure MTBc by agreement between spoligotyping and PNB/catalase, while 11 were NTM or a mixture MTBc/NTM. On these pure MTBc isolates, the MPT64 test conducted on day-zero of culture-positivity correctly identified most MTBC isolates (93.2%, 300/322), but failed to detect 24% of L5 isolates (18/75) versus 2% (4/202) of L4 (OR=15.6(5.3-45.8), p<0.0001), with improved sensitivity for L5 detection on repeat testing after 14 days. The L5-wide non-synonymous SNP in the *mpt64* gene may explain the poor performance of the MPT64 test for L5.

Conclusion: The MPT64 test has a lower sensitivity for detecting L5 isolates of the MTBc, and can be considered as a first-screening test that should be confirmed by another identification method when negative in countries with L5. Given the microbiological bias in both isolation and identification of MAF lineages, diagnostics with high sensitivity for direct testing on clinical material

are preferable.

Keywords: Rapid MPT64 identification test; *Mycobacterium tuberculosis* complex lineages; *Mycobacterium africanum* West African 1 (Lineage 5); non-tuberculous mycobacteria; *mpt64* gene (Rv1980c); non-synonymous SNP

Introduction

Tuberculosis (TB) remains a public health problem, especially in low-resource countries where 95% of global tuberculosis is detected [1].

Microscopic detection of acid-fast bacilli (AFB) has been the main TB diagnostic tool for more than a century, yet it cannot distinguish MTBc (*Mycobacterium tuberculosis* complex) from NTM (non-tuberculous mycobacteria) [2]. The advent of the molecular Cepheid® GeneXpert MTB/RIF test partially resolved this problem as it can confirm the presence of MTBc but not NTM, and is also more sensitive than AFB-microscopy for TB diagnosis [2]. Nevertheless, compared to culture, GeneXpert MTB/RIF still has a lower sensitivity for the diagnosis of TB in smear-negative pulmonary and extra-pulmonary specimens [3, 4] as well as for TB in children [4]. Culture thus remains the most sensitive laboratory test for the diagnosis of TB. Isolation of strains in culture is also a prerequisite for conventional phenotypic drug-susceptibility testing (DST) techniques. When culture is positive, rapid differentiation of MTBc from NTM is necessary for prompt TB treatment initiation to ensure better outcome. Moreover, the differentiation of MTBc from NTM is crucial for a valid interpretation of resistance patterns in phenotypic DST assays, as NTM can be intrinsically resistant to anti-TB drugs and be mistaken for multidrug-resistant MTBc.

Many techniques have been described for the differentiation of MTBc from NTM, such as growth in the presence of para-nitro-benzoic acid (PNB), heat-labile catalase test, cord-formation ability, hybridization with specific molecular probes and High Performance Liquid Chromatography [5, 6]. Due to their speed and simplicity, the most popular tests are the rapid and simple immuno-chromatographic methods, which yield result in fifteen minutes [2]. The MPT64

antigen test has been reported in a systematic review [7], to be highly sensitive and specific for the identification of MTBc. Few studies included strains isolated from solid medium [7, 8] or measured the test performance in different MTBc lineages [9].

Human-adapted MTBc members comprise *M. tuberculosis sensus stricto* and *M. africanum* (MAF). The latter is subdivided in MAF West-African 1 and 2, also called Lineage 5 (L5) and 6 (L6) respectively and are mostly restricted to the West-African region, where they cause up to 40% of TB [10]. Recently, a study in The Gambia, where L6 is prevalent, found that the MPT64-antigen test has low sensitivity to detect L6 isolates grown in automated liquid Mycobacterial Growth Indicator Tube (MGIT) [9], while the sensitivity of the test for the detection of L5 isolates remains unknown. Nevertheless, a recent study found a substitution (I43N) in the *mpt64* gene of L5 isolates in Ghana [11]; and Gagneux et al. [12] suggested that L5 has a non-synonymous single nucleotide polymorphism (nSNP; not specified) in the *mpt64* gene, which could negatively impact the performance of the MPT64 antigen test for L5 detection, affecting countries where this lineage is endemic.

We evaluated the performance of the MPT64-antigen rapid identification test for L5 detection on isolates from solid culture medium in Benin, where L5 causes up to one third of human TB [13, 14], and compared available L5- and L6 genomes for nSNPs in the *mpt64* gene.

Materials and Methods

Patient selection and specimens

Presumptive TB patients were systematically screened with AFB (acid-fast-bacilli) microscopy nationwide from the 24 largest TB clinics in Benin from April to December 2016. TB patients with smear-positive sputa were prospectively recruited. In each TB clinic, all retreatment TB patients and a sample of new TB patients (the next four diagnosed after a retreatment TB patient) were included. Their sputa were collected before they initiated TB treatment, and shipped to the National Reference Laboratory for Mycobacteria (Laboratoire de Référence des Mycobactéries) in Cotonou, Benin, where further laboratory analyses were performed.

Specimen preparation: sputum decontamination and culture of mycobacteria

Sputa were decontaminated using the Petroff method [5] (2% NaOH final for 15 minutes and neutralized using 1N HCl containing phenol red), followed by centrifugation at 3000g and 4°C for 20 minutes. The pellet was resuspended in 1.5 mL of Phosphate-Buffered-Saline (PBS). Per patient's sputum, two standard Löwenstein Jensen (LJ) slants and one LJ slant without glycerol but supplemented with 0.5% sodium-pyruvate were inoculated, incubated at 37°C and read weekly for 13 weeks (90 days) before reporting a negative culture. An aliquot of the sediment was stored at -20°C for direct spoligotype analysis.

MPT64 antigen rapid identification test

The 'SD Bioline TB Ag MPT 64 Rapid' assay (Standard Diagnostics, Korea) was used for identification of isolates from culture-positive specimens following the manufacturer's instructions [15]. The test was performed at Day 0 (D0) and, if negative, was repeated 14 days later (D14). D0 was defined as the date when

primary colonies were large enough that they could be scraped from the slant. The test was repeated at D14 on colonies remaining on the primary slant after it had been re-incubated at 37°C. A few colonies were suspended in 200 µL of the extraction buffer provided in the kit or in the condensation fluid (if any) of the slant [15]. Then, 100 µL of this suspension was deposited in the sample well of the MPT64 device [15]. After 15 minutes' incubation, the MPT64 was reported negative if a color signal line appeared only in the control band and not in the test band and positive if a color signal line appeared in the test band as well[15]. For MPT64-positive isolates, the intensity of the positive color signal line was also recorded as "strong" (strong intensity) or "faint" (weaker intensity). MPT64 devices were double-read by another person who was blinded for the results of the first reading, mainly because of faint positivity signals. The MPT64 antigen test was performed without prior knowledge of spoligotype analysis results (spoligotype analysis for all specimens was performed in batches after availability of MPT64 results). The MPT64 test, spoligotype analysis and conventional identification were done blindly by different persons.

Conventional identification with PNB/Catalase along with DST

Identification with PNB on LJ medium was performed along with 1st line DST. Mycobacterial suspensions were inoculated on LJ medium with a final concentration of 500 µg/mL PNB and on LJ control slants, as described by the International Union Against Tuberculosis and Lung Diseases (IUATLD) [5]. First line DST included, as usual, rifampicin, isoniazid, streptomycin and ethambutol respectively at the critical concentrations of 40 µg mL^{-1}, 0.2 µg mL^{-1}, 4 µg mL^{-1}, 2 µg mL^{-1} using the proportion method on LJ medium [5, 16]. As recommended by the IUATLD [5], when an isolate is PNB-susceptible, and resistant to at least

one of the 1st line drug, it was tested for the production of 68°C-labile catalase to rule out NTM.

DNA extraction and spoligotyping

For each specimen, 200 µL of decontaminated sputum from the aliquot stored at -20°C was heat-inactivated at 100°C for 5 minutes [17], followed by DNA-extraction in 300 µL elution buffer using the Promega Maxwell®16 Tissue DNA purification Kit (AS1030)[18] with the Promega Maxwell 16 machine model AS2000 version 4.9, as described before [19]. A mycobacterial sediment known to be PCR-positive for MTBc was included as positive extraction control, and distilled water as negative extraction control. Spoligotyping (PCR followed by hybridization) was then performed as previously described [19, 20]. The *M. tuberculosis* H37Rv and *M. bovis* BCG reference strains were included as positive PCR-controls and distilled water as negative control. The MTBc lineage was assigned using the online tool TBlineage database [21].

Determination of NTM species and prevalence in the study population

The Genotype® *Mycobacterium* CM version 1 (Hain Lifescience) [16], molecular species identification test was used to identify presumed NTM isolates or MTBc/NTM mixtures based on PNB, catalase and spoligotyping results. DNA was extracted from the isolates by heat-inactivation (100°C for 20 minutes)[17, 19]. PCR followed by hybridization of the PCR products on the Genotype® Hain CM strip, were performed as described by the manufacturer [22]. The H37Rv reference strain was included as positive control and distilled water as negative control. After hybridization, the strips were fixed on an interpretation sheet, and interpreted following the species' profiles provided by the manufacturer.

SNP analysis of *mpt64* gene in L5

To better understand the MPT64 test performance obtained for L5 strains in our study, we investigated whether mutations are present in their *mpt64* gene (Rv1980c) using Snippy 3.1 [23] with H37Rv (NC_000962.3) as reference MTBc genome. Available whole genomes sequences (FastQ files, Supplementary material S1) from 25 strains belonging to L5 previously isolated in different countries were used (21 from Benin, 2 from Ivory-Coast, 2 from Guinea). As control group, we also checked the presence of mutations in the *mpt64* gene of 18 L6 strains (2 from Benin, 4 from Burkina-Faso, 3 from Guinea, 5 from Ivory-Coast, 4 from Senegal) for which genomes were available. For genome sequencing, multiplexed Illumina libraries were prepared following manufacturers guidelines using custom multiplex tags. Pooled samples were sequenced on an Illumina HiSeq2000 using the V4 kit to produce paired-end reads of 100 bp in length. We aimed to attain an average depth of coverage of ~100-fold for each sample.

Statistical analyses

We used Stata®12.0 (StataCorp, USA) for statistical analyses. The Mc Nemar exact Chi-squared test was used to compare paired data, and the Fisher's exact test for independent data.

Odds-ratios (OR), sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR-) were all calculated along with 95% confidence-interval (CI). The difference was considered significant when the two-sided p-value was below 0.05.

Results

Specimens testing flow

The specimen flow chart is presented in Fig. 1.

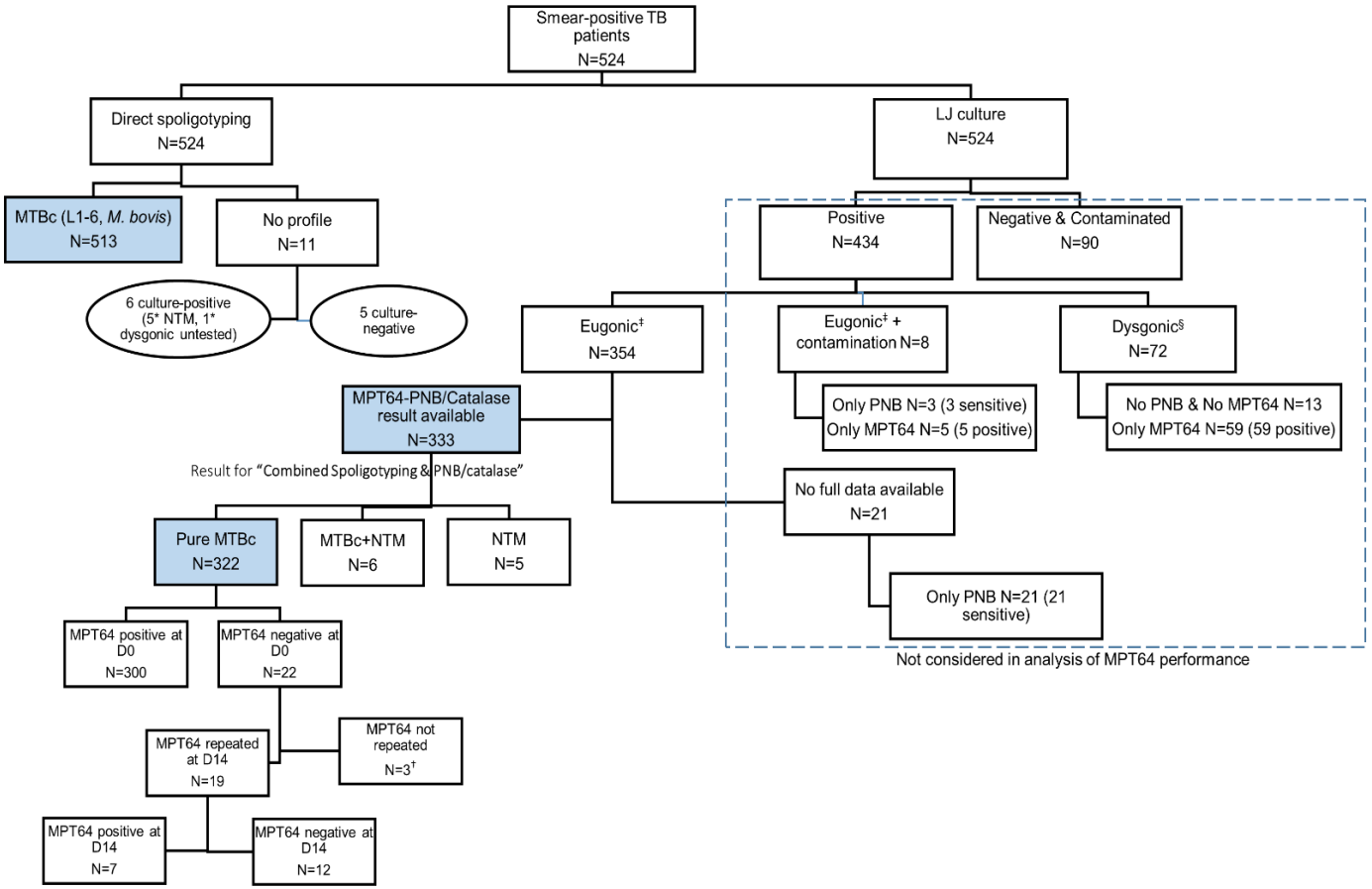

Fig. 1 Specimen flow chart

Figure legend

*: Using PNB/catalase and Genotype CM

†: MPT64 could not be repeated at D14 as very few colonies remained on the LJ slant, which were used for DST and PNB/catalase

‡: Eugonic isolate: isolate with easily scrapable colonies that easily grow in subculture

§: Dysgonic isolate: isolate with very small dry flat or convex [27], hardly scrapable colonies, not improved by subculture. Dysgonic colonies can be hardly visible.

Overall performance of MPT64 test

MPT64 at D0. In total, 397 /434 (91.5%) positive cultures had MPT64 results available (Fig. 1, Table 1). Using either spoligotyping [19, 20] or PNB/catalase [5] as reference standard for MTBc identification, 93.4% (366/392; 95%CI: 90.4-95.6) and 93.2% (300/322; 95%CI: 89.8-95.7) of MTBc isolates were correctly classified as MTBc by the MPT64 test, respectively (Table 1). The PNB/catalase results showed that 6 isolates (2 MPT64-positive and 4 MPT64-negative) identified as MTBc using spoligotyping were identified phenotypically as NTM (PNB/catalase), suggesting the presence of MTBc and NTM in the same specimen. This could be due to a mixed infection MTBc/NTM (Table 1). Excluding these isolates from the analysis, 93.2% (300/322; 95%CI:89.8-95.7) of MTBc isolates were correctly classified as MTBc by the MPT64 test on D0 with a specificity of 80% (4/5; 95%CI:28.4-99.5 (Table 1).

MPT64 at D14. The MPT64 test was repeated at D14 for 19 of the 22 isolates with a negative MPT64 test at D0 (Fig. 1). Seven of the 19 isolates became MPT64-positive, increasing the

sensitivity of the MPT64 test to 96.2% (307/319; 95%CI: 93.5-98) considering 'PNB/catalase and spoligotyping agreement' as the reference standard (Table 1).

Species of NTM identified. Of 11 isolates phenotypically identified as not being pure MTBc, 6 were considered mixtures (MTBc/NTM) based on the double banding profile on the Genotype *Mycobacterium* CM strip and/or a spoligotyping profile from the sputum specimen combined with growth of the isolate on the PNB tube. The 5 pure NTM isolates included one *M. kansasii*, one *M. abscessus* (or *M. immunogenum*), one *M. malmoense* (or *M. haemophilum* or *M. palustre* or *M. nebraskense*), one *M. scrofulaceum* (or *M. paraffunicum*

or *M. parascrofulaceum*) and one *M. species*. Among the MTBc/NTM mixtures, two isolates were positive in the MPT64 assay (Table 1). They were confirmed as mixtures by Genotype CM and the spoligotype banding profiles, revealing L4 + *M. fortuitum* and L6 + *M. species*. For the 4 remaining presumed mixtures, spoligotyping of the sputum suggested co-existence of MTBc, whereas no MTBc was detected in the isolates with only the NTM probe of the Genotype CM reacting and growth on PNB. Combined results suggested the following mixtures: *M. bovis* + *M. intracellulare*, L4 + *M. fortuitum* and twice L6 + *M. intracellulare*. The overall proportion of NTM identified was 3.3% (11/333). *M. intracellulare* followed by *M. fortuitum* were the most observed.

Table 1. Overall performance of MPT64 antigen test at day 0 (D0) and at day 14 (D14)

MPT64	Spoligotyping			PNB/Catalase			Combined "PNB/Catalase + Spoligotyping"				Sensitivity (95%CI) Specificity (95%CI) LR+ (95%CI) LR- (95%CI)
MPT64 at D0	MTBc	No profile (NTM or test failure)	Total	MTBc	NTM	Total	**"Pure" MTBc**	NTM	**Mixture spol MTBc + pnb NTM**	Total	**Sen**: **93.2** (89.8 to 95.7) **Spe**: **80** (**28.4** to 99.5) **LR+: 4.7** (0.8 to 26.9) **LR-: 0.09** (0.05 to 0.16)
+ (MTBc)	366	1	367	300	3	303	300	1[#]	2[*]	303	
- (NTM)	26	4	30	22	8	30	22	4	4[†]	30	
Total	392	5	397	322	11	333	322	5	6	333	
MPT64 at D0/D14							MTBc	NTM	**Mixture spol MTBc + pnb NTM**	Total	**Sen**: **96.2** (93.5 to 98) **Spe**: **80** (**28.4** to 99.5) **LR+: 4.8** (0.8 to 27.8) **LR-: 0.05** (0.02 to 0.1)
+ (MTBc)							307	1[#]	2[*]	310	
- (NTM)							12	4	4[†]	20	
Total							319[‡]	5	6	330[‡]	

[#]: Identified as MTBc by MPT64 but NTM by PNB/catalase and no spoligo bands or MTBc found using Genotype *Mycobacterium* CM, which identified this isolate as *M. scrofulaceum (or parafunicum or parasrofulaceum)*.

*****: MTBc/NTM mixtures confirmed using 'Genotype *Mycobacterium* CM' on the isolates.

†: Identified as NTM by MPT64 and PNB/catalase (agreement). Spoligotyping of specimens detected the presence of MTBc (two L6, one L4, one *M. bovis*), yet no mixture (MTBc/NTM) found using Genotype CM on the isolates;

‡: Three isolates with MPT64-negative at D0, had MPT64 not repeated at D14, so not included in the table

Performance of MPT64 test across MTBc lineages

For the following comparisons, NTMs and mixtures were excluded; only pure MTBc (PNB/catalase and spoligotyping confirmed) isolates (322/333, Table 1) were included.

MPT64 at D0 in MTBc isolates. At D0, MPT64 positivity varied significantly across lineages (p<0.001, Table 2). MPT64 was positive for all isolates from lineages 1, 2, 3, 6 and *M. bovis* whereas almost one quarter (24%) of MAF L5 isolates were MPT64-negative, and 2% of L4 isolates. MPT64-positive isolates were significantly under-represented among L5 versus L4 isolates, the most prevalent lineage (OR=0.06, 95%CI (0.02-0.19), p<0.0001, Table 2), corresponding to a 15.6-fold odds (95%CI: 5.3-45.8) of MPT64 false-negativity in L5 isolates (p<0.0001).

MPT64 at D14 in MTBc isolates. The 19 isolates with MPT64-negative at D0 repeated at D14 included 16 L5 and 3 L4. MPT64 became positive for an additional 7 (43.8%) L5 isolates, significantly increasing the positivity of MPT64 from 76% at D0 to 87.7% at D14 (RR=1.1, (95%CI:1.0-1.2), p=0.016, Exact Mc Nemar). None of the three L4 isolates with negative MPT64 at D0 became positive. Despite the positivity at D14 of some previously MPT64-negative isolates, MPT64 positivity still varied across lineages (Table 2). The difference in MPT64 false-negativity between L5 versus L4 isolates was still strongly significant with 12.3% (9/73) of L5 isolates versus 1.5% (3/201) from L4 remaining negative at D14 (false-negativity OR=9.3, 95%CI:2.6-32.6, p=0.0005).

Variation of the intensity of the positivity signal (color) line of the MPT64 cartridge. At D0, in total 19 (6.3%) of the 300 MPT64-positive isolates had a faint signal. Most of them were from L5 (24.6%, 14/57), followed by L6 (12.5%, 1/8) versus only 1.5% (3/198) from L4 (Table 3). The odds of obtaining a faint

signal in L5 isolates was 21.2-fold (95%CI: 6.2-71.5) the odds of that in L4 isolates ($p<0.0001$) did not decrease after D14 repeat (OR=21.7 95%CI: 6.5-72.2, $p<0.0001$) (Table 3). The proportion of isolates with a faint signal was also higher in the MAF group (L5+L6) versus 'non-MAF' ($p<0.0001$), or ancestral lineages group (L1+L5+L6) compared to modern lineages ($p<0.0001$, Table 3).

Table 2. Performance of MPT64 test across MTBc lineages. Only pure MTBc isolates (detected using PNB/catalase and spoligotyping) are included in this comparison.

PNB/Catalase (All MTBc)				
	MPT64 at D0			OR with 95%CI, p*
MTBc Lineages	Positive (MTBc)	Negative	Total	
L1	15 (100%)	0	15	-
L2	14 (100%)	0	14	-
L3	3 (100%)	0	3	-
L4	198 (98%)	4	202	• OR=**0.06** (0.02 to 0.19) • p<**0.0001**
L5	57 (**76%**)	18	75	
L6	8 (100%)	0	8	-
M. bovis	5 (100%)	0	5	-
Total	300	22	322	overall **p<0.001**
	MPT64 at D0-14			
L1	15(100%)	0	15	-
L2	14 (100%)	0	14	-
L3	3 (100%)	0	3	-
L4	198 (98.5%)	3 †	201	• OR=**0.11** (0.03 to 0.38) • p=**0.0005**
L5	64 (**87.7%**)	9 ‡	73	
L6	8 (100%)	0	8	-
M. bovis	5 (100%)	0	5	-
Total	307 k	12	319	overall **p=0.018**

*: p-values were calculated using Fischer exact test.
†: MPT64 not done at D14 for one specimen with MPT64 negative at D0, so excluded
‡: MPT64 not done at D14 for 2 specimens with MPT64 negative at D0, so excluded; MPT64 became positive for 7 specimens.

Table 3. Variation of the intensity of the positivity signal line of the MPT64 strip across MTBc lineages.

Intensity of MPT64 cartridge positivity signal band (MPT64 positive specimens)				
	MPT64 positive at D0			OR (95%CI), p[I]
MTBc Lineages	**Strong** signal	**Faint** signal	Total	
L1	14	1 (6.7%)	15	-
L2	14	0	14	-
L3	3	0	3	-
L4	195	3 (1.5%)	198	• OR=**21.2** (6.2 to 71.5) • p<**0.0001**
L5	43	14 (24.6%)	57	
L6	7	1 (12.5%)	8	-
M. bovis	5	0	5	-
Total	281	19	300	
Modern MTBc (L2+L3+L4)	212	3(1.4%)	215	• OR=**17.7** (5.3 to 58.4) • p<**0.0001**
Ancestral MTBc (L1 +L5+L6)	64	16 (20%)	80	
Other than MAF (L1+L2+L3+L4)	226	4 (1.7%)	230	• OR=**16.9** (5.6 to 50.6) • p<**0.0001**
MAF (L5+L6)	50	15 (23.1%)	55	
	MPT64 positive at D0-14			
L1	14	1 (6.7%)	15	-
L2	14	0	14	-
L3	3	0	3	-
L4	195	3 (1.5%)	198	• OR=**21.7** (6.4 to 72.2) • p**<0.0001**
L5	48	16(25%)	64	
L6	7	1 (12.5%)	8	-
M. bovis	5	0	5	-
Total	286	21	307	

Modern MTBc (L2+L3+L4)	212	3 (1.4%)	215	• OR=**18.4** (5.6 to 60.3) • p**<0.0001**
Ancestral MTBc (L1 +L5+L6)	69	18 (20.7%)	87	
Other than MAF (L1+L2+L3+L4)	226	4 (1.7%)	230	• OR=17.5(5.9 to 51.4) • p**<0.0001**
MAF (L5+L6)	55	17 (23.6%)	72	

l: p-values were calculated using the Fischer exact test

Variation of the time from inoculation (start of incubation) to the realization of the MPT64 test (Δt). The median Δt at D0 was shortest for L1, L3 and L4 (3 weeks) followed by L2 and *M. bovis* (4 weeks), L5 (6 weeks), and L6 (8 weeks). Including repeated testing, the median Δt increased for L5 to 8 weeks, while it remained similar for the other lineages. (Table S2).

SNP analysis of *mpt64* gene in L5

All 25 of the L5 isolate genomes shared the same non-synonymous SNP (nSNP) 128T>A (I43N) in the *mpt64* gene (Supplementary material S1). One of the L5 genomes had in addition, a synonymous SNP 519G>A (V197V). Among the L6 isolate genomes, there was no mutation in the *mpt*64 gene for 17 of the 18 isolates analyzed. One genome harbored a synonymous SNP 81C>T (T27T) in the gene.

Lineage distribution among population of isolates included versus excluded in the MPT64 comparison

The lineage distribution in the isolates (pure MTBc) included in the "MPT64 vs Combined Spoligotyping &PNB/Catalase" comparison differed significantly to the distribution in isolates excluded for unavailability of MPT64 or/and PNB/catalase result (poor quality for subsequent tests) (Table 4). MAF L5 and

L6 were significantly over-represented in that excluded group relative to the included group (OR=**4.5** (95% CI: 2.7-7.5), p<0.0001). The excluded group of isolates (n=80) included 72 dysgonic and 8 partially contaminated isolates. When comparing the included group to only the dysgonic isolates among the excluded group, we found that the dysgonic nature of isolates was strongly associated with the lineage, with L5 and L6 over-represented among dysgonic isolates (OR=5.7**,** 95%CI (3.3-9.9), p<0.0001), especially L6 (OR=9.5**,** 95%CI (3.9-23.1), p<0.0001) (Table 4)

Table 4. MTBc isolate culture outcome across lineages: lineage distribution among population of isolates included versus excluded ("dysgonic/partially contaminated") in the MPT64/PNB comparison

MTBc lineages	All specimens (culture positive, negative and contaminated) %(no)	All culture-positive specimens %(no)	**Included isolates (Pure MTBc)** %(no)	**Included vs Excluded for quality (dysgonic, partial contamination)**				**Included (eugonic) vs Excluded (dysgonic only)**			
				Excluded isolates (dysgonic/partially contaminated) %(no)	**OR** Excluded/Include, **95%CI**	**%Difference** (Excluded-Included), **95%CI**	**p**	**Dysgonic among excluded** %(no)	**$OR_{Dysgonic}$** among Excluded/ Eugonic (Included) **, 95%CI**	**%Difference** [Dysgonicamong Excluded - Eugonic (Included)], **95%CI**	**p**
Total	N=513*	N=428†	**N=322**	**N=79‡**				**N=72**			
L1	3.9 (20)	4.2 (18)	4.7 (15)	3.8 (3)	0.8 (0.2 to 2.7)	-0.9 (-5.7 to 3.9)	1	2.8 (2)	0.6 (0 to 2.4)	-1.9(-6.3 to 2.6)	0.749
L2	4.3 (22)	4.7 (20)	4.3 (14)	3.8 (3)	0.9 (0.3 to 2.9)	-0.6 (-5.3 to 4.2)	1	2.8 (2)	0.6(0 to 2.5)	-1.6(-6 to 2.8)	0.747
L3	0.8 (4)	0.7 (3)	0.9 (3)	0 (0)	0 (0 to 5.3)	-0.9 (-2 to 0.1)	1	0 (0)	0(0 to 5.8)	-0.9(-2 to 0.1)	1
L4	52.4 (269)	56.1 (240)	**62.7** (202)	**29.1** (23)	**0.2** (0.1 to 0.4)	**-33.6**(-44.9 to -22.3)	**<0.0001**	**25** (18)	**0.2**(0.1 to 0.4)	**-37.7**(-49 to -26.4)	**<0.0001**
L5	27.9 (143)	26.4 (113)	**23.3** (75)	**43.0** (34)	**2.5** (1.5 to 4.2)	**19.7** (7.9 to 31.6)	**0.0007**	**47.2** (34)	**3**(1.7 to 5)	**23.9**(11.5 to 36.4)	**0.0001**
L6	8.4 (43)	6.1 (26)	**2.5** (8)	**17.7** (14)	**8.5** (3.5 to 20.5)	**15.2** (6.6 to 23.8)	**<0.0001**	**19.4** (14)	**9.5**(3.9 to 23.1)	**17**(7.7 to 26.3)	**<0.0001**
M. bovis	2.3 (12)	1.9 (8)	1.6 (5)	2.5 (2)	1.7 (0 to 7.5)	1 (-2.7 to 4.7)	0.628	2.8 (2)	1.8(0 to 8.3)	1.2(-2.8 to 5.3)	0.616
Other than MAF	63.7 (327)	67.5 (289)	**74.2** (239)	**39.2** (31)	**0.2** (0.1 to 0.4)	**-35** (-46.8 to -23.2)	**<0.0001**	**33.3**(24)	**0.2**(0.1 to 0.3)	**-40.9**(-52.8 to -29)	**<0.0001**

MAF	36.3 (186)	32.5 (139)	**25.8** (83)	**60.8** (48)	**4.5** (2.7 to 7.5)	**35** (23.2 to 46.8)	**66.7** (48)	**5.8**(3.3 to 9.9)	**40.9**(29 to 52.8)

p-values were calculated using Fischer exact test *****: There were 11 specimens that did not yield a spoligotype profile among the 524 specimens

†: There were 6 specimens that did not yield a spoligotype profile among the 434 culture-positive specimens.

‡: There is one specimen with failed spoligotyping among the 80 dysgonic/partially contaminated) specimens (101 excluded in total – 21 with PNB available and MPT64 missing).

Discussion

Our evaluation in a nationwide prospective study of LJ based primary cultures from smear-positive TB patients in Benin suggests lower performance of the SD Bioline® MPT64 antigen test for rapid identification of the L5 of MTBc.

This lower sensitivity to confirm L5 as MTBc, leading to its misclassification as NTM. Repeating the test 14 days after a first negative result improved the sensitivity for L5 detection significantly, yet not completely. The few L4 isolates that tested negative at D0 (1.5%) remained negative at D14 testing, which could mean that those L4 isolates and the L5 isolates with MPT64 negative at D14 need an incubation time beyond 14 days after D0, or could point to mutations in the *mpt64* gene, as found in L4 (including a 63-bp deletion [24]) in other studies [24, 25]. The MPT64 test positivity at D0 for L5 isolates cultured on LJ (solid) medium in our study (76%) was similar to the findings of Ofori-Anyinam et al. on L6 isolates cultured in MGIT (liquid) medium (78.4%) at the same D0 testing time-point, with a similar increase in positivity at D14 for L5 in our study (87.7%) relative to L6 at D15 (90.2%)[9]. In our study, however all L6 isolates (n=8) were positive in the MPT64 test, even at D0. This observed difference – albeit on a small number – may be explained by the lower performance of MPT64 for L6 cultured in MGIT than for isolates cultured on LJ, as mycobacterial growth is more rapid in MGIT (liquid) than on LJ (solid) medium, allowing the production of a higher quantity of MPT64 protein on LJ versus MGIT. Indeed, a shorter incubation time posed an independent risk for false negative MPT64 test in the MGIT based study [9].

SNP analysis of the *mpt64* gene in L5 genomes confirmed that L5 isolates harbor a nSNP in this gene [11] (also confirmed in all the 367 L5 genomes available in another genome collection from various countries- personal

communication with Coscolla M), leading to a modification of the amino-acid chain of the MPT64 produced by L5 (I43N), probably impacting the protein structure. Jiang et al. [24] found that nSNPs in the *mpt64* gene rarely changed the structure and function of the protein, in contrast to a 63-bp deletion (amino-acids 66-86) observed mostly in L4 but also in some L1 isolates [24-26].

Oettinger et al. created five monoclonal antibodies (C24b1-3, L24b4-5) that reacted with four epitopes of the MPT64 antigen [26]. The MPT64 antigen's epitope for the C24b3 antibody consists of two structural domains found in the sequences Ala1-Leu43 and Ala108-Ser152 [26]. The Ile43Asn mutation (I43N) found in all L5 isolates genomes coincides with the final amino-acid of the first part of this epitope. This could lead to partial binding, lower adherence or prevent the binding of the C24b3 antibody to its partially-modified epitope in the MPT64 antigen; resulting in a faint positivity-signal band or negative MPT64 test, if that antibody was used in the development of the MPT64 test. The mouse antibodies (at least 3) [15] used in the development of the MPT64 test are not specified, but are likely included among these five detected by Oettinger et al. [26] as this paper is cited in the SD Bioline MPT64 sheet [15]. This can explain why despite the non-synonymous mutation in the *mpt64* gene of L5, there were a significant proportion of isolates with faint positivity signal bands, isolates with strong positivity signal bands, and a significant proportion of MPT64-negative isolates (even after D14 repeat). Nevertheless, changing the mouse antibody of which the corresponding MPT64 epitope is mutated in L5 isolates may improve the sensitivity for L5. It was reported that the MPT64 test has lower sensitivity for L6 strains [9], although no missense *mpt64* mutations were identified in L6 genomes. So another mechanism may account for the lower sensitivity in MAF L5 as for MAF L6 strains, in which the expression of the *mpt64* gene was lower than in *M. tuberculosis* sensus stricto [9]. More

extended gene expression and regulation studies should be conducted in order to confirm these possible causes for the lower performance of MPT64 tests in L5 and L6 MAF lineages.

Importantly, in West-Africa where MAF is common, MPT64-negative tests should be confirmed by another identification method (such as Genotype Hain CM, IS6110 PCR, spoligotyping, Cepheid GeneXpert, PNB/catalase) before being classified as NTM. If available, molecular analyses (Genotype *Mycobacterium* CM, IS6110-PCR, spoligotyping, GeneXpert MTB/RIF (on a diluted bacterial suspension)) should be prioritized for this confirmation, as these are more rapid and sensitive (can identify MTBc in mixture MTBc/NTM isolates) than PNB/catalase. If only PNB/catalase is available, further MPT64 testing at least at day 14 after the first testing can be done (while PNB/catalase is underway) as the result (if positive) could be obtained more rapidly than PNB/catalase.

In our study, MAF isolates are more likely to be dysgonic, as previously reported [27, 28], particularly for L6. Gehre et al. found that L6 has non-synonymous mutations in genes related to growth in culture (*aceE, recA, Rv2112, Rv0862*) that may explain its attenuated growth in culture [28]. A previous study also described that L5 is less likely to grow in culture compared to *M. tuberculosis* sensu stricto [19]. Our findings suggest that, in addition to the lower ability of MAF strains to yield a positive culture, the MAF isolates successfully grown are more likely to be dysgonic, jeopardizing phenotypic post-culture tests. These findings highlight the need for simple phenotypic or genotypic diagnostic tests that can be directly applied on patient sputa/uncultured specimens, to detect all species of the MTBc, distinguish MTBc strains from NTM, detect possible mixed MTBc/NTM infections, and drug-resistance against multiple antibiotics.

One strength of our study is that it was conducted in a setting where L5 is common, allowing the prospective assessment of the performance of the

MPT64 test for identifying this lineage as MTBc member. Testing on primary isolation reflects how the MPT64 test is used in routine practice, typically on cultures from smear-negative or GeneXpert MTB/Rif-negative specimens, or extra-pulmonary specimens, and for rapid screening identification before phenotypic DST. Our testing was essentially blinded to the strain lineage, which was determined later and by different technicians. Possible limitations of our study include the fact that we only used one of the four commercially available MPT64 assays (Capillia, SD Bioline, MGIT TBc, TB Check), although a study on L6 did not find a difference between MPT64 tests from different manufacturers [9]. We expect that our findings are therefore generalizable to other settings, especially in West- and Central Africa, where L5 is prevalent.

We therefore strongly recommend that, the difference in MTBc lineage characteristics should be considered in diagnostics development, so tests will perform equally well in ancestral lineages such as L5 and L6. Furthermore, their performance should be evaluated in different settings, especially where such ancestral lineages are common.

Authors statements

Funding information

This work was supported by funds from the Directorate General for Development (DGD), Belgium (FA4) to CNS, BCdJ and DA, and the European Research Council-INTERRUPTB starting grant nr.311725 to BCdJ, LR, BO. The funders had no role in study design, data collection and interpretation or the decision to submit the work for publication.

Acknowledgments

The MAF genomes analyzed in this study were sequenced by the Welcome Trust Sanger Institute. We thank Julian Parkhill (Welcome Trust Sanger

Institute, United Kingdom) for contributing in genome sequencing, and Conor Meehan (Institute of Tropical Medicine, Antwerp, Belgium) for advising on genome analysis, and providing feedback on the manuscript.

Conflict of interest

All authors declare there is no conflict of interest regarding the study and its findings.

Ethical statement

This study is part of the BeniDiT study that has been approved by the Ethics Committee of Benin, and the ones of the Institute of Tropical Medicine and the University of Antwerp, Belgium. Its registration number on ClinicalTrials.gov is NCT02744469. All included patients provided written informed-consent. Moreover, all sputa were anonymized before laboratory analyses.

Abbreviations

AFB: acid-fast bacilli; **CI**: confidence interval; **DST**: drug susceptibility testing; **D0**: date when primary colonies were large enough that they could be scraped from the culture medium slant; **D14**: 14 day after D0; **HCl**: hydrochloric acid; **IUATLD**: International Union Against Tuberculosis and Lung diseases; **LJ**: Löwenstein-Jensen medium; **L1**: Lineage 1; **L2**: Lineage 2; **L3**: Lineage 3; **L4**: Lineage 4; **L5**: Lineage 5; **L6**: Lineage 6; **MAF**: *Mycobacterium africanum* (or *M. africanum*); **MGIT**: mycobacterial growth indicator tube; ***M. tuberculosis***: *Mycobacterium tuberculosis*; **MTBc**: *Mycobacterium tuberculosis* complex; **NaOH**: sodium hydroxide; **LR-**: negative likelihood ratio; **nSNP**: non-synonymous SNP; **NTM**: non-tuberculous mycobacteria; **OR**: odds ratio; **LR+**:

positive likelihood ratio; **PNB**: para-nitro-benzoic acid; **Sen**: sensitivity; **Spe**: specificity; **TB**: tuberculosis

References

1. **Raviglione MC, Snider DE, Jr., Kochi A**. Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. *Jama* 1995;273(3):220-226.
2. **WHO**. TB Diagnostics and Laboratory Services: Information note (laboratory info note GF R 11 ver 2). http://www.who.int/tb/dots/lab.pdf [accessed.
3. **Sharma SK, Kohli M, Yadav RN, Chaubey J, Bhasin D et al.** Evaluating the Diagnostic Accuracy of Xpert MTB/RIF Assay in Pulmonary Tuberculosis. *PLoS One* 2015;10(10):e0141011.
4. **Zeka AN, Tasbakan S, Cavusoglu C**. Evaluation of the GeneXpert MTB/RIF assay for rapid diagnosis of tuberculosis and detection of rifampin resistance in pulmonary and extrapulmonary specimens. *J Clin Microbiol* 2011;49(12):4138-4141.
5. **Rieder H L VA, Kam K M , Kim S J, Chonde T M , Trébucq A , Urbanczik R**. *Priorities for Tuberculosis Bacteriology Services in Low-Income Countries*, Second edition
ed: International Union Against Tuberculosis and Lung Disease (The Union) Second edition 2007.
6. **Leao S C MA, Mejia M G I, Palomino J C, Robledo R J, Telles M A d S, Portaels F**. *Practical hanbook for phenotypic and genotypic identification of mycobacteria* 2004.
7. **Yin X, Zheng L, Lin L, Hu Y, Zheng F et al.** Commercial MPT64-based tests for rapid identification of Mycobacterium tuberculosis complex: a meta-analysis. *J Infect* 2013;67(5):369-377.
8. **Hopprich R, Shephard L, Taing B, Kralj S, Smith A et al.** Evaluation of (SD) MPT64 antigen rapid test, for fast and accurate identification of Mycobacterium tuberculosis complex. *Pathology* 2012;44(7):642-643.
9. **Ofori-Anyinam B, Kanuteh F, Agbla SC, Adetifa I, Okoi C et al.** Impact of the Mycobaterium africanum West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased Sensitivity of Rapid Identification Tests in The Gambia. *PLoS Negl Trop Dis* 2016;10(7):e0004801.
10. **de Jong BC, Antonio M, Gagneux S**. Mycobacterium africanum--review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis* 2010;4(9):e744.
11. **Otchere I, Harris S, Busso S, Asante-Poku A, Osei-Wusu S et al.** The First population structure and comparative genomics analysis of Mycobacterium africanum strains from Ghana reveals higher diversity of Lineage 5. *International Journal of Mycobacteriology*, Article 2016;5(5):80-81.
12. **Gagneux S, Small PM**. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *The Lancet Infectious Diseases* 2007;7(5):328-337.
13. **Gehre F, Antonio M, Faihun F, Odoun M, Uwizeye C et al.** The first phylogeographic population structure and analysis of transmission dynamics of M. africanum West African 1--combining molecular data from Benin, Nigeria and Sierra Leone. *PLoS One* 2013;8(10):e77000.

14. **Affolabi D, Anyo G, Faihun F, Sanoussi N, Shamputa IC et al.** First molecular epidemiological study of tuberculosis in Benin. *Int J Tuberc Lung Dis* 2009;13(3):317-322.
15. **SD**. SD Bioline TB Ag MPT64 Rapid. http://www.biosystemsantioquia.com.co/images/docs/reactivos/enfermedades-infecciosas/otras/08fk50-sd-tb-ag-mpt64-d.pdf [accessed 2018].
16. Handbook on TB laboratory diagnostic methods for the European Union, Stockholm: ECDC; 2016 (2016).
17. **Bemer-Melchior P DHB**. Inactivation of Mycobacterium tuberculosis for DNA Typing Analysis *JCM* 1999;July 1999: p. 2350–2351
18. **Promega**. Maxwell 16 DNA Purification Kits Technical Manual. https://worldwide.promega.com/resources/protocols/technical-manuals/0/maxwell-16-dna-purification-kits-protocol/ [accessed 2018].
19. **Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B**. Genotypic characterization directly applied to sputum improves the detection of Mycobacterium africanum West African 1, under-represented in positive cultures. *PLoS Negl Trop Dis* 2017;11(9):e0005900.
20. **Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D et al.** Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J Clin Microbiol* 1997;35(4):907-914.
21. **Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL et al.** TB-Lineage: an online tool for classification and analysis of strains of Mycobacterium tuberculosis complex. *Infect Genet Evol* 2012;12(4):789-797.
22. **Lifescience H**. Genotype Mycobacterium CM. http://www.hain-lifescience.de/en/products/microbiology/mycobacteria/ntm/genotype-mycobacterium-cm.html [accessed 09 May 2018.
23. **Seemaan T**. Snippy: rapid haploid variant calling and core SNP phylogeny. https://github.com/tseemann/snippy [accessed 2018].
24. **Jiang Y, Liu H, Wang H, Dou X, Zhao X et al.** Polymorphism of antigen MPT64 in Mycobacterium tuberculosis strains. *J Clin Microbiol* 2013;51(5):1558-1562.
25. **Hirano K, Aono A, Takahashi M, Abe C**. Mutations Including IS6110 Insertion in the Gene Encoding the MPB64 Protein of Capilia TB-Negative Mycobacterium tuberculosis Isolates. *Journal of Clinical Microbiology* 2004;42(1):390-392.
26. **Oettinger T, Andersen AB**. Cloning and B-cell-epitope mapping of MPT64 from Mycobacterium tuberculosis H37Rv. *Infection and immunity* 1994;62(5):2058-2064.
27. **Pattyn SR, Portaels F, Spanoghe L, Magos J**. Further studies on African strains of Mycobacterium tuberculosis: comparison with M. bovis and M. microti. *Annales des Societes belges de medecine tropicale, de parasitologie, et de mycologie* 1970;50(2):211-227.
28. **Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML et al.** Deciphering the growth behaviour of Mycobacterium africanum. *PLoS Negl Trop Dis* 2013;7(5):e2220.

Supplementary materials

Table S1. Accession number of genomes analyzed on NCBI

Isolate ID	Study accession	Genome accession	Lineage	Country of isolation	SNP analysis mpt64	Amino-acid change in *mpt64* gene
ITM_030231	PRJEB4884, ERP004216	ERR439930	L6	Senegal	no mutation	
ITM_060418	PRJEB4884, ERP004216	ERR439933	L6	Guinea	no mutation	
ITM_060422	PRJEB4884, ERP004216	ERR439934	L6	Guinea	no mutation	
ITM_060423	PRJEB4884, ERP004216	ERR439935	L6	Guinea	no mutation	
ITM_060430	PRJEB4884, ERP004216	ERR439936	L5	Guinea	128T>A (non-synonymous)	Ile43Asn
ITM_060789	PRJEB4884, ERP004216	ERR439937	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060792	PRJEB4884, ERP004216	ERR439939	L5	Benin	128T>A (non-synonymous); 519G>A (synonymous)	Ile43Asn; Val197Val
ITM_060793	PRJEB4884, ERP004216	ERR439940	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060795	PRJEB4884, ERP004216	ERR439941	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060799	PRJEB4884, ERP004216	ERR439944	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060800	PRJEB4884, ERP004216	ERR439945	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060801	PRJEB4884, ERP004216	ERR439946	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060802	PRJEB4884, ERP004216	ERR439947	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060803	PRJEB4884, ERP004216	ERR439948	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060808	PRJEB4884, ERP004216	ERR439949	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060811	PRJEB4884, ERP004216	ERR439950	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060814	PRJEB4884, ERP004216	ERR439951	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060816	PRJEB4884, ERP004216	ERR439952	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060819	PRJEB4884, ERP004216	ERR439953	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060823	PRJEB4884, ERP004216	ERR439955	L5	Benin	128T>A (non-synonymous)	Ile43Asn

ITM_060830	PRJEB4884, ERP004216	ERR439956	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060831	PRJEB4884, ERP004216	ERR439957	L6	Benin	no mutation	
ITM_060833	PRJEB4884, ERP004216	ERR439958	L6	Benin	no mutation	
ITM_060834	PRJEB4884, ERP004216	ERR439959	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060837	PRJEB4884, ERP004216	ERR439960	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060839	PRJEB4884, ERP004216	ERR439961	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060846	PRJEB4884, ERP004216	ERR439962	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060872	PRJEB4884, ERP004216	ERR439963	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_060873	PRJEB4884, ERP004216	ERR439964	L5	Benin	128T>A (non-synonymous)	Ile43Asn
ITM_081443	PRJEB4884, ERP004216	ERR439969	L5	Guinea	128T>A (non-synonymous)	Ile43Asn
ITM_08496	PRJEB4884, ERP004216	ERR439976	L6	Ivory-Coast	no mutation	
ITM_08540	PRJEB4884, ERP004216	ERR439977	L6	Ivory-Coast	no mutation	
ITM_08552	PRJEB4884, ERP004216	ERR439978	L6	Ivory-Coast	no mutation	
ITM_08555	PRJEB4884, ERP004216	ERR439979	L6	Ivory-Coast	no mutation	
ITM_08616	PRJEB4884, ERP004216	ERR439980	L5	Ivory-Coast	128T>A (non-synonymous)	Ile43Asn
ITM_101231	PRJEB4884, ERP004216	ERR439981	L6	Senegal	no mutation	
ITM_101679	PRJEB4884, ERP004216	ERR439986	L6	Senegal	81C>T (synonymous)	Thr27Thr
ITM_101682	PRJEB4884, ERP004216	ERR439987	L6	Senegal	no mutation	
ITM_07519	PRJEB4884, ERP004216	ERR439990	L6	Burkina-Faso	no mutation	
ITM_07522	PRJEB4884, ERP004216	ERR439991	L6	Burkina-Faso	no mutation	
ITM_07525	PRJEB4884, ERP004216	ERR439992	L6	Burkina-Faso	no mutation	
ITM_07533	PRJEB4884, ERP004216	ERR439994	L6	Burkina-Faso	no mutation	
ITM_08448	PRJEB4884, ERP004216	ERR439995	L6	Ivory-Coast	no mutation	
ITM_08557	PRJEB4884, ERP004216	ERR439996	L5	Ivory-Coast	128T>A (non-synonymous)	

Table S2. Variation of the time from inoculation (incubation start) to the realization of MPT64 across MTBc lineages (Δt)

MTBc lineages	MPT64 at D0	Time from inoculation (incubation start) to the realization of MPT64 at D0 (day)													Total	Median time to MPT64-D0	Median time to MPT64-D14
		W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13			
L1	Positive†	-	2	**6**	1	2	4	-	-	-	-	-	-	-	**15**	W3	W3
	Negative‡	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**		
L2	Positive†	-	1	1	**11**	-	1	-	-	-	-	-	-	-	**14**	W4	W4
	Negative‡	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**		
L3	Positive†	-	-	3	-	-	-	-	-	-	-	-	-	-	**3**	W3	W3
	Negative‡	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**		
L4	Positive†	1	22	**99**	42	3	18	-	7	-	1	-	-	5	**198**	W3	W3
	Negative‡	-	-	2	1	-	-	-	1*	-	-	-	-	-	**4**		
L5	Positive†	-	-	4	13	1	10	1	**21**	-	7	-	-	-	**57**	W6	W8
	Negative‡	-	-	3	2	-	8*	-	2*	-	2	-	-	1	**18**		
L6	Positive†	-	-	-	-	-	3	-	3	-	-	-	-	2	**8**	W8	W8
	Negative‡	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**		
M. bovis	Positive†	-	-	-	**2**	-	1	-	1	-	1	-	-	-	**5**	W4	W4
	Negative‡	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**		
Total		1	25	**118**	72	6	45	1	35		11	0	0	8	322	W4	W4

W: week; - : no specimens (0); † MPT64 done only at D0; ‡ MPT64 repeated at D14; *MPT64 not done at D14 for one specimen with MPT64 negative at D0

PART 2

Understanding *M. africanum* West African 1 (Lineage 5) epidemiology and genomic characteristics

Chapter 5

Genotypic characterization directly applied to sputum improves the detection of *Mycobacterium africanum* West African 1, under-represented in positive cultures

C. N'Dira Sanoussi[1, 2*], Dissou Affolabi[2], Leen Rigouts[1, 3], Séverin Anagonou [2], Bouke de Jong [1*]

[1] Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium,

[2] Laboratoire de Référence des Mycobactéries, Centre National Hospitalier Universitaire de Pneumo-Phtisiologie de Cotonou, National Tuberculosis Programme, Cotonou, Benin, [3] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

*Corresponding authors

PLoS Negl Trop Dis. 2017 Sep 1;11(9): e0005900.doi: 10.1371/journal.pntd.0005900

https://doi.org/10.1371/journal. pntd.0005900

https://www.ncbi.nlm.nih.gov/pubmed/28863143/

Conceived and designed the research: **CNS**, DA, LR, BCdJ

Performed the experiments: **CNS**

Data analysis: **CNS**, BCdJ; Supervision: DA, LR, SA, BCdJ

Writing original draft: **CNS**

Writing, review and editing: **CNS**, DA, LR, SA, BCdJ

Contributed reagents, materials, analysis tools: DA, LR, SA, BCdJ

Abstract

Background

This study aimed to compare the prevalence of *Mycobacterium tuberculosis* complex (MTBc) lineages between direct genotyping (on sputum) and indirect genotyping (on culture), to characterize potential culture bias against difficult growers.

Methodology/ Principal findings

Smear-positive sputa from consecutive new tuberculosis patients diagnosed in Cotonou, (Benin) were included, before patients had started treatment. An aliquot of decontaminated sputum was used for direct spoligotyping, and another aliquot was cultured on Löwenstein Jensen (LJ) medium (90 days), for indirect spoligotyping. After DNA extraction, spoligotyping was done according to the standard method for all specimens, and patterns obtained from sputa were compared versus those from the derived culture isolates.

From 199 patient's sputa, 146 (73.4%) yielded a positive culture. In total, direct spoligotyping yielded a pattern in 98.5% (196/199) of the specimens, versus 73.4% (146/199) for indirect spoligotyping on cultures. There was good agreement between sputum- and isolate derived patterns: 94.4% (135/143) at spoligotype level and 96.5% (138/143) at (sub)lineage level. Two of the 8 pairs with discrepant pattern were suggestive of mixed infection in sputum.

Ancestral lineages (Lineage 1, and *M. africanum* Lineages 5 and 6) were less likely to grow in culture (OR=0.30, 95%CI (0.14 to 0.64), p=0.0016); especially Lineage 5 (OR=0.37 95%CI (0.17 to 0.79), p=0.010). Among modern lineages, Lineage 4 was over-represented in positive-culture specimens (OR=3.01, 95%CI (1.4 to 6.51), p=0.005).

Conclusions/ Significance

Ancestral lineages, especially *M. africanum* West African 1 (Lineage 5), are less likely to grow in culture relative to modern lineages, especially *M. tuberculosis* Euro-American (Lineage 4). Direct spoligotyping on smear positive sputum is effective and efficient compared to indirect spoligotyping of cultures. It allows for a more accurate unbiased determination of the population structure of the *M. tuberculosis* complex.

Keywords: Spoligotyping, sputum, culture isolate, *M. tuberculosis* complex, *M. africanum*, ancestral lineages, population structure

Author summary

The vast majority (95%) of tuberculosis (TB) patients worldwide live in low-income countries, including in West-Africa. Typing the bacteria responsible for TB (tuberculosis; *Mycobacterium tuberculosis* complex) is important for targeted TB control. Typing is usually performed on isolates obtained after the culture isolation of TB bacteria in the sputa from patients. However, cultures can be false negative, and some 'ancestral' strains, only found in West-Africa (*Mycobacterium africanum*), require more time (90 days versus the usual 56 days) to grow in culture. To characterize potential culture bias against such "difficult growers", we compared the performance of direct typing (on sputum) relative to its yield on culture isolates. We found that ancestral types of TB bacteria were significantly less likely to grow in culture despite the 90-day incubation. This suggests that typing results of cultured isolates are not representative of the diversity in the population of TB bacteria causing disease in patients. Typing sputum directly is effective and can be used for a more

precise, unbiased determination of the proportion of different TB bacteria in a population.

Introduction

Tuberculosis (TB), caused by bacteria of the *Mycobacterium tuberculosis* complex (MTBc), remains a public health problem. Globally, over 8 million new patients with TB disease arise each year, including 2 million deaths. The vast majority (95%) of global TB is detected in limited-resource countries [1], including West-Africa. Each year in Benin, over 4000 cases of TB are detected, and the incidence of smear-positive pulmonary TB is 39 per 100000 inhabitants.

Genotypic characterization is important in order to understand the population structure of the MTBc for better insights into endemic- and epidemic strains and to identify instances of nosocomial transmission or laboratory contamination. *M. tuberculosis sensu stricto* and *M. africanum* sub-species within the MTBc have been subdivided into 7 main lineages of human importance [2,3]. These 7 MTBc lineages are classified as ancestral (or 'ancient') (Lineages 1, 5, 6) [4,5], intermediate (Lineage 7) [3,4] and modern lineages (Lineages 2,3,4) [4]. Lineage 5 (*M. africanum* West African 1) and Lineage 6 (*M. africanum* West african 2) are only found in West- and Central Africa, where they cause up to 40% of all TB [6,7]. Recent reports suggested a decrease in prevalence of *M. africanum* in some West-African countries [8–10]. Whether methodological issues explain the apparent disappearance of *M. africanum* has not been excluded to date.

For the determination of the population structure of the MTBc, genotyping is usually applied on culture isolates [11]. *M. africanum* grows

significantly slower than the other members of the MTBc (*M. tuberculosis sensu stricto*) [12] and cultures should be incubated for 90 days rather than the usual 56 days, before reporting a negative result [13]. However, even this extended incubation time may not permit recovery of *M. africanum* isolates at the same rate as *M. tuberculosis*, and thus bias the population structure derived from cultured isolates, especially in settings where *M. africanum* is endemic. Differences in expression of genes involved in metabolism pathways of the various MTBc lineages may also affect their growth in culture, as recently reported for *M. africanum* Lineage 6 which has an under-expression for the gene (*Dos R*) involved in adaptation to lower oxygen tension relative to to Lineage 4 [14]. For isolation, of some MTBc species, including *M. africanum*, the need for pyruvate to support growth in culture [15] has been known for a long time [16].

Few studies evaluated genotyping, such as spoligotyping, directly on clinical specimens such as sputa [17,18], sputum smears [19], paraffin wax-embedded tissues [20] or mummified remains of human [20]. Only one study from Brazil, where *M. africanum* is not endemic, compared spoligotyping on sputum to spoligotyping from the respective isolates [21]. Moreover, to the best of our knowledge, no study has investigated whether the proportional prevalence of MTBc lineages differs among specimens with a positive culture versus culture-negative specimens.

In this study, we determined the performance of spoligotyping on sputum ('direct spoligotying') relative to its yield on culture ('indirect spoligotyping') for genotypic characterization of MTBc, and evaluated for a potential culture bias against difficult-growers, even when incubation was prolonged to enhance detection of *M. africanum*.

Methods

Ethical considerations

This study is part of the BeniDiT study that has been approved by the national ethics committee of Benin, the Institutional Review Board of the Institute of Tropical Medicine of Antwerp, Belgium and the ethics committee of the University of Antwerp. It is registered on ClinicalTrials.gov under the registration number NCT02744469. All sputa were anonymized before laboratory analyses.

Patients/specimens and laboratory analyses

Smear-positive sputa from consecutive new TB patients diagnosed in the Centre National Hospitalier Universitaire de Pneumo-Phtisiologie in Cotonou, (Benin) were prospectively included (Fig 1), before patients initiated TB treatment. Laboratory analyses were conducted in the National Reference Laboratory for Mycobacteria (Laboratoire de Référence des Mycobactéries) in Cotonou, Benin.

Preparation of sediment aliquots, culture and sediment microscopy. When patients were identified as having (direct) acid-fast bacilli (AFB) positive pulmonary TB by fluorescent microscopy, their smear-positive sputum was included in the study. Each sputum was decontaminated using the Petroff method (15 minutes in an equal volume of 4% NaOH corresponding to 2% NaOH final concentation, neutralized with 1N HCl containing phenol red). Pellets obtained after centrifugation (3000 g, 4°C, 20 min) were resuspended with 2 mL phosphate-buffered saline. An aliquot of decontaminated sputum was used to prepare a smear, another aliquot was cultured on Löwenstein-

Jensen (LJ, 2 slants of LJ containing 0.75% glycerol and 1 slant of LJ containing 0.5% pyruvate) and incubated for 90 days (13 weeks) at 37°C before being reported as negative, and another aliquot was used for direct spoligotyping. Slides from decontaminated sputa were auramine stained and read on a fluorescence microscope for acid-fast bacilli grading as previously described [22].

DNA extraction from culture isolates. If culture was positive, DNA was extracted from culture isolates by transferring a loop of bacilli into 300 µL molecular grade water, followed by heat-inactivation for 20 minutes at 100°C [23].

DNA extraction from decontaminated sputa (sediments). DNA was extracted by the Promega Maxwell®16 Tissue DNA purification kit AS1030 [24] after a prior heat-inactivation (at 100°C for 5 minutes) [23] of 200 µL decontaminated sputum and digestion with 50 µL of 20mg/mL proteinase K (at 62°C overnight) in 200 µL of lysis buffer (10 mM tris-HCl pH 7.5, 10 mM NaCl, 10 mM EDTA, 0.5% SDS), using the Maxwell 16 machine model AS2000 ver 4.9 (Promega) [24]. DNA was eluted in 300µL Maxwell elution buffer [24]. Positive (mycobacterial sediment known to be PCR-positive) and negative (water) controls were included for DNA extraction from sputum.

Spoligotyping. Spoligotyping was done according to the standard method previously described by Kamerbeek *et al.* [25] on in-house prepared membranes for all samples (sputa and culture isolates) (Fig 1). Each PCR reaction (50µL) contained 5 µL of DNA from the sputum or culture isolate. Specimen flow and laboratory analyses are summarized in Fig 1. The reference *M. tuberculosis* strains H37Rv, *M. bovis* BCG and a negative control (molecular grade water) were included in each PCR and hybridization run.

For each patient, the spoligotype pattern from sputum was compared to the one from the respective isolate. If discrepant spacers were identified, the process (DNA extraction, amplification, hybridization for spoligotyping) was repeated from both the sputum and the isolate for confirmation.

Lineage assignment

Spoligotype patterns were recorded in an Excel file using a binary code (1 for presence of a given spacer and 0 for the absence of a given spacer). Entered profiles were verified and validated by an independent person. The persons who typed and validated the data were blinded to the spoligotype pattern of the corresponding sputum or isolate. The Excel file was loaded into the TBlineage database http://tbinsight.cs.rpi.edu/run_tb_lineage.html [26] for lineage assignment. Sub-lineages (spoligotype families) were obtained by loading the Excel file with the spoligotype patterns in the SPOTCLUST database http://tbinsight.cs.rpi.edu/run_spotclust.html [27]

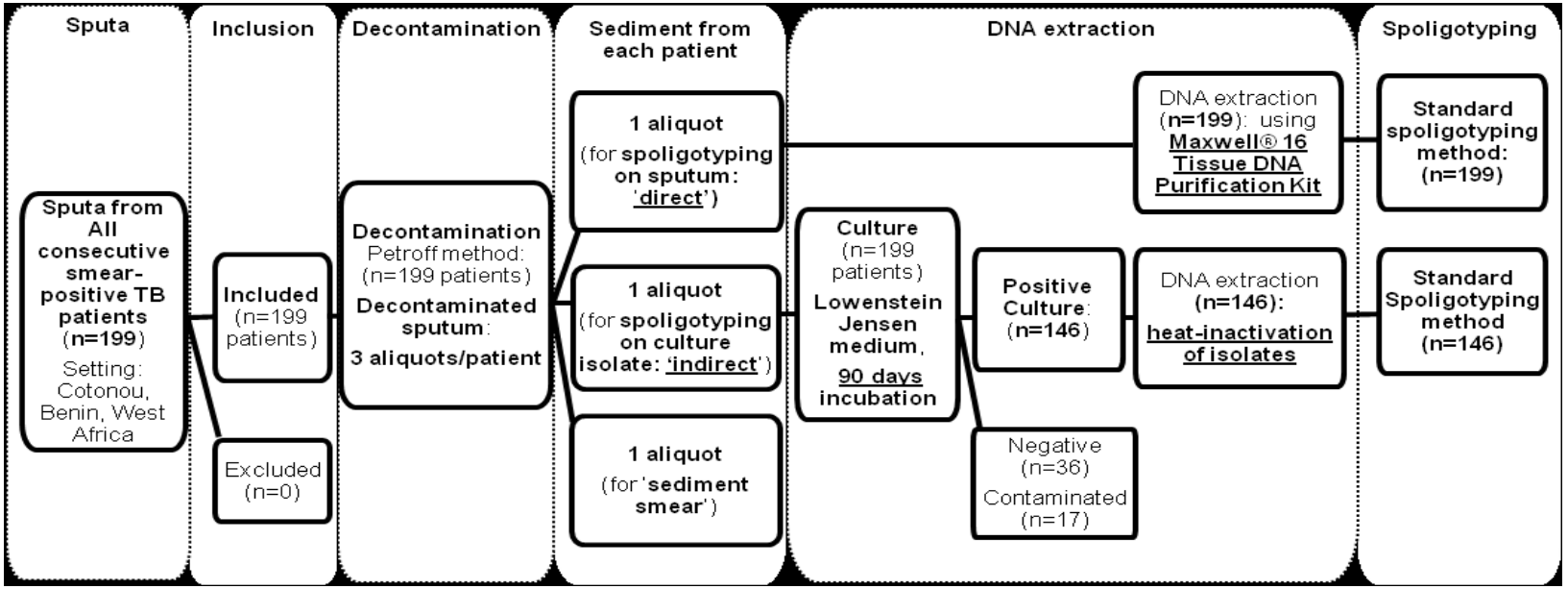

Fig 1. Patients, specimens flow diagram and laboratory analyses.

