



OPEN

## Sputum microbiota profiles of treatment-naïve TB patients in Uganda before and during first-line therapy

David Patrick Kateete<sup>1,9</sup>✉, Monica M. Mbabazi<sup>1</sup>, Faith Nakazzi<sup>1</sup>, Fred A. Katabazi<sup>1</sup>, Edgar Kigozi<sup>1</sup>, Willy Ssengooba<sup>2</sup>, Lydia Nakiyingi<sup>3,4</sup>, Sharon Namiro<sup>5</sup>, Alphonse Okwera<sup>6</sup>, Moses L. Joloba<sup>1</sup> & Adrian Muwonge<sup>7,8,9</sup>✉

Information on microbiota dynamics in pulmonary tuberculosis (TB) in Africa is scarce. Here, we sequenced sputa from 120 treatment-naïve TB patients in Uganda, and investigated changes in microbiota of 30 patients with treatment-response follow-up samples. Overall, HIV-status and anti-TB treatment were associated with microbial structural and abundance changes. The predominant phyla were *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Fusobacteria* and *Actinobacteria*, accounting for nearly 95% of the sputum microbiota composition; the predominant genera across time were *Prevotella*, *Streptococcus*, *Veillonella*, *Haemophilus*, *Neisseria*, *Alloprevotella*, *Porphyromonas*, *Fusobacterium*, *Gemella*, and *Rothia*. Treatment-response follow-up at month 2 was characterized by a reduction in abundance of *Mycobacterium* and *Fretibacterium*, and an increase in *Ruminococcus* and *Peptococcus*; month 5 was characterized by a reduction in *Tannerella* and *Fusobacterium*, and an increase in members of the family *Neisseriaceae*. The microbiota core comprised of 44 genera that were stable during treatment. Hierarchical clustering of this core's abundance distinctly separated baseline (month 0) samples from treatment follow-up samples (months 2/5). We also observed a reduction in microbial diversity with 9.1% (CI 6–14%) of the structural variation attributed to HIV-status and anti-TB treatment. Our findings show discernible microbiota signals associated with treatment with potential to inform anti-TB treatment response monitoring.

Tuberculosis (TB) is a persistent global public health problem and one of the top 10 causes of death worldwide<sup>1,2</sup>. Nearly half a million new TB cases have been reported in Uganda since 2010<sup>1</sup>, and the TB incidence in the country has surpassed that of HIV-infection<sup>3</sup>. While introduction of the Xpert MTB/RIF assay revolutionized the diagnosis of TB globally<sup>4</sup>, treatment still hinges on long treatment regimens i.e., 6–24 months depending on whether treating drug susceptible TB or drug resistant TB<sup>5</sup>. The standard first-line treatment regimen comprises of an intensive phase of 2 months treatment with isoniazid, rifampicin, pyrazinamide and ethambutol, followed by a continuation phase of 4 months treatment with isoniazid and rifampicin<sup>6,7</sup>. Following commencement of treatment with anti-TB drugs, sputum microscopy for identification of mycobacteria (in form of acid-fast bacilli, AFB) or where affordable, sputum culturing for *Mycobacterium tuberculosis* growth is done during the treatment period, usually at months 2 and 5 to monitor treatment response. Attaining sputum sterilisation, also known as sputum smear-conversion or sputum culture-conversion (i.e., from AFB smear/culture-positive to AFB smear/culture-negative) at months 2 or 5 after initiating treatment is a known cardinal index of treatment

<sup>1</sup>Department of Immunology and Molecular Biology, School of Biomedical Sciences, Makerere University College of Health Sciences, Kampala, Uganda. <sup>2</sup>BSL-3 Mycobacteriology Laboratory, Department of Medical Microbiology, School of Biomedical Sciences, Makerere University College of Health Sciences, Kampala, Uganda. <sup>3</sup>Department of Medicine, School of Medicine, Makerere University College of Health Sciences, Kampala, Uganda. <sup>4</sup>Infectious Diseases Institute, Makerere University College of Health Sciences, Mulago Hospital Complex, Kampala, Uganda. <sup>5</sup>Makerere University Lung Institute, Makerere University College of Health Sciences, Kampala, Uganda. <sup>6</sup>TB Clinics, Mulago National Referral Hospital, Kampala, Uganda. <sup>7</sup>Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh, UK. <sup>8</sup>Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, UK. <sup>9</sup>These authors contributed equally: David Patrick Kateete and Adrian Muwonge. ✉email: david.kateete@mak.ac.ug; adrian.muwonge@roslin.ed.ac.uk

success. In this regard, the World Health Organization (WHO) guidelines state that “a patient whose sputum was AFB smear-positive or culture-positive at the beginning of treatment but becomes smear-negative or culture-negative in the last month of treatment and on at least one previous occasion is declared cured” of TB<sup>6,8</sup>. Despite the importance of sputum-smear/sputum-culture conversion in monitoring TB treatment response<sup>9</sup>, they have low sensitivity and in sub-Saharan Africa culturing is mainly done in reference/regional laboratories but not as routine. Therefore, identification of microbiological factors in a sputum sample, the cornerstone for diagnosing TB, and their association with anti-TB treatment response, could be useful in unravelling new approaches for improving treatment response monitoring<sup>7</sup>.

The microbiota are microorganisms—bacteria, fungi, protozoa and viruses that live on the skin and mucosa of humans and other mammals. Their role in induction, maintenance, disruption and modulation of the immune response has recently come into focus with the advent of the human microbiome project<sup>10</sup>. The microbiota also exist in the lung<sup>11</sup>, the predilection site for *M. tuberculosis* and perhaps influence its behaviour in a variety of ways—for example via signalling<sup>11–13</sup>, which could lead to positive interactions (synergism) and/or negative interactions (competition)<sup>14</sup>. Therefore, a sound understanding of the microbiota dynamics in TB is necessary given their emerging importance in human and animal health<sup>11,15</sup>. In this context, microbial profiling in TB could advance our knowledge of TB pathogenesis (infection vs. active disease) or unravel new ways in which TB diagnostics and management can be improved<sup>7,11–13</sup>. Although microbiota shifts have been associated with several conditions and infectious diseases including pulmonary TB<sup>11,12,16,17</sup>, there is a general lack of knowledge on microbiota and disease in sub-Saharan Africa<sup>11</sup> where there is high burden of both TB and HIV.<sup>18</sup> For the first time, we report the sputum microbial composition of treatment-naïve TB patients in Uganda, and the impact of first-line anti-TB drugs on the microbiota using sputum as proxy for the lung environment<sup>19</sup>. As well, we describe the microbial changes associated with critical transitions of anti-TB treatment: pre-treatment (baseline) and treatment response follow-up at months 2 and 5.

## Results and discussion

**Patients and samples.** We enrolled 120 treatment-naïve TB patients at Mulago National Referral Hospital (Mulago hospital) in Kampala, Uganda, in the period between 2016 and 2018. Table 1 summarises the clinical and demographic characteristics of the patients; the mean age was 33 years—majority were male and residents of greater Kampala metropolitan area (i.e., Kampala city proper and the neighbouring districts of Wakiso, Mukono, Mpigi, Buikwe and Luweero), Supplementary Fig. S1 online.

**Taxonomic composition of the sputum microbiota.** High-throughput sequencing of the variable region of the *16S rRNA* gene generated a total of 9,316,821 sequence reads from the 205 sputum samples i.e., 120 (baseline), 44 (month 2), and 41 (month 5). After filtering and quality control, we retained 8,638,640 sequences representing 191 samples—106 (baseline), 44 (month 2), and 41 (month 5) that were analysed; 14 baseline samples were removed due to poor sequence quality, Table 1. The retained high-quality sequences yielded 8,180 operational taxonomic units (OTUs), 18 phyla and 333 genera.

