



# Successful expansion of *Mycobacterium tuberculosis* Latin American and Mediterranean sublineage (L4.3/LAM) in Tunisia mainly driven by a single, long-established clonal complex

Mohamed Amine Skhairia, Naira Dekhil, Besma Mhenni, Saloua Ben Fradj,  
Helmi Mardassi

## ► To cite this version:

Mohamed Amine Skhairia, Naira Dekhil, Besma Mhenni, Saloua Ben Fradj, Helmi Mardassi. Successful expansion of *Mycobacterium tuberculosis* Latin American and Mediterranean sublineage (L4.3/LAM) in Tunisia mainly driven by a single, long-established clonal complex. *International Journal of Infectious Diseases*, 2021, 103, pp.220 - 225. 10.1016/j.ijid.2020.11.195 . hal-03492930

**HAL Id: hal-03492930**

**<https://hal.science/hal-03492930>**

Submitted on 1 Apr 2022

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License



# Successful expansion of *Mycobacterium tuberculosis* Latin American and Mediterranean sublineage (L4.3/LAM) in Tunisia mainly driven by a single, long-established clonal complex

Mohamed Amine Skhairia<sup>1</sup>, Naira Dekhil<sup>1</sup>, Bisma Mhenni, Saloua Ben Fradj, Helmi Mardassi\*

Unit of Typing and Genetics of Mycobacteria, Laboratory of Molecular Microbiology, Vaccinology, and Biotechnology Development, Institut Pasteur de Tunis, Université Tunis El Manar, Tunis, Tunisia



## ARTICLE INFO

### Article history:

Received 19 October 2020

Received in revised form 24 November 2020

Accepted 26 November 2020

### Keywords:

*Mycobacterium tuberculosis*

L4.3/LAM

Genotyping

Clonal complex

Whole-genome sequencing

Evolution

## ABSTRACT

**Objectives:** To explore the evolutionary history of *Mycobacterium tuberculosis* Latin American and Mediterranean (L4.3/LAM) sublineage in Tunisia, where it predominates.

**Methods:** High-resolution genotyping of 252 L4.3/LAM clinical strains was undertaken, and whole-genome sequencing was performed on 31 representative isolates.

**Results:** Genotyping data coupled with Bayesian analyses split the Tunisian L4.3/LAM strain collection into two divergent entities (65.07% vs 34.92%): a major subpopulation, dominated by a single clonal complex (CC), TUN4.3\_CC1 (94.51%); and a minor subpopulation, dominated by TUN4.3\_CC2 (42.04%). TUN4.3\_CC1 is clearly thriving in Tunisia, accounting for 61.5% of the L4.3/LAM sublineage. TUN4.3\_CC1 displayed higher mean allelic richness compared with TUN4.3\_CC2 and predominated throughout the entire region, indicating a long-established history. The very low proportion of drug resistance among TUN4.3\_CC1 isolates is indicative of their intrinsic ability to spread successfully in the host population. Genomic analyses further confirmed the clear genetic separation between the two main CCs (pairwise fixation index 0.56), and suggested the relatively ancient origin of TUN4.3\_CC1. Consistent with its successful expansion, TUN4.3\_CC1 showed reduced mean pairwise genetic distance between genomes. **Conclusions:** These findings link the successful expansion of L4.3/LAM in Tunisia to a single long-established clone.

© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Tuberculosis (TB) is a major human scourge largely caused by *Mycobacterium tuberculosis*, a pathogen that primarily affects the lungs. TB can also be caused by other bacterial species closely related to *M. tuberculosis*, all of which are referred to as the '*M. tuberculosis* complex' (MTBC) (Brosch et al., 2002; Brites and Gagneux, 2017). According to the latest estimates, TB is the main cause of infectious disease deaths worldwide, with an estimated 10 million cases and 1.5 million deaths annually (World Health Organization, 2019). The rapid expansion of multi-drug-resistant TB (MDR-TB) and the amplifying effect of a parallel human

immunodeficiency virus (HIV) epidemic severely restrict the likelihood of eradicating TB (Eldholm et al., 2016).

The MTBC encompasses seven human-adapted phylogenetic lineages and two lineages adapted to various wild and domestic animal species (Gagneux, 2018). Among the human-adapted MTBC lineages, some have been shown to occur globally (L2 and L4), and others are geographically restricted (L5, L6 and L7), suggesting generalist and specialist phenotypes (Stucki et al., 2016). The phylogeographical picture of these lineages reflects a longstanding host–pathogen interaction (Comas et al., 2013; Brites and Gagneux, 2015).

The current TB pandemic is mainly driven by the two generalist lineages: L2 and L4. While L2 is expanding rapidly through continents, fuelling the MDR-TB epidemic, the more widely dispersed L4 sublineage is likely to be endemic and less prone to cause cross-border MDR-TB expansion (Merker et al., 2015; Stucki et al., 2016; Brynildsrud et al., 2018). Global dispersal of the L4 lineage has been shown to be intimately linked to historical

\* Corresponding author at: Unit of Typing and Genetics of Mycobacteria, Institut Pasteur de Tunis, 13 Place Pasteur BP74, 1002, Tunis, Tunisia.

