



OPEN

# Whole genome sequencing of clinical samples reveals extensively drug resistant tuberculosis (XDR TB) strains from the Beijing lineage in Nigeria, West Africa

Idowu B. Olawoye<sup>1,2</sup>, Jessica N. Uwanibe<sup>1,2</sup>, Chioma N. Kunle-Ope<sup>3</sup>, Olabisi F. Davies-Bolorunduro<sup>3</sup>, Temitope A. Abiodun<sup>3</sup>, Rosemary A. Audu<sup>3</sup>, Babatunde L. Salako<sup>3</sup> & Christian T. Happi<sup>2,1</sup>✉

Multi-drug (MDR) and extensively drug-resistant (XDR) tuberculosis (TB) continues to be a global public health problem especially in high TB burden countries like Nigeria. Many of these cases are undetected and go on to infect high risk individuals. Clinical samples from positive rifampicin resistant Xpert®MTB/Rif assay were subjected to direct whole genome sequencing and bioinformatics analysis to identify the full antibiotics resistance and lineage profile. We report two (2) XDR TB samples also belonging to the East-Asian/Beijing family of lineage 2 *Mycobacterium tuberculosis* complex from clinical samples in Nigeria. Our findings further reveal the presence of mutations that confer resistance to first-line drugs (rifampicin, isoniazid, ethambutol and pyrazamide), second-line injectables (capreomycin, streptomycin, kanamycin and/or amikacin) and at least one of the fluoroquinolones (ofloxacin, moxifloxacin, levofloxacin and/or ciprofloxacin) in both samples. The genomic sequence data from this study not only provide the first evidence of XDR TB in Nigeria and West Africa, but also emphasize the importance of WGS in accurately detecting MDR and XDR TB, to ensure adequate and proper management treatment regimens for affected individuals. This will greatly aid in preventing the spread of drug resistance TB in high burden countries.

*Mycobacterium tuberculosis* complex (MTBC) which is responsible for tuberculosis (TB) remains a global public health concern especially in low-income countries as it is the leading cause of death from a bacterial infectious disease<sup>1</sup>. In 2019, it was estimated that ten million people worldwide contracted the tubercle disease and about 1.2 million (HIV negative) of these people died as a result of complications that arose from TB<sup>1</sup>. MTBC is grouped into different lineages, known as *M. tuberculosis* sensu stricto (lineage 1, 2, 3, 4 and, 7), *M. africanum* (lineage 5 and 6), the recently discovered lineages 8, 9<sup>2,3</sup>, and, several other animal-associated habitat (*M. bovis*, *M. microti*, *M. pinnipedii*, *M. orygis* and, others)<sup>4</sup>. Some of these lineages are geographically restricted, such as lineages 5, 6, 7 and, 9<sup>2,5,6</sup>, whilst some are globally distributed such as lineages 1, 2, 3 and, 4<sup>7</sup>.

Lineage 2, also known as the Beijing lineage is associated with the increased spread of multi-drug resistant TB in Eurasia, and has since spread across the world in successive waves due to human migration<sup>8</sup>. Genome analysis have reported multiple introductions of the Beijing lineage in Africa dating back to 300 years<sup>9</sup>. In addition, the Beijing lineage have been reported in the West African region, using spoligotyping and whole genome sequencing<sup>10–12</sup>, but with no report in Nigeria yet using molecular techniques<sup>13,14</sup>.

<sup>1</sup>African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Osun State, Nigeria. <sup>2</sup>Department of Biological Sciences, Faculty of Natural Sciences, Redeemer's University, Ede, Osun State, Nigeria. <sup>3</sup>Centre for Tuberculosis Research (CTBR), Microbiology Department, Nigerian Institute of Medical Research (NIMR), Yaba, Lagos State, Nigeria. ✉email: happi@run.edu.ng

	Sex	Age	HIV Status	INH	RIF	FLQ	KAN/AMK/CAP/VAN	Low level KAN	Treatment type
1	Male	48	Negative	S	R	S	S	S	New TB case
2	Male	26	Negative	S	R	S	S	S	New TB case
3	Male	34	Negative	S	R	S	S	S	New TB case
4	Male	45	Negative	S	R	S	S	S	New TB case
5	Male	49	Negative	S	R	S	S	S	New TB case
6	Female	NA	Positive	S	R	S	S	S	New TB case
7	Male	25	Negative	S	R	S	S	S	New TB case
8	Female	31	Positive	R	R	S	S	S	New TB case
9	Male	23	Negative	S	R	S	S	S	New TB case
10	Male	20	Negative	R	R	S	S	S	New TB case

**Table 1.** Patient demography and rapid diagnostic molecular drug resistance outcome. NA not available, S sensitive, R resistant, FLQ fluoroquinolones, KAN kanamycin, AMK amikacin, CAP capreomycin, VAN vancomycin.