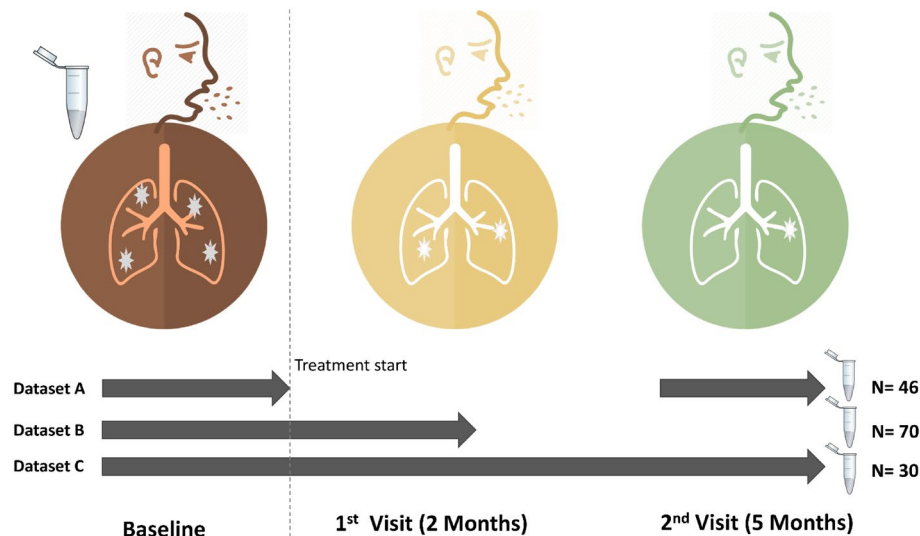
Figure 1 shows the sputum samples collected from patients at baseline (i.e., month 0 prior to anti-TB treatment commencement) and treatment response follow-up visits (months 2 and 5). Of the 106 patients enrolled, 30 had samples across the three-sampling points – Fig. 1, dataset C; 70 had samples at baseline and month 2 (dataset B), 46 had samples at baseline and month 5 (dataset A)—the microbial analyses reported in this study focus on dataset C to allow for within and between patient comparison of microbiota changes. However, to fully explore the breadth and depth of microbial characteristics for TB patients, we explored the entire dataset (Fig. 1) and show that a sputum sample on average generated 44,992 sequences, 6,580 OTUs, 4 phyla and 36 genera (also see Supplementary Figs. S2, S3 and S4). The predominant phyla detected were *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Fusobacteria* and *Actinobacteria*, accounting for nearly 95% of the sputum microbiota composition.

While we observed similar microbial composition at phylum level in the sputa of 30 patients with samples across the three-sampling points, the composition of the bacterial genera in treatment follow-up samples varied compared to baseline samples, Fig. 2. We observed a dramatic reduction in number of sequences mapping to the genus *Mycobacterium* at month 2, probably due to the effect of anti-TB drugs. Overall, 10 major genera—*Prevotella*, *Streptococcus*, *Veillonella*, *Haemophilus*, *Neisseria*, *Alloprevotella*, *Porphyromonas*, *Fusobacterium*, *Gemella* and *Rothia*, were detected across the sampling period. *Megasphaera* was detected only at baseline (month 0); *Stomatobaculum* was absent at baseline but occurred at months 2 and 5. Furthermore, *Leptotrichia* and *Actinobacillus* were present at baseline and month 2 but undetected at month 5, while *Oribacterium* and *Johnsonella* exclusively occurred at month 5, Fig. 2. Overall, 617 OTUs and 91 genera were shared between patients across the three-sampling points. The changes in abundance of some of the dominant genera i.e., *Megasphaera*, *Stomatobaculum*, *Leptotrichia* and *Johnsonella* could represent microbial shifts worth exploiting to improve monitoring of TB treatment in the future.

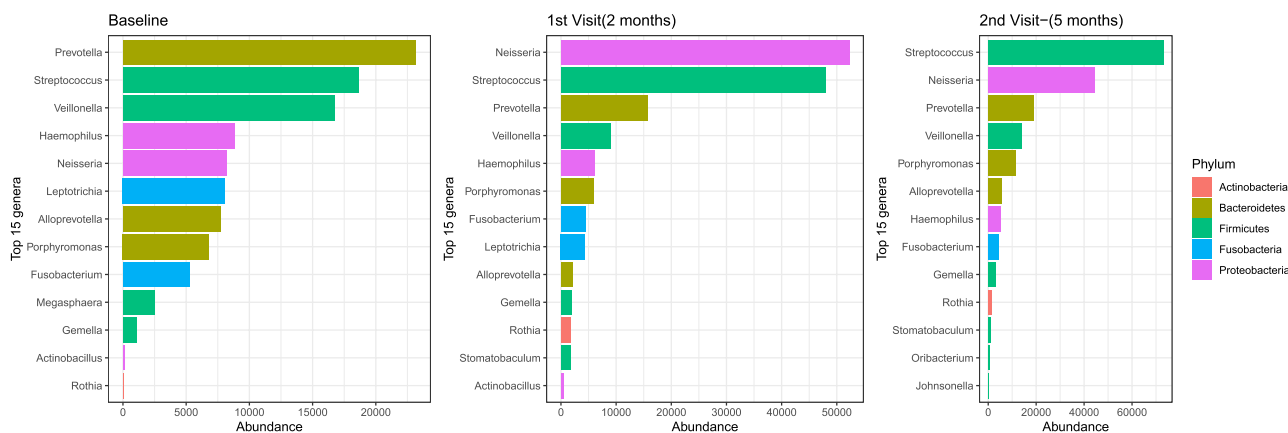
Although sputum microbial characteristics of pulmonary TB patients have been explored before<sup>11,17,20,21</sup>, for the first time we report on this from a country where HIV-associated TB is prevalent<sup>1,22</sup> and ranked by the WHO alongside 29 other countries as having the highest burden of TB globally<sup>1</sup>. Overall, our study generated nearly 16 times the number of OTUs reported in Asia<sup>16,17</sup>, probably due to differences in amplicon sequencing technology used i.e., MiSeq/illumina sequencing platform vs. 456/pyrosequencing in previous studies. Equally, these differences may signify a higher sputum microbiota diversity for Ugandan TB patients: indeed, the microbiota diversity observed in this study accounts for nearly 48% of the diversity in the human oral microbiome database <http://www.homd.org>. Generally, the sputum microbiota composition of pulmonary TB patients has been previously described as complex and more diverse than that of healthy control participants<sup>23</sup>, and our report seems to conform to this trend i.e., the sputum microbiota of TB patients in Uganda appears as complex

Variable	Patients, n = 106 <sup>a</sup> (%)	Proportion by gender (%)	
		Female, n = 38 (36)	Male, n = 68 (64)
<b>HIV status</b>			
Positive	23 (22)	10 (44)	13 (57)
Negative	83 (78)	28 (34)	55 (66)
<b>Chest X-ray</b>			
Cavities	69 (65)	25 (36)	44 (64)
No cavities	37 (35)	13	24
<b>Medication</b>			
HIV-positive on ART	16 (60)	8 (50)	8 (50)
History of antibiotics use	22 (18.3)	8 (36)	14 (64)
Others <sup>b</sup>	23 (19.2)	11 (48)	12 (52)
None reported	45 (42)	11	34
<b>Comorbidity</b>			
Hypertension	10 (9)	2 (20)	8 (80)
Smoking	26 (25)	1 (4)	25 (96)
Consume alcohol	40 (38)	11 (28)	29 (73)
None reported	30 (28)	24	6
<b>BMI</b>			
Low ( $\leq 18$ )	59 (56)	19 (32)	40 (68)
Medium ( $\geq 19-25$ )	43 (41)	16 (37)	27 (63)
High ( $\geq 26-29$ )	4 (4)	3 (8)	1 (2)
<b>Highest level of education</b>			
None	6 (6)	5 (83)	1 (17)
Primary	21 (20)	9 (43)	12 (57)
Secondary (O' / A' levels)	79 (75)	24 (30)	55 (70)
<b>District of residence</b>			
Kampala	61 (58)	24 (39)	37 (61)
Wakiso	35 (33)	12 (34)	23 (66)
Others <sup>c</sup>	10 (9)	2 (20)	8 (80)
<b>Ancestry</b>			
Bantu	92 (87)	30 (33)	62 (67)
Non-Bantu	14 (13)	8 (57)	6 (43)
<b>Employment status</b>			
Employed	94 (89)	31 (33)	63 (67)
Not employed	12 (11)	9 (75)	3 (25)
<b>Conventional TB diagnostics</b>			
<i>AFB</i>			
No AFB	16 (15)	7 (44)	9 (56)
Scanty	17 (16)	6 (35)	11 (65)
1 + AFB	10 (9)	2 (20)	8 (80)
2 + AFB	29 (27)	12 (41)	17 (59)
3 + AFB	34 (32)	11 (32)	23 (68)
<i>Xpert MTB/RIF</i>			
Positive	106	35	71
Negative	NA	NA	NA
<i>LJ culture</i>			
Positive (%)	83 (78.3)	23 (28)	60 (72)
Negative (%)	23 (22)	15 (65)	8 (35)
<b>Sampling point (month)</b>			
0 (baseline before treatment)	106 (100)	46 (43.4)	60 (56.6)
2 (1st treatment response follow-up) <sup>d</sup>	44 (41.5)	15 (34.1)	29 (65.9)
5 (2nd treatment response follow-up) <sup>c</sup>	41 (38.7)	16 (39)	25 (61)

**Table 1.** Clinical and demographic characteristics of pulmonary TB patients enrolled. ART antiretroviral treatment, AFB acid fast bacilli, BMI body mass index, LJ Lowenstein-Jensen medium, NA not applicable. <sup>a</sup>Fourteen baseline samples were removed due to poor sequence quality leaving 106 patients (one sample per patient) with sequences that were analysed (120 – 14 = 106). <sup>b</sup>Refers to medicines other than antibiotics and antiretrovirals (ARVs). <sup>c</sup>Refers to participants staying in districts other than Kampala and Wakiso. <sup>d</sup>Of the 106 patients enrolled and sampled at baseline, the available samples from treatment response monitoring at months 2 and 5 represented 30 patients i.e., 44 and 41, respectively.