E-mail address: [helmi.merdassi@pasteur.rns.tn](mailto:helmi.merdassi@pasteur.rns.tn) (H. Mardassi).

<sup>1</sup> These authors contributed equally.

migration out of Europe, with intense dispersal to Africa and the Americas accompanying the European colonial expansion that started in the 16th century (Brynildsrud et al., 2018). Clinical observations and molecular data have shown that L4 consists of at least 10 sublineages (L4.1.1–L4.10), with considerable variation in the transmission success of different L4 strains in clinical settings (Coscolla and Gagneux, 2014; Ke et al., 2018).

Among the L4 sublineages, the L4.3 Latin American and Mediterranean (L4.3/LAM) sublineage is the most frequently encountered worldwide (Stucki et al., 2016), particularly in the Euro Mediterranean region and Latin America. The global phylogeography of this generalist sublineage revealed three global genotypes that could be defined by large deletions: RD115 (the most ancestral), RD174 (best known as 'RD-Rio') and spoligotype SIT33. However, certain L4.3/LAM genotypes tend to be geographically confined, such as spoligotype SIT388 which is endemic in Japan (Mokrousov et al., 2016). The existence of additional differentiated genotypes that are tightly associated with their host population is very likely. These geographically confined L4.3/LAM clones could have evolved from one of the three global generalist genotypes as a result of local adaptation (Mokrousov et al., 2016; Stucki et al., 2016; Brynildsrud et al., 2018).

In Tunisia, L4.3/LAM is the most prevalent sublineage, accounting for more than 37% of all cases of TB. Within this sublineage, the LAM9 genotype was found to be the most common (Namouchi et al., 2008). In Morocco, a neighbouring country of the Maghreb, L4.3/LAM is even more successful, representing almost 50% of the *M. tuberculosis* population (Bouklata et al., 2015; Stucki et al., 2016). Therefore, it is tempting to speculate that L4.3/LAM is particularly fit to Maghrebians. By combining molecular typing and whole-genome sequencing (WGS), this study explored the evolutionary history of L4.3/LAM in a Maghrebian country, Tunisia, in order to better understand the molecular basis underlying its relative success.

## Methods

### *M. tuberculosis* L4.3/LAM strain collection

*M. tuberculosis* isolates were obtained from the culture collection of the Unit of Typing and Genetics of Mycobacteria (UTGM) of Institut Pasteur, Tunis. The clinical strains, which had been isolated as part of UTGM's TB diagnostic activity, originated predominantly from three geographical regions in Northern Tunisia (Bizerte, Tunis and Zaghouan). Primary isolation and culturing of *M. tuberculosis* isolates was performed using Löwenstein-Jensen (LJ) medium, as described elsewhere (Kent and Kubica, 1985). The colonies were harvested from the surface of LJ medium, suspended, and preserved at  $-70^{\circ}\text{C}$  in Middlebrook 7H9 medium with 50% glycerol. Genomic DNA was prepared from recultured frozen stocks for molecular typing or WGS.

Isolates of the L4.3/LAM sublineage were identified by polymerase chain reaction targeting a LAM-specific IS6110 insertion site (Sampson and Warren, 1999; Marais et al., 2006). A collection comprising 252 L4.3/LAM clinical strains, spanning the period 2001–2017, was derived and used in this study (Table S1, see online supplementary material).

Among these 252 L4.3/LAM isolates, the results of drug susceptibility testing (DST) were available for 166 isolates. DST was determined by the proportion method on LJ medium, as described previously (Dekhil et al., 2016). The anti-TB drugs tested were: isoniazid, rifampicin, streptomycin, ethambutol and pyrazinamide at concentrations of 0.2, 40, 4.0, 2.0 and 200  $\mu\text{g/mL}$ , respectively. The criterion used for drug resistance was growth of  $\geq 1\%$  of the bacterial population on media containing the critical concentration of each drug. MDR-TB was defined as resistance to

both isoniazid and rifampicin, the two most effective front-line antituberculosis drugs. All other resistance profiles were referred to as 'non-MDR'.

### Genotyping and phylogenetic inference

High-resolution genotyping of 252 L4.3/LAM isolates was undertaken using the 24-loci mycobacterial interspersed repetitive units-variable numbers of tandem repeats (24-loci MIRU-VNTR) approach (Supply et al., 2006). Minimum spanning trees (MST) based on 24-loci MIRU-VNTR results were computed using a temporary BioNumerics 7.6.3 evaluation licence from Applied Maths (Sint-Martens-Latem, Belgium) with permission to publish. Clusters were defined as nodes containing strains with identical 24-loci MIRU-VNTR haplotypes, and clonal complexes (CC) were defined as groups with a maximum difference of two loci. The rate of clustered strains was used as an indicator for the extent of recent transmission (Merker et al., 2015).