Multi-drug resistant TB (MDR TB) is the type of TB bacteria that is resistant to two of the most important drugs to treat TB also known as first-line drugs: isoniazid (INH) and rifampicin (RIF). The updated definition of extensively drug-resistant TB (XDR TB) is a rare type of MDR TB that is resistant to INH, RIF and any fluoroquinolone, and at least one additional Group A drug (levofloxacin, moxifloxacin, bedaquiline and/or, linezolid) which are the most powerful second-line drugs used for drug resistant TB management<sup>15</sup>. The emergence of drug resistant and extensively drug resistant MTBC worldwide has largely affected public health and global economy, as many cases go unreported or undetected in many low resource settings and cause further spread and misdiagnosis<sup>16</sup>. In 2019, the incidence of MDR TB resistant to rifampicin (RR-TB) in Nigeria was estimated at 21,000 with 16 laboratory confirmed XDR TB cases, with reports that about 2% of global MDR TB cases acquire multiple resistance to antibiotics to become XDR TB<sup>1</sup>. This statistics may be underplayed as many high burden TB countries do not have diagnostic capacities to capture the real scenario, as such it is crucial to use cutting edge technologies to screen drug-resistant (DR) or MDR TB samples for the presence of XDR acquiring mutations in these countries<sup>17</sup>.

Just as many other West African countries, Nigeria has deployed rapid molecular techniques to detect RR-TB and MDR TB by using Xpert<sup>®</sup>MTB/Rif assay (Cepheid, USA) and GenoType MTDRplus Line Probe Assay (Hain Lifescience GmbH, Germany)<sup>18</sup>. However, these techniques come with limitations when compared to whole genome sequencing (WGS), as the latter can detect drug resistant mutations outside the target zone of rapid tests and produce much more reliable and robust drug resistance profile of TB<sup>19</sup>.

In this study, we use whole genome sequencing to investigate TB positive clinical samples that were confirmed RR-TB with Xpert<sup>®</sup>MTB/Rif assay (Cepheid, USA). We also use bioinformatics to classify the bacteria into lineages in order to identify the strains or clonal complexes present in Nigeria.