Statistical analyses

Data was analyzed using the statistical software Stata/IC 12.0 (StataCorp). The two-group proportion test or the Fisher Exact test was used to analyze independent data. Mc Nemar Chi2 test was used to compare paired proportions. Two-sided p-values were calculated and for differences in proportion, odds ratios were calculated along with 95% confidence interval. Differences were considered statistically significant when $p<0.05$.

Results

Success of direct spoligotyping versus indirect spoligotyping

From the 199 recruited TB patients and their sputum samples, 146 (73.4%) yielded a positive culture, whereas 36 (18.1%) remained negative and 17 (8.5%) were contaminated. Spoligotype patterns were obtained for all the 146 culture isolates, and for 196 of the 199 sputa, yielding an overall success for direct spoligotyping of 98.5%. All of the extraction controls and amplification/hybridization controls yielded expected results, and repeat spoligotyping for discordant results between sputum and culture confirmed the original patterns. Stratified by culture result, direct spoligotyping reached a success of 100% (53/53) for culture-negative or contaminated sputa, and 98% (143/146) for culture-positive sputa. Microscopy was negative in 6 sediments after decontamination, while all the others had positive microscopy. Of the 6 microscopy negative sediments, 3 failed direct spoligotyping and 2 others had a negative culture.

Spoligotype patterns were available for 98.5% of sputa versus 73.4% of cultures (Table 1).

Table 1. Availability of spoligotype patterns depending on the spoligotyping method used.

		Direct spoligotyping		Total	$D_{Culture - Sputum}$ (95% CI),
		Yes	No		
Indirect spoligotyping	**Yes**	143	3 [a]	146 (73.4%)	• D: **-25.1%** (-32.1 to -18.1)
	No	53 [b]	0	53 [b]	
Total		196 (98.5%)	3 [a]	199	

[a] The 3 specimens with failed direct spoligotyping were culture positive and successfully typed indirectly

[b] Culture-negative or contaminated specimens

D: Difference

Spoligotype patterns from direct spoligotyping versus indirect spoligotyping

Comparison between respective direct and indirect spoligotypes showed 94.4% (135/143) agreement. In total three types of discrepancies were observed (Fig 2): mixed infection with one pattern found in sputum and the other found in the culture isolate (n=3, discrepancy 5-7), mixed infection with overlapping spoligotype patterns in sputum (n=2, discrepancies 1 and 4), and false negative (missing) spacers in sputum (n=3, discrepancies 2, 3 and 8). Five (5) of these patterns led to inter-lineage discrepancies, and three (3) to intra-lineage discrepancies

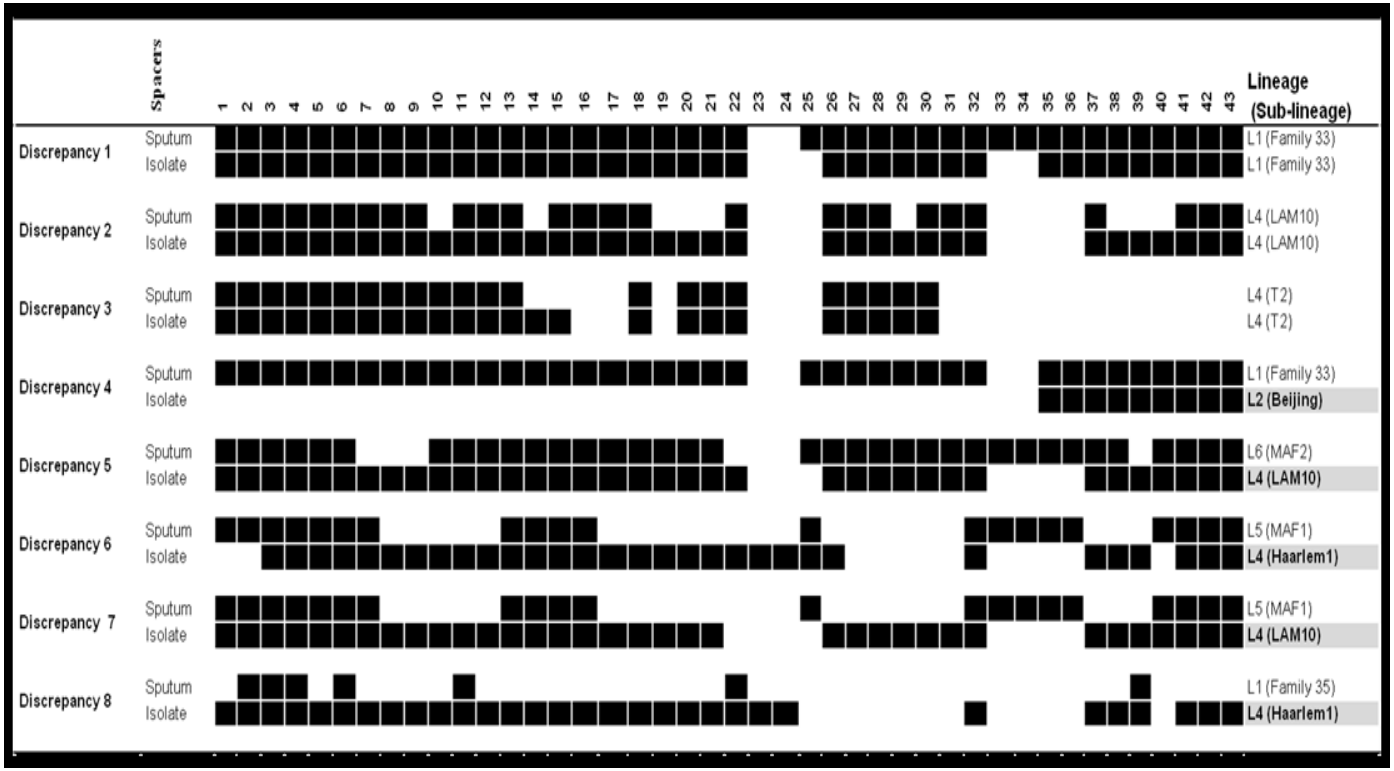

Fig 2. Discrepancies: spoligotype profile, lineage and sub-lineage. For inter-lineage discrepancies, sub(lineages) observed in isolates are shaded grey. The five inter-lineage discrepant pairs ("discrepancy 4-8") showed patterns suggestive of a simultaneous presence of ancestral and modern lineages, while these yielded only the ancestral lineage in sputum and only the modern lineage in culture. Three (discrepancy 5-7) of these five inter-lineage pairs showed this (ancestral *M. africanum* in sputum and modern Lineage 4 in culture), without any other possible explanation, while the other two (discrepancy 4 and 8) can also be interpreted as follows. Inter-lineage pair 8 and intra-lineage pairs 2 and 3 showed patterns suggestive of false negative spacers in sputum (spacer present in isolate but absent in sputum). Intra-lineage pair 1 and inter-lineage pair 4 showed patterns suggestive of overlapping spoligotype signatures in sputum (discrepancy 1 and 4) and/or in isolate (discrepancy 1). "Discrepancy 4" suggested an overlapping of Lineages 2 and 4 signatures in sputum, with only the Lineage 2 grown in culture. "Discrepancy 1" was suggestive of overlapping spoligotype signatures in sputum and in culture isolate that could be a mixture of Lineages 2 and 4.

Distribution of lineages in culture positive versus culture negative sputa

The distribution of lineages in culture-positive sputa versus directly in sputum with unsuccessful culture differed, with Lineage 5 (*M. africanum West African 1*) being significantly less prevalent among culture-positive sputa (OR=0.48 95%CI (0.24 to 0.94) p=0.033, Table 2). This association became more significant when contaminated cultures were excluded from the analysis (OR= 0.37, 95%CI (0.17 to 0.8), 21% vs 41.7%, p=0.011, Table 2). Ancestral lineages (Lineages 1, 5 and 6) were significantly less present among culture-positive sputa (OR=0.33, 95%CI (0.16 to 0.7), 37.1% vs 63.9%, p=0.004, Table 2). Lineage 4 (*M. tuberculosis Euro-American*), a modern lineage, was most overrepresented in culture-positive sputa (OR=2.81, 95%CI (1.30 to 6.03)55.2% vs 30.5%, p=0.008, Table 2).

Table 2. Effect of prior culture on spoligotyping analysis for MTBc lineage detection.

Lineages/Groups	All patients [a] % (n=199)	All sputa [b] % (n=196)	Culture positive sputa [b] % (n=143)	Culture Negative & Contaminated sputa				Culture Negative sputa (only)			
				% (n=53)	Odds ratio ($Odd_{Pos\ cult}$ / $Odd_{Neg\ \&\ Cont\ cult}$) with 95% CI	% Difference ($P_{Pos\ cult}$ - $P_{Neg\ \&\ Cont\ cult}$) with 95% CI	p-value [c]	% (n=36)	Odds ratio ($Odd_{Pos\ cult}$ / $Odd_{Neg\ cult}$) with 95% CI	% Difference ($P_{Pos\ cult}$ -$P_{Neg\ cult}$) with 95% CI	p-value [c]
Lineage 1 (Indo-Oceanic)	8.0	8.2	9.1	5.7	1.67 (0.46 to 6.04)	3.4 (-5.2 to 12.1)	0.565 *	8.3	1.1 (0.3 to 4.1)	0.8 (-9.7 to 11.2)	1 *
Lineage 2 (East Asian *Beijing*)	5.6	5.6	6.3	3.8	1.71 (0.36 to 8.09)	2.5 (-4.7 to 9.8)	0.730 *	5.6	1.14 (0.23 to 5.55)	0.7 (-8.1 to 9.5)	1 *
Lineage 3 (East African Indian)	1.0	1.0	1.4	0	-	1.4 (-1.8 to 4.6)	0.1 *	0	-	1.4 (-2.5 to 5.3)	1 *
Lineage 4 (Euro-American)	51.8	51.5	**55.2**	41.5	1.74 (0.92 to 3.29)	13.7 (-2.1 to 29.5)	0.087	**30.5**	**2.81** (1.30 to 6.03)	**24.7** (6.4 to 43.0)	**0.008**
Lineage 5 (*M. Africanum* West African 1)	25.1	25.0	**21**	**35.8**	**0.48** (0.24 to 0.94)	**-14.9** (-28.6 to -1.2)	**0.033**	41.7	**0.37** (0.17 to 0.8)	**-20.7** (-36.6 to -4.8)	**0.011**
Lineage 6 (*M. Africanum* West African 2)	8.5	8.7	7	13.2	0.49 (0.18 to 1.36)	-6.2 (-15.1 to 2.7)	0.17	13.9	0.47 (0.15 to 1.43)	-6.9 (-17.1 to 3.3)	0.182
Modern lineages (L2 + L3 + L4)	58.3	58.2	**62.9**	**45.3**	**2.05** (1.09 to 3.87)	**17.7** (2.1 to 33.2)	**0.026**	**36.1**	**3.0** (1.43 to 6.31)	**26.8** (8.7 to 44.9)	**0.004**
Ancestral lineages (L1 + L5 + L6)	41.7	41.8	**37.1**	**54.7**	**0.49** (0.26 to 0.92)	**-17.7** (-33.2 to -2.1)		**63.9**	**0.33** (0.16 to 0.7)	**-26.8** (-44.9 to -8.7)	
Other than *M. africanum* (L1 + L2 + L3 + L4)	66.3	66.3	**72.0**	**50.9**	**2.48** (1.30 to 4.72)	**21.1** (6.2 to 36.0)	**0.006**	**44.4**	**3.22** (1.55 to 6.7)	**27.6** (10.3 to 44.9)	**0.002**
***M. africanum* (L5 + L6)**	33.7	33.7	**28**	**49.1**	**0.40 (0.21 to 0.77)**	**-21.1** (-36.0 to -6.2)		**55.6**	**0.31** (0.15 to 0.65)	**-27.6** (-44.9 to -10.3)	

[a] Based on sputum results in 196 patients, and culture results in 3 patients (for whom direct spoligotyping failed: 2 'Lineage 4' strains and 1 'Lineage 5' strain). [b] Direct spoligotyping (on sputa). **L**: Lineage

[c] p-values were calculated using the two-group proportion test (independent groups). * p-values were calculated using the Fisher Exact test (independent groups).

Excluding discrepant spoligotypes between direct and indirect spoligotype analysis, the association gained further statistical significance. The odds of detecting ancestral lineages in positive-cultures was 0.30 fold (95% CI (0.14 to 0.64); p=0.0016) less in positive-cultures relative to negative cultures, especially Lineage 5 (OR=0.37 95%CI (0.17 to 0.79); p=0.010) (S1 Table). Modern lineages were inversely more represented in positive-culture specimens (OR=3.31, 95%CI (1.57 to 6.99), p=0.0016), especially Lineage 4 (OR=3.01, 95%CI (1.4 to 6.51), p=0.005) (S1 Table).

The prevalence of L1, L5, L6 tended to be higher among culture-negative specimens (respectively 8.3%, 41.7%, 13.9%; S1_Table) than in culture-positive specimens (7.4%, 20.7%, 6.7%; S1_Table). In contrast the prevalence of L2, L3, L4 tended to be lower among culture-negative specimens (5.6%, 0%, 30.5%) than in culture-positive specimens (6.7%, 1.5%, 57.0%; S1_Table). This justified the analysis in subgroup of ancestral and modern lineages. The distribution of sub-lineages (families) within Lineage 4 showed that LAM 10, LAM 9, LAM 1, T1, T2, Haarlem 1, Haarlem 2, Haarlem 3, X3 families were present in new TB patients in Cotonou. This distribution of Lineage 4 families did not differ significantly in culture-positive versus culture-negative sputa (S2 Table).

Incubation time to culture positivity across lineages

Almost all positive cultures were positive within 8 weeks of incubation, while prolonged incubation only yielded one additional positive culture. This was a Lineage 5/ *M. africanum West African 1* strain.

Among positive cultures, over half (5/9: 55.5%) of the Lineage 6/ *M. africanum West African 2* cultures became positive between 6 to 8 weeks of incubation,

whereas most of positive cultures from other lineages specimens were positive within 6 weeks: 10/11 (90.9%) for Lineage 1, 10/10 (100%) for Lineage 2, 2/2 (100%) for Lineage 3, 83/85 (97.6%) for Lineage 4 and 28/29 (95.6%) for Lineage 5.

Despite the prolonged incubation period, over a third of specimens from each *M. africanum* lineage remained culture negative (34.1% for Lineage 5 and 35.7% for Lineage 6), while for other lineages, none (Lineage 3) or fewer specimens (21.4% for Lineage 1, 16.7% for Lineage 2, 11.5% for Lineage 4) remained negative (Table 3). The sediment smear of the culture negative specimens from Lineage 1 and 2 had low AFB-grading or were negative whereas nearly all (14/15) the culture negative specimens from Lineage 5 had high smear grading (S3 Table).

Table 3. Time to culture positivity (on LJ medium) across lineages.

Lineages/Groups	Time to culture positivity [a]			Culture-negative sputa[b], n (%)	Total, n
	< 6 weeks, n	6 -8 weeks, n	> 8 - 13 weeks, n		
Lineage 1 (Indo-Oceanic)	**10**	1	0	3 (21.4)	14
Lineage 2 (East Asian *Beijing*)	**10**	0	0	2 (16.7)	12
Lineage 3 (East African Indian)	**2**	0	0	0 (0)	2
Lineage 4 (Euro-American)	**83**	2	0	11 (11.5)	96
Lineage 5 (*M. Africanum* West African 1)	**28**	0	**1**	**15 (34.1)**	44
Lineage 6 (*M. Africanum* West African 2)	4	**5**	**0**	**5 (35.7)**	14

[a] Lineages were determined using indirect spoligotyping (culture isolates)

[b] Direct spoligotyping (on sputa) used. Sputa with contaminated culture were not included

Discussion

Our results show that indirect spoligotyping provided spoligotype profiles for all 146 culture-positive specimens (73.4%), while direct spoligotyping provided spoligotyping profiles for 50 more sputa (+ 25.1% of all 199 specimens, 95% CI (18.1% to 32.1%)) that would not otherwise be genotyped in the absence of an isolate. Direct spoligotyping on sputum after semi-automated DNA extraction using Maxwell DNA tissue purification kit, has a high sensitivity (98.5% (196/199)) to detect MTBc genotypes. The 98.5% (196/199) overall availability of spoligotype profiles in our study is higher than the 90.9% (159/175) found on smear-positive sputa by Goyal et *al.* in Ghana (p=0.001) [18] and the 49.1% (28/57) found by Heyderman et *al.* in Zimbabwe [17]. This could be explained by the variability of methods used for DNA extraction from sputa and/or the variability in PCR reagents mix. The overall availability of spoligotype profiles on sputa in our study (98.5%) is also higher than the 77.7% (41/53) found by Suresh et *al.* and 90.5% (19/21) by Zanden et *al.* on smears [19,28], which likely have less mycobacterial DNA than a 200 µL sputum sample.

The fact that- within mixed infections- ancestral lineages are found with direct spoligotyping on sputum, suggests that the load of ancestral lineage bacilli *in vivo* exceeds the load of the modern lineage bacilli, with subsequent out-competition in culture by the latter. Sarkar et *al* also found that Lineage 4 grows more rapidly (in liquid medium) than other lineages including Lineage 1, an ancestral lineage [29]. Moreover, Gehre et *al.* found that Lineage 6, another ancestral lineage, grows more slowly than MTBc lineages other than *M. africanum* in liquid medium [12].

Sputum provided the most representative population distribution of lineages of the MTBc in new TB patients in Cotonou, with more TB due to ancestral lineages, including *M. africanum*. This distribution did not alter when the three isolates which sputum failed direct spoligotyping were added (two from Lineage 4 and one from Lineage 5; Table 2). The 'most true' distribution is the one combining profiles obtained directly from sputum, complemented by profiles on isolates from failed direct spoligotyping, and includes: 8.0% (16/199) for Lineage 1, 5.6% (11/199) for Lineage 2, 1% (2/199) for Lineage 3, 51.8% (103/199) for Lineage 4, 25.1% (50/199) for Lineage 5, 8.5% (17/199) for Lineage 6, or 41.7% for ancestral lineages, and 33.7% for *M. africanum* (Table 2). This distribution would have been different if smear-negative specimens were also genotyped, as it had been previously reported that *M. africanum* is more likely to be found in lower grade smear-positive specimens [30], and Lineage 6 is associated with HIV infection [31], which is in turn associated with smear-negativity [32–34].

The comparison of the distribution of MTBc lineages in a similar population, also consisting of consecutive smear-positive new pulmonary TB patients aged at least 15 years old of Cotonou in year 2005-2006 on cultured isolates [9,35], to the one obtained in our study indirectly on cultured isolates from similar patients in Cotonou 10 years later, showed that the previous prevalence of Lineage 1 (7.7%), Lineage 2 (10.3%), Lineage 3 (0%), Lineage 6 (6.2%) are similar to our findings in this study (respectively: 7.5%, 6.8%, 1.4% and 6.2%). Yet the prevalence of Lineage 4 (42.3% in year 2005-2006) has increased to 58.2% (difference: +15.9%), and Lineage 5 prevalence (30.9% in year 2005-2006) has decreased to 19.9% (difference: -11.0%). While we demonstrate that the present L5 prevalence of 19.9% on indirect genotyping is an underestimate, even the present 'true' L5 prevalence of 25.1% on direct

genotyping would constitute a decline from the L5 prevalence of 30.9% on indirect genotyping in 2005-2006. Other authors also reported a decrease of *M. africanum* [8,9].

Our results show that rates of culture isolation from smear-positive pulmonary TB patients are lower for Lineages 5 and 6 of the MTBc, despite prolonged incubation of cultures for 90 days [13]. Extending the incubation time beyond 6 weeks enhanced isolation of Lineage 6 (between 6-8 weeks) yet did not further augment the isolation rate. Ancestral lineages, especially Lineage5/*M. africanum West African 1* are 'difficult-growers' in culture relative to modern lineages, such as Lineage 4. The decreased odds of ancestral lineages to grow in culture could partly be due to culture procedures (culture medium or decontamination method) that were originally developed for modern lineages prevalent in Europe. Ofori-Anyinam et al reported that Lineage 6 as compared to Lineage 4, is more adapted to microaerobic growth [14] which may be the reason for its impaired growth on solid media such as LJ used in this study. Furthermore Gehre et al found that Lineage 6 has mutations in genes that lead to its attenuated growth in vitro[12]. Such genetic analyses need to be conducted on Lineage 5 in order to understand the reasons for its difficult growth in vitro. Further studies should also be conducted on other lineages to find out the genetic basis of their in vitro growth pattern. To the best of our knowledge, this is the first demonstration that ancestral lineages are underrepresented in positive cultures. Direct spoligotyping is thus more appropriate for unbiased determination of MTBc population structure in settings where ancestral lineages, including *M. africanum,* are common.

The implications of our findings also affect MTBc population structures generated with different typing methods, including whole genome sequencing.

Such studies tend to be culture-based, given the ongoing limitations of sequencing entire MTBc genomes directly from clinical material. While direct genome sequencing is technically feasible given sufficient coverage, in practice the associated costs are prohibitive. Studies to date have shown limited coverage, precluding SNP cut-offs for molecular epidemiological studies [36]. Optimized methods to sequence genomes directly from clinical material are thus urgently needed.

One strength of this study is the prolonged incubation time, to maximize the yield of *M. africanum* in culture. Other strengths include the paired design for the comparison of direct spoligotyping versus indirect spoligotyping and the inclusion of multiple controls and blinding of operators, and the fact that the study was conducted in a setting where *M. africanum* is prevalent. A limitation is that only LJ medium was used, and we do not know whether other medium, such as liquid medium (known to enable the growth of more non-tuberculous mycobacteria) may also favor the growth of ancestral MTBc lineages. This study was conducted only on fresh unshipped acid-fast bacilli positive sputa from new TB patients. Culture positivity may be worse if sputa had to be shipped from peripheral laboratories to a reference or central laboratory where spoligotyping can be done. Another limitation is that the number of specimens with Lineages 1, 2, 3, 6 among culture negative specimens under-powered the estimation of any difference in the prevalence of these individual lineage among culture-negative versus culture-positive specimens. So, although no evidence of such difference in prevalence among culture-negative versus – positive specimens was found in Lineages 1, 2, 3, 6 in the present study, such difference could be tested for in settings with higher prevalence of these lineages.

In conclusion, ancestral lineages especially *M. africanum* West African 1 (Lineage 5), are less likely to grow in culture, unlike modern lineages especially *M. tuberculosis* Euro-American (Lineage 4). Direct spoligotyping on sputum is effective, and saves effort and time compared to indirect spoligotyping of cultures. It has an important gain in sensitivity, especially for ancestral lineages that may not yield a positive culture, allowing a more precise unbiased determination of the population structure of the MTBc. It can also be used for specimens from patients under TB treatment and other specimens in which culture may be negative or contaminated. While differences in culture isolation technique and reliance on indirect spoligotyping may partially account for the reduction in the prevalence of *M. africanum* observed in several West African countries [8,9], comparison of our findings with the genotyping study from Cotonou 10 years ago suggests that the decline in *M. africanum* is not explained by the lower sensitivity of culture isolation. The potential decline of *M. africanum* lineages will be addressed in more depth in a larger ongoing study on the population structure of the *M. tuberculosis* complex in Benin, in which direct genotyping will be applied, given the findings presented in this manuscript. Further studies must be conducted to investigate whether culture procedures (medium, decontamination) can be optimized for growth of ancestral lineages. Additional studies should address the frequency and role, if any, of a mixed infection between an ancestral- and modern lineage in the faster spread of modern lineages [4] and disappearance of ancestral lineages [8,9].

Aknowledgments. We are grateful to Mathieu Odoun (Laboratoire de Référence des Mycobactéries, Cotonou, Benin) for his assistance for data verification and validation.

References

1. MC R, DE S, Jr, Kochi A. Global epidemiology of tuberculosis: Morbidity and mortality of a worldwide epidemic. JAMA [Internet]. 1995 Jan 18;273(3):220–6. Available from: http://dx.doi.org/10.1001/jama.1995.03520270054031
2. Gagneux S, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis. 2007;7(May):328–37.
3. Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial lineages causing pulmonary and extrapulmonary Tuberculosis, Ethiopia. Emerg Infect Dis. 2013;19(3):460–3.
4. Coscolla M, Gagneux S. Consequences of genomic diversity in mycobacterium tuberculosis. Semin Immunol [Internet]. 2014;26(6):431–44. Available from: http://dx.doi.org/10.1016/j.smim.2014.09.012
5. Brosch R, Gordon S V, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci [Internet]. 2002;99(6):3684–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=122584&tool=pmcentrez&rendertype=abstract
6. de Jong BC, Antonio M, Gagneux S. Mycobacterium africanum-review of an important cause of human tuberculosis in West Africa. PLoS Negl Trop Dis. 2010;4(9).
7. Niobe-eyangoh SN, Kuaban C, Cunin P, Thonnon J, Sola C, Rastogi N, et al. Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon. J Cinical Microbiol. 2003;41(6):2547–53.
8. Koro FK, Simo YK, Piam FF, Noeske J, Gutierrez C, Kuaban C, et al. Population dynamics of tuberculous bacilli in cameroon as assessed by spoligotyping. J Clin Microbiol. 2013;51(1):299–302.
9. Gehre F, Antonio M, Faïhun F, Odoun M, Uwizeye C, de Rijk P, et al. The First Phylogeographic Population Structure and Analysis of Transmission Dynamics of M. africanum West African 1- Combining Molecular Data from Benin, Nigeria and Sierra Leone. PLoS One. 2013;8(10):8–13.
10. Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, et al. The Guinea-Bissau family of Mycobacterium tuberculosis complex revisited. PLoS One [Internet]. 2011 Apr 20 [cited 2017 Jan 13];6(4):e18601. Available from: http://dx.doi.org/10.1371/journal.pone.0018601
11. Kato-Maeda M, Metcalfe JZ, Flores L. Genotyping of Mycobacterium tuberculosis: application in epidemiologic studies. Future Microbiol [Internet].

2011;6(2):203–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21366420%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4296029

12. Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML, Mulders W, et al. Deciphering the Growth Behaviour of Mycobacterium africanum. PLoS Negl Trop Dis. 2013;7(5).

13. Castets M. [Mycobacterium africanum (author's transl)]. Med Trop (Mars). 1979;39(2):145–8.

14. Ofori-Anyinam B, Dolganov G, Van T, Davis JL, Walter ND, Garcia BJ, et al. Significant under expression of the DosR regulon in M.??tuberculosis complex lineage 6 in sputum. Tuberculosis. 2017;104:58–64.

15. Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, et al. The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: Implications for in vivo growth. Mol Microbiol. 2005;56(1):163–74.

16. Stonebrink B. Tubercle bacilli and Pyruvic acid. Proc Tuberc Res Counc R NetherlandsTuberculosis Assoc 1957; 4467-74. (2).

17. Heyderman RS, Goyal M, Roberts P, Ushewokunze S, Zizhou S, Marshall BG, et al. Pulmonary tuberculosis in Harare, Zimbabwe: analysis by spoligotyping. Thorax. 1998;53(October 2016):346–50.

18. Goyal M, Lawn S, Afful B, Acheampong JW, Griffin G, Shaw R. Spoligotyping in molecular epidemiology of tuberculosis in Ghana. J Infect [Internet]. 1999 May 1 [cited 2017 Jan 13];38(3):171–5. Available from: http://www.journalofinfection.com/article/S0163445399902463/fulltext

19. Suresh N, Arora J, Pant H, Rana T, Singh UB. Spoligotyping of Mycobacterium tuberculosis DNA from Archival Ziehl-Neelsen-stained sputum smears. J Microbiol Methods [Internet]. 2007 Feb [cited 2017 Jan 13];68(2):291–5. Available from: http://www.sciencedirect.com/science/article/pii/S0167701206002661

20. Parwati I, Van Crevel R, Van Soolingen D, Van Der Zanden A. Application of Spoligotyping to Noncultured Mycobacterium tuberculosis Bacteria Requires an Optimized Approach [3]. J Clin Microbiol. 2003;41(11):5350–1.

21. Cafrune P, Possuelo L. Prospective study applying spoligotyping directly to DNA from sputum samples of patients suspected of having tuberculosis. Can J … [Internet]. 2009;900(October):895–900. Available from: http://www.nrcresearchpress.com/doi/abs/10.1139/W09-033

22. Rieder H, Deun A, Kam K, Kim S, Chonde T, Trébucq A, et al. Priorities for Tuberculosis Bacteriology Services in Low-Income Countries [Internet]. Tuberculosis. 2007. Available from: http://www.tbrieder.org/publications/books_english/red_book.pdf

23. Drugeon HB. Inactivation of Mycobacterium tuberculosis for DNA Typing Analysis Inactivation of Mycobacterium tuberculosis for DNA Typing Analysis. 1999;37(7):2350–2.

24. Manual T. Maxwell ® 16 DNA Purification Kits. Components. :0–17.

25. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol. 1997;35(4):907–14.

26. Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, et al. TB-Lineage: An online tool for classification and analysis of strains of Mycobacterium tuberculosis complex. Infect Genet Evol [Internet]. 2012;12(4):789–97. Available from: http://dx.doi.org/10.1016/j.meegid.2012.02.010

27. Vitol I, Driscoll J, Kreiswirth B, Kurepina N, Bennett KP. Identifying Mycobacterium tuberculosis complex strain families using spoligotypes. Infect Genet Evol. 2006;6:491–504.

28. Soolingen D Van, Schouls LM. Use of DNA Extracts from Ziehl-Neelsen-Stained Slides for Molecular Detection of Rifampin Resistance and Spoligotyping of Mycobacterium tuberculosis Use of DNA Extracts from Ziehl-Neelsen-Stained Slides for Molecular Detection of Rifampin Resistance and S. 2003;41(3):1101–8.

29. Sarkar R, Lenders L, Wilkinson K a., Wilkinson RJ, Nicol MP. Modern lineages of Mycobacterium tuberculosis exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. PLoS One. 2012;7(8):6–13.

30. Tientcheu LD, Maertzdorf J, Weiner J, Adetifa IM, Mollenkopf H-J, Sutherland JS, et al. Differential transcriptomic and metabolic profiles of M. africanum- and M. tuberculosis-infected patients after, but not before, drug treatment. Genes Immun [Internet]. 2015 Jul;16(5):347–55. Available from: http://dx.doi.org/10.1038/gene.2015.21

31. de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M and AR. NIH Public Access. FEMS Immunol Med Microbiol [Internet]. 2010;58(1):102–5. Available from: http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2922882&blobtype=pdf

32. Kibiki GS, Mulder B, Van Der Ven AJAM, Sam N, Boeree MJ, Van Der Zanden A, et al. Laboratory diagnosis of pulmonary tuberculosis in TB and HIV endemic settings and the contribution of real time PCR for M. tuberculosis in bronchoalveolar lavage fluid. Trop Med Int Heal. 2007;12(10):1210–7.

33. Chartier L, Leng C, Sire JM, Le Minor O, Saman M, Bercion R, et al. Factors

Associated with Negative Direct Sputum Examination in Asian and African HIV-Infected Patients with Tuberculosis (ANRS 1260). PLoS One. 2011;6(6):2–7.

34. Organization WHO. Recommendations for HIV-prevalent and resource-constrained settings. World Health Organization [Internet]. 2006; Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Recommendations+for+HIV-prevalent+and+resource-constrained+settings#5

35. Affolabi D, Anyo G, Faihun F, Sanoussi N, Shamputa IC RL et al. Première étude d ' épidémiologie moléculaire de la tuberculose au Bénin. Int J Tuberc Dis. 2009;13(3):317–22.

36. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ. Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer. PeerJ [Internet]. 2014;2:e585. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4179564/pdf/peerj-02-585.pdf

Supporting information

S1 Table. Effect of prior culture on spoligotyping analysis for MTBc lineage detection: Sensitivity analysis excluding all (spoligotype and sub(lineage) level) discrepancies (between direct versus indirect spoligotyping). Direct spoligotyping is used for all sputa in this comparison.

Lineages/Groups	Culture positive specimens [a, b] % (n=135)	Culture Negative & Contaminated specimens				Culture Negative specimens (only)			
		% (n=53)	Odds ratio ($Odd_{Pos\ cult}$ / $Odd_{Neg\ \&\ Cont\ cult}$) with 95% CI	Difference ($P_{Pos\ cult}$ - $P_{Neg\ \&\ Cont\ cult}$) with 95% CI	p-value [c]	% (n=36)	Odds ratio ($Odd_{Pos\ cult}$ / $Odd_{Neg\ cult}$) with 95% CI	Difference ($P_{Pos\ Cult}$ - $P_{Neg\ cult}$) with 95% CI	p-value [c]
Lineage 1 (Indo-Oceanic)	7.4	5.7	1.33 (0.35 to 5.04)	1.7 (-6.3 to 9.8)	1 *	8.3	0.88 (0.23 to 3.39)	-0.9 (-10.7 to 8.8)	0.738 *
Lineage 2 (East Asian *Beijing*)	6.7	3.8	1.82 (0.39 to 8.6)	2.9 (-4.6 to 10.4)	0.731 *	5.6	1.21 (0.25 to 5.9)	-1.1 (-7.9 to 10.1)	1 *
Lineage 3 (East African Indian)	1.5	00	-	01.5 (-1.8 to 4.8)	1 *	0	-	1.5 (-2.5 to 5.4)	1 *
Lineage 4 (Euro-American)	**57.0**	41.5	1.87 (0.98 to 3.55)	15.5 (-0.4 to 0.3)	0.055	**30.5**	**3.01** (1.4 to 6.51)	**26.5** (8.0 to 44.9)	**0.005**
Lineage 5 (*M. Africanum* West African 1)	**20.7**	**35.8**	**0.47** (0.23 to 0.94)	**-15.1** (-28.9 to -1.3)	**0.031**	**41.7**	**0.37** (0.17 to 0.79)	**-20.9** (-36.9 to -4.9)	**0.010**
Lineage 6 (*M. Africanum* West African 2)	06.7	13.2	0.47 (0.17 to 1.31)	-6.5 (-15.4 to 2.3)	0.148	13.9	0.44 (0.14 to 1.38)	-7.2 (-17.3 to 2.9)	0.160
Modern lineages (L2 + L3 + L4)	**65.2**	**45.3**	**2.26** (1.19 to 4.3)	**19.9** (4.3 to 35.5)	**0.012**	**36.1**	**3.31** (1.57 to 6.99)	**29.1** (10.9 to 47.2)	**0.0016**

Ancestral lineages (L1 + L5 + L6)	**34.8**	**54.7**	**0.44** (0.23 to 0.84)	**-19.9** (-35.5 to -4.3)		**63.9**	**0.30** (0.14 to 0.64)	**-29.1** (-47.2 to -10.9)	
Other than *M. africanum* (L1 + L2 + L3 + L4)	**72.6**	**50.9**	**2.55** (1.33 to 4.89)	**21.6** 6.6 to 36.7)	**0.005**	**44.4**	**3.31 (1.58 to 6.93)**	**28.1** (10.8 to 45.5)	**0.0015**
***M. africanum* (L5 + L6)**	**27.4**	**49.1**	**0.39** (0.2 to 0.75)	**-21.6** (-36.7 to -6.6)		**55.6**	**0.30** (0.14 to 0.63)	**-28.2** (-45.5 to -10.8)	

[a] All (8) discrepancies (between direct versus indirect spoligotyping) were excluded.

[b] Direct spoligotyping (on sputa) used.

[c] p-values were calculated using the two-group proportion test (independent groups).

[*] p-values were calculated using the Fisher Exact test (independent groups).

S2 Table. Distribution of sub-lineages (families) within Lineage 4 depending on culture result

Families (sublineages) within L4	All (L4) sputa [a] n=101	All (L4) isolates [b] n=85	Culture positive sputa [a] n=79	Culture Negative & Contaminated sputa [d]			Culture Negative sputa (only) [d]		
				n=22	% Difference ($P_{Pos\ cult}$ - $P_{Neg\ \&\ Cont\ cult}$) with 95% CI	p-value [c]	n=11	% Difference ($P_{Pos\ cult}$ - $P_{Neg\ cult}$) with 95% CI	p-value [c]
LAM 1	2	2	2	0	2.5 (-4.1 to 9.1)	1	0	2.5 (-6.8 to 11.9)	1
LAM 9	4	4	4	0	5.1 (-4.2 to 14.3)	0.574	0	5.1 (-8 to 1.8)	1
LAM 10	68 (67.3%)	55 (64.7%)	52 (65.8%)	16 (72.7%)	-6.9 (-29.2 to 15.4)	0.615	10 (90.9%)	-25.1 (-54.4 to 4.3)	0.162
T1	15 (14.8%)	13 (15.3%)	12 (15.2%)	3 (13.6%)	1.6 (-15.3 to 18.4)	1	1 (9.1%)	6.1 (16.2 to 28.4)	1
T2	1	1	1	0	1.3 (-3.4 to 6)	1	0	1.3 (-5.4 to 7.9)	1
Haarlem 1	4	6	4	0	5.1 (-4.2 to 14.3)	0.574	0	5.1 (-8.0 to 18.1)	1
Haarlem 2	1	1	1	0	1.3 (-3.4 to 6)	1	0	1.3 (-5.4 to 7.9)	1
Haarlem 3	4	2	2	2	-6.6 (-15.8 to 2.7)	0.206	0	2.5 (-6.8 to 11.9)	1
X3	2	1	1	1	-3.3 (-9.9 to 3.3)	0.390	0	1.3 (-5.4 to 7.9)	1

S3 Table. AFB-microscopy of sediments in positive and negative cultures across lineages of the MTBc

Lineages/Groups	AFB-microscopy in SEDIMENT									
	Culture positive specimens [a]					**Culture negative** specimens [b]				
	+++	**++**	**+**	**Scanty**	**Negative**	**+++**	**++**	**+**	**Scanty**	**Negative**
Lineage 1 (Indo-Oceanic)	4	5	3	-	-	-	**-**	**-**	**2**	**1**
Lineage 2 (East Asian *Beijing*)	3	5	2	-	-	-	-	1	1	-
Lineage 3 (East African Indian)		2	-	-	-	-	-	-	-	-
Lineage 4 (Euro-American)	28	36	11	4	3	1	5	4	1	-
Lineage 5 (*M. Africanum* West African 1)	5	19	5	1	1	**11**	**3**	-	-	1
Lineage 6 (*M. Africanum* West African 2)	1	7	1	-	-	1	2	1	1	-

[a] Lineages were determined using indirect spoligotyping (culture isolates)

[b] Direct spoligotyping (on sputa) used. Sputa with contaminated culture were not included.

- means 0 (zero) specimen.

Chapter 6

First insight into a nationwide genotypic diversity of *Mycobacterium tuberculosis* among previously treated pulmonary tuberculosis cases in Benin, West Africa

Dissou Affolabi[1,2], **N'Dira Sanoussi[2]**, Sergio Codo[1], Fréderic Sogbo[1], Prudence Wachinou[1], Faridath Massou[1,2], Aderemi Kehinde,[3,4], Séverin Anagonou[1,2]

[1]Faculty of Health Sciences, Abomey-Calavi University, Cotonou, Benin
[2]National Reference Laboratory for Mycobacteria, Cotonou, Benin
[3]Department of Medical Microbiology & Parasitology, College of Medicine, University of Ibadan, Ibadan, Nigeria
[4]Department of Medical Microbiology & Parasitology, University College Hospital, Ibadan, Nigeria

Can J Infect Dis Med Microbiol. 2017;2017:3276240.

doi:10.1155/2017/3276240

https://doi.org/10.1155/2017/3276240

https://www.ncbi.nlm.nih.gov/pubmed/28713434/

Conceived and designed the research: DA, **NS**, SC
Performed the experiments: **NS**, SC
Data analysis: DA, **NS**, SC
Supervision: DA, SA
Writing original draft: DA
Writing, review and editing: DA, **NS**, SC, FS, PW, FM, AK,
Contributed reagents, materials, analysis tools: DA, SA

Abstract

Background. Molecular studies on tuberculosis (TB) are rare in low-resource countries like Benin where data on molecular study on previously treated TB cases is unavailable.

Material and methods. From January to December 2014, all smear and culture positive previously treated pulmonary TB patients from all TB clinics were systematically recruited. Drug susceptibility testing as well as spoligotyping were performed on all isolates.

Results. Of the 100 patients recruited, 71 (71.0%) were relapse cases, 24 (24.0%) were failure while five (5.0%) were default cases. Resistance rate to any first-line drug was 40.0% while 12.0% of strains were multidrug resistant (MDR) and no strain was extensively drug resistant (XDR). A total of 40 distinct spoligotypes were found corresponding to a genotypic diversity of 40.0%. ST61 was the most predominant spoligotype with a prevalence of 33.0%. In all, 31 single spoligotypes and nine clusters were observed with two to 33 strains per cluster giving a clustering rate of 69.0%. Euro-American (Lineage 4) was the most prevalent (74.0%) and Lineage 2 was associated with resistance to streptomycin.

Conclusion. This first insight of genetic diversity of previously treated pulmonary TB patients in Benin showed a relatively high genetic diversity of *Mycobacterium tuberculosis*.

Key Words: *Mycobacterium tuberculosis*, spoligotyping, previously treated, Benin

Introduction

Tuberculosis (TB) remains a global public health problem. According to the World Health Organization (WHO), an estimated number of 10.4 million new cases occurred in the world in 2015 [1]. The African Region recorded the highest incidence rate, almost twice that of the world [1]. In Benin in West Africa, 4,092 cases were detected in 2015 [2].

Despite use of standardized treatment regimens and a well-established National TB program (NTP) in the country, the treatment success rate as well as the number of previously treated cases (failure, relapse and default) has remained stable over years [2]. In contrast to new cases, previously treated cases are much more likely to harbour multidrug-resistant (MDR) strains, defined as resistance to rifampicin (R) and isoniazid(H) and their characteristics may differ from those of new cases [3,4].

Molecular tools are useful for better understanding TB transmission dynamics in a given area. Nevertheless, molecular studies on TB are scarce in high incidence, low-income countries [5]. In Benin, the only molecular epidemiologic study available to our knowledge recruited only TB new cases in one city [6,7]. The scarcity of these studies in TB endemic countries is partly due to lack of resources and relative complexity of some molecular techniques. Among them, spoligotyping has the advantage of being relatively simple, inexpensive and generally sufficient as a first approach of molecular epidemiology of TB [8].

In this study, we aimed to evaluate a nationwide genotypic diversity of *Mycobacterium tuberculosis* complex strains in previously treated pulmonary TB patients in Benin, using spoligotyping technique.

Materials and methods

Setting. Benin is a country with a size of 114,763 square kilometres and an estimated population of 11 million. It has 70 TB facilities spread all over the country and a well-established National TB Program. Every year, about 4,000 TB cases including new and previously treated cases, are detected in the country [2].

Specimens. A total of 100 isolates obtained from 100 sputum samples collected from smear positive previously treated pulmonary TB patients all over the country, were sent to the National Reference Laboratory (NRL) in Cotonou for processing. Previously treated TB patients were from, relapse (n = 71), failure (n = 24) and default (n = 5) cases. Two sputum samples were collected (spot and early morning) from each patient, stored at 4^0C and sent in a cool box to the NRL within a week. Upon arrival at the NRL, the two samples were processed for culture but only one strain per patient was used for drug susceptibility testing (DST) and DNA fingerprinting. Samples were systematically collected between January and December 2014, and for each of them, demographic data was retrieved while after obtaining consent from each patient, HIV screening was performed on blood using rapid immunochromatography based tests: Alere Determine HIV-1/2 ® (Alere

Medical, Japan) was used for HIV screening while samples that were reactive were confirmed by ImmunoComb HIV 1&2 BiSpot® (Orgenics, France).

Culture and DST. Samples were decontaminated using the Petroff method and cultured on Löwenstein-Jensen (LJ) media [9]. The *M. tuberculosis* isolates (one per patient) were tested for susceptibility against rifampicin (R), isoniazid (H), streptomycin (S) and ethambutol (E) using the proportion method on LJ medium at the following concentrations: 40 µg/mL, 0.2 µg/mL, 4µg/mL and 2 µg/mL respectively [9]. Internal quality control was routinely performed while annual external quality assurance was carried out by the WHO Supranational Reference Laboratory at the Institute of Tropical Medicine, Antwerp, Belgium. In case of resistance to R, DST for 2nd line drugs was performed using the proportion method on LJ medium at the following concentrations: kanamycin (30 µg/mL), capreomycin (40 µg/mL), amikacin (40 µg/mL) and ofloxacin (2 µg/mL) [10]. All strains were stored upon routine processing at -80°C and subcultured on LJ for spoligotyping.

Spoligotyping. DNA was extracted by making a suspension of bacteria with a loop of colonies into 300 µL of molecular grade water followed by heating at 100°C for 20 minutes. Spoligotyping was performed as previously described [11]. *Mycobacterium tuberculosis* H37Rv was used as a positive control while molecular grade water served as negative control. Spoligotype patterns obtained were then translated into binary code with 1 and 0 for presence and absence of "spacer" and then entered on an Excel file. From these codes, lineages and families of strains were determined using TB lineage database http://tbinsight.cs.rpi.edu/run_tb_lineage.html [12] and the SPOTCLUST

database http://tbinsight.cs.rpi.edu/run_spotclust.htm [13] respectively. Spoligotype data were compared to the SITVIT WEB database (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/) [Demay, 2012] to determine the Spoligotype International Type (SIT) if already described.

Data analysis. Data were analyzed using EpiData 3.1. Chi-square and Fisher exact test were used to compare proportions. P-value < 0.05 was considered significant.

Results

In total, 100 viable strains (one single strain per patient) were used for spoligotyping. They were 71, 24 and 5 isolates from relapse, failure and default patients respectively. In total, 74 (74.0%) isolates were from male patients while 26 (26.0%) were from females. HIV positivity rate was 15.2%, all of whom were infected with HIV1.

Resistance pattern to first- line drugs by type of previously treated cases is presented in Table 1. Resistance rate to any first-line drug was 40.0% while 12.0% of strains were multidrug resistant (MDR). In addition, two other strains were resistant to R but not to H; one monoresistant to R and another one resistant to both R and S. Thus, resistance rate to R was 14.0%. MDR rates were 20.8% and 9.9% for failure and relapse cases respectively, while none was found among defaulters. Second-line DST results were available for nine MDR strains of which, six (66.7%) were susceptible, two (22.2%) showed resistance

to ofloxacin, one (11.1%) to kanamycin while none was resistant to both fluoroquinolones and injectable drugs. Thus, no strain was extensively drug resistant (XDR) (Table 2).

Table 1. Resistance pattern of strains to first-line drugs

Type of resistance	Failure (n=24) n (%)	Relapse (n=71) n (%)	Default (n=5) n (%)	Total (n=100) n (%)
Susceptible to all drugs	12 (50.0)	45 (63.4)	3 (60.0)	60 (60.0)
Mono-resistance				
H	1 (4.2)	1 (1.4)	0 (0.0)	2 (2.0)
S	3 (12.5)	10 (14.1)	1 (20.0)	14 (14.0)
R	0 (0.0)	0 (0.0)	1 (20.0)	1 (1.0)
E	1 (4.2)	1 (1.4)	0 (0.0)	2 (2.0)
Total	5 (20.8)	12 (16.9)	2 (40.0)	19 (19.0)
Multi-Drug resistance				
HR	0 (0.0)	2 (2.8)	0 (0.0)	2 (2.0)
HRE	1 (4.2)	1 (1.4)	0 (0.0)	2 (2.0)
HRS	1 (4.2)	1 (1.4)	0 (0.0)	2 (2.0)
HRES	3 (12.5)	3 (4.2)	0 (0.0)	6 (6.0)
Total	5 (20.8)	7 (9.9)	0 (0.0)	12 (12.0)
Other patterns				
HS	1 (4.2)	1 (1.4)	0 (0.0)	2 (2.0)
HSE	0 (0.0)	2 (2.8)	0 (0.0)	2 (2.0)
RS	0 (0.0)	1 (1.4)	0 (0.0)	1 (1.0)
ES	1 (4.2)	3 (4.2)	0 (0.0)	4 (4.0)
Total	2 (8.3)	7 (9.9)	0 (0.0)	9 (9.0)

H: Isoniazid; E: Ethambutol; S: Streptomycin; R: Rifampicin.

Table 2. Resistance patterns to second - line drugs on MDR strains

Type of resistance	MDR strains n =9 n (%)
Susceptible to all second-line drugs	6 (66.7)
Mono-resistance	
Ofloxacin	2 (22.2)
Kanamycin	1 (11.1)
Capreomycin	0
Amikacin	0
Total	**3 (33.3)**
XDR	0

XDR: extensively drug resistant

A total of 40 distinct spoligotypes were found corresponding to a genotypic diversity of 40.0%. Of these, 21 (52.5%) corresponded to spoligotypes already identified in the SITVIT database and had shared-type (ST) denominations (SIT) while 19 (47.5%) were newly found spoligotypes. ST61, ST53 and ST1 were the most predominant spoligotypes with prevalence rates of 33.0%, 13.0% and 8.0% respectively. In this study, 31 single spoligotypes and nine clusters were observed with 2 to 33 strains per cluster giving a rate of 69.0% (Table 3).

Most prevalent families were LAM 10, T1 and *M. africanum* West-African 1 with a prevalence of 46.0%, 17.0% and 12.0% respectively. For lineages, the more prevalent were Euro-American (Lineage 4), *M. africanum* West-African 1

(Lineage 5) and East-Asian (Lineage 2) with prevalence rates of 74.0%, 12.0% and 8.0% respectively. Interestingly, one strain was identified as *M. bovis* representing 1.0% of the total strains tested (Table 4).

Table 3. Strains per family

Family	Spoligotype	ST	Strains n (%)
Family33	761777767775771	U	1 (1.0%)
	777777777763771	54	1 (1.0%)
Family34	777777770000000	46	1 (1.0%)
Beijing	000000000003771	1	8 (8.0%)
CAS	703777740001171	1199	1 (1.0%)
LAM1	677777607760771	20	1 (1.0%)
LAM9	377777607760771	177	1 (1.0%)
LAM10	777777743760771	61	33 (33.0%)
	767740741760751	U	1 (1.0%)
	777770343760771	U	1 (1.0%)
	777677743760771	U	3 (3.0%)
	777770343740771	U	2 (2.0%)
	777777743460771	772	3 (3.0%)
	777777742760771	U	1 (1.0%)
	777777743760731	403	1 (1.0%)
	777777743740771	U	1 (1.0%)
T1	777777777760771	53	13 (13.0%)
	777777777760731	51	1 (1.0%)
	737777777760731	848	1 (1.0%)
	777777757760771	44	1 (1.0%)
	737777777760531	U	1 (1.0%)
T2	777417707700000	U	1 (1.0%)
T4	777740017760771	159	1 (1.0%)
Haarlem1 (H1)	777777770020731	316	1 (1.0%)
Haarlem2 (H2)	000000000020731	U	1 (1.0%)
Haarlem3 (H3)	777777777720731	49	2 (2.0%)
	777777777720771	50	2 (2.0%)
Family36	000000007760771	4	1 (1.0%)
***M. africanum* West-African 1**	774077607777071	331	3 (3.0%)
	674077717777071	U	1 (1.0%)
	774077400603031	U	1 (1.0%)
	770002607777071	U	1 (1.0%)
	574077607777071	319	1 (1.0%)
	774077600000071	U	1 (1.0%)
	374077607777031	U	1 (1.0%)
	574017607777071	U	1 (1.0%)
	774040077777071	U	1 (1.0%)
	774077777777071	438	1 (1.0%)
***M. africanum* West-African 2**	700000377777671	U	1 (1.0%)
M. bovis	000040000200000	U	1 (1.0%)

ST: shared type, U: unknown

Table 4. Strains per lineage

Lineage	Denomination	n	%
1	Indo-Oceanic	3	3.0
2	East-Asian	8	8.0
3	East-African-Indian	1	1.0
4	Euro-American	74	74.0
5	*M. africanum* West-African 1	12	12.0
6	*M. africanum* West-African 2	1	1.0
M. bovis	*M. bovis*	1	1.0
Total		**100**	**100.0**

By comparing characteristics of patients within lineages, we found no association between sex, HIV status, types of treatment and lineages however, drug-resistance particularly resistance to S was associated with lineages distribution. Strains belonging to Lineage 2 were more likely to be resistant to S than the other strains (p = 0.001) (Table 5).

Table 5. Association between lineages and resistance

Lineage	Resistance to S n (%)	No resistance to S n (%)
1	1 (33.3)	2 (66.7)
2	7 (87.5)	1 (12.5)
3	1 (100.0)	0 (0.0)
4	17 (23.0)	57 (77.0)
5	4 (40.0)	6 (60.0)
6	0 (0.0)	1 (100.0)
M. bovis	1 (100.0)	0 (0.0)
Total	**31 (31.6)**	**67 (68.4)**

S: Streptomycin

Discussion

There are still several gaps in understanding TB dynamics in Africa. For example, the reason why *M. africanum* is mainly restricted to the Western and Central parts of the continent remains unclear [5,7]. Studies using molecular tools may be useful in this respect. Unfortunately, the few molecular studies available were either limited to a city or a region or only focused on new TB cases and if previously treated cases were included, the number was usually low [15,16].

In this study, we carried out a nationwide molecular study on previously treated pulmonary TB cases detected in Benin over a period of one year. In total, 40 different spoligotypes were found, corresponding to a genotypic diversity of 40.0%. This percentage was higher than the 19.1% found by Ouassa et al in previously treated cases in Côte d'Ivoire but quite similar to 35.1% obtained on the genetic diversity in a mixed population of new and previously treated cases in Rwanda [15,16]. A genotypic diversity of 49.0% was reported in 2005 among new cases in Cotonou, the biggest city in Benin suggesting that genetic diversity was similar among new and previously treated cases [6]. However, the previous study among new cases was carried out 10 years ago and distribution of spoligotypes in new cases might have changed over time. In addition, the national figure might be different to what was obtained in Cotonou.

This study showed that the most frequent spoligotype was ST61 (33%) belonging to the Latino-American and Mediterranean (LAM) family. This finding was similar to what was previously reported in the same country in 2005 indicating that ST 61 was the most prevalent spoligotype in new cases [6]. This

same genotype was previously described to be prevalent in countries within the West African coast [17].

At a lineage level, lineage 4 was the most prevalent (74.0%). A high prevalence of lineage 4 was also found in both new and previously treated cases at a similar rate in Ethiopia (72.4%) and in Guinea (78.8%) [18,19]. In comparison with other lineages, lineage 4 appears to have certain characteristics that promote its rapid expansion. For *M. bovis*, the prevalence (1.0%) is similar to those found elsewhere in a mixed population of new and previously treated cases in Ethiopia (1.2%), Nigeria (1.0%) and Mali (0.8 %) [20,21,22]. These low proportions could be explained by the fact that *M. bovis* is usually involved in extra-pulmonary TB in humans, whereas most of these studies, including the present one, were on pulmonary TB [20,21,22].

A significant association was found between lineage 2 (Beijing strains) and resistance to streptomycin ($p = 0.001$). This same association was observed in a study of new cases in 2005 in Cotonou where an outbreak characterised by Information Geographical System was identified [23]. The positive correlation between lineage 2 and resistance to streptomycin suggests a clonal distribution of Beijing strains in Benin.

This study showed that in Benin, molecular signatures of strains causing TB retreatment cases are similar to those causing new cases. The reasons why these strains did not respond to first- line TB treatment are likely to be related to human and environmental factors than the intrinsic molecular characteristics of strains. Therefore, for TB control, national TB programs in

Benin as well as in neighbouring countries should make efforts to reduce the impact of these factors in order to decrease the number of TB retreatment cases.

In this study, the resistance rate to R was 14.0%. This rate was slightly higher than what was observed by Ade *et al* in Benin in 2013 (10%) but comparable to that of the national drug resistance survey in 2010 [24,25]. Furthermore, this rate was similar to what was reported by Homolka *et al* among previously treated cases in Sierra Leone (14.4%) but lower than 43.7% reported by Dia *et al* in Senegal [26,27]. Among MDR cases, 33.3% were pre-XDR but no XDR strain was found, contrary to findings from Burkina Faso, Ethiopia and many other countries in sub-Saharan Africa [28,29,30]. This absence of XDR strains in this study could be explained by the rigorous management of the MDR-TB program in Benin with strict application of directly observed therapy during the whole nine months' treatment course. However, the threat of XDR-TB is always present with the emergence of pre-XDR cases and a need for more vigilance cannot be over-emphasized.

In conclusion, this first insight into the genetic diversity of TB in previously treated cases in Benin showed a genetic diversity of 40.0% with most strains belonging to lineage 4, similar to previous data in new TB cases. Occurrence of retreatment cases is more likely to be related to human and environmental factors than the intrinsic molecular characteristics of strains.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

List of abbreviations

DST	**:**	Drug Susceptibility Test
E	:	Ethambutol
H	:	Isoniazid
LJ	:	Löwenstein Jensen
LAM	:	Latino American and Mediterranean
MDR	:	Multi drug resistant
NRL	:	National Reference Laboratory
R	:	Rifampicin
SIT	:	Spoligotype International Type
S	:	Streptomycin
ST	:	Shared Type
TB	:	Tuberculosis
WHO	:	World health Organization
XDR	:	Extensively drug resistant

References

1. World Health Organization. *Global tuberculosis report 2016*. WHO; Geneva, Switzerland:2016.
2. Programme National contre la tuberculose. *Rapport annuel 2015*. PNT Benin, Cotonou, Benin, 2016.
3. Caminero JA. Multidrug-resistant tuberculosis: epidemiology, risk factors and case finding. *Int J Tuberc Lung Dis.* 2010;14:382-390.
4. Rockwood N, Abdullahi LH, Wilkinson RJ, Meintjes G. Risk Factors for Acquired Rifamycin and Isoniazid Resistance: A Systematic Review and Meta-Analysis. *PLoS One.* 2015;10(9):e0139017.
5. de Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*--review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis*. 2010;4:e744.
6. Affolabi D, Anyo G, Faïhun F, Sanoussi N, Shamputa IC, Rigouts L, *et al.* First molecular epidemiological study of tuberculosis in Benin. *Int J Tuberc Lung Dis*. 2009;13:317-322.
7. Gehre F, Antonio M, Faïhun F, Odoun M, Uwizeye C, de Rijk P*, et al*. The first phylogeographic population structure and analysis of transmission dynamics of *M. africanum* West African 1--combining molecular data from Benin, Nigeria and Sierra Leone. *PLoS One*. 2013;8 :e77000.
8. Ei PW, Aung WW, Lee JS, Choi GE, Chang CL. Molecular Strain Typing of *Mycobacterium tuberculosis*: a Review of Frequently Used Methods. *J Korean Med Sci. 2016*;31:1673-1683.
9. World Health Organization. *Laboratory services in tuberculosis control. Part III. Culture.* Geneva, Switzerland: WHO, 1998.
10. World Health Organization, 2012. http://www.stoptb.org/wg/gli/assets/documents/Updated%20critical%20concentration%20table_1st%20and%202nd%20line%20drugs.pdf
11. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S,*et al*. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. 1997; 35: 907–914.
12. Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, Bennett KP. Tb-lineage. http://Tbinsight.cs.rpi.edu/about_tblineage.html
13. Vitol I., J. Driscoll, B. Kreiswirth, N. Kurepina, and K. P. Bennett, "Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes," *Infect. Genet. Evol.* 2006; 6:491–504.
14. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. 2012. SITVITWEB - A publicly available international multimarker database for studying

Mycobacterium tuberculosis genetic diversity and molecular epidemiology. *Infect Genet Evol*. 12 : 755-766
15. Ouassa T, Borroni E, Loukou GY, Faye-Kette H, Kouakou J, Menan H, et al. High prevalence of shared international type 53 among Mycobacterium tuberculosis complex strains in retreated patients from Côte d'Ivoire. PLoS One. 2012;7:e45363.
16. Gafirita J, Umubyeyi A N, Asiimwe B B. A first insight into the genotypic diversity of *Mycobacterium tuberculosis* from Rwanda. BMC ClinPathol. 2012; 12: 20.
17. Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, Van de Perre P, Carriere C, Bañuls AL. First molecular epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. J Clin Microbiol. 2007;45:921-927.
18. Garedew L, Mihret A, Mamo G, Abebe T, Firdessa R, Bekele Y and ?al. Strain diversity of mycobacteria isolated from pulmonary tuberculosis patients at DebreBirhan Hospital, Ethiopia. Int J Tuberc Lung Dis. 2013; 17(8):1076-1081.
19. Ejo M, Gehre F, Barry M D, Sow O, Bah N M, Camara M, *et al*. First insights into circulating Mycobacterium tuberculosis complex lineages and drug resistance in Guinea. Infect Genet Evol. 2015; 33: 314-319.
20. Traore, Diarra B, Dembele B P P, Somboro A M, Hammond A S, Siddiqui S, et al. Molecular strain typing of Mycobacterium tuberculosis complex in Bamako, Mali. Int J Tuberc Lung Dis. 2012; 16(7): 911-916.
21. Lawson L, Zhang J, Gomgnimbou MK, Abdurrahman ST, Moullec SL, Mohamed F, et al. A molecular epidemiological and genetic diversity study of tuberculosis in Ibadan, Nnewi and Abuja, Nigeria. Plos One. 2012; 7 (6):1-11.
22. Nuru A, Mamo G, Worku A, Admasu A, Medhin G, Pieper R, *et al*. Genetic diversity of Mycobacterium tuberculosis Complex isolated from tuberculosis patients in Bahir Dar City and its surroundings, Northwest Ethiopia. Biomed Res Int. 2015; 2015:1-9.
23. Affolabi D, Faïhun F, Sanoussi N, Anyo G, Shamputa IC, Rigouts L, et al.Possible Outbreak of Streptomycin-Resistant *Mycobacterium tuberculosis* Beijing in Benin. Emerg Infect Dis. 2009; 15: 1123–1125.
24. Affolabi D, Faihun F, Migeo G K, Adè S S, Adè G, Makpenon A,et al. Anti-tuberculosis drug resistance among new and previously treated pulmonary tuberculosis patients in Benin. 44th World Conference on Lung Health, 2013, Paris , France.
25. Ade S, Adjibodé O, Wachinou P, Toundoh N, Awanou B, Agodokpessi G, *et al.* Characteristics and Treatment Outcomes of Retreatment Tuberculosis Patients in Benin. Tuberc Res Treat. 2016; 2016:1468631.
26. Homolka S, Post E, Oberhauser B, George AG, Westman L, Dafae F, *et al*. High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. BMC Microbiol. 2008; 8: 103-111.

