

# Metagenomics of tuberculosis infection in Hong Kong

WY Lam, MK Cheung, WYW Fung, PTW Law, KM Kam, CH Au, WY Nong, D Hwang, RCY Chan, HS Kwan, SKW Tsui \*

## KEY MESSAGES

1. Taxonomic assignment of the sequence reads using the Ribosomal Database Project Classifier showed that *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteria* were the major bacterial phyla recovered from sputum. They composed over 98% of the microbial community in both the tuberculosis (TB) and control samples.
2. Upon genera identification, 16 major bacterial genera were recovered in the sputum samples. The most abundant genera in the TB samples were *Neisseria* (28.0%), *Streptococcus* (27.8%), and *Prevotella* (16.8%), whereas those for the controls were *Streptococcus* (31.8%), *Neisseria* (22.0%), and *Prevotella* (14.4%). *Neisseria* and *Prevotella* were more dominant in the TB samples and *Streptococcus* was more dominant in the controls.
3. The microbial diversity was similar in both the TB and the control sputum samples. There was no relationship between microbial diversity and disease state. This may be because different diseases may involve different sets of microbiota.

Hong Kong Med J 2016;22(Suppl 7):S15-7

RFCID project number: 09080442

<sup>1</sup> WY Lam, <sup>2</sup> MK Cheung, <sup>1</sup> WYW Fung, <sup>2</sup> PTW Law, <sup>3</sup> KM Kam, <sup>2</sup> CH Au, <sup>2</sup> WY Nong, <sup>4</sup> D Hwang, <sup>5</sup> RCY Chan, <sup>2</sup> HS Kwan, <sup>1,6</sup> SKW Tsui

<sup>1</sup> School of Biomedical Sciences, The Chinese University of Hong Kong

<sup>2</sup> School of Life Sciences, The Chinese University of Hong Kong

<sup>3</sup> Centre for Health Protection, Department of Health

<sup>4</sup> Department of Laboratory Medicine and Pathology, University of Toronto, Canada

<sup>5</sup> Department of Microbiology, The Chinese University of Hong Kong

<sup>6</sup> Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong

\* Principal applicant and corresponding author: kwtsui@cuhk.edu.hk

## Introduction

*Mycobacterium tuberculosis* is the causative agent of tuberculosis (TB) that kills about two million persons annually.<sup>1</sup> In addition, two billion people are estimated to be latently infected with this organism.<sup>2</sup> In Hong Kong as at November 2012, there were more than 4900 TB cases, representing more than 31% of the incidence of notifiable diseases.<sup>3</sup> The proximity of Hong Kong to other regions with a high incidence of multidrug-resistant (MDR)-TB or extensively drug-resistant (XDR)-TB strains undermines local TB control efforts.<sup>4</sup> This study aimed to determine the microbial communities in sputum during primary TB infection by comparing TB with non-TB sputum samples.

## Methods

This study was conducted from November 2009 to July 2012 and was approved by the Joint CUHK-NTEC Clinical Research Ethics Committee of the Prince of Wales Hospital in Hong Kong.

Sputum samples were collected from the Tuberculosis Reference Laboratory of the Hong Kong Government. All samples were from Hong Kong Chinese patients free of HIV. No anti-TB or other antibiotic medication had been given to

patients in the 4 weeks prior to sputum collection. Smear-positive and culture-positive TB sputum samples were collected from 13 males and 8 females aged 20 to 66 years. Non-TB sputum samples were collected from 6 males and 8 females aged 22 to 82 years. The sputum samples were treated with a 3% NaOH solution, neutralised with a phosphate buffer and then centrifuged. About 0.2 mL supernatant was subjected to genomic DNA extraction using the QIAamp DNA Mini Kit according to the manufacturer's protocol (Qiagen, Valencia [CA], USA).

Polymerase chain reaction (PCR) was performed by composite primers with adaptors and sample-specific multiplex identifiers, flanking the hypervariable V1-V2 region of the 16S rRNA gene using the Platinum PCR SuperMix High Fidelity Kit (Invitrogen, Carlsbad [CA], USA). The PCR profile consisted of an initial denaturation of 30 s at 94°C, followed by 28 cycles of 30 s at 94°C, 30 s at 55°C, and 40 s at 68°C. PCR products were purified using the MinElute PCR Purification Kit (Qiagen, Valencia [CA], USA) and the quality of the purified products was assessed using Agilent Bioanalyzer 2100. Pyrosequencing of the purified PCR products was performed using the Roche/454 GS FLX Titanium platform.