## Results

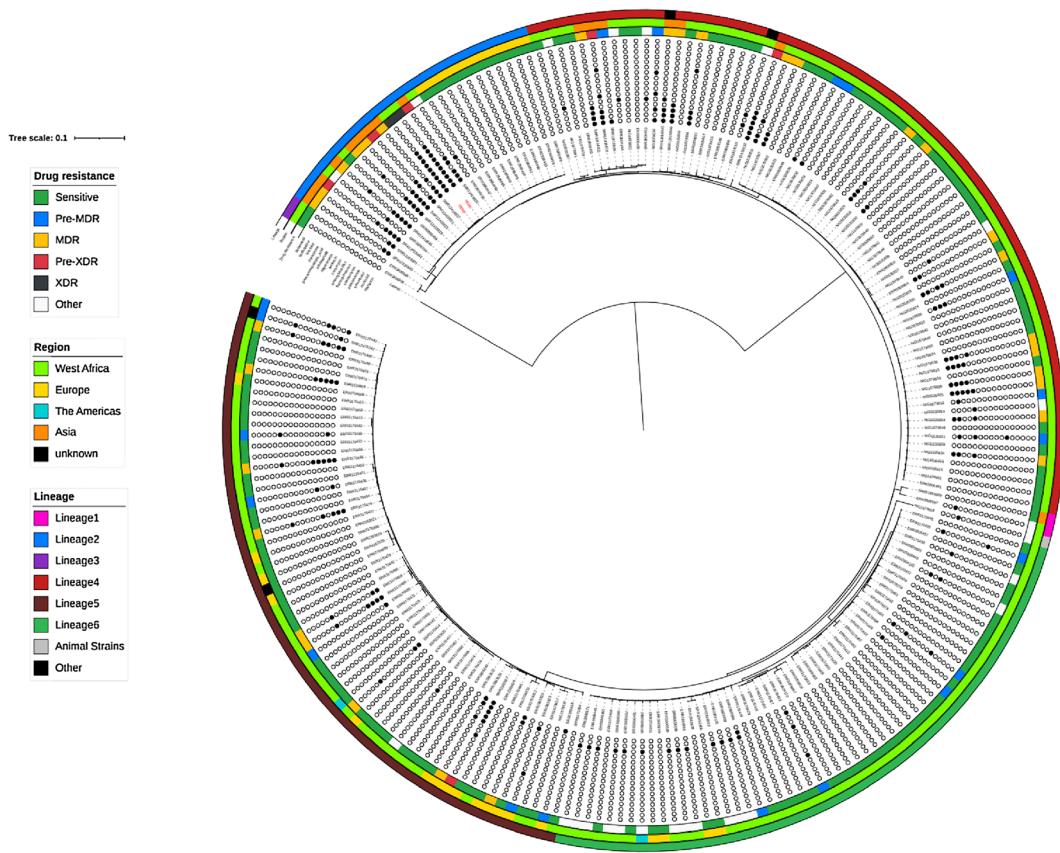
**Patient demography and molecular drug resistance outcome.** Out of ten (10) de-identified patients, 20% (2/10) were females and 20% (2/10) tested positive to HIV. The ages of the patients were between 20 and 49 years. All samples (100%) tested positive to resistance to RIF using Xpert<sup>®</sup>MTB/Rif assay and 20% tested positive to resistance to INH and none of the samples tested positive to fluoroquinolones (FLQ) or second-line drug injectables (kanamycin, amikacin, capreomycin and/or, vancomycin) using Hain Lifescience GenoType MTBDRplus and MTBDRsI line probe assays (Table 1).

**Whole genome sequencing and quality control.** Out of 10 samples that were quantified using Qubit and Agilent 2100 Bioanalyzer, only 4 libraries were prioritized for whole genome sequencing as they had high concentration of DNA that passed the loading concentration on the Illumina iSeq 100. Two (2) out of four (4) samples passed the 20 $\times$  average depth, and 220 genomes downloaded from SRA passed the quality control (QC) threshold used for all samples. The average depth of the two newly sequenced TB samples from Nigeria (TB3qc and TB8qc) were 71.66 $\times$  and 63.59 $\times$  using SAMtools. The average depth of the other SRA genomes in this study were within  $\pm 40\times$  of the two newly sequenced samples.

**Phylogenetic analysis and in-silico drug resistance profiling of TB samples.** Geographical location of all MTBC genomes were retrieved from SRA and grouped into: The Americas, Europe, Asia and, West Africa. Phylogenetic analysis of core genomes containing SNPs of all samples in this study that passed QC steps showed that TB3qc and TB8qc clustered alongside other global lineage 2 genomes as seen in Fig. 1.

Drug resistance profiling of TB3qc and TB8qc classified them as XDR TB and lineage classification as Lineage 2.2.1 also known as Beijing or East-Asian lineage. Analysis of drug mutations associated with resistance to anti-TB drugs show that these samples have mutations against all first and second line drug classes (Table 2).

Full drug resistance profile of TB3qc and TB8qc from TB-profiler and Mykrobe predictor in CSV and JSON files containing genome position, locus tag, estimated fraction, coverage of genotype and, other mutations



**Figure 1.** Maximum likelihood phylogeny tree of MTBC samples inferred from their core genomes with lineage classification, region where samples were collected and drug resistance profile. The samples coloured red are genomes generated from this study.

associated with drug resistance are available in Supplementary Data 1, 2, 3 and 4 such as P239L and L707I substitution mutations in *katG*.

TB3qc and TB8qc drug susceptibility phenotypic test outcome classified them as drug resistant and pre-XDR respectively and the patients treatment outcome was followed up from the Lagos State MDR TB treatment centre (Table 3).