27. Dia ML, Sow AI, Cisse MF, Gueye Pal, Ba F, Cisse NN, *et al*. Resistance profile of mycobacteria isolated from patients undergoing retreatment in Senegal. J Infect Dis Epidemiol. 2016; 2 (2): 12-15.
28. Agonafir M, Lemma E, Wolde-Meskel D, Goshu S, Santhanam A, Girmachew F, *et al*. Phenotypic and genotypic analysis of multidrug-resistant tuberculosis in Ethiopia. Int J Tuberc Lung Dis. 2010; 14(10):1259-65.
29. Saleri N, Badoum G, Ouedraogo M, Dembele SM, Nacanabo R, Bonkoungou G. Extensively drug resistant tuberculosis, Burkina Faso. Emerg Infect Dis. 2010; 16 (5): 840-842.
30. Gehre F, Otu J, Kendall L, Forson A, Kwara A, Kudzawu S, *et al*. The emerging threat of pre-extensively drug-resistant tuberculosis in West Africa: preparing for large scale tuberculosis research and drug resistance surveillance. BMC Medicine. 2016; 14: 160-172.

Chapter 7

Mycobacterium africanum Lineage 5 is associated with Gbe ethnicity and overrepresented among new tuberculosis patients in Southern Benin

C. N'Dira Sanoussi[1,2,3], Dissou Affolabi[1], Leen Rigouts[2,3], Mathieu Odoun[1], Karamatou Arekpa[1], Ousman Bodi[1], Pim de Rijk[2], Koen Vandelannoote[4], Claudiane Adigbonon[1], Cécile Uwizeye[2], Moulikatou Ali Ligali[1], Bouke C. de Jong[2]

[1] Laboratoire de Référence des Mycobactéries, Cotonou, Benin
[2] Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium
[3] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[4] Department of Microbiology and Immunology, Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, 3000, Australia

In preparation

Conceived and designed the research: **CNS**, DA, LR, BCdJ

Performed the experiments: **CNS**, MO, KA, OB, CA, MAL, BCdJ

Data analysis: **CNS**, KV, BCdJ

Supervision: DA, LR, BCdJ; Provided specific technical training: PdR, CU

Writing original draft: **CNS**

Writing, review and editing: **CNS**, DA, LR, MO, KA, OB, PdR, KV, CA, CU, MAL, BCdJ; Contributed reagents, materials, analysis tools: DA, LR, BCdJ

Abstract

There is a suggestion that the prevalence of *M. africanum* which is restricted to West and Central Africa is decreasing. The sole molecular epidemiology investigation of tuberculosis (TB) conducted in Benin, the country with the highest prevalence of *M. africanum* West-African 1 (Lineage 5, L5) worldwide, was conducted more than 10 years ago, in new TB patients of only one town. Treatment failure and relapse/reactivation can be caused by drug-resistant MTBC strains, but also by drug-susceptible strains. Whether the lineage of the MTBC strain is associated with TB recurrence (treatment failure, relapse/reactivation) remains unknown. We conducted the first nationwide study of the molecular epidemiology of TB in Benin, to measure the unbiased population structure of *Mycobacterium tuberculosis* complex (MTBC) lineages among new and previously-treated TB patients, evaluated the implications of MTBC lineages in TB retreatment/recurrence, and finally identified risk factors associated with lineages.

In total 1530 smear-positive TB patients were prospectively included between April 2016 and January 2018. Patients' characteristics and clinical data were recorded using a questionnaire, while direct spolygotyping followed by PhyloSNP (when necessary) were used for lineage determination. GeneXpert MTB/RIF, culture and phenotypic susceptibility testing were performed.

L5, L6, and *M. bovis* were under-represented in positive cultures with a higher proportion of isolates of dysgonic nature, relative to L4. Previous TB treatment was significantly associated with older age, lower baseline body-weight and

smear-microscopy positivity grade, lower culture positivity, higher drug resistance, lower treatment success rate, lower L5 and higher L4 prevalence compared to new patients.

L5 was associated with older age, female sex, Gbe ethnicity, decreased TB relapse, and possibly acquisition of RIF resistance. In previously-treated patients, L5 was associated with increased rifampicin-resistance and a lower grade of microscopy positivity relative to L4. L6 was associated with Peulh and Bariba ethnicities relative to L4, but not with cattle contact. The distribution of lineages varied significantly by regions in Benin, with L5 associated with decreasing population density relative to L4. After 10 years, the prevalence of L1 and L5 decreased in Cotonou, whereas that of L4 increased, and that of L6 did not significantly differ.

In conclusion, the distribution of MTBC lineages significantly varied between new and previously-treated patients, with L5 and L6 associated (differently) with patient ethnicity and the population density, and with L5 and L1 being replaced by L4 in Cotonou over time. Molecular epidemiology studies should proportionally include new and previously-treated TB patients, and analyze both groups separately, applying sputum-based genotyping techniques.

Keywords: *M. tuberculosis* complex, lineage 5, *M. africanum*, direct spoligotyping, PhyloSNP, ethnicity, population density

Introduction

Tuberculosis (TB) is caused by the bacteria of the *M. tuberculosis* complex (MTBC), which comprises 7 human-adapted lineages in addition to animal-adapted lineages such as *M. bovis*. The 7 human-adapted lineages include *M. tuberculosis sensu stricto* lineages (L) L1 (Indo-oceanic), L2 (East-Asian, includes Beijing), L3 (East-African-Indian), L4 (Euro-american), *M. africanum* L5 (*M. africanum* West-African 1) and L6 (*M. africanum* West-African 2), and L7 (Ethiopia)[1]. Global phylogeography shows that L2 and L4 are most widespread, L1 and L3 have an intermediary distribution, and L5, L6 and L7 are geographically most restricted [1]. L5 and L6 constitute *M. africanum*, and are restricted to Western and Central-African countries, while L7 is restricted to Ethiopia. The distribution of L5 and L6 varies within West-Africa, with L5 mostly found in the Eastern-region (Eastern-West-Africa) while L6 is mostly found in Western-West-Africa [2].

Molecular epidemiology of diseases including tuberculosis (TB) enable a deeper understanding of disease transmission dynamics, circulating strains, and association of strains with host factors. This is useful for informing TB control for improved strategies and policy. Molecular epidemiology studies are not common in resource-limited countries as they require advanced infrastructure, not widely available in those countries. In Benin, a West-African country, the first such study was conducted in the year 2005-2006, on new TB patients of only one town (Cotonou); in this study an outbreak of a streptomycin-resistant Beijing (L2) strain was identified [3,4]. The prevalence of the different lineages in Cotonou was 7.7% for L1, 10.3% for L2, no L3, 42.3% for L4, 31% for L5 and

6% for L6, amounting to 37% for MAF (L5 + L6) [3,5], with the L5 prevalence being the highest worldwide [2,5,6]. No nationwide study has been conducted to date in Benin.

In Cameroon - an L5 endemic country - and in Guinea-Bissau and Burkina –Faso - L6 endemic countries - it was reported that these *M. africanum* lineages were decreasing in prevalence [7–10]. The L5 prevalence in Cameroon, which was estimated at 56% in 1971 (by biochemical/phenotypic identification), decreased to 9% between 1997-1998 (by molecular identification on cultured isolates) [7], and has almost disappeared from the country in 2013 [8]. However, recent (2012) studies reported L5 at 14%-33% from neighboring Nigeria regions (by molecular identification on cultured isolates) [11,12]. Furthermore, in Ghana, the L5 prevalence was stable over many years [13] and in The Gambia the L6 prevalence did not change over a 7-year period [14]. The question that arises is, first, whether methodological issues explain the apparent disappearance of *M. africanum* (L5 and L6), and, if real, whether ancestral lineages including *M. africanum* are more adapted to low density populations (as also hypothetized by Comas et al [15][16]), suggesting that they may be disappearing as population density increases.

The reasons underlying the restriction of L5 and L6 to West and Central –Africa remain unknown.

It was recently reported that the MTBC evolved in Africa from an environmental organism to a pathogen [17] that had been coevolving with humans since before they left Africa 70 thousand years ago [15,16,18,19], with L6 and L5 being the earliest (respectively 73 and 67 thousand years ago) MTBC lineages

to branch off, followed by L1, L7 and modern lineages (L2, L3, L4) [16]. L5 is mostly restricted to the countries around the Gulf of Guinea (Ghana, Togo, Benin, Nigeria) [2]. A study in Ghana showed that L5 is associated with Ewe ethnicity, which people are also present in countries of the East of Ghana, as well as people of related ethnic groups (Fon, Aja, Gen (Mina), Goun, Phla-Phera); also called Gbe languages [13,20,21]. This association of L5 with Ewe ethnicity might be a reason for its geographical restriction, and supports the hypothesis that differential host genetic susceptibility plays a key role in the geographic restriction of L5 [13].

Host factors (such as ethnicity) are known to play a role in the clinical presentation of tuberculosis. Ancestral lineages (L1, L5, L6) have reduced transmissibility compared to modern lineages [1,22,23]. Compared to L4, L6 was associated with HIV positivity [20], older age of patients, severe malnutrition, and worse disease on chest X-ray [14,24], while no difference was observed regarding the presence of a BCG scar[14]. In other studies, however, no association between L5 + L6 and HIV infection was found [25]. Furthermore, L6 is associated with a slower progression to disease compared to *M. tuberculosis sensu stricto* [26]. Although, many studies have been conducted on L6 regarding TB presentation, little is known on L5.

Treatment failure and relapse/reactivation can be caused by drug-resistant MTBC strains, but also by drug-susceptible strains. This could be due to non-adherence to treatment, or to increased host susceptibility. Whether the lineage of the MTBC strain is associated with TB recurrence (treatment failure, relapse/reactivation) remains unknown.

The techniques used for MTBC strain typing include spoligotype (CRISPR, Clustered Regulatory Short Palindromic Repeats) analysis [27,28], MIRU-VNTR (Mycobacterial Interspersed Repetitive Unit, Variable Number of Tandem Repeats)[29], IS*6110*-RFLP (Restriction Fragment Length Polymorphism) [30], regions of difference (RD) identification [31], whole genome sequencing (WGS), or lineage-specific SNP (single nucleotide polymorphism) analysis [28,31,32]. WGS has the highest resolution and can be used for transmission studies [31,33–35], yet is expensive and poorly available in resource-limited countries. Spoligotyping - despite a lower resolution for transmission studies - in most cases suffices for reliable lineage assignment [35–37], including in strains with a single or few copies of IS*6110* like *M. africanum* [27,38] It is cheaper than other techniques and more easily implementable. Most genotyping techniques are applied on positive cultures [35]. Recently, we reported that culture-based genotyping techniques biased the population structure of the MTBC, as ancestral lineages including L5 and L6 are less likely to grow in culture [39]. Interestingly, we reported that spoligotyping directly applied on sputum (direct spoligotyping) is a good alternative for unbiased determination of the MTBC population structure (**Chapter 5**) [39].

We determined the population structure of the MTBC in a nationwide cohort of enhanced TB surveillance, in new and previously-treated TB patients, evaluated the implications of MTBC lineages in TB retreatment/recurrence, identified which host and environmental factors are associated with lineage distribution, and finally compared the prevalence of L5 to the previous study from Cotonou.

Methods

Ethical considerations and study registration

This study (BeniDiT study: Benin population Diversity of TB and Implications) was approved by the national ethics committee of Benin, the institutional review board of the Institute of Tropical Medicine (ITM) of Antwerp, and the ethics committee of the University of Antwerp in Belgium. It was registered on the public registry ClinicalTrials.gov under the identifier NCT02744469. Patients were included in the study after providing informed consent, and their specimens were pseudonymized before the study's laboratory analyses.

Study design & setting, TB clinics & participants selection, specimens and questionnaire

The study was a prospective observational cohort study. New and previously-treated, microscopy-positive TB patients (aged 15 years or more) who consented were included. Sputa were collected before they started TB treatment. Patients with extra-pulmonary TB only were excluded. All TB patients were treated in TB clinics as per routine practice.

Sample size calculation

We assumed that the lineage prevalence determined among 2005-2006 isolates from new patients in Cotonou (7.7% for L1, 10.3% for L2, no L3, 42.3% for L4, 31% for L5 and 6% for L6) were representative estimates for Benin.

Assuming an L5 prevalence of 31% in new patients (Cotonou data 2005-2006), to detect a relative difference in L5 proportion of at least 30% in previously-treated patients (i.e. 21.7%) with

- a power of 80%,
- a two-sided significance level α of 5%, and
- a patient inclusion ratio of 1 previously-treated/4 new,

a total sample size of 1190 participants (238 previously-treated + 952 new patients) was required for statistical analyses (*Stata IC 12.0,* comparison of two proportions- two independent groups).

Allowing for 20% refusal/withdrawal/laboratory test failure, an adjusted total sample size of 1490 participants (298 previously-treated patients + 1192 new patients) was needed.

This sample size was also sufficient to detect a relative difference of at least 30% between MAF (L5+L6) among new (37% in Cotonou 2005-2006) and previously treated patients with a power of 89.54%, and to detect detect a relative difference of at least 30% in L5 proportion (prevalence) between urban (31% in Cotonou 2005-2006) and rural areas with a power of 87.39%.

TB clinics and patient's selection

On average (2013-2014), 233 previously-treated patients and 3104 AFB-positive new TB patients are detected each year in Benin. Therefore, all previously-treated patients detected during the recruitment period of the study were included, supplemented with a random sample of new patients by including the first 4 new patients detected after detection of a previously-treated patient.

The selection of participating TB clinics was based on a reasoned choice. To ensure a strong yet realistic (cost, time, personnel) quality control of data to be

recorded in TB clinicss, we selected a limited number of TB clinics where most TB patients are diagnosed, instead of including all TB clinicss (57 in the year 2015 when the study was being designed). With this smaller number of TB clinics, all (former) departments of Benin should be represented, a similar number of TB clinics to be included per department, and the number of participants that can be recruited during the planned recruitment period (2 years at most) in the selected TB clinics should reach at least the study sample size. So, based on the number of (new and previously-treated) patients diagnosed yearly in each of the Benin TB clinics (Table S1), the 4 largest TB clinics per former department were selected: Atlantique/Littoral, Oueme/Plateau, Mono/Couffo, Zou/Collines, Borgou/Alibori, Atacora/Donga. In 2013-2014, an average of 172 previously-treated patients (74% of previously-treated patient in Benin) and 2336 new patients were detected per year in those 24 selected TB clinics, amounting to 2508 TB patients (75% of the average total TB patients detected in Benin).

Including all previously-treated patients, to reach the total of 298 previously-treated patients, recruitment in those selected 24 TB clinics was planned to be completed after 21 months.

Patient recruitment during the study, specimen collection and questionnaire

Patients were recruited for 21 months (from April 2016 to January 2018) from the 4 biggest TB clinics of each of the 6 regions in Benin, amounting to 24 TB clinics where 75% of all TB patients in Benin are diagnosed (Table S1). All previously-treated TB patients were invited and those who consented were included in the study. When a previously-treated patient was included, the four

next new TB patients diagnosed (and consenting) were also included. Included patients responded to a questionnaire on their sociodemographic information (current and usual residence addresses, migration information, ethnicity) and clinical history (hospitalization leading to TB diagnosis, previous TB history). The presence of a BCG scar was checked, and the HIV result was recorded if done. A third sputum sample was collected from the patient (when possible) and the three sputum specimens from each patient were shipped with the completed questionnaires to the Supranational Reference Laboratory for Mycobacteria (Laboratoire de Référence des Mycobactéries, LRM) in Cotonou, Benin, for further laboratory analysis, utilizing the established shipment system for routine surveillance of drug resistance in Benin.

Study implementation and monitoring

Personnel of the participating TB clinics (clinical and laboratory staff) were trained on the study procedures (participant recruitment, consent seeking, questionnaire administration, filling in study documents, sputum shipment) prior to the study initiation. The TB clinics were also followed-up along the study through quarterly National TB Program routine supervision, telephone calls, and an interim meeting with all the participating TB clinics.

Laboratory analyses

Upon receipt of sputa in LRM, the three specimens from each participant were pooled to make one specimen per patient, before further laboratory analyses. Microscopy was performed on the pooled sputum, after which one aliquot was used for GeneXpert MTB/RIF (hereafter referred to as Xpert, Cepheid, USA) and another decontaminated. The obtained sediment was used for direct

spoligotyping and culture. When culture was positive, phenotypic identification using PNB (para-nitro-benzoic acid) and phenotypic drug susceptibility testing (pDST) were performed. Those laboratory analyses were performed on all new and previously-treated patients' specimens; except Xpert and culture that were performed on all previously-treated patients' specimens as well as on the specimens of the first 896 new TB patients enrolled in the study (April 2016 to August 2017, 16 months).

Microscopy was performed using auramine staining and read with a fluorescent microscope using the IUATLD/WHO grading scale (International Union Against Tuberculosis and Lung Diseases/ World Health Organization) [41].

Xpert MTB/RIF was performed as recommended by the manufacturer [42] directly on the sputum. The result for the presence of MTBC in the specimen was recorded as well as the susceptibility or resistance to rifampicin (RIF). When RIF resistance was identified, the TB clinic was immediately informed and the patient was initiated on treatment with the short course multidrug-resistant regimen[43].

Specimen decontamination. For each patient, the pooled sputum was decontaminated for 15 min with 2% NaOH (final concentration) and neutralized with 1N HCl containing phenol red (Petroff method [41]). The mixture was centrifuged at x 3000 g and 4°C for 20 min, and the pellet was suspended with 1.5 mL phosphate buffered saline. The sediment obtained was used for culture and spoligotyping, and the remainder was stored at -20°C.

Culture and pDST. The sediment was inoculated on two Löwenstein Jensen (LJ) medium slants (0.75% glycerol) and one LJ without glycerol but supplemented with 0.5% pyruvate, and incubated at 37°C for 13 weeks (as recommended by Castets *et al*. for the growth of *M. africanum* [44]) before reporting a negative culture. If the culture was positive, phenotypic DST was performed using the proportion method on LJ medium with 1st line drugs (all isolates), and 2nd line drugs (if the strain was resistant to RIF). The recommended critical concentrations for the drugs were used [45].

Direct spoligotyping. For DNA extraction, 200 µL of the sediment was inactivated at 95°C for 10 min [46] and then digested with 50 µL of 20 mg/mL proteinase K in 200 µL of lysis buffer (10 mM tris-HCl pH 7.5, 10 mM NaCl, 10 mM EDTA, 0.5% SDS) at 62°C overnight, as previously described [39]. Then DNA was extracted (in 300 µL Maxwell elution buffer) from the digested sample using the semi-automated Promega Maxwell 16 tissue DNA purification kit (AS1030) on the Maxwell 16 machine AS2000 version 4.9 [47]. A PCR of the direct repeat region, followed by the hybridization of amplicons on a membrane (in house prepared) was performed as previously described [27]. Spoligotype profiles were loaded in the TBlineage database [48] and TB miner [49] for lineage assignment. Discrepant lineage assignments were checked for lineage signature patterns by an experienced person in lineage assignment, and unresolvable patterns (likely containing no specific lineage signature or more than one lineage signature) were tested with the 'PhyloSNP' approach (lineage-specific SNP detection).

The **PhyloSNP** analysis is based on lineage-specific SNP identification. The test was designed for culture isolates (Carcino *et al*., manuscript submitted). We

implemented the test for culture isolates and further optimized it for application on sputum, using reference isolates from lineage 1 through 6 (BCCM/ITM mycobacteria (Belgian Coordinated collection of Microorganism mycobacteria collection) [50]), and sputa with their respective positive culture from the ITM collection. The optimized settings differed slightly from those used by Carcino (Supplement S1). After validation of the 'direct PhyloSNP' test, we analyzed the specimens for which the spoligotype-patterns were difficult to assign to a specific lineage. The test consists of two PCRs, one targeting L1-L2-L5 specific SNPs and the other for the L3-L4-L6 SNPs region, followed by Sanger sequencing (outsourced to Baseclear, The Netherlands). The sequences were aligned to the H37Rv *M. tuberculosis* reference strain using CLC sequence viewer 8 [51] and SnapGene viewer 4.2.9 (for curves pics visualization [52]), in order to identify lineage-specific SNPs as well as double peaks indicating mixed infections, if any. Specimens that were still non-assignable to a lineage after attempted PhyloSNP were classified as "unclear".

Treatment outcome, patient weight and data missing on questionnaire

The patient's weight before and after the end of the TB treatment was recorded. Treatment outcome and information unfilled on the questionnaire (missing data) were collected from patients' files in the participating TB clinics whenever possible.

Checking/inquiring for relapse within one year after TB cure, and contact with cattle

Patients received a telephone call from TB clinic personnel one year after the end of their TB treatment to ask for symptoms of relapse (cough confirmed as TB in a TB clinic), and previous or current contact with cattle. We classified patients as “with high contact with cattle”, when their current or/and previous occupation involved contact with cattle and as “with low contact with cattle” when they lived in an area where cattle circulate or/and visit areas where cattle are bred.

Preparation of data for analyses

Classification of patient’s usual residence and population density. The usual residence areas provided on the questionnaires were classified into department, district, sub-districts (“arrondissements”) and village, based on the Benin list of villages within sub-districts /districts/ departments kindly provided by the National Centre for Teledetection in Benin (SENATEL). Population density data of districts were collected from the last census in Benin (2013) [53][54]. The population size of each sub-district [53] with the area size were used to calculate the population density of each sub-district.

Classification of ethnicity. For each patient, the ethnic group of his/her father and mother were recorded in the questionnaire. The ethnicity of the patient was considered as the combination of his/her father and mother ethnicity(ies). So a patient has a single ethnicity when his/her father and mother have the same ethnicity; and a mixed ethnicity when his/her father nd mother have different ethnicity. Paients with mixed ethnicity were excluded from ethnicity analysis. Furthermore, ethnicity was grouped as Western-, Central-, or Eastern–Gbe as previously described [21]. Western-Gbe includes Ewe, Mina (Gen), and

Ouatchi ethnicity [55,56], Central-Gbe includes Adja (Aja) and Eastern-Gbe includes Adjarra (Ajra), Aizo (Ayizo), Fon, Goun (Gun), Kotafon, Mahi (Maxi), Sahoue (Saxwe), Se, Setto, Toffin, Tori (Toli), Weme, Pedah (Xwela), and Xwla ethnicity [21].

Exclusion of redundant patient inclusion. We checked whether the same patient was included in the study more than one time in the same period of time, and excluded repeat patients from analysis.

Trends of lineage prevalence over time

In order to be able to compare the sputum-based lineages prevalence from this study (direct spoligotyping), to the isolates-based prevalence among new patients in Cotonou more than 10 years ago (2005-2006; indirect spoligotyping) [3][57], we estimated what the prevalence of the previous study would be if direct spoligotyping had been used. To this end, we calculated the proportional estimates based on the sputum- and isolates-based prevalences found in the sputa with respective isolates during the direct spoligotyping validation study [39].

Statistical analyses

Statistical analyses were performed using Stata IC 12 (StataCorp). Independent categorical variables were compared using the Chi2 test. The test of two sample independent proportions was used to compare independent data, along with the odds ratio (OR) and difference (Diff) with 95% confidence interval (IC). Paired t-test was used for the comparison of paired means, and the t-test for independent mean comparison to compare weight mean across lineages. 95%

CI were calculated for t-test as well. Multivariable analysis was performed using logistic regression to adjust for potential confounders. The difference was considered significant when the p-value was below 0.05.

Results

Flow chart for participants/specimens, laboratory analyses

In total, 1530 TB patients, including 312 previously-treated and 1218 new TB patients, were included in the study. Of these, 1437 had interpretable genotypes that allowed for bacterial lineage assignment (Fig 1), and 1196 were processed for culture and Xpert.

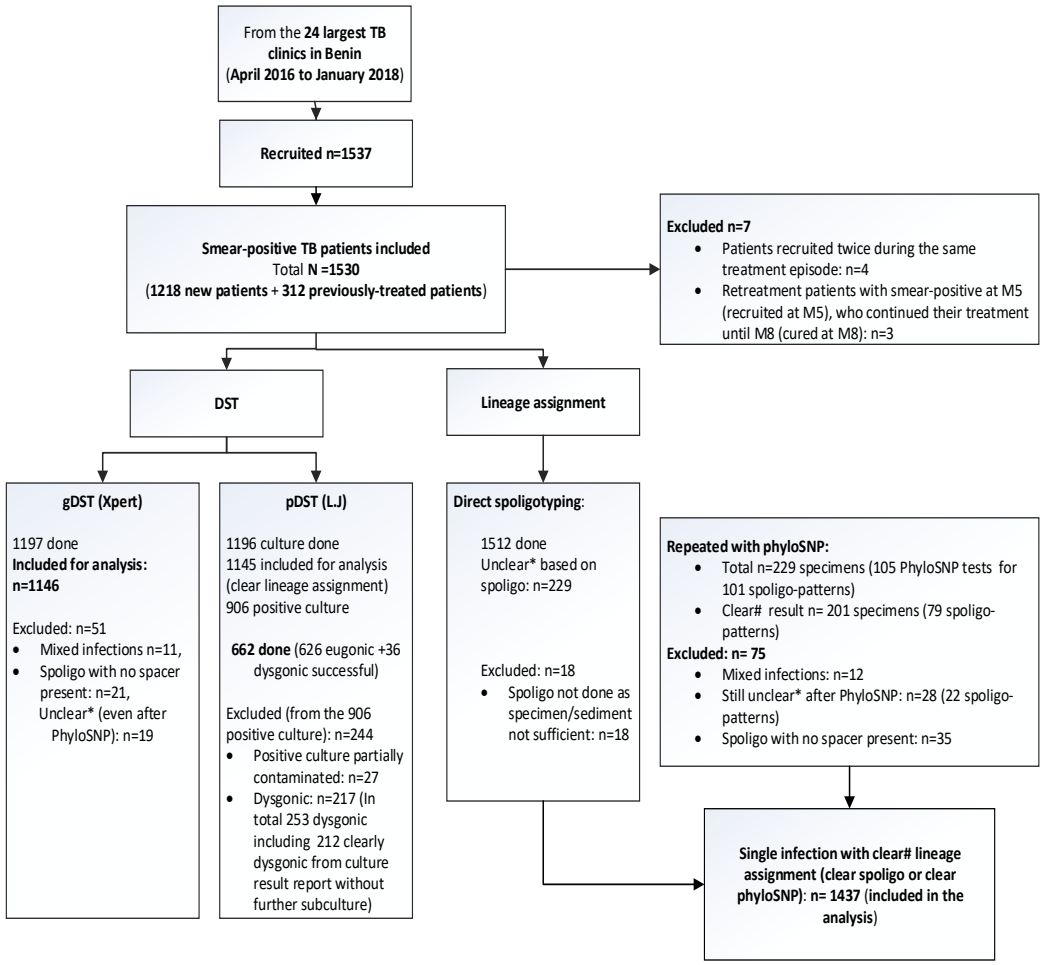

Figure 1. Flow chart for specimens and laboratory mycobacterial analyses

***unclear**: difficult lineage assignment (for spoligo: presence of many or no lineage signatures, for PhyloSNP: no lineage specific SNP identified); **# Clear**: successful lineage assignment

gDST: genotypic DST; **pDST**: phenotypic DST; **DST**: drug susceptibility testing

Comparison of previously-treated patients to new patients

Baseline characteristics

Older patients (35-84 years) were significantly over-represented among previously-treated patients (67.4% vs 53% in new patients; OR: 1.8 (1.4 to 2.4), p < 0.00001; Table 1). Also male patients were significantly over-represented in this group (78.2% vs 69%, OR: 1.6 (1.2 to 2.2), p=0.001). In addition, previously-treated patients were more likely to be hospitalized for other reasons leading to the diagnosis of a new episode of TB (29.2% among previously-treated patients vs 12.4% among new patients, OR: 2.9 (2.1 to 4), p< 0.0001). New patients and previously-treated patients were similar in terms of HIV status, presence of BCG scar, immigration (Table 1 (continued1)), ethnicity (Table 1 (continued 2)), and contact with cattle (Table 1 (continued 4)).

Baseline patient weight and difference in weight gain after treatment

The mean baseline weight was significantly lower in previously-treated (49.9 kg) compared to new (51.7 kg) patients (Diff: -1.8 (-3.1 to -0.5), p=0.006, Table 1 (continued 3)). However, the weight-gain from baseline to the end of TB treatment (programmatic 6 months' treatment for new patients and 8 months for previously-treated patients) did not significantly differ (Diff: -0.1 (-0.9 to 0.7), p=0.785, Table 1 (continued 3)), although new patients gained more weight in the first five months of treatment (5 kg, versus 4.1 kg in previously-treated, Diff: -0.9 (-1.6 to -2.4), p=0.008, Table 1 (continued 3)).

TB diagnostic results

Microscopy. The AFB (acid fast bacilli) positivity grade of patient's baseline sputum microscopy was significantly lower among previously-treated TB patients (1+positive and scanty microscopy combined 31.4% of previously-treated versus 20.8% in new patients, Diff: 10.6% (4.2 to 17.1), OR: 1.7 (1.3 to 2.4), p=0.0005, Table 1 (continued 4)).

Xpert MTB/RIF. Most patients had a positive Xpert (MTBC present) and the positivity rate was similar in previously-treated (96.4%) and new (98.2%) patients. The prevalence of RIF-resistance, however, was higher among previously-treated patients (4%) than in new (0.7%) patients (Diff: 3.3% (1 to 5.6), OR: 6.1 (2.4 to 15.9), p<0.00001, Table 1 (continued 5)).

Culture. Culture positivity was lower in previously-treated patients than in new patients (53.4% vs 65.4%, Diff: -12% (-18.4 to 5.6), p<0.0001, Table 1 (continued 6)). After repeating culture on specimens not positive (contaminated and negative) at the first attempt, the culture positivity increased (albeit still lower in previously-treated patients) and became 68% in previously-treated patients versus 81.7% in new patients (Diff: -30.4% (-35.8 to -24.9), p<0.00001). Excluding contaminated cultures, the culture positivity was also significantly lower in previously-treated patients (72.2%) compared to new (86.4%) patients (Diff: -14.3 (19.9 to -8.6), p<0.00001).

Drug resistance, treatment outcome and relapse after cure. The prevalence of resistance to RIF, isoniazid (INH) and streptomycin (STR) was significantly higher in previously-treated compared to new patients, with respectively 6%, 6.7%, 21.6% of previously-treated patients with resistant strains versus 0.8%, 2.9%,

13% of new patients (p<0.00001 for RIF, p=0.031 for INH, p=0.012 for STR, Table 1 (continued 7)). However, the prevalence of EMB resistance did not differ in previously-treated patients (1.3%) compared to new patients (1.4%, Table 1 (continued 7)).

The treatment success rate was significantly lower for previously-treated (80%) patients versus new (87.6%) patients (Diff: -7.1 (-12.7 to -1.5), OR: 0.6 (0.4 to 0.9), p=0.006, Table 1 (continued 7)). There was no difference in relapse within one year after being cured from TB between new and previously treated-patients (Table 1 (continued 7)).

Nationwide distribution of lineages in new versus previously-treated patients

Nationwide the prevalence of lineage among new patients was 1.7% for L1, 3.8% for L2, 1% for L3, 52.9% for L4, 31.1% for L5, 8.1% for L6 and 1.3% for *M. bovis*. Among previously-treated patients the nationwide prevalence was 0.7% for L1, 4.6% for L2, 1.1% for L3, 66.9% for L4, 21% for L5, 5.3% for L6 and 0.4% for *M. bovis* (Table 1 (continued 8)). L4 was the commonest but less frequent among new patients (52.9% versus 66.9% among previously-treated, Diff: 14% (7.8 to 20.2), OR: .1.8 (1.4 to 24), p< 0.00001, Table 1 (continued 8)), followed by L5 which was more frequent among new patients (31.1% versus 21% in previously-treated patients, Diff:-10.1% (-15.6 to -4.6), OR:0.6 (0.4 to 0.8), p=0.0008). *M. bovis* was detected in 1.3% of new patients and 0.4% of previously-treated patients (1.1% in the population combined). For the other lineages, the prevalence in previously-treated vs new patients was not significantly different.

Table 1. Comparison of patient characteristics, lineage distribution, diagnostic performance between new and previously treated patients (continued on following pages)

Patient characteristics		New %	Previously-treated %	Total %	Diff (95% CI)	OR (95% CI)	p
Age (years)	Total	100 (n=1201)	100 (n=307)	100 (n=1508)			
	15-24	19.2	7.8	16.8			
	25-34	27.9	24.8	27.3			
	35-44	24	29.3	25.1			
	45-54	16.5	19.9	17.2			**< 0.0001**
	55-64	8.2	9.1	8.4			
	65-74	3.3	7.8	4.2			
	75-84	1	1.3	1.1			
	15-34	**47**	32.6	44.1	**-14.5 (-20.4 to -8.5)**	**0.5 (0.4 to 0.7)**	**< 0.00001**
	35-84	53	**67.4**	55.9	**14.5 (8.5 to 20.4)**	**1.8 (1.4 to 2.4)**	
Sex	Total	100 (n=1218)	100 (n=312)	100 (n=1530)			
	Male	69	**78.2**	70.8	**9.2 (4 to 14.5)**	**1.6 (1.2 to 2.2)**	**0.001**
	Female	**31**	21.8	29.2	**-9.2 (-14.5 to -4)**	**06 (0.5 to 08)**	

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 1**)

Patient characteristics		**New %**	**Previously-treated %**	**Total %**	**Diff (95% CI)** p-treated vs new	**OR (95% CI)**	**p**
HIV status	**Total clear**	100 (n=1063)	100 (n=284)	100 (n=1347)			
(Excl indeterminate n=9)	**Positive**	11.9	15.1	12.5	3.3 (-1.3 to 7.9)	1.3 (0.9 to 1.9)	0.137
BCG scar	**Total clear**	100 (n=1042)	100 (n=266)	100 (n=1308)			
(Excl doubtful n=146)	**Yes**	59	57.9	58.8	-1.1 (-7.8 to 5.5)	1 (07 to 1.3)	0.739
Migrant in Benin	**Total**	100 (n=1198)	100 (n=306)	100 (n=1504)			
	Yes	13.6	13.7	13.4	0.1 (-4.2 to 4.4)	1 (0.7 to 1.5)	0.957
Hospitalization leading to TB diagnosis	Total	100 (n=1014)	100 (n=271)	100 (n=1285)			
	Yes	12.4	**29.2**	16	**16.7 (10.9 to 22.5)**	**2.9 (2.1 to 4)**	**< 0.0001**

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 2**)

Patient characteristic		**New %**	**P-treated %**	**Total %**	**Diff (95% CI)** p-treated vs new	**OR (95% CI)**	**P**
Ethnicity	**Total**	100 (n=980)*	100 (n=254)	100 (n=1234)			
	Single ethnicity (same for father & mother): not listed	See Table S1	See Table S1	See Table S1			0.101
	Total	100 (n=1089)*	100 (n=282)	100 (n=1371)			
	Gbe	79.7 (868)	78.7 (222)	79.5 (1090)	-9.8 (-6.3 to 4.4)	0.9 (0.7 to 1.3)	0.716
	Non-Gbe	20.3 (221)	21.3 (60)	20.5 (281)			
	Total (Gbe type)	100 (n=1041)	100 (n=272)	100 (n=1313)			0.342
	Western-Gbe	2.7 (28)	4.8 (13)	3.1 (41)	2.1 (-0.6 to 4.8)	1.8 (0.9 to 3.5)	0.078
	Central-Gbe	9 (94)	8.5 (23)	8.9 (117)			
	Eastern-Gbe	67.1 (698)	64.7 (176)	66.6 (874)			
	Non-Gbe	21.2 (221)	22.1 (60)	21.4 (281)			

*The total number of patients for the Gbe /non-Gbe comparison is higher than the individual single ethnicity comparison as patients with two individual Gbe ethnicity (one for father and one for mother) were not included among single ethnicity, and are now included in the Gbe /non-Gbe comparison as the two different individual ethnicities are both Gbe and then form a single Gbe.

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 3**)

Patient characteristics		**New %**	**P-treated %**	**Total %**	**Diff (95% CI)** p-treated vs new	**OR (95% CI)**	**p**
Baseline weight (kg)	Total	n=1016	n=258	n=1274			
	Mean Weight M0	51.7	49.9	50.2	**-1.8 (-3.1 to -0.5)**	NA	**0.006**
	St. deviation	9.9	9.4	9.5			
Gain in weight from baseline to end of treatment	**To End treatment**						
	Total	n=749	n=170	n=919			
	Mean Weight at End	55.7	58.7	56.2			
	St. dev (Mean Weight at End)	9.8	10.8	10			
	Diff Weight (End-M0)	6.1	6		-0.1 (-0.9 to 0.7)		0.785
	St. dev Weight (End-M0)	4.7	4.3				
	To M5						
	Total	n=735	n=180	n=815			
	Mean weight at M5	54.8	56.3	55.1			
	St. dev (Mean weight at M5)	9.5	10.5	9.7			
	Diff weight (M5-M0)	5	4.1		**-0.9 (-1.6 to -2.4)**		**0.008**
	St. dev Weight (M5-M0)	4.4	3.6				

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 4**)

		New **%**	**Previously-treated %**	**Total** **%**	**Diff (95% CI)** p-treated vs new	**OR (95% CI)**	**P**
Contact with cattle	Total	100 (n=330)	100 (n=80)	100 (n=410)			0.382
	High	13.6	8.8	12.7	-4.9 (-12.1 to 2.3)	0.6 (0.3 to 1.4)	0.239
	Low	21.2	18.8	20.7	-2.5 (-12.1 to 7.2)	0.9 (0.5 to 1.6)	0.626
	No	65.2	72.5	66.6	7.3 (-3.7 to 18.4)	1.4 (0.8 to 2.4)	0.211
Diagnostics							
Baseline patient' sputum (pool/mix) microscopy	**Individual grade**						
	Total	100 (n=901)	100 (n=239)	100 (n=1140)			**< 0.0001**
	+++	41.2	34.3	39.7	**-6.9 (-13.7 to -0.04)**	0.7 (0.6 to 1)	*0.054*
	++	37.3	31.4	36.1	-5.9 (-12.6 to 0.8)	0.8 (0.6 to 1)	0.091
	+	14.1	**20.1**	15.4	**6 (0.4 to 11.6)**	**1.5 (1.1 to 2.2)**	**0.022**
	Scanty	6.7	**11.3**	7.6	**4.6 (0.3 to 9)**	**1.8 (1.1 to 2.9)**	**0.016**
	Negative	0.8	**2.9**	1.2	**2.2 (-0.06 to 4.4)**	**3.9 (1.4 to 10.6)**	**0.007**
	High vs low positive						

Total	100 (n=901)	100 (n=239)	100 (n=1140)			
High pos (+++ & ++)	**78.5**	65.7	75.8	**-12.8 (-19.4 to -6.2)**	**0.5 (0.4 to 0.7)**	**< 0.00001**
Low pos (+ & scanty)	20.8	**31.4**	23	**10.6 (4.2 to 17.1)**	**1.7 (1.3 to 2.4)**	**0.0005**
Negative	0.8	**2.9**	1.2	**2.2 (-0.06 to 4.4)**	**3.9 (1.4 to 10.6)**	**0.007**

Cattle contact:

"high"= current or/and previous occupation involved contact with cattle

"low"= patient lives in an area where cattle circulate or/and visit areas where cattle are breeded.

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 5**)

Diagnostics		**New %**	**P-treated %**	**Total %**	**Diff (95% CI)**	**OR (95% CI)**	**P**
Xpert positivity	Total	100 (n=896)	100 (n=309)	100 (n=1205)			
	Positive	98.2 (880)	96.4 (298)	97.8 (1178)	-1.8 (-4 to 0.5)	0.5 (0.2 to 1.1)	0.069
Xpert RIF resistance	Total	100 (n=880)	100 (n=298)	100 (n=1178)			
	RR	0.7	**4**	1.5	**3.3 (1 to 5.6)**	**6.1 (2.4 to 15.9)**	**< 0.00001**
	RS	**99.3**	96	98.5			
Culture	**All**						
	Total	N=896	N=309	N=1205			**< 0.0001**
	Pos	**81.7** (732)	68 (210)	78.2 (942)	**-30.4 (-35.8 to -24.9)**	**0.2 (0.2 to 0.3)**	**< 0.00001**
	contaminated	5.5 (49)	5.8 (18)	5.6 (67)			
	Excluding contaminated						
	Total	N=847	N=291	N=1138			**< 0.0001**
	Pos	**86.4** (732)	**72.2** (210)	82.8 (942)	**-14.3 (-19.9 to -8.6)**	**0.4 (0.3 to 0.6)**	**< 0.00001**
	Original results (without repeat)						
	Total	N=896	N=309	N=1205			
	Pos	**65.4** (586)	53.4 (165)	62.3 (751)	**-12 (-18.4 to -5.6)**	**0.6 (0.5 to 0.8)**	**< 0.0001**

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 6**)

		New %	Previously-treated %	Total %	Diff (95% CI)	OR (95% CI)	P
Diagnostics							
Culture	**Repeated**	100 (n=203)	100 (n=77)	100 (n=280)			
	positive	72.9 (148)	58.4 (45)	68.9 (193)			0.064
	contaminated	8.9 (18)	14.3 (11)	10.4 (29)			
	Not repeated	n=639	n=232	n=925			
Isolate quality	Total	100 (n=707)	100 (n=209)	N=916			
	dysgonic	29.7 (210)	26.3 (55)	28.9 (265)	-3.4 (-10.2 to 3.5)	0.8 (0.6 to 1.2)	0.343
	eugonic	70.3 (497)	73.7 (154)	71.1 (651)			
Phenotypic DST	**PNB**	100 (n=535)	100 (n=160)	100 (n=695)			
	S	**98.3**	95	97.6	**-3.3 (-6.9 to 0.2)**	**0.3 (0.1 to 0.8)**	**0.017**
	R	1.7	**5**	2.5	**3.3 (-0.2 to 6.9)**	**3.1 (1.2 to 7.9)**	
	RIF	100 (n= 528)	100 (n=151)	100 (n=679)			
	R	0.8	**6**	1.9	**5.2 (1.4 to 9.1)**	**8.3 (2.7 to 25.8)**	**< 0.00001**

	S	**99.2**	94	98.1			
INH		N=522	N=150	N=672			
	R	2.9	**6.7**	3.7	**3.8 (-0.4 to 8)**	**2.4 (1.1 to 5.4)**	**0.031**
	S	**97.1**	93.3	96.3			

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 7**)

			New **%**	**Previously-treated %**	**Total** **%**	**Diff (95% CI)**	**OR (95% CI)**	**P**
Diagnostics	**EMB**		100 (n=511)	100 (n=149)	100 (n=660)			
		R	1.4	1.3	1.4	-0.02 (-2.1 to 2.1)	1 (0 to 4.2)	0.980
		S	98.6	98.7	98.6			
	STR		100 (n=521)	100 (n=148)	100 (n=669)			
		R	13.2	**21.6**	15.1	**8.4 (1.1 to 15.6)**	**1.8 (1.1 to 2.8)**	**0.012**
		S	**86.8**	78.4	84.9			
Treatment outcome	Total		100 (n=884)	100 (n=225)	100 (n=1109)			0.088
	Cured		**87.6**	80.4	86.1	**-7.1 (-12.7 to -1.5)**	**0.6 (0.4 to 0.9)**	**0.006**
	M5 smear-negative		5.8	8.9	6.4	3.1 (-0.9 to 7.1)	1.6 (0.9 to 2.7)	0.088
	Failed		2.7	3.6	2.9	0.8 (-1.8 to 0.03)	1.3 (0.6 to 2.9)	0.5
	Died		2	3.6	2.3	1.5 (-1.1 to 4.1)	1.8 (0.8 to 4)	0.179
	LTFU		1.9	3.6	2.3	1.6 (-1 to 4.2)	1.9 (0.8 to 4.3)	0.141

Relapse within one year after cure	Total	100 (n=394)	100 (n=82)	100 (n=476)			
	Relapsed TB	2.8	2.4	2.7	0.4 (-3.3 to 4.1)	-	0.858
	No & Non TB	97.2	97.6	97.3			

Table 1. Comparison of patient characteristics, lineage distribution, diagnosis performance by type of patient (**continued 8**)

		New **%**	**Previously-treated %**	**Total** **%**	**Diff (95% CI)** p-treated vs new	**OR (95% CI)**	**P**
Lineages	Total	100 (n=1156)	100 (n=281)	100 (n=1437)			**0.001**
	L1	1.7	0.7	1.5	-1 (-2.2 to 0.2)		0.217
	L2	3.8	4.6	4	0.8 (-1.9 to 3.5)		0.537
	L3	1	1.1	1	0.1 (-1.2 to 1.4)		0.881
	L4	52.9	**66.9**	55.7	**14 (7.8 to 20.2)**	**1.8 (1.4 to 2.4)**	**<0.00001**
	L5	**31.1**	21	29.1	**-10.1 (-15.6 to -4.6)**	**0.6 (0.4 to 0.8)**	**0.0008**
	L6	8.1	5.3	7.6	-2.8 (-5.9 to 0.3)		0.111
	M. bovis	1.3	0.4	1.1	-0.9 (-1.9 to 0.1)		0.199

Association with MTBC lineages

Patients characteristics

***Patient age*.** The combined analysis of new and previously-treated patients showed that the distribution of lineages varied across patient's age group, with L4 significantly associated with younger age of patient (p=0.002 for 15-54 vs 55-84 years old), and L5 mostly found in older patients, yet not significantly compared to all other lineages (p=0.1 for 15-54 vs 55-84 years old, Table 2.1). But relative to L4, L5 was over-represented among older (55-84 years) patients (OR=1.5 (1.1 to 2.2), p=0.016, Table 2.1). Adjusting for patient type, sex, HIV status and sputum ABF grade, confirmed that L5 prevalence is positively correlated with patient age when compared to either L4 (p=0.007) or non L5 lineages (p=0.028, Table 3).

Patient sex. For the combined analysis of new and previously-treated patient, male patients were under-represented among patients with L5 (64.6%) and L3 (46.7%) strains respectively, compared to L4 patients (74.8% for L4 patients, p=0.014 for L3 vs L4 and p=0.0002 for L5 vs L4, Table 2.2). After stratification by patient type, those trends were also observed in new and previously-treated patients. The association was similarly significant for the new patients group, but not significant for the previously-treated patients group (Table 2.2). After adjusting for patient type, age, HIV status and sputum ABF grade, L5 was strongly over-represented among female patients (adjusted OR (aOR): 1.8 (1.3 to 2.5), p<0.0001 vs L4 and aOR: 1.7 (1.3 to 2.3), p=0.001 vs non-L5 lineages, Table 3). The proportion of male patients among L6 patients was similar to that among L4 patients, also after adjusting for potential confounders (Table 4).

***HIV status*.** The combined analysis of new and previously-treated patients showed that L5 is not significantly associated with HIV-positivity. The prevalence of L5 was 36.2% among HIV-positive patients versus 28.3% among HIV-negative ones (Diff: 0.8 (-0.2 to 16.1), OR: 1.4 (1 to 2.1), p=0.045, Table 2.3). After adjusting for patient type, sex, age and sputum ABF grade, the difference was clearly not significant (p=0.288 vs L4 and p=0.355 vs non-L5 lineages, Table 3). Hospitalization (leading to the diagnosis of TB) and the presence of BCG scar were not associated with lineage.

Immigration. L5 was less common among patients who are migrants in Benin compared to non-migrants (17.3% vs 30.7%, Diff: -13.4 (-19.3 to -7.4), OR: 0.5 (0.3 to 0.7), p=0.0001). The duration of stay in Benin (less than 3 months to more than 2 years) did not affect the L5 prevalence among migrants (Table 2.4).

Contact with cattle. The distribution of lineages was not associated with reported contact with cattle (Table 2.5).

Table 2.1. Repartition of lineage by patient age group

	Age (years)									Age (broader grouping based on lineage prevalence similarity in smaller groups)		
Lineages	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75-84**	**Total**	**p**	**15-54**	**55-84**	**P (15-54 vs 55-84)**
L1	0.8 (2)	0.8 (3)	1.7 (6)	2.1 (5)	3.4 (4)	1.8 (1)	0	1.5 (21)				
L2	4 (10)	3.8 (15)	4.6 (16)	4.1 (10)	4.3 (5)	1.8 (1)	0	4 (57)				
L3	2 (5)	0.3 (1)	0.6 (2)	0.8 (2)	3.4 (4)	0	7.7 (1)	1.1 (15)				
L4	61.7 (153)	59.8 (235)	53.2 (185)	55.4 (134)	42.7 (50)	50 (28)	46.2 (6)	55.8 (791)		57.4 (707)	45.2 (84)	**0.002** **Diff: -12.3 (-19.9 to -4.6)** **OR: 0.6 (0.4 to 0.8)**
L5	25.8 (64)	28.8 (113)	29.9 (104)	28.9 (70)	35 (41)	30.4 (17)	46.2 (6)	29.3 (415)		28.5 (351)	34.4 (64)	0.1***** 5.9 (-1.4 to 13.2) 1.3 (0.9 to 1.8)
L6	4.4 (11)	6.4 (25)	8.6 (30)	7.4 (18)	8.6 (10)	16.1 (9)	0	7.3 (103)				
M. bovis	1.2 (3)	0.3 (1)	1.4 (5)	1.2 (3)	2.6 (3)	0	0	1.1 (15)	**0.03**			
Total	**100 (248)**	**100 (393)**	**100 (348)**	**100 (242)**	**100 (117)**	**100 (56)**	**100 (13)**	**100 (1417)**		**100 (1231)**	**100 (186)**	
New patients	100 (n=225)	100 (n=324)	100 (n=266)	100 (n=186)	100 (n=92)	100 (n=36)	100 (n=11)	1140 (n=100)				
L1	0.9 (2)	0.9 (3)	2.3 (6)	2.2 (4)	4.4 (4)	0	0	1.7 (19)				
L2	3.6 (8)	4 (13)	4.1 (11)	3.2 (6)	5.4 (5)	2.8 (1)	0	3.9 (44)				
L3	2.2 (5)	0	0.8 (2)	0.5 (1)	4.4 (4)	0	0	1.1 (12)				
L4	61.8 (139)	58 (188)	48.9 (130)	51.6 (96)	38 (35)	36.1 (13)	45.5 (5)	53.2 (606)				
L5	26.7 (60)	29.9 (97)	32.3 (86)	32.3 (60)	35.9 (33)	41.7 (15)	54.6 (6)	31.3 (357)				
L6	3.7 (8)	6.8 (22)22	9.8 (26)	9.1 (17)	8.7 (8)	19.4 (7)	0	7.7 (88)				

M. bovis	1.3 (3)	0.3 (1)	1.9 (5)	1.1 (2)	3.3 (3)	0	0	1.2 (14)	**0.002**	
Previously-treated	100 (n=23)	100 (n=69)	100 (n=82)	100 (n=56)	100 (n=25)	100 (n=20)	100 (n=2)	100 (n=277)		
L1	0	0	0	1.8 (1)	0	5 (1)	0	0.7 (2)		
L2	8.7 (2)	2.9 (2)	6.1 (5)	7.1 (4)	0	0	0	4.7 (13)		
L3	0	1.5 (1)	0	1.8 (1)	0	0	50 (1)	1.1 (3)		
L4	60.9 (14)	68.1 (47)	67.1 (55)	67.9 (38)	60 (15)	75 (15)	50 (1)	66.8 (185)		
L5	17.4 (4)	23.2 (16)	22 (18)	17.9 (10)	32 (8)	10 (2)	0	20.9 (58)		
L6	13 (3)	4.4 (3)	4.9 (4)	1.8 (1)	8 (2)	10 (2)	0	5.4 (15)		
M. bovis	0	0	0	1.8 (1)	0	0	0	0.4 (1)		
									< 0.0001	

*For all patients, **older patients** were significantly **over-represented in** patients with **L5** (15.4%, 64/415) **relative to** patients with **L4** strains (10.6%, 84/79); **Diff=4.8%** (0.7 to 8.9), **OR=1.5** (1.1 to 2.2), **p=0.016.**

p-treated : previously treated patients

Table 2.2. Repartition of lineage by patient sex

All patients						
Lineages		**Sex**				
	Male	**Female**	**Total**	**P**	**Diff (95% CI)**	**OR (95% CI)**
L1	1.8 (18)	1 (4)	1.5 (22)		in male	male
L2	3.4 (35)	5.3 (22)	4 (57)			
L3	0.7 (7)	**1.9 (8)**	1 (15)	**0.014**	**-28.1 (-53.5 to -2.7)** Male: **46.7** vs 74.8	**0.3 (0.1 to 08)**
L4	**58.4 (598)**	48.9 (202)	55.7 (800)	**-**	**-**	**-**
L5	26.4 (270)	**35.8 (148)**	29.1 (418)	**0.0002**	**-10.2 (-15.6 to -4.7)** Male: **64.6** vs 74.8	**0.6 (0.5 to 0.8)**
L6	8.2 (84)	6.1 (25)	7.6 (109)	0.6	2.3 (-6.1 to 10.7) Male: 77.1 vs 74.8	1.1 (0.7 to 1.8)
M. bovis	1.2 (12)	1 (4)	1.1 (16)	**0.001**		
Total	100 (1024)	100 (413)	100 (1437)			
New patients						
L1	2 (16)	1.1 (4)	1.7 (20)			
L2	3.2 (26)	5.1 (18)	3.8 (44)			
L3	0.6 (5)	**2** (7)	1 (12)	**0.018**	**-31 (- 59.2 to -2.9)** Male: **41.7** vs 72.7	**0.3 (0.1 to 0.8)**
L4	**55.4** (445)	47.3 (167)	52.9 (612)	-	-	-
L5	28.5 (229)	**36.8** (130)	31.1 (359)	**0.004**	**8.9 (-15 to -2.8)** Male: **63.8** vs 72.7	**0.7 (0.5 to 0.9)**

L6	8.8 (71)	6.5 (23)	8.1 (94)	0.566	2.8 (6.6 to -12.2) Male:75.5 vs 72.7	1.2 (0.7 to 1.9)
M. bovis	1.4 (11)	1.1 (4)	1.3 (15)	**0.007**		
Total	100 (803)	100 (353)	100 (1156)			
Previously-treated						
L1	0.9 (2)	0	0.7 (2)			
L2	4.1 (9)	6.7 (4)	4.6 (13)			
L3	0.9 (2)	1.7 (1)	1.1 (3)	0.518	-14.7 (-68.3 to 38.9) Male: **66.7** vs 81.4	-
L4	69.2 (153)	58.3 (35)	66.9 (188)	-	-	-
L5	18.6 (41)	30 (18)	21 (59)	*0.052*	-11.9 (-24.9 to 1.1) Male:**69.5** vs 81.4	0.5 (0.3 to 1)
L6	5.9 (13)	3.3 (2)	5.3 (15)	0.61	5.3 (-12.8 to 23.4) Male: 86.7 vs 81.4	-
M. bovis	0.5 (1)	0	0.4 (1)	0.410		
Total	100 (221)	100 (60)	100 (281)			

Table 2.3. Repartition of lineage by HIV status, BCG-scar, hospitalization (leading to TB diagnosis)

Lineages	HIV %(n)			BCG scar %(n)			Hospitalized prior the diagnosis of TB % (n)		
	Positive	**Negative**	P	**Yes**	**No**	**P**	**Yes**	**No**	**P**
ALL PATIENTS	100 (n=149)	100 (n=1117)		100 (n=720)	100 (n=506)		100 (n=193)	100 (n=1008)	
L1	0.7 (1)	1.6 (18)		1.3 (9)	1.2 (6)		2.1 (4)	1.4 (14)	
L2	4.7 (7)	3.6 (40)		3.9 (28)	4.7 (24)		3.6 (7)	3.8 (38)	
L3	1.3 (2)	1.2 (13)		1.1 (8)	1.2 (6)		1.6 (3)	0.9 (9)	
L4	49 (73)	56.5 (631)	0.08 Diff: -7.5 (-16 to 1) OR: 0.7 (0.5 to 1)	57.1 (411)	53.2 (269)	0.174 Diff: 3.9 (-1.7 to 9.6) OR: 1.2 (0.9 to 1.5)	57.5 (111)	56.9 (573)	
L5	**36.2** (54)	**28.3** (316)	**0.045*** Diff: 0.8 (-0.2 to 16.1) OR:1.4 (1 to 2.1)	27.4 (197)	31 (157)	0.163 Diff: -3.7 (-8.8 to 1.5) OR:0.8 (0.7 to 1.1)	25.4 (49)	28.4 (286)	
L6	6.0 (9)	7.8 (87)		7.9 (57)	7.9 (40)		8.8 (17)	7.6 (77)	
M. bovis	2.0 (3)	1.1 (12)		1.4 (10)	0.8 (4)		1 (2)	1.1 (11)	
			0.327			0.721			0.916
NEW	100 (n=113)	100 (n=896)		100 (n=576)	100 (n=410)		100 (n=121)	100 (n=839)	
L1	0.9 (1)	1.8 (16)		1.4 (8)	1.5 (6)		2.5 (3)	1.6 (13)	
L2	4.4 (5)	3.5 (31)		3.7 (21)	4.9 (20)		3.3 (4)	3.6 (30)	
L3	0.9 (1)	1.2 (11)		1 (6)	1.2 (5)		0.8 (1)	1 (8)	
L4	46 (52)	53.5 (479)	0.135	54.5 (314)	50.2 (206)	0.186	49.6 (60)	54.7 (459)	
L5	38.9 (44)	30.5 (273)	0.068	29.3 (169)	32.9 (135)	0.229	33.1 (40)	29.7 (249)	
L6	6.2 (7)	8.4 (75)		8.5 (49)	8.3 (34)		9.1 (11)	8.2 (69)	
M. bovis	2.7 (3)	1.2 (11)		1.6 (9)	1 (4)		1.7 (2)	1.3 (11)	
			0.396			0.755			0.946
P-TREATED	100 (n=36)	100 (n=221)		100 (n=144)	100 (n=96)		100 (n=72)	100 (n=169)	
L1	0	0.9 (2)		0.7 (1)	0		1.4 (1)	0.6 (1)	
L2	5.6 (2)	4.1 (9)		4.9 (7)	4.2 (4)		4.2 (3)	4.7 (8)	
L3	2.8 (1)	0.9 (2)		1.4 (2)	1 (1)		2.8 (2)	0.6 (1)	
L4	58.3 (21)	68.8 (152)	0.215	67.4 (97)	65.6 (63)	0.778	70.8 (51)	67.5 (114)	
L5	27.8 (10)	19.5 (43)	0.253	19.4 (28)	22.9 (22)	0.516	12.5 (9)	21.9 (37)	

L6	5.6 (2)	5.4 (12)			5.6 (8)	6.3 (6)			8.3 (6)	4.7 (8)		
M. bovis	0	0.5 (1)			0.7 (1)	0			0	0		
			0.794				0.932					0.316

*****For all patients, the proportion of **HIV-positive patients** among patients with **L5** (14.6%, 54/370) was **not** significantly different from that of patients with **L4** (10.4%, 73/704) strains; Diff=4.2% **(-0.002 to 8.5)**, OR=1.5 (**1.01 to 2.1**)**, p=0.042;**

p-treated: previously-treated TB patients

Table 2.4. Repartition of lineage among migrants versus non-migrants

Lineages	Migrated or not			Migrant (duration of stay in Benin)				Non-migrant	p
	Migrant % (n)	Non-migrant % (n)	P	< 3 months % (n)	3-11 months % (n)	12-23 months % (n)	≥ 24 months % (n)		
ALL PATIENTS									
L1	2.1 (4)	1.5 (18)		0	0	9.1 (1)	3.5 (3)	1.5 (18)	
L2	3.7 (7)	4.1 (50)		1.8 (1)	0	18.2 (2)	4.6 (4)	4.1 (50)	
L3	0.5 (1)	1.1 (13)		0	0	0	0	1.1 (13)	
L4	**67 (128)**	54.1 (662)	**0.0009** **Diff: 12.9 (5.7 to 20.1)** **OR: 1.7 (1.2 to 2.4)**	69.4 (39)	73.3 (11)	72.7 (8)	65.5 (57)	**54.1 (662)**	
L5	17.3 (33)	**30.7 (375)**	**0.0001** **Diff: -13.4 (-19.3 to -7.4)** **OR: 0.5 (0.3 to 0.7)**	17.9 (10)	13.3 (2)	0	17.2 (15)	**30.7 (375)**	
L6	8.4 (16)	7.5 (92)	0.679 Diff: 0.8 (-3.3 to 5.1) OR: 1.1 (0.6 to 1.9)	7.1 (4)	13.3 (2)	0	9.2 (8)	7.5 (92)	
M. bovis	1.1 (2)	1.1 (13)	**0.012**	3.6 (2)	0	0	0	1.1 (13)	**0.025**
Total	**100 (191)**	**100 (1223)**		**100 (56)**	**100 (15)**	**100 (11)**	**100 (87)**	**100 (1223)**	

Table 2.5. Repartition of lineage in regards to contact with cattle

Lineages	Contact with cattle			Total	P
	High % (n)	Low % (n)	No % (n)		
ALL PATIENTS					
L1	6.5 (3)	0	1.2 (3)	1.6 (6)	
L2	4.4 (2)	1.4 (1)	3.2 (8)	3 (11)	
L3	2.2 (1)	0	1.2 (3)	1.1 (4)	
L4	47.8 (22)	57.5 (42)	50.6 (126)	51.6 (190	0.268
L5	23.9 (11)	28.8 (21)	25.3 (63)	25.8 (95)	
L6	15.2 (7)	12.3 (9)	16.5 (41)	15.5 (57)	
M. bovis	0	0	2 (5)	1.4 (5)	
Total	**100 (46)**	**100 (73)**	**100 (249)**	**100 (368)**	

Ethnicity *(single for father and mother)*. Lineage was strongly associated with ethnicity ($p < 0.0001$, Table S2). L5 was strongly associated with the Gbe ethnic continuum (33.2% among Gbe vs 17.2% among non-Gbe patients, $p<0.00001$). Conversely, L1, L6 and *M. bovis* were significantly under-represented among Gbe-patients ($p=0.00004$ for L1, $p=0.0001$ for L6 and $p=0.0003$ for *M. bovis*, Table 2.6). Stratification in Western-, Central- and Eastern-Gbe, showed that L5 was strongly associated with Central- ($p < 0.00001$) and Eastern-Gbe ($p < 0.00001$), but not Western-Gbe ($p=0.507$) which includes the Ewe ethnic group. The repartition of L5 by type of Gbe ethnicity across residence department of patients showed that in Littoral and Atlantique (2 southern departments of Benin), L5 was strongly associated with Eastern-Gbe ($p=0.0009$) in Littoral, yet no association was seen between L5 and any type of the Gbe ethnic continuum (Table S3) in Atlantique. Overall, in Atlantique L5 was not associated with Gbe ethnicity (38.2% among Gbe patients versus 33.3% among non-Gbe patients, $p=0.7$, data not shown); but in Littoral L5 was associated with Gbe ethnicity (27.1% among Gbe patients versus 8.6% among non-Gbe patients, $p=0.02$, data

not shown). However, Atlantique is one of the Benin departments with the highest prevalence of L5 (36.9% in Atlantique vs 23.5% in Littoral, Table S4). The two departments differ in population density. Also, most Gbe people live in the South of Benin where L5 is most prevalent (Fig 2 and 3). We then adjusted for those potential confounders. After adjusting for population density of patient residence area and residence department (region) of patient (Southern region: Littoral, Atlantique, Mono, Couffo, Oueme, Plateau, Zou; Central region: Collines; Northern region: Borgou, Alibori, Atacora, Donga), the association of L5 with Gbe ethnicity was confirmed (relative to L4 aOR: 1.7 (1.1 to 2.9), p=0.031; vs non-L5 lineages aOR: 1.8 (1.1 to 2.9), p=0.023; Table 5). Similarly, L5 compared to non-L5 lineages was confirmed significantly associated with Eastern-Gbe and Central-Gbe, but not Western-Gbe; independent of population density of patient residence and geographical situation of patient residence (for Eastern-Gbe aOR: 1.8 (1.1 to 3), p=0.021; for Central-Gbe aOR: 2 (1.1 to 3.8), p=0.028; Table 5). Compared to L4, L5 was again significantly associated with Eastern-Gbe (aOR: 1.7 (1.05 to 2.9), p=0.032), but not with Central-Gbe (aOR: 1.8 (0.98 to 3.5), p=0.06) or Western-Gbe (p=0.802, Table 5). Comparing the individual ethnicities, in the univariate analysis L5 was significantly over-represented in 8 ethnicities relative to L4 and 13 ethnicities relative to non-L5 lineages (Table 6). After adjusting for population density of patient residence, and geographical situation of residence department (South, Centre, North), L5 was still significant over-represented among 4 ethnicities relative to L4 (Aizo, Sahoue, Weme, Kotafon) and among the 13 ethnicities relative to non-L5 lineages (Table 6).