**Figure 1.** Participants' enrolment, sampling and treatment response follow-up at months 2 and 5. A total of 120 pulmonary TB patients were enrolled at month 0 (baseline) prior to treatment commencement with first-line anti-TB drugs; of these, 70 had sputum samples at months 0 and 2 (dataset B), 46 had samples at months 0 and 5 (dataset A), while 30 had samples at months 0, 2 and month 5 (dataset C). The microbiome analyses reported in this study focus on dataset C. Dark red, orange and green icons depict months 0, 2 and 5, respectively.



**Figure 2.** The 15 major bacterial genera detected across the three-sampling points.

and diverse as reported for TB patients in Asia<sup>16,17,23,24</sup>. For example, there was an average of 6580 OTUs and 36 genera in a sputum sample in this study vs. 602 OTUs and 12 genera in a sputum sample in India<sup>17</sup>. Although at high taxonomic level (i.e., phylum) the microbiota profile of TB patients in Uganda is comparable to the one reported in previous studies (e.g., the five major taxa described in this study—*Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria* and *Proteobacteria*, were previously reported as major phyla in Asia comprising of  $\geq 98\%$  of the sputum microbiota composition of pulmonary TB patients<sup>16,17,23</sup>), genus level analysis reveals considerable differences in diversity and distribution of the major genera detected in Ugandan vs. Asian TB patients<sup>16,17,23,24</sup> i.e., *Prevotella*, *Streptococcus*, *Veillonella*, *Haemophilus*, *Neisseria*, *Alloprevotella*, *Porphyromonas*, *Fusobacterium*, *Gemella*, *Rothia*, *Leptotrichia*, *Megasphaera*, *Stomatobaculum*, *Oribacterium*, and *Johnsonella*. *Prevotella*, *Streptococcus* and *Veillonella* are the three most prevalent genera in this study, which is in agreement with the findings of a previous study in China<sup>16</sup>; however in India, one study reported *Streptococcus*, *Neisseria* and *Veillonella* as the three most prevalent genera<sup>17</sup>. Of note, *Prevotella* is consistently reported as a major component of sputum microbiota in chronic obstructive pulmonary disease (COPD), cystic fibrosis, pulmonary TB and lower respiratory tract infections<sup>16,23</sup>. Furthermore, while genera like *Leuconostoc*, *Lactobacillus*, *Corynebacterium*, *Bacillus*, *Acinetobacter*, *Granulicatella*, *Actinomyces*, *Anoxybacillus*, *Klebsiella*, *Pilibacter*, *Abiotrophia*, *Paucisalibacillus*, *Stenotrophomonas*, *Cupriavidus*, *Pseudomonas*, *Thermus*, *Sphingomonas*, *Brevundimonas*, *Brevibacillus*, *Methylobacterium*, *Diaphorobacter*, *Comamonas*, *Mobilicoccus*, *Granulicatella* and *Fervidicoccus* were described as prevalent and abundant in pulmonary TB patients in India<sup>17</sup> and/or China<sup>23,24</sup>, they are not among the 15 major genera reported in our study. These differences likely represent inherent environmental characteristics between

the two geographic locations; the granular role on TB treatment outcomes is beyond the scope of this study but merits further study.

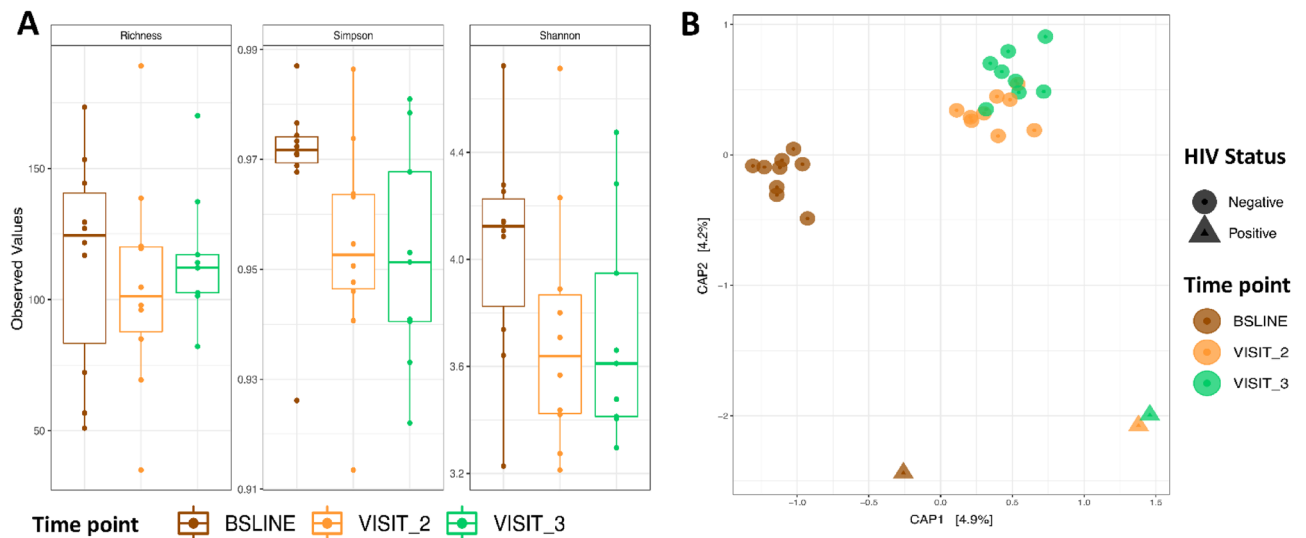
**A comparison of amplicon sequencing and conventional TB diagnostics.** In this study, there was a 0.61 (CI 0.50–0.71) and 0.63 (CI 0.51–0.73) agreement between the detection of *Mycobacterium* sequences and positive outcome on microscopy and culture (Löwenstein-Jensen—LJ), respectively. In other words, detection of *Mycobacterium* sequences in a microbiome dataset correlated with sputum-smear positivity and/or sputum-culture positivity; a Kappa test for agreement showed a 61% agreement between amplicon sequencing and microscopy, and a 63% agreement between amplicon sequencing and culture. However, this potential diagnostic utility i.e., high positive predictive value of 0.71 (CI 0.59–0.81) and 0.75 (CI 0.63–0.91) respectively, is eroded by poor specificity (see Supplementary Table S1 and Supplementary Fig. S2 online). Therefore, while amplicon sequencing ably identified smear-positive and/or culture-positive TB patients, its performance with smear-negative and/or culture-negative TB samples was poor. Furthermore, while amplicon sequencing is 10 times more expensive than microscopy (e.g., the cost for amplicon sequencing is \$25 [US dollars] per sample vs. ≤ \$2.5 per sample for microscopy), it is considerably less expensive than culture (e.g., the cost for TB culture in resource-limited settings is on average \$62.1 per sample<sup>25</sup>). Therefore, amplicon sequencing could be a good diagnostic alternative in resource-poor settings where culture is very expensive.