Mean allelic richness in each L4.3/LAM CC was calculated based on a rarefaction procedure implemented in HP-RARE Version 1.0 which compensates for sampling disparities (Kalinowski, 2005).

### Population structure analyses

The population structure of L4.3/LAM was determined using a Bayesian approach implemented in STRUCTURE 2.3.4 based on 24-loci MIRU-VNTR data. The parameters of interest were set using a Markov chain Monte Carlo algorithm with a burn-in of  $10^5$  iterations and a run length of  $10^6$ . pophelper R Version 2.3.0 was used to generate barplots and to calculate delta *K* using the Evanno method (Francis, 2017), with the optimal value of *K* deduced using CLUMPAK (Kopelman et al., 2015).

### Whole-genome sequencing

To gain in-depth insight into the evolutionary history of L4.3/LAM prevailing in Tunisia, 17 and 11 isolates representative of TUN4.3\_CC1 and TUN4.3\_CC2, respectively, and three strains with an intermediate position between the two CCs, as revealed by the 24-loci MIRU-VNTR-based MST, underwent WGS.

Genomic DNA was extracted from *M. tuberculosis* colonies cultured in LJ medium, and subjected to Illumina paired-end (PE) sequencing using NextSeq. The data generated were analysed using an in-house automated pipeline, including open source software. Briefly, FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) was used for quality control of raw sequencing data. Both adapters and low-quality bases with a Phred quality score  $<20$  were clipped using Trimmomatic (Bolger et al., 2014), and only reads with a minimum length of 36 bases were used in the following analysis. Three different aligners—Burrows-Wheeler Alignment Tool (Li and Durbin, 2009), Novoalign (<http://www.novocraft.com/products/novoalign/>) and SMALT (<http://www.sanger.ac.uk/resources/software/SMALT/>)—were used to map PE reads to the genome sequence of *M. tuberculosis* H37Rv reference strain (Genbank: AL123456.3). Local re-alignment and de-duplication of alignment files were performed using the Genome Analysis Toolkit (GATK) (McKenna et al., 2010) and Picard (<https://broadinstitute.github.io/picard/>). VCF files, including single nucleotide polymorphisms (SNPs) and insertions/deletions, were generated from each alignment file using GATK. Variants were then annotated referring to the TubercuList server (Lew et al., 2011). SNPs located in repetitive sequences, such as PE/PPE genes and mobile elements, and those up to five base pairs from each other were excluded from the analysis. Only SNPs that were sequenced at a depth  $\geq 10$  and present in  $\geq 70\%$  of the reads were included in this study. SpolPred (Coll et al., 2012) was used to

predict spoligotypes *in silico*. Raw data were deposited in the European Nucleotide Archive (Accession No. PRJEB39509).

### Genomic analyses

Phylogenetic relationships between the Tunisian isolates were inferred by constructing an MST based on 1719 polymorphic sites using a temporary BioNumerics 7.6.3 evaluation licence (Applied Maths).

Pairwise SNP distances were calculated using the ape package and the dna.dist function in R Version 3.6.3. For comparison of pairwise numbers of SNPs between L4.3/LAM strains, the SNP distance between each pair of isolates was calculated, and a distribution of the pairwise distance was plotted. Wilcoxon rank sum test and Kruskal–Wallis test were used to test for differences as the data did not show a normal distribution.

## Results

### 24-loci MIRU–VNTR typing of L4.3/LAM

In total, 252 L4.3/LAM clinical strains from Tunisia were included in this study (Table S1, see online supplementary material). Although these originated mainly from Northern Tunisia, the collection was considered to reflect the prevailing situation throughout Tunisia as it represented three large areas, including Tunis. Furthermore, these clinical strains had been collected over a long period (17 years).

The whole collection was typed using 24-loci MIRU–VNTR, the highly discriminatory typing modality of *M. tuberculosis*. Overall, 179 MIRU–VNTR patterns were obtained, with a clustering rate of 0.289. The discriminatory power of each MIRU–VNTR locus was assessed using the Hunter and Gaston Discriminatory Index (HGDI). Four loci were highly discriminatory: Miru26 and Mtub39 had the best discriminatory power (HGDI = 0.77), followed by Miru40 (HGDI = 0.72) and Mtub04 (HGDI = 0.64).

### Population structure of *M. tuberculosis* L4.3/LAM sublineage in Tunisia

To gain insight into the population structure of L4.3/LAM in Tunisia, a Bayesian analysis was undertaken using STRUCTURE based on 24-loci MIRU–VNTR data. Two divergent subpopulations were disclosed. Multiple runs with *K* values ranging from 2 to 10

revealed maximal posterior probability at *K* = 2 (Figure 1), suggesting that all L4.3/LAM strains were descendants of two distinct ancestor populations.