L6 was strongly associated with the ethnic groups Bariba (p< 0.00001, OR: 7.6 (2.7 to 12.3)), Peulh (also called Fulani, p < 0.00001, OR: 5 (2.4 to 10.1)), Gando (p=0.004 OR: 8.6 (0 to 40)), Ahoussa (p=0.007, OR: 4.2 (1.4 to 12.3)), Waama (p=0.034, OR: 4.8 (0 to 20.8), Table 2.7 and Table S2). After adjusting for population density of patient residence, and geographical situation of residence department, L6 remained over-represented among patients of Peulh (Fulani, aOR: 3.6 (1.3 to 9.9), p=0.012) and Bariba (aOR: 5 (1.1 to 23.2), p=0.041) ethnicities relative to L4 (Table 6).

Table 2.6. Repartition of lineage by ethnicity (Gbe versus non-Gbe)

	Ethnicity (father and mother)	Lineages							Total % (n)	p
		L1 % (n)	L2 % (n)	L3 % (n)	L4 % (n)	L5 % (n)	L6 % (n)	M. bovis % (n)		
ALL PATIENTS										
	Gbe	0.9 (9)	4.5 (46)	0.7 (7)	54.3 (559)	**33.2** (342)	5.8 (60)	0.6 (6)	100 (1029)	**< 0.0001**
	Non-Gbe	**3.8** (10)	2.7 (7)	1.9 (5)	58.8 (154)	17.2 (45)	**13** (34)	**2.7** (7)	100 (262)	
	P (Gbe vs Non-Gbe)	**0.0004** **Diff: -2.9 (-5.3 to -0.6)** **OR: 0.2 (0.1 to 0.5)**	0.190 Diff: 1.8 (-0.5 to 4.1) OR: 1.7 (0.8 to 3.7)			**< 0.00001** **Diff: 16.1 (10.7 to 21.5)** **OR: 2.4 (1.7 to 3.4)**	**0.0001** **Diff: -7.1 (-11.5 to -2.8)** **OR: 0.4 (0.3 to 0.6)**	**0.003** **Diff: -2.1 (-4.1 to -0.1)** **OR: 0.2 (0.1 to 0.6)**		
	Total	1.5 (19)	4.1 (53)	0.9 (12)	55.2 (713)	30 (387)	7.3 (94)	1 (13)	100 (1291)	
Single ethnicity (same ethnicity for father and mother)	**Western Gbe**	**0**	**8.1 (3)**	0	59.5 (22)	21.6 (8)	**10.8 (4)**	0	100 (37)	**< 0.0001**
	Central Gbe	**0**	1.8 (2)	0	54.5 (61)	**37.5 (42)**	4.5 (5)	1.8 (2)	100 (112)	
	Eastern Gbe	1.1 (9)	**4.5 (37)**	0.6 (5)	54.6 (450)	**33.3 (275)**	5.5 (45)	0.5 (4)	100 (825)	
	Non-Gbe	**3.8 (10)**	2.7 (7)	1.9 (5)	58.8 (154)	17.2 (45)	**13 (34)**	**2.7 (7)**	100 (262)	
	P (Western-Gbe vs Non-Gbe)	0.227 -3.8 (-6.1 to -1.5) OR: 0 (0 to 2.7)	0.085 Diff: 5.4 (-3.6 to 14.4) OR: 3.2 (0.9 to 12)			0.507 Diff: 4.4 (-9.6 to 18.5) OR: 1.3 (0.6 to 3.1)	0.711 Diff: -2.2 (-13 to 8.6) OR: 0.8 (0.3 to 2.3)	0.314 Diff: -2.7 (-4.6 to -0.7) OR: 0 (0 to 3.9)		
	P (Central-Gbe vs Non-Gbe)	**0.036** **Diff: -3.8 (-6.1 to -1.5)** **OR: 0 (0 to 0.9)**	0.609 Diff: -0.9 (-4 to 2.2) OR: 0.7 (0 to 2.9)			**<0.00001** **Diff: 20.3 (10.3 to 30.4)**	**0.014** **Diff: -8.5 (-14.1 to -2.9)** **OR: 0.3 (0.1 to 0.8)**	0.609 Diff: -0.9 (-4 to 2.2) OR: 0.7 (0 to 2.9)		

						OR: 2.9 (1.8 to 4.8)				
	P (Eastern-Gbe vs Non-Gbe)	**0.003** **Diff: -2.7 (-5.2 to -0.3)** **OR: 0.3 (0.1 to 0.7)**	0.195 Diff: 1.8 (-0.6 to 4.2) OR: 1.7 (0.8 to 3.8)			**<0.00001** **Diff: 16.2 (10.6 to 21.7)** **OR: 2.4 (1.7 to 3.4)**	**< 0.00001** **Diff: -7.5 (-11.9 to -3.2)** **OR: 0.4 (0.2 to 0.6)**	**0.0021** **Diff: -2.2 (-4.2 to -0.2)** **OR: 0.2 (0.1 to 0.6)**		
	Total	1.5 (19)	4 (49)	0.8 (10)	55.6 (687)	29.9 (370)	7.1 (88)	1.1 (13)	100 (1236)	
NEW PATIENTS										
	Gbe	1 (8)	4.2 (35)	0.6 (5)	52.1 (430)	35.4 (292)	6.1 (50)	0.6 (5)	100 (825)	**< 0.0001**
	Non-Gbe	4.3 (9)	2.4 (5)	1.9 (4)	54.8 (115)	19.1 (40)	14.3 (30)	3.3 (7)	100 (210)	
	P (Gbe vs Non-Gbe)	**< 0.0007** **Diff: -3.3 (-6.1 to -0.5)** **OR: 0.2 (0.1 to 0.6)**				**< 0.00001** **Diff: 16.3 (10.1 to 22.6)** **OR: 2.3 (1.6 to 3.4)**	**0.0001** **Diff: -8.2 (-13.2 to -3.2)** **OR: 0.4 (0.2 to 0.6)**	**0.001** **Diff: -2.7 (-5.2 to-0.2)** **OR: 0.2 (0.1 to 0.5)**		
	Total	1.6 (17)	3.9 (40)	0.9 (9)	52.7 (545)	332 (32.1)	7.7 (80)	1.2 (12)	100 (1035)	
P-TREATED										
	Gbe	0.5 (1)	5.4 (11)	1 (2)	63.2 (129)	24.5 (50)	4.9 (10)	0.5 (1)	100 (204)	0.269
	Non-Gbe	1.9 (1)	3.9 (2)	1.9 (1)	75 (39)	9.6 (5)	7.7 (4)	0	100 (52)	
	Total	0.8 (2)	5.1 (13)	1.2 (3)	65.6 (168)	21.5 (55)	5.5 (14)	0.4 (1)	100 (256)	

Western-Gbe: includes Ewe, Mina (Gen), Ouatchi (a form of Ewe). **Central-Gbe**: includes Adja (Aja). **Eastern –Gbe**: includes Adjarra (Ajra), Aizo (Ayizo), Fon, Goun (Gun), Kotafon, Mahi (Maxi), Sahoue (Saxwe), Se, Setto, Toffin, Tori (Toli), Weme, Pedah (Xwela), Xwla.

Total with clear lineage assignment: n=1428

Gbe among L5 vs L4: p<0.00001, 88.4% in L5 vs 78.4% in L4, **Diff=10%** (5.6 to 14.4), OR: 2.1 (1.5 to 3)

Gbe among L6 vs L4: p=0.0017, 63.8% in L6 vs 78.4% in L4**, Diff=-14.6%** (-24.7 to -4.4), OR=0.5 (0.3 to 0.8)

Gbe among L1 vs L4: p=0.0014, 47.4% in L1 vs 78.4% in L4, **Diff=-31%** (-53.7 to -8.4), OR=0.2 (0.1 to 0.6)

Gbe among *M. bovis* vs L4: p=0.0055, 46.2% in *M. bovis* vs 78.4% in L4, **Diff=-32.2%** (-59.5% to -5), OR=0.2 (0.1 to 0.7)

Table 2.7. Ethnicity association with L6

Ethnicity (same for father and mother)	L6 % (n)	Total % (n)	p	Diff (95% CI)	OR (95% CI)
Ahoussa	**20 (4)**	100 (20)	**0.007**	**14.4 (-3.2 to 31.9)**	**4.2 (1.4 to 12.3)**
Bariba	**31.3 (5)**	100 (16)	**< 0.00001**	**25.6 (2.9 to 48.4)**	**7.6 (2.7 to 21.7)**
Gando	**33.3 (2)**	100 (6)	**0.004**	**27.7 (-10.1 to 65.4)**	**8.6 (0 to 40)**
Peulh	**22.9 (11)**	100 (48)	**< 0.00001**	**17.3 (5.3 to 29.2)**	**5 (2.4 to 10.1)**
Waama	**22.2 (2)**	100 (9)	**0.034**	**16.6 (-10.6 to 43.8)**	**4.8 (0 to 20.8)**
Other (different to above ethnicity)	5.6 (60)	100 (1063)	-	-	-
Total	**7.2 (84)**	100 (1162)			

Geographical distribution of lineage, patient residence & population density

Overall, L4 is more widespread in all the regions in Benin, whereas L5 is mostly found in the South of Benin and the Western-North regions, and L6 in the Eastern-North regions (Fig.2).

The distribution of lineages strongly varies across population density of patients' district of residence (p <0.0001, Table S5). L4 patients were significantly over-represented in residence districts of lower (11-49 and 50-87 populations/km^2) and highest population densities (739-788 and 739-1642 populations/km^2, Tables S4 and S6). The residence department with the highest L5 prevalence included Atlantique (36.9%), Mono (45.9%), Couffo (38.8%) and Plateau (43.5%), which were of intermediary population density (Fig 3, Table S4). L1 and L6 patients were significantly over-represented in lower population density districts (Fig 3, Table S5, Table S6). The residence departments with highest L6 prevalence included Alibori (23.7%) and Borgou (27.8%) and are both of lower population density (Fig 3, Table S4). L2 was significantly over-

represented in residence districts with the highest population densities, while *M. bovis* was significantly over-represented in lower population density districts (Fig.3, Table S6).

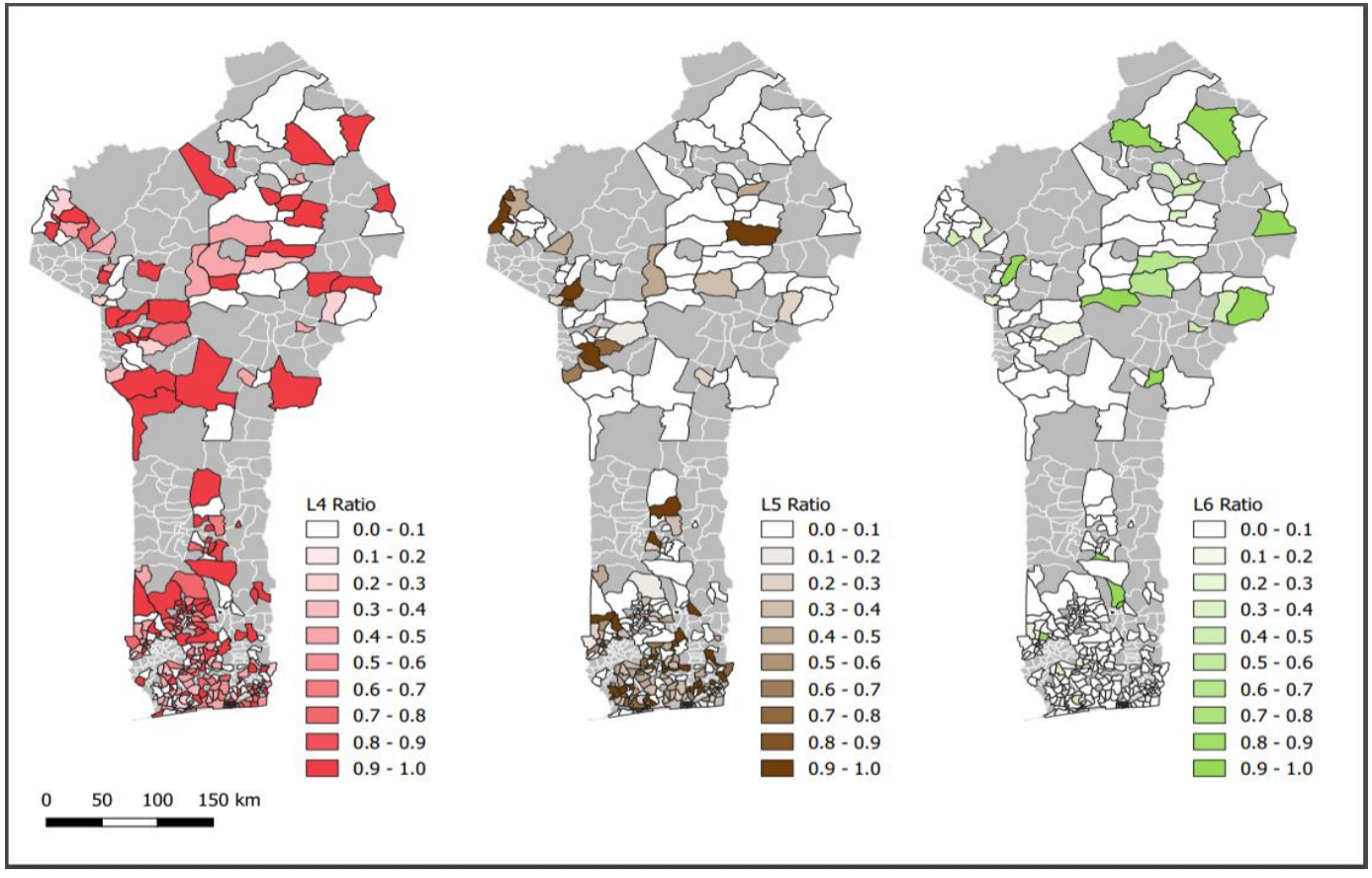

Figure 2. Geographical distribution of L4, L5 and L6 in Benin

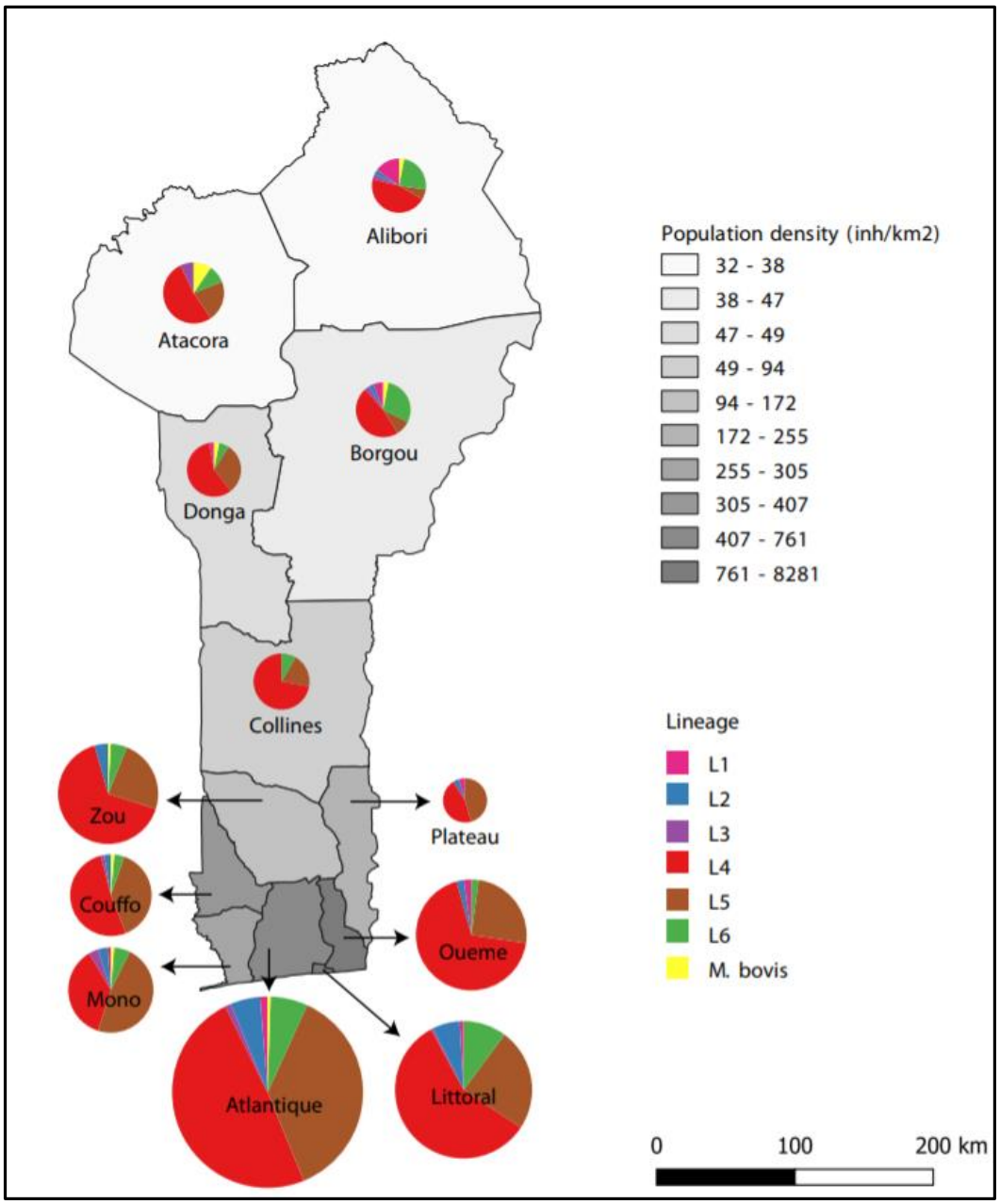

Figure 3. Distribution of MTBC lineages per residence department of patient along with population density of the departments. The numeric values of the lineages prevalence per department are presented in Table S3.

Testing for a trend in the population density of patient residence for L5 and L6 in an univariate analysis was not significant. After adjusting geographical situation of patient residence (South, Centre, North) and patient ethnicity, L5 prevalence decreases when the population density increases (relative to L4, aOR: 0.9 (0.8 to 0.99), p=0.033; and relative to non-L5 lineages aOR: 0.9 (0.8 to 0.97), p=0.011; Table 5), whereas L6 prevalence increases when the population density increases (relative to L4 aOR: 1.3 (1.03 to 1.7), p=0.026; relative to L6 aOR: 1.4 (1.05 to 1.7), p=0.016; Table 6).

The distribution of lineage significantly varied by department of residence. Relative to L4, L5 was significantly over-represented in Atlantique, Mono, Couffo, Alibori (Table 7), while L6 was significantly over-represented in Borgou, Alibori and under-represented in Oueme (Table 8) and L1 overrepresented in Alibori (Table 9). L5 was significantly over-represented in Atlantique (aOR: 1.7 (1.002 to 2.9), p=0.049) and Mono (aOR: 2.7 (1.2 to 6), p=0.012), independent of population density, ethnicity, type of patient, patient sex and age (Table 7). Similarly, L6 was significantly under-represented in Oueme (aOR: 0.2 (0.05 to 0.7), p=0.012), independent of population density, ethnicity, type of patient, patient sex and age (Table 8).

Baseline patient weight and difference in weight gain after treatment

The weight before treatment initiation did not significantly differ among patients with L2-L6 and *M. bovis* strains. But patients with L1 strains had a significantly lower baseline weight relative to patients with L4 strains (mean weight: 42.7 kg for L1 patients vs 50.6 kg for L4 patients, Diff: -7.9 kg (-13.7 to -

2.1), p=0.008, Table 2.8). The weight-gain from baseline to the end of the TB treatment did not differ across lineages (Table 2.8).

Distribution of lineage in category of previously-treated patients

L5 was less likely to be found in relapse patients than in new patients (17% vs 31.1%, Diff: -14.1 (-20 to -8.2), p=0.0001), while no difference was seen among previously-treated patients that failed or defaulted and new patients (respectively 28.4%, 35%, 31.1%). Conversely L4 was strongly over-represented among relapse patients (69.6% vs 52.9%, p< 0.00001, Table 2.9). After adjusting for Sex, age, HIV, sputum AFB-grade L5 was confirmed under-represented among relapse patients (aOR: 0.4 (0.3 to 0.7), p<0.0001, Table 3) relative to L4.

Treatment outcome

TB treatment outcome was associated with lineage (p=0.026, Table 2.10). While treatment success rates were similar in L2 through L6 and *M. bovis* (respectively 87.9% for L2, 90.9% for L3, 85% for L4, 89.4% for L5, 85.7% for L6 and 91.7% for *M. bovis*), it was significantly lower for patients with L1 strains, with a strongly higher failure rate (23.1%) relative to that of patients with L4 (3%) strains (Diff: 20.1% (-2.1 to 43.8), p< 0.00001, Table 2.10). After adjusting for RIF resistance by Xpert, HIV, patient type before treatment (new, relapse, failure, default), sex, age, residence department (the 12 departments), L1 was significantly associated with increased failure rates relative to L4 (aOR: 43.7 (4.6 to 418.4), p=0.001) and other non-L1 lineages (p<0.0001, Table 10). In he univariate analysis L6 treatment outcome was not significantly different compared to L4. However, after adjusting for RIF resistance by Xpert, HIV, patient type before treatment (new, relapse, failure, default), sex, age, L6 was significantly

associated with increased death rates (aOR: 3.6 (1.01 to 12.2), p=0.048, Table 10) relative to L4.

Occurrence of relapse within one year after being cured from current TB episode

No association was found between lineage and occurrence of TB relapse within one year after the patient was cured from the current TB episode (Table 2.11).

Table 2.8. Change in patient weight from treatment initiation to the end of the treatment

	ALL PATIENTS										
						Comparison weight end vs weight M0				Comparison of mean weigh diff (End-M0) across lineage (weight gain)	
	Sample size n	**Mean Weight M0 (kg)**	**St dev Mean weght M0**	**Diff (Mean Weight M0 Lx vs L4)**	**P (Mean Weight M0 Lx vs L4)**	**Mean Weight End**	**Mean Weigh Diff End – M0 (95% CI)**	**St dev of mean diff**	**P (weight End vs M0) (paired t-test)**	**Diff[mean diff End-M0 of Lx – mean diff End-M0 of L4) 95%CI**	**P (mean Diff Lx vs L4) (independent sample t-test)**
All	918	50.2				56.2	6.1 (5.6 to 6.4)	4.6	**< 0.00001**		
L1	10	**42.7**	10	**-7.9 (-13.7 to -2.1)**	**0.008**	47.3	4.6 (0.7 to 8.5)	5.5	**0.027**	-1.5 (-4.4 to 1.4)	0.315
L2	28	53.1	14.9	2.5 (-1.2 to 6.2)	0.185	59.4	6.3 (4.5 to 8)	4.6	**< 0.00001**	0.2 (-1.6 to 1.9)	0.862
L3	9	49.9	12.9	-0.7 (-6.9 to 5.5)	0.824	54.6	4.7 (0.007 to 9.3)	6.1	**0.0497**	-1.4 (-4.5 to 1.7)	0.363
L4	477	50.6	9.3	-	-	56.7	6.1 (5.7 to 6.5)	4.6	**< 0.00001**	-	
L5	266	49.5	9.3	-1.1 (-2.5 to 0.3)	0.123	55.5	6.1 (5.6 to 6.6)	4.4	**< 0.00001**	-0.01 (-0.7 to 0.7)	0.975
L6	68	50.1	8.8	-0.5 (-2.9 to 1.9)	0.677	56.4	6.3 (5.3 to 7.3)	4	**< 0.00001**	0.2 (-9.4 to 1.4)	0.716
M. bovis	11	47.4	6.7	-3.2 (-8.7 to 2.3)	0.257	54.1	6.7 (0.4 to 13)	9.4	**0.039**	0.6 (-2.2 to 3.5)	0.664

Table 2.9. Repartition of lineages by new and main categories of previously-treated patients

Lineages	Patient type										
	New	P-treated									P (new vs failure vs relapse vs default)
	% (n)	Failure % (n)	Failure vs new		Relapse % (n)	Relapse vs new		Default % (n)	Default vs new		
			Diff (95%CI)	p		Diff (95%CI)	p		Diff (95%CI)	p	
L1	1.7 (20)	1.5 (1)	-0.2 (-3.2 to 2.8)	0.902	0.5 (1)	-1.2 (-2.4 to 0)	0.207	0	-1.7 (-2.4 to -0.1)	0.557	
L2	3.8 (44)	4.5 (3)	0.7 (-4.4 to 5.8)	0.772	5.2 (10)	1.4 (-1.9 to 4.7)	0.357	0	-3.8 (-4.9 to -2.7)	0.374	
L3	1 (12)	0	-1 (-1.6 to -0.4)	0.411	1.6 (3)	0.6 (-1.3 to 2.5)	0.456	0	-1 (-1.6 to -0.4)	0.653	
L4	**52.9** (612)	62.2 (41)	9.3 (-2.7 to 21.3)	0.138	**69.6** (135)	**16.7 (9.6 to 23.8)**	**<0.00001**	60 (12)	7.1 (-14.6 to 28.8)	0.528	
L5	**31.1** (359)	28.4 (19)	-2.7 (-13.8 to 8.4)	0.642	**17** (33)	**-14.1 (-20 to -8.2)**	**0.0001**	**35** (7)	3.9 (-17.2 to 25)	0.709	
L6	8.1 (94)	4.5 (3)	-3.6 (-8.8 to 1.6)	0.288	5.7 (11)	-2.4 (-6 to 1.2)	0.248	5 (1)	-3.1 (-12.8 to 6.6)	0.613	
M. bovis	1.3 (15)	0	-1.3 (-2 to -0.6)	0.348	0.5 (1)	-0.8 (-2 to 0.4)	0.341	0	-1.3 (-2 to -0.6)	0.608	
											0.042
Total	**100 (1156)**	**100 (67)**			**100 (194)**			**100 (20)**			

L5 vs L4 (among type of p-treated):

Among previously-treated, proportion of **failure** among L5 (32.2, 19/59) vs L4 (41/188), p=0.104, OR: 1.7 (0.9 to 3.2), Diff: 10.4% (-2.9 to 23.7%).

Proportion of **relapse** among L5 (55.9%, 33/59) vs L4 (71.8%, 135/188), **p=0.023, OR: 0.5 (0.3 to 0.9), Diff: -15.9% (-30.1 to -1.7)**

Proportion of **default** among L5 (11.9%, 7/59) vs L4 (6.4%, 12/188), p=0.168, OR: 2 (0.7 to 5.1), Diff: 5.5% (-3.5 to 14.4).

Failure vs relapse: Proportion **L5** among **failure (28.4%) vs relapse patients (17%)**, **p=0.045**, **OR:1.9 (1.01 to 3.7), Diff: 11.3% (-6.7 to 23.4)**. **L4** among failure (61.2%) vs relapse (69.6%), p=0.21, OR:0.7 (0.4 to 1.2), Diff: -8.4% (-21.7 to 5)

Default vs relapse: Proportion **L5** among **default (35%) vs relapse patients (17%)**, **p=0.049**, **OR:2.6 (1.002 to 6.9), Diff: 18% (-3.6 to 39.6)**. **L4** among default (60%) vs relapse (69.6%), p=0.379, OR:0.7 (0.3 to 1.6), Diff: -9.6% (-32 to 12.8)

Failure vs default: Proportion **L5** among failure (28.4%) vs default patients (35%), p=0.569, OR:0.7 (0.3 to 2.1), Diff: -6.6% (-30.2 to 16.9). **L4** among failure (61.2%) vs default (60%), p=0.924, OR:1.1 (0.4 to 2.9), Diff: -1.2% (-23.2 to 25.6)

Table 2.10. Treatment outcome by lineage (for over one month of treatment taken)

Lineages	Treatment outcome					Total % (n)	P (outcome vs lineages)	Cured p-value Diff (95% CI)	Failed p-value Diff (95% CI)	Died p-value Diff (95% CI)
	Cured %(n)	M5 neg %(n)	Failed %(n)	Died %(n)	LTFU %(n)					
ALL PATIENTS										
L1	69.2 (9)	7.7 (1)	**23.1** (3)	0	0	100 (13)		0.118 -15.8 (-41.1 to 9.5)	**<0.00001** **20.1 (-2.1 to 43.8)**	0.6 -2.1 (-3.2 to -0.9)
L2	87.9 (29)	6.1 (2)	3 (1)	0	3 (1)	100 (33)		0.654 2.9 (-8.7 to 14.4)	0.767 0.8 (-5.2 to 6.8)	0.404 -2.1 (-3.2 to -0.9)
L3	90.9 (10)	9.1 (1)	0	0	0	100 (11)		0.587 5.9 (-11.4 to 23.1)	0.616 -2.2 (-3.4 to -1)	0.63 -2.1 (-3.2 to -0.9)
L4	85 (494)	8.3 (48)	2.2 (13)	2.1 (12)	2.4 (14)	100 (581)		-	-	-
L5	89.4 (287)	3.7 (12)	2.5 (8)	2.2 (7)	2.2 (7)	100 (321)		0.065 4.4 (-0.1 to 8.8)	0.808 0.3 (-1.8 to 2.3)	0.908 0.1 (-0.9 to 2.1)
L6	85.7 (66)	6.5 (5)	1.3 (1)	5.2 (4)	1.3 (1)	100 (77)		0.873 0.7 (-7.6 to 9)	0.592 -0.9 (-3.7 to 1.9)	0.094 3.1 (-2 to 8.2)
M. bovis	91.7 (11)	0	0	8.3 (1)	0	100 (12)		0.522 6.6 (-9.3 to 22.5)	0.6 -2.2 (-3.4 to -1)	0.142 6.3 (-9.4 to 21.9)
							0.026			
Total	**86.5 (906)**	**6.6 (69)**	**2.5 (26)**	**2.3 (24)**	**2.2 (23)**	**100 (1048)**				

Table 2.11. Occurrence of relapse within one year after TB is cured

Lineages	Cough within 1 year after TB cure		Total	Diff (95% CI)	P
	TB relapse % (n)	**No + Non TB** % (n)			
ALL PATIENTS					
L1	0	100 (8)	100 (8)		
L2	0	100 (14)	100 (14)		
L3	0	100 (5)	100 (5)		
L4	3.7 (10)	96.3 (257)	100 (267)	-	
L5	0.8 (1)	99.2 (129)	100 (130)	**-3 (-5.7 to -0.2)**	0.09
L6	3.9 (1)	96.2 (25)	100 (26)		
M. bovis	0	100 (5)	100 (5)		
					0.669
Total	**2.6 (12)**	**97.4 (443)**	**100 (455)**		

Were excluded: “unscreened cough”, “died” (died after cure and died after M5 negative in a p-treated patient, and “unreachable”.

TB diagnostic results

Microscopy. Among new patients, AFB-smear grade did not differ across lineages. However, among previously-treated patients L5 specimens were significantly over-represented among AFB-scanty sputa relative to L4 specimens (21.6% vs 7.6%, Diff: 13.9% (1.8 to 26), OR: 3.3 (1.4 to 8.1), p=0.007; Table 2.12). After stratification by type of previously-treated patients, L5 was actually significantly over-represented among failure patients with AFB-scanty sputa compared to relapse (50% vs 6.9%, p=0.0009) and default patients, unlike L4 (12.1% in failure vs 4.8% in relapse, p=0.14; Table S7). Furthermore, among

failure patients with ABF-scanty sputa, L5 prevalence was significantly higher than L4 prevalence (50% vs 12.1%, Diff:37.9% (11 to 64.8), OR: 7.3 (1.8 to 29), p=0.0004; Table S7).

Xpert MTB/RIF. Xpert positivity did not differ across lineage with the positivity ranging from 98.6 to 100% in new patients' specimens and 98.3 to 100% in previously-treated patients' specimens. Among new patients, 0.7% of the positive specimens were RIF-resistant (6/684) compared to 3.6% (10/278) of previously-treated patients. Of these 10 RIF-resistant patients, five were from L5 (8.6%, 5/58) and the other five were from L4 (2.7%, 5/187) (Diff: 5.9% (-1.6 to13.5), OR: 3.4 (1 to 11.5), p=0.0455, Table 2.13).

Culture. The positivity of culture varied significantly across lineage (Table S8) even after repeat of contaminated and negative specimens to improve positivity yields (Table 2.14). L5, L6 and *M. bovis* were strongly under-represented among positive cultures (Tables 2.14 and S8). Furthermore, L5, L6 and *M. bovis* positive cultures were strongly over-represented among dysgonic isolates (respectively p <0.00001, p< 0.00001, p=0.042; Table 2.15). That under-representation in positive cultures and over-representation among dysgonic isolates was even more strongly confirmed after adjusting for patient type (new, relapse, failure, default), sputum AFB-grade (Table 3 for L5 and Table 4 for L6). For culture positivity, relative to L4, the aOR for L5 was 0.4 (0.2 to 0.6), p<0.0001 (Table 3) and that of L6 was 0.1 (0.08 to 0.3), p<0.0001 (Table 4). Regarding the dysgonic nature of positive cultures, relative to L4, the aOR for L5 was 9.7 (6.6 to 14.2), p<0.0001 (Table 3) and that of L6 was 8.6 (4.3 to 17.1), p<0.0001 (Table 4).

Phenotypic DST. Among new and previously-treated patients, L2 was strongly associated respectively with STR resistance (70.8% of new patients and 100% of previously-treated patients) and EMB resistance (15.8% of new patients and 16.7% of previously-treated patients). As was seen with Xpert, in previously-treated patients, L5 was significantly over-represented (17.7%, 3/17) among RIF-resistant isolates relative to L4 (4.4%, 5/114 for L4; Diff: 13.3% (-5.2 to 31.8), OR: 4.7 (1.1 to 20), p=0.033, Table 2.16). All three L5 RIF-resistant specimens were from previously-treated patients, including two failure patients and one relapse patient. One of the two failure patients, a 30 years old female, was previously enrolled in the study as a new patient. The Xpert result for the initial specimen was RIF-sensitive, but the culture was negative. Interestingly, the spoligotyping-pattern when she failed treatment and was re-enrolled as failure patient, was the same as when she was a new TB patient (101111100000001111110000111111111110001111), suggesting that the L5 strain she was infected with possibly acquired RIF resistance.

Table 2.12. Distribution of sputum AFB microscopy by lineage

Lineages	Sputum AFB microscopy % (n)					Total	p	Scanty Lx vs L4 Diff (95% CI) OR (95% CI)	p (scanty Lx vs L4)
	+++	++	+	scanty	neg				
ALL PATIENTS	40.6 (n=440)	36.4 (n=394)	15.2 (n=165)	7.1 (n=77)	0.7 (n=8)	100 (n=1084)			
L1	41.2 (7)	29.4 (5)	29.4 (5)	0	0	100 (17)			
L2	40.5 (17)	35.7 (15)	16.7 (7)	4.8 (2)	2.4 (1)				
L3	54.6 (6)	27.3 (3)	9.1 (1)	9.1 (1)	0	100 (11)			
L4	40.4 (243)	37.9 (228)	13.8 (83)	7.1 (43)	0.8 (5)	100 (602)			
L5	39.5 (126)	34.8 (111)	16.9 (54)	8.2 (26)	0.6 (2)	100 (319)			
L6	45 (36)	32.5 (26)	17.5 (14)	5 (4)	0	100 (80)			
M. bovis	38.5 (5)	46.2 (6)	7.7 (1)	7.7 (1)	0	100 (13)			
							0.971		
NEW PATIENTS	41.9 (n=362)	37.8 (n=327)	14 (n=121)	**6.1 (n=53)**	0.2 (n=2)	100 (n=865)			
L1	46.7 (7)	33.3 (5)	20 (3)	0	0	100 (15)			
L2	42.4 (14)	39.4 (13)	15.2 (5)	3 (1)	0	100 (33)			
L3	55.6 (5)	33.3 (3)	0	11.1 (1)	0	100 (9)			
L4	41.5 (190)	39.5 (181)	11.8 (54)	7 (32)	02 (1)	100 (458)			
L5	40.3 (108)	36.6 (98)	17.2 (46)	5.6 (15)	0.4 (1)	100 (268)			
L6	47.8 (33)	30.4 (21)	17.4 (12)	4.4 (3)	0	100 (69)			
M. bovis	38.5 (5)	46.2 (6)	7.7 (1)	7.7 (1)	0	100 (13)			
							0.971		
P-TREATED	35.6 (n=78)	30.6 (n=67)	20.1 (n=44)	**11 (n=24)**	2.7 (n=6)	100 (n=219)			
L1	0	0	100 (2)	0	0	100 (2)			
L2	33.3 (3)	22.2 (2)	22.2 (2)	11.1 (1)	11.1 (1)	100 (9)		Diff: 2.4 (-16.8 to 21.5) OR: 1.3 (0 to 9.2)	0.788

L3	50 (1)	0	50 (1)	0	0	100 (2)		
L4	36.8 (53)	32.6 (47)	20.1 (29)	**7.6 (11)**	2.8 (4)	100 (144)	-	
L5	35.3 (18)	25.5 (13)	15.7 (8)	**21.6 (11)**	2 (1)	100 (51)	**Diff: 13.9 (1.8 to 26)** **OR: 3.3 (1.4 to 8.1)**	**0.007**
L6	27.3 (3)	45.5 (5)	18.2 (2)	9.1 (1)	0	100 (11)	Diff: 1.5 (-16.1 to 19) OR: 1.2 (0 to 8.2)	0.862
M. bovis	0	0	0	0	0	0		
							0.355	

Sensitivity analysis: among p-treated, **L5 vs L4 in scanty+neg** yielded 23.5%(L5) vs 10.4%(L4), **p=0.02,** OR=2.6 (1.2 to 6), Diff=13.1 (0.4 to 25.8). See **Table S7** for stratification by p-treated type

Table 2.13. Xpert MTB/RIF (classical) result by lineage

Lineages	Xpert MTB/RIF				
	Positive	Total	P	RIF R % relative to positive (n)	P
ALL PATIENTS					
L1	100 (20)	100 (20)		0	
L2	100 (43)	100 (43)		0	
L3	100 (12)	100 (12)	0.927	0	0.964
L4	99.7 (647)	100 (649)		1.6 (10)	
L5	99.7 (321)	100 (322)		1.6 (5)	
L6	98.9 (87)	100 (88)		1.2 (1)	
M. bovis	100 (12)	100 (12)		0	
Total	**99.7 (1142)**	**100 (1146)**		**1.4 (16)**	
New patients					
L1	100 (18)	100 (18)		0	
L2	100 (31)	100 (31)		0	
L3	100 (9)	100 (9)		0	
L4	99.8 (460)	100 (461)		1.1 (5)	**-**
L5	100 (263)	100 (263)		0	0.09
L6	98.6 (72)	100 (73)		1.4 (1)	0.822
M. bovis	100 (11)	100 (11)		0	0.728
			0.559		0.697
Total	99.8 (864)	100 (866)			
P-treated patients					
L1	100 (2)	100 (2)		0	
L2	100 (12)	100 (12)		0	
L3	100 (3)	100 (3)		0	
L4	99.5 (187)	100 (188)		2.7 (5)	-
L5	98.3 (58)	100 (59)		**8.6** (5)	**0.0455** **Diff: 5.9 (-1.6 to 13.5)** **OR: 3.4 (1 to 11.5)**
L6	100 (15)	100 (15)		0	0.521
M. bovis	100 (1)	100 (1)		0	0.868
			0.980		0.433
Total	**99.3 (278)**	**100 (280)**		**3.6 (10)**	

Table 2.14. Culture result by lineage (final result)

Lineages	LJ Culture (up to 90 days)				LJ Culture (up to 90 days)- excluding contamination		
	Positive % (n)	Contaminated	Total % (n)	P	Positive %	Total % (n)	p
ALL PATIENTS							
L1	85 (17)	0	100 (20)		85 (17)	100 (20)	
L2	81.4 (35)	**7** (3)	100 (43)		87.5 (35)	100 (40)	
L3	91.7 (11)	0	100 (12)	**< 0.0001**	91.7 (11)	100 (12)	**< 0.0001**
L4	85.8 (556)	3.9 (25)	100 (648)		89.3 (556)	100 (623)	
L5	71.4 (230)	**8.7** (28)	100 (322)		**78.2** (230)	100 (294)	
L6	56.8 (50)	**6.8** (6)	100 (88)		**61** (50)	100 (82)	
M. bovis	58.3 (7)	**16.7** (2)	100 (12)		**70** (7)	100 (10)	
Total	**79.1 (906)**	**5.6 (64)**	**100 (1145)**		**83.8 (906)**	**100 (1081)**	
NEW							
L1	94.4 (17)	0	100 (18)		94.4 (17)	100 (18)	
L2	87.1 (27)	9.7 (3)	100 (31)		96.4 (27)	100 (28)	
L3	88.9 (8)	0	100 (9)		88.9 (8)	100 (9)	
L4	88.7 (409)	3.7 (17)	100 (461)	-	92.2 (409)	100 (444)	-
L5	76.8 (202)	7.6 (20)	100 (263)	**<0.00001** **Diff: -11.9 (-18.8 to -6.1)** **OR: 0.4 (0.3 to 0.6)**	83.1 (202)	100 (243)	**0.0003** **Diff: -9 (-14.3 to -3.7)** **OR: 0.4 (0.3 to 0.7)**
L6	60.3 (44)	8.2 (6)	100 (73)	**< 0.00001** **Diff: -28.4 (-40 to -16.9)**	65.7 (44)	100 (67)	**< 0.00001**

				OR: 0.2 (0.1 to 0.3)			**Diff: -26.4 (-38.1 to -14.8)**
M. bovis	63.6 (7)	9.1 (1)	100 (11)	**0.011** **Diff: -25.1 (-53.7 to 3.5)** **OR: 0.2 (0.1 to 0.7)**	70 (7)	100 (10)	**OR: 0.2 (0.1 to 0.3)** **0.013** **Diff: -22.1 (-50.6 to 6.4)** **OR: 19.9 (5.3 to 73.8)**
				< 0.0001			**< 0.0001**
Total	**82.5 (714)**	**5.4 (47)**	100 (866)		**87.2 (714)**	**100 (819)**	
P-TREATED							
L1	0	0	100 (2)		0	100 (2)	
L2	66.7 (8)	0	100 (12)		66.7 (8)	100 (12)	
L3	100 (3)	0	100 (3)		100 (3)	100 (3)	
L4	**78.6** (147)	4.3 (8)	100 (187)	-	**82.1** (147)	100 (179)	
L5	**47.5** (28)	13.6 (8)	100 (59)	**< 0.00001** **Diff: -31.2 (-45.2 to -17.1)** **OR: 0.2 (0.1 to 0.5)**	**54.9** (28)	100 (51)	**< 0.00001** **Diff: -36.2 (-49.9 to -22.5)** **OR: 0.2 (0.1 to 0.3)**
L6	**40** (6)	0	100 (15)	**0.0008** **Diff: -38.6 (-64.1 to -13.1)** **OR: 0.2 (0.1 to 0.5)**	**40** (6)	100 (15)	**0.0001** **Diff: -42.1 (-67.5 to -16.7)** **OR: 0.1 (0.1 to 0.4)**
M. bovis	0	100 (1)	100 (1)		0	0	
				< 0.0001			**< 0.0001**
Total	**68.8 (192)**	6.1 (17)	100 (279)		73.3 (192)	100 (262)	

Table 2.15. Quality/nature of isolates (from LJ culture) by lineage

Lineages	Quality (nature) of isolate				
	Dysgonic % (n)	**Eugonic**	**Diff in proportion of dysgonic relative to L4 (95% CI)**	**OR (95% CI)**	**P**
ALL PATIENTS					
L1	18.8 (3)	81.3 (13)	3.7 (-15.6 to 23.1)	1.3 (0.4 to 4.4)	0.681
L2	15.2 (5)	84.9 (28)	0.1 (-12.5 to 12.7)	1 (0.4 to 2.6)	0.983
L3	9.1 (1)	90.1 (10)	-5.9 (-23.2 to 11.3)	0.6 (0 to 3.5)	0.585
L4	15 (82)	85 (464)	-	-	-
L5	60.9 (134)	39.1 (86)	**45.9 (38.8 to 53)**	**8.8 (6.2 to 12.6)**	**< 0.00001**
L6	54.4 (25)	45.7 (21)	**39.3 (24.6 to 54)**	**6.7 (3.6 to 12.5)**	**< 0.00001**
M. bovis	42.9 (3)	57.1 (4)	**27.8 (-8.9 to 64.6)**	**4.2 (1 to 17.3)**	**0.042**
					< 0.0001
Total	**28.8 (253)**	**71.2 (626)**			

Table 2.16. Repartition of lineage by phenotypic resistance to antibiotics

Lineages	Phenotypic DST											
	RIF R % (n)	Total % (n)	p	INH R	Total % (n)	p	STR R	Total	p	EMB R	Total	p
ALL PATIENTS												
L1	0	100 (12)		8.3 (1)	100 (12)		0	100 (12)		0	100 (12)	
L2	0	100 (30)		0	100 (30)		**76.7** (23)	100 (30)	**< 0.00001 Diff: 63.8 (48.4 to 79.3) OR: 22.3 (9.3 to 53)**	**16** (4)	100 (25)	**< 0.00001 Diff: 15.3 (0.9 to 29.7) OR: 28.3 (6.6 to 120.5)**
L3	0	100 (10)		0	100 (10)		0	100 (10)		0	100 (9)	
L4	1.8 (8)	100 (456)	-	4 (18)	100 (455)	-	12.9 (58)	100 (451)	-	0.7 (3)	100 (448)	-
L5	2.3 (3)	100 (129)		3.2 (4)	100 (124)		11.9 (15)	100 (126)		1.6 (2)	100 (125)	
L6	**4.8** (1)	100 (21)	0.322	**9.5** (2)	100 (21)	0.214	4.8 (1)	100 (21)		0	100 (21)	
M. bovis	0	100 (4)		0	100 (3)		0	100 (4)		0	100 (0)	
			0.894			0.626			**< 0.0001**			**< 0.0001**
Total	**1.8 (12)**	**100 (662)**		**3.8 (25)**	**100 (655)**		**14.8 (97)**	**100 (654)**		**1.4 (9)**	**100 (643)**	
NEW PATIENTS												
L1	0	100 (12)		8.3 (1)	100 (12)	0.337	0	100 (12)		0	100 (2)	
L2	0	100 (24)		0	100 (24)		**70.8** (17)	100 (24)	**< 0.00001 Diff: 70.8 (53.8 to 87.9) OR: 37.8 (12.6 to 112.5)**	**15.8** (3)	100 (19)	**< 0.00001** Diff: 14.9 (-1.5 to 31.3) **OR: 20.8 (4.4 to 98.2)**

L3	0	100 (7)		0	100 (7)		0	100 (7)		0	100 (6)	
L4	0.9 (3)	100 (342)	-	3.2 (11)	100 (341)	-	11.2 (38)	100 (338)	-	0.9 (3)	100 (336)	
L5	0	100 (112)	0.237	0.9 (1)	100 (108)	0.197	10.1 (11)	100 (109)		0.9 (1)	100 (108)	
L6	5.6 (1)	100 (18)	0.151	11.1 (2)	100 (18)	0.081	5.6 (1)	100 (18)	0.528	0	100 (18)	
M. bovis	0	100 (4)	0.823	0	100 (3)		0	100 (4)		0	100 (3)	
			0.352			0.228			**< 0.0001**			**< 0.0001**
Total New	**0.8 (4)**	**100 (519)**		2.9 (15)	100 (513)		**13.1 (67)**	**100 (512)**		**1.4 (7)**	**100 (502)**	
P-TREATED PATIENTS												
L1	0	0		0	0		0	0		0	0	
L2	0	100 (6)		0	100 (6)		**100** (6)	100 (6)	**0.009 Diff: 32.3 (3.1 to 61.5) OR: 4.7 (1.4 to 15.2)**	**16.7** (1)	100 (6)	**< 0.00001** Diff: 16.7 (-13.2 to 46.5)
L3	0	100 (3)		0	100 (3)		0	100 (3)		0	100 (3)	
L4	4.4 (5)	100 (114)	-	6.1 (7)	100 (114)		17.7 (20)	100 (113)	-	0	100 (112)	
L5	**17.7** (3)	100 (17)	**0.033 Diff: 13.3 (-5.2 to 31.8) OR: 4.7 (1.1 to 20)**	**18.8** (3)	100 (16)	0.076 Diff: 12.6 (-7 to 32.2) OR: 3.5 (0.9 to 14.3)	23.5 (4)	100 (17)	0.564	**5.9** (1)	100 (17)	0.01 Diff: 5.9 (-5.3 to 17.1) -
L6	0	100 (3)	0.711	0	100 (3)		0	100 (3)		0	100 (3)	
M. bovis	0	0		0	0		0	0		0	0	
			0.223			0.354			**< 0.0001**			**0.007**
Total P-treated	**5.6 (8)**	**100 (143)**		**7 (10)**	**100 (142)**		**21.1 (30)**	**100 (142)**		**1.4 (2)**	**100 (141)**	

Table 3. Multivariable analyses for L5 association with other variables (patients' characteristics)

Variable 1		Variable 2		Crude OR (95%CI)	Crude p	Adjusted for	Adjusted OR (95% CI)	Adjusted P*
lineage		**Patient type**						
L5	nonL5 (ref)	p-treated	New (ref)	**0.6 (0.4 to 0.8)**	**0.001**	Sex, age, HIV, sputum AFB-grade	**0.6 (0.4 to 0.9)**	**0.013**
	nonL5 (ref)	relapse	New (ref)	**0.5 (0.3 to 0.7)**	**<0.0001**	Sex, age, HIV, sputum AFB-grade	**0.5 (0.3 to 0.7)**	**0.002**
		failure		0.9 (0.5 to 1.5)	0.643		0.9 (0.5 to 1.8)	0.812
		default		1.2 (0.5 to 3)	0.706		1.7 (0.5 to 5.5)	0.370
lineage		**sex**						
L5	nonL5 (ref)	female	Male (ref)	**1.6 (1.2 to 2)**	**<0.0001**	Patient type (new, relapse, failure, default), age, HIV, sputum AFB-grade	**1.7 (1.3 to 2.3)**	**0.001**
lineage		**age**						
L5	nonL5 (ref)	As continuous variable		1.007 (1 to 1.02)	0.103	Patient type (new, relapse, failure, default), sex, HIV, sputum AFB-grade	**1.01 (1.001 to 1.02)**	**0.028**
lineage		**HIV**						
L5	nonL5 (ref)	positive	Negative (ref)	**1.4 (1.01 to 2.1)**	**0.046**	Patient type (new, relapse, failure, default), sex, age, sputum AFB-grade	1.2 (0.8 to 1.9)	0.355

lineage		**Microscopy sputum (AFB-grade)**						
L5	nonL5 (ref)	+++,++,+pos	scanty (ref)	1.2 (0.8 to 2)	0.391	Patient type (new, relapse, failure, default), sex, age, HIV	1.3 (0.7 to 2.2)	0.386
		Patient type						
L5	L4 (ref)	p-treated	new	**0.5 (0.4 to 0.7)**	**<0.0001**	Sex, age, HIV, sputum microcopy	**0.5 (0.4 to 0.8)**	**0.003**
	L4 (ref)	relapse	new	**0.4 (0.3 to 0.6)**	**<0.0001**	Sex, age, HIV, sputum AFB-grade	**0.4 (0.3 to 0.7)**	**<0.0001**
		failure		0.8 (0.5 to 1.4)	0.409		0.8 (0.4 to 1.5)	0.465
		default		1 (0.4 to 2.5)	0.991		1.6 (0.5 to 5.3)	0.458
		sex				-		
L5	L4 (ref)	female	Male (ref)	**1.6 (1.3 to 2.1)**	**<0.0001**	Patient type (new, relapse, failure, default), age, HIV, sputum AFB-grade	**1.8 (1.3 to 2.5)**	**<0.0001**
		Age						
L5	L4 (ref)	As continuous variable		**1.01 (1.002 to 1.02)**	**0.013**	Patient type (new, relapse, failure, default), sex, HIV, sputum AFB-grade	**1.01 (1.004 to 1.03)**	**0.007**
		HIV						

L5	L4 (ref)	positive	Negative (ref)	**1.5 (1.013 to 2.2)**	**0.042**	Patient type (new, relapse, failure, default), sex, age, sputum AFB-grade	1.3 (0.8 to 2)	0.288
lineage		**Microscopy sputum (AFB-grade)**						
L5	L4 (ref)	+++,++,+pos	Scanty (ref)	1.2 (0.7 to 1.9)	0.587	Patient type (new, relapse, failure, default), sex, age, HIV	1.1 (0.6 to 2)	0.660
		Culture						
L5	nonL5 (ref)	positive	Negative (ref)	**0.6 (0.42 to 0.8)**	**0.003**	Patient type (new, relapse, failure, default), sputum AFB-grade	**0.5 (0.4 to 0.8)**	**0.002**
L5	L4 (ref)	positive	Negative (ref)	**0.4 (0.3 to 0.6)**	**<0.0001**	Patient type (new, relapse, failure, default), sputum AFB-grade	**0.4 (0.2 to 0.6)**	**<0.0001**
		Isolates quality						
L5	nonL5 (ref)	dysgonic	Eugonic (ref)	**7.1 (5.1 to 9.9)**	**<0.0001**	Patient type (new, relapse, failure, default), sputum AFB-grade	**7.6 (5.3 to 10.8)**	**<0.0001**
L5	L4 (ref)	dysgonic	Eugonic (ref)	**8.8 (6.2 to 12.6)**	**<0.0001**	Patient type (new, relapse, failure, default), sputum AFB-grade	**9.7 (6.6 to 14.2)**	**<0.0001**
		Patient type						

dysgonic	Eugonic (ref)	relapse	New (ref)	0.8 (0.6 to 1.2)	0.376	sputum AFB-grade	0.8 (0.5 to 1.2)	0.344
		failure		1 (0.4 to 2.1)	0.936		1.3 (0.6 to 3)	0.520
		default		0.6 (0.2 to 2.3)	0.505		0.91 (0.2 to 3.5)	0.891

*Using logistic regression

Ref: reference group (referent)

Table 4. Multivariable analyses for L6 association with other variables (patients' characteristics)

Variable 1		Variable 2		Crude OR (95%CI)	Crude p	Adjusted for	Adjusted OR	Adjusted p
lineage		**Patient type**						
L6	nonL6 (ref)	p-treated	New (ref)	0.6 (0.4 to 1.1)	0.115	Sex, age, HIV, sputum AFB-grade	0.6 (0.3 to 1.3)	0.220
	nonL6 (ref)	Relapse	New (ref)	0.7 (0.4 to 1.3)	0.239	Sex, age, HIV, sputum AFB-grade	0.6 (0.3 to 1.4)	0.243
		Failure		0.5 (0.2 to 1.7)	0.290		0.6 (0.1 to 2.7)	0.522
		Default		0.6 (0.1 to 4.5)	0.614		1.1 (0.1 to 8.6)	0.937
lineage		**Sex**						
L6	nonL6 (ref)	Female	Male (ref)	0.7 (0.5 to 1.1)	0.165	Patient type (relapse, failure, default), age, HIV, sputum AFB-grade	0.8 (0.4 to 1.4)	0.340
lineage		**Age**						
L6	nonL6 (ref)	As continuous variable		**1.02 (1.002 to 1.03)**	**0.019**	Patient type (relapse, failure, default), sex, HIV, sputum AFB-grade	1 (1 to 1.03)	0.337
lineage		**HIV**						
L6	nonL6 (ref)	Positive	Negative (ref)	0.8 (0.4 to 1.5)	0.450	Patient type (relapse, failure, default), sex, age, sputum AFB-grade	0.9 (0.4 to 2.1)	0.876
lineage		**Microscopy sputum (AFB-grade)**						
L6	nonL6 (ref)	+++,++,+pos	scanty (ref)	0.7 (0.2 to 1.9)	0.440	Patient type (relapse, failure, default), sex, age, HIV	0.4 (0.1 to 1.7)	0.219

		Patient type						
L6	L4 (ref)	p-treated	new	**0.5 (0.3 to 0.9)**	**0.024**	Sex, age, HIV, sputum microcopy	0.5 (0.2 to 1.02)	0.059
	L4 (ref)	Relapse	new	0.5 (0.3 to 1.01)	0.057	Sex, age, HIV, sputum AFB-grade	0.5 (0.2 to 1.1)	0.074
		Failure		0.5 (0.1 to 1.6)	0.223		0.5 (0.1 to 2.1)	0.345
		Default		0.5 (0.1 to 4.2)	0.559		1.1 (0.1 to 9.1)	0.953
		Sex				-		
L6	L4 (ref)	Female	Male (ref)	0.9 (0.5 to 1.4)	0.601	Patient type (relapse, failure, default), age, HIV, sputum AFB-grade	0.9 (0.5 to 1.7)	0.867
		Age						
L6	L4 (ref)	As continuous variable		**1.02 (1.007 to 1.04)**	**0.003**	Patient type (relapse, failure, default), sex, HIV, sputum AFB-grade	1.01 (1 to 1.03)	0.116
		HIV						
L6	L4 (ref)	Positive	Negative (ref)	0.9 (0.4 to 1.9)	0.763	Patient type (relapse, failure, default), sex, age, sputum AFB-grade	1.1 (0.5 to 2.5)	0.875
lineage		**Microscopy sputum (AFB-grade)**						
L6	L4 (ref)	+++,++,+pos	Scanty (ref)	0.7 (0.2 to 1.9)	0.469	Patient type (relapse, failure, default), sex, age, HIV	0.4 (0.1 to 1.7)	0.223
		Culture						
L6	nonL6 (ref)	Positive	Negative (ref)	**0.3 (0.2 to 0.4)**	**<0.0001**	Patient type (relapse, failure, default), sputum AFB-grade	**0.2 (0.1 to 0.4)**	**<0.0001**

L6	L4 (ref)	Positive	Negative (ref)	**0.2 (0.1 to 0.3)**	**<0.0001**	Patient type (relapse, failure, default), sputum AFB-grade	**0.1 (0.08 to 0.3)**	**<0.0001**
		Isolates quality						
L6	nonL6 (ref)	Dysgonic	Eugonic (ref)	**3.2 (1.7 to 5.8)**	**<0.0001**	Patient type (relapse, failure, default), sputum AFB-grade	**3.4 (1.8 to 6.4)**	**<0.0001**
L6	L4 (ref)	Dysgonic	Eugonic (ref)	**6.7 (3.6 to 12.6)**	**<0.0001**	Patient type (relapse, failure, default), sputum AFB-grade	**8.6 (4.3 to 17.1)**	**<0.0001**

Table 5. Multivariable analyses for L5 association with ethnicity, population density, department of patient residence

Variable 1		Variable 2		Crude OR (95%CI)	Crude p	Adjusted for	Adjusted OR (95%CI)	Adjusted p
lineage		**Ethnic group**						
L5	nonL5 (ref)	Gbe	Non gbe (ref)	**2.4 (1.7 to 3.4)**	**< 0.0001**	- Population density of residence (ordinal) - Department of residence (south, centre, north)	**1.8 (1.1 to 2.9)**	**0.023**
	L4 (ref)	Gbe	Non gbe (ref)	**2.1 (1.4 to 3)**	**< 0.0001**	- Population density of residence (ordinal) - Department of residence (south, centre, north)	**1.7 (1.1 to 2.9)**	**0.031**
L5	nonL5 (ref)	East-Gbe	Non gbe (ref)	**2.4 (1.7 to 3.4)**	**< 0.0001**	- Population density of residence (ordinal) - Department of residence (south, centre, north)	**1.8 (1.1 to 3)**	**0.021**
		Centr-Gbe		**2.9 (1.8 to 4.8)**	**< 0.0001**		**2 (1.1 to 3.8)**	**0.028**
		West-Gbe		1.3 (0.6 to 3.1)	0.509		1.1 (0.4 to 2.7)	0.851
	L4 (ref)	East-Gbe	Non gbe (ref)	**2.1 (1.5 to 3)**	**< 0.0001**	- Population density of residence (ordinal)	**1.7 (1.05 to 2.9)**	**0.032**
		Centr-Gbe		**2.4 (1.4 to 3.9)**	**0.001**		1.8 (0.98 to 3.5)	0.06
		West-Gbe		1.2 (0.5 to 3)	0.624		1.1 (0.4 to 2.)	0.802

					-	Department of residence (south, centre, north)		
lineage		**Population density**						
L5	nonL5 (ref)	As ordinal variable	0.99 (0.9 to 1.1)	0.922	-	Ethnic group (non gbe, gbe)	**0.9 (0.8 to 0.97)**	**0.011**
			2.9 (2 to 4.4) for categorical intermediary vs low	**<0.0001**	-	Department of residence (south, centre, north)		
			1.8 (1.2 to 2.7) for categorical high vs low	**0.006**				
	L4 (ref)	As ordinal variable	0.97 (0.9 to 1.05)	0.484	-	Ethnic group (non gbe, gbe)	**0.9 (0.8 to 0.99)**	**0.033**
			2.6 (1.7 to 4) for categorical intermediary vs low	**<0.0001**	-	Department of residence (south, centre, north)		
			1.5 (1.01 to 2.3) for categorical high vs low	**0.042**				

L5	nonL5 (ref)	As ordinal variable	0.99 (0.9 to 1.1) **2.9 (2 to 4.4)** for categorical intermediary vs low **1.8 (1.2 to 2.7)** for categorical high vs low	0.922 **<0.0001** **0.006**	-	Ethnic group (**non gbe, East-gbe, centr-gbe, west-gbe**) - Department of residence (south, centre, north)	**0.9 (0.8 to 0.99)**	**0.028**
	L4 (ref)	As ordinal variable	0.97 (0.9 to 1.05) **2.6 (1.7 to 4)** for categorical intermediary vs low **1.5 (1.01 to 2.3)** for categorical high vs low	0.484 **<0.0001** **0.042**	-	Ethnic group (**non gbe, East-gbe, centr-gbe, west-gbe**) - Department of residence (south, centre, north)	**0.9 (0.8 to 0.99)**	**0.049**
lineage		**Department of residence**						
L5		Centre	0.5 (0.2 to 1.2)	0.122			0.4 (0.2 to 1.2)	0.094

	nonL5 (ref)	North	South (ref)	**0.4 (0.3 to 0.6)**	**<0.0001**	-	Ethnic group (gbe, non gbe) - Population density (as ordinal variable)	**0.5 (0.2 to 0.97)**	**0.042**
	L4 (ref)	Centre	South (ref)	0.5 (0.2 to 1.1)	0.075	-	Ethnic group (gbe, non gbe)	0.4 (0.1 to 1.1)	0.064
		North		**0.5 (0.3 to 0.9)**	**0.009**	-	Population density (as ordinal variable)	0.6 (0.3 to 1.4)	0.243
L5	nonL5 (ref)	Centre	South (ref)	0.5 (0.2 to 1.2)	0.122	-	Ethnic group (**non gbe, East-gbe, centr-gbe, west-gbe**)	0.4 (0.2 to 1.2)	0.106
		North		**0.4 (0.3 to 0.6)**	**<0.0001**	-	Department of residence (south, centre, north)	0.5 (0.2 to 1)	0.054
	L4 (ref)	Centre	South (ref)	0.5 (0.2 to 1.1)	0.075	-	Ethnic group (**non gbe, East-gbe, centr-gbe, west-gbe**)	0.4 (0.2 to 1.1)	0.068
		North		**0.5 (0.3 to 0.9)**	**0.009**	-	Department of residence (south, centre, north)	0.6 (0.3 to 1.4)	0.26