## Discussion

In this study, we were able to use next generation sequencing (NGS) technology to identify the presence of the Beijing family of MTBC lineage 2 and extensively drug-resistant tuberculosis in Nigeria for the first time. Maximum likelihood phylogeny from variable regions in core genome alignments and also using characteristic regions of differences such as the presence of RD105, RD207 and, RD181 in the genome grouped them as lineage 2.2.1. In addition, these genomes were classified as extensively drug-resistant TB (XDR TB) as they have defining mutations that confer resistance to all first-line drugs (isoniazid, rifampicin, ethambutol and, pyrazinamide), at least one second-line anti-TB injectables or aminoglycosides (capreomycin, streptomycin, kanamycin and/or, amikacin) and at least one fluoroquinolone (ofloxacin, moxifloxacin, levofloxacin and/or, ciprofloxacin) (Table 2). Resistance R-value below 0.3 for drug classes analyzed with Resistance Sniffer implies that the strain has higher likelihood to be sensitive to said drug, with number of resistance sites present in the genome in square brackets (Table 2).

The presence of the Beijing lineage in Nigeria is no surprise as it has been detected in Benin Republic, which shares the western border with Nigeria<sup>10</sup> and more so, it may have resulted from the continuous increase of economic relations between Nigeria and China over the past decades<sup>20</sup>. The East-Asian or Beijing lineage of MTBC is well known for multi-drug resistance, increased virulence and, fast disease progression unlike other MTBC lineages<sup>8,9</sup>, with another study seeing a level of association with HIV<sup>21</sup>.

In our findings, one of the patients who tested positive to HIV was diagnosed as MDR TB whilst the other was diagnosed as drug-resistant TB as both samples tested negative to second-line drugs using rapid molecular tests which many TB labs in Nigeria solely rely on. However, both samples were confirmed to be XDR TB with whole genome sequence analysis. The reason for this disparity is that some false negative results can be produced by the MTBDRsI strip-based assay as some second-line drug mutations may not be built in the panel which can be detected using whole genome sequence data by screening mutations that are not part of the LPA as seen in

		Sample ID	
		TB3qc	TB8qc
<b>Lineage, Family and Spoligotype</b>		Lineage 2.2.1; East-Asian (Beijing)	Lineage 2.2.1; East-Asian (Beijing)
<b>Region of difference (RD)</b>		RD105, RD207 and RD181	RD105, RD207 and RD181
<b>Drug resistance group</b>		XDR TB	XDR TB
<b>Drug genotypic resistance (TB-Profiler)</b>			
Rifampicin	<i>rpoB</i>	S450L	S450L
Isoniazid	<i>katG</i>	S315T, c1561 del	S315T
Pyrazinamide	<i>pncA</i>	D12G	D12G
Ethambutol	<i>embB</i>	Q497R	Q497R
Streptomycin	<i>rpsL</i>	K43R	K43R
	<i>rrs</i>	c799t	c799t, g888a
Fluoroquinolones	<i>gyrA</i>	A90V, S91P	A90V, S91P
Amikacin	<i>rrs</i>	c1402a	c1402a, r1484t
Kanamycin	<i>rrs</i>	c1402a	c1402a, r1484t
Capreomycin	<i>rrs</i>	c1402a	c1402a, r1484t
Cycloserine	<i>ald</i>	—	—
Bedaquiline	<i>Rv0678</i>	—	—
Delamanid	<i>ddn</i>	—	—
Clofazimine	<i>Rv0678</i>	—	—
<b>Drug genotypic resistance (Mykrobe)</b>			
Amikacin	<i>rrs</i>	—	—
Capreomycin	<i>rrs</i>	C1402X	—
Ciprofloxacin	<i>gyrA</i>	A90V, S91P	A90V, S91P
Ethambutol	<i>embB</i>	Q497R	Q497R
Isoniazid	<i>katG</i>	S315G	S315G
Kanamycin	<i>rrs</i>	—	—
Moxifloxacin	<i>gyrA</i>	A90V, S91P	A90V, S91P
Ofloxacin	<i>gyrA</i>	A90V, S91P	A90V, S91P
Pyrazinamide	<i>pncA</i>	D12G	D12G
Rifampicin	<i>rpoB</i>	S450L	S450L
Streptomycin	<i>rpsL</i>	K43R	K43R
<b>Drug genotypic resistance (MTB Resistance Sniffer)</b>			
Amikacin		0.25 +/ - 0.06 [26 sites]	0.25 +/ - 0.06 [26 sites]
Capreomycin		0.16 +/ - 0.04 [27 sites]	0.16 +/ - 0.04 [27 sites]
Cycloserin		0.06 +/ - 0.02 [22 sites]	0.06 +/ - 0.02 [22 sites]
Ethambutol		0.17 +/ - 0.05 [26 sites]	0.17 +/ - 0.05 [26 sites]
Ethionamide		0.29 +/ - 0.08 [23 sites]	0.28 +/ - 0.08 [21 sites]
Floroquinolones		0.41 +/ - 0.07 [43 sites]	0.42 +/ - 0.07 [40 sites]
Isoniazid		0.81 +/ - 0.1 [26 sites]	0.81 +/ - 0.1 [26 sites]
Kanamycin		0.79 +/ - 0.12 [25 sites]	0.73 +/ - 0.12 [22 sites]
Ofloxacin		0.27 +/ - 0.07 [26 sites]	0.27 +/ - 0.07 [26 sites]
Para-amino salicylic acid		0.48 +/ - 0.1 [28 sites]	0.5 +/ - 0.11 [27 sites]
Pyrazinamide		0.27 +/ - 0.07 [26 sites]	0.28 +/ - 0.07 [25 sites]
Rifampicin		0.41 +/ - 0.08 [26 sites]	0.41 +/ - 0.08 [26 sites]
Streptomycin		0.92 +/ - 0.09 [26 sites]	0.92 +/ - 0.09 [26 sites]