We also note that smear-negative and culture-negative samples produced comparatively more sequence reads than the smear-positive and culture-positive samples (see Supplementary Table S1 and Supplementary Fig. S2 online). This was a consistent observation for which the implication is currently not clear but could be attributed to low yield of the extracted total DNA from the smear-positive and culture-positive samples relative to the smear-negative and culture-negative samples. On the other hand, as respiratory infections are caused by bacteria and viruses that frequently interact with each other (reviewed by Bosch et al.<sup>14</sup>), perhaps growth/presence of *M. tuberculosis* is antagonistic to growth/presence of certain bacteria or vice-versa. However, this notwithstanding, there was an expected concordance between the *Mycobacterium* sequence count and colony forming units (CFU) on LJ cultures i.e., the sputum culture negative samples with higher total bacterial sequence reads also yielded culture-negative results.

**Sputum microbial diversity.** To understand the impact of anti-TB treatment on microbial structure, we analysed dataset C (Fig. 1) that depicts the 30 patients with samples across the three sampling points – baseline, months 2 and 5. The impact on microbial structure was determined by assessing Alpha and Beta diversities, all shown in Fig. 3: In brief, we observed a significant difference in microbiota diversity in treatment follow-up samples (collected at months 2 and 5) compared to baseline samples (collected before initiating anti-TB treatment) at alpha diversity level on indices like Richness, Simpson and Shannon, Fig. 3A. The largest change in diversity was observed between baseline and month 2 samples, which highlights the impact of TB treatment on the microbial community and specifically the genus *Mycobacterium* as noted above. The constrained ordination in Fig. 3B (and Fig. 7B) shows the microbial structural changes during treatment, here patients clustered by HIV status (i.e., HIV-positive vs. HIV-negative) and sampling point (baseline vs. months 2 and 5), explaining 9.1% of the variance. Indeed, the PERMANOVA multivariable analysis indicates that TB treatment accounts for 4–6% of the total effect-size on the microbial structure shown in Supplementary Table S2 (online) and Supplementary Fig. S3 (online) i.e., body mass index (BMI) (4%, 3%, 11%), nutritional status (3.5%, 0.5%, 0.7%), and HIV status (4%, 3.4%, 1.5%) changes in Bray–Curtis, Weighted and unWeighted unfrac distances, respectively. These variables cumulatively explain 9.4%–13.1% of the total variance i.e., microbial structural shift across the sampling period. Supplementary Fig. S5 (online) depicts two bacterial families with the highest variability relative to HIV status while Supplementary Fig. S6 (online) depicts the microbial diversity of all patients regardless of visits.

Since a large proportion of the explained variance was attributed to sampling point, probably a significant decrease in the *Mycobacterium* reads at months 2 and 5, as expected from anti-TB treatment, could be producing this signal. Indeed by month 2 there was a general reduction in total microbiota abundance, which could be attributed to the administering of rifampicin, an antibiotic with a broad-spectrum of activity<sup>11</sup>. Even after controlling for factors like sex, HIV status and BMI, this marked change in abundance remained. However, from the clustering depicted in Fig. 3B (and Fig. 7B), the microbiota grouped patients not only according to their clinic visits but also according to their HIV- and nutritional status hence, other factors<sup>26</sup> also contribute to the observed signal and these can be investigated in future for an insight into anti-TB treatment response monitoring for example, by examining the characteristics of patients with group-specific and/or stage-specific bacterial families/genera e.g., *Stomatobaculum*, *Oribacterium* and *Johnsonella* that increased during treatment and were missing at baseline (see Supplementary Fig. S5 online). Overall, the differential clustering of patients in this study contrasts with reports from Asia, in which no obvious clustering of TB patients was observed<sup>16</sup> but it is important to note that unlike this study, all the TB patients in Asian studies were HIV-negative<sup>16,17,24</sup>.

**Core and accessory microbiota.** To examine how the microbial components are affected by anti-TB treatment, we analysed dataset C in Fig. 1 and tracked components defined as ‘core’ (i.e., the genera present in eighty percent of the samples at a given sampling point<sup>27</sup>—in this case baseline, months 2 and 5) and ‘accessory’ (the difference between the core and richness, which provides insight into the characteristics of the transient microbiota). We characterised and tracked the size and composition of core and accessory microbiota in sputum samples (n = 90) of 30 patients collected at baseline (month 0) and treatment follow-up visits (months 2 and 5), Figs. 4 and 5, respectively. The change in abundance of core members is shown in Fig. 5B; interestingly, the core size was remarkably stable with an average of 44 genera over the sampling period. It is noteworthy that *Mycobacterium*’s membership of the core was restricted to baseline samples before commencement of anti-TB treatment



**Figure 3.** Microbial diversity analysis during first-line anti-TB treatment. **(A)** Depicts Alpha diversity for indices like Richness, Simpson and Shannon. **(B)** Depicts Beta diversity based on constrained ordination of the Bray–Curtis distances. Brown, orange and green represent months 0 (baseline), 2 and 5, respectively. The circular and triangular shapes in **(B)** denote HIV-negative and HIV-positive, respectively. Note this figure is based dataset C of Fig. 1 (n = 90 samples, 30 from each sampling point).

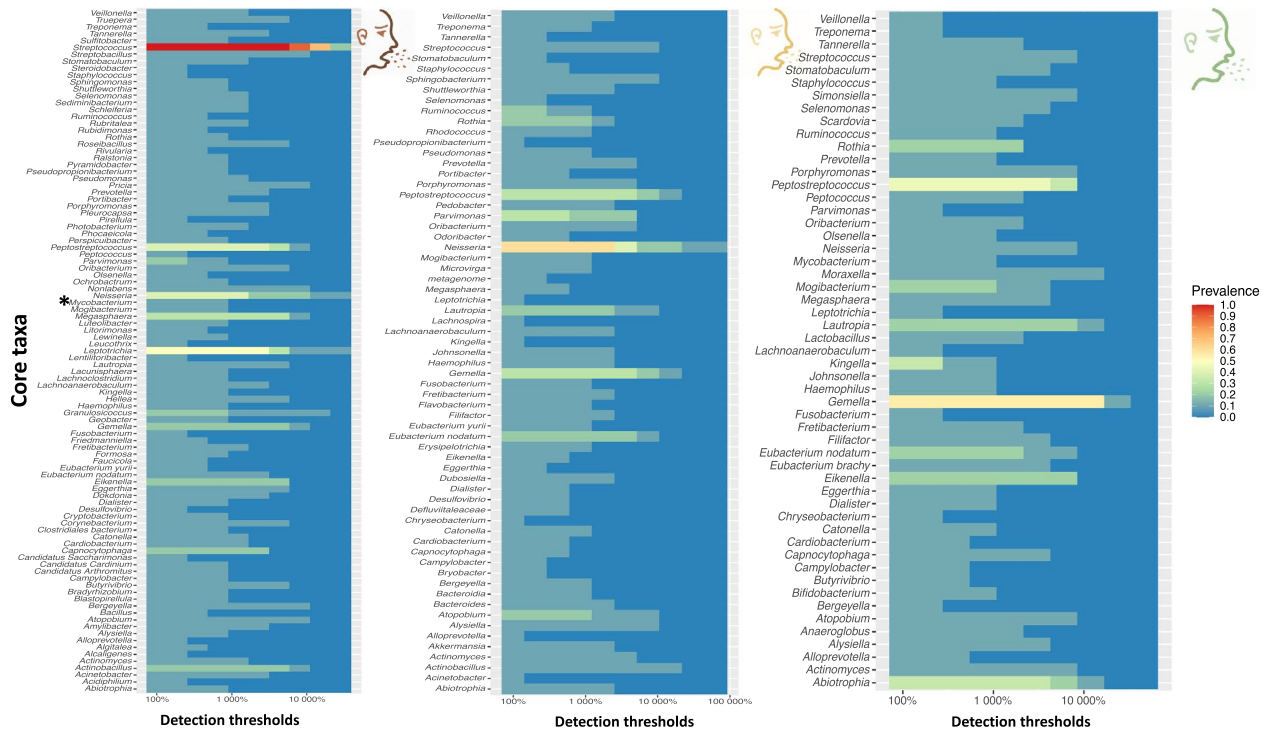
(Fig. 4). Indeed, at baseline before initiating treatment, *Mycobacterium* was a member of the core accounting for 1 in 500 bacteria in sputum. By months 2 and 5, the proportion of *Mycobacterium* fell to 1 in 5000 and 1 in 10,000, respectively, and its membership shifted to the accessory microbiota. This observation is in line with the clustering of patients by Ziehl–Neelsen (ZN) sputum smear status therefore, it supports sputum as a sample of choice for anti-TB treatment response monitoring.