Table 6. Multivariable analyses for L5 and L6 association with individual ethnicity, and L6 association with population density, department of patient residence

Variable 1		Variable 2		Crude OR (95%CI)	Crude p	Adjusted for		Adjusted OR	Adjusted p
lineage		**Ethnic group**							
L5	nonL5 (ref)	1=Fon	Peulh	**5.2 (1.8 to 14.8)**	**0.002**	-	Population density of residence (ordinal)	**6.3 (1.3 to 30)**	**0.021**
		2=Aizo		**6.6 (2.2 to 19.3)**	**0.001**			**8.2 (1.7 to 39.9)**	**0.009**
		3=Goun		**4.9 (1.6 to 14.9)**	**0.006**			**6.7 (1.3 to 33.8)**	**0.022**
		4=Sahoue		**11 (3.5 to 34.3)**	**<0.0001**	-	Department of residence (south, centre, north)	**12.9 (2.5 to 65.7)**	**0.002**
		6=Pedah		**6.3 (1.6 to 24.1)**	**0.007**			**8.5 (1.4 to 50.3)**	**0.019**
		7=Weme		**11 (2.9 to 42.3)**	**<0.0001**			**15.7 (2.6 to 94.6)**	**0.003**
		10=Kotafon		**13.2 (2.8 to 63.3)**	**0.001**			**16.6 (2.4 to 117)**	**0.005**
		11=Toffin		**13.7 (2.6 to 72.8)**	**0.002**			**13.8 (1.7 to 110.2)**	**0.013**
		13=Adja		**6.6 (2.2 to 19.7)**	**0.001**			**8.1 (1.6 to 39.5)**	**0.010**
		18=Nago		**7.3 (2 to 26.9)**	**0.003**			**9.3 (1.7 to 51.5)**	**0.011**
		27=Ditamari		**33 (2.8 to 395.6)**	**0.006**			**40.7 (3.1 to 526)**	**0.005**
		30=Kotocoli		**11 (1.2 to 100.4)**	**0.034**			**13.4 (1.4 to 132)**	**0.026**
		33=Yoa		**8.8 (1.7 to 46.6)**	**0.011**			**11 (1.9 to 65.1)**	**0.008**
		Other (non significant)		Not significant				Not significant	
	L4 (ref)	1=Fon	Peulh	**3 (1.02 to 9.1)**	**0.047**	-	Population density of residence (ordinal)	3.7 (0.8 to 17.6)	0.100
		2=**Aizo**		**3.9 (1.3 to 12)**	**0.018**			**4.9 (1.007 to 24)**	**0.049**
		3=Goun		2.7 (0.8 to 8.6)	0.098			3.6 (0.7 to 18.1)	0.124
		4=**Sahoue**		**6.5 (2 to 21.4)**	**0.002**			**7.8 (1.5 to 40.1)**	**0.014**
		6=Pedah		**4.7 (1.1 to 19.5)**	**0.035**			5.7 (0.9 to35.2)	0.060

		7=**Weme**		**5.8 (1.4 to 23.6)**	**0.013**	-	Department of	**8.2 (1.3 to 49.8)**	**0.023**
		10=**Kotafon**		**6.3 (1.3 to 31.1)**	**0.024**		residence (south,	**7.7 (1.1 to 54.4)**	**0.040**
		11=Toffin		**6.6 (1.2 to 35.7)**	**0.030**		centre, north)	6.4 (0.8 to 51.3)	0.079
		13=Adja		**3.6 (1.2 to 11.3)**	**0.027**			4.5 (0.9 to 22.1)	0.063
		18=Nago		3.8 (0.98 to 14.4)	0.054			4.7 (0.8 to 26.2)	0.077
		27=Ditamari		1				1	
		30=Kotocoli		5.3 (0.6 to 49)	0.145			6.3 (0.6 to 63.8)	0.117
		33=Yoa		4.2 (0.8 to 22.9)	0.097			5.1 (0.8 to 31.5)	0.078
		Other (non significant)		Non significant				Not significant	
L6	nonL6 (ref)	Peulh	Fon (ref)	**3.9 (1.7 to 8.7)**	**0.001**	-	Population density of residence	**3.6 (1.3 to 9.6)**	**0.012**
		Bariba		**5.9 (1.8 to 18.6)**	**0.002**		(ordinal)	3.9 (0.95 to 15.9)	0.058
		Gando		**6.5 (1.1 to 37.6)**	**0.037**	-	Department of	6.3 (0.9 to 41.6)	0.057
		Other (non significant) grouped		0.8 (0.5 to 1.4)	0.399		residence (south, centre, north)	Ref	Ref
						(Ref: Fon + Other non significant)			
	L4 (ref)	Peulh	Fon (ref)	**4 (1.7 to 9.5)**	**0.001**	-	Population density	**3.6 (1.3 to 9.9)**	**0.012**
		Bariba		**7.7 (2.1 to 28.9)**	**0.002**		of residence	**5 (1.1 to 23.2)**	**0.041**
		Gando		3.9 (0.7 to 22.4)	0.132		(ordinal)	3.7 (0.5 to 22.9)	0.196
		Other (non significant) grouped		0.8 (0.5 to 1.4)	0.404	-	Department of residence (south, centre, north)	Ref	Ref

					(Ref: Fon + Other non significant)			
		Population density						
L6	nonL6 (ref)	As ordinal variable	0.9 (0.8 to 1.1)	0.234	-	Ethnicity (individual)	**1.4 (1.05 to 1.7)**	**0.016**
			categorical		-	Department of residence (south, centre, north)		
			0.4 (0.2 to 0.7) for intermediary vs low	**0.001**				
				0.014				
			0.5 (0.3 to 0.9) for high vs low					
	L4 (ref)	As ordinal variable	0.9 (0.8 to 1)	0.167	-	Ethnicity (individual)	**1.3 (1.03 to 1.7)**	**0.026**
			categorical		-	Department of residence (south, centre, north)		
			0.5 (0.3 to 0.9) for intermediary vs low	**0.023**				
				0.028				
			0.6 (0.3 to 0.9) for high vs low					
		Department of residence			-			

L6	nonL6 (ref)	Centre	South	1.3 (0.4 to 4.5)	0.640	-	Ethnicity (individual)	1.3 (0.1 to 10.8)	0.832
		North		**2.9 (1.8 to 4.7)**	**<0.0001**	-	Population density (as ordinal)	2.8 (0.6 to 12.4)	0.187
	L4 (ref)	Centre	South	1 (0.3 to 3.4)	0.991	-	Ethnicity	0.6 (0.04 to 8.7)	0.721
		North		**2.8 (1.7 to 4.6)**	**<0.0001**		(individual)	3.2 (0.7 to 13.9)	0.123
						-	Population density (as ordinal)		

Table 7. Multivariable analyses for the association of L5 with patient residence department

Variable 1		Variable 2		Crude OR (95%CI)	Crude p	Adjusted for		Adjusted OR	Adjusted p
Lineage		**Residence department**							
L5	nonL5 (ref)	2=Atlantique,	Littoral (ref)	**1.9 (1.3 to 2.7)**	**<0.0001**	-	Sex	**1.7 (1.002 to 2.8)**	**0.049**
		3=oueme,		1 (0.6 to 1.7)	0.906	-	Age		0.633
		4=plateau,		**2.5 (1.04 to 6)**	**0.041**	-	Type patient (new, relapse, failure, default)		0.194
		5=mono,		**2.7 (1.6 to 4.7)**	**<0.0001**			**2.4 (1.1 to 5)**	**0.022**
		6=couffo,		**2 (1.2 to 3.4)**	**0.009**				0.302
		7=zou,		1 (0.6 to 1.6)	0.875	-	Ethnicity (single/same for father and mother)		0.427
		8=collines,		0.8 (0.3 to 1.9)	0.590				0.548
		9=borgou,		0.2 (0.1 to 1)	0.050				0.221
		10= alibori,		**0.2 (0.04 to 0.8)**	**0.021**	-	Population density of patient residence*		0.200
		11=atacora,		0.9 (0.4 to 1.9)	0.713				0.796
		12=donga		1.3 (0.6 to 2.9)	0.515				0.638
L5	L4 (ref)	2=Atlantique,	Littoral (ref)	**1.8 (1.3 to 2.7)**	**0.002**	-	Sex	**1.7 (1.002 to 2.9)**	**0.049**
						-	Age		
		3=oueme,		0.9 (0.5 to 1.4)	0.561	-	Type patient (new, relapse, failure, default)		0.321
		4=plateau,		2.2 (0.9 to 5.6)	0.085				0.326
		5=mono,		**2.9 (1.7 to 5.1)**	**<0.0001**			**2.7 (1.2 to 6)**	**0.012**
		6=couffo,		**1.8 (1.04 to 3.1)**	**0.035**	-	Ethnicity (single/same for father and mother)		0.341
		7=zou,		0.8 (0.5 to 1.4)	0.545				0.334
		8=collines,		0.7 (0.3 to 1.6)	0.362				0.442
		9=borgou,		0.4 (0.1 to 1.4)	0.165				0.557

	10= alibori,	0.3 (0.06 to 1.3)	0.104	-	Population density of patient residence*		0.527
	11=atacora,	1 (0.4 to 2.2)	0.924				0.544
	12=donga	1.2 (0.5 to 2.8)	0.625				0.644

* note: HIV not associated with L5 nor residence department, so HIV is not a confounder between L5 and residence department

Table 8. Multivariable analyses for the association of L6 with patient residence department

Variable 1		Variable 2		Crude OR (95%CI)	Crude p	Adjusted for		Adjusted OR	Adjusted p
Lineage		**Residence department**							
L6	nonL6(ref)	2=Atlantique,	Littoral (ref)	0.6 (0.3 to 1.001)	0.051	-	Sex		0.398
						-	Age		
		3=oueme,		**0.2 (0.1 to 0.7)**	**0.006**	-	Type patient (new, relapse, failure, default)	**0.2 (0.05 to 0.7)**	**0.016**
		4=plateau,		1				1	
		5=mono,		0.5 (0.2 to 1.4)	0.201				0.961
		6=couffo,		0.4 (0.1 to 1.2)	0.105	-	Ethnicity (single/same for father and mother)		0.812
		7=zou,		0.5 (0.2 to 1.2)	0.122				0.724
		8=collines,		0.8 (0.2 to 2.7)	0.668				0.821
		9=borgou,		**3.2 (1.4 to 7.4)**	**0.006**	-	Population density of patient residence*		0.095
		10= alibori,		**2.6 (1.1 to 6.1)**	**0.029**				0.318
		11=atacora,		0.9 (0.3 to 2.6)	0.786				0.699
		12=donga		0.5 (0.1 to 2.2)	0.370				0.650
L6	L4 (ref)	2=Atlantique,	Littoral (ref)	0.7 (0.4 to 1.2)	0.217	-	Sex		0.682
						-	Age		
		3=**oueme**,		**0.2 (0.1 to 0.6)**	**0.004**	-	Type patient (new, relapse, failure, default)	**0.2 (0.05 to 0.7)**	**0.012**
		4=plateau,		1				1	
		5=mono,		0.8 (0.3 to 2.3)	0.703				0.526
		6=couffo,		0.5 (0.2 to 1.5)	0.194	-	Ethnicity (single/same for father and mother)		0.986
		7=zou,		0.5 (0.2 to 1.1)	0.090				0.558
		8=collines,		0.6 (0.2 to 2.2)	0.465				0.985
		9=borgou,		**3 (1.2 to 7.3)**	**0.015**				0.101

	10= alibori,	**2.9 (1.1 to 7.1)**	**0.024**	-	Population density of patient residence*		0.198
	11=atacora,	0.9 (0.3 to 2.9)	0.914				0.633
	12=donga	0.5 (0.1 to 2.5)	0.425			0.7 (0.1 to 8.2)	0.744

* note: HIV not associated with L6 nor residence department, so HIV is not a confounder between L5 and residence department

Table 9. Association of L1 with patient residence department

Lineage		**Residence department**		OR (95%CI)	P
L1	nonL1(ref)	2=Atlantique,	Littoral (ref)	Not significant	
		3=oueme,		Not significant	
		4=plateau,		Not significant	
		5=mono,		Not significant	
		6=couffo,		Not significant	
		7=zou,		Not significant	
		8=collines,		Not significant	
		9=borgou,		6.8 (0.9 to 50.1)	0.059
		10= **alibori**,		**17.6 (3.3 to 94.3)**	**0.001**
		11=atacora,		Not significant	
		12=donga		Not significant	
L1	L4 (ref)	2=Atlantique,	Littoral (ref)	Not significant	
		3=oueme,		Not significant	
		4=plateau,		Not significant	
		5=mono,		Not significant	
		6=couffo,		Not significant	
		7=zou,		Not significant	
		8=collines,		Not significant	
		9=borgou,		7.5 (0.99 to 56.6)	0.051
		10= **alibori**,		**20 (3.6 to 110.3)**	**0.001**
		11=atacora,		Not significant	
		12=donga		Not significant	

Table 10. Multivariable analyses for the association of L5, L6 and L1 with treatment outcome

Variable 1		Variable 2		Crude OR (96%CI)	Crude p	Adjusted for	Adjusted OR (96% CI)	Adjusted p
lineage		**Treatment outcome**						
L5	nonL5 (ref)	failed	cured	1 (0.4 to 2.2)	0.922	- RIF resistance by Xpert,	1.1 (0.4 to 2.8)	0.814
		died		0.9 (0.4 to 2.2)	0.794	- HIV,	0.9 (0.3 to 2.3)	0.766
		ltfu		0.9 (0.4 to 2.3)	0.899	- Patient type before treatment (new, relapse, failure, default),	0.8 (0.2 to 2.6)	0.661
		completed		**0.6 (0.2 to 0.9)**	**0.016**		**0.4 (0.2 to 1)**	**0.038**
						- Sex,		
						- Age		
L5	L4 (ref)	failed	cured	1.1 (0.4 to 2.6)	0.899	- RIF resistance by Xpert,	1.4 (0.6 to 3.9)	0.493
		died		1 (0.4 to 2.6)	0.993	- HIV,	1 (0.3 to 3)	0.969
		ltfu		0.9 (03 to 2.2)	0.749	- Patient type before treatment (new, relapse, failure, default),	0.6 (0.2 to 2)	0.406
		completed		**0.4 (0.2 to 0.8)**	**0.011**		**0.4 (0.2 to 0.9)**	**0.026**
						- Sex,		
						- Age		
L6	nonL6 (ref)	failed	cured	0.6 (0.1 to 3.8)	0.611	- RIF resistance by Xpert,	0.6 (0.1 to 4.6)	0.628
		died		2.6 (0.8 to 7.7)	0.097	- HIV,	3.1 (1 to 9.8)	0.069
		ltfu		0.6 (0.1 to 4.4)	0.696	- Patient type before treatment (new, relapse, failure, default),	1	
		completed		1 (0.4 to 2.6)	0.991		1.2 (0.4 to 3.4)	0.797

						-	Sex,		
						-	Age		
L6	L4 (ref)	failed	cured	0.6 (0.1 to 4.6)	0.698	-	RIF resistance by Xpert,	0.8 (0.1 to 7)	0.876
		died		2.6 (0.8 to 8)	0.123	-	HIV,	**3.6 (1.01 to 12.2)**	**0.048**
		ltfu		0.6 (0.1 to 4.1)	0.648	-	Patient type before treatment (new, relapse, failure, default),	1	
		completed		0.8 (0.3 to 2)	0.610			0.9 (0.3 to 2.7)	0.863
						-	Sex,		
						-	Age		
L1	nonL1	failed	cured	**13 (3.3 to 61.2)**	**<0.0001**	-	RIF resistance by Xpert,	**11.9 (2.8 to 60.6)**	**0.001**
		died		1		-	HIV,	-	-
		ltfu		1		-	Patient type before treatment (new, relapse, failure, default),	-	-
		completed		1.6 (0.2 to 11.7)	0.719			-	-
						-	Sex,		
						-	Age		
L1	nonL1 (ref)	failed	cured	**13 (3.3 to 61.2)**	**<0.0001**	-	RIF resistance by Xpert,	**47.7 (6.7 to 398)**	**<0.0001**
						-	HIV,		
						-	Patient type before treatment (new, relapse, failure, default),		
						-	Sex,		
						-	Age		

						-	**Residence department** (the 12 departments in Benin)		
L1	L4 (ref)	failed	Cured	**12.7 (3.1 to 62.3)**	**<0.0001**	-	RIF resistance by Xpert,	**18.8 (3.6 to 98.9)**	**0.001**
		died		1		-	HIV,	-	-
		ltfu		1		-	Patient type before treatment (new, relapse, failure, default),	-	-
		completed		1.1 (0.1 to 9.2)	0.900	-	Sex,	-	-
						-	Age		
L1	L4 (ref)	failed	cured	**12.7 (3.1 to 62.3)**	**<0.0001**	-	RIF resistance by Xpert,	**43.7 (4.6 to 418.4)**	**0.001**
						-	HIV,		
						-	Patient type before treatment (new, relapse, failure, default),		
						-	Sex,		
						-	Age		
						-	**Residence department** (the 12 departments in Benin)		

Trends of lineage prevalence over time

Lineage distribution varied significantly across TB clinics for new and previously-treated patients (Tables S9, S10, S11). Comparison of the lineage prevalence of the current sputum-based study and the 2005-2006 isolates-based study in the CNHU-PPC clinic in Cotonou, showed that L5 and L1 are decreasing (for L5 Diff: -9.4 (-17 to -1.2), p=0.023; and for L1 p=0.0006, Table S12, Fig S1). Conversely, L4 prevalence is increasing (Diff:16% (7.4 to 24.6), p=0.0003). Also when the prevalence was not adjusted for sputum-based spoligotyping, the observation remained the same for L1 and L4 (Table S12, Fig S1).

Discussion

We conducted a nationwide study of the molecular epidemiology of TB in Benin, the country with the highest L5 prevalence worldwide [2,3,5,6], to measure for the first time the unbiased population structure of MTBC lineages among new and previously-treated TB patients (over 1500 from 2016 to 2018), and assess whether L5 prevalence is decreasing, and also assess whether L5 is associated with TB retreatment/recurrence, as well as possible association between lineages and TB presentation, diagnosis and outcome.

Our findings showed that in Benin, either in new or previously-treated patients, L4 is the commonest lineage followed by L5. L5 is significantly under-represented among previously-treated patients, unlike L4.

The current distribution of MTBC lineages among new patients in Cotonou compared to the adjusted previous (2005-2006) prevalence also among new patients in the same clinic [3], showed that L5 prevalence significantly decreased. Without correcting for the potential culture bias, the over-time

change in L5 prevalence was not significant, yet this is an unfair comparison of things of different nature, as L5 is under-represented in positive cultures both in the current and previous study [39]. The prevalence of L1 decreased as well, whereas the L4 prevalence increased significantly, when adjusted to sputum-based estimates or not, likely in replacement of the decreasing L5 and L1. This confirms the hypothesis of transmission out-competition between modern and ancestral lineages [23]. The fact that L5 prevalence was found stable over 8 years in Ghana [58] could be due to the non-distinction between new and previously-treated patients (inclusion just as TB patients), and non-proportional (differential) inclusion of new and previously-treated patients at the time-points compared. That would tend to cancel/reduce the significant L5 prevalence change, as we found more L5 in new and less in previously-treated patients. The L5 stability in Ghana could also be due to the isolate-based typing approach [58]. In our study, the prevalence of L6 was not significantly different among new and previously-treated patients, and no change over time was observed. Also in the Gambia, the prevalence of L6 remained stable over many years [14], albeit no new and previously-treated patients were not analyzed separately in that study. That overtime decline in L5 prevalence in Cotonou, could be due to the reduced (recent) transmission of L5 as found in urban area in Ghana (rural area not investigated) [23], the slower progression to disease found for L6 [26] if also applicable for L5 and finally the increasing population density in Cotonou/Littoral (665100 population in 2002 vs 678874 populations in 2013 https://www.insaebj.org/images/docs/insaestatistiques/demographiques/mortalite/Mortalite.pdf), as our findings showed that L5 is more adapted to decreasing population density areas. In this study, the overtime trend of lineage prevalence was compared only in Cotonou/Littoral a region of high population

density where previous data on lineage distribution in new patients are available. The trends of L5 prevalence overtime in areas with intermediary or low population density should also be investigated.

Our findings show that, in Benin, new and previously-treated TB patients differ not only in terms of age, sex and treatment outcome, but also in lineage distribution, stressing the importance of integrating the patient's TB treatment history in such analysis. Another important aspect is whether the population selected for the study is representative. The distribution of lineages greatly varied across regions in Benin, as it has been observed in Ghana [58] and Senegal [59], and can be quite variable within a region. We thus recommend, for molecular epidemiological studies, a proportional (to the number of TB patients diagnosed in routine care) sampling of the regions within the study area that includes sufficient new and previously-treated patients for a stratified analysis.

In our study population, L5 was associated with decreasing population density while L6 was associated with increasing population density, after adjustment for potential confounders in a multivariable analysis. Comas and colleagues had previously hypothetized that one of the reasons why *M. africanum* L5 and L6 are geographically restricted to West-Africa could be that they are more adapted to lower population densities [16]. In our study, both L5 and L6 were also associated to specific ethnic groups, with L5 being associated with Gbe ethnic continuum, albeit not in a systematic manner; and Aizo, sahoue, Weme, Kotafon that are part of Eastern-Gbe ethnicities. In one of two southern departments, Littoral (Cotonou), L5 was associated with Gbe ethnicity, while in the second department, Atlantique - with a high L5 prevalence - the association did not hold. The two departments differ in population density, which is

intermediary in Atlantique, and higher in Littoral. But after adjusting for population density and geographical situation of residence department (South, Centre, North), the association of L5 with Gbe ethnicity remained significant, especially with L5 associated with Eastern-Gbe relative to L4. So, after adjusting for potential confounders, L5 was associated with both population density (decreasing when L5 increases) and Gbe ethnicity (especially Eastern-Gbe). Interestingly, L5 (relative to L4) was over-represented in Atlantique and Mono independent of ethnicity, population density of patient residence, patient type, age and sex; which means that this association is only due to the location/department (or something specific to it).

Gbe communities live in the South-Eastern region of West-Africa including from west to east: South-Eastern Ghana, Southern Togo, Southern Benin and South-Western Nigeria [21]. Our findings showed that within the Gbe ethnic continuum, L5 was associated with Eastern-gbe relative to L4 (and Central- and Eastern-Gbe relative to non-L5 lineages), but not to Western-Gbe which includes the Ewe ethnic group reported in Ghana to be associated with L5 [13,20]. This also correlates with the fact that L5 is mostly prevalent in Eastern West-Africa. As a person is made of 50% of the genetic information from each of his/her two parents, in our study we classified a patient in a specific ethnic group, when the patient's father and mother shared the same ethnicity, and excluded from the ethnic group analysis patients whose parents have different ethinicity. Although this approach does not take into account the ethnic variation of the generation(s) before the patient's parents, it makes our findings stronger compared to the approach assigning ethnicity based on his/her father only (which is common practice in epidemiological studies). We found after adjusting for potential confounders that L6 is significantly associated with Peulh (Fulani [60]) and Bariba ethnicities relative to L4. In The Gambia, L6 was also

found associated with Fulani ethnicity [26]. Fulani are cattle herders. Also in Benin, L6 was over-represented in the north where cattle herding and farming are principal activities and where *M. bovis* is prevalent. L6 and *M. bovis* are phylogenetically closely related, as part of the clade of animal-adapted MTBC clade [1]. Our hypothesis is that *M. bovis* descended from the rest of the MTBC in West African cattle herders. Fulani people (herdsmen) are nomadic or semi-nomadic, and are often found in small groups in the Sahel or semi-arid regions[60]. The association of L6 with the Fulani ethnicity could also partly explain why Sahel countries such as Mali, Senegal, Gambia, Burkina-Faso are L6-endemic countries [2,6]. However, in this study we did not find an association with lineage and reported cattle contact.

Previously-treated TB was significantly associated with lower sputum-microscopy grade at baseline, compared to new patients. Furthermore, previously-treated L5 patients were more likely to produce paucibacillary sputum compared to those with L4 strains. This could be a problem for TB diagnostic in previously-treated patients in general, and especially for sputum-based testing to overcome the problem of reduced in vitro growth of *M. africanum* [39]. Also the fact that L5 is actually over-represented among failure patients with AFB-scanty sputa compared to relapse and default patients, unlike L4 (no difference among failure, relapse and default), might suggest that L5 is implicated in false microscopy-detected TB treatment failures (dead bacilli), and that needs to be further investigated.

Result from in vitro culture was lineage dependent, with L5, L6 and *M. bovis* under-represented in positive cultures relative to L4, and a higher likelihood for dysgonic growth, corroborating on a larger sample size what we previously reported [39][61]. This was also confirmed after adjusting for potential

confounders such as patient type (relapse, failure, default), sputum AFB-grade. The reduced growth in culture was also observed for L6 in The Gambia in liquid medium (MGIT; unpublished data). For long *M. africanum and M. bovis* are known for the dysgonic nature of their isolates[62]. *M. bovis* cannot use glycerol as sole carbon source, and it is recommended to use pyruvate supplemented medium to allow an optimal growth in vitro. That pyruvate supplementation is meant to favor the growth (eugonic isolates) of L5, L6 as well, as they have with *M. bovis* a SNP in the *pykA* gene encoding the pyruvate kinase necessary for the metabolism of carbohydrates such as glycerol and pyruvate[63]. However, despite we used a standard and pyruvate-supplemented LJ slant for each specimen, we still got dysgonic isolates, even though not all strains of these lineages were dysgonic.

L5-patients were less likely to have a previous treatment history, mainly less TB relapse. However, among previously-treated patients but not in new patients, L5 was significantly associated with higher RIF-resistance rate. This could be due to acquisition of RIF resistance during first-line TB treatment, as most (4/5) of the RIF-resistant L5 strains in our study were identified from patients who failed treatment. Acquisition of RIF resistance was document in one patient who was included in the study as a new patients and subsequently failed treatment despite good treatment adherence. This is potentially threatening for L5-endemic countries, albeit current numbers are too low and more detailed strain typing is required to draw firm conclusions

Despite the fact that L5-patients were less likely to have a previous relapse compared to L4, we did not find any association with a lineage and relapse during the one-year follow up, potentially due to the short follow-up duration. Checking for relapse after a period longer than one year after the TB cure could

be more informative, although recurrence of TB beyond 2 years after cure in high endemic countries, is mostly due to exogenous reinfection rather than endogenous reactivation.

As previously reported for L6 [14], L5 as well is associated with older age of patients relative to L4 (also confirmed after adjusting for patient type (relapse, failure, default), sex, HIV, sputum AFB-grade). However, in this study the L6 association with older age and HIV positivity [14,20] were not found. This could be due to a lower power of the study for L6 analysis as L6 is not common in Benin. Also, we did not find any significant association of L5 with HIV positivity, unlike L6 found to be associated with HIV positivity in Ghana and The Gambia [14].

L5 was over-represented in female patients relative to L4 (also confirmed after adhusting for patient type (relapse, failure, default), age, HIV, sputum AFB-grade). Whether this finding points to a source/reservoir other than human for L5 that could be related to an activity/occupation done by women, remains unclear. In contrast, the proportion of males among L6 patients was similar to that of L4 patients, with male over-represented in L6 relative to L5. This observation could hint to different reservoirs of L5 and L6, if any.

In this study, L2 (Beijing) was associated with STR and EMB resistance. In the 2005-2006 study in Cotonou, an outbreak with a STR-resistant Beijing strain was identified [4]. Further genotyping beyond spoligotyping - that can't differentiate strains of the Beijing lineage - and contact tracing are needed to elucidate whether this observation is linked to an outbreak as well.

L1 was associated with lower treatment success and increased treated failure that is also confirmed after adjusting for RIF resistance by Xpert, HIV status,

patient type before treatment (new, relapse, failure, default), residence department (the 12 departments), sex and age (p=0.001). Measures should be taken in Benin to enhance directly observed treatment strategies (DOTS) and encourage treatment compliance especially in Alibori where there is the highest prevalence of L1 in Benin.

Likewise, after adjusting for RIF resistance by Xpert, HIV status, patient type before treatment (new, relapse, failure, default), L6 became associated with increased patient death during TB treatment relative to L4 (p=0.048). A recent study in Mali by Diarra et al showed that L6 is associated with a slower smear conversion during TB treatment relative to L4, which is also indicative of a decreased response to the current TB treatment regimen for L6 strains. Further studies should be conducted to evaluate the effectiveness of the current TB treatment drug regimen for L6.

The strengths of this study include its prospective nature, with a representative sampling of all regions, a proportional sampling of new and previously-treated patients, and the use of sputum-based genotyping (direct spoligotyping followed by direct PhyloSNP for specimens with uncommon spoligotypes). As limitations, some information such as contact with cattle and immigration were collected by questioning patients and their response could not be verified. For ethnicity, this problem is unlikely as the information about parents' region of origin on the questionnaire was used to cross-check the information given by the patient, as ethnicity is closely linked with region of origin. The variable 'hospitalization leading to TB diagnosis' could be misunderstood by some nurses and considered as hospitalization for TB treatment, despite the monitoring (supervision, meeting, telephone calls). For future studies, recording patient occupation could help in finding possible sources of infection

(environmental, animal) of L5 and L6. Another limitation is that some specimens classified as 'unclear' after PhyloSNP were PCR-negative for the two PhyloSNP PCRs, although they were confirmed MTBC members by the presence of spacers on the spoligotyping membrane. If an outbreak of a strain occurred during the study in an area (TB clinic, or village, or town), this could inflate the specific lineage prevalence trends in that area. Direct WGS is the best alternative for genotyping yet it is not yet available and researchers are struggling for that.

In conclusion, the distribution of MTBC lineages significantly varied between new and previously-treated patients. The distribution of L5 and L6 was associated with patient ethnicity and population density. L5, L6 and *M. bovis* were less likely to grow in culture, and mostly yield isolates of dysgonic nature when they did. In Cotonou, L5 and L1 prevalence are decreasing and being replaced by L4. Molecular epidemiology studies should proportionally include new and previously-treated TB patients, and analyze both groups separately. Sputum-based genotyping techniques (such as direct spoligotyping followed by direct PhyloSNP when necessary) should be used in such studies (for lineage determination). Further investigation should be conducted on the possible association of L5 with the acquisition of RIF-resistance and the possibility of false failure detected using microscopy (auramine, Ziehl-neelsen) in patients with L5 strains.

Acknowledgments

We thank all the personnel of the 24 participating TB clinics (listed in Table 1), for their participation into the study. We also thank Houéto O. Félix (SENATEL) for providing the updated list of villages in arrondissements within districts of each department in Benin. We finally than the personnel of the radiology service of the National University Hospital for Lung diseases in Cotonou (CNHU-PPC) for their technical assistance with radiofilm development for spoligotype-analysis.

References

[1] Coscolla M, Gagneux S. Consequences of genomic diversity in mycobacterium tuberculosis. Semin Immunol 2014;26:431–44. doi:10.1016/j.smim.2014.09.012.

[2] Gehre F, Kumar S, Kendall L, Ejo M, Secka O, Ofori-Anyinam B, et al. A Mycobacterial Perspective on Tuberculosis in West Africa: Significant Geographical Variation of M. africanum and Other M. tuberculosis Complex Lineages. PLoS Negl Trop Dis 2016;10. doi:10.1371/journal.pntd.0004408.

[3] Affolabi D, Anyo G, Faïhun F, Sanoussi N, Shamputa IC, Rigouts L, et al. First molecular epidemiological study of tuberculosis in Benin. Int J Tuberc Lung Dis 2009;13.

[4] Affolabi D, Faïhun F, Sanoussi N, Anyo G, Shamputa IC, Rigouts L, et al. Possible outbreak of streptomycin-resistant Mycobacterium tuberculosis Beijing in Benin. Emerg Infect Dis 2009;15. doi:10.3201/eid1507.080697.

[5] Gehre F, Antonio M, Faïhun F, Odoun M, Uwizeye C, de Rijk P, et al. The First Phylogeographic Population Structure and Analysis of Transmission Dynamics of M. africanum West African 1- Combining Molecular Data from Benin, Nigeria and Sierra Leone. PLoS One 2013;8:8–13. doi:10.1371/journal.pone.0077000.

[6] De Jong BC, Antonio M, Gagneux S. Mycobacterium africanum-Review of an Important Cause of Human Tuberculosis in West Africa 2010. doi:10.1371/journal.pntd.0000744.

[7] Ngo Niobe-Eyangoh S, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, et al. Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon. J Clin Microbiol 2003;41:2547–53. doi:10.1128/JCM.41.6.2547-2553.2003.

[8] Koro FK, Simo YK, Piam FF, Noeske J, Gutierrez C, Kuaban C, et al. Population dynamics of tuberculous bacilli in cameroon as assessed by spoligotyping. J Clin Microbiol 2013;51:299–302. doi:10.1128/JCM.01196-12.

[9] Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, et al. The Guinea-Bissau family of Mycobacterium tuberculosis complex revisited. PLoS One 2011;6:e18601. doi:10.1371/journal.pone.0018601.

[10] Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, et al. First Molecular Epidemiology Study of Mycobacterium tuberculosis in Burkina Faso. J Clin Microbiol 2007;45:921–7. doi:10.1128/JCM.01918-06.

[11] Lawson L, Zhang J, Gomgnimbou MK, Abdurrahman ST, Le Moullec S, Mohamed F, et al. A molecular epidemiological and genetic diversity study of tuberculosis

in Ibadan, Nnewi and Abuja, Nigeria. PLoS One 2012;7:e38409. doi:10.1371/journal.pone.0038409.

[12] Thumamo BP, Asuquo AE, Abia-Bassey LN, Lawson L, Hill V, Zozio T, et al. Molecular epidemiology and genetic diversity of Mycobacterium tuberculosis complex in the Cross River State, Nigeria. Infect Genet Evol 2012;12:671–7. doi:10.1016/j.meegid.2011.08.011.

[13] Asante-Poku A, Yeboah-Manu D, Otchere ID, Aboagye SY, Stucki D, Hattendorf J, et al. Mycobacterium africanum Is Associated with Patient Ethnicity in Ghana. PLoS Negl Trop Dis 2015. doi:10.1371/journal.pntd.0003370.

[14] de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, et al. Differences between tuberculosis cases infected with *Mycobacterium africanum* , West African type 2, relative to Euro-American *Mycobacterium tuberculosis* : an update. FEMS Immunol Med Microbiol 2010;58:102–5. doi:10.1111/j.1574-695X.2009.00628.x.

[15] Comas I, Hailu E, Kiros T, Bekele S, Mekonnen W, Gumi B, et al. Population Genomics of Mycobacterium tuberculosis in Ethiopia Contradicts the Virgin Soil Hypothesis for Human Tuberculosis in Sub-Saharan Africa. Curr Biol 2015;25. doi:10.1016/j.cub.2015.10.061.

[16] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet 2013;45. doi:10.1038/ng.2744.

[17] Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol 2018;16. doi:10.1038/nrmicro.2018.8.

[18] Gagneux S. Host-pathogen coevolution in human tuberculosis. Philos Trans R Soc Lond B Biol Sci 2012;367:850–9. doi:10.1098/rstb.2011.0316.

[19] Wikipedia. Early human migrations n.d. https://en.wikipedia.org/wiki/Early_human_migrations.

[20] Asante-Poku A, Otchere ID, Osei-Wusu S, Sarpong E, Baddoo A, Forson A, et al. Molecular epidemiology of Mycobacterium africanum in Ghana. BMC Infect Dis 2016;16:385. doi:10.1186/s12879-016-1725-6.

[21] Kluge Angela. A sociolinguistic survey of the Gbe language communities of Benin and Togo: Gbe language family overview | SIL International. SIL Electron Surv Reports 2011. https://www.sil.org/resources/publications/entry/41567 (accessed May 21, 2019).

[22] Sharma A, Bloss E, Heilig CM, Click ES. Tuberculosis Caused by Mycobacterium africanum, United States, 2004-2013. Emerg Infect Dis 2016;22:396–403. doi:10.3201/eid2203.151505.

[23] Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, Darko Otchere I, et al. Reduced transmission of Mycobacterium africanum compared to Mycobacterium tuberculosis in urban West Africa. Int J Infect Dis 2018;73:30–42. doi:10.1016/j.ijid.2018.05.014.

[24] de Jong BC, Hill PC, Brookes RH, Otu JK, Peterson KL, Small PM, et al. Mycobacterium africanum: a new opportunistic pathogen in HIV infection? AIDS 2005;19:1714–5. doi:10.1097/01.aids.0000185991.54595.41.

[25] Meyer CG, Scarisbrick G, Niemann S, Browne ENL, Chinbuah MA, Gyapong J, et al. Pulmonary tuberculosis: Virulence of Mycobacterium africanum and relevance in HIV co-infection. Tuberculosis 2008;88:482–9. doi:10.1016/J.TUBE.2008.05.004.

[26] Jong BC de, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by M. tuberculosis lineage in The Gambia. J Infect Dis 2008;198:1037. doi:10.1086/591504.

[27] Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14. doi:10.1016/S0305-4403(02)00239-X.

[28] Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One 2009;4:e7815. doi:10.1371/journal.pone.0007815.

[29] Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis † Downloaded from. J Clin Microbiol 2006;44:4498–510. doi:10.1128/JCM.01392-06.

[30] Thierry D. S6110, an IS-like element of Mycobacterium tuberculosis complex. vol. 18. n.d.

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

[32] Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R. High Resolution Discrimination of Clinical Mycobacterium tuberculosis Complex Strains Based on Single Nucleotide Polymorphisms. PLoS One 2012;7:39855. doi:10.1371/journal.pone.0039855.

[33] Meehan CJ, Moris P, Kohl TA, Pečerska J, Akter S, Merker M, et al. The relationship between transmission time and clustering methods in

Mycobacterium tuberculosis epidemiology. EBioMedicine 2018;37:410–6. doi:10.1016/j.ebiom.2018.10.013.

[34] Wyllie DH, Davidson JA, Grace Smith E, Rathod P, Crook DW, Peto TEA, et al. A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying Mycobacterium tuberculosis Transmission: A Prospective Observational Cohort Study. EBioMedicine 2018;34:122–30. doi:10.1016/j.ebiom.2018.07.019.

[35] Comas I. Genomic epidemiology of tuberculosis. Adv. Exp. Med. Biol., vol. 1019, 2017. doi:10.1007/978-3-319-64371-7_4.

[36] De Jong BC, Antonio M, Awine T, Ogungbemi K, De Jong YP, Gagneux S, et al. Use of spoligotyping and large sequence polymorphisms to study the population structure of the Mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in the gambia??? J Clin Microbiol 2009;47:994–1001. doi:10.1128/JCM.01216-08.

[37] Stucki D, Gagneux S. Single nucleotide polymorphisms in Mycobacterium tuberculosis and the need for a curated database. Tuberculosis (Edinb) 2013;93:30–9. doi:10.1016/j.tube.2012.11.002.

[38] De Jong BC, Antonio M, Gagneux S. Mycobacterium africanum-Review of an Important Cause of Human Tuberculosis in West Africa n.d. doi:10.1371/journal.pntd.0000744.

[39] Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. Genotypic characterization directly applied to sputum improves the detection of Mycobacterium africanum West African 1, under-represented in positive cultures. PLoS Negl Trop Dis 2017;11:e0005900. doi:10.1371/journal.pntd.0005900.

[40] Kamakoli MK, Khanipour S, Hadifar S, Ghajavand H, Farmanfarmaei G, Fateh A, et al. Challenge in direct Spoligotyping of Mycobacterium tuberculosis: a problematic issue in the region with high prevalence of polyclonal infections. BMC Res Notes 2018;11:486. doi:10.1186/s13104-018-3579-z.

[41] Rieder HL, Van A, Kai D, Kam M, Kim SJ, Chonde TM, et al. Priorities for Tuberculosis Bacteriology Services in Low-Income Countries Second edition 2007 International Union Against Tuberculosis and Lung Disease. 2007.

[42] Cepheid. Xpert MTB/RIF package insert. REF GXMTB/RIF-US-10. 2015.

[43] WHO. The shorter MDR-TB regimen. 2016.

[44] Castets M. [Mycobacterium africanum (author's transl)]. Med Trop (Mars) 1979;39:145–8.

[45] Rieder H, Deun A, Kam K, Kim S, Chonde T, Trébucq A, et al. Priorities for

Tuberculosis Bacteriology Services in Low-Income Countries. 2007.

[46] Bemer-Melchior P, Drugeon HB. Inactivation of Mycobacterium tuberculosis for DNA typing analysis. J Clin Microbiol 1999;37:2350–1.

[47] Manual T. Maxwell ® 16 DNA Purification Kits. Components n.d.:0–17.

[48] Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, et al. TB-Lineage: An online tool for classification and analysis of strains of Mycobacterium tuberculosis complex. Infect Genet Evol 2012;12:789–97. doi:10.1016/j.meegid.2012.02.010.

[49] TBminer n.d. https://info-demo.lirmm.fr/tbminer/ (accessed July 7, 2019).

[50] BCCM/ITM Mycobacteria Collection | BCCM Belgian Coordinated Collections of Microorganisms n.d. http://bccm.belspo.be/about-us/bccm-itm (accessed July 7, 2019).

[51] CLC Sequence Viewer USER MANUAL Manual for CLC Sequence Viewer 8.0.0. n.d.

[52] SnapGene Viewer | Free software for plasmid mapping, primer design, and restriction site analysis n.d. https://www.snapgene.com/snapgene-viewer/ (accessed July 7, 2019).

[53] INSAE Benin (Institut National de la Statistique et de l'Analyse économique). RGPH4: Que retenir des effectifs de population en 2013? (Résultats définitifs). 2015.

[54] INSAE Benin (Institut National de la Statistique et de l'Analyse économique). INSAE - Statistiques Démographiques n.d. https://www.insae-bj.org/statistiques/statistiques-demographiques (accessed July 7, 2019).

[55] Wikipedia. Waci language n.d.

[56] Gbe, Waci | Ethnologue n.d. https://www.ethnologue.com/language/wci (accessed July 7, 2019).

[57] Affolabi D, Sanoussi N, Codo S, Sogbo F, Wachinou P, Massou F, et al. First Insight into a Nationwide Genotypic Diversity of Mycobacterium tuberculosis among Previously Treated Pulmonary Tuberculosis Cases in Benin, West Africa. Can J Infect Dis Med Microbiol 2017;2017. doi:10.1155/2017/3276240.

[58] Yeboah-Manu D, Asare P, Asante-Poku A, Otchere ID, Osei-Wusu S, Danso E, et al. Spatio-Temporal Distribution of Mycobacterium tuberculosis Complex Strains in Ghana. PLoS One 2016;11:e0161892. doi:10.1371/journal.pone.0161892.

[59] Diop S, de Medeiros D, de Medeiros G, Baylet R, Sankalé M. [Incidence and

geographic distribution of Mycobacterium africanum in Senegal]. Bull Soc Med Afr Noire Lang Fr 1976;21:50–6.

[60] Wikipedia. Fula people. n.d.

[61] Sanoussi CN, De Jong BC, Odoun M, Arekpa K, Ligali MA, Bodi O, et al. Low sensitivity of the MPT64 identification test to detect lineage 5 of the Mycobacterium tuberculosis complex 2018. doi:10.1099/jmm.0.000846.

[62] Leao S, Martin A, Mejia GI, Portaels F, Leao SC, Martin A, Mejia GI, Palomino JC, Robledo J, Telles MAS P-TF. Practical handbook for the phenotypic and genotypic identification of mycobacteria. 2004:77–126. doi:10.1007/BF01045823.

[63] Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, et al. The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: Implications for in vivo growth. Mol Microbiol 2005;56:163–74. doi:10.1111/j.1365-2958.2005.04524.x.

Table S1. Selected (participating) TB clinics

Region	4 largest TB clinics (largest in number of patients)
Atlantic/Littoral	CNHU-PPC (Cotonou)
	Abomey-Calavi
	Allada
	Ouidah
Oueme/Plateau	CPP-Akron (Porto-Novo)
	Pobè
	Sakété
	Avrankou
Mono/Couffo	Aplahoué
	Comè
	Houeyogbé
	Bopa
Zou/Collines	Abomey
	Bohicon
	Zagnanado
	Dassa-Zoumé
Borgou/Alibori	CSC Parakou
	Bembèrèkè HE
	Kandi
	Nikki HSS
Atacora/Donga	Djougou
	Tanguiéta
	Natitingou
	Matéri

Table S2. Repartition of lineage by ethnicity (mixed ethnicity father-mother)

	Ethnicity (father and mother)	Lineages							Total % (n)	p
		L1 % (n)	L2 % (n)	L3 % (n)	L4 % (n)	L5 % (n)	L6 % (n)	M. bovis % (n)		
Single ethnicity (same ethnicity for father and mother)	**Adja**	0	1.8 (2)	0	54.5 (61)	37.5 (42)	4.5 (5)	1.8 (2)	100 (112)	**< 0.0001**
	Ahoussa	0	5 (1)	0	60 (12)	10 (2)	20 (4)	5 (1)	100 (20)	
	Aizo	1.4 (2)	4.9 (7)	0.7 (1)	50.7 (73)	37.5 (54)	4.2 (6)	0.7 (1)	100 (144)	
	Bariba	0	12.5 (2)	0	31.3 (5)	6.3 (1)	**31.3** (5)	**18.8 (3)**	100 (16)	
	Berba	**0**	0	13.3 (2)	60 (9)	20 (3)	6.7 (1)	0	100 (15)	
	Dendi	0	0	0	66.7 (6)	11.1 (1)	11.1 (1)	**11.1 (1)**	100 (9)	
	Ditamari	0	0	0	0	**75 (3)**	25 (1)	0	100 (4)	
	Ewe	0	33.3 (1)	0	66.7 (2)	0	0	0	100 (3)	
	Fon	0.3 (1)	4.4 (13)	0.3 (1)	55.1 (162)	32 (94)	7.1 (21)	0.7 (2)	100 (294)	
	Gando	0	0	0	66.7 (4)	0	**33.3 (2)**	0	100 (6)	
	Goun	0	2.3 (2)	0	60.2 (53)	30.7 (27)	6 (5)	1.1 (1)	100 (88)	
	Holli	0	50 (1)	0	0	**50 (1)**	0	0	100 (2)	
	Ibo	0	0	0	100 (3)	0	0	0	100 (3)	
	Idatcha	0	0	0	65.2 (15)	21.7 (5)	13 (3)	0	100 (23)	
	Kotafon	0	0	**0**	45.5 (5)	**54.6 (6)**	0	0	100 (11)	
	Kotocoli	0	0	0	50 (2)	**50 (2)**	0	0	100 (4)	
	Lokpa	0	0	0	100 (2)	0	0	0	100 (2)	
	Mahi	3.6 (1)	10.7 (3)	3.6 (1)	15 (53.6)	21.4 (6)	7.1 (2)	0	100 (28)	
	Mina	0	4.2 (1)	0	58.3 (14)	25 (6)	12.5 (3)	0	100 (24)	
	Nago	0	4 (1)	0	56 (14)	40 (10)	0	0	100 (25)	
	Ouatchi	0	11.1 (1)	0	55.6 (5)	22.2 (2)	11.1 (1)	0	100 (9)	
	Pedah	0	13.6 (3)	0	40.9 (9)	36.4 (8)	9.1 (2)	0	100 (22)	
	Peulh	**16.7** (8)	2.1 (1)	6.3 (3)	43.8 (21)	8.4 (4)	**22.9 (11)**	0	100 (48)	

	Sahoue	0	3.2 (2)	0	40.3 (25)	**50 (31)**	6.5 (4)	0	100 (62)
	Setto	25 (1)	0	0	75 (3)	0	0	0	100 (4)
	Toffin	0	0	0	44.4 (4)	**55.6 (5)**	0	0	100 (9)
	Tori	0	0	0	73.7 (14)	26.3 (5)	0	0	100 (19)
	Waama	0	0	0	55.6 (5)	11.1 (1)	**22.2 (2)**	**11.1 (1)**	100 (9)
	Weme	5 (1)	0	0	45 (9)	**50 (10)**	0	0	100 (20)
	Xwla	0	5.6 (1)	0	72.2 (13)	22.2 (4)	0	0	100 (18)
	Yoa	0	0	0	55.6 (5)	44.4 (4)	0	0	100 (9)
	Yom	0	0	0	80 (4)	0	0	**20 (1)**	100 (5)
	Yoruba	4.1 (2)	2 (1)	0	69.4 (34)	16.3 (8)	8.2 (4)	0	100 (49)
	Zerma	0	0	0	**100 (13)**	0	0	0	100 (13)
	Other	0	3 (1)	3 (1)	75.8 (25)	12.1 (4)	3 (1)	3 (1)	100 (33)
	TOTAL	**1.4 (16)**	**3.8 (44)**	**0.8 (9)**	**55.6 (646)**	**30 (349)**	**7.2 (84)**	**1.2 (14)**	**100 (1162)**

Table S3. Repartition of L5 by type of Gbe ethnicity per department of residence

Department		L5	Total	p
Atlantique	**Western Gbe**	44.4 (4)	100 (9)	0.586
	Central Gbe	42.1 (8)	100 (19)	0.804
	Eastern Gbe	37.8 (129)	100 (341)	0.725
	Non-Gbe	33.3 (5)	100 (15)	-
Littoral	**Western Gbe**	0	100 (14)	
	Central Gbe	20 (1)	100 (5)	
	Eastern Gbe	30.3 (40)	100 (132)	**0.009**
	Non-Gbe	8.6 (3)	100 (35)	**Diff: 21.7 (9.6 to 33.9)**
				OR: 4.6 (1.4 to 15)
Oueme	**Western Gbe**	25 (1)	100 (4)	
	Central Gbe	100 (1)	100 (1)	
	Eastern Gbe	25.7 (27)	100 (105)	0.636
	Non-Gbe	16.7 (4)	100 (24)	
Plateau	**Western Gbe**	0	0	
	Central Gbe	0	0	
	Eastern Gbe	23.3 (3)	100 (11)	0.479
	Non-Gbe	55.6 (8)	100 (20)	
Mono	**Western Gbe**	33.3 (3)	100 (9)	
	Central Gbe	0	100 (4)	
	Eastern Gbe	51.6 (32)	100 (62)	
	Non-Gbe	**0**	0	
Couffo	**Western Gbe**	0	0	
	Central Gbe	39.7 (29)	100 (73)	
	Eastern Gbe	40 (4)	100 (10)	
	Non-Gbe	**0**	100 (1)	
Zou	**Western Gbe**	0	0	
	Central Gbe	100 (1)	100 (1)	
	Eastern Gbe	22.5 (25)	100 (111)	0.895
	Non-Gbe	20 (1)	100 (5)	

Collines	**Western Gbe**	0	0	
	Central Gbe	0	100 (1)	
	Eastern Gbe	27.3 (3)	100 (11)	0.338
	Non-Gbe	13.7 (3)	100 (22)	
Borgou	**Western Gbe**	0	0	
	Central Gbe	0	0	
	Eastern Gbe	33.3 (1)	100 (3)	0.225
	Non-Gbe	6.7 (2)	100 (30)	
Alibori	**Western Gbe**	0	0	
	Central Gbe	0	0	
	Eastern Gbe	0	0	
	Non-Gbe	7.1 (2)	100 (28)	
Atacora	**Western Gbe**	0	0	
	Central Gbe	0	0	
	Eastern Gbe	0	100 (2)	
	Non-Gbe	25 (8)	100 (32)	
Donga	**Western Gbe**	0	0	
	Central Gbe	0	100 (1)	
	Eastern Gbe	0	0	
	Non-Gbe	27.3 (6)	100 (22)	

Table S4. Repartition of lineages by usual residence department of patient

	Patient department of usual residence	Lineages							Total % (n)	P (All) P (Benin only)
		L1 % (n)	L2 % (n)	L3 % (n)	L4 % (n)	L5 % (n)	L6 % (n)	M. bovis % (n)		
Benin	Atlantique	1.1 (5)	**4.6 (20)**	0.9 (4)	49.7 (218)	**36.9 (162)**	6.4 (28)	0.5 (2)	100 (439)	
	Littoral	0.9 (2)	**6.8 (16)**	0.4 (1)	57.7 (135)	23.5 (55)	10.7 (25)	0	100 (234)	
	Oueme	2.6 (4)	2.6 (4)	0	68.2 (105)	24 (37)	2.6 (4)	0	100 (154)	
	Plateau	**4.3 (1)**	**4.3 (1)**	0	47.8 (11)	**43.5 (10)**	0	0	100 (23)	
	Mono	1.2 (1)	3.5 (3)	**3.5 (3)**	38.8 (33)	**45.9 (39)**	5.9 (5)	1.2 (1)	100 (85)	
	Couffo	0	2.3 (2)	1.2 (1)	52.3 (45)	**38.8 (33)**	4.7 (4)	1.2 (1)	100 (86)	
	Zou	0.8 (1)	4.1 (5)	0	65.9 (81)	22.8 (28)	5.7 (7)	0.8 (1)	100 (123)	
	Collines	0	0	0	72.2 (26)	19.4 (7)	8.3 (3)	0	100 (36)	
	Borgou	**5.6 (2)**	2.8 (1)	2.8 (1)	50 (18)	8.3 (3)	**27.8 (10)**	2.8 (1)	100 (36)	
	Alibori	**13.2 (5)**	2.6 (1)	2.6 (1)	44.7 (17)	5.3 (2)	**23.7 (9)**	**7.9 (3)**	100 (38)	**< 0.0001**
	Atacora	0	0	**7 (3)**	53.5 (23)	20.9 (9)	9.3 (4)	**9.3 (4)**	100 (43)	**< 0.0001**
	Donga	2.9 (1)	2.9 (1)	0	57.1 (20)	28.6 (10)	5.7 (2)	2.9 (1)	100 (35)	
	Total Benin	**1.7 (22)**	**4.1 (54)**	**1.1 (14)**	**55 (732)**	**29.7 (395)**	**7.6 (101)**	**1.1 (14)**	**100 (1332)**	
Other countries	Cameroon	0	0	0	100 (1)	0	0	0	100 (1)	
	Gabon	0	0	0	100 (2)	0	0	0	100 (2)	
	Ivory-Coast	0	0	0	100 (2)	0	0	0	100 (2)	
	Mali	0	0	0	100 (1)	0	0	0	100 (1)	
	Niger	0	0	0	85.3 (5)	0	0	**16.7 (1)**	100 (6)	
	Nigeria	0	1.3 (1)	0	61.8 (47)	26.3 (20)	9.2 (7)	1.3 (1)	100 (76)	
	Togo	0	**14.3 (1)**	0	57.1 (4)	14.3 (1)	**14.3 (1)**	0	100 (7)	
	Total	**1.5 (22)**	**3.9 (56)**	**1 (14)**	**55.6 (794)**	**29.2 (416)**	**7.6 (109)**	**1.1 (16)**	**100 (1427)**	

Table S5 Repartition of lineage by population density of patient residence district

Lineages	Population density of residence district (population/km^2)						Total % (n)	P
	11-49	**50-87**	**88-433**	**434-738**	**739-1642**	**1643-8595**		
L1	**5 (4)**	**3.3 (4)**	1.5 (5)	0	1.5 (5)	1.3 (4)	1.6 (22)	
L2	2.5 (2)	0.8 (1)	4.1 (14)	3.5 (5)	4.4 (15)	5.6 (17)	4.1 (54)	
L3	2.5 (2)	2.5 (3)	0.6 (2)	1.4 (2)	1.2 (4)	0.3 (1)	1.1 (14)	
L4	**57.5 (46)**	**57.9 (70)**	48.8 (166)	49.7 (71)	**57.7 (196)**	**59.2 (180)**	54.9 (729)	
L5	16.3 (13)	18.2 (22)	**38.5 (131)**	**37.8 (54)**	29.7 (101)	24.3 (74)	29.7 (395)	
L6	**13.8 (11)**	**12.4 (15)**	5.9 (20)	4.9 (7)	5.6 (19)	9.2 (28)	7.5 (100)	
M. bovis	2.5 (2)	5 (6)	0.6 (2)	2.8 (4)	0	0	1.1 (14)	
								< 0.0001
Total	100 (80)	100 (121)	100 (340)	100 (143)	100 (340)	100 (304)	100 (1328)	

Table S6. Repartition of lineage by population density of patient residence district (broader grouping of population density based on similarity on lineage prevalence in smaller grouping categories)

Lineages	Population density of patient residence district (km²)						
	11-87	**88-738**	**739-8595**	**Total % (n)**	**p (11-87 vs 88-738)**	**P (11-87 vs 739-8595)**	**P (88-738 vs 739-8595)**
L1	**4 (8)**	1 (5)	1.4 (9)	1.6 (22)	**0.010** **Diff: 2.9 (0.1 to 5.8)** **OR: 4 (1.3 to 11.7)**	**0.023** **Diff: 2.6 (-0.3 to 5.4)** **OR: 2.9 (1.1 to 7.4)**	0.587 Diff: -0.4 (-1.6 to 0.9) OR: 0.7 (0.3 to 2.1)
L2	1.5 (3)	3.9 (19)	5 (32)	4.1 (54)	0.099 Diff: -2.4 (-4.9 to -0.03) OR: 0.4 (0.1 to 1.2)	**0.031** **Diff: -3.5 (-5.8 to -1.1)** **OR: 0.3 (0.1 to 0.9)**	0.408 Diff: -1 (-3.4 to 1.4) OR: 0.8 (0.4 to 1.4)
L3	2.5 (5)	0.8 (4)	0.8 (5)	1.1 (14)	0.083 Diff: 1.7 (-0.6 to 4) OR: 3.1 (0.9 to 10.6)	0.050 Diff: 1.7 (-0.5 to 4) **OR: 3.3 (1 to 10.6)**	0.923 Diff: 0.1 (-1 to 1.1) OR: 1.1 (0.3 to 3.7)
L4	**57.7 (116)**	49.1 (237)	**58.4 (376)**	54.9 (729)	**0.039** **Diff: 8.6 (0.5 to 16.8)** **OR: 1.4 (1 to 2)**	0.866 -0.7 (-8.4 to 7.1) OR: 1 (0.7 to 1.3)	**0.002** **Diff: -9.3 (-15.2 to -3.5)** **OR: 0.7 (0.5 to 0.9)**
L5	17.4 (35)	**38.3 (185)**	27.2 (175)	29.7 (395)	**< 0.00001** **Diff: -20.9 (-27.7 to -14.1)** **OR: 0.3 (0.2 to 0.5)**	**0.005** **Diff: -9.8 (-16 to -3.5)** **OR: 0.6 (0.4 to 0.8)**	**< 0.0001** **Diff: 11.1 (5.6 to 16.7)** **OR: 1.7 (1.3 to 2.1)**
L6	**12.9 (26)**	5.6 (27)	7.3 (47)	7.5 (100)	**0.001** **Diff: 7.3 (2.3 to 12.4)** **OR: 2.5 (1.4 to 4.4)**	**0.013** **Diff: 5.6 (0.6 to 10.7)** **OR: 1.9 (1.1 to 3.1)**	0.252 Diff: -1.7 (-4.6 to 1.2) OR: 0.8 (0.5 to 1.2)
M. bovis	4 (8)	1.2 (6)	0	1.1 (14)	**0.021** Diff: 2.7 (-1.3 to 5.6) **OR: 3.3 (1.2 to 9.2)**	**<0.00001** **Diff: 4 (1.3 to 6.7)** **OR: -**	**0.005** **Diff: 1.2 (0.3 to 2.2)** **OR: -**
Total	100 (201)	100 (483)	100 (644)	100 (1328)			

Table S7. Distribution of sputum AFB microscopy by lineage among previously-treated patients: stratification by type of previously-treated patient

Lineages	Sputum AFB microscopy % (n)					Total	Scanty Lx vs L4	Scanty&Neg Lx vs L4
	+++	++	+	scanty	neg		p Diff (95% CI) OR (95% CI)	p Diff (95% CI) OR (95% CI)
P-TREATED types								
FAILURE	17 (n=9)	18.9 (n=10)	24.5 (n=13)	24.5 (n=13)	15.1 (n=8)	100 (n=53)		
L1	0	0	100 (1)	0	0	100 (1)		
L2	0	0	0	0	100 (1)	100 (1)		
L3	0	0	0	0	0	0		
L4	24.2 (8)	18.2 (6)	27.3 (9)	12.1 (4)	18.2 (6)	100 (33)	-	-
L5	6.3 (1)	18.8 (3)	18.8 (3)	**50 (8)**	6.3 (1)	100 (16)	**0.004** **37.9 (11 to 64.8)** **7.3 (1.8 to 29)**	0.08 25.9 (-3 to 54.9) 3 (0.9 to 9.9)
L6	0	50 (1)	0	50 (1)	0	100 (2)		
M. bovis	0	0	0	0	0	0		
RELAPSE	41.5 (n=63)	34.9 (n=53)	18.4 (n=28)	5.3 (n=8)	0	100 (n=152)		
L1	0	0	100 (1)	0	0	100 (1)		
L2	37.5 (3)	25 (2)	25 (2)	12.5 (1)	0	100 (8)		
L3	50 (1)	0	50 (1)	0	0	100 (2)		
L4	41.4 (43)	36.5 (38)	17.3 (18)	4.8 (5)	0	100 (104)	-	
L5	44.8 (13)	34.5 (10)	13.8 (4)	6.9 (2)	0	100 (29)	0.656 2.1 (-8 to 12.2) 1.5 (0 to 7)	
L6	37.5 (3)	37.5 (3)	25 (2)	0	0	100 (8)		
M. bovis	0	0	0	0	0	0		

DEFAULT	42.9 (n=6)	28.6 (n=4)	21.4 (n=3)	7.1 (n=1)	0	100 (n=14)	
L1	0	0	0	0	0	0	
L2	0	0	0	0	0	0	
L3	0	0	0	0	0	0	
L4	28.6 (2)	42.9 (3)	28.6 (2)	0	0	100 (7)	-
L5	66.7 (4)	0	16.7 (1)	16.7 (1)	0	100 (6)	0.261
							16.7 (-13.2 to 465)
							-
L6	0	100 (1)	0	0	0	100 (1)	
M. bovis	0	0	0	0	0	0	

Table S8. Culture result by lineage (without repeat)

Lineages	LJ Culture (up to 90 days)				
	Positive % (n)	Total % (n)	P	Diff (95% CI)	OR (95% CI)
ALL PATIENTS					
L1	80 (16)	100 (20)			
L2	67.4 (29)	100 (43)			
L3	75 (9)	100 (12)			
L4	**70.4** (456)	100 (648)	-	-	-
L5	**53.4** (172)	100 (322)	**< 0.00001**	**-17 (-23.4 to -10.5)**	**0.5 (0.4 to 0.6)**
L6	**43.2** (38)	100 (88)	**< 0.00001**	**-27.2 (-38.1 to -16.3)**	**0.3 (0.2 to 0.5)**
M. bovis	**33.3** (4)	100 (12)	**0.006**	**-37 (-63.9 to -10.1)**	**0.2 (0.1 to 0.7)**
			< 0.0001		
Total	63.2 (724)	100 (1145)			
NEW					
L1	88.9 (16)	100 (18)			
L2	67.7 (21)	100 (31)			
L3	66.7 (6)	100 (9)			
L4	**74** (341)	100 (461)	-	-	-
L5	**57.4** (151)	100 (263)	**< 0.00001**	**-16.6 (-23.7 to -9.4)**	**0.5 (0.3 to 0.6)**
L6	**48** (35)	100 (73)	**< 0.00001**	**-26 (-38 to -13.9)**	**0.3 (0.2 to 0.5)**
M. bovis	**36.4** (4)	100 (11)	**0.005**	**-37.6 (-66.3 to -8.9)**	**0.2 (0.1 to 0.7)**
			< 0.0001		
Total	66.3 (574)	100 (866)			
P-TREATED					
L1	0	100 (2)			