**Table 2.** In-silico drug-resistant analysis of MTBC samples and Lineage report.

Feliciano et al.<sup>22</sup>. After WGS, phenotypic drug susceptibility test (DST) was done to correlate the results from rapid molecular tests and TB3qc still remained drug resistant TB, while TB8qc tested as pre-XDR TB.

This infers that a number of XDR TB cases will go undetected using LPAs as we can see in the Mykrobe predictor JSON output file for TB3qc (Supplementary Data 2) where the allele responsible for resistance to isoniazid has 100% coverage and a median depth of 39 and R-value of 0.81 from Resistance Sniffer (Table 2). However, this sample tested sensitive to isoniazid through LPA and DST test. Other advantages of WGS is that new drug-resistant mutations can also be detected across the entire genome and not just target genes as seen under ‘other variants report’ section in TB-Profiler CSV file outputs of TB3qc and TB8qc (Supplementary Data 1 and 3) where other drug resistance mutations present in the genomes are listed. Furthermore, antimicrobial

	Smear result	Culture result		First line drug			Second line drug					Treatment outcome
		Solid	Liquid	INH	RIF	EMB	LEV	MOX	KAN	AMK	CAP	
TB3qc	3+	POS	POS	S	R	S	S	S	S	S	S	Stopped treatment
TB8qc	2+	POS	POS	R	R	R	R	S	S	S	DEAD	

**Table 3.** Laboratory results, phenotypic drug sensitivity test and treatment outcome from the two sequenced samples. *POS* positive, *INH* isoniazid, *RIF* rifampicin, *EMB* ethambutol, *LEV* levofloxacin, *MOX* moxifloxacin, *KAN* kanamycin, *AMK* amikacin, *CAP* capreomycin.

resistance (AMR) tools used for genome analysis are constantly being improved, such as one of the software used in this research which was developed with a drug resistance database extracted from over 17,000 MTBC strains with their phenotypic and genotypic data<sup>23</sup>.

Following up on the patients treatment outcome from the MDR TB treatment centre in Lagos State revealed that the patient who tested positive to HIV and had an XDR TB died as a result of complications from both infections, while the other patient stopped treatment due to severe side effects arising from the treatment regimen. This patient can potentially go on to infect others with this highly infectious and drug resistant strain of TB.

The major discovery in this study is that we were able to identify the presence of the Beijing family of lineage 2 MTBC in Nigeria for the first time using WGS from direct clinical samples. This shows that there is more genetic diversity of MTBC in Nigeria than presently known. In addition to that, we were also able to characterize two (2) XDR TB from drug-resistant and multi-drug resistant TB strains that were initially tested using LPA and Xpert<sup>®</sup>MTB/Rif assays. This shows the high resolution of WGS for AMR detection in clinical diagnosis and public health. In addition, WGS analysis show that the XDR TB samples detected in this study are sensitive to new and repurposed drugs such as delamanid, bedaquiline and cycloserine, which implies that they can be managed with either of these.