The ‘core’ comprised of genera like *Streptococcus*, *Veillonella*, *Neisseria*, *Fusobacterium*, *Lachnoanaerobaculum*, *Atopobium*, *Peptostreptococcus* and *Leptotrichia*. Of these, *Streptococcus* was the most abundant but unlike *Veillonella*, its abundance was consistent before and after initiating treatment. Conversely, genera like *Neisseria* showed a steady increase over time. Figure 6 shows that most of the core genera are normal flora hence the similar patterns across sampling points. Further investigation of the abundance of each of the microbiota components showed subtle differences; clusters like (i) and (v) in Fig. 5 mapped to treatment groups while (iv) and (v) had a comparatively depleted core. Although not exclusive, clusters (i) and (iv) were predominated by patients who had started anti-TB treatment. Figure 5A examines the signatures of change in normal flora that appear to be less discriminatory as no cluster exclusively maps to a time point.

Core microbiota as characterised in this study is a close approximation of the normal flora, a significant contributor to maintenance of the integrity of several anatomical sites including the oral cavity and lung<sup>11</sup>. Dysbiosis (a deviation from normal flora) has been associated with dysregulation of the immune response, which alters the environment in favour of invading/foreign bacteria<sup>10</sup>; we noted a high biomass of the accessory microbiota, which supports the notion that dysbiosis promotes proliferation of bad/foreign bacteria<sup>10</sup> e.g., *Actinobacillus*, *Bergeyella* and *Fretibacterium* that were components of the accessory microbiota. Overall, these findings suggest that the accessory microbiota could also be playing a critical role in augmenting sputum microbiota dynamics but the pathobiological implication of this remains unclear.

**Sputum microbial abundance and clinical covariates.** To quantitatively investigate the impact of treatment, we used a Poisson mixed effects regression model, with the abundance of bacterial families as the outcome variable. Figure 7 shows the fixed and random components in panels A, B and C, respectively. We observed a dramatic change (~9 log reduction) in the abundance of families such as *Peptostreptococcaceae* (#13) and *Ruminococcaceae* (#17) at month 2, panel A. Other families like *Streptococcaceae* (#1), *Leptotrichiaceae* (#12), *Lachnospiraceae* (#14), *Erysipelotrichaceae* (#24) and *Campylobacteraceae* (#21) significantly changed between the three sampling points. Interestingly, when we account for the variation due to our explanatory variables, the change in abundance of *Mycobacterium* observed before becomes less dramatic. Generally, the model based on dataset A shows inherent variations in abundance of core microbes but within this, anti-TB treatment associated changes are discernible, panel A. The random component of this model (Fig. 7B,C) suggests that an individual sample from a TB patient accounts for, on average, 0.27 (se = 0.52), 0.53 (se = 0.73) and 0.04 (se = 0.21) of the variation but this changes to 0.18 (se = 0.42), 0.41 (se = 0.64), 0.03 (se = 0.18) respectively, when we account for the fixed effects component at the three sampling points. Therefore, clinical covariates explain most of the variance at baseline and month 2 whilst most of the variation is explained by the sample at month 5, Fig. 7B,C.

Figure 8 shows the changes in microbiota composition relative to baseline samples. Here, we observe a significant difference in the Shannon and Simpson diversity indices between the baseline and first treatment follow-up visit (month 2) but not the second treatment follow-up visit (month 5), panel A. We also observe



**Figure 4.** Characterizing the core microbiota during anti-TB treatment. The three panels depict composition of the core genera during treatment as defined by QIIME-2 microbiome package. Note that the genus *Mycobacterium* is a member of the core before initiating treatment but not during/after treatment. This analysis is based dataset C of Fig. 1.

distinct clustering by treatment follow-up period on the constrained ordination analysis, panel B. We note that BMI is associated with microbial abundance shifts between baseline and month 2, characterised by a reduction in abundance of genera like *Mycobacterium* (#25), *Olsenella* (#13) and *Haemophilus* (#11), and an increase in genera/families like *Ruminococcus* (#22) and *Lachnospiraceae* (#12) at month 2 relative to baseline, panel C. At Month 5, there was a reduction in genera like *Tannerella* (#21), *Atopobium* and *Olsenella* (#13), and an increase in *Haemophilus* (#14), *Peptostreptococcus* (#17), *Ruminococcus* (#19), *Mycobacterium* (#24) and core members of the families *Neisseriaceae* (#2) and *Lachnospiraceae* (12), panels C and D.

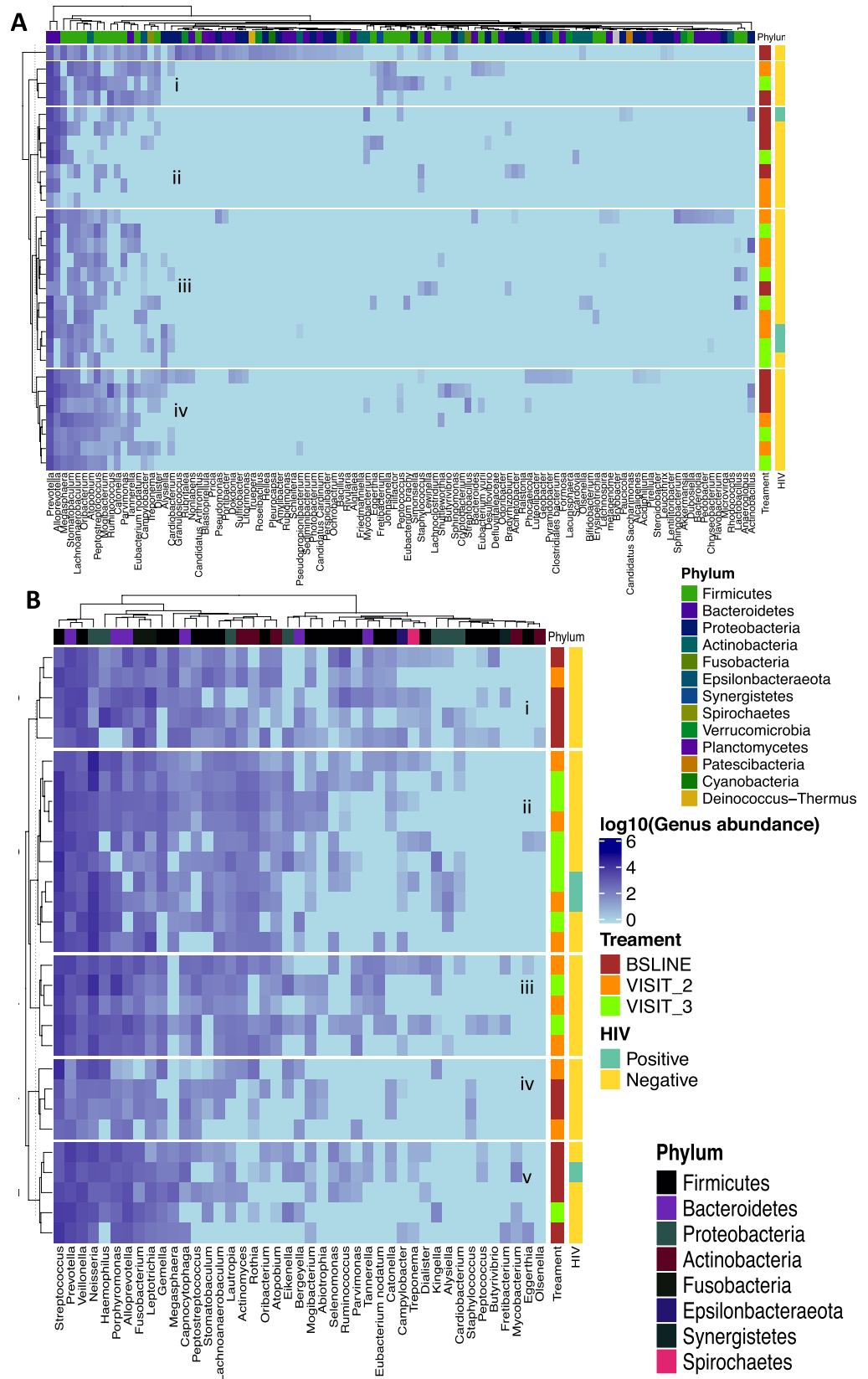
Finally, we acknowledge certain limitations in this study: first, it is difficult to normalise for sputum collection and processing, which has the potential to introduce variations. Using sputum to represent the lung environment only provides an approximation—we therefore recommend a comparison with bronchoalveolar lavage (BAL) samples, as well as stool to investigate the emerging pathogenic associations between lung function and gut microbiota (the ‘gut-lung axis’) <sup>28</sup>. Second, we did not include samples from healthy individuals as controls to compare with the microbiota from TB patients. It is noteworthy including such controls comes with complicated ethical issues; nonetheless, a longitudinal study allowed us to account for within and between patient dynamics across time. In reality, studies that have used sputum microbiota of TB patients and age-matched healthy controls <sup>15,16,20,21</sup> suggest limited utility given the similarities observed.