These two subpopulations, referred to herein as ‘L4.3\_TUN1’ and ‘L4.3\_TUN2’, were disproportionately prevalent (65.07% vs 34.92%) and significantly differentiated (pairwise fixation index = 0.53) (Table 1).

### A particular CC of *M. tuberculosis* L4.3/LAM predominated in the Tunisian population

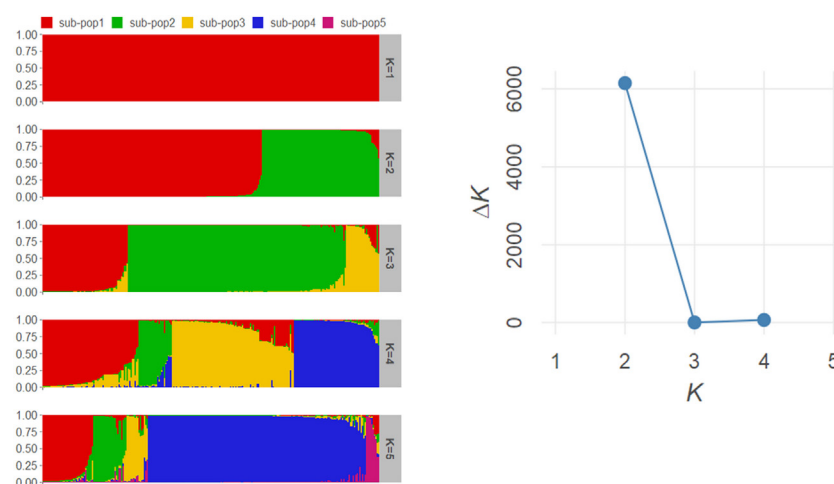
An MST based on 24-loci MIRU–VNTR data was constructed to explore the population structure of L4.3/LAM in Tunisia in more detail. Two subpopulation entities are clearly discernible in the MST depicted in Figure 2A, confirming the population dichotomy revealed by STRUCTURE analyses.

In total, the L4.3/LAM population in Tunisia comprised two main CCs (TUN4.3\_CC1 and TUN4.3\_CC2), eight minor CCs (TUN4.3\_CC3 to TUN4.3\_CC10) and 36 singletons (Table 1 and Figure 2A). TUN4.3\_CC1 and TUN4.3\_CC6 were restricted to the predominant population L4.3\_TUN1, while TUN4.3\_CC2 and the remaining seven CCs were specific to L4.3\_TUN2. The two main CCs (TUN4.3\_CC1 and TUN4.3\_CC2) prevailed in their respective subpopulations (94.51% and 42.04%, respectively), representing a total of 77% of the Tunisian L4.3/LAM population (61.5% and 14.68%, respectively). Hence, TUN4.3\_CC1 is by far the most prevalent CC of L4.3 LAM in Tunisia. TUN4.3\_CC1 tended to display a higher mean allelic richness than TUN4.3\_CC2 (Figure 2B), and predominated in the three regions under consideration, evolving through successive transmission peaks (Figure 3A). This finding is in line with its increased recent transmission rate, which was estimated to be twice that of TUN4.3\_CC2 (0.36 vs 0.19).

DST data were available for 166 clinical strains, 129 of which (77.71%) showed a pansusceptible profile. Only five strains (3.01%) were MDR, while the remaining 32 strains displayed resistance to one (16.86% monoresistance rate) or more (2.42%) antituberculosis drugs. As shown in Figure 3B, the two main CCs included a very small proportion of MDR strains.

### WGS-based analysis of Tunisian *M. tuberculosis* L4.3/LAM

To better establish the genetic relationships within and between L4.3/LAM clinical strains in Tunisia, 17 TUN4.3\_CC1 isolates and 11 TUN4.3\_CC2 isolates underwent WGS. In addition,

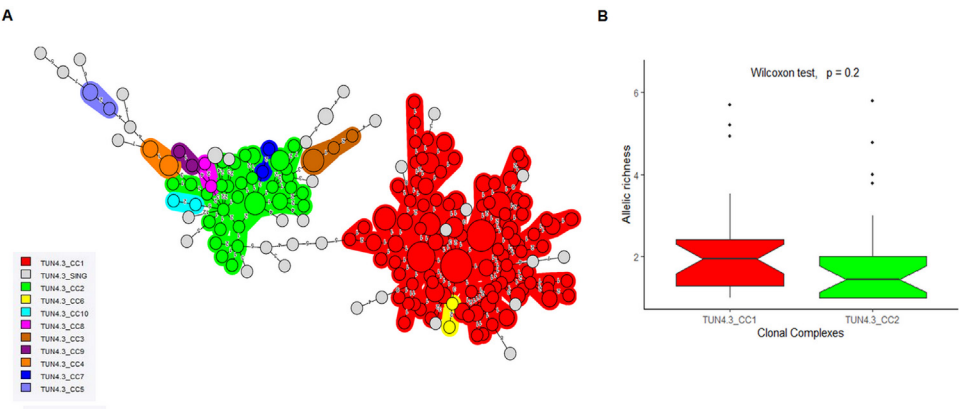


**Figure 1.** Clustering assignments of the 252 L4.3/Latin American and Mediterranean (LAM) isolates from Tunisia based on ADMIXTURE 1.3.0 (Alexander et al., 2009) for inferred *K* values ranging from 1 to 5. Each individual is represented by a single vertical line divided into *K* coloured segments, where *K* is the number of ancestral populations. The figure shown for a given *K* is based on the highest probability run of 10 STRUCTURE runs at that *K*.