L2	66.7 (8)	100 (12)			
L3	100 (3)	100 (3)			
L4	**61.5** (115)	100 (187)	-	-	-
L5	**35.6** (21)	100 (59)	**0.0005**	**-25.9 (-40 to -11.8)**	**0.3 (0.2 to 0.6)**
L6	**30** (3)	100 (15)	**0.002**	**-41.5 (-62.9 to -20.1)**	**0.2 (0.05 to 0.5)**
M. bovis	0	100 (1)			
			< 0.0001		
Total	**53.8** (150)	100 (279)			

Table S9. Repartition of lineage by TB clinic (new and previously-treated patients combined)

Regions	TB clinics	Lineages							Total	p
		L1 n(%)	L2	L3	L4	L5	L6	Bovis		
Atlantique/Littoral	Allada	1 (0.8)	5 (3.7)	0	67 (50)	**52 (38.8)**	8 (6)	1 (0.8)	134 (100)	<0.0001
	Ab-Calavi	2 (2)	3 (2.9)	2 (2)	56 (54.9)	35 (34.3)	4 (3.9)	0 (0)	102 (100)	
	CNHU-PPC (Cotonou)	6 (1.4)	**30 (6.8)**	3 (0.7)	251 (56.7)	115 (26)	38 (8.6)	0	443 (100)	
	Ouidah	0	1 (1.7)	0	22 (37.9)	**30 (51.7)**	4 (6.9)	1 (1.7)	58 (100)	
	total	**9 (1.2)**	**39 (5.3)**	**5 (0.7)**	**396 (53.7)**	**232 (31.5)**	**54 (7.3)**	**2 (0.3)**	**737 (100)**	
Oueme/Plateau	Akron (Porto-Novo)	4 (2.4)	1 (0.6)	0	107 (63.4)	50 (29.8)	5 (3)	1 (0.6)	168 (100)	
	Avrankou	0	0	0	**9 (75)**	3 (25)	0	0	12 (100)	
	Pobe	0	1 (20)	0	3 (60)	1 (20)	0	0	5 (100)	
	Sakete	0	0	0	1 (25)	**3 (75)**	0	0	4 (100)	
	total	**4 (2.1)**	**2 (1.1)**	**0**	**120 (63.5)**	**57 (30.2)**	**5 (2.6)**	**1 (0.5)**	**189 (100)**	
Mono/Couffo	Aplahoue	0	1 (1.2)	0	47 (56.6)	30 (36.1)	3 (3.6)	2 (2.4)	83 (100)	
	Bopa	0	1 (2.6)	0	19 (50)	**16 (42)**	2 (5.3)	0	38 (100)	
	Come	0	**3 (7.1)**	2 (4.8)	14 (33.3)	**19 (45.2)**	4 (9.5)	0	42 (100)	
	Houeyogbe	0	0	**1 (20)**	1 (20)	**3 (60)**	0	0	5 (100)	
	total	**0**	**8 (4.8)**	**3 (1.8)**	**81 (48.2)**	**68 (40.5)**	**9 (5.4)**	**2 (1.2)**	**168 (100)**	
Zou/Collines	Abomey	1 (1.1)	5 (5.5)	1 (1.1)	57 (62.6)	22 (24.2)	4 (4.4)	1 (1.1)	91 (100)	
	Bohicon	0	1 (2.9)	0	**25 (71.4)**	7 (20)	2 (5.7)	0	35 (100)	
	Dassa	0	0	0	**25 (71.4)**	7 (20)	3 (8.6)	0	35 (100)	
	Zagnanado	0	**1 (8.3)**	0	8 (66.7)	1 (8.3)	2 (16.7)	0	12 (100)	
	total	**1 (0.6)**	**7 (4)**	**1 (0.6)**	**115 (66.5)**	**37 (21.4)**	**11 (6.4)**	**1 (0.6)**	**173 (100)**	
Borgou/Alibori	Bembereke	**4 (6.7)**	2 (3.3)	2 (3.3)	31 (51.7)	3 (5)	**14 (23.3)**	**4 (6.7)**	60 (100)	
	Kandi	**2 (11.1)**	**1 (5.6)**	0	8 (44.4)	2 (11.1)	**5 (27.8)**	0	18 (100)	
	Nikki-hzss	**1 (10)**	0	0	5 (50)	1 (10)	**3 (30)**	0	10 (100)	
	Parakou	0	0	**1 (20)**	2 (40)	0	**1 (20)**	**1 (20)**	5 (100)	
	total	**7 (7.5)**	**3 (3.2)**	**3 (3.2)**	**46 (49.5)**	**6 (6.5)**	**23 (24.7)**	**5 (5.4)**	**93 (100)**	

Atacora/Donga	Djougou	1 (3.1)	0	0	18 (56.3)	10 (31.3)	2 (6.3)	1 (3.1)	32 (100)
	Materi	0	0	**2 (13.3)**	8 (53.3)	4 (26.7)	1 (6.8)	0	15 (100)
	Natitingou	0	1 (4)	1 (4)	12 (48)	4 (16)	3 (12)	**4 (16)**	25 (100)
	Tanguieta	0	0	0	4 (80)	0	**1 (20)**	0	5 (100)
	total	**1 (1.3)**	**1 (1.3)**	**3 (3.9)**	**42 (54.5)**	**18 (23.4)**	**7 (9.1)**	**5 (6.5)**	**77 (100)**
Total		**22 (1.5)**	**57 (4)**	**15 (1)**	**800 (55.7)**	**418 (29.1)**	**109 (7.6)**	**16 (1.1)**	**1437 (100)**

Table S10. Repartition of lineages by TB clinics for new TB patients

New TB patients											
Regions	**TB clinics**	**Lineages n(%)**							**Total**	**p**	
		L1	**L2**	**L3**	**L4**	**L5**	**L6**	**bovis**			**MAF n (%)**
Atlantique/Littoral	Allada	1 (0.9)	4 (3.5)	0	54 (47.8)	46 (40.7)	7 (6.2)	1 (0.9)	113 (100)		53 (46.9)
	Ab-Calavi	2 (2.5)	3 (3.7)	2 (2.5)	40 (49.4)	31 (38.3)	3 (3.7)	0	81 (100)		34 (42)
	CNHU-PPC (Cotonou)	5 (1.4)	22 (6.2)	2 (0.6)	198 (55.5)	98 (27.5)	32 (9)	0	357 (100)		130 (36.5)
	Ouidah	0	0	0	18 (37.5)	25 (52.1)	4 (8.3)	1 (2.1)	48 (100)		29 (60.4)
	Total Atl/Lit	**8 (1.3)**	**29 (4.8)**	**4 (0.7)**	**310 (51.8)**	**200 (33.4)**	**46 (7.7)**	**2 (0.3)**	**599 (100)**		**246 (41.1)**
Oueme/Plateau	Akron (Porto-Novo)	4 (2.9)	1 (0.7)	0	87 (63)	40 (29)	5 (3.6)	1 (0.7)	138 (100)		45 (32.6)
	Avrankou	0	0	0	6 (66.7)	3 (33.3)	0	0	9 (100)		3 (33.3))
	Pobe	0	1 (25)	0	2 (50)	1 (25)	0	0	4 (100)		1 (25)
	Sakete	0	0	0	1 (25)	**3 (75)**	0	0	4 (100)		**3 (75)**
	Total Oue/Pla	**4 (2.6)**	**2 (1.3)**	**0**	**96 (61.9)**	**47 (30.3)**	**5 (3.2)**	**1 (0.6)**	**155 (100)**		**52 (33.5)**
Mono/Couffo	Aplahoue	0	1 (1.5)	0	36 (54.6)	24 (36.4)	3 (4.6)	2 (3)	66 (100)		27 (41)
	Bopa	0	1 (3.3)	0	14 (46.7)	13 (43.3)	2 (6.7)	0	30 (100)	**<0.0001**	15 (50)
	Come	0	3 (8.8)	2 (5.9)	10 (29.4)	17 (50)	2 (5.9)	0	34 (100)		19 (55.9)
	Houeyogbe	0	0	0	1 (25)	**3 (75)**	0	0	4 (100)		**3 (75)**
	Total Mono/Couf	**0**	**5 (3.7)**	**2 (1.5)**	**61 (45.5)**	**57 (42.5)**	**7 (5.2)**	**2 (1.5)**	**134 (100)**		**64 (47.7)**
Zou/Collines	Abomey	1 (1.4)	4 (5.6)	1 (1.4)	43 (59.7)	19 (26.4)	4 (5.6)	0	72 (100)		23 (32)
	Bohicon	0	0	0	16 (64)	7 (28)	2 (8)	0	25 (100)		9 (36)
	Dassa	0	0	0	**19 (70.4)**	7 (25.9)	1 (3.7)	0	27 (100)		8 (29.6)
	Zagnanado	0	**1 (11.1)**	0	6 (66.7)	1 (11.1)	1 (11.1)	0	9 (100)		2 (22.2)
	Total Zou/Coll	**1 (0.8)**	**5 (3.8)**	**1 (0.8)**	**84 (63.2)**	**34 (25.6)**	**8 (6)**	**0**	**133 (100)**		**42 (31.6)**
Borgou/Alibori	Bembereke	4 (7.8)	2 (3.9)	1 (2)	24 (47.1)	3 (5.9)	13 (25.5)	4 (7.8)	51 (100)		16 (31.4)
	Kandi	1 (7.1)	1 (7.1)	0	5 (35.7)	2 (14.3)	5 (35.7)	0	14 (100)		7 (50)
	Nikki-hzss	**1 (12.5)**	0	0	3 (37.5)	1 (12.5)	**3 (37.5)**	0	8 (100)		4 (50)
	Parakou	0	0	**1 (25)**	1 (25)	0	1 (25)	**1 (25)**	4 (100)		1 (25)
	Total Borg/Alib	**6 (7.8)**	**3 (3.9)**	**2 (2.6)**	**33 (4.3)**	**6 (7.8)**	**22 (28.6)**	**5 (6.9)**	**77 (100)**		**28 (36.4)**
	Djougou	1 (4.2)	0	0	11 (45.8)	9 (37.5)	2 (8.3)	1 (4.2)	24 (100)		11 (45.8)

	Materi	0	0	2 (16.7)	6 (50)	3 (25)	1 (8.3)	0	12 (100)		4 (33.3)
Atacora/Donga	Natitingou	0	0	1 (5.3)	9 (47.4)	3 (15.8)	2 (10.5)	**4 (21.1)**	19 (100)		5 (26.3)
	Tanguieta	0	0	0	2 (66.7)	0	1 (33.3)	0	3 (100)		1 (33.3)
	Total Atac/Donga	**1 (1.7)**	**0**	**3 (5.2)**	**28 (48.3)**	**15 (25.9)**	**6 (10.3)**	**5 (8.6)**	**58 (100)**		**21 (36.2)**
Total	Total	**20 (1.7)**	**44 (3.8)**	**12 (1)**	**612 (52.9)**	**359 (31.1)**	**94 (8.1)**	**15 (1.3)**	**1156 (100)**		**453 (39.2)**

Table S11. Repartition of lineages by TB clinics for previously-treated TB patients

Previously-treated TB patients											
Regions	***TB clinics***	***Lineages n(%)***							***Total***	***p***	
		L1	***L2***	***L3***	***L4***	***L5***	***L6***	***bovis***			***MAF n (%)***
Atlantique/Littoral	*Allada*	*0*	*1 (4.8)*	*0*	*13 (61.9)*	*6 (28.6)*	*1 (4.8)*	*0*	*21 (100)*		*7 (33.4)*
	Ab-Calavi	*0*	*0*	*0*	*16 (76.2)*	*4 (19.1)*	*1 (4.8)*	*0*	*21 (100)*		*5 (20.1)*
	CNHU-PPC (Cotonou)	*1 (1.2)*	*8 (9.3)*	*1 (1.2)*	*53 (61.6)*	*17 (19.8)*	*6 (7)*	*0*	*86 (100)*		*23 (26.8)*
	Ouidah	*0*	*1 (10)*	*0*	*4 (40)*	*5 (50)*	*0*	*0*	*10 (100)*		***5 (50)***
	Total Atl/Lit	***1 (0.7)***	***10 (7.2)***	***1 (0.7)***	***86 (62.3)***	***32 (23.2)***	***8 (5.8)***	***0***	***138 (100)***		***40 (29)***
Oueme/Plateau	*Akron (Porto-Novo)*	*0*	*0*	*0*	*20 (66.7)*	*10 (33.3)*	*0*	*0*	*30 (100)*		*10 (33.3)*
	Avrankou	*0*	*0*	*0*	***3 (100)***	*0*	*0*	*0*	*3 (100)*		*0*
	Pobe	*0*	*0*	*0*	***1 (100)***	*0*	*0*	*0*	*1 (100)*		*0*
	Sakete	*0*	*0*	*0*	*0*	***0***	*0*	*0*	*0*		***0***
	Total Oue/Pla	***0***	***0***	***0***	***24 (70.6)***	***10 (29.4)***	***0***	***0***	***34 (100)***		***10 (29.4)***
Mono/Couffo	*Aplahoue*	*0*	*0*	*0*	*11 (64.7)*	*6 (35.3)*	*0*	*0*	*17 (100)*		*6 (35.3)*
	Bopa	*0*	*0*	*0*	*5 (62.5)*	***3 (37.5)***	*0*	*0*	*8 (100)*		*3 (37.5)*
	Come	*0*	*0*	*0*	*4 (50)*	*2 (25)*	*2 (25)*	*0*	*8 (100)*	***<0.0001***	***4 (50)***
	Houeyogbe	*0*	*0*	*1 (100)*	*0*	***0***	*0*	*0*	*1 (100)*		***0***
	Total Mono/Couf	***0***	***0***	***1 (2.9)***	***20 (58.8)***	***11 (32.4)***	***2 (5.9)***	***0***	***34 (100)***		***13 (38.3)***
Zou/Collines	*Abomey*	*0*	*1 (5.3)*	*0*	*14 (73.7)*	*3 (15.8)*	*0*	***1 (5.3)***	*19 (100)*		*3 (15.8)*
	Bohicon	*0*	*1 (10)*	*0*	*9 (90)*	*0*	*0*	*0*	*10 (100)*		*0*
	Dassa	*0*	*0*	*0*	***6 (75)***	*0*	*2 (25)*	*0*	*8 (100)*		*2 (25)*
	Zagnanado	*0*	***0***	*0*	*2 (66.7)*	*0*	***1 (33.3)***	*0*	*3 (100)*		*1 (33.3)*
	Total Zou/Coll	***0***	***2 (5)***	***0***	***31 (77.5)***	***3 (7.5)***	***3 (7.5)***	***1 (2.5)***	***40 (100)***		***6 (15)***
Borgou/Alibori	*Bembereke*	*0*	*0*	***1 (11.1)***	*7 (77.8)*	*0*	*1 (11.1)*	*0*	*9 (100)*		*1 (11.1)*
	Kandi	***1 (25)***	*0*	*0*	*3 (75)*	*0*	*0*	*0*	*4 (100)*		*0*
	Nikki-hzss	***0***	*0*	*0*	***2 (100)***	*0*	*0*	*0*	*2 (100)*		*0*
	Parakou	*0*	*0*	***0***	***1 (100)***	*0*	*0*	***0***	*1 (100)*		*0*

	Total Borg/Alib	***1 (6.3)***	***0***	***1 (6.3)***	***13 (81.3)***	***0***	***1 (6.3)***	***0***	***16 (100)***	***1 (6.3)***
	Djougou	*0*	*0*	*0*	*7 (87.5)*	*1 (12.5)*	*0*	*0*	*8 (100)*	*1 (12.5)*
	Materi	*0*	*0*	*0*	*2 (66.7)*	*1 (33.3)*	*0*	*0*	*3 (100)*	*1 (33.3)*
Atacora/Donga	*Natitingou*	*0*	***1 (16.7)***	*0*	*3 (50)*	*1 (16.7)*	*1 (16.7)*	***0***	*6 (100)*	*2 (33.4)*
	Tanguieta	*0*	*0*	*0*	***2 (100)***	*0*	*0*	*0*	*2 (100)*	*0*
	Total Atac/Donga	***0***	***1 (5.3)***	***0***	***14 (73.7)***	***3 (15.8)***	***1 (5.3)***	***0***	***19 (100)***	***4 (21.1)***
Total	*Total*	***2 (0.7)***	***13 (4.6)***	***3 (1.1)***	***188 (66.9)***	***59 (21)***	***15 (5.3)***	***1 (0.4)***	***281 (100)***	***74 (26.3)***

Table S12. Temporal variation of lineage prevalence in new TB patient in Cotonou, Benin

New TB patients in Cotonou							
Lineages	**This study direct-spol (2016-2018) % (n=357)**	**Direct spol evaluation study (2016)**		**Previous study (2005-2006, indirect spol) % (n=194)**	**Previous study (2005-2006) Estimated prevalence if using direct spol % (n=194)**	**Comparison current prevalence versus prevalence in 2005-2006 P-value Difference(95%CI)**	
		Direct spol % (n=196)	**Indirect spol % (n=143)**			**Current sputum-based vs previous culture-based**	**Current sputum-based vs previous estimated sputum-based**
L1	1.4	8.2	9.1	7.7	6.9	**0.0002 (0.0001** for Ha: D<0**)** **-6.3 (-10.2 to -2.4)**	**0.0006 (0.0003** for Ha: D<0**)** **-5.5 (-9.3 to -1.7)**
L2	6.2	5.6	6.3	10.3	9.2	0.083 -4.1 (-9.1 to 0.9)	0.195 -3 (-7.8 to 1.8)
L3	0.6	1	1.4	0	0	0.28 0.6 (-0.2 to 1.4)	0.28 0.6 (-0.2 to 1.4)
L4	**55.5**	51.5	55.2	**42.3**	**39.5**	**0.003** **13.2 (4.5 to 21.9)**	**0.0003 (0.0002** for Ha: D>0**)** **16 (7.4 to 24.6)**
L5	**27.5**	25	21	**31**	**36.9**	0.386 -3.5 (-11.5 to 4.5)	**0.023 (0.011** for Ha: D<0**)** **-9.4 (-17.6 to -1.2)**
L6	9	8.7	7	6	7.5	0.214 3 (-1.5 to 7.5)	0.546 1.5 (-3.2 to 6.2)

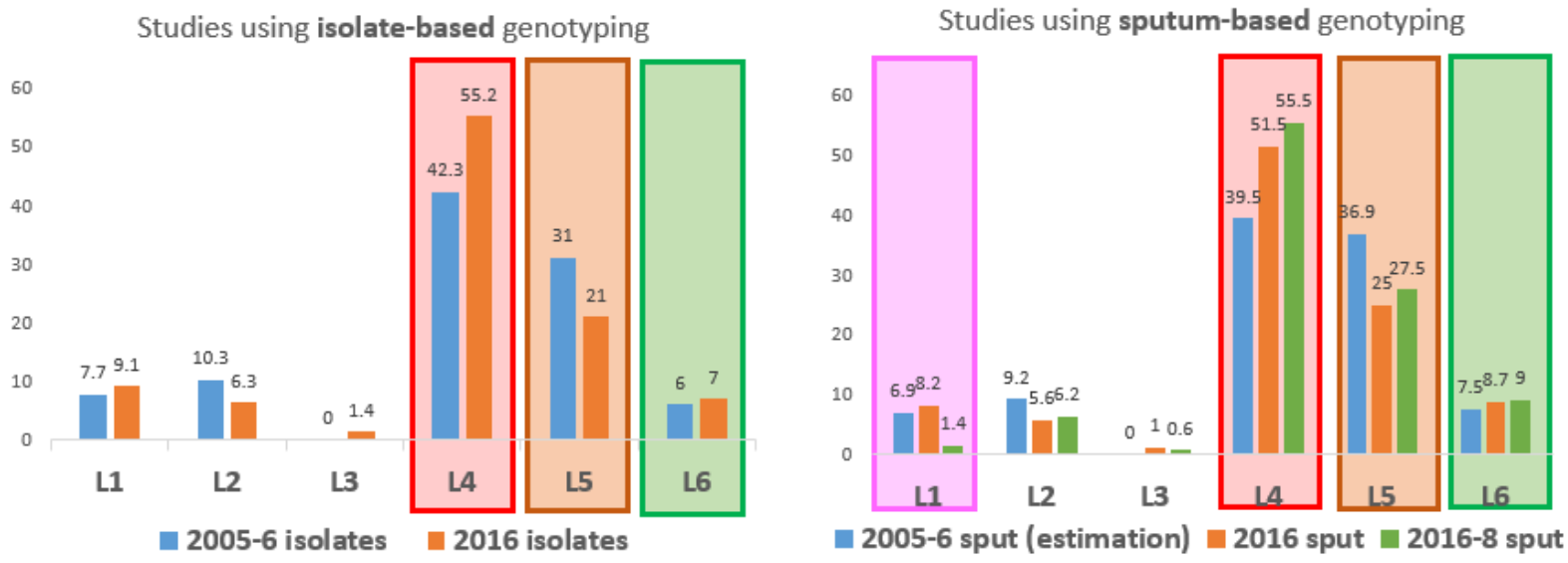

Figure S1. Temporal variation of lineage prevalence in new TB patient in Cotonou, Benin

sput: sputum

SUPPLEMENT S1

Optimization of L125, L346 PCR for PhyloSNP (SNP-based lineage detection)

BOX 1. PCR reactional mix for isolates

For **PCR L125** on **ISOLATES**:	For **PCR L346** on **ISOLATES**:
• **25 pmol/50 µL** (reaction) PCR mix • 40 µL PCR mix per reaction (+**10 µL** sample DNA) • **No Q solution** • **Annealing T°=67.4°C** • 45 cycles	• **25 pmol/50 µL** (reaction) PCR mix • 40 µL PCR mix per reaction (+**10 µL** sample DNA) • **with Q solution** • **Annealing T°=67.4°C** • 45 cycles

BOX 2. PCR reactional mix for sputum

For **PCR L125** on **SPUTUM**:	For **PCR L346** on **SPUTUM**:
• **25 pmol/50 µL** (reaction) PCR mix • 40 µL PCR mix per reaction (+**10 µL** sample DNA) • **with Q solution** • **Annealing T°=64.9°C** • 60 cycles	• **25 pmol/50 µL** (reaction) PCR mix • 40 µL PCR mix per reaction (+**10 µL** sample DNA) • **No Q solution** • **Annealing T°=69.3°C** • 60 cycles

Amplification program

Initial denaturation	95°C	3 min
denaturation	95°C	30 s
annealing	Choose in boxes 1 or 2 (above) depending on the type of sample	30 s
extension	72°C	30 s
Final extension	72 °C	10 min

Table. Primer used for Sanger sequencing amplification and lineage specific SNP and position (Source: Carcino et al)

MTBC lineage	SNP position[1]	Nucleotide change	Gene	Primers 5'-3'	Amplicon size
LINEAGES 1,2,5	L1 4357773	G/A	Rv3878/Rv3879c	F-ACCCTCAACAACCACAACGT R-CGACACTACCGATCAGCGTT	386bp
	L2 4357804	T/G			
	L5 4357657	G/A			
LINEAGES 3,4,6	L3 1281984	G/A	Rv1155/intergenic region	F-GATGGTCATACGCCGTTGCT R-CTCTTGCGGGGACTTCGATT	402bp
	L4 1281771	C/T			
	L6 1281685	C/G			

Chapter 8

The genomic diversity of *Mycobacterium tuberculosis* lineage 5 (*Mycobacterium africanum* West-African 1)

C. N'Dira Sanoussi[1,2,3], Mireia Coscolla[4], Boatema Ofori-Anyinam[5], Patrick Beckert[6], Isaac Otchere[7], Pim de Rijk[2], Martin Antonio[5], Stefan Niemann[6], Julian Parkhill[8], Simon Harris[8], Dorothy Yeboah-Manu[7], Sébastien Gagneux[9], Leen Rigouts[2,3], Dissou Affolabi[1], Bouke C. de Jong[2], Conor J. Meehan[2,10]

[1] Laboratoire de Référence des Mycobactéries, Cotonou, Benin
[2] Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium
[3] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[4] University of Valencia, Spain
[5] Center for Global Health Security and Diplomacy, Ottawa, Canada,
[6] Research Center Borstel, Molecular and Experimental Mycobacteriology, Borstel, Germany
[7] Noguchi Memorial Institute for Medical Research, Legon, Accra, Ghana
[8] Wellcome Trust Sanger Institute, United Kingdom
[9] Swiss Tropical and Public Health Institute, Basel, Switzerland
[10] School of Chemistry and Biosciences, University of Bradford, Bradford, UK

In preparation

Conceived and designed the research: **CNS**, MC, SN, JP, SH, DYM, SG, LR, DA, BCdJ, CJM

Performed the research: **CNS**, MC, CJM

Data analysis: **CNS**, MC, BOA, LR, BCdJ, CJM

Supervision: MC, LR, BCdJ, CJM

Writing original draft: **CNS**

Writing, review and editing: **CNS**, MC, BOA, PB, IO, PdR, MA, SN, JP, SH, DYM, SG, LR, DA, BCdJ, CJM

Contributed reagents, materials, analysis tools: MC, PB, IO, SN, JP, SH, DYM, SG, LR, DA, BCdJ, CJM

Abstract

Several phenotypic and genotypic factors separate *Mycobacterium africanum* (including lineage 5 (L5)) from the other human-adapted lineages of the *M. tuberculosis* complex. For whole genome sequencing, standard practice is to use H37Rv (an L4 strain) as the reference genome for strain mapping from all lineages. We assessed whether the H37Rv reference genome is biologically representative of L5 (gene content), and whether it is suitable as a reference genome for L5.

Gene content differences were assessed between complete PacBio genomes from three L5 strains, from Benin (current study), Gambia, and Nigeria (Phelan J. et al, 2018), against H37Rv.

In addition, 205 L5 genomes (Illumina reads) from various countries (de novo sequences and public data) were mapped against the PacBio H37Rv and L5 genomes to identify additional genetic diversity using the MTBseq pipeline.

Ten genes were present in all three Pacbio L5 genomes but absent in H37Rv. Those ten genes included two PE/PPE family genes and two insertion sequences. Four of those ten (4/10) L5-specific genes were common in all 205 Illumina L5 genomes, while 6 were found in 81-98% (165-201) of them. Importantly for L5 epidemiology, in the four L5 specific genes common to all L5, single nucleotide polymorphisms were detected in some Illumina L5 strains (3-7/205; 1.5-3.4%). The functional group of the four genes present in all L5 genomes included one PE/PPE, two hypothetical proteins (possibly IS*256* and CAAX conserved domain) and a gene of unknown function (*Mb2048c*). The IS*256* and *Mb2048c* were also present in the complete genome (reference) of L6 and *M. bovis*.

On the other hand, nine genes were present in H37Rv but absent in the complete PacBio L5 genomes. No PE/PPE family gene and insertion sequence was included in those 9 genes. Five of them (5/9) were absent in all L5 genomes, while three were found in 0.5-1.5% (1-3) of the Illumina genomes and one in 99.5% (204). The five genes absent in all L5 genomes included *Rv1977* (conserved hypotheticals, implicated in bacterial survival), *Rv1979c* (cell wall and cell prosesses, associated with clofazimine and bedaquiline resistance), *Rv1993* (conserved hypotheticals, associated with virulence), *Rv1995* (conserved hypotheticals, implicated in oxygen transport), *Rv2073c* (intermediary metabolism and respiration, probable short-chain deshydrogenase). The latter was also absent in the complete genome of L6 and *M. bovis*. Similarly, *Rv2074* found in only one Illumina L5 genome, was also absent in the complete genome of L6 and *M. bovis*. That gene (*Rv2074,* pyridoxamine-5-phosphate oxydase) is implicated in the biosynthesis of pyridoxine (vitamin B6) which is essential for bacterial survival, virulence and also involved in immune-evasive mechanisms to allow bacterial persistence. Interestingly, not all PacBio L5 genomes gave a similar gene-count, with the Nigerian strain missing two regions containing 30 genes, also missed in 6 Illumina L5 strains from Benin and Ghana that belong to a sublineage L5.3.2.

Overall, L5 gene content was different from H37Rv, with several L5 genes absent in H37Rv and SNPs present in these genes in some L5 strains. Therefore, using an L5 as reference-genome for L5 genome analyses and epidemiology is preferable. However, high within-lineage gene content variability suggests that the pan-genome of *M. tuberculosis* may be larger than previously thought, suggesting a reference-free approach may be needed. The reduced growth of L5 strains in culture could partly be explained by the absence- in the vast majority of L5 strains- of genes implicated in bacterial survival and in vitro

growth, and similarly the suggested lower virulence (immunogenicity, transmissibility) of L5 could be explained by the absence of genes associated with increased virulence. Our genomic finding of possible clofazimine (and bedaquiline) resistance of L5 strains should be tested in vitro.

Keywords: gene content, *M. africanum*, Lineage 5, H37Rv, reference genome, genome analyses, within lineage gene content variability

Introduction

Tuberculosis (TB) is caused by bacteria of the *Mycobacterium tuberculosis* complex (MTBC) and consists of 7 human-adapted lineages and several animal-adapted lineages.

The human-adapted lineages are subdivided into ancestral lineages (L1, L5 and L6), an intermediate lineage (L7) and modern lineages (L2, L3, L4)(Coscolla and Gagneux 2014). L5 and L6 are also called *M. africanum* (*M. africanum* west-african 1 and 2 respectively) and are restricted to West Africa where they cause up to 40% of human TB (de Jong, Antonio, and Gagneux 2010). L7 is geographically restricted to Ethiopia, while the modern lineages are more widespread (Gagneux 2017). Reasons for the geographical restriction of *M. africanum* remain unclear.

Several phenotypic and genotypic factors separate L5 and L6 from the other human-adapted lineages. Lineage 6 and the closely related animal strains/lineages such as *M. bovis* are reported to be less virulent than modern MTB lineages in population-based studies (de Jong et al. 2008; Magnus 1966) and in genome studies (Gonzalo-Asensio et al. 2014), primarily owing to a mutation in the PhoPR regulon, resulting in decreased ESAT-6 secretion, which

possibly explains the decreased virulence of those strains (Gonzalo-Asensio et al. 2014). Infection with L6 slowly progresses to TB disease, and is associated with impaired immunity in some settings but not all (e.g. HIV infection (de Jong et al. 2008)). In contrast, little is known about the virulence and progression of L5.

Some TB diagnostics have a lower performance for L5 and L6 strains, compared to modern lineages (Sanoussi et al. 2018; Ofori-Anyinam et al. 2016) and these lineages are less likely to grow in culture (Sanoussi et al. 2017), with isolates of dysgonic nature on solid medium (Leao et al. 2004; Pattyn et al. 1970; Sanoussi et al. 2018). Mutations in genes essential for growth in culture partly explain that difficult growth of L6 in culture (Gehre et al. 2013), yet for L5 the reasons for the difficult growth are unclear.

Whole genome sequencing of the MTBC often involves mapping short sequence reads to a reference genome, usually H37Rv (Cole et al. 1998; Camus et al. 2019; Meehan et al. 2019). However, since H37Rv is a laboratory-adapted L4 strain, it might not be representative of other MTBC lineages for genome analyses, including for L5. The members of the MTBC evolved from an environmental organism to an obligate pathogen through genome reduction and acquisition of new genes (Gagneux 2018) and it is known that some differences in gene content exist between lineages (Meehan et al. 2019; Bifani et al. 2000; Periwal et al. 2015; Kato-Maeda et al. 2001; O'Toole and Gautam 2017). However, little is known about the gene differences between L4 and L5 and the potential limitations this may impose for in-depth analysis of L5 genomes studies (e.g. sub-lineage detection and transmission tracking). Therefore, we assessed whether the H37Rv reference genome is biologically representative of L5 (gene content), and whether it is suitable as a reference

genome for L5. To this end, we analyzed three complete PacBio genomes of L5 and compared them to H37Rv and closely related lineages (L6, *M. bovis*), in order to find the particularities of L5. These gene content differences were then checked in a multi-country collection of short-read (Illumina) L5 genomes for confirmation. We then evaluated whether *H37Rv* serves well as reference genome for L5 genome epidemiological studies. The potential functions of the L5-specific and -missing genes were also identified.

Materials and Methods

Genomes

Complete genomes (PacBio SMRT). Three complete L5 genomes, all sequenced with the PacBio SMRT technology, were analyzed. One genome was from a Benin isolate (PcbL5Ben; sequenced in this study), one from an isolate from The Gambia (PcbL5Gbia, WBB1453_11-00429-1) (Phelan et al. 2018), and one from Nigeria (PcbL5Nig, WBB1454_IB091-1) (Phelan et al. 2018). The previously published reference/complete genomes H37Rv (L4) (NC_000962.3)(NCBI 2017), L6 (GM041182, Genbank accession n°: FR878060.1, GCF_0001593225.1_ASM159322v1) (Bentley et al. 2012) and *M. bovis* (AF2122/97, accession: LT708304.1) (Malone et al. 2017) were also included.

Whole genomic DNA extraction. Genomic DNA extraction was performed on growth from fresh Löwenstein-Jensen slants using the semi-automated Maxwell 16 Cell DNA purification kit in the Maxwell 16 machine, or from the late exponential phase of growth in 7H9 medium, using the CTAB method (Belisle and Sonnenberg 1998)(Coscolla et al, in preparation).

Sequencing (PacBio SMRT) and construction of PcbL5Ben. The PacBio genome was assembled using the in-house Sanger pipeline and checked for sufficient quality and coverage

L5 Illumina reads. In total 205 L5 genomes from various countries were included in the study. The genomes originated from South, East and Central Africa, but primarily (n=157) from two regions within West Africa: the Western part of West Africa (including The Gambia, Senegal, Mauritania, Sierra Leone, Ivory Coast, Liberia, Guinea and Mali) and the Eastern part of West Africa (including Ghana, Nigeria, Benin Niger and Burkina Faso) (Coscolla et al, in preparation).

Illumina sequencing. Sequencing libraries were prepared using NEXTERA XT DNA Preparation Kit (Illumina, San Diego, USA). Multiplexed libraries were paired-end sequenced on Illumina HiSeq2500 (Illumina, San Diego, USA) with 151 or 101 cycles at the Genomics Facility Basel, Germany (Coscolla et al, in preparation).

Annotation

The fasta sequences of the complete genomes were annotated using Prokka-1.12 (Seemann 2014) based on the reference genome H37Rv annotation.

Gene presence-absence analysis

Gene content differences were assessed with an all-vs-all BLASTn approach, using BLAST+ (version+2.8.1 (Altschul et al. 1990; Camacho et al. 2009)) (Fig.1). For a specific genome-genome comparison, the following procedure was used: the gene sequences of genome 1 (ffn file) were compared to those of genome 2 using the default cut-offs to look for any homology for each gene. Those genes found in genome 1 and not in genome 2 were then compared to the complete

genome (fna file) of genome 2 to look for pseudogenes using BLASTn. These pseudogenes were then confirmed using tBLASTn of the genome 1 protein sequences (faa file) compared to the complete genome (fna file) of genome 2. This procedure was used to compare all three L5 genomes to each other as well as each to H37Rv. Those genes that are present or absent in H37Rv (or pseudogenes in either) were compared in a similar manner to the L6 and *M. bovis* reference strains to determine if these genes/pseudogenes are L5-specific (present in L5, but not in H37Rv, L6 and *M. bovis*).

Mapping of L5 illumina reads to H37Rv vs PacBio L5 complete genome. After a quality control of L5 illumina reads, that involved checking the purity and depth coverage, those genomes without exogenous contamination and with at least 30x coverage, were mapped respectively to H37Rv and each of the 3 PacBio L5 genomes, using the MTBseq pipeline (Kohl et al. 2018). The depth mapping coverage of these Illumina genomes against the reference genomes (percent read mapping, unambiguous coverage mean) was compared between the three L5 PacBio genomes and between each L5 PacBio genome and H37Rv. Mapping statistics parameters such as percent unambiguous total base, uncovered bases, SNP, deletions, insertions, substitutions, percent genes mapped to reference, were also compared between L5 PacBio genomes and H37Rv.

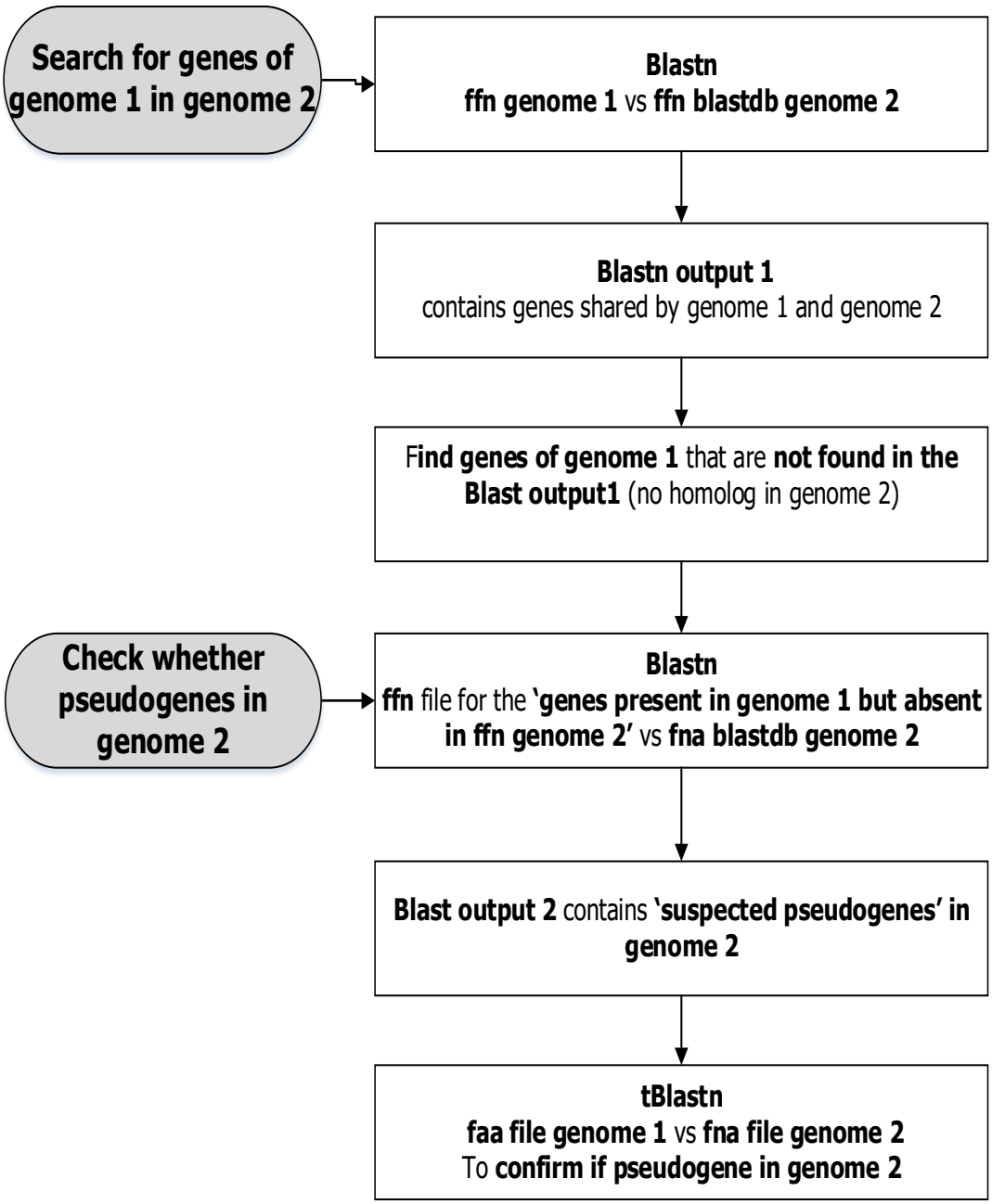

Figure 1. Methodology for finding gene presence-absence between two genomes

Checking L5 unique and lacking genes in other L5 genomes, and SNP in those confirmed as L5 specific. The genes found present or absent in L5 based on complete genome comparison with H37Rv were checked for their expected presence or absence in the L5 Illumina genomes. Using the Position Tables produced by the MTBseq pipeline, a gene was considered absent if 95% of its position in the genome had less than 8 reads covering them. From this data, a gene presence/absence matrix was generated for L5 illumina genomes mapped to H37Rv. The same was done for the mapping of L5 illumina genomes to each of the three L5 Pacbio genomes. Genes found to be L5-specific (present in Pacbio and Illumina genomes) were also checked for SNPs in these genes (in L5 illumina genomes), using each of the L5 Pacbio genomes as a reference.

Determination of putative function of genes differentially present or absent in the L5 complete genomes

To find the (putative) function of genes present or absent only in L5, and genes present in L5 but pseudogenes in H37Rv/L6/*M. bovis* or vice-versa, the fasta sequence of the genes were searched against the NR database of NCBI using BLASTX-2.8.1+ (Altschul et al. 1997) as well as against the Tuberculist database (http://tuberculist.epfl.ch), Mycobrowser (https://mycobrowser.epfl.ch/) and literature. Furthermore, the gene function group/class was found using the COG database (Tatusov et al. 2000) and Mycobrowser.

Visualization of the complete genomes

The ffn files of H37Rv, the three L5 PacBio and the L6 complete genomes, were aligned using progressive Mauve in Mauve (ver 20150226) to identify rearrangements and examine synteny.

Results

Gene content in the complete genomes

Using the H37Rv annotation, the number of genes – including duplicates - in the Pacbio L5 genomes was 4189 genes in the PcbL5Ben genome, 4162 in PcbL5Gbia, and 4134 in PcbL5Nig versus 4126 in H37RV, 4126 in the reference L6 genome and 4059 in the reference *M. bovis* genome. Hence, the largest difference observed in gene count among the three L5 isolates from this study was 55 genes, among the human-adapted lineages (H37Rv, L5, L6) 63 genes, and 130 among the human- and animal adapted lineage (*M. bovis*).

PacBio-based genome structure comparison

The visualization of the structure of the genomes showed that all three L5 genomes have a region (~33334 bp between 1683331-1716665; Fig.2) absent in H37Rv (gap between 1683331-1716665 on Fig.2). Furthermore, the Nigerian L5 genome missed an additional region (~33334 bp, region between 1966666-2000000 in Fig.2) present in the Benin and Gambian L5 genomes (Fig.2).

PacBio-based (pseudo-)gene presence/absence

The Benin genome contained 5 genes that were neither present in the Nigerian or Gambian genome, while the latter contained respectively 1 and 2 unique genes. The Benin and Gambian strain contained each 33 genes that were not present in the Nigerian strain (Fig.3, Table 1, Table S1). Interestingly, 32 of these genes were also present in H37Rv (Fig.4, Table 1, Table S1). Ten genes that were shared by the three L5 genomes were absent in H37Rv, while 9 genes present in H37Rv were not present in any of the three L5 genomes.

We qualified as ‘suspected pseudogenes’ any gene sequences that were not found in the ffn file (sequences for coding regions) of a genome but found in its fasta (fna) file (coding and non-coding regions of the genome sequence) that are not yet confirmed as pseudogenes as described above in the methods.

Six genes shared by the 3 Pacbio L5 were confirmed pseudogenes in H37Rv (Table S2). Three genes present in H37Rv were confirmed pseudogenes in the 3 Pacbio L5 genomes (Table S3). Two genes (*Rv2073c*, *Rv2074*) were only present in H37Rv but absent in L5, L6 and *M. bovis* complete genomes (Table 4).

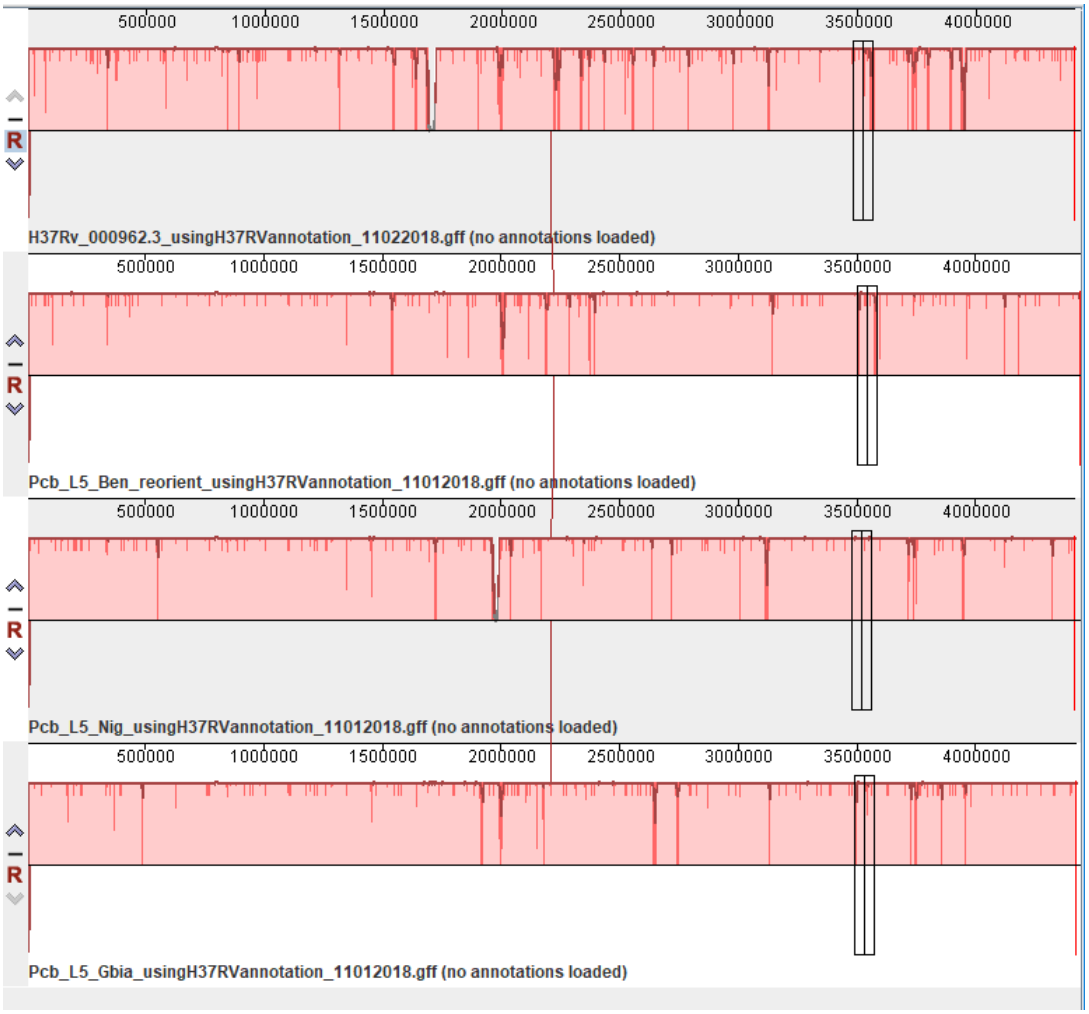

Figure 2. Visualization of possible rearrangements using Mauve pipeline: H37Rv (L4) genome and three complete PacBio genomes of *M. africanum* L5 isolates from Benin, The Gambia and Nigeria.

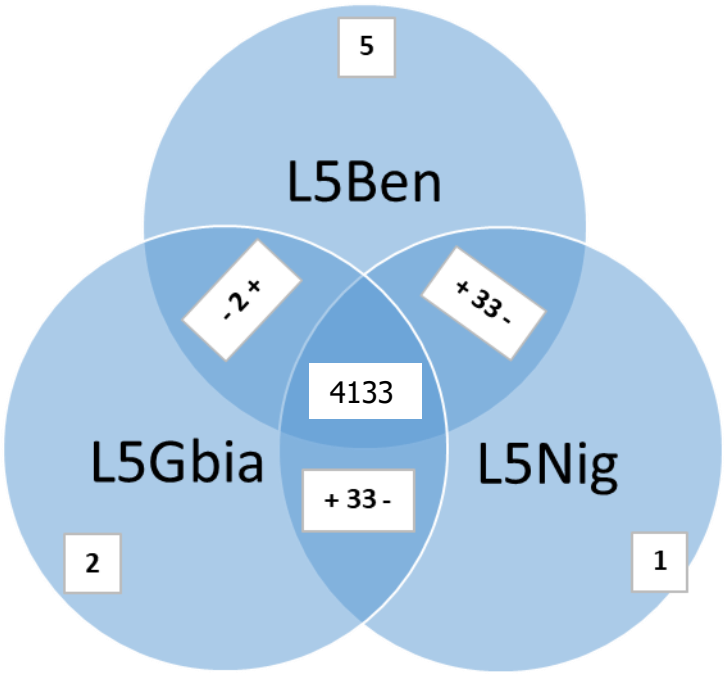

Figure 3. Gene count difference between three complete (PacBio SMRT) genomes of *M. africanum* L5 from Benin, The Gambia and Nigeria: L5Ben, L5Gbia and L5Nig.

The numbers are the count/number of gene difference (present or absent) between genomes. The sign (+) indicates the presence of genes while (-) indicates the absence of genes. For example, the "- 2 +" gene difference between L5Gbia and L5Ben, indicates that 2 genes are present in the L5Ben genome but absent in the L5Gbia genome. 5, 2, 1 are the number of genes only found in respectively L5Ben, L5Gbia, L5Nig.

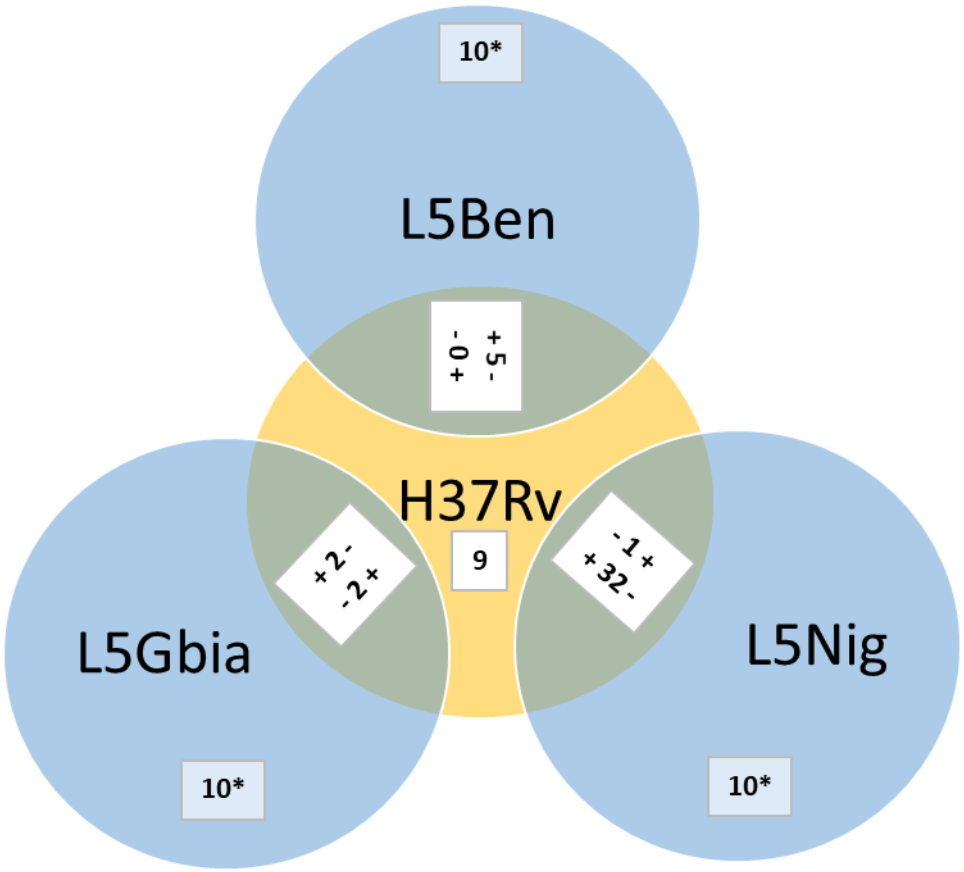

Figure 4. Gene count difference between *M. tuberculosis* H37Rv (L4) genome and three complete PacBio genomes of *M. africanum* L5 isolates from Benin, The Gambia and Nigeria.

***** Those 10 genes are present in each of the three L5 genomes, but absent in H37RV.

Table 1. Gene content difference between *M. tuberculosis* H37Rv (L4) genome and complete PacBio genomes of three *M. africanum* L5 isolates from Benin, The Gambia and Nigeria

		Present in			
		PcbL5Ben	**PcbL5Gbia**	**PcbL5Nig**	**H37Rv**
Absent in	**PcbL5Ben**		2[B]	1[C]	**9[D]**
	PcbL5Gbia	2 + 5[A]		1[C]	2 + **9[D]**
	PcbL5Nig	33 + 5[A]	33 + 2[B]		32 + **9[D]**
	H37Rv	**10[*]** + 5[A]	**10[*]** + 2[B]	**10[*]** + 1[C]	
		4189	4162	4134	4126

A: include 5 only present in PcbL5Ben

B: includes 2 only present in PcbL5Gbia

C: includes 1 only present in PcbL5Nig

D: includes 9 only present in H37Rv

*****: includes 10 genes shared by the PcbL5 (Ben, Gbia, Nig) and absent in H37Rv

Table 2. Mapping of 205 Illumina L5 genomes to the *M. tuberculosis* H37Rv (L4) genome and complete PacBio genomes of three *M. africanum* L5 isolates from Benin, The Gambia and Nigeria (mapping statistics/estimates)

	H37Rv	**PcbL5Ben**	**PcbL5Gbia**	**PcbL5Nig**
Reads				
Mean of Percent L5 illumina reads mapped to	96.9	**97.5**	97.3	96.5
Mean of unambiguous coverage mean	121.9	122.7	**122.8**	122.5
Bases				
Mean of Percent unambiguous total bases	0.983	0.988	**0.989**	0.982
Mean uncovered	30444.8	18801.7	**14657.5**	35215.4
Mean SNP	2210	529.9	**513.2**	**503.7**
Mean deletions	373.3	95.9	**95.7**	97.2
Mean insertions	239.2	**77.3**	107.6	**60.8**
Mean substitutions (including stop codons)	1193	**0**	**0**	**0**
Genes				
Mean of Percent gene mapped (presence)	99.59	99.72	**99.79**	99.76
L5 illumina having all the ref. genome genes % (n=205)	0	2 (4)	**7.3 (15)**	0
Mean gene count difference between L5 Illumina genomes (gene count per Illumina genome minus minimum gene count)	29.8	38.3	**33.4**	33

Gene presence/absence, related SNPs and functional groups of lineage specific genes in Illumina genomes of a wider set of clinical isolates

The mapping estimates of the Illumina-generated genomes are presented in Table 2. Mapping quality of mapping to a PacBio L5 reference is superior to the H37Rv reference approach (Table 2, Fig.5). Within the PacBio L5 genomes, the Benin and the Gambian had similar mapping estimates that were better than those of the Nigerian, likely due to the large gene differences between this L5 genome and the two others.

The genes specific to PcbL5Ben (5), PcbL5Gbia (2), PcbL5Nig (1) were each found at various rates among Illumina L5 genomes (77.6%-97.1%, 159-199/205, Table 3).

Table 3. Presence in 205 Illumina L5 genomes of genes detected in only one of the three PacBio L5 complete genomes from Benin, Nigeria or The Gambia

L5 complete genomes	Genes	Present in Illumina L5 % (n=205)
	PcbL5_01893	**77.6** (159)
	PcbL5_01894	**77.6** (159)
PcbL5Ben	PcbL5_01895	**77.6** (159)
	PcbL5_02043	**95.1** (195)
	PcbL5_03043	**97.1** (199)
PcbL5Gbia	PcbL5_02028	**95.1** (195)
	PcbL5_03020	**97.1** (199)
PcbL5Nig	PcbL5_02001	**95.1** (195)

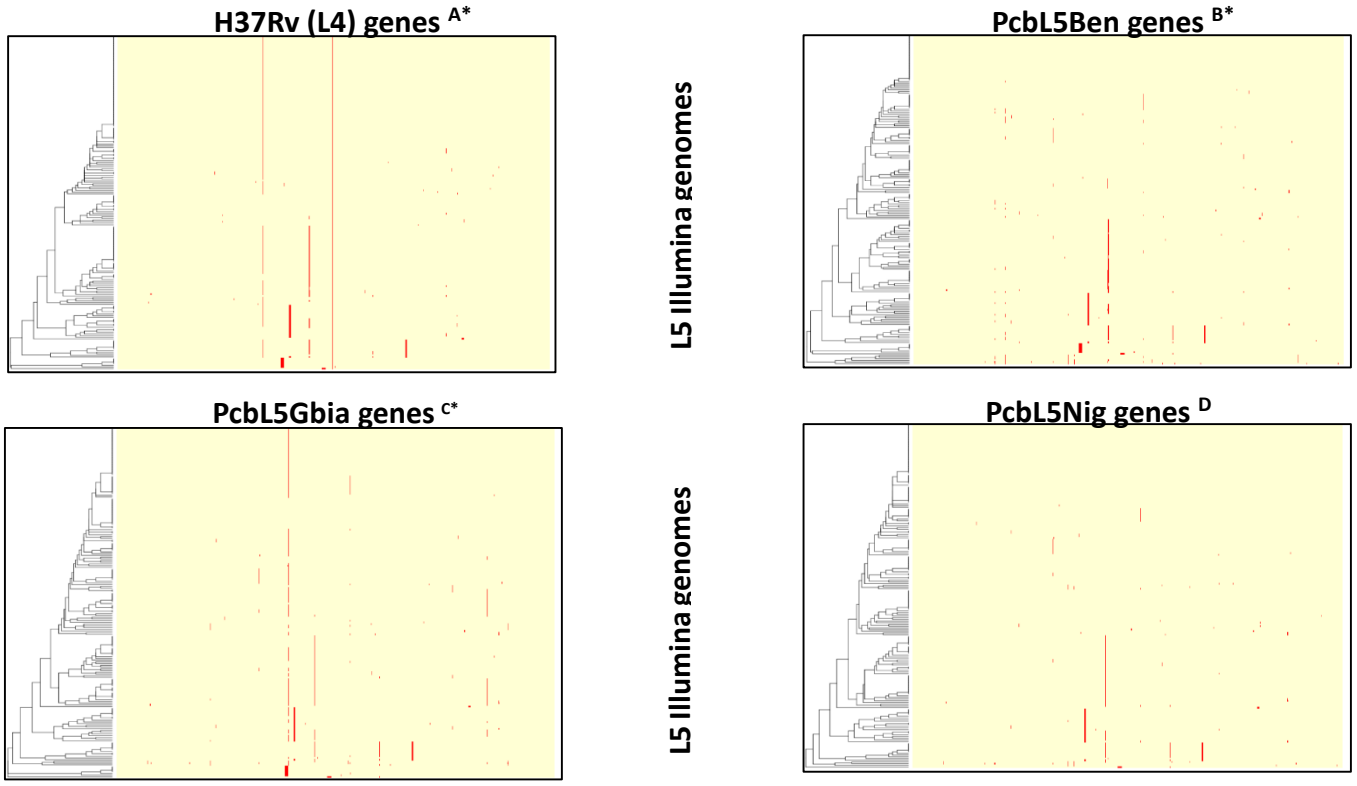

Figure 5. Mapping of 205 Illumina genomes (L5_ig) of *M. africanum* L5 from various countries to *M. tuberculosis* H37Rv (L4) and three *M. africanum* L5 isolates from Benin, The Gambia and Nigeria as reference (L5_cg_Ben; L5_cg_Gbia; L5_cg_Nig).

(A) **5**/9 genes present in H37Rv but absent in the three L5_cg were **absent in all 205 L5_ig.**
(B) **5**/5 genes **present in L5_cg_Ben but absent in the two other L5_cg** were **present in 159-199/205 L5_ig.**
(C) **2**/2 genes **present in L5_cg_Gbia but absent in the two other L5_cg** were **present in 195 & 199/205 L5_ig.**
(D) **1**/1 gene **present in L5_cg_Nig but absent in the two other L5_cg** was **present in 195/205 L5_ig.**
* The 32-33 genes absent in L5_cg_Nig but present in L5_cg_Ben, L5_cg_Gbia and *H37Rv* were also absent in 6/205 L5_ig (Fig.9).

.... **or** **....** gene absent

Interestingly, 2.9% of the Illumina L5 genomes (6/205) had similar patterns of large gene loss as the Nigerian PacBio as they missed 30 of the 33 genes present in the Benin and Gambian PacBio genomes and H37Rv. Those 30 genes constituted one region of 18 genes and another region of 12 genes. The six Nigerian-like Illumina L5 genomes formed a monophyletic group, within the other L5 genomes (Fig. 6), suggesting a single loss of these gene clusters, although those six Illumina L5 genomes originated from several different countries. These two blocks contain genes whose annotations include *mutB*, *mazE4*, *mazF4* and other (Table S4).

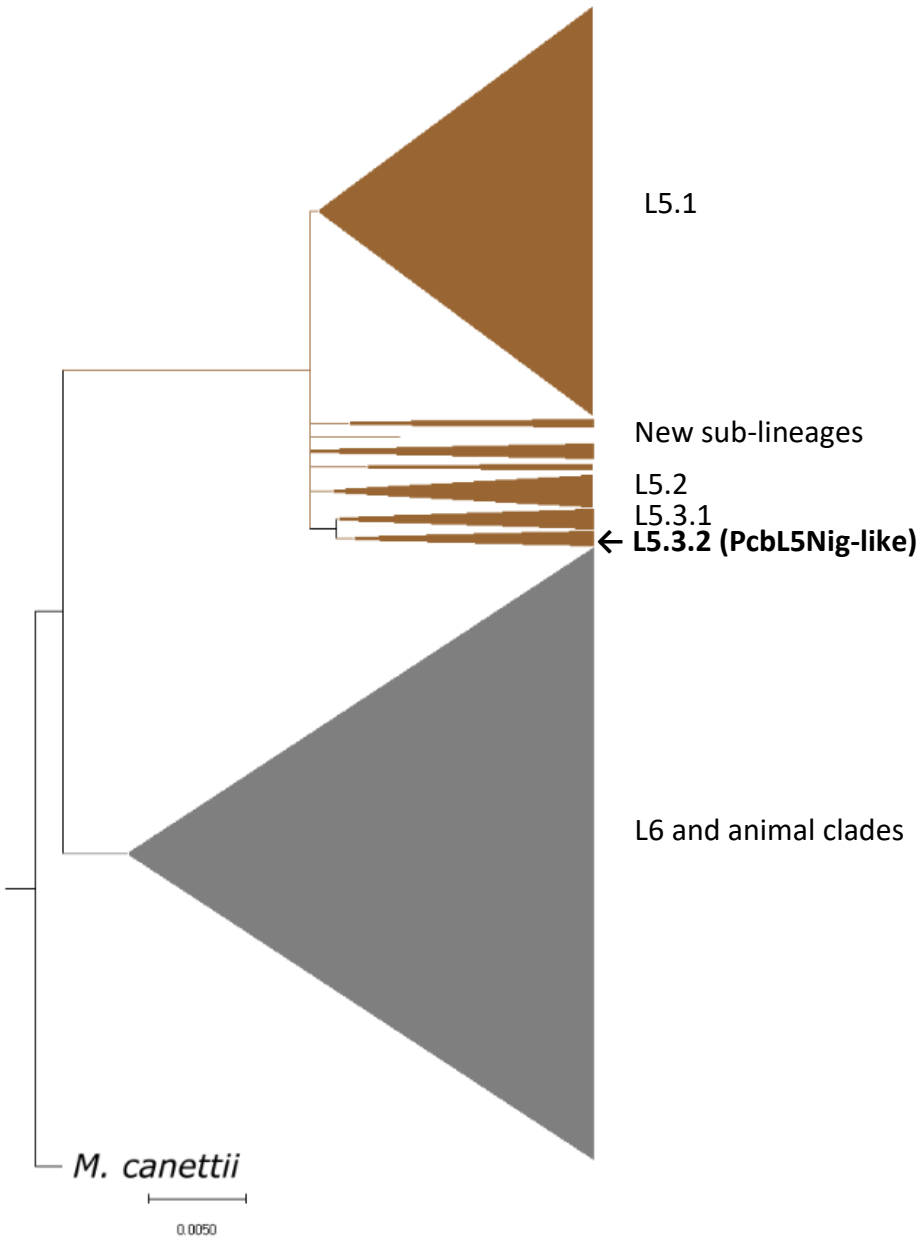

Figure 6. Phylogenetic tree showing the 6 Illumina L5 isolates (L5.3.2) similar to the PacBio genome of the Nigerian L5 strain (PcbL5Nig).