### Conclusions

In a TB endemic setting with high prevalence of HIV/AIDS, sputum microbiota parameters of treatment-naïve TB patients map to patient characteristics regarding clinic visits, HIV and nutritional status. Moreover, the sputum microbiota core size does not change during treatment; in contrast, the accessory microbiota to which *Mycobacterium* belongs, changes significantly. Taken together, the changes in microbial structure, microbial components (core and accessory), attributable variance and magnitude and direction of change suggest that treatment with first-line anti-TB drugs has a significant impact on sputum microbiota, which is largest at month 2 of treatment response follow-up. This study demonstrates the potential utility of clinical metagenomics as diagnostic approach to inform anti-TB treatment response monitoring in resource-limited settings where culture remains expensive.

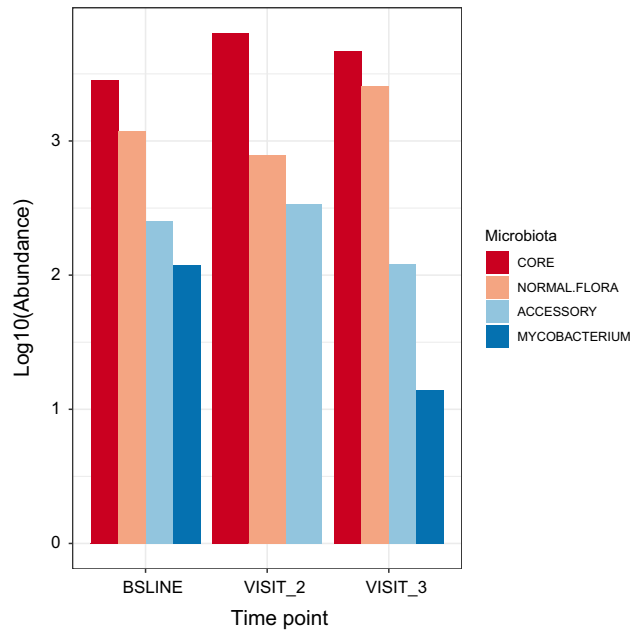
### Methods

**Study setting.** This longitudinal study was conducted at Mulago hospital in Kampala, Uganda, between 2016 and 2018. Mulago is the largest public hospital in Uganda with 1500 beds, a TB clinic and a multidrug resistant (MDR)-TB treatment centre for the country. Around 5000 TB patients are treated at the clinic every year, of whom one third are retreated patients <sup>29</sup>. We randomly enrolled 120 treatment-naïve TB patients 18 years and older who were confirmed to have pulmonary TB, collected sputum and profiled them (i.e., collecting and

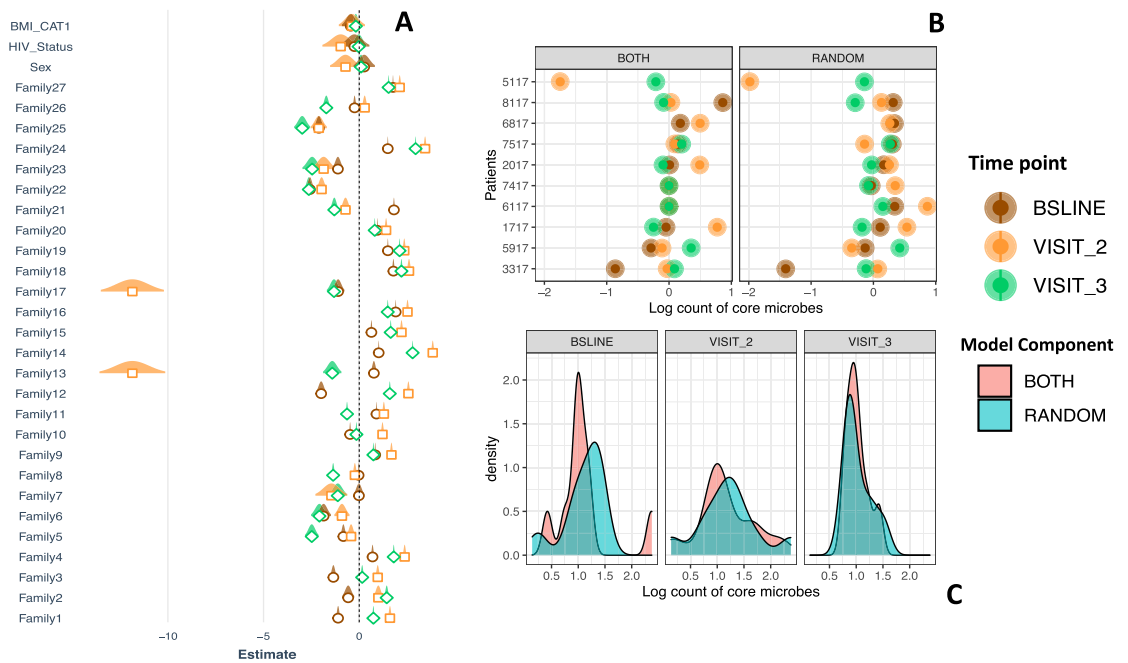


**Figure 5.** Hierarchical clustering of the normal flora (A) and core (B) at genera level. This clustering allows us to identify groups (i)–(v) that correspond to clinical covariates i.e., treatment response follow-up time and HIV status. This figure is based on dataset C of Fig. 1.