With the falling cost of NGS technologies nowadays, WGS has the potential to bolster clinical diagnosis of TB in low-to-middle income countries and high TB burden regions as it will greatly improve the accuracy of antimicrobial resistance (AMR) detection.

Although this study is limited by the small sample size, it provide new insights into TB genetic diversity in Nigeria and West Africa, despite budgetary restrictions to perform WGS.

## Conclusion

In this study, we show the robustness of WGS in public health and diagnosis of drug resistance tuberculosis. The data generated from this work shows how extensively drug resistance could be missed when using routine rapid molecular diagnosis assays only. Bioinformatics analysis can reveal sensitivity or resistance to new drugs by checking for the presence or absence of mutations responsible for such phenotypic reactions. We conclude that the integration of NGS platforms into the established, but siloed, clinical diagnosis, to support the detection of extensively drug resistant tuberculosis in high burden countries, will not only help in suggesting treatment options to MDR/XDR TB patients, but will also provide exciting opportunities for public health surveillance.

## Methods

**Sample collection, DNA extraction and Molecular drug resistance assay.** Random sputum samples ( $n=10$ ) that were collected in the month of July, 2019 from patients who were referred for diagnosis at the Centre of Tuberculosis Research (CTBR), a TB reference laboratory in Lagos State. Patients who had severe TB symptoms and tested positive with Xpert<sup>®</sup>MTB/RIF (Cepheid, USA) to detect the presence of MTBC and resistance to RIF were recruited for this study. These samples were further subjected to DNA extraction using the GenoLyse kit (Hain Lifescience GmbH, Germany). The genomic bacterial DNA further underwent Line Probe Assay (LPA) using GenoType MTBDRplus version 2.0 (Hain Lifescience GmbH, Germany) to detect samples resistant to RIF and INH. Second-line drug resistance molecular test was done using GenoType MTBDRsl VER 2.0 (Hain Lifescience GmbH, Germany) strip-based assay. Samples of interest were cultured on Lowenstein Jensen (LJ) media and Middlebrook 7H9 broth using BACTEC<sup>™</sup> MGIT machine (BD, Erembodegem, Belgium) and tested using proportion method for phenotypic drug susceptibility with first and second line drugs.

These molecular processes were carried out at CTBR, Microbiology Department, Nigerian Institute of Medical Research (NIMR), Lagos.

**Library preparation and whole genome sequencing.** DNA extracts were quantified using Qubit fluorometer (ThermoFisher Scientific) using the dsDNA High sensitivity assay and diluted to 0.2 ng/μl. Sequencing libraries were prepared using Nextera XT DNA preparation kit (Illumina, USA) and libraries were quantified with Qubit fluorometer and the library length was estimated using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Library preparation protocol was adopted from the US CDC PulseNet SOP<sup>24</sup>. Individual libraries were loaded on the Illumina iSeq 100 (Illumina, USA) with 151 cycles to generate paired end reads using the manufacturer's instructions (Illumina, USA). TB genome sequencing was done at the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Nigeria.

**Bioinformatics analysis.** The paired-end reads were demultiplexed using the GenerateFASTQ module version 2.0 on the iSeq 100 Software System Suite local run manager. 230 MTBC genomes were retrieved from NCBI SRA using sra-toolkit v2.10.9 (<https://github.com/ncbi/sra-tools>) with paired-end reads and Illumina platform as the selection criteria from the following BioProjects: PRJEB15857, PRJEB25506, PRJNA300846, PRJEB28842, PRJNA633244, PRJNA480117, PRJEB27244, PRJEB36076, PRJNA534674 and, PRJNA655747.

All reads were processed with the Bacterial Genome Pipeline (BAGEP)<sup>25</sup>, which does quality control on raw reads, taxonomic classification and variants detection by mapping reads to the reconstructed ancestral MTBC sequence<sup>26</sup> with minimum base quality set at 20, minimum site depth for calling alleles set at 10 (default value) and also excluding duplicate reads. Average coverage of the genomes were calculated using SAMtools v1.12<sup>27</sup> and genomes that are  $\leq 20\times$  was excluded from further downstream analysis.