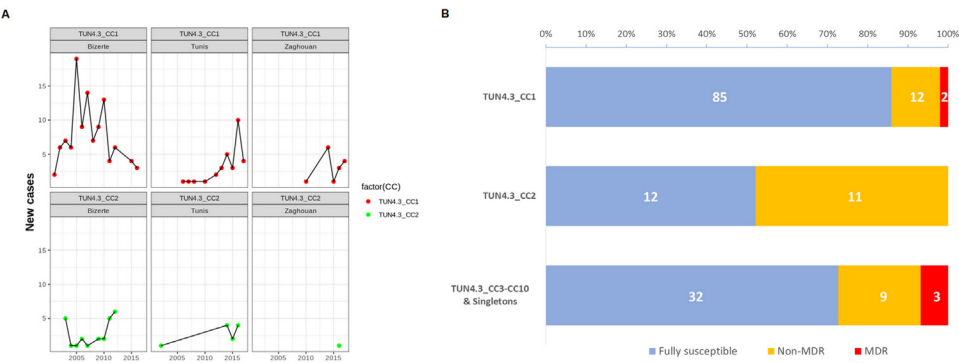
**Table 1**  
Population structure of L4.3/Latin American and Mediterranean (LAM) sublineage in Tunisia.

Population	Clonal complex	n (%)	Np	Na	Clustering rate	F <sub>ST</sub>
L4.3_TUN1 n = 164 (65.07%)	TUN4.3_CC1	155 (94.51%)	0.08	2.16	0.367	0.5399
	TUN4.3_CC6	2 (1.21%)	0	1.08	0	
	Singletons	7 (4.26%)	–	–	–	
	TUN4.3_CC2	37 (42.04%)	0.04	1.97	0.189	
L4.3_TUN2 n = 88 (34.92%)	TUN4.3_CC3	6 (6.81%)	0.03	1.13	0.5	0.333
	TUN4.3_CC4	5 (5.68%)	0	1.04	0.6	
	TUN4.3_CC5	3 (3.4%)	0	1.08	0.333	
	TUN4.3_CC7	2 (2.27%)	0	1.08	0	
	TUN4.3_CC8	2 (2.27%)	0	1.08	0	
	TUN4.3_CC9	2 (2.27%)	0.05	1.04	0	
	TUN4.3_CC10	2 (2.27%)	0.04	1.08	0	
	Singletons	29 (32.95%)	–	–	–	

Np, average number of private alleles per locus; Na, average number of alleles per locus; F<sub>ST</sub>, pairwise fixation index to evaluate the degree of population differentiation.



**Figure 2.** Population structure of *Mycobacterium tuberculosis* L4.3/Latin American and Mediterranean (LAM) sublineage in Tunisia based on 24-loci mycobacterial interspersed repetitive units-variable numbers of tandem repeats (MIRU-VNTR) data. (A) Minimum spanning tree based on 24-loci MIRU-VNTR copy numbers illustrating evolutionary relationships between the 252 *M. tuberculosis* L4.3/LAM isolates from Tunisia. Clonal complexes are represented by different colours. The size of a circle is proportional to the number of isolates sharing the same pattern. The number of allele changes between two patterns is indicated. (B) Boxplot of allelic richness of *M. tuberculosis* TUN4.3\_CC1 and TUN4.3\_CC2 calculated by a rarefaction procedure implemented using HP-RARE 1.0. Significant differences were calculated using the Wilcoxon test. Boxes correspond to median values  $\pm$  quartiles of allelic richness; dots show minimum/maximum values.