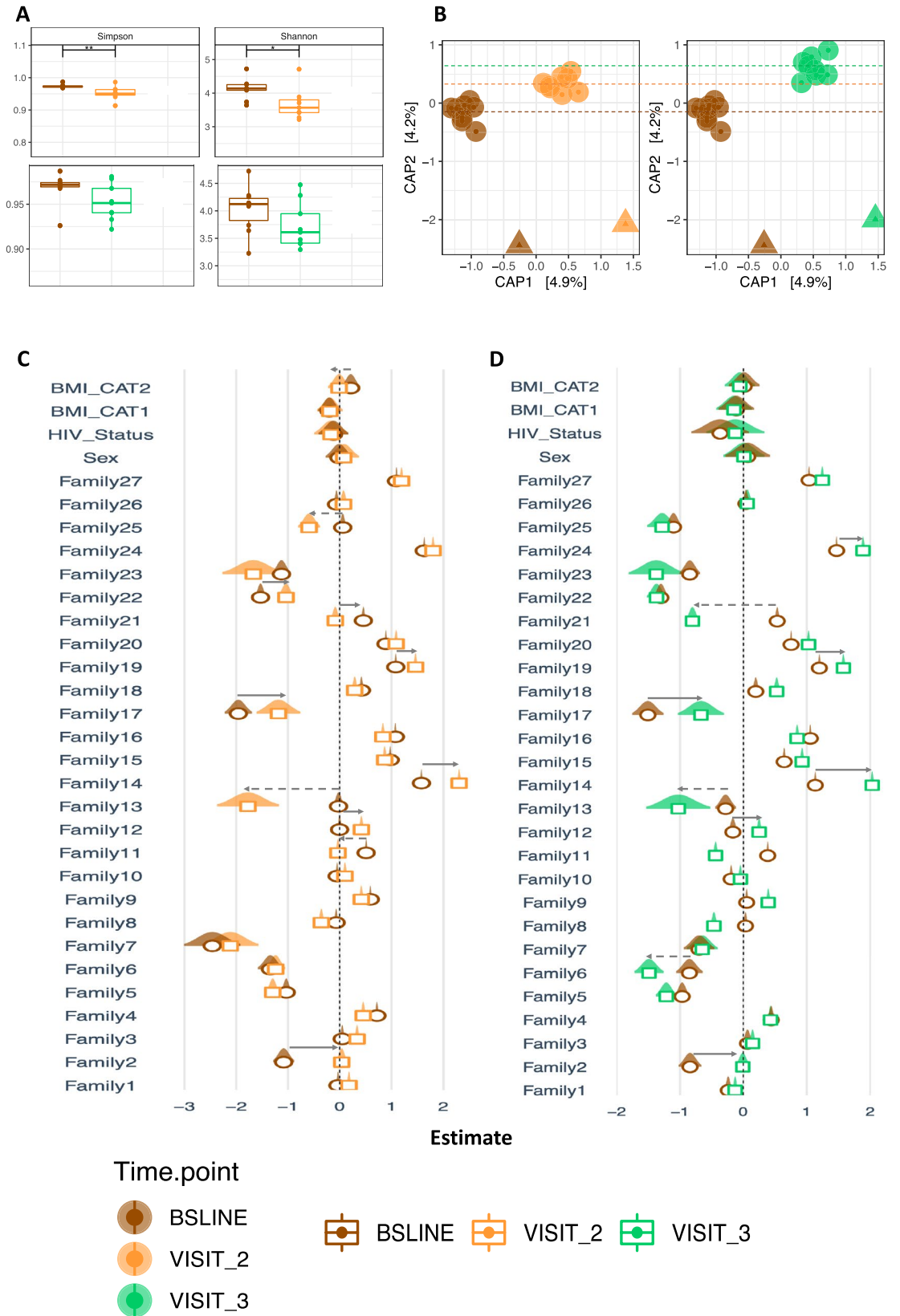


**Figure 6.** Changes in microbial components i.e., core and accessory microbiota; oral disease associated microbiota and *Mycobacterium*. This figure is based database C of Fig. 1.



**Figure 7.** Analysis of microbial-abundance. (A) Depicts change in abundance across the treatment period accounting for BMI, HIV status and gender. Brown, orange and green depict months 0 (baseline), 2 and 3, respectively. (B) depict the variance attributed to an individual sample (random intercept of our mixed model) and that due to clinical variables (fixed component of the model), the latter was larger than the former, especially between months 0 (baseline) and 2 (C). This figure is based dataset C of Fig. 1.

organizing clinical and demographic data in a way that enables its mapping to the microbiota) before initiating anti-TB treatment which formed the baseline for the study. Pulmonary TB disease was clinically diagnosed by Pulmonologists at the TB clinic and confirmed with the Xpert MTB/RIF assay on sputum; however, ZN staining for AFB detection and LJ culturing was performed for all samples at the BSL-3 Mycobacteriology Laboratory, Makerere University College of Health Sciences; the Mulago Hospital TB clinics and Makerere University laboratories are located on the same campus (for sample work-flow see Supplementary Fig. S7 online). Additional



**Figure 8.** Analysis of the magnitude and direction of change relative to baseline. This figure is based database C of Fig. 1.

samples from treatment-response follow-up visits (at months 2 and 5) were available for 30 patients; these were compared to baseline samples to unravel the microbiota changes and dynamics after initiating treatment. To account for potential cofounders, we included a wide range of clinical and lifestyle variables.

Sputum samples were collected with consent as part of routine clinical care however, to ensure consistency in quality and quantity of the sputum collected, sputum induction was performed by an expert Pulmonologist at the TB clinic as described previously<sup>30</sup>. Specimen containers with sputum were tightly closed and placed in plastic biohazard bags and immediately transported to the BSL-3 Mycobacteriology laboratory in ice-cool boxes, where the samples were processed as previously described<sup>31</sup> for AFB detection and culturing on LJ medium. Briefly, to 5–10 ml of specimen, an equal volume of a mixture of NALC-NaOH at a concentration of 3% NALC and 6% NaOH was added, vortexed for 5 min and incubated at room temperature for 15 min. The digested sample was diluted to the 50 ml mark with phosphate buffer (pH 6.8), mixed thoroughly and centrifuged at 4000g for 15 min. The sediment was adjusted to 2.5 ml with sterile phosphate buffer saline (PBS); 0.5 ml of the suspension was used to inoculate LJ slants. A portion of the sputum sediment (in ~0.5 ml) was sent to the Molecular Diagnostics laboratory, adjacent to the BSL-3 Mycobacteriology laboratory for molecular analysis. Chromosomal DNA was extracted from the pellets by using the AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions.

**DNA sequencing and sequence analysis.** High throughput sequencing of sputum DNA targeting the V3-V4 region of the 16S *rRNA* gene was performed at the Makerere University Genomics/Molecular Diagnostics Laboratory and the Integrated Microbiome Resource (IMR), Dalhousie University Canada (<https://imr.bio>)<sup>32,33</sup> on a MiSeq platform (illumina Inc.). Sequence reads were processed by using the Quantitative Insights into Microbial Ecology version 2 software (QIIME-2)<sup>34</sup>. Briefly, raw paired-end sequence reads and metadata were combined to generate a QIIME-2 object (artefact) for subsequent analysis<sup>34</sup>. Paired sequence reads were filtered for quality, dereplicated, chimeras removed and denoised by using DADA2<sup>34</sup>. This generated amplicon sequence variant (ASV) tables (previously known as operational taxonomic unit [OUT] tables) and representative sequences (in this study, the terms “ASVs” and “OTUs” are used interchangeably). The OTU tables were used to estimate Alpha and Beta diversity indices at an OTU minimum depth threshold of 2000. This was followed by a step to identify potential contamination. Alpha diversity indices included observed OTUs and Shannon<sup>35</sup>. Beta diversity was estimated by using Bray–Curtis, weighted and unweighted uniFrac distances<sup>34</sup>. Phylogenetic analysis was inferred with phyloseq package v1.26.0<sup>36</sup> and Metacoder v0.3.2-based microbiome analysis in R statistical package (v3.5.1)<sup>15</sup> during beta diversity estimation. This was done by aligning representative sequences, filtering out non-informative sites and generating a rooted tree using MAFFT and FASTTREE<sup>34</sup> software. The taxonomic classification of OTUs was achieved by using a naïve Bayes classifier trained on the most recent SILVA database at 97% similarity<sup>37</sup>. First, the training dataset was extracted by using primers used for sequencing the samples<sup>32</sup>, and the resultant dataset used to train the classifier for taxonomically assigning the OTUs. Visualisation of taxonomic abundance was done by using ggplot2 in R.

**Microbiota community structure and clinical covariates.** To assess the explanatory power/effect size of variables for changes in the microbiota community structure, we used sampling point at months 2 and 5 as proxy for monitoring the effect of anti-TB treatment on the microbiota. We performed a constrained analysis of distance-based redundancy (CAP); in this case we used Bray–Curtis as the outcome variable, HIV status and sampling point as the explanatory variables. We compared the Shannon diversity index and Richness using ANOVA and Kruskal–Wallis depending on normality<sup>35</sup>. In addition, we compared Shannon diversity index and Richness using ANOVA and Kruskal Wallis depending on normality<sup>35</sup>. For inferential analysis, we adopted a Poisson mixed effects regression in LME4 package in R using microbial core abundance as the outcome variable; HIV status, BMI, gender and microbial family as the explanatory variables; and sample identification (ID) as the random variable. Here we assessed the abundance at family level to minimise the number of covariate patterns. The families selected corresponded to the microbiota core membership across the sampling period to ensure equal representation during the participants' visits to the clinic. To examine the dynamics across the three-sampling points, we analysed dataset A; for changes relative to baseline, we analysed datasets B and C (Fig. 1). For cumulative effect size of all the explanatory variables, we used PERMANOVA models with an Adonis function (9999 permutations) in phyloseq v1.26.0<sup>36</sup> to identify microbiota structure-associations.

**Characterising the core and accessory microbiota.** To determine the composition of the core and accessory microbiota<sup>27,38</sup>, we used the microbiome package in R <https://github.com/microbiome/microbiome>. To track changes in the core as proportion of richness across sampling points, we used genera that were common to all cores for the three sampling points i.e., baseline, months 2 and 5.

**Ethical approval and consent to participate.** Ethical approval was provided by the Research and Ethics Committee of the School of Biomedical Sciences, Makerere University College of Health Sciences (reference #s SBS381/SBS542). Written informed consent was obtained from all the participants prior to enrolment into the study. All study methods were performed in accordance with the relevant guidelines and regulations.

### Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files). The raw sequence data was deposited with links to BioProject accession number PRJNA564562 in the NCBI BioProject database <https://www.ncbi.nlm.nih.gov/bioproject/>.