**Drug resistance prediction and lineage classification.** TBProfiler v3.0.4<sup>23</sup> was used to analyze all 222 genomes in-silico to infer phylogenetic lineages and drug resistance profiles using small variants and big deletions associated with drug resistance that are present in a robust library of recently updated mutation database which contains new anti-TB drugs such as cycloserine, delamanid, bedaquiline, clofazimine, para-aminosalicylic acid, linezolid and ethionamide [commit b2af444 on 21st December 2020 (<https://github.com/jodyphelan/tbdb>)] using freebayes<sup>28</sup> as the variant calling option. In order to validate the results, Mykrobe predictor<sup>29</sup> and Resistance Sniffer<sup>30</sup> was used to analyze the newly sequenced Nigerian samples for comparison.

**Phylogenetic reconstruction.** Outputs from all the variant call analysis with Snippy were combined to produce core genome alignment of all 223 samples (including *Mycobacterium canettii* NC\_015848.1) were generated and aligned using Snippy-core (<https://github.com/tseemann/snippy>) by excluding regions such as PPE, PE-PGRS, insertion sequences, phages, repeat sequences and, regions that are at least 50 bp long. The core genome alignment containing polymorphisms was used to infer a maximum-likelihood phylogenetic tree using IQTREE<sup>31</sup> v1.6.12 general time reversible model (GTR) and ultrafast bootstrap value of 10,000. The phylogenetic tree was rooted to *M. canettii* and annotated with Interactive tree of life (iTOL) v5<sup>32</sup>.

**Ethical approval.** The study was performed in accordance with the Declaration of Helsinki, and the protocol was approved by the ethical review boards of Redeemer's University (Osun State, Nigeria) (RUN-IREC/19/012) and Nigerian Institute of Medical Research (Lagos State, Nigeria) (IRB/19/053). Sputum samples used for this study were obtained under a waiver of consent granted by the Institutional Review Board of the Nigerian Institute of Medical Research (ref: IRB/19/053).

## Data availability

Raw sequence data generated in this study are available from the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the Project Accession No. PRJNA725394 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA725394/>).

Received: 2 July 2021; Accepted: 17 August 2021

Published online: 30 August 2021

## References

1. Global tuberculosis report 2020. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO (2020).
2. Coscolla, M. *et al.* Phylogenomics of *mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb. Genomics* **7**, 1–14 (2021).
3. Ngabonziza, J. C. S. *et al.* A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat. Commun.* **11**, 1–11 (2020).
4. Brites, D. *et al.* A new phylogenetic framework for the animal-adapted *mycobacterium tuberculosis* complex. *Front. Microbiol.* **9**, 2820 (2018).
5. de Jong, B. C., Antonio, M. & Gagneux, S. *Mycobacterium africanum*-review of an important cause of human tuberculosis in West Africa. *PLoS Negl. Trop. Dis.* **4**, e744 (2010).
6. Blouin, Y. *et al.* Significance of the identification in the horn of Africa of an exceptionally deep branching *mycobacterium tuberculosis* clade. *PLoS ONE* **7**, e52841 (2012).
7. Hershberg, R. *et al.* High functional diversity in *mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* **6**, e311 (2008).
8. Merker, M. *et al.* Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* **47**, 242–249 (2015).
9. Rutaihwa, L. K. *et al.* Multiple introductions of *mycobacterium tuberculosis* lineage 2–Beijing into Africa over centuries. *Front. Ecol. Evol.* <https://doi.org/10.3389/fevo.2019.00112> (2019).
10. Affolabi, D. *et al.* First molecular epidemiological study of tuberculosis in Benin. *Int. J. Tuberc. Lung Dis.* **13**, 317–322 (2009).
11. Ofori-Anyinam, B. *et al.* Comparative genomics shows differences in the electron transport and carbon metabolic pathways of *Mycobacterium africanum* relative to *Mycobacterium tuberculosis* and suggests an adaptation to low oxygen tension. *Tuberculosis* **120**, 101899 (2020).
12. Senghore, M. *et al.* Evolution of *Mycobacterium tuberculosis* complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. *Sci. Rep.* **10**, 1–9 (2020).
13. Cadmus, S., Hill, V., Van Soolingen, D. & Rastogi, N. Spoligotype profile of *Mycobacterium tuberculosis* complex strains from HIV-positive and -negative patients in Nigeria: A comparative analysis. *J. Clin. Microbiol.* **49**, 220–226 (2011).
14. Adesokan, H. K., Streicher, E. M., van Helden, P. D., Warren, R. M. & Cadmus, S. I. B. Genetic diversity of *Mycobacterium tuberculosis* complex strains isolated from livestock workers and cattle in Nigeria. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0211637> (2019).
15. World Health Organization. WHO announces updated definitions of extensively drug-resistant tuberculosis. <https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis> (2021).