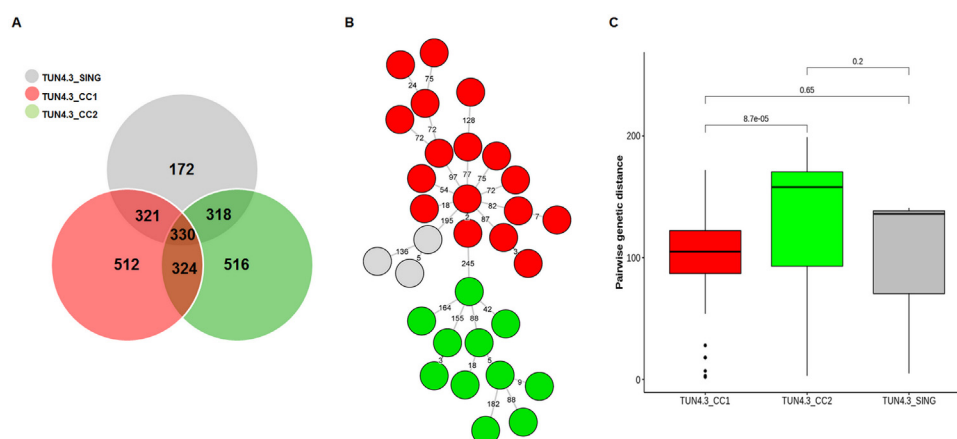


**Figure 3.** Incidence and drug resistance patterns of *Mycobacterium tuberculosis* L4.3/Latin American and Mediterranean (LAM) cases. (A) Geographic variations in incident cases attributed to TUN4.3\_CC1 and TUN4.3\_CC2 isolates, 2011–2017. (B) Proportions of L4.3/LAM multi-drug-resistant (MDR) strains among the identified clonal complexes (CC). Note the very low proportion of MDR strains. The numbers inside the bars indicate the number of strains with drug susceptibility test results.

three singleton isolates with phylogenetic positions intermediate between the two main CCs underwent WGS. Comparative genomics revealed that each of the two main CCs has accrued a substantial number of specific SNPs (>500), considerably exceeding the number of shared SNPs ( $n = 324$ ). The WGS-based MST clearly separated TUN4.3\_CC1 and TUN4.3\_CC2 into two distinct entities, with  $\geq 245$  SNPs between the two most related isolates (Figure 4B). The pairwise fixation index value of 0.56 ( $>0.33$ )

between the two CCs lent further support to their significant separation, despite the fact that they are evolving in the same demogeographic context. The WGS-based MST also established a phylogenetic link between the singleton isolates and the predominant CC (TUN4.3\_CC1). The mean pairwise genetic distance between pairs of isolates was significantly higher in TUN4.3\_CC2 than in TUN4.3\_CC1 (129 vs 104 SNPs, respectively) (Kruskal–Wallis test,  $P < 0.0001$ ) (Figure





**Figure 4.** Whole-genome sequence analysis of 31 representative *Mycobacterium tuberculosis* L4.3/Latin American and Mediterranean (LAM) isolates from Tunisia. (A) Venn diagram illustrating the number of shared and specific single nucleotide polymorphisms between the two main clonal complexes and singletons. (B) Minimum spanning tree based on the whole-genome sequence of isolates from *M. tuberculosis* TUN4.3\_CC1 (red) and TUN4.3\_CC2 (green), as well as three singletons (grey). (C) Box plots showing pairwise genetic distances (number of polymorphisms) for L4.3/LAM strains of TUN4.3\_CC1 and TUN4.3\_CC2 (*P*-values are from Wilcoxon rank sum tests). Each box corresponds to the 25% and 75% quantiles, the black line represents the median, and the whiskers extend to 1.5 times the interquartile range.

4C). Evidence of recent transmission (difference <10 SNPs between genome pairs) was observed in seven cases (three TUN4.3\_CC1, three TUN4.3\_CC2 and one singleton).

Comparative genome analysis of the two main CCs further revealed a striking difference in terms of evolutionary history. While all TUN4.3\_CC2 isolates harboured the RD115 deletion, TUN4.3\_CC1 isolates were all devoid of this deletion (Figure S1, see online supplementary material). Furthermore, both CCs lacked the prevalent RD-Rio deletion.

## Discussion

The current global picture of the epidemiology of MTBC reflects a strong association with the human host. However, it is obvious that the phylogenetically 'modern' MTBC lineages (L2 and L4) are more successful in terms of geographical spread compared with the 'ancient' lineages (Gagneux, 2012; Comas et al., 2013). Detailed population genetics studies have uncovered the relative success of particular MTBC clones or strains in specific host populations. This is likely to be the consequence of local adaptive evolution, as evidenced by the identification of positively selected mutations in the *M. tuberculosis* genome, and the interactions observed between human genetic polymorphisms and *M. tuberculosis* genotypes (Comas et al., 2013; Brites and Gagneux, 2015; Merker et al., 2015; Brynildsrud et al., 2018; Holt et al., 2018; Chiner-Oms et al., 2019). By performing population genetics and genomics analyses of L4.3/LAM (the predominant sublineage) in Tunisia, this study provides a concrete example of a CC that is particularly fit to the host population.

Analysis of the population structure of L4.3/LAM in Tunisia uncovered two well-differentiated entities: a highly prevalent subpopulation consisting of a single CC (TUN4.3\_CC1); and a second, much less prevalent subpopulation, mainly represented by TUN4.3\_CC2. Overall, phylogenetic analyses indicated a clear differentiation between the two main CCs, and this finding is consistent with the fact that they can be segregated based on RD115 deletion (a hallmark of L4.3/LAM that expanded globally through European colonial migrations) (Mokrousov et al., 2016; Stucki et al., 2016; O'Neill et al., 2019).