Four of the ten genes present in the three PacBio L5 but absent in H37Rv were found in all the 205 L5 Illumina genomes while the other were found in 81-98% (165-201) of the L5 Illumina genomes (Table 4). These four genes include: a hypothetical protein possibly an IS*256* transposase, *Mb2048c*, PE35, and a hypothetical protein possibly related to the CAAX conserved domain. Two of these four genes were present in L5 genomes only (PE35 and hypothetical protein (possibly CAAX conserved domain)) while the remaining two (*Mb2048c* and hypothetical protein (possibly IS256 transposase)) were found in L6 and *M. bovis* as well (Table 4). Importantly for phylogenetic purposes, a minority of the 205 L5 Illumina genomes showed SNPs in these four genes (1.5-3.4 % depending on the gene) (Table 4).

On the other hand, five (*Rv1977*, *Rv1979c*, *Rv1993c*, *Rv1995*, *Rv2073c*) of the 9 genes present in H37Rv and absent in the three PacBio L5 genomes were confirmed absent in all the 205 L5 Illumina genomes while three were found in a minority of the of the L5 Illumina genomes (0.5-1.5%; Table 5, Fig. 5). Four of these genes (*Rv1977*, *Rv1979c*, *Rv1993c*, *Rv1995*) were absent in L5 only (i.e. present in L6 and *M. bovis*) while the fifth (*Rv2073c*) was absent in L6 and *M. bovis* as well (Table 5).

Importantly, in the four L5-specific genes that were present in all the L5 Illumina genomes but lacking in H37Rv, SNPs were found in 3-7 strains, depending on the gene (1.5-3.4% of the 205 L5 Illumina genomes, Table 4).

The classification of the L5-specific genes into functional groups based on Mycobrowser and Tuberculist databases, is summarized in Table 4. PE/PPE protein family (possibly PE35) is involved in antigenic variation (NCBI) and virulence (STRING database). A CAAX conserved domain is implicated in post-translation modification by attaching to the isoprenoid proteins in the process

called prenylation (NCBI). Prenylation allows transduction signals to enable CAAX-box proteins (that do not have a transmembrane domain) to fulfil their various function (Gao, Liao, and Yang 2009). The IS*256* transposase in staphylococci and enterococci is involved in resistance to antibiotics, and in virulence (recombination for adaptation, invasion) (Murugesan et al. 2018; Gu et al. 2005). *Mb2048c* is classified in the functional group "unknown function" (Mycobrowser) and is expressed during exponential growth in Sauton's minimal medium (NCBI, Mycobrowser).

The functional group of the four genes absent in L5 only (PacBio and Illumina; *Rv1977*, *Rv1979c*, *Rv1993c*, *Rv1995*) are respectively "conserved hypotheticals", "cell wall and cell processes", "conserved hypotheticals" and "conserved hypotheticals" (Mycobrowser, Table 5). Of note, *Rv1979c* is associated with clofazimine and bedaquiline resistance (Zhang et al. 2015; Ghodousi et al. 2019), two of the drugs used for the treatment of RIF-resistant/MDR TB (WHO 2016, 2019).

The functional group for the H37Rv-specific gene *Rv2073c* (short chain dehydrogenase) absent in L5 (PacBio and Illumina), L6 and *M. bovis* (PacBio) is "intermediary metabolism and respiration". It is an NAD(P)-dependent oxidoreductase (NCBI) that play essential roles in central metabolic pathways including catalyzing a wide range of redox reactions using different substrates (Kavanagh et al. 2008; Sellés Vidal et al. 2018). The second gene of H37Rv (*Rv2074*, pyridoxamine-5-phosphate oxydase: vitamin B6 (pyridoxine)) absent in L6, *M. bovis* and the complete genomes of L5 (present in 1/205 (0.5%) Illumina L5 genomes), is also classified in the functional group "intermediary metabolism and respiration" and involved in pyridoxine biosynthesis (Mycobrowser). Vitamin B6 is essential for survival and virulence of *M.*

tuberculosis (Dick et al. 2010) and in immune-evasive mechanisms to allow bacterial persistence (its cofactor F420 dependent biliverdin reductase) (Ahmed et al. 2016; Selengut and Haft 2010)(Table 5).

The genes *Rv1978* and *Rv1994c* present in H37Rv and the complete genomes of L6 and *M. bovis*, were absent in the 3 complete genomes of L5, yet present in very few illumina L5 genomes (3/205, 1.5% for *Rv1978*; 1/205, 0.5% for *Rv1994c*). They are in the functional groups "conserved hypotheticals" and "regulatory proteins" respectively. *Rv1978* is required for bacterial survival in macrophages, and non-essential for in vitro growth of H37Rv (Rengarajan, Bloom, and Rubin 2005) (Mycobrowser, Tuberculist). *Rv1994c* is involved in the regulation and transport (efflux) of toxic metal especially copper which is toxic in excess (Samanovic et al. 2012). The functional similarity between *Rv1994c* (*cmtR*) and copper sensor *csoR*, suggests that the disruption of *Rv1994c* may hamper in vitro growth, and survival during the chronic phase of infection just as the disruption of *csoR* does (Marcus et al. 2016)(Rowland and Niederweis 2012)(Ward, Hoye, and Talaat 2008)

Table 4. Presence /absence in 205 Illumina L5 genomes of the 10 genes shared by the three PacBio L5 complete genomes but absent in the *M. tuberculosis* H37Rv genome

Genes present in the 3 L5 complete genomes but absent in H37Rv	Present (confirmed) in Illumina L5 genome n=205 (%)	SNP in L5 Illumina for genes absent in H37Rv but present in all L5 Illumina			Gene function			Present in complete genome of
		PcbL5Ben as ref n=205 (%)	PcbL5Gbia as ref n=205 (%)	PcbL5Nig as ref n=205 (%)	Functional group (Mycobrowser)	genes	Summary of role based on the literature	
PcbL5Ben_1364	191 (93.2)				ABC transporter ATP-binding protein/permease (Prokka)	Transmembrane ATP-binding protein ABC transporter	Conserved domains: FHA and CcmA (NCBI) FHA (forkhead associated domain, COG1716 (T)) binds pSer, pThr, pTyr (Signal transduction mechanisms) CcmA (COG1131 (V)), ABC-type multidrug transport system, ATPase component (Defense mechanisms (NCBI))	L5, bovis
PcbL5Ben_1365	191 (93.2)				Same as PcbL5Ben_1364	Same as PcbL5Ben_1364	Same as PcbL5Ben_1364	L5, bovis
PcbL5Ben_2129#	**205 (100)**	4 (2)	4 (2)	205* (100)	**PE/PPE**	PPE, possibly PE35 (77% id with PE35 from *M. braziliense*)	PE35 implicated in virulence (STRING database) and PE family is suggested to be related to antigenic variation of MTBC (NCBI)	**Only L5**
PcbL5Ben_2130#	**205 (100)**	7 (3.4)	7 (3.4)	7 (3.4)	Not in Mycobrowser	**Hypothetical protein possibly CAAX conserved domain** (NCBI)	Involved in post-translation modification by attaching to the isoprenoid proteins in the process called prenylation (NCBI). Most of CAAX box proteins do not have a transmembrane domain, thus, the prenylation process is crucial for the function of many signal transduction proteins (Gao, Liao, and Yang 2009)	**Only L5**
PcbL5Ben_2181	**205 (100)**	3 (1.5)	3 (1.5)	3 (1.5)	Not in Mycobrowser	**Hypothetical protein, possibly**	IS256 transposase is implicated in resistance to antibiotics and virulence (recombination	L5, L6, bovis

						IS256 transposase (90% id, NCBI)	for adaptation, invasion)(Murugesan et al. 2018; Gu et al. 2005)	
PcbL5Ben_2182	**205 (100)**	6 (2.9)	6 (2.9)	6 (2.9)	**Unknown function (Mb2048c**)	*Mb2048c*	Function unknown (expressed during exponential growth in Sauton's minimal media (NCBI)). Gene absent in H37Rv (Mycobrowser)	L5, L6, bovis
PcbL5Ben_3555	167 (81.5)				Intermediary metabolism and respiration	moaA3 (NCBI)	*moaA3* (molybdenum cofactor biosynthesis protein subunit MoaA , Cyclic pyranopterin monophosphate synthase (Prokka)) is contained in a IS6110 sequence deleted in H37Rv, present in *M. tuberculosis*_18b (Mycobrowser)	L5, L6, bovis
PcbL5Ben_3556	165 (80.5)				Not in Mycobrowser	Hypothetical protein, possibly DnrI superfamily (NCBI)	*DnrI*: DNA binding transcriptional activator of the SARP family (signal transduction mechanisms, COG3629 (NCBI))	L5, L6, bovis
PcbL5Ben_3557	165 (80.5)				Peptide synthase (Prokka)	Hypothetical protein, possibly IS6110 transposase (99-100% id NCBI)		L5, L6, bovis
PcbL5Ben_3669	201 (98)				PE/PPE	Hypothetical protein		L5, L6

*****In total 210 in the 205 L5 Illumina genome: more than one SNP in the gene for some strains

Present in L5 only, not present in L6 nor *M. bovis*

Table 5. Presence/absence in 205 Illumina L5 genomes of the 9 genes present in the *M. tuberculosis* H37Rv genome but absent in the three PacBio L5 complete genomes.

Genes present in H37Rv but absent in the 3 L5 complete genomes	Present in Illumina L5 genomes n=205 (%)	Absence (confirmed) in Illumina L5 genomes n=205 (%)	Gene function			Absent in complete genome of	Present in complete genome of
			Functional group (Mycobrowser)	genes	Summary of role based on the literature		
Part of Rv1899c (H37Rv_2002)	204 (99.5)	1 (0.5)	Hypothetical protein (35 aa) part of *Rv1899c* (343 aa, Cell wall and cell processes)(Tuberculist)	Hypothetical protein, part of Rv1899c (LppD: 4-hydroxybuturate dehydrogenase 173 aa) and also LpqI beta-hexoaminidase precursor (Herrmann et al. 2000)	Possibly bacterial membrane lipoprotein (LppD and LpqI)	L5, L6, bovis	**H37Rv only**
Rv1977 (H37Rv_2085)	**0**	**205 (100)**	**Conserved hypotheticals**	M48 peptidase family protein including as homologs: CAAX prenyl protease (*Htpx* (COG0501 O, heat shock stress response protein) M48-Ste24p-like	Bacterial survival (bacterial self-degradation to eliminate abnormal membrane proteins, htpx upregulated at high temperature, and also after 96 hours of starvation (NCBI))	**L5 only**	L6, bovis, H37Rv
Rv1978 (H37Rv_2086)	3 (1.5)	202 (98.5)	Conserved hypotheticals,	M48-2C-Ste24p (100% id), class I SAM-dependent methyltransferase (99.65% id), type 11 methyltransferase (97.1% id)	Bacterial survival in macrophages, and non-essential for in vitro growth of H37Rv (Rengarajan, Bloom, and Rubin 2005)(Mycobrowser, Tuberculist)	L5 only	L6, bovis, H37Rv

Rv1979c (H37Rv_2087)	**0**	**205 (100)**	**Cell wall and cell processes**	APC family permease	-Involved in the transport of **clofazimine** (and **bedaquiline**), associated with clofazimine and bedaquiline resistance (Zhang et al. 2015)(Ghodousi et al. 2019) -Disruption of the gene provides an in vitro growth advantage to H37Rv	**L5 only**	L6, bovis, H37Rv
Rv1993c (H37Rv_2107)	**0**	**205 (100)**	**Conserved hypotheticals**	*Rv0968* =*DUF1490* similar to *Rv1993c* (NCBI)	Similar to *Rv0968* which is included in the operon *cosR-Rv0968-ctpV* where *cosR* and *ctpV* are virulence associated (Rademacher and Masepohl 2012)	**L5 only**	L6, bovis, H37Rv
Rv1994c (H37Rv_2108)	1 (0.5)	204 (99.5)	Regulatory proteins	HTH transcriptional regulator *cmtR*	Operon *cmtR-Rv1993c-ctpG* is similar to *csoR-Rv0968-ctpV*. *cmtR* (*Rv1994c*) has the same role as *csoR*. These operons all function in the regulation and transport (efflux) of toxic metal especially copper which is toxic in excess (Samanovic et al. 2012). Disruption may hamper in vitro growth, and survival during chronic phase of infection as *csoR* absence	L5 only	L6, bovis, H37Rv

					(Marcus et al. 2016; Rowland and Niederweis 2012; Ward, Hoye, and Talaat 2008).		
Rv1995 (H37Rv_2109)	**0**	**205 (100)**	**Conserved hypothetical**	**Hemerythrin-domain containing protein (NCBI)**	- Involved in oxygen transport (NCBI)	**L5 only**	L6, bovis, H37Rv
Rv2073c (H37Rv_2188)	**0**	**205 (100)**	**Intermediary metabolism and respiration**	**Probable short chain dehydrogenase**	SDR family NAD(P)-dependent oxidoreductase (NCBI): catalyzes a wide range of reactions and substrate(Kavanagh et al. 2008; Sellés Vidal et al. 2018)	L5, L6, bovis	**H37Rv only**
Rv2074 (H37Rv_2189)	1 (0.5)	204 (99.5)	**Intermediary metabolism and respiration (Pyridoxamine, Vit B6)**	Pyridoxamine-5-phosphate oxidase (Mycobrowser, Tuberculist, NCBI) Its cofactor (Selengut and Haft 2010) F420 dependent biliverdin reductase (99.2-100% id, NCBI)(Ahmed et al. 2016)	- Biosynthesis of (pyridoxal phosphate and) pyridoxine (vitamin B6, Mycobrowser) which is essential for survival and virulence of M. tuberculosis (Dick et al. 2010) - Its cofactor implicated in Immuno-evasive mechanism to allow bacterial persistence (Ahmed et al. 2016; Selengut and Haft 2010)	L5, L6, bovis	H37Rv only

Discussion

We assessed whether the H37Rv, which is an L4 strain currently used worldwide as reference for all lineages genomes analyses, is representative of L5 genomes in terms of gene content, and whether it is reliable to use it for L5 genome analyses, especially for transmission detection.

We found that some genes present in all the L5 isolates (PacBio and Illumina genomes) of our multi-country collection, were absent in the H37Rv genome. Furthermore, those genes contain SNPs that could be used to identify additional diversity between L5 strains. This has strong implications when using precise SNP cut-offs to define transmission clusters (Walker et al. 2013; Meehan et al. 2018) as SNPs in these genes would be missed with H37Rv mapping, resulting in incorrect transmission estimates. Our findings suggest that using an L5 genome as a reference would increase the resolution of L5 genome analyses over the H37Rv approach. This is in line with other reports that some genes present in MTBC isolates were absent in the H37Rv genome (Periwal et al. 2015)(O'Toole and Gautam 2017), and with their recommendation to use additional reference genomes different to H37Rv (Norman et al. 2019; O'Toole and Gautam 2017).

Contrary, Lee *et al*. - after comparing various lineages as reference genome for L4 genome analysis - concluded that there is no need to use a lineage-specific reference genome (Lee and Behr 2016). However, their observation was based on the analysis of a set of modern strains (L4) using an L4 and more ancestral lineages as a reference. They did not analyze L5 genomes using various lineages as reference. Although our comparison of L5 to L4 (H37Rv) showed that each of the 2 lineages has genes absent in the other, a difference in the repartition of sub-lineages in the genomes to analyze, the number of lineage-specific genes

in the lineage to analyze versus the reference genome lineage, and the presence of SNPs in the genes specific to the lineage (genomes) to analyze are important factors that can make a difference in the result. For instance, in this study the PacBio L5 Benin (PcbL5Ben) genome had 15 genes absent in H37Rv with 4 of them present in all Illumina L5 with SNP in some isolates, while 5 of 9 H37Rv-genes were absent in all L5. The other 11 genes of PcbL5Ben were variably present in the Illumina L5 and would still add to the diversity captured when using an L5 as a reference genome compared to H37Rv. Furthermore, our mapping statistics confirmed that the mapping of the Illumina L5 genomes to PacBio L5 as reference was better than when mapped to H37Rv.

Although the use of an L5 genome as reference would have many benefits over H37Rv, gene content differences were still observed within the L5 lineage. Each of the three PacBio L5 genomes had genes unshared with the two others that were all found in Illumina L5 genomes (thus excluding exogenous contamination). Furthermore, 33 genes shared by the Benin and Gambian PacBio genomes were absent in PcbL5Nig despite its closed genome status. The absence of 30 of those genes was also found in six Illumina L5 genomes, confirming the good quality of the PcbL5Nig genome and showing that it is representative of some circulating L5 strains. Deletions between MTBC isolates were previously reported (Kato-Maeda et al. 2001), but not explicitly within a lineage. The phylogenetic tree also grouped the 6 isolates into a monophyletic group of the L5.3 sublineage (Coscolla et al, in preparation) with these 30 genes probably a marker of L5.3 sublineage. This suggests that both lineages and sub-lineages may have regions of difference, indicating a need for more closed genomes of MTBC sub-lineages. The variable presence of genes within one lineage suggests that an L5 pangenome-based reference genome capturing all the known diversity may be required instead of the “best representative

unmodified" L5 reference genome approach. The MTBC-wide pangenome approach was also suggested before (Periwal et al. 2015; Meehan et al. 2019) but has its own drawbacks of between-strain comparisons and mapping of reads. Other approaches proposed include the inferred ancestral genome representative of MTBC lineages (Goig et al. 2018; Comas et al. 2013; Meehan et al. 2018) and *de novo* assembly reference free approach (Maretty et al. 2017; Iqbal et al. 2012; Meehan et al. 2019).

The H37Rv genes *Rv1977*, *Rv1979*, *Rv1993* and *Rv1995* absent in L5 only (PacBio and Illumina genomes), but present in L6 and *M. bovis* are respectively implicated in bacterial survival, transport of clofazimine and bedaquiline through the bacterial membrane (Mycobrowser), virulence and oxygen transport (Mycobrowser). This implies that L5 would less likely survive, less likely to be virulent, and more likely to be resistant to clofazimine and bedaquiline, compared to L4, L6 and *M. bovis*, provided that other genes specific to L6 and/or *M. bovis* do not have an attenuating effect. Resistance to clofazimine and bedaquiline which are 2nd line drugs used for the treatment of rifampicin-resistant patients is also associated with mutations in the genes *Rv0678* and *Rv2535c* (*pepQ*)*, atpE* (Zhang et al. 2015; Yew et al. 2017; Ghodousi et al. 2019; Almeida et al. 2016).

The H37Rv genes *Rv1978*, *Rv1994c* and *Rv2074*, absent in the vast majority of Illumina L5 genomes are all implicated in bacterial survival (Table 4). In addition, *Rv1994c* is needed for *in vitro* growth and *Rv2074* enables an immune-evasive mechanism to allow bacterial persistence (Ahmed et al. 2016; Selengut and Haft 2010) and the biosynthesis of vitamin B6 essential for virulence (Dick et al. 2010). Therefore, the absence of these genes in most L5 isolates suggests that L5 would be less likely to survive (in macrophages, Rv1978), less grow or grow

slower *in vitro*, and be less immune-evasive, less persistent, and less virulent than L4 (at least H37Rv). This confirms the suggestion that *M. africanum* (including L5) is less virulent than modern lineages (Coscolla and Gagneux 2014) and also previous findings on the reduced ability of L5 to grow in culture (Sanoussi et al. 2017). The gene *Rv2074* - implicated in the immune-evasive mechanism for bacterial persistence and the biosynthesis of vitamin B6 essential for survival and virulence - is absent in most L5 and in the complete genome of L6 and *M. bovis* as well. This suggests that L6 and *M. bovis* as well would be less likely to be immuno-evasive, to be persistent and to survive than L4; and explain the findings that L6 is less virulent than L4.

A limitation of this study is that the PacBio and Illumina genomes were derived from positive cultures, excluding possible minority L5 strain diversity as L5 is more represented in negative cultures (Sanoussi et al. 2017). Whole genome sequencing applied directly on sputum is more and more needed, especially for ancestral lineages (including L5 and L6) where growth in culture is problematic (negative culture or dysgonic isolates that are challenging to be grown for DNA extraction (Sanoussi et al. 2018, 2017)). Also, our study included only a single complete genome of L4, L6 and *M. bovis*, while these lineages may have their own variability similar to the intra-L5 variability we observed in our study.

Further studies could be conducted using an L5 reference genome to identify sub-lineages of L5 in L5-endemic countries and to find characteristics (host factors, host origin within West-Africa or other) associated with sub-lineages. For other lineages, studies similar to the present study should be conducted to find the best reference genome for each lineage, and/or construct a MTBC-wide pangenome.

In conclusion, the use of an L5 reference genome is preferable for L5 genome analyses for epidemiology (transmission), phylogeny and sub-lineage determination. The high within-lineage gene content variability suggests the pangenome of MTBC may be larger than previously thought, implying a reference-free genome assembly approach may be needed.

References

Ahmed, F Hafna, A Elaaf Mohamed, Paul D Carr, Brendon M Lee, Karmen Condic-Jurkic, Megan L O'mara, and Colin J Jackson. 2016. "Rv2074 Is a Novel F 420 H 2-Dependent Biliverdin Reductase in Mycobacterium Tuberculosis." https://doi.org/10.1002/pro.2975.

Almeida, Deepak, Thomas Ioerger, Sandeep Tyagi, Si-Yang Li, Khisimuzi Mdluli, Koen Andries, Jacques Grosset, Jim Sacchettini, Eric Nuermberger, and Janssen Pharmaceutica. 2016. "Mutations in PepQ Confer Low-Level Resistance to Bedaquiline and Clofazimine in Mycobacterium Tuberculosis." https://doi.org/10.1128/AAC.00753-16.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Altschul, Stephen F, Thomas L Madden, Alejandro A Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research*. Vol. 25. Oxford University Press. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/pdf/253389.pdf.

Belisle, John T., and Michael G. Sonnenberg. 1998. "Isolation of Genomic DNA from Mycobacteria." In *Mycobacteria Protocols*, 31–44. New Jersey: Humana Press. https://doi.org/10.1385/0-89603-471-2:31.

Bentley, Stephen D., Iñaki Comas, Josephine M. Bryant, Danielle Walker, Noel H. Smith, Simon R. Harris, Scott Thurston, et al. 2012. "The Genome of Mycobacterium Africanum West African 2 Reveals a Lineage-Specific Locus and Genome Erosion Common to the M. Tuberculosis Complex." Edited by Pamela L. C. Small. *PLoS Neglected Tropical Diseases* 6 (2): e1552. https://doi.org/10.1371/journal.pntd.0001552.

Bifani, P, S Moghazeh, B Shopsin, J Driscoll, A Ravikovitch, and B N Kreiswirth. 2000. "Molecular Characterization of Mycobacterium Tuberculosis H37Rv/Ra Variants: Distinguishing the Mycobacterial Laboratory Strain." *Journal of Clinical Microbiology* 38 (9): 3200–3204. https://www.ncbi.nlm.nih.gov/pubmed/10970357.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10 (December): 421. https://doi.org/10.1186/1471-2105-10-421.

Camus, Jean-Christophe, Melinda J Pryor, Claudine Me, and Stewart T Cole. 2019. "Re-Annotation of the Genome Sequence of Mycobacterium Tuberculosis H37Rv."

Vol. 24. www.sanger.ac.uk.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, et al. 1998. “Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence.” *Nature* 393 (6685): 537–44. https://doi.org/10.1038/31159.

Comas, Iñaki, Mireia Coscolla, Tao Luo, Sonia Borrell, Kathryn E Holt, Midori Kato-Maeda, Julian Parkhill, et al. 2013. “Out-of-Africa Migration and Neolithic Coexpansion of Mycobacterium Tuberculosis with Modern Humans.” *Nature Genetics* 45 (10). https://doi.org/10.1038/ng.2744.

Coscolla, Mireia, and Sebastien Gagneux. 2014. “Consequences of Genomic Diversity in Mycobacterium Tuberculosis.” *Seminars in Immunology* 26 (6): 431–44. https://doi.org/10.1016/j.smim.2014.09.012.

DeJong, Bouke C, Philip C Hill, Alex Aiken, Timothy Awine, Martin Antonio, Ifedayo M Adetifa, Dolly J Jackson-Sillah, et al. 2008. “Progression to Active Tuberculosis, but Not Transmission, Varies by M. Tuberculosis Lineage in The Gambia.” *J Infect Dis* 198 (7): 1037–43. https://doi.org/10.1086/591504.

Dick, Thomas, Ujjini Manjunatha, Barbara Kappes, and Martin Gengenbacher. 2010. “Vitamin B6 Biosynthesis Is Essential for Survival and Virulence of Mycobacterium Tuberculosis.” *Molecular Microbiology* 78 (4): 980–88. https://doi.org/10.1111/j.1365-2958.2010.07381.x.

Gagneux, Sebastien. 2017. *Strain Variation in the Mycobacterium Tuberculosis Complex: Its Role in Bio...: Discovery Service for ITM Library*. Edited by Sebastien Gagneux. Advances in Experimental Medicine and Biology, v. 1019. http://eds.b.ebscohost.com/eds/detail/detail?vid=0&sid=19804620-6ed7-43ac-996b-c04baea11da9%40sessionmgr120&bdata=JnNpdGU9ZWRzLWxpdmU%3D#db=nlebk&AN=1632545.

———. 2018. “Ecology and Evolution of Mycobacterium Tuberculosis.” *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro.2018.8.

Gao, Juehua, Jie Liao, and Guang-Yu Yang. 2009. “Review Article CAAX-Box Protein, Prenylation Process and Carcinogenesis.” *Am J Transl Res*. Vol. 1. www.ajtr.org/AJTR905003.

Gehre, Florian, Jacob Otu, Kathryn DeRiemer, Paola Florez de Sessions, Martin L. Hibberd, Wim Mulders, Tumani Corrah, Bouke C. de Jong, and Martin Antonio. 2013. “Deciphering the Growth Behaviour of Mycobacterium Africanum.” *PLoS Neglected Tropical Diseases* 7 (5). https://doi.org/10.1371/journal.pntd.0002220.

Ghodousi, Arash, Alamdar Hussain Rizvi, Aurangzaib Quadir Baloch, Abdul Ghafoor,

Faisal Masood, Mehmood Qadir, Emanuele Borroni, Alberto Trovato, Sabira Tahseen, and Daniela Maria. 2019. “Acquisition of Cross-Resistance to Bedaquiline and Clofazimine Following Treatment for Tuberculosis in 2 Pakistan.” https://doi.org/10.1101/630012.

Goig, Galo A, Silvia Blanco, Alberto L Garcia-Basteiro, and Iñaki Comas. 2018. “Pervasive Contaminations in Sequencing Experiments Are a Major Source of False Genetic Variability: A Mycobacterium Tuberculosis Meta-Analysis.” https://doi.org/10.1101/403824.

Gonzalo-Asensio, Jesús, Wladimir Malaga, Alexandre Pawlik, Catherine Astarie-Dequeker, Charlotte Passemar, Flavie Moreau, Françoise Laval, et al. 2014. “Evolutionary History of Tuberculosis Shaped by Conserved Mutations in the PhoPR Virulence Regulator.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (31): 11491–96. https://doi.org/10.1073/pnas.1406693111.

Gu, J, H Li, M Li, C Vuong, M Otto, Y Wen, and Q Gao. 2005. “Bacterial Insertion Sequence IS256 as a Potential Molecular Marker to Discriminate Invasive Strains from Commensal Strains of Staphylococcus Epidermidis.” *The Journal of Hospital Infection* 61 (4): 342–48. https://doi.org/10.1016/j.jhin.2005.04.017.

Herrmann, Jean Louis, Robin Delahay, Alex Gallagher, Brian Robertson, and Douglas Young. 2000. “Analysis of Post-Translational Modification of Mycobacterial Proteins Using a Cassette Expression System.” *FEBS Letters* 473 (3): 358–62. https://doi.org/10.1016/S0014-5793(00)01553-2.

Iqbal, Zamin, Mario Caccamo, Isaac Turner, Paul Flicek, and Gil Mcvean. 2012. “De Novo Assembly and Genotyping of Variants Using Colored de Bruijn Graphs.” https://doi.org/10.1038/ng.1028.

Jong, Bouke C De, Martin Antonio, and Sebastien Gagneux. n.d. “Mycobacterium Africanum-Review of an Important Cause of Human Tuberculosis in West Africa.” Accessed May 21, 2019. https://doi.org/10.1371/journal.pntd.0000744.

Kato-Maeda, Midori, Jeanne T Rhee, Thomas R Gingeras, Hugh Salamon, Jorg Drenkow, Nat Smittipat, and Peter M Small. 2001. “Comparing Genomes within the Species Mycobacterium Tuberculosis.” https://doi.org/10.1101/gr166401.

Kavanagh, K L, H Jçrnvall, B Persson, and U Oppermann. 2008. “The SDR Superfamily: Functional and Structural Diversity within a Family of Metabolic and Regulatory Enzymes” 171: 77. https://doi.org/10.1007/s00018-008-8588-y.

Kohl, Thomas Andreas, Christian Utpatel, Viola Schleusener, Maria Rosaria De Filippo, Patrick Beckert, Daniela Maria Cirillo, and Stefan Niemann. 2018. “MTBseq: A Comprehensive Pipeline for Whole Genome Sequence Analysis of Mycobacterium Tuberculosis Complex Isolates.”

https://doi.org/10.7717/peerj.5895.

Leao, Sylvia, Anandi Martin, Gloria Isabel Mejia, Francoise Portaels, and Por- Taels F Leao SC, Martin A, Mejia GI, Palomino JC, Robledo J, Telles MAS. 2004. “Practical Handbook for the Phenotypic and Genotypic Identification of Mycobacteria.,” 77–126. https://doi.org/10.1007/BF01045823.

Lee, Robyn S, and Marcel A Behr. 2016. “Does Choice Matter? Reference-Based Alignment for Molecular Epidemiology of Tuberculosis.” *Journal of Clinical Microbiology* 54 (7): 1891–95. https://doi.org/10.1128/JCM.00364-16.

Magnus, K. 1966. “Epidemiological Basis of Tuberculosis Eradication 3. Risk of Pulmonary Tuberculosis after Human and Bovine Infection *.” *Bull. Org. Mond. Sante*. Vol. 35. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2476032/pdf/bullwho00607-0027.pdf.

Malone, Kerri M, Damien Farrell, Tod P Stuber, Olga T Schubert, Ruedi Aebersold, Suelee Robbe-Austerman, and Stephen V Gordon. 2017. “Updated Reference Genome Sequence and Annotation of Mycobacterium Bovis AF2122/97.” *Genome Announcements* 5 (14): e00157-17. https://doi.org/10.1128/genomeA.00157-17.

Marcus, Sarah A, Sarah W Sidiropoulos, Howard Steinberg, and Adel M Talaat. 2016. “CsoR Is Essential for Maintaining Copper Homeostasis in Mycobacterium Tuberculosis.” https://doi.org/10.1371/journal.pone.0151816.

Maretty, Lasse, Jacob Malte Jensen, Bent Petersen, Jonas andreas sibbesen, siyang Liu, Palle Villesen, Laurits skov, et al. 2017. “Sequencing and de Novo Assembly of 150 Genomes from Denmark as a Population Reference.” *Nature Publishing Group* 548. https://doi.org/10.1038/nature23264.

Meehan, Conor J., Pieter Moris, Thomas A. Kohl, Jūlija Pečerska, Suriya Akter, Matthias Merker, Christian Utpatel, et al. 2018. “The Relationship between Transmission Time and Clustering Methods in Mycobacterium Tuberculosis Epidemiology.” *EBioMedicine* 37 (November): 410–16. https://doi.org/10.1016/j.ebiom.2018.10.013.

Meehan, Conor J, Galo A Goig, Thomas A Kohl, Lennert Verboven, Anzaan Dippenaar, Matthew Ezewudo, Maha R Farhat, et al. 2019. “Whole Genome Sequencing of Mycobacterium Tuberculosis: Current Standards and Open Issues.” *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0214-5.

Murugesan, Saravanan, Stalin Mani, Indhumathy Kuppusamy, and Padma Krishnan. 2018. “Role of Insertion Sequence Element Is256 as a Virulence Marker and Its Association with Biofilm Formation among Methicillin-Resistant Staphylococcus Epidermidis from Hospital and Community Settings in Chennai, South India.”

Indian Journal of Medical Microbiology 36 (1): 124. https://doi.org/10.4103/ijmm.IJMM_17_276.

NCBI. 2017. "Mycobacterium Tuberculosis H37Rv, Complete Genome." 2017. https://www.ncbi.nlm.nih.gov/nuccore/NC_000962.3/.

Norman, Anders, Dorte Bek Folkvardsen, Søren Overballe-Petersen, and Troels Lillebaek. 2019. "Complete Genome Sequence of Mycobacterium Tuberculosis DKC2, the Predominant Danish Outbreak Strain." https://doi.org/10.1128/MRA.

O'Toole, Ronan F., and Sanjay S. Gautam. 2017. "Limitations of the Mycobacterium Tuberculosis Reference Genome H37Rv in the Detection of Virulence-Related Loci." *Genomics* 109 (5–6): 471–74. https://doi.org/10.1016/j.ygeno.2017.07.004.

Ofori-Anyinam, Boatema, Fatoumatta Kanuteh, Schadrac C. Agbla, Ifedayo Adetifa, Catherine Okoi, Gregory Dolganov, Gary Schoolnik, et al. 2016. "Impact of the Mycobaterium Africanum West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased Sensitivity of Rapid Identification Tests in The Gambia." *PLoS Neglected Tropical Diseases* 10: 1–12. https://doi.org/10.1371/journal.pntd.0004801.

Periwal, Vinita, Ashok Patowary, Shamsudheen Karuthedath Vellarikkal, Anju Gupta, Meghna Singh, Ashish Mittal, Shamini Jeyapaul, et al. 2015. "Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of Mycobacterium Tuberculosis Pangenome." https://doi.org/10.1371/journal.pone.0122979.

Phelan, Jody, Paola Florez De Sessions, Leopold Tientcheu, Joao Perdigao, Diana Machado, Rumina Hasan, Zahra Hasan, et al. 2018. "Methylation in Mycobacterium Tuberculosis Is Lineage Specific with Associated Mutations Present Globally." *Scientific Reports* 8 (1): 1–7. https://doi.org/10.1038/s41598-017-18188-y.

Rademacher, Corinna, and Bernd Masepohl. 2012. "Copper-Responsive Gene Regulation in Bacteria." https://doi.org/10.1099/mic.0.058487-0.

Rengarajan, Jyothi, Barry R Bloom, and Eric J Rubin. 2005. "Genome-Wide Requirements for Mycobacterium Tuberculosis Adaptation and Survival in Macrophages." *PNAS June*. Vol. 7. www.pnas.orgcgidoi10.1073pnas.0503272102.

Rowland, Jennifer L, and Michael Niederweis. 2012. "Resistance Mechanisms of Mycobacterium Tuberculosis against Phagosomal Copper Overload." https://doi.org/10.1016/j.tube.2011.12.006.

S.R. Pattyn, F. Portaels, L. Spanoghe, and J. Magos. 1970. "Further Studies on African Strains of Mycobacterium Tuberculosis. Comparison with M. Bovis and M.

Microti." Ann. Soc. belge Méd. Trop. https://www.ncbi.nlm.nih.gov/pubmed/4996738.

Samanovic, Marie I., Chen Ding, Dennis J. Thiele, and K. Heran Darwin. 2012. "Copper in Microbial Pathogenesis: Meddling with the Metal." *Cell Host & Microbe* 11 (2): 106–15. https://doi.org/10.1016/j.chom.2012.01.009.

Sanoussi, C. N'Dira, Dissou Affolabi, Leen Rigouts, Séverin Anagonou, and Bouke de Jong. 2017. "Genotypic Characterization Directly Applied to Sputum Improves the Detection of Mycobacterium Africanum West African 1, under-Represented in Positive Cultures." *PLOS Neglected Tropical Diseases* 11 (9): e0005900. https://doi.org/10.1371/journal.pntd.0005900.

Sanoussi, C. N'Dira, Bouke C De Jong, Mathieu Odoun, Karamatou Arekpa, Moulikatou Ali Ligali, Ousman Bodi, Simon Harris, et al. 2018. "Low Sensitivity of the MPT64 Identification Test to Detect Lineage 5 of the Mycobacterium Tuberculosis Complex." https://doi.org/10.1099/jmm.0.000846.

Seemann, T. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics* 30 (14): 2068–69. https://doi.org/10.1093/bioinformatics/btu153.

Selengut, Jeremy D, and Daniel H Haft. 2010. "Unexpected Abundance of Coenzyme F 420-Dependent Enzymes in Mycobacterium Tuberculosis and Other Actinobacteria †." *JOURNAL OF BACTERIOLOGY* 192 (21): 5788–98. https://doi.org/10.1128/JB.00425-10.

Sellés Vidal, Lara, Ciarán L. Kelly, Paweł M. Mordaka, and John T. Heap. 2018. "Review of NAD(P)H-Dependent Oxidoreductases: Properties, Engineering and Application." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1866 (2): 327–47. https://doi.org/10.1016/j.bbapap.2017.11.005.

Tatusov, Roman L, Michael Y Galperin, Darren A Natale, and Eugene V Koonin. 2000. "The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution." *Nucleic Acids Research*. Vol. 28. http://www.ncbi.nlm.nih.gov/COG.

Walker, Timothy M, Camilla LC Ip, Ruth H Harrell, Jason T Evans, Georgia Kapatai, Martin J Dedicoat, David W Eyre, et al. 2013. "Whole-Genome Sequencing to Delineate Mycobacterium Tuberculosis Outbreaks: A Retrospective Observational Study." *The Lancet Infectious Diseases* 13 (2): 137–46. https://doi.org/10.1016/S1473-3099(12)70277-3.

Ward, Sarah K, Elizabeth A Hoye, and Adel M Talaat. 2008. "The Global Responses of Mycobacterium Tuberculosis to Physiological Levels of Copper †." *JOURNAL OF BACTERIOLOGY* 190 (8): 2939–46. https://doi.org/10.1128/JB.01847-07.

WHO. 2016. "The shorter MDR-TB regimen." www.who.int/tb.

———. 2019. *WHO Consolidated Guidelines on Drug-Resistant Tuberculosis*

Treatment. http://apps.who.int/bookorders.

Yew, Wing Wai, Dachao Liang, Denise P Chan, Wanliang Shi, and Ying Zhang. 2017. "Molecular Mechanisms of Clofazimine Resistance in Mycobacterium Tuberculosis." *J Antimicrob Chemother* 72: 2943–44. https://doi.org/10.1093/jac/dkx227.

Zhang, Shuo, Jiazhen Chen, Peng Cui, Wanliang Shi, Wenhong Zhang, and Ying Zhang. 2015. "Identification of Novel Mutations Associated with Clofazimine Resistance in Mycobacterium Tuberculosis." https://doi.org/10.1093/jac/dkv150.

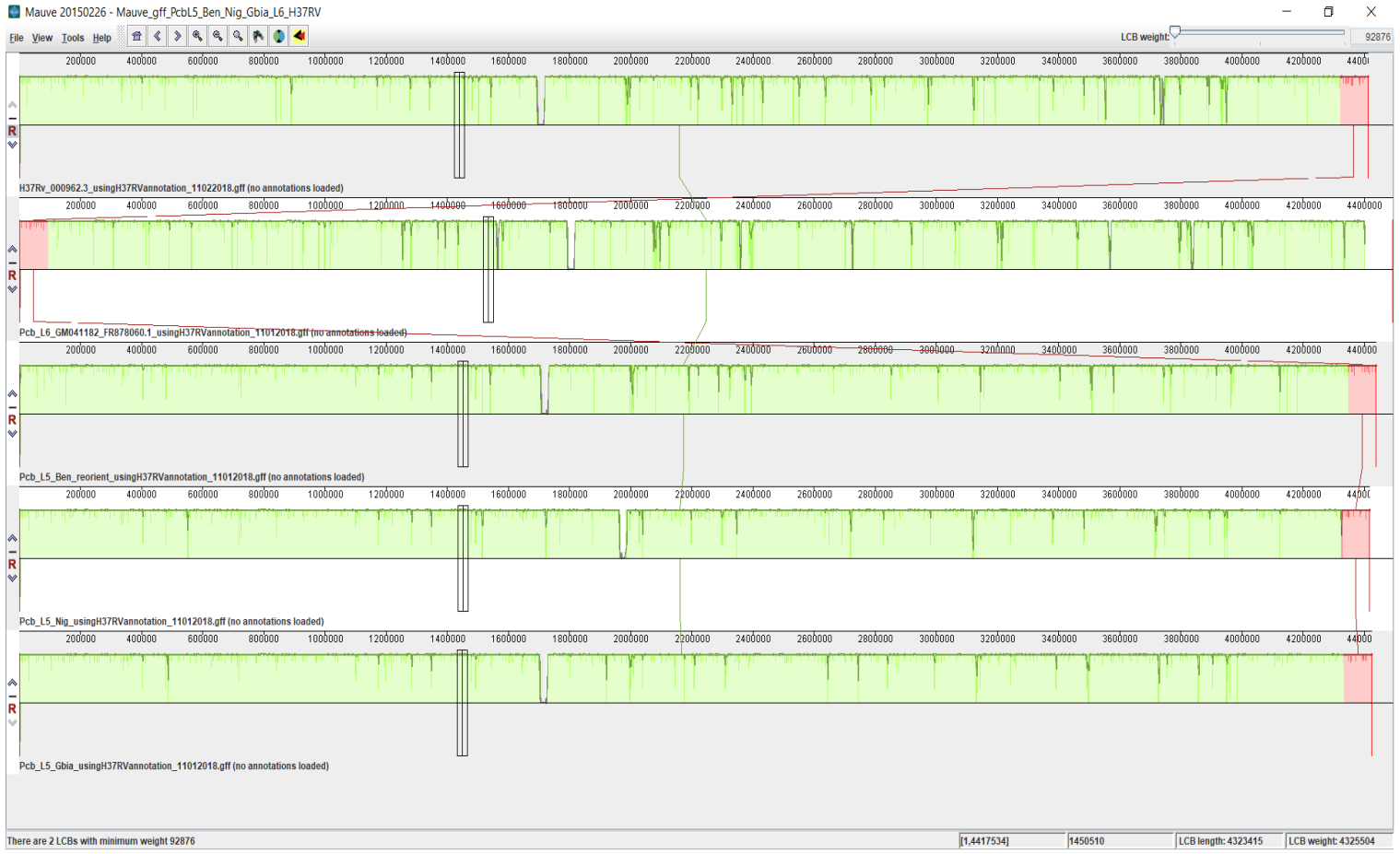

Fig S1. Visualization of possible rearrangements using Mauve pipeline: H37Rv, L6 reference genome and the 3 Pacbio L5 (complete) genomes

Table S1. Gene difference within the three PacBio L5 complete genomes from Benin, Nigeria or The Gambia, and between these three genomes and *M. tuberculosis* H37Rv

Genes	Present in PcbL5Ben	Present in PcbL5Gbia	Present in PcbL5Nig	Present in H37Rv
Unshared	PcbL5Ben_01893, PcbL5Ben_01894, PcbL5Ben_01895, PcbL5Ben_02043, PcbL5Ben_03043	PcbL5Gbia_02028, PcbL5Gbia_03020	PcbL5Nig_02001	**H37Rv_02002, H37Rv_02085, H37Rv_02086, H37Rv_02087, H37Rv_02107, H37Rv_02108, H37Rv_02109, H37Rv_02188, H37Rv_02189**
Absent in PcbL5Ben		PcbL5Gbia_unshared	PcbL5Nig_unshared	H37Rv_unshared
Suspected pseudogenes in PcbL5Ben		PcbL5Gbia_00036, PcbL5Gbia_00955, PcbL5Gbia_02045, PcbL5Gbia_03308, PcbL5Gbia_03376	PcbL5Nig_00011, PcbL5Nig_00036, PcbL5Nig_00953, PcbL5Nig_01794, PcbL5Nig_01853, PcbL5Nig_01896, PcbL5Nig_02018, PcbL5Nig_02119, PcbL5Nig_02291, PcbL5Nig_02467	H37Rv_00012, H37Rv_00036, H37Rv_00528, H37Rv_01371, H37Rv_01801, H37Rv_02020, H37Rv_02463, H37Rv_02556, H37Rv_02951, H37Rv_03349, H37Rv_03797

Absent in PcbL5Gbia	PcbL5Ben_01443, PcbL5Ben_01444, PcbL5Ben_unshared	PcbL5Gbia_unshared	PcbL5Nig_unshared	H37Rv_01413, H37Rv_01414, H37Rv_unshared
Suspected pseudogenes in PcbL5Gbia	PcbL5Ben_01064, PcbL5Ben_02378, PcbL5Ben_02486, PcbL5Ben_03113		PcbL5Nig_00011, PcbL5Nig_01794, PcbL5Nig_01853, PcbL5Nig_01896, PcbL5Nig_02119, PcbL5Nig_02291, PcbL5Nig_02467	H37Rv_00012, H37Rv_00528, H37Rv_01371, H37Rv_01801, H37Rv_02329, H37Rv_02463, H37Rv_02556, H37Rv_02951, H37Rv_03797,
Absent in PcbL5Nig	PcbL5Ben_01617, PcbL5Ben_01618, PcbL5Ben_01619, PcbL5Ben_01620, PcbL5Ben_01621, PcbL5Ben_01622, PcbL5Ben_01623, PcbL5Ben_01624, PcbL5Ben_01625, PcbL5Ben_01626, PcbL5Ben_01627, PcbL5Ben_01628, PcbL5Ben_01629, PcbL5Ben_01630, PcbL5Ben_01631, PcbL5Ben_01632,	PcbL5Gbia_01604, PcbL5Gbia_01605, PcbL5Gbia_01606, PcbL5Gbia_01607, PcbL5Gbia_01608, pcbL5Gbia_01609, PcbL5Gbia_01610, PcbL5Gbia_01611, PcbL5Gbia_01612, pcbL5Gbia_01613, PcbL5Gbia_01614, pcbL5Gbia_01615, PcbL5Gbia_01616, PcbL5Gbia_01617, PcbL5Gbia_01618, pcbL5Gbia_01619,		H37Rv_01581, H37Rv_01582, H37Rv_01583, H37Rv_01584, H37Rv_01585, H37Rv_01586, H37Rv_01587, H37Rv_01588, H37Rv_01589, H37Rv_01590, H37Rv_01591, H37Rv_01592, H37Rv_01593, H37Rv_01594, H37Rv_01595, H37Rv_01596,

	PcbL5Ben_01633, PcbL5Ben_01634,	PcbL5Gbia_01620, PcbL5Gbia_01621,	H37Rv_01597, H37Rv_01598,
	PcbL5Ben_01636, PcbL5Ben_01637, PcbL5Ben_01638, PcbL5Ben_01639, PcbL5Ben_01640, PcbL5Ben_01641, PcbL5Ben_01642, PcbL5Ben_01643, PcbL5Ben_01644, PcbL5Ben_01645, PcbL5Ben_01646, PcbL5Ben_01647,	PcbL5Gbia_01623, PcbL5Gbia_01624, PcbL5Gbia_01625, pcbL5Gbia_01626, PcbL5Gbia_01627, PcbL5Gbia_01628, PcbL5Gbia_01629, PcbL5Gbia_01630, PcbL5Gbia_01631, PcbL5Gbia_01632, pcbL5Gbia_01633, PcbL5Gbia_01634,	H37Rv_01600, H37Rv_01601, H37Rv_01602, H37Rv_01603, H37Rv_01604, H37Rv_01605, H37Rv_01606, H37Rv_01607, H37Rv_01608, H37Rv_01609, H37Rv_01610,
	PcbL5Ben_01891, PcbL5Ben_01892, PcbL5Ben_02248, PcbL5Ben_unshared	PcbL5Gbia_01881, PcbL5Gbia_01882, PcbL5Gbia_02228, PcbL5Gbia_unshared	H37Rv_01853, H37Rv_01854, H37Rv_02202 H37Rv_unshared
Suspected pseudogenes in PcbL5Nig	PcbL5Ben_01064, PcbL5Ben_01718, PcbL5Ben_02486	PcbL5Gbia_01705, PcbL5Gbia_03308, PcbL5Gbiaa_03376	H37Rv_00528, H37Rv_01371, H37Rv_01679, H37Rv_02463, H37Rv_02556, H37Rv_02951, H37Rv_03349, H37Rv_03797

Absent in H37Rv	**PcbL5Ben_01364, PcbL5Ben_01365, PcbL5Ben_02129, PcbL5Ben_02130, PcbL5Ben_02181, PcbL5Ben_02182, PcbL5Ben_03555, PcbL5Ben_03556, PcbL5Ben_03557, PcbL5Ben_03669** PcbL5Ben_unshared	**PcbL5Gbia_01359, PcbL5Gbia_01360, PcbL5Gbiaa_02112, PcbL5Gbia_002113, PcbL5Gbia_02163, PcbL5Gbiaa_02164, PcbL5Gbia_03527, PcbL5Gbia_03528, PcbL5Gbiaa_03529, PcbL5Gbia_03643**, PcbL5Gbia_unshared	**PcbL5Nig_01357, PcbL5Nig_01358, PcbL5Nig_02085, PcbL5Nig_02086, PcbL5Nig_02137, PcbL5Nig_02138, PcbL5Nig_03500, PcbL5Nig_03503, PcbL5Nig_03504, PcbL5Nig_03616,** PcbL5Nig_unshared
Suspected pseudogenes in H37Rv	PcbL5Ben_00749, PcbL5Ben_00828, PcbL5Ben_01064, PcbL5Ben_01108, PcbL5Ben_02128, PcbL5Ben_02486, PcbL5Ben_03113, PcbL5Ben_03141, PcbL5Ben_03667	PcbL5Gbia_00750, PcbL5Gbia_00826, PcbL5Gbiaa_00955, PcbL5Gbia_01103, PcbL5Gbia_01879, PcbL5Gbiaa_02111, PcbL5Gbia_03117, PcbL5Gbia_03308, PcbL5Gbiaa_03641	PcbL5Nig_00748, PcbL5Nig_00824, PcbL5Nig_00953, PcbL5Nig_01100, PcbL5Nig_01846, PcbL5Nig_01853, PcbL5Nig_01896, PcbL5Nig_02084, PcbL5Nig_02119, PcbL5Nig_02291, PcbL5Nig_02467, PcbL5Nig_03064, PcbL5Nig_03092, PcbL5Nig_03614

Table S2. Checking whether genes in the PacBio L5 complete genomes but suspected pseudogenes in *M. tuberculosis* H37Rv genome are really pseudogenes.

Gene	Gene in PcbL5Ben	Gene in PcbL5Gbia	Gene in PcbL5Nig	Length (aa)	Locus in H37Rv	Remarks	Conclusion
1	**PcbL5Ben_00749**	**PcbL5Gbia_00750**	**PcbL5Nig_00748**	101	798857-799150	M1V, L96p, H97t, del NCD (aa 99-101)	**Pseudogene in H37Rv**
2	PcbL5Ben_00828	PcbL5Gbia_00826	PcbL5Nig_00824	34	863155-863256	No change	Not pseudogene in H37Rv
3	Not suspected pseudogene	PcbL5Gbia_00955	PcbL5Nig_00953	204	993212-992602	Inverted, M1V	**Pseudogene in H37Rv**
4	PcbL5Ben_01064	Not suspected pseudogene	Not suspected pseudogene	34	1102546-1102647	S23r	Not pseudogene in H37Rv
5	**PcbL5Ben_01108**	**PcbL5Gbia_01103**	**PcbL5Nig_01100**	70	1148218-1148397	M1V	**Pseudogene in H37Rv**
6	**PcbL5Ben_01889**	**PcbL5Gbia_01879**	**PcbL5Nig_01846**	68	1981341-1981341	M1V	**Pseudogene in H37Rv**
7	**PcbL5Ben_02128**	**PcbL5Gbia_02111**	**PcbL5Nig_02084**	92	2219418-2219318	Inverted, alignment showed aa 59-92, no mutation	**Pseudogene in H37Rv**
8	Not suspected pseudogene	Not suspected pseudogene	PcbL5Nig_01853 is PcbNig_1896	41	2038813-2038706	Inverted, alignment showed aa 59-92, no mutation	**Pseudogene in H37Rv**

9	PcbL5Ben_02486	Not suspected pseudogene	Not suspected pseudogene	41	2582440-2582318	Inverted, M1V	**Pseudogene in H37Rv**
10	**PcbL5Ben_03141**	**PcbL5Gbia_03117**	**PcbL5Nig_03092**	111	3291532-3291350	Inverted, del aa 1-aa60 (60 first aa)	**Pseudogene in H37Rv**
11	Not suspected pseudogene	Not suspected pseudogene	PcbL5Nig_02119	46	2256620-2256483	Inverted, No change	**Pseudogene in H37Rv**
12	Not suspected pseudogene	Not suspected pseudogene	PcbL5Nig_02291	33	24345567-2434665	No change	Not pseudogene in H37Rv
13	Not suspected pseudogene	Not suspected pseudogene	PcbL5Nig_02467	94	2614352-2614071	Inverted	**Pseudogene in H37Rv**
14	PcbL5Ben_03113	Not suspected pseudogene	PcbL5Nig_03060	204	3232652-3232867	Del aa1-aa56, N74D, del aa129-aa204	**Pseudogene in H37Rv**
15	Not suspected pseudogene	PcbL5Gbia_03308	Not suspected pseudogene	58	3487871-3487698	Inverted, M1V	**Pseudogene in H37Rv**
16	**PcbL5Ben_03667**	**PcbL5Gbia_03641**	**PcbL5Nig_03614**	99	3841840-3842109	M1V, all aa mutated (insertion substitution) from aa26-aa81, del aa82-aa98	**Pseudogene in H37Rv**

Table S3. Checking whether genes present in H37Rv but suspected pseudogenes in the 3 Pacbio L5 are really pseudogenes

Gene in H37Rv	Length (aa)	Locus in H37Rv	Locus in PcbL5Ben	Locus in PcbL5Gbia	Locus in PcbL5Nig	Remarks	Conclusion
H37Rv_02463	42	2604950-2605078	2631987-2632112	2614961-2615086	2603370-2603495	No change	Not pseudogene
H37Rv_02556	67	2718844-2719047	2743020-2742820	2729937-2729737	2716988-2716788	Inverted, nSNP I66L	**Pseudogene in the 3 Pacbio L5**
H37Rv_02951	55	3115741-3115908	3137926-3137762	3126133-3125969	3112866-3112702	Inverted, nSNP M1V, L55F	**Pseudogene in the 3 Pacbio L5**
H37Rv_03797	200	4053036-4053638	4076665-4077105	4064716-4065156	4055914-4056354	del aa 1-53	**Pseudogene in the 3 Pacbio L5**

Table S4. Genes absent in the Nigerian complete PacBio genome (PcbL5Nig) and some Illumina L5 genomes (n=6), but present in the Benin and Gambian PacBio genomes (PcbL5Ben, PcbL5Gbia), *M. tuberculosis* H37Rv and all other Illumina L5 genomes.

ID (this study)	Rv number	Functional group (Mycobrowser)	Gene name (Mycobrowser)	Product (Mycobrowser)	Function (Mycobrowser)
H37Rv_01581	*Rv1493*	Lipid metabolism	*mutB*	Probable methylmalonyl-CoA mutase large subunit MutB (MCM)	Involved in propionic acid fermentation
H37Rv_01582	*Rv1494*	Virulence, detoxification, adaptation	*mazE4*	Possible antitoxin MazE4	Possible mazE4, antitoxin, part of toxin-antitoxin (TA) operon with Rv1495. Non-essential gene for in vitro growth of H37Rv
H37Rv_01583	*Rv1495*	Virulence, detoxification, adaptation	*mazF4*	Possible toxin MazF4	Sequence-specific mRNA cleavage
H37Rv_01584	*Rv1496*	Cell wall and cell processes	*Rv1496*	Possible transport system kinase	Possibly involved in transport (possibly arginine)
H37Rv_01585	*Rv1497*	Intermediary metabolism and respiration	*lipL*	Probable esterase LipL	Function unknown, but supposed involvement in lipid metabolism
H37Rv_01586	*Rv1498A*	Conserved hypotheticals	*Rv1498A*	Conserved protein	Function unknown
H37Rv_01587	*Rv1498c*	Intermediary metabolism and respiration	*Rv1498c*	Probable methyltransferase	Causes methylation
H37Rv_01588	*Rv1499*	Conserved hypotheticals	*Rv1499*	Hypothetical protein	Function unknown

H37Rv_01589	*Rv1500*	Intermediary metabolism and respiration	*Rv1500*	Probable glycosyltransferase	Function unknown
H37Rv_01590	*Rv1501*	Conserved hypotheticals	*Rv1501*	Conserved hypothetical protein	Function unknown
H37Rv_01591	*Rv1502*	Unknown	*Rv1502*	Hypothetical protein	Function unknown
H37Rv_01592	*Rv1505c*	Conserved hypotheticals	*Rv1505c*	Conserved hypothetical protein	Function unknown. It has some similarity to hypothetical proteins and glycosylases
H37Rv_01593	*Rv1506c*	Unknown	*Rv1506c*	Hypothetical protein	Function unknown. Non-essential gene for in vitro growth of H37Rv
H37Rv_01594	*Rv1507A*				
H37Rv_01595	*Rv1507c*	Conserved hypotheticals	*Rv1507A*	Hypothetical protein	Function unknown
H37Rv_01596	*Rv1508A*	Conserved hypotheticals	*Rv1508A*	Conserved hypotheticals	Function unknown. Highly similar to central part of glycosyl transferases from various mycobacteria and eubacteria
H37Rv_01597	*Rv1508c*	Cell wall and cell processes	*Rv1508c*	Probable membrane protein	Function unknown. Predicted to be in the GT-C superfamily of glycosyltransferases
H37Rv_01598	*Rv1509*	Unknown	*Rv1509*	Hypothetical protein	Function unknown
H37Rv_01599	*Rv1510* (present in PcbL5Nig but missing in the	Cell wall and cell processes	*Rv1510*	Probable conserved membrane protein	Function unknown

	PcbL5Nig-like Illumina L5)				
H37Rv_01600	*Rv1511*	Intermediary metabolism and respiration	*gmdA*	GDP-D-mannose dehydratase GmdA (GDP-mannose 4,6 dehydratase)	Function unknown, probably involved in nucleotide-sugar metabolism
H37Rv_01601	*Rv1512*	Intermediary metabolism and respiration	*epiA*	Probable nucleotide-sugar epimerase EpiA	Function unknown, probably involved in nucleotide-sugar metabolism
H37Rv_01602	*Rv1513*	Conserved hypotheticals	*Rv1513*	Conserved protein	Function unknown, similar to hypothetical proteins from several organisms
H37Rv_01603	*Rv1514c*	Conserved hypotheticals	*Rv1514c*	Conserved hypothetical protein	Function unknown; similar to other hypothetical protein and to putative colanic acid biosynthesis glycosyl transferase
H37Rv_01604	*Rv1515c*	Conserved hypotheticals	*Rv1515c*	Conserved hypothetical protein	Function unknown
H37Rv_01605	*Rv1516c*	Intermediary metabolism and respiration	*Rv1516c*	Probable sugar transferase	Function unknown, involved in cellular metabolism. Non essential for in vitro growth.
H37Rv_01606	*Rv1517*	Cell wall and cell processes	*Rv1517*	Conserved hypothetical membrane protein	Function unknown
H37Rv_01607	*Rv1518*	Conserved hypotheticals	*Rv1518*	Conserved hypothetical protein	possibly glycosyl transferase involved in exopolysaccharide synthesis, similar to several hypothetical proteins and glycosyl transferases from diverse organisms

H37Rv_01608	*Rv1519*	Conserved hypotheticals	*Rv1519*	Conserved hypothetical protein	Function unknown
H37Rv_01609	*Rv1520*	Intermediary metabolism and respiration	*Rv1520*	Probable sugar transferase	Funtion unknown; thought to be involved in cellular metabolism
H37Rv_01610	*Rv1521*	Lipid metabolism	*fadD25*	Probable fatty-acid-AMP ligase FadD25 (fatty-acid-AMP synthetase) (fatty-acid-AMP synthase)	Function unknown; involved in lipid degradation
	Rv1522c (present in PcbL5Nig but missing in the PcbL5Nig-like Illumina L5)	Cell wall and cell processes	*mmpL12*	Probable conserved transmembrane transport protein MmpL12	Function unknown. Thought to be involved in fatty acid transport.

Chapter 9

General discussion and conclusions

This thesis aimed to increase our understanding of *M. tuberculosis* West African 1 (Lineage 5) epidemiology and genomic characteristics, and identified numerous novel associations that shed new light on this distinct *M. tuberculosis* complex member. In addition, in the process of conducting the cohort study, we identified technical advances for improved unbiased diagnosis of TB and related molecular epidemiological studies.

The nationwide genetic diversity of the MTBC in Benin was first determined retrospectively using stored culture isolates from previously-treated TB patients (**Chapter 6**). We then investigated whether growth in culture is lineage dependent, and compared spoligotyping directly applied on sputum (direct) to spoligotyping applied on respective positive cultures (indirect). We found that L5 is under-represented in positive cultures (**Chapter 5**). We therefore used direct spoligotyping for lineage determination for the prospective nationwide population structure study. These results are presented in **Chapter 7** and include the clinical implications of strain differences. Besides the identified culture bias, we also identified decreased performance of the rapid MPT64 antigen identification test for the detection of L5 as an MTBC member in positive cultures (**Chapter 4**). Comparative genomics was used to find the difference in gene content between L5 and the currently used reference genome H37Rv (L4), and the implications of these differences for genome sequence analysis and phenotypic characteristics (**Chapter 8**). Regarding sputum shipment to remote reference laboratories for surveys/studies or further routine diagnostic analyses, we compared CPC and OMNIgene.SPUTUM for short- and long-term storage of sputum for subsequent culture (**Chapter 2**). CPC, OMNIgene.SPUTUM and ETOH were compared for one-month storage of sputum at room temperature for subsequent molecular TB diagnostic testing

and we provided an algorithm to help the user in the choice of the storage reagent depending on the samples' bacillary burden and planned types of molecular- and/or phenotypic tests (**Chapter 3**). The methods evaluated in this thesis, along with the methods used as reference for the comparisons, are presented in Table 1.

We found that ancestral lineages - including L5 - are less likely to grow in culture (**Chapter 5**) [1]. Thus, the selective pressure of culture, resulting in false-negative cultures, introduces a bias in culture-based estimations of bacterial prevalence such as molecular epidemiological studies and drug-resistance surveys. For molecular epidemiological studies, our findings clearly showed the difference in lineage distribution when using sputum-based genotyping versus culture-based genotyping (**Chapter 5**).