Received: 5 April 2020; Accepted: 14 December 2021

Published online: 29 December 2021

## References

- World Health Organization. *Global Tuberculosis Report 2021* (WHO, 2021).
- World Health Organization, W. H. O. The top 10 causes of death. 2014. *Fact sheet* (2018).
- Global Tuberculosis Community Advisory Board. *TB infections hit 90,000 a year in Uganda, overtake HIV rates*. <https://www.tbonline.info/posts/2019/3/20/tb-infections-hit-90000-year-uganda-overtake-hiv-r/> (2019).
- Theron, G. *et al.* Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *The Lancet* **383**, 424–435 (2014).
- Wipperfman, M. F. *et al.* Antibiotic treatment for Tuberculosis induces a profound dysbiosis of the microbiome that persists long after treatment is completed. *Sci. Rep.* **7**, 1–11 (2017).
- WHO. (World Health Organization Geneva, 2017).
- Rockwood, N., du Bruyn, E., Morris, T. & Wilkinson, R. J. Assessment of treatment response in tuberculosis. *Expert Rev Respir Med* **10**, 643–654. <https://doi.org/10.1586/17476348.2016.1166960> (2016).
- Ambreen, A., Jamil, M., Rahman, M. A. u. & Mustafa, T. Viable Mycobacterium tuberculosis in sputum after pulmonary tuberculosis cure. *BMC Infectious Diseases* **19**, 923. <https://doi.org/10.1186/s12879-019-4561-7> (2019).
- Su, W.-J., Feng, J.-Y., Chiu, Y.-C., Huang, S.-F. & Lee, Y.-C. Role of 2-month sputum smears in predicting culture conversion in pulmonary tuberculosis. *Eur. Respir. J.* **37**, 376–383. <https://doi.org/10.1183/09031936.00007410> (2011).
- Consortium, I. H. i. R. N. The integrative human microbiome project. *Nature* **569**, 641–648 (2019).
- Hong, B.-Y. *et al.* Microbiome changes during tuberculosis and antituberculous treatment. *Clin. Microbiol. Rev.* **29**, 915–926 (2016).
- Namasivayam, S., Sher, A., Glickman, M. S. & Wipperfman, M. F. The microbiome and tuberculosis: early evidence for cross talk. *MBio* **9**, e01420–e11418 (2018).
- Pechal, J. L., Schmidt, C. J., Jordan, H. R. & Benbow, M. E. A large-scale survey of the postmortem human microbiome, and its potential to provide insight into the living health condition. *Sci. Rep.* **8**, 1–15 (2018).
- Bosch, A. A. T. M., Biesbroek, G., Trzcinski, K., Sanders, E. A. M. & Bogaert, D. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog* **9**, e1003057–e1003057. <https://doi.org/10.1371/journal.ppat.1003057> (2013).
- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998).
- Cheung, M. K. *et al.* Sputum microbiota in tuberculosis as revealed by 16S rRNA pyrosequencing. *PLoS One* **8** (2013).
- Krishna, P., Jain, A. & Bisen, P. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. *Eur. J. Clin. Microbiol. Infect. Dis.* **35**, 1205–1210 (2016).
- Hermans, S. M. *HIV and tuberculosis co-infection in Uganda: clinical management, immune reconstitution and health service delivery*. (2012).
- Garcia, B. J. *et al.* Sputum is a surrogate for bronchoalveolar lavage for monitoring Mycobacterium tuberculosis transcriptional profiles in TB patients. *Tuberculosis* **100**, 89–94 (2016).
- Botero, L. E. *et al.* Respiratory tract clinical sample selection for microbiota analysis in patients with pulmonary tuberculosis. *Microbiome* **2**, 29 (2014).
- Luo, M. *et al.* Alternation of gut microbiota in patients with pulmonary tuberculosis. *Front. Physiol.* **8**, 822 (2017).
- Organization, W. H. Global tuberculosis report 2016. 2016. *Google Scholar*, 214 (2018).
- Cui, Z. *et al.* Complex sputum microbial composition in patients with pulmonary tuberculosis. *BMC Microbiol.* **12**, 276. <https://doi.org/10.1186/1471-2180-12-276> (2012).
- Wu, J. *et al.* Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PLoS one* **8** (2013).
- Audibert, M. *et al.* Is introducing rapid culture in the diagnostic algorithm of smear-negative tuberculosis cost-effective? (2013).
- Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
- Astudillo-García, C. *et al.* Evaluating the core microbiota in complex communities: a systematic investigation. *Environ. Microbiol.* **19**, 1450–1462 (2017).
- Budden, K. F. *et al.* Emerging pathogenic links between microbiota and the gut–lung axis. *Nat. Rev. Microbiol.* **15**, 55–63. <https://doi.org/10.1038/nrmicro.2016.142> (2017).
- Bwanga, F., Haile, M., Joloba, M. L., Ochom, E. & Hoffner, S. Direct nitrate reductase assay versus microscopic observation drug susceptibility test for rapid detection of MDR-TB in Uganda. *PLoS One* **6** (2011).
- Bell, D., Leckie, V. & McKendrick, M. The role of induced sputum in the diagnosis of pulmonary tuberculosis. *J. Infect.* **47**, 317–321 (2003).
- Muchwa, C. *et al.* Evaluation of Capilia TB assay for rapid identification of Mycobacterium tuberculosis complex in BACTEC MGIT 960 and BACTEC 9120 blood cultures. *BMC. Res. Notes* **5**, 44. <https://doi.org/10.1186/1756-0500-5-44> (2012).
- Comeau, A. M., Douglas, G. M. & Langille, M. G. Microbiome helper: a custom and streamlined workflow for microbiome research. *MSystems* **2** (2017).
- Walters, W. *et al.* Improved bacterial 16S rRNA gene (V4 and V4–5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *Msystems* **1**, e00009-00015 (2016).
- Bolyen, E. *et al.* QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. Report No. 2167–9843, (PeerJ Preprints, 2018).
- McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS one* **8** (2013).
- Foster, Z. S., Sharpton, T. J. & Grünwald, N. J. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput. Biol.* **13**, e1005404 (2017).
- Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
- Deo, P. N. & Deshmukh, R. Oral microbiome: Unveiling the fundamentals. *J. Oral Maxillofac. Pathol.* **23**, 122–128. [https://doi.org/10.4103/jomfp.JOMFP\\_304\\_18](https://doi.org/10.4103/jomfp.JOMFP_304_18) (2019).

## Acknowledgements

We thank: Professor Alison Elliott of Makerere University Infection & Immunity Programme (MUII) at the MRC/UVRI & LSHTM Uganda Research Unit, Entebbe, for useful comments on the manuscript; staff at the Mulago Hospital TB treatment centre and the BSL-3 Mycobacteriology laboratory at Makerere University for technical assistance; Harriet Nakayiza & Geraldine Nalwadda (Makerere University College of Health Sciences), and Joshua Mandre (MUII-plus, Entebbe) for administrative support.

## Author contributions

D.P.K. and A.M. conceived the study, analysed and interpreted the data and wrote the manuscript. S.N. and A.O. recruited study participants and collected sputum samples from the patients. M.M.M., F.N., E.K., W.S., and F.K.A. performed the molecular and microbiological procedures. D.P.K., L.N., W.S. and M.L.J. supervised the study protocol. M.M.M. performed part of the bioinformatics analysis (under supervision of D.P.K., A.M., and L.N.) in partial fulfilment of the requirements for the award of the degree of Master of Science in Immunology & Clinical Microbiology of Makerere University. All authors read and approved the final manuscript.

## Funding

This project is part of the EDCTP2 programme supported by the European Union (grant number TMA2018CDF-2357-MTI-Plus). The project was also supported in part by the DELTAS Africa Initiative (grant # 107743/Z/15/Z), the Africa Centre of Excellence in Materials, Product Development & Nanotechnology (MAPRONANO) (Project ID Number: P151847, IDA Number: 5797-UG) to MMM, and the Erasmus Mobility travel Grant to DPK. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS), Alliance for Accelerating Excellence in Science in Africa (AESA), and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant no. 107743) and the UK Government. AM's contribution was funded by his BBSRC Future Leader fellowship (BB/P007767/1) and Wellcome ISSF3 grant (1S3-RI.0919/20) for developing tools for clinical metagenomics in LMICs. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 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/s41598-021-04271-y>.

**Correspondence** and requests for materials should be addressed to D.P.K. or A.M.

**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