Given the history of Tunisia as a former French colony (1881–1956), the introduction of L4.3/LAM could have occurred upon establishment of the French community. Indeed, a considerable influx of TB cases to Tunisia during the French protectorate is very likely as Europe was experiencing the peak of its TB epidemic at that time (Bates and Stead, 1993; Kay et al., 2015). The present-day L4.3/

LAM population in Tunisia may also include newly introduced European L4.3/LAM strains, notably because of intense contemporary migration between the Maghreb and Europe, particularly France, Italy and Germany. As TUN4.3\_CC2 isolates harboured the RD115 deletion but preserved the RD-Rio region, this CC likely represents the 'out-of-Europe' pool of the Tunisian L4.3/LAM population.

However, given the very likely African origin of L4 (O'Neill et al., 2019), one should not dismiss the possibility that a subset of L4.3/LAM strains may have originated locally or, as suggested previously (O'Neill et al., 2019), may have been imported to the Maghreb through regional land routes linking Eastern, Western and Northern Africa, particularly during the era of Roman conquest. In this respect, an investigation addressing the geospatial distribution of *M. tuberculosis* genotypes in Africa revealed the Eastern Mediterranean origin of L4.3/LAM prevailing in Northern Africa (Chihota et al., 2018). This finding is consistent with a previous study which assigned a Maghrebian origin for the L4.3/LAM sublineage (Mokrousov et al., 2016). This indigenous origin is very plausible for TUN4.3\_CC1, the most prevalent L4.3/LAM CC in Tunisia and the Maghreb. Indeed, TUN4.3\_CC1 displayed an intact RD115 region, the deletion of which is shared by virtually all L4.3/LAM strains that accompanied the historical migration out of Europe. As RD115 deletion is the most ancestral event in the expansion history of L4.3/LAM, its absence in TUN4.3\_CC1, along with the retention of the RD-Rio region, indicates its early origin relative to TUN4.3\_CC2, and hence its closest relationship with the progenitor of L4.3/LAM. On the other hand, TUN4.3\_CC1 tended to exhibit higher mean allelic richness than TUN4.3\_CC2; this finding is indicative of a long period of diversification. Strikingly, by considering those isolates with known spoligotype patterns, TUN4.3\_CC2 was found to be comprised predominantly of the LAM9 genotype (92.5%), while the more prevalent TUN4.3\_CC1 was much more diverse, consisting of LAM9 (55.68%), LAM4 (31.81%), LAM5 (7.95%) and LAM3 (4.54%) (data not shown). This spoligotypic diversity of TUN4.3\_CC1 strongly suggests that it may have been evolving locally for a longer period compared with TUN4.3\_CC2.

TUN4.3\_CC1 appears to be particularly fit to the Tunisian population. Indeed, aside from being highly prevalent and long established, this CC expanded more rapidly in the host population than TUN4.3\_CC2, as revealed by the lower mean pairwise genetic distance between genomes. Moreover, TUN4.3\_CC2 was not particularly associated with drug resistance. Therefore, its success in the host population reflects an intrinsic ability to spread, rather

than a consequence of increased fitness due to the compensation of costs of drug resistance.

In summary, this study explored the evolutionary history of L4.3/LAM in Tunisia. Globally, this is the most widespread L4 sublineage of *M. tuberculosis*, and it is the most successful in the Maghreb. The success of this sublineage in Tunisia was found to stem from a single, long-established CC which appears to be intrinsically fit to the local host population. Identification of this CC will certainly further our knowledge on processes leading to the emergence of highly fit *M. tuberculosis* clones.

## Funding

Tunisian Ministry of Higher Education and Scientific Research (Grant LR16IPT01).

## Ethical approval

No interventions were performed specifically for this study. DNA was obtained from the strain collection of the Unit of Typing and Genetics of Mycobacteria of the Institut Pasteur, Tunis, as part of its routine diagnostic activity permitted by the Tunisian Ministry of Health, in the context of the national TB programme. Only fully anonymized data were processed, and hence no further ethical clearance was required.

## Conflict of interest

None declared.

## Data availability statement

General characteristics and genotypic data (24-loci MIRU-VNTR) of the 252 Tunisian L4.3/LAM strains included in this study are provided in Table S1 (see online supplementary material). The WGS of the 31 L4.3/LAM Tunisian isolates has been deposited in GenBank (Accession No. PRJEB39509).

## Authors' contributions

HM conceived and designed this study. MAS, ND, BM and SBF performed culture of *M. tuberculosis* isolates and isolation of genomic DNA. MAS and ND carried out 24-loci MIRU-VNTR. MAS, ND and HM performed genomic analyses, interpreted the data and wrote the manuscript. All authors reviewed and approved the manuscript.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijid.2020.11.195>.