16. Chisompola, N. K., Streicher, E. M., Muchemwa, C. M. K., Warren, R. M. & Sampson, S. L. Molecular epidemiology of drug resistant *Mycobacterium tuberculosis* in Africa: A systematic review. *BMC Infect. Dis.* **20**, 344 (2020).
17. Velayati, A. A., Farnia, P. & Farahbod, A. M. Overview of drug-resistant tuberculosis worldwide. *Int. J. Mycobacteriol.* **5**, S161 (2016).
18. Adejumo, O. A. *et al.* The pattern of comorbidity and its prevalence among drug-resistant tuberculosis patients at treatment initiation in Lagos, Nigeria. *Trans. R. Soc. Trop. Med. Hyg.* **114**, 415–423 (2020).
19. Katale, B. Z. *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* isolates and clinical outcomes of patients treated for multidrug-resistant tuberculosis in Tanzania. *BMC Genomics* **21**, 174 (2020).
20. Oyeranti, G. A., Babatunde, M. A. & Ogunkola, E. O. An analysis of China-Nigeria investment relations. *J. Chin. Econ. Foreign Trade Stud.* **4**, 183–199 (2011).
21. Caws, M. *et al.* Beijing genotype of *mycobacterium tuberculosis* is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. *J. Clin. Microbiol.* **44**, 3934–3939 (2006).
22. Feliciano, C. S. *et al.* Accuracy of whole genome sequencing versus phenotypic (MGIT) and commercial molecular tests for detection of drug-resistant *Mycobacterium tuberculosis* isolated from patients in Brazil and Mozambique. *Tuberculosis* **110**, 59–67 (2018).
23. Phelan, J. E. *et al.* Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med.* **11**, 41 (2019).
24. CDC. Whole Genome Sequencing (WGS) | PulseNet Methods | PulseNet | CDC. <https://www.cdc.gov/pulsenet/pathogens/wgs.html> (2016).
25. Olawoye, I. B., Frost, S. D. W. & Happi, C. T. The Bacteria Genome Pipeline (BAGEP): An automated, scalable workflow for bacteria genomes with Snakemake. *PeerJ* **8**, e10121 (2020).
26. Comas, I. *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* **42**, 498–503 (2010).
27. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
28. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv Prepr. ArXiv1207.3907* (2012).
29. Hunt, M. *et al.* Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res.* **4**, 191. <https://doi.org/10.12688/wellcomeopenres.15603.1> (2019).
30. Muzondwa, D., Mutshembele, A., Pierneef, R. E. & Reva, O. N. Resistance Sniffer: An online tool for prediction of drug resistance patterns of *Mycobacterium tuberculosis* isolates using next generation sequencing data. *Int. J. Med. Microbiol. IJMM* **310**(2), 151399. <https://doi.org/10.1016/j.ijmm.2020.151399> (2020).
31. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
32. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkab301> (2021).

## Acknowledgements

We thank the Nigerian Institute of Medical Research (NIMR) for the collaboration to conduct this study. We are also grateful for the facility at the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Nigeria that enabled us carry this study further.

## Author contributions

I.B.O., C.T.H.—Designed the study. C.N.K., O.F.D., T.A.A.—Collected samples and performed rapid molecular diagnostic and drug susceptibility assays. I.B.O., J.N.U.—Performed whole genome sequencing. B.L.S., R.A.A.—Supervised the project, provided resources and reviewed draft. I.B.O.—Performed bioinformatics analysis and data curation. C.T.H., I.B.O., J.N.U.—Wrote manuscript. All authors read the manuscript and agreed with the content.

## Funding

This work is made possible by support from Flu Lab and a cohort of generous donors through TED's Audacious Project, including the ELMA Foundation, MacKenzie Scott, the Skoll Foundation, and Open Philanthropy. This work was also supported by Grants from the National Institute of Allergy and Infectious Diseases (<https://www.niaid.nih.gov>), NIH-H3Africa (<https://h3africa.org>) (U01HG007480 and U54HG007480 to C.T.H), the World Bank Grant (worldbank.org) (ACE IMPACT Project to C.T.H).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96956-7>.

**Correspondence** and requests for materials should be addressed to C.T.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021