For drug-resistance surveys using culture-based techniques (pDST, indirect WGS), the specimens with a false-negative culture would be missed. As a result, the drug-resistance trends would be based on specimens with positive culture only; this likely results in biased estimates and trends, as was observed in our nationwide study. For previously-treated patients with L5 strains, the prevalence of RIF-resistance was 8.6% (5/58 sputa) using the sputum-based Xpert, while it was higher using the culture-based pDST (17.7%, 3/17 cultures), leading to overestimation of the resistance burden in this lineage. The same was observed for L6 strains (among new patients, 1.4% RIF-resistance using Xpert vs 4.4% using pDST), whereas L4 was less affected (among previously-treated patients, 2.7% RIF-resistance using Xpert vs 4.4% using pDST), similar to other lineages. This showed that a sputum-based drug resistance survey (DRS) is preferable over culture-based DRS in *M. africanum* endemic settings. Furthermore, the culture bias suggests that treatment monitoring of patients

(on 2nd line treatment), would be less informative when culture is negative in settings where *M. africanum* L5 and L6 are endemic. To make matters worse, even when specimens containing *M. africanum* (L5 and L6) are culture-positive, the isolates are mostly of dysgonic quality (also confirmed on the larger dataset from the nationwide molecular epidemiology study **Chapter 7**), and are disproportionally false negative on the rapid MPT64 identification test, possibly due to a non-synonymous SNP in the *mpt64* gene across all L5 isolates (**Chapter 4**). Dysgonic isolates complicate successful culture-based analyses (pDST, WGS), even after tentative recovery through subculture. All those observations underline the urgent need to optimize diagnostics directly applied to sputum for all purposes (diagnosis, drug resistance detection, strain typing and identification). Next generation sequencing, especially WGS [2], is a good solution to yield results for all such analyses, including for a more extensive screen for drug resistance mutations than the drugs usually tested in pDST. It can also provide relevant information for individualized treatment modifications for patients in whom a standard treatment regimen may not be optimal [3]. However, WGS is to date still culture based rather than from clinical samples, although encouraging developments suggest that direct genome sequencing is becoming feasible. Also, the genotypic prediction of phenotypic resistance is most reliable for 1st line drugs, and international consortia, such as CRYpTIC and ReSeqTB are still working to determine the correlation between phenotypic resistance and genotypic mutations in 2nd line and new TB drugs[4,5]. Especially for new drugs, like bedaquiline and delamanid, the multiplicity of genes, and the relatively high diversity of mutations observed in these genes- with incomplete data on their association with phenotypic resistance and clinical relevance- complicate gDST for these drugs. Therefore, WHO reiterates the need for urgent capacity strengthening for phenotypic DST

for new drugs [6], which compounds the culture challenges mentioned above in L5- and L6 endemic areas.

Although culture and pDST can take a long time to yield a result, it remains useful to optimize current mycobacterial culture media for the growth of *M. africanum*. Despite the use of Löwenstein-Jensen medium supplemented with pyruvate in our study (beneficial for the growth of *M. africanum* and *M. bovis* [7]) and a longer incubation time (13 weeks as recommended by Castets for the growth of *M. africanum* [8] versus the usual 8 weeks), *M. africanum* and *M. bovis* were less likely to grow in culture, with mostly dysgonic isolates. The reduced growth in culture was also observed in The Gambia when using MGIT for primary isolation (9 of 14 MGIT negative sputa contained L6 strains (64.3%) identified by spoligotype analysis of the sputum sediments); PhD thesis Boatema Ofori-Anyinam, University of Antwerp, 2017). The possible benefit, if any, of an extended incubation time for MGIT (like the extended incubation used for LJ in this study) and/or an optimization of these media for growth of *M. africanum* (L5 and L6) warrants further evaluation. Identifying the causes of the reduced growth, would facilitate such optimization. The supplementation of mycobacterial culture medium with vitamin B6, which is encoded by the *Rv2074* gene that we found missing in most *M. africanum* and *M. bovis,* could be a first approach for such optimization.

Xpert MTB/RIF (Cepheid, USA), which is currently used for initial diagnosis of TB, only determines resistance to RIF. However, Xpert -as all nucleic acid amplification techniques- cannot differentiate between live and dead bacilli, which is problematic when the highly sensitive new generation Xpert MTB/RIF Ultra (Cepheid, USA) is applied for the diagnosis of TB in previously-treated patients [9], as they can continue to shed MTBC DNA from dead bacilli many

years after they are cured [10]. We found that L5 was over-represented among AFB-scanty sputa from failure patients admitted for subsequent TB treatment (50% vs 12.1% for L4). This may imply that those L5 specimens contained remaining dead bacilli and that some of these patients were falsely diagnosed as treatment failure patients by microscopy, although in Benin two microscopy-positive sputa are required to confirm a treatment failure, and patients' sputum is cultured before starting a retreatment regimen. Although this hypothesis needs to be investigated, its confirmation would further underline the need to distinguish dead from live bacilli, especially in L5-endemic settings. So, an ideal solution could be a nucleic acid amplification/sequencing technique that can distinguish between dead and live TB bacilli. Some techniques, like the use of propidium-monoazide, were previously described, but are not widely used [11,12]. Also, RNA-based biomarkers could allow for detection of live bacilli, as RNA degrades much faster than DNA, once bacteria die. In the frame of the EDCTP-funded DIAMA project, coordinated from Benin, RIF-resistant patients' treatment monitoring using a pre-RNA/mRNA approach is being evaluated in four African countries [13].

We also found that in specimens with mixed infection of ancestral and modern lineages, the modern lineage outgrew the ancestral ones in culture (**Chapter 5**). So, a nucleic acid amplification/sequencing alternative to culture should also be able to detect minority populations within mixed infections, both mixtures of different or of the same lineages.

The lower sensitivity of the MPT64 antigen identification test to detect L5 as a member of the MTBC in positive cultures (**Chapter 4**)[14] was also reported for L6 [15]. Four MPT64 tests exist and include the WHO recommended Capilia TB-Neo (Tauns Laboratories, Numazu, Japan), TB Ag MPT64 Rapid Test (SD Bioline,

Kyonggi-do, South Korea), BD MGIT TBcID (Becton Dickinson Microbiology Systems, Sparks, USA) [6], and the TBCheck MPT64 assay (Hain Lifescience). In this thesis we evaluated the TB Ag MPT64 Rapid Test (SD Bioline, Kyonggi-do, South Korea) (**chapter 4**). The MPT64 test BD MGIT TBcID was previously compared to the TB Ag MPT64 Rapid Test (SD Bioline) for the differentiation of L6 from NTM, and no difference was found regarding their performance for L6 [15]. As L5 has a non-synonymous SNP in the *mpt64* gene, the performance of an MPT64 test for L5 will depend on whether or not the epitopes of the MPT64 protein used in the MPT64 assay cover the modified amino-acid region. For L6, the reason for the lower sensitivity of MPT64 tests is the lower expression of the *mpt64* gene relative to *M. tuberculosis sensu stricto* [15], which could also be an additional reason for the lower sensitivity for L5. This however was not investigated in our study. Consequently, in MAF (L5, L6) endemic regions, for whatever MPT64 test is used, negative results should be confirmed by another identification method such as the Genotype CM, IS*6110* PCR, spoligotyping, Xpert on diluted bacterial suspension or the phenotypic PNB/catalase tests. Nonetheless, MTBC diagnostics development, whether phenotypic or genotypic, should take into account their performance on geographically restricted lineages including L5, L6, L7, and the newly discovered L0 and L8 (Ngabonziza et al, in preparation; Coscolla et al, in preparation).

Specimens with a spoligotype-pattern difficult to assign to a lineage (after an experienced knowledgeable person had verified spoligotype signatures for lineage assignment done by online lineage assignment databases), were tested with PhyloSNP analysis (PCR followed by sequencing for detection of lineage-specific SNPs in Lineages 1 through -6, **Chapter 7**). Applying PhyloSNP analysis helped in assigning the lineage to isolates with difficult to interpret spoligo-

patterns, including mixed infections (together constituting 1% of specimens). Therefore, while waiting for the development/optimization of new (sputum-based) techniques (WGS or other) capable of solving the culture bias and identification limitations, “direct spolygotyping” (genotyping) supplemented with “direct PhyloSNP” can be used for strain typing directly from sputum for unbiased population structure determination. For now, some NGS-based genotypic techniques such as Deeplex®Myc-TB (GenoScreen, France) [13,16] are being evaluated for their performance for drug resistance detection and strain typing, directly from sputum. Transmission studies are more complicated, as higher resolution typing techniques like WGS are required, which is not yet consistently feasible directly from sputum. Ideally, an accurate transmission study in settings where *M. africanum* is endemic would be based on “direct WGS”, with as large a sampling fraction (geographically and over time) as possible to avoid missing links in the chains of transmission.

All in all, there is a need to revise current algorithms of TB diagnosis and treatment monitoring (of patients with RIF-resistant strains) in *M. africanum*, especially L5, endemic countries.

Our findings show that, in Benin, new and previously-treated TB patients differ in terms of age, sex, treatment outcome, and lineage distribution (**Chapter 7**). Studies on the molecular epidemiology of TB need to distinguish the patient’s TB treatment history. Another important aspect is whether the population selected for the study is representative. Our findings showed that lineage distribution varies greatly across regions within the same country, and can be quite variable within a region (**Chapter 7**). We thus recommend, for molecular epidemiological studies, a proportional (to the number of TB patients diagnosed in routine care) sampling of the regions within the study area that

includes sufficient new and previously-treated patients for a stratified analysis. Depending on the purpose (diagnostic results with treatment implications, versus for research findings such as a drug resistance survey or molecular epidemiological study), and depending on the microscopy grade of the specimens and the methods that will be applied, we recommended the suitable reagents for the storage of sputum specimens at ambient temperature for up to one month or further for molecular analyses (**Chapter 2** and **3**). CPC was good for the storage of sputa for one month before culture, while OMNIgene.SPUTUM (commercial reagent) was performant for the storage of sputum for up to 8 days only, with reduced culture positivity after one-month storage [17]. For molecular TB diagnostic testing CPC, OMNIgene and ethanol show good performance, but the latter cannot be used if culture is needed as well. The length of the PCR-target, paucibacillary baseline microscopy, higher ambient temperatures, and reagent costs are also factors to take into account in the choice of sputum storage reagents (**Chapter 3**) [18].

An alternative for molecular diagnostics in settings where Xpert is done as initial testing, is the use of remnant DNA from the Xpert cartridge for further molecular testing such as 2nd line LPAs or spoligotyping when the Xpert bacilli burden is at least medium (Ct ≤24), but it cannot be used for 1st line LPA (MTBDR*plus*) or targeted sequencing of the *gyr*A and *rrs* genes [19]. The aspiration of remnant DNA from the Xpert cartridge requires a dedicated bench and qualified personnel to avoid cross-contamination with Xpert amplicons. When such facility and qualified personnel are not available in a laboratory (for example peripheral or intermediary laboratories where no molecular biology platform exists), Xpert cartridges with at least medium bacilli load could be stored at 2-8°C (if available), and shipped to a reference laboratory.

Our data on sputum storage as well as the algorithm we proposed (**chapter 3**) for the choice of sputum storage reagent, showed that for subsequent long ragment PCR, there is the need to know whether the sputum is AFB-scanty microscopy or not in order to choose the appropriate storage reagent. Where Xpert is used for initial diagnosis, an Xpert result with “low” (Ct 22-28) or “very low” (Ct >28) bacilli burden could be considered as AFB-scanty microscopy [18,20] while the “high” (Ct <16) or “medium” (Ct 16-22) will be considered as “+++, ++, + positive microscopy ”. Following this association, sputum specimens can be stored using the recommended storage reagent in the algorithm we proposed, based on their Xpert result, without the need for smear microscopy.

Our findings show that previously-treated patients had a lower sputum-microscopy grade at baseline than new patients. In addition, among the previously-treated patients, those with L5 strains were more likely to produce paucibacillary (ABF-scanty) sputum compared to those with L4 strains. This could be a further challenge for a “direct WGS” approach in these patient groups, given the limited amount of mycobacterial DNA present in ‘scanty’ specimens.

On the other hand, patients with L5 strains were less likely to have a previous treatment history, mainly due to less TB relapse. However, among previously-treated patients but not in new patients, L5 was significantly associated with higher RIF-resistance rates (either with the sputum-based Xpert or positive-culture based pDST). Coscolla et al. similarly found from the analyses of a multi-country collection of L5 genomes that L5 was associated with resistance, yet the sequencing was culture-based, and it is unclear how those patients were selected from the countries’ TB patient population (Coscolla et al, in preparation). The fact that our findings are based on a prospective study with

proportional sampling representative of new- and previously-treated TB patients in Benin, supports the association of L5 with RIF-resistance. This could be due to acquisition of RIF resistance (selection of resistant mutant) during first-line TB treatment, as most (4/5) of the RIF-resistant L5 strains in our study were identified from patients who failed treatment. An increased propensity to develop RIF resistance is potentially threatening for L5 endemic countries, albeit current numbers are too low and more detailed strain typing is required to draw firm conclusions. Furthermore, we found from L5 genome analysis that all L5 strains miss the *Rv1979c* gene, which is one of the genes containing variants (mutations or (complete) deletion) associated with resistance to clofazimine and bedaquiline [21,22], two key drugs used for the treatment of RIF-resistant TB [23,24]. Therefore, RIF-resistant and MDR-TB treatment that includes clofazimine and bedaquiline might potentially fail in L5 patients, increasing treatment failure rates in L5-endemic countries.

To ensure optimized management of RIF-resistant TB in Benin, we propose the following:

1. Replacement of microscopy by Xpert as initial test (including in peripheral TB clinics) for early diagnosis of RIF-resistant TB
2. Performing direct *rpoB* gene sequencing [25] or repeat Xpert to confirm RIF-resistant cases in this low RIF-resistance prevalence setting
3. Performing direct *rpoB* gene sequencing [25] or repeat Xpert to detect any acquisition of RIF-resistance during the course of treatment in case of continued positive smear microscopy or lack of clinical improvement (at month 2 or 3, for instance), especially for

patients with L5 (if known from baseline direct analysis such as spoligotyping or Deeplex)

4. Expanding patient centered approaches for improved adherence to avoid resistance acquisition (one size does not fit all), including self-administered and DOT (direct observed treatment) based options, such as the choice between community- or home-based DOT or VOT (video observed treatment)[26,27].

The short course (9 month) MDR-TB treatment regimen that has been used in Benin since 2007 yields highly satisfactory treatment outcome results and only needs adjustment if resistance occurs to fluoroquinolones or injectables, which can be detected by LPAs directly on smear-positive sputum or on Xpert DNA remnant [19]. Nevertheless, WGS directly on sputum would be useful for the prediction of resistance to many 2nd line and even new TB drugs, in order to find alternative susceptible drugs to treat patients with resistance or intolerance to 2nd line drugs [23], or another standardized regimen after safety and effectiveness/superiority is proven in clinical trials.

WHO is calling for the use of universal gDST (not only in previously-treated patients) for at least RIF (using Xpert or 1st line LPA) and fluoroquinolones in case RIF resistance is detected (using 2nd line LPA), with the possibility of addressing the potential discrepancies of such new techniques using pDST [6]. WHO is also calling for research on new (next generation sequencing (NGS)-based) gDST techniques for drug resistance diagnosis [23,28]. While gDST can overcome the culture-related problems (like reduced positivity and dysgonic isolates for *M. africanum*), they have their own disadvantages such as:

1. Inability to determine the viability of bacteria; especially needed for TB diagnosis in previously-treated patients, and treatment monitoring of MDR/RIF-resistant patients,
2. Difficulty to implement such techniques in non-reference laboratories where it should be deployed by unskilled personnel, if not available as a low-tech or point-of-care technique
3. Challenge to distill information relevant for patient care by unskilled health care workers from the obtained sequencing data, if data are not analyzed and interpreted in an automated (off-line) way,
4. Involvement of unknown resistance conferring mechanisms (genes) for some 2nd line and new drugs.

On the other hand, gDST can potentially provide data faster and for drugs that are phenotypically not tested in some laboratories. Whole genome sequencing data also comprises information for newly introduced drugs and can contribute research data to unravel unknown drug resistance mechanisms, both for existing and new drugs. Sputum-based WGS is still a challenge and sputum-based targeted NGS (direct deep sequencing) may not yield sufficiently high quality sequences [28]. pDST, although lengthy, remains important to provide susceptibility data – preferably minimal inhibitory concentration data - for less known or new drugs. This requires a positive culture with good growth, which is a challenge in L5 (MAF) settings, due to the false-negative and dysgonic nature of the primary isolation in culture. For now, pDST and direct gDST/NGS-based techniques (targeted or WGS) are complementary approaches and both are needed in L5 (MAF) endemic settings, while encouraging research on the optimization of solid/liquid medium for MAF culture, the understanding of new

drug resistance mechanisms, the optimization of sputum-based gDST/targeted NGS/ WGS, and the development of point-of-care viability tests.

In our study population, L5 was under-represented in previously-treated TB, especially relapse, and overrepresented in female, Gbe ethnicity, and older age patients. Its prevalence is inversely related to the population density of patient residence areas. L5 was however not associated with HIV positivity, the presence of BCG vaccine scar, baseline patient weight and TB treatment outcome. For L6 as well, we did not find any association with HIV positivity, in contrast to the findings in Ghana, the United States and the Gambia [29–31]. This could be due to a lower power of the study for L6 analysis, as neither HIV co-infection nor L6 are common in Benin.

L5 prevalence increases with the decrease of population density while L6 prevalence increases with the increase of population density. Both L5 and L6 were associated with specific ethnic groups; Peulh (Fulani) and Bariba for L6, and Gbe ethnic group (especially Eastern-Gbe) for L5.

To summarize, from a programmatic view, the implications of strain differentiation in L5 (broadly MAF) endemic countries include:

1. Biased culture-based diagnosis due to false-negative culture results for L5, L6 and *M. bovis* containing specimens, especially for extra-pulmonary TB [23], where TB diagnosis is mostly based on culture and where MAF is overrepresented [31].
2. Likewise, false-negative cultures lead to unavailability of pDST results for 2nd line drugs and lack of possibility to phenotypically check genotypically detected RIF and/or INH resistance (via Xpert MTB/IRF and/or 1st line LPA)[6] and test pDST for new drugs.

3. Falsely diagnosing TB as NTM due to the use of the rapid MPT64 Antigen test to differentiate MTBC from NTM in positive cultures. Consequently, the patient with a missed TB diagnosis will be at risk of inappropriate treatment modifications, including poor outcomes and further transmission of TB in the community.
4. Biased MDR/RIF-resistant treatment monitoring by culture, due to falsely negative cultures. Conversely, there may also be false treatment failure defined using classical microscopy during treatment monitoring among RIF-sensitive L5 patients, due to the ongoing secretion of dead bacilli. The proportion of patients "failing" at 5M (month 5) who are culture negative and may not need further treatment needs to be confirmed in further studies.
5. Possibility of increased resistance acquisition by L5 strains during treatment with 1st line drugs (RIF-sensitive regimen), although this needs to be confirmed in further studies. If confirmed, patients with L5 strains may require more intense diagnostic monitoring for (acquired) RIF-resistance.

Comparison of chest X-ray profiles of TB patients with L5 or L6 strains may differ from those with other strains: given their propensity for extrapulmonary disease and/or microaerophilic growth, there may be less predominance of apical lesions, where oxygen tension is highest.

In conclusion, the algorithm for TB diagnostics testing in *M. africanum* (especially L5) endemic countries could be improved, favoring direct molecular diagnostics for unbiased results. A general lesson that can be drawn from work in this thesis and by others, is that the performance of diagnostic tests should be validated in a broad variety of settings – including those with geographically

limited lineages - ideally before implementation for patient diagnosis/care, or as post-marketing validation. The reduced growth of L5 strains in culture could partly be explained by the absence - in the vast majority of L5 strains - of genes implicated in bacterial survival and in vitro growth (mainly *Rv1994c*). Efforts should be increased to advance the quality of "direct WGS". In the meantime, population structure studies could be based on "direct genotyping" (spoligotyping), followed by "direct PhyloSNP" analysis for uncommon patterns. The use of an L5-specific reference genome may help in sub-lineage classification and the determination of L5 sub-lineage distribution across West African L5 countries to better understand the L5 origin. A reference free, *de novo* assembly approach is preferable over the currently used reference genome H37Rv for genome analyses. The distribution of lineages is significantly different in new versus previously-treated patients. L5 is less likely to cause TB relapse, yet is possibly associated with the acquisition of RIF-resistance, while genomic predictions showed that it is also possibly associated with resistance to clofazimine and bedaquiline, included among the drugs used to treat RIF-resistant/MDR-TB. These findings should be confirmed with phenotypic MIC determinations. Solutions should be implemented to avoid this potential threat to TB treatment in L5-endemic countries. The suggested lower virulence (immunogenicity, transmissibility) of *M. africanum* [32,33] may be partly explained in L5 by the absence of genes associated with bacterial survival in macrophages (*Rv1978),* in stress conditions (high temperature, starvation, *Rv1977)* and during the chronic phase of infection *(Rv1994c, Rv1993c),* and genes associated with immune-evasion and virulence (*Rv2074*, vitamin B6: pyridoxine). In Cotonou (Benin), within 10 years, the L5 prevalence significantly declined by 9.4 % (95% CI:-17.6 to -1.2) as well as L1, while L4 prevalence increased by 16% (95% CI: 7.4 to 24.6). In Benin, the L5 and L6 geographical

distribution are not driven by the same factors. While L5 prevalence increases with decreasing host population density, surprisingly L6 prevalence increases with increasing population density, after controlling for patient ethnicity and residence region. L5 and L6 distribution is explained both by specific population density and specific ethnicities. Future studies on the dynamics of L5 and L6 transmission may include analysis of sympatric versus allopatric transmission [34,35]. Extending such studies to other countries in the region may form the basis for studies on host-pathogen co-evolution, including potential non-human reservoirs of L6, and on the likely environment in which the Most Recent Common Ancestor of ancestral MTBc lineages evolved in human- specific lineages L5 and L6 versus animal-specific lineages such as *M. bovis*.

Table 1. Methods evaluated in the thesis and methods used as reference

Purpose	**Method evaluated**	**Reference used**	**Conclusion (relative to reference used)#**	**Chapter**
Specimen storage				
Sputum storage for mycobacterial culture: **contamination rate**	CPC 8 CPC 28 OMNI 8 OMNI 28	Fresh sputum*	Not different Not different **Lower contamination** **Lower contamination**	**2**
	CPC 28 OMNI 8 OMNI 28	CPC 8	Not different Not different Not different	**2**
	OMNI 8 OMNI 28	CPC 28	Not different Not different	**2**
	OMNI 28	OMNI 8	Not different	**2**
Sputum storage for mycobacterial culture: **culture positivity rate**	CPC 8 CPC 28 OMNI 8 OMNI 28	Fresh sputum*	Not different Not different Not different **Lower positivity rate**	**2**
	CPC 28 OMNI 8	CPC 8	**Lower positivity rate** Not different	**2**

	OMNI 28		**Lower positivity rate**	
	OMNI 8 OMNI 28	CPC 28	Not different **Lower positivity rate**	**2**
	OMNI 28	OMNI 8	**Lower positivity rate**	**2**
28-day storage of sputum for Xpert MTB/RIF (short fragment PCR): **positivity**	CPC 28 OMNI 28	ETOH 28	Not different Not different	**3**
	OMNI 28	CPC 28	Not different	**3**
28-day of sputum for 1764 bp nested *rpoB* PCR (long fragment): **positivity**	CPC 28 OMNI 28	ETOH 28	Not different **Higher positivity rate**	**3**
	OMNI 28	CPC 28	Not different	**3**
Storage of **AFB-scanty** microscopy sputum for Xpert MTB/RIF (short fragment PCR): **positivity**	CPC 28 OMNI 28	ETOH 28	Not different Not different	**3**
	OMNI 28	CPC 28	Not different	**3**
Storage of **AFB-scanty** microscopy sputum for 1764 bp nested *rpoB* PCR (long fragment): **positivity**	CPC 28 OMNI 28	ETOH 28	**Higher positivity rate** **Higher positivity rate**	**3**
	OMNI 28	CPC 28	Not different	**3**
Storage of sputum for Xpert MTB/RIF (short fragment PCR): **Ct value D28**	CPC 28 OMNI 28	ETOH 28	**Higher Ct value (ie lower bacilli load)** **Higher Ct value (ie lower bacilli load)**	**3**
	OMNI 28	CPC 28	Not different	**3**
Storage of sputum for Xpert MTB/RIF (short fragment PCR): **Ct value D28 vs D0**	ETOH 28	Fresh sputum (before ETOH is added)	**Lower Ct value (ie higher bacilli load)**	**3**
	CPC 28	Fresh sputum (before CPC is added)	Not different	**3**

	OMNI 28	Fresh sputum (before OMNI is added)	Not different	**3**
Species identification				
Rapid differentiation of MTBC from NTM in culture isolates	MPT64 Ag test	Combined PNB/catalase + Spoligotyping	L5 isolates (24%) misclassified as NTM	**4**
MTBC strain typing				
Lineage determination	Direct spoligotyping (sputum-based)	Indirect spoligotyping (isolate-based)	Good agreement (95%) **Spoligo-patterns available for nearly all sputa (98.5%) by direct spoligotyping vs 73.4% sputa by indirect spoligotyping**	**5**
Difference in MTBC population structure in negative vs positive culture specimens (gain obtained from **direct spoligotyping**)	Lineage distribution (using **direct spoligotyping**) among culture negative specimens	Lineage distribution (using **direct spoligotyping**) among culture positive specimens	**L5 under-represented in positive culture** specimens	**5**
Lineage determination for uncommon spoligotype patterns	Direct spoligotyping	Direct spoligotyping + direct PhyloSNP for uncommon spoligo-patterns	**Direct PhyloSNP (after spoligotyping)** confirms lineage assignment for **uncommon spoligo-patterns**	**7**
Whole genome analysis				
Genome comparison	L5 complete genomes	H37Rv (L4) reference genome (NC_ 000962.3)	**Different gene content**	**8**
	L5 complete genomes	L6 complete reference genome (GM041182)	**Different gene content**	**8**

	L5 complete genomes	*M. bovis* complete reference genome (LT708304.1)	**Different gene content**	**8**

* Fresh sputum (no additive) cultured within 1 day after collection; CPC 8/28 = samples were processed after 8/28 days' storage in cetylpiridinium choloride; OMNI 8/28 = samples were processed after 8/28 days' storage in OMNIgene.SPUTUM; MTBC = *Mycobacterium tuberculosis* complex; # Conclusions in bold are significantly different ($p<0.05$); not different = not significantly different ($p>0.05$)

References

[1] Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. Genotypic characterization directly applied to sputum improves the detection of Mycobacterium africanum West African 1, under-represented in positive cultures. PLoS Negl Trop Dis 2017;11:e0005900. doi:10.1371/journal.pntd.0005900.

[2] Su M, Satola SW, Read TD. Genome-Based Prediction of Bacterial Antibiotic Resistance. J Clin Microbiol 2019;57:e01405-18. doi:10.1128/JCM.01405-18.

[3] WHO. Global tuberculosis report 2018. 2018.

[4] CRYPTIC consortium. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med 2018;379:1403–15. doi:10.1056/NEJMoa1800474.

[5] Ezewudo M, Borens A, Chiner-Oms Á, Miotto P, Chindelevitch L, Starks AM, et al. Integrating standardized whole genome sequence analysis with a global Mycobacterium tuberculosis antibiotic resistance knowledgebase 2018. doi:10.1038/s41598-018-33731-1.

[6] World Health Organization. GLI Practical Guide to TB Laboratory Strengthening 2017:130.

[7] Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, et al. The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: Implications for in vivo growth. Mol Microbiol 2005;56:163–74. doi:10.1111/j.1365-2958.2005.04524.x.

[8] Castets M. [Mycobacterium africanum (author's transl)]. Med Trop (Mars) 1979;39:145–8.

[9] WHO | Next-generation Xpert® MTB/RIF Ultra assay recommended by WHO. WHO 2017.

[10] Boyles TH, Hughes J, Cox V, Burton R, Meintjes G, Mendelson M. False-positive Xpert® MTB/RIF assays in previously treated patients: need for caution in interpreting results. Int J Tuberc Lung Dis 2014;18:876–8. doi:10.5588/ijtld.13.0853.

[11] Nikolayevskyy V, Miotto P, Pimkina E, Balabanova Y, Kontsevaya I, Ignatyeva O, et al. Utility of propidium monoazide viability assay as a biomarker for a tuberculosis disease. Tuberculosis 2015;95:179–85. doi:10.1016/j.tube.2014.11.005.

[12] Kayigire XA, Friedrich SO, Karinja MN, van der Merwe L, Martinson NA, Diacon AH. Propidium monoazide and Xpert MTB/RIF to quantify Mycobacterium

tuberculosis cells. Tuberculosis 2016;101:79–84. doi:10.1016/j.tube.2016.08.006.

[13] DIAMA consortium. DIAgnostics for Multidrug Resistant Tuberculosis in Africa (DIAMA). 2017. doi:https://clinicaltrials.gov/ct2/show/NCT03303963.

[14] Sanoussi CN, De Jong BC, Odoun M, Arekpa K, Ligali MA, Bodi O, et al. Low sensitivity of the MPT64 identification test to detect lineage 5 of the Mycobacterium tuberculosis complex 2018. doi:10.1099/jmm.0.000846.

[15] Ofori-Anyinam B, Kanuteh F, Agbla SC, Adetifa I, Okoi C, Dolganov G, et al. Impact of the Mycobaterium africanum West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased Sensitivity of Rapid Identification Tests in The Gambia. PLoS Negl Trop Dis 2016;10:1–12. doi:10.1371/journal.pntd.0004801.

[16] From tuberculosis clinical samples to drug resistance profile. n.d.

[17] Affolabi D, Sanoussi N, Sossou A, Nys T, Bodi O, Esse M, et al. Performance of OMNIgene•SPUTUM (DNA Genotek) and cetylpyridinium chloride for sputum storage prior to mycobacterial culture. J Med Microbiol 2018;67:798–805. doi:10.1099/jmm.0.000745.

[18] Sanoussi CN, de Jong BC, Affolabi D, Meehan CJ, Odoun M, Rigouts L. Storage of Sputum in Cetylpyridinium Chloride, OMNIgene.SPUTUM, and Ethanol Is Compatible with Molecular Tuberculosis Diagnostic Testing. J Clin Microbiol 2019;57:e00275-19. doi:10.1128/JCM.00275-19.

[19] Venter R, Derendinger B, De Vos M, Pillay S, Dolby T, Simpson J, et al. Mycobacterial genomic DNA from used Xpert MTB/RIF cartridges can be utilised for accurate second-line genotypic drug susceptibility testing and spoligotyping OPEN n.d. doi:10.1038/s41598-017-14385-x.

[20] Najjingo I, Muttamba W, Kirenga BJ, Nalunjogi J, Bakesiima R, Olweny F, et al. Comparison of GeneXpert cycle threshold values with smear microscopy and culture as a measure of mycobacterial burden in five regional referral hospitals of Uganda- A cross-sectional study. PLoS One 2019;14. doi:10.1371/journal.pone.0216901.

[21] Zhang S, Chen J, Cui P, Shi W, Zhang W, Zhang Y. Identification of novel mutations associated with clofazimine resistance in Mycobacterium tuberculosis 2015. doi:10.1093/jac/dkv150.

[22] Ghodousi A, Hussain Rizvi A, Quadir Baloch A, Ghafoor A, Masood F, Qadir M, et al. Acquisition of cross-resistance to Bedaquiline and Clofazimine following treatment for Tuberculosis in 2 Pakistan 2019. doi:10.1101/630012.

[23] WHO. WHO consolidated guidelines on drug-resistant tuberculosis treatment. 2019.

[24] WHO. THE SHORTER MDR-TB REGIMEN FEATURES OF THE SHORTER MDR-TB REGIMEN REGIMEN COMPOSITION 4-6 Km-Mfx-Pto-Cfz-Z-H high-dose-E / 5 Mfx-Cfz-Z-E. 2016.

[25] Rigouts L, Nolasco O, De Rijk P, Nduwamahoro E, Van Deun A, Ramsay A, et al. Newly developed primers for comprehensive amplification of the rpoB gene and detection of rifampin resistance in Mycobacterium tuberculosis. J Clin Microbiol 2007;45:252–4. doi:10.1128/JCM.01489-06.

[26] Stop TB Partnership. Technologies for improving TB medication adherence & treatment outcomes. n.d. doi:10.5588/Ijtld.12.0080.

[27] TREATMENT OF TUBERCULOSIS Guidelines for treatment of drug-susceptible tuberculosis and patient care. 2017.

[28] The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide 2018. n.d.

[29] Asante-Poku A, Otchere ID, Osei-Wusu S, Sarpong E, Baddoo A, Forson A, et al. Molecular epidemiology of Mycobacterium africanum in Ghana. BMC Infect Dis 2016;16:385. doi:10.1186/s12879-016-1725-6.

[30] de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, et al. Differences between tuberculosis cases infected with *Mycobacterium africanum* , West African type 2, relative to Euro-American *Mycobacterium tuberculosis* : an update. FEMS Immunol Med Microbiol 2010;58:102–5. doi:10.1111/j.1574-695X.2009.00628.x.

[31] Sharma A, Bloss E, Heilig CM, Click ES. Tuberculosis Caused by Mycobacterium africanum, United States, 2004-2013. Emerg Infect Dis 2016;22:396–403. doi:10.3201/eid2203.151505.

[32] Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, Darko Otchere I, et al. Reduced transmission of Mycobacterium africanum compared to Mycobacterium tuberculosis in urban West Africa. Int J Infect Dis 2018;73:30–42. doi:10.1016/j.ijid.2018.05.014.

[33] Coscolla M, Gagneux S. Consequences of genomic diversity in mycobacterium tuberculosis. Semin Immunol 2014;26:431–44. doi:10.1016/j.smim.2014.09.012.

[34] Gagneux S, Deriemer K, Van T, Kato-Maeda M, De Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. 2006.

[35] Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, et al. HIV Infection Disrupts the Sympatric Host-Pathogen Relationship in Human Tuberculosis. PLoS Genet 2013;9. doi:10.1371/journal.pgen.1003318.

Chapter 10

Perspectives: future research

The findings of this thesis showed that *M. africanum* L5, L6 and *M. bovis* are less likely to grow in culture, and when they do, they mostly yield isolates of dysgonic morphology (**Chapters 5, 7 and 4**)[1,2], despite the use of pyruvate-supplemented medium as recommended for their isolation [3]. That culture bias, the reduced performance of dysgonic isolates for downstream culture-based diagnostic testing, and the low sensitivity of some diagnostic tests for *M. africanum* (MPT64 antigen identification test [2], **Chapter 4**), could be overcome by using direct testing on clinical specimens. Direct genotypic techniques for detection, identification, susceptibility testing and strain typing - with possibility of detecting mixed infections - are required. Direct WGS would be the preferred method as it allows both diagnostic testing – including an expanded "resistotype" - and high resolution genomic analyses for transmission studies, offering the possibility of finding transmission links in real time. Integration of direct WGS in a system capable to provide essential results for routine patient care without particular bioinformatics skills would be most useful for TB control, while also allowing use of the whole genome sequence for research purposes. It would be useful to further elucidate the biology of the MTBC and mycobacteria in general, especially the geographically most restricted ones that are affected by the culture bias problem. Also, because of the unexpected MTBC genome diversity even within a lineage (**Chapter 8**), the sequencing technique should be able to successfully detect all lineages, including in paucibacillary samples. But for now high quality and affordable sputum-based WGS is not yet available, and efforts should be made to have it available soon.

Culture is meant to inform the microbiologist whether the mycobacteria are alive or dead. Highly sensitive, sputum-based techniques capable of providing that information are also needed to replace culture for treatment monitoring.

Ideally, this viability test could be combined with the direct diagnostic assay, making use of the same sample preparation.

To ensure good quality specimens for comprehensive diagnostics, comparing the performance of CPC, OMNIgene.SPUTUM or ethanol for long-term sputum storage at ambient temperatures beyond 37°C are needed, as this is reality in many countries. In Benin -where the mean ambient temperature is ~28°C-31°C (**Chapter 3**) [4], our comparison with successful targeted *rpoB* sequencing after storage at 37°C remains relevant, but could be expanded to genomic DNA extraction and direct WGS.

With regard to treatment success and (development of) drug-resistance, our data showed an association of L5 with RIF resistance, and confirmed absence of the *Rv1979c* gene [5–7] in all L5 isolates (**Chapter 8**). Further studies could be done to investigate the extent of RIF-resistance acquisition in L5 strains and treatment outcome in patients previously diagnosed with multidrug-/rifampicin-resistant L5 isolates, especially in L5-endemic countries. The MICs clofazimine and bedaquiline for L5 isolates should be studied, to correlate with the genomic findings of this thesis. Besides, the impact on MICs caused by a broad variety of mutations in other genes associated with clofazimine and bedaquiline resistance (*Rv0678, Rv2535c* (*pepQ*), *atpE*)[5–8] remains insufficiently documented as of today.

In addition, it should be investigated whether the significantly higher proportion of ABF-scanty sputa (microscopy) from L5 failure patients is indicative of false treatment failure detected using microscopy.

Finally, experimental gene supplementation studies could be conducted on the L5 strain, to see if the supplementation of *Rv1993c*, *Rv1994c* and *Rv2074*

(vitamin B6: pyridoxine) [9,10] in culture medium would improve in vitro growth, and also if the supplementation of *Rv1977*, *Rv1978*, *Rv1993c*, *Rv1994c, Rv2074* [9,10] would increase L5 virulence (survival in macrophage during chronic stage of infection, immune-invasion).

References

[1] Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. Genotypic characterization directly applied to sputum improves the detection of Mycobacterium africanum West African 1, under-represented in positive cultures. PLoS Negl Trop Dis 2017;11:e0005900. doi:10.1371/journal.pntd.0005900.

[2] Sanoussi CN, De Jong BC, Odoun M, Arekpa K, Ligali MA, Bodi O, et al. Low sensitivity of the MPT64 identification test to detect lineage 5 of the Mycobacterium tuberculosis complex 2018. doi:10.1099/jmm.0.000846.

[3] Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, et al. The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: Implications for in vivo growth. Mol Microbiol 2005;56:163–74. doi:10.1111/j.1365-2958.2005.04524.x.

[4] Sanoussi CN, de Jong BC, Affolabi D, Meehan CJ, Odoun M, Rigouts L. Storage of Sputum in Cetylpyridinium Chloride, OMNIgene.SPUTUM, and Ethanol Is Compatible with Molecular Tuberculosis Diagnostic Testing. J Clin Microbiol 2019;57:e00275-19. doi:10.1128/JCM.00275-19.

[5] Zhang S, Chen J, Cui P, Shi W, Zhang W, Zhang Y. Identification of novel mutations associated with clofazimine resistance in Mycobacterium tuberculosis 2015. doi:10.1093/jac/dkv150.

[6] Yew WW, Liang D, Chan DP, Shi W, Zhang Y. Molecular mechanisms of clofazimine resistance in Mycobacterium tuberculosis. J Antimicrob Chemother 2017;72:2943–4. doi:10.1093/jac/dkx227.

[7] Ghodousi A, Hussain Rizvi A, Quadir Baloch A, Ghafoor A, Masood F, Qadir M, et al. Acquisition of cross-resistance to Bedaquiline and Clofazimine following treatment for Tuberculosis in 2 Pakistan 2019. doi:10.1101/630012.

[8] Almeida D, Ioerger T, Tyagi S, Li S-Y, Mdluli K, Andries K, et al. Mutations in pepQ Confer Low-Level Resistance to Bedaquiline and Clofazimine in Mycobacterium tuberculosis 2016. doi:10.1128/AAC.00753-16.

[9] Ahmed FH, Mohamed AE, Carr PD, Lee BM, Condic-Jurkic K, O'mara ML, et al. Rv2074 is a novel F 420 H 2-dependent biliverdin reductase in Mycobacterium tuberculosis 2016. doi:10.1002/pro.2975.

[10] Dick T, Manjunatha U, Kappes B, Gengenbacher M. Vitamin B6 biosynthesis is essential for survival and virulence of Mycobacterium tuberculosis. Mol Microbiol 2010;78:980–8. doi:10.1111/j.1365-2958.2010.07381.

Acknowledgments

To my promotors and co-promotors:

Prof. Dr. Bouke de Jong, Prof. Dr. Leen Rigouts, Prof. Dr Dissou Affolabi, Dr. Conor Meehan.

It is difficult to express my gratitude by words, yet I will try. Thank you for your continuous support, guidance and great availability. I am lucky to have you as promotors. I have learned and gained a lot from each of you. Each of you has particular talents and I am very happy that I got a bit of all those talents. Thank you so much!

To the staff of the Mycobacteriology Unit of the ITM: I thank you all for your support of any kind and especially Viviane, Karin, Ciska, Cécile, Pim, Sarah, Praha with whom I closely worked.

My peers at the Mycobacteriology Unit of ITM: thank you

I thank the staff of the ITM student service, for all the support.

I thank the National TB Program of Benin for the support in conducting the BeniDiT study

I also thank the staff of the TB clinics staff who participated in the BeniDiT study.

To LRM staff: I thank all of you for your support of any kind.

Dad, mom, thank you from the bottom of my heart. You have always been present and supportive whenever I needed it and even when I did not expect it. You are so thoughtful. I am happy that you are my parents.

Sisters and brother, thank you for your support. I am happy to have you.

Baraka, Nouri, Mathieu; you have been very supportive during all the journey. Thank you very much for your deep support and understanding of my long periods of absence, my unavailability even when I was present. You understood my needs and supported me until the end. We made it together! Deep thanks. May the Blessings Be!

Last but not the least, I am very grateful to DGD Belgium for funding this PhD research.

List of publications

1- **Sanoussi CN**, BC de Jong, D Affolabi, CJ Meehan, M Odoun, L Rigouts (2019) **Storage of sputum in cetylpyridinium chloride, OMNIgeneSPUTUM and ethanol is compatible with molecular tuberculosis diagnostic testing.** J Clin Microbiol. 2019 Jun 25;57(7). pii: e00275-19.doi:10.1128/JCM.00275-19 ; https://www.ncbi.nlm.nih.gov/pubmed/31092592

2- **Sanoussi CN**, BC de Jong, M Odoun, K Arekpa, M Ali Ligali, O Bodi, S Harris, B Ofori-Anyinam, D Yeboah-Manu, ID Otchere, A Asante-Poku, S Anagonou, S Gagneux, M Coscolla, L Rigouts, and D Affolabi (2018) **Low sensitivity of the MPT64 identification test to detect lineage 5 of the Mycobacterium tuberculosis complex**. J Med Microbiol. 2018 Dec; 67(12):1718-1727.doi:10.1099/jmm.0.000846. https://www.ncbi.nlm.nih.gov/pubmed/30388066

3- Affolabi D, **N Sanoussi**, A Sossou, T Nys, O Bodi, M Esse, S Houeto, F Massou, BC de Jong, L. Rigouts. (2018) **Performance of OMNIgene•SPUTUM (DNA Genotek) and cetylpyridinium chloride for sputum storage prior to mycobacterial culture**. J Med Microbiol. 2018 May2.doi:10.1099/jmm.0.000745. https://www.ncbi.nlm.nih.gov/pubmed/29717969

4- **Sanoussi CN**, D Affolabi, L Rigouts, S Anagonou, B de Jong (2017) **Genotypic characterization directly applied to sputum improves the detection of *Mycobacterium africanum* West African 1, under-represented in positive cultures**. PLoS Negl Trop Dis 11(9): e0005900. https://doi.org/10.1371/journal.pntd.0005900

5- Otchere ID, M Coscollá, L Sánchez-Busó, A Asante-Poku, D Brites, C Loiseau, C Meehan, S Osei-Wusu, A Forson, C Laryea, AI Yahayah, A Baddoo, GA Ansa, SY Aboagye, P Asare, S Borrell, F Gehre, P Beckert, TA Kohl, **N Sanoussi**, C Beisel, M Antonio, S Niemann, BC de Jong, J Parkhill, SR Harris, S Gagneux and D Yeboah-Manu (2018) **Comparative genomics of *Mycobacterium africanum* Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches**. *Scientific Reports* volume 8, Article number: 11269 (2018). https://www.nature.com/articles/s41598-018-29620-2

6- Affolabi D, **N Sanoussi**, S Codo, F Sogbo, P Wachinou, F Massou, Kehinde, S Anagonou (2017) **First Insight into a Nationwide Genotypic Diversity of**

Mycobacterium tuberculosis **among Previously Treated Pulmonary Tuberculosis Cases in Benin, West Africa**. Canadian Journal of Infectious Diseases and Medical Microbiology Volume 2017, Article ID 3276240, 6 pages https://doi.org/10.1155/2017/3276240

7- Affolabi D., **N. Sanoussi**, M. Odoun, A. Martin, L. Koukpemedji, J. C. Palomino, L. Kestens, S. Anagonou and F. Portaels.(2013) **Rapid low-cost identification of *Mycobacterium tuberculosis complex* using p-nitro-benzoic acid (PNB) as inhibitor and the resazurin microplate assay (REMA): A preliminary study**. African Journal of Microbiology research, Vol. 7(24), pp. 3135-3138, 11 June, 2013, DOI: 10.5897/AJMR10.278, ISSN 1996-0808 ©2013 Academic Journals, http://www.academicjournals.org/AJMR

8- Affolabi D., **N. Sanoussi**, K. Vandelannoote, M. Odoun, F. Faïhun, G. Sopoh, S. Anagonou, F. Portaels and M. Eddyani. (2012) **Effect of Decontamination, DNA Extraction, and Amplification Procedures on the Molecular Diagnosis of *Mycobacterium ulcerans* Disease (Buruli Ulcer)**. J. Clin. Microbiol. 2012, 50(4):1195. DOI: 10.1128/JCM.05592-11.

9- Affolabi D., F. Faihun, **N. Sanoussi**, G. Anyo , C.I Shamputa, L. Rigouts, L. Kestens, S. Anagonou, F. Portaels. (2009) **Possible outbreak of streptomycin-resistant *Mycobacterium tuberculosis Beijing* in Benin**. Emerg Infect Dis. 2009 Jul; 15(7):1123-5.

10- Affolabi D., G. Anyo, F. Faihun, **N. Sanoussi**, C.I. Shamputa, L. Rigouts, L. Kestens, S. Anagonou, F. Portaels (2009) **First molecular epidemiology study of tuberculosis in Benin**. Int J Tuberc Lung Dis. 2009 Mar; 13(3):317-22.

11- Affolabi D, M. Odoun, **N. Sanoussi**, A. Martin, J.C. Palomino, L. Kestens, S. Anagonou, F. Portaels. (2008) **Rapid and inexpensive detection of multidrug-resistant *Mycobacterium tuberculosis* with the nitrate reductase assay using liquid medium and direct application to sputum samples**. J Clin Microbiol. 2008 Oct; 46 (10): 3243-5. Epub 2008 Aug 27.

12- Affolabi D, **N. Sanoussi**, M. Odoun, A. Martin, L. Koukpemedji, J.C. Palomino, L. Kestens, S. Anagonou, F. Portaels. (2008) **Rapid detection of multidrug-resistant *Mycobacterium tuberculosis* in Cotonou (Benin) using two low-cost colorimetric methods: resazurin and nitrate reductase assays**. J Med Microbiol. 2008 Aug;57 (Pt 8): 1024-7.

13- Affolabi D., M. Odoun, **C. N. Sanoussi**, B. Tanimomo-Kledjo, D. Saîzonou, K. Soumaïla, L. Kestens, S.Y. Anagonou, F. Portaels. (2008) **Bulk staining of smears: no demonstrated risk of bacilli transfer from a positive to a negative smear**. Int J Tuberc Lung Dis. 2008 Jun; 12(6):683-5.

Manuscripts in preparation

1- **Sanoussi** CN, M Coscolla, B Ofori-Anyinam, P Beckert, I Otchere, P de Rijk, M Antonio, S Niemann, J Parkhill, S Harris, D Yeboah-Manu, S Gagneux, L Rigouts, D Affolabi, BC de Jong, CJ Meehan. **The genomic diversity of *Mycobacterium tuberculosis* lineage 5 (*Mycobacterium africanum* West-African 1).** In preparation

2- **Sanoussi CN**, D Affolabi, L Rigouts, M Odoun, K Arekpa, O Bodi, P de Rijk, K Vandelannoote, C Adigbonon, C Uwizeye, M Ali Ligali, BC de Jong. ***Mycobacterium africanum* Lineage 5 is associated with Gbe ethnicity and overrepresented among new tuberculosis patients in Southern Benin.** In preparation

Congress contributions (oral presentations, posters)

1- **Sanoussi CN**, L. Rigouts, M. Ali Ligali, O. Bodi, K. Arekpa, M. Odoun, S. Houéto, F. Massou, H. Dedehouanou, B. C. de Jong, D. Affolabi (2019) **Relative decrease in rifampicin-resistant tuberculosis in Benin**. Abstract **accepted** for presentation at the *50th (2019) World Conference on Lung Health of the International Union against Tuberculosis and Lung Diseases (The Union)***,** Hyderabad, India, 30 October – 2 November 2019

2- **Sanoussi CN**, M. Coscolla, B. Ofori-Anyinam, P. Beckert, I. D. Otchere, P. de Rijk, M. Antonio, S. Niemann, J. Parkhill, S. Harris, D. Yeboah-Manu, S. Gagneux, L. Rigouts, D. Affolabi, B. C. de Jong, C. J. Meehan (2019) **Genomic characterization of *Mycobacterium africanum* West-African 1 (Lineage 5).** Poster presented (abstract accepted) at the 40th (*2019) Annual ESM (European Society of Mycobacteriology) congress,* Valencia, Spain, 30 June – 3 July 2019

3- **Sanoussi CN**, D Affolabi, M Odoun, BC de Jong, L Rigouts (2018) **Comparison of one-month storage of sputum in CPC, ethanol and OMNIgene.SPUTUM for molecular analyses**. Oral presentation (after abstract accepted) at the *49th* (2018) *World Conference on Lung Health of the International Union against Tuberculosis and Lung Diseases (The Union)*, The Hagues, Netherland, 23-27 October 2018

4- **Sanoussi C. N.**, D. Affolabi, L. Rigouts, S. Anagonou, B. C. de Jong. (2017) **Performance de la caractérisation génotypique du bacille tuberculeux directement à partir des crachats**. Oral presentation (abstract accepted) at the *''1ères Journées scientifiques sur la Tuberculose au Bénin (JTB Bénin 2017)* '', 02-03 November 2017, Cotonou, Benin.

5- **Sanoussi C. N**., D. Affolabi, M. Odoun, K. Arekpa, S. Houéto, O. Bodi, S. Anagonou, L. Rigouts, B. C. de Jong. (2017) **First study of the nationwide population structure of the *Mycobacterium tuberculosis* complex in Benin**. Poster presented (abstract accepted) at the 38th (2017) annual congress of the *European Society for Mycobacteriology (ESM)*, 25 -28 June 2017, Sibenik, Croatia.

6- **Sanoussi N.**, D. Affolabi, L. Rigouts, S. Anagonou, B. de Jong. (2016) **Performance of spoligotyping applied directly on sputum for genotypic characterization of *Mycobacterium tuberculosis* complex**. *IJTLD*, vol.20, n°11, nov 2016, suppl 1, page S137; Poster presented (abstract accepted) at the *47th World Conference on Lung Health of the International Union against Tuberculosis and Lung Diseases (The Union)*, Liverpool, UK, 26-29 October 2016, PD-609-27

Conferences / congresses attendance

1- *50th (2019)* ***World Conference on Lung Health of the International Union against Tuberculosis and Lung Diseases (The Union),*** Hyderabad, India, 30 October – 2 November 2019 (In preparation abstract accepted for poster presentation)

2- *40th (2019)* ***Annual ESM (European Society of Mycobacteriology) congress,*** Valencia, Spain, 30 June – 3 July 2019

3- *49th* (2018) ***World Conference on Lung Health of the International Union against Tuberculosis and Lung Diseases (The Union)***, The Hagues, Netherland, 23-27 October 2018

4- *"**1ères Journées scientifiques sur la Tuberculose au Bénin** (JTB Bénin 2017)* '', 02-03 November 2017, Cotonou, Benin

5- 38th (2017) **annual congress of the *European Society for Mycobacteriology (ESM)***, 25 -28 June 2017, Sibenik, Croatia

Training and courses during PhD studies

2019 — 2019 PhD training weeks (3 weeks), Institute of Tropical Medicine (ITM), Antwerp, Belgium

Topics of training attended:

1- Workshop **Research integrity (short introduction and dilemma's)**
2- Workshop **Authorship ethics (real life cases)**

3- **Handling biosamples in research**
4- **Good Documentation Practices**
5- Worshop **'The floor is yours, because life is too short for bad presentations'**
6- Workshop **'Hurray, another group just published the research findings that we are about to report'**
7- **How to get the most out of peer review (or not)?**
8- **Leadership & academic work: necessity or necessary evil?**
9- **Measure your impact: a quick guide to bibliometrics**
10- **How to get your research funded?**
11- Discussion: **'Career choices: the mobility imperative'**

25, 28, 30/01/2019 — Course: **Analysis of grouped or longitudinal data using linear mixed models**, Lecturers: Erik Fransen and Ella Roelant, StatUa, Campus Groenenburger, Antwerp Doctoral School, University of Antwerp

15, 17, 18/01/2019 — Course: **R workshop**, Lecturers: Erik Fransen and Ella Roelant, StatUa, Campus Groenenburger, Antwerp Doctoral School, University of Antwerp

15, 16, 30 /11/2018 — Course: **Grow your future career**, Lecturer: Robin Lefebvre, Campus Middleheim room A301, Antwerp Doctoral School, University of Antwerp, Belgium

7,12, 23/11/2018 — Course: **Applied communication**, Lecturer: Gerrit Van Aken, Campus Middleheim room A301, Antwerp Doctoral School, University of Antwerp, Belgium

19/11/2018 — Course: **E-sources for the Biomedical sciences**, lecturer: Lejeune Barbara, Campus Drie Eiken, room 312, University of Antwerp.

25/04/2018 — Training on "**IATA DGR CLASS 6.2 infectious substance**" (successful with certificate), Lecturer: Sarah Sander, SGS Belgium N.V., course at ITM Antwerp Belgium

5, 12, 19, 26/03/2018 — Course: '**English for communicating effectively in academic context'**, lecturer: Daniel Sossi (Lingapolis), Campus Middleheim, Antwerp Doctoral School, University of Antwerp, Belgium

7-28 /02/2018	Course: **"Molecular data for Infectious Diseases" (MID) short-course, face-to-face part** (in ITM, Antwerp), organized by the Department of Biomedical Sciences of ITM, Antwerp, Belgium. **Major** for face to face part. Final mark (face-to-face and online courses): **79.4/100** (rank: 2nd)
2nd/02/2018	Attended the course on **"Presentation skills and storytelling: how to get your (research) message across"**; 2018 PhD training weeks, ITM, Antwerp, Belgium
6th/10/2017 and 14th/12/2017	Participated in **Biosafety training including use of BSL2 and BSL3 laboratories**; by Leen Rigouts (14th/12/2017) from ITM, Belgium and Michele Driesen (6th/10/2017) from ITM, Belgium; held in LRM, Cotonou, Benin
1st – 28th/11/2017	Participated and passed the online part (**distance learning part**) of the **"Molecular data for Infectious Diseases" (MID)** short-course, organized by the Department of Biomedical Sciences of ITM, Antwerp, Belgium. Mark: 7.6/10
10th – 11th/10/2016	Attended the course on **''Introduction to Linux''** in University of Antwerp, lecturer: Stefan Becuwe
26th /05/2015	Successful completion of training for Pipetting course **"Rainin pipetting seminar"**; Antwerp, Belgium Certificate issued by N. V Mettler Toledo S.A, Zaventem, Belgium
19th/05/2015	Successful completion of training for **"Safe transport of Dangerous Goods by Air 6.2 Infectious Substances and Class 9 (Other related substances) for Shippers"** (IATA transportation course), Antwerp, Belgium Certificate issued 29 May 2015, by World Courier Belgium

Research stays / residencies

September – October 2019	PhD Research stay in the **Institute of Tropical Medicine (ITM), Antwerp, Belgium** (1 month stay)
July –September 2019	PhD Research stay in the **Laboratoire de Référence des Mycobactéries (LRM), Cotonou, Benin** (2 months stay)
June -July 2019	PhD Research stay in the **ITM, Antwerp, Belgium** (1 month stay)
February –June 2019	PhD Research stay in the **LRM, Cotonou, Benin** (4 months stay)
10/2018 – 02/2019	PhD Research stay in the **ITM, Antwerp, Belgium** (data analysis and writing, laboratory analyses/works, hands-on in bioinformatics, 4 months stay)
April - October 2018	PhD Research stay in the **LRM, Cotonou, Benin** (6 months)
January– April 2018	PhD Research stay in the **ITM, Antwerp, Belgium** (including hands-on in bioinformatics, 3 months stay)
07/2017- 01/2018	PhD Research stay in the **LRM, Cotonou, Benin** (patient recruitment and study sites supervision, laboratory analyses, 7 months)
05/2017- 07/2017	PhD Research stay in the **ITM, Antwerp, Belgium** (including hands-on in bioinformatics, 2 months)
Sept - Oct 2016	PhD Research stay in **ITM, Antwerp, Belgium** (including hands-on in bioinformatics, 1.5 months)
09/ 2015- 09/ 2016	PhD Research stay in **LRM, Cotonou, Benin** (patient recruitment and study sites supervision, laboratory analyses, 1 year)
June– August 2015	Pre-PhD Research stay in the **LRM, Cotonou, Benin** (Ethics submissions, Laboratory works, 3 months)
March – May 2015	Pre-PhD Research stay in **ITM, Antwerp, Belgium** (Write and submit PhD research proposal to ITM PhD Committee, laboratory training and works, 3 months)

Guidance of theses

Co-supervision of:

- Thesis “Évaluation de la performance de deux milieux de conservation des expectorations *(OMNI·gene sputum et chlorure de cétylpyridinium)* pour la culture des mycobactéries”, Adelaïde Sossou (2016-2017, MSc Pharmaceutical Sciences, Faculty of Health Sciences, University of Abomey-Calavi, Cotonou, Benin)
- Thesis “Evaluation du test rapide MPT64 pour l’identification des mycobactéries tuberculeuses au Bénin”, Karamatou Arekpa (2016-2017, BSc Human Biology: Biomedical analyses, Ecole Polytechnique d’Abomey-Calavi (EPAC), University of Abomey-Calavi, Abomey-Calavi, Benin)

Curriculum vitae

Education

September 2015 to October 2019	**PhD student (Biomedical Sciences)** Fields covered by PhD topic: **Microbiology, genetics and molecular biology, tuberculosis (TB) diagnostics, molecular epidemiology of TB, genome analysis, bioinformatics** Sandwich PhD with research in: - Laboratoire de Référence des Mycobactéries: **LRM** (National Reference Laboratory for Mycobacteria), Cotonou, Benin - Institute of Tropical Medicine (**ITM**), Mycobacteriology Unit, Antwerp, Belgium **University of Antwerp, Belgium** **Selected after competitive application** for the PhD grant.
March - August 2015	**Pre-doctoral student** ITM, Mycobacteriology Unit, Antwerp, Belgium and LRM, Cotonou, Benin
2011 –2014	**Master of Science in Clinical Trials** Final mark: **3.85/5 GPA** (grade point average) **180 credits** **London School of Hygiene and Tropical Medicine (LSHTM)**; University of London, International Programmes, London, United Kingdom
2009-2011	**Master of Science** in Genetics, Biotechnologies and Biological Resources. Field of specialization: **Genetics and Molecular Biology Major** with final mark **16.77/20** Faculty of Sciences and Techniques (FAST), University of Abomey-Calavi (UAC), Benin
2007-2009	Postgraduate diploma in Natural Sciences. Field of specialization: Physiology Faculty of Sciences et Techniques (FAST), University of Abomey-Calavi (UAC), Benin

2006-2007 Bachelor in Natural Sciences. Field of specialization: Physiology
Faculty of Sciences and Techniques (FAST), University of Abomey-Calavi, Benin (UAC)

2002 – 2006 **Postgraduate diploma in Bio-Medical Analyses (Human Biology Techniques), (diplôme d'ingénieur des travaux en Analyses Bio-médicales)**
Final mark: **74.63/100**
Polytechnic School of Abomey-Calavi (Ecole Polytechnique d'Abomey-Calavi (EPAC))/ University of Abomey –Calavi, Benin (UAC)
Pass the competitive examination for entry in Bio-Medical Analyses at the Polytechnic School of Abomey-Calavi (EPAC)

Other theses

1- **Assessment of the performance of the molecular genetic test Hain GenoType MTBDRplus for the early and rapid detection of multiresistant-tuberculosis in Benin**. Long dissertation for the degree of *Master in Genetics, Biotechnologies and Biological Resources. Option: Genetics and Molecular Biology*. Defence the 5th August 2011, University of Abomey-Calavi, Benin.

2- **Preliminary study for the development of a rapid test for the detection of *Mycobacterium ulcerans* resistance to rifampicin.** Long dissertation for the Postgraduate diploma in Natural Sciences (*Maîtrise es Sciences Naturelles), Option: Physiology*, Defence the 3rd July 2009, University of Abomey-Calavi, Benin.

3- **Comparison of 4 decontamination techniques for the culture of *Mycobacterium ulcerans*.** Long dissertation for the Postgraduate Diploma in Bio-medical Analyses, Defence the10th August 2006, University of Abomey-Calavi, Benin.

Professional experience

September 2006 - February 2015 : **Laboratory scientist** (biotechnologist Biomedical Sciences) at the **National Reference Laboratory for Mycobacteria (Laboratoire de Référence des Mycobactéries (LRM)), Cotonou, Benin /** Centre National Hospitalier Universitaire de Pneumo-Phtisiologie (CNHU-PPC), National Tuberculosis Programme, Benin: Standard operating Procedures (SOP) development, research works, actively involved in the implementation of first genotypic diagnosis tests in LRM, realization phenotypic and genotypic testing (drug susceptibility testing, strain typing), molecular diagnosis tests , culture, biochemical and serological blood tests, laboratory supervision for tuberculosis and Buruli ulser diagnosis, technical training of other laboratory personnel.

Other

Since June 2018	**Peer-reviewer** of articles in scientific journals
August 2014	Trained on **sequencing (PCR and sequence analysis)** of genes involved in drug-resistance of *M. tuberculosis*, LRM, Cotonou, Benin
June 2012	Participated in the **International Course of mycobacteriology applied to National Tuberculosis Programmes needs**, Cotonou, Benin, organized by International Against Tuberculosis and Lung Diseases (IUTLD) and National Tuberculosis Program of Benin
March 2011	Trained on **Quality Assurance** by IQLS at LRM, Cotonou.
January 2008	Trained on **diagnosis of Buruli ulcer by molecular biology**. Mycobacteriology Unit, ITM, Antwerp, Belgium.

Parameters of estim

1- **Selected after an international competitive call for application** for a grant for **PhD** in Biomedical Sciences (funded by **DGD Belgium**)
2- **Major of my class** for the face-to-face part of the international **short-course** "Molecular Data for Infectious Diseases **(MID)", Institute of Tropical Medicine, Antwerp Belgium**
3- **Major of my class** for the **Master of Science** in Genetics, Biotechnologies and Biological Resources. (Specialization: Genetics and Molecular Biology), Faculty of Sciences and Techniques, university of Abomey-Calavi, Benin
4- **Pass the competitive examination for a national grant to study** in Biomedical Sciences (**Postgraduate diploma**) - Bio-Medical Analyses at the Polytechnic School of Abomey-Calavi (EPAC), University of Abomey-Calavi, Benin

Contact

Email: **ndirasanoussi@gmail.com**