## References

- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19: 1655–64.
- Bates JH, Stead WW. The history of tuberculosis as a global epidemic. *Med Clin N Am* 1993;77:1205–17.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
- Bouklata N, Supply P, Jaouhari S, Charof R, Seghrouchni F, Sadki K, et al. Molecular typing of Mycobacterium tuberculosis complex by 24-locus based MIRU-VNTR typing in conjunction with spoligotyping to assess genetic diversity of strains circulating in Morocco. *PLoS One* 2015;10:e0135695.
- Brites D, Gagneux S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. *Immunol Rev* 2015;264:6–24.
- Brites D, Gagneux S. The nature and evolution of genomic diversity in the Mycobacterium tuberculosis complex. *Adv Exp Med Biol* 2017;1019:1–26.
- Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *PNAS* 2002;99:3684–9.

- Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. *Sci Adv* 2018;4: eaat5869.
- Chihota VN, Niehaus A, Streicher EM, Wang X, Sampson SL, Mason P, et al. Geospatial distribution of Mycobacterium tuberculosis genotypes in Africa. *PLoS One* 2018;13:e0200632.
- Chiner-Oms Á, Sánchez-Busó L, Corander J, Gagneux S, Harris SR, Young D, et al. Genomic determinants of speciation and spread of the Mycobacterium tuberculosis complex. *Sci Adv* 2019;5: eaaw3307.
- Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNerney R, et al. SpolPred: rapid and accurate prediction of Mycobacterium tuberculosis spoligotypes from short genomic sequences. *Bioinformatics* 2012;28:2991–3.
- Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic co-expansion of Mycobacterium tuberculosis with modern humans. *Nat Genet* 2013;45:1176–82.
- Coscolla M, Gagneux S. Consequences of genomic diversity in Mycobacterium tuberculosis. *Semin Immunol* 2014;26:431–44.
- Dekhil N, Meftahi N, Mhenni B, Ben Fraj S, Haltiti R, Belhaj S, et al. MDR-TB outbreak among HIV-negative Tunisian patients followed during 11 years. *PLoS One* 2016;11:e0153983.
- Eldholm V, Rieux A, Monteserin J, Lopez JM, Palmero D, Lopez B, et al. Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. *eLife* 2016;09:5.
- Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Res* 2017;17:27–32.
- Gagneux S. Host–pathogen coevolution in human tuberculosis. *Philos Trans R Soc Lond B Biol Sci* 2012;367:850–9.
- Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol* 2018;16:202–13.
- Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, et al. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat Genet* 2018;50:849–56.
- Kalinowski ST. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 2005;5:187–9.
- Kay GL, Sergeant MJ, Zhou Z, Chan JZ-M, Millard A, Quick J, et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat Commun* 2015;6:6717.
- Ke W, Lp W, Jr L, Jm R, R Z-C, A G, et al. Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC Med* 2018;16:196.
- Kent PT, Kubica GP. Public Health Mycobacteriology. A Guide for the Level III Laboratory. Atlanta, GA: Centers for Disease Control, US Department of Health and Human Services; 1985.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 2015;15:1179–91.
- Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList – 10 years after. *Tuberculosis* 2011;91:1–7.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- Marais BJ, Victor TC, Hesselink AC, Barnard M, Jordaán A, Brittle W, et al. Beijing and Haarlem genotypes are overrepresented among children with drug-resistant tuberculosis in the Western Cape Province of South Africa. *J Clin Microbiol* 2006;44:3539–43.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nat Genet* 2015;47:242–9.
- Mokrousov I, Vyazovaya A, Iwamoto T, Skiba Y, Pole I, Zhdanova S, et al. Latin-American-Mediterranean lineage of Mycobacterium tuberculosis: human traces across pathogen's phylogeography. *Mol Phylogenet Evol* 2016;99:133–43.
- Namouchi A, Karboul A, Mhenni B, Khabouchi N, Haltiti R, Ben Hassine R, et al. Genetic profiling of Mycobacterium tuberculosis in Tunisia: predominance and evidence for the establishment of a few genotypes. *J Med Microbiol* 2008;57:864–72.
- O'Neill MB, Shockey A, Zarley A, Aylward W, Eldholm V, Kitchen A, et al. Lineage specific histories of Mycobacterium tuberculosis dispersal in Africa and Eurasia. *Mol Ecol* 2019;28:3241–56.
- Sampson SL, Warren RM, Richardson M, van der Spuy GD, van Helden PD. Disruption of coding regions by IS6110 insertion in Mycobacterium tuberculosis. *Tuber Lung Dis* 1999;79:349–59.
- Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* 2016;48:1535–43.
- 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. *J Clin Microbiol* 2006;44:4498–510.
- World Health Organization. Global Tuberculosis Report 2019. Geneva: WHO; 2019 Available at: <https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714-eng.pdf?> [Accessed 30 July 2020].