Raw sequencing reads were demultiplexed, quality-filtered, and analysed using QIIME 1.4.0. Quality-filtered reads were clustered into operational taxonomic units (OTUs) at 97% similarity. Taxonomic assignment of representative OTUs was performed using the Ribosomal Database Project (RDP) Classifier at a 0.8 confidence threshold against the Greengenes core set. DNA reads obtained from pyrosequencing against databases of known sequences using a comparison tool such as BLAST were also performed. The dataset was ratified before alpha diversity calculations. Principal coordinate analysis was performed using the weighted and unweighted UniFrac distances.

## Results

### Manipulation of samples and pyrosequencing results

For pyrosequencing, a total of 964 556 raw 16S rRNA reads were obtained. The filtering process removed about 14% of the raw sequencing reads. There were about 499 000 and 331 000 high-quality reads for the TB and control samples, respectively. The average read length was about 370 bp, after removal of the primer sequences. The average number of qualified reads was 22 660 for the TB group and 23 667 for the control group. The sequencing data were then submitted to the National Centre of Biotechnology Information short read archive with an accession number SRA058505.

### Microbial diversity in the sputum samples microbiota

The qualified sequence reads were taxonomically assigned using the RDP Classifier. The major bacterial phyla identified were *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteria*. These bacterial phyla comprised over 98% of the microbial community in both the TB and control samples. Among the five major bacterial phyla identified, *Firmicutes*, *Proteobacteria*, and *Bacteroidetes* were the most abundant groups, comprising 37.6%, 31.2%, and 19.2% in the TB samples and 43.6%, 27.1%, and 17.0% in the controls, respectively. The relative abundance of *Actinobacteria* and *Fusobacteria* in the two groups was similar. *Proteobacteria* and *Bacteroidetes* were more dominant in the TB samples, whereas *Firmicutes* was more dominant in the controls. Further elaboration of the data revealed a large variation in the community structure among the tested individuals. This was supported by principal coordinate analysis, in which no obvious differential clustering was observed between members of the TB and control groups.

Upon genera identification, 16 major bacterial genera were recovered in the sputum samples.

The most abundant genera in the TB samples were *Neisseria* (28.0%), *Streptococcus* (27.8%), and *Prevotella* (16.8%), whereas those for the controls were *Streptococcus* (31.8%), *Neisseria* (22.0%), and *Prevotella* (14.4%). *Neisseria* and *Prevotella* were more dominant in the TB samples, and *Streptococcus* was more dominant in the controls. *Lactococcus*, *Pseudomonas* and unclassified *Enterobacteriaceae* were less dominant and prevalent in the TB samples. Within the 16 major genera, only *Actinomyces*, *Fusobacterium*, *Leptotrichia*, *Prevotella*, *Streptococcus*, and *Veillonella* were recovered in all the TB samples, and all these genera, except *Leptotrichia* were found in all the 36 samples. Furthermore, comprehensive data analysis revealed that there were inter-member variations in the community structure at the genus level.

The three major genera: *Neisseria*, *Streptococcus*, and *Prevotella* identified were dominated by two to three major OTUs. Two OTUs comprised more than 90% of the dominant genus *Neisseria* in both the TB samples and controls. By BLAST search, OTU 6783 was found to be similar to an uncultured clone NSV3Q1b18 and OTU 7988 was similar to an uncultured clone 7H59. For *Streptococcus*, OTU 1734 was similar to a *S mitis* clone; OTU 6370 was similar to a *S parasanguinis* clone, and OTU 7204 was highly similar to the *S salivarius*. Interestingly, two other OTUs—OTU 2225 and OTU 7999—comprised more than one-third of the genus *Prevotella* in both TB samples and controls.

Genera *Moryella*, *Mogibacterium*, and *Oribacterium* were the less represented taxa. They were enriched in the TB samples ( $P < 0.05$ ). In addition, a genus belonging to the unclassified *Lactobacillales* was enriched in the control samples ( $P < 0.05$ ). Nonetheless, at the species level, eight OTUs, including those showing 99% identity to *Prevotella melaninogenica*, *Lactobacillus crispatus*, *Streptococcus anginosus*, and *Parvimonas micra*, were enriched in the TB samples ( $P < 0.05$ ). Two other OTUs, including one showing 99% similarity to the *Aggregatibacter aphrophilus* species, were enriched in the controls ( $P < 0.05$ ).

## Discussion

Five major bacterial phyla were identified from the sputum samples: *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteria*. They are also found in other human body sites such as the oral cavity, skin, colon, the sputum of patients with cystic fibrosis and in the bronchial tract of patients with chronic obstructive pulmonary disease. This collectively suggests the prevalence of these major phyla in normal and diseased lung microbiota.<sup>6</sup>

Among the 16 major genera recovered,

only *Actinomyces*, *Fusobacterium*, *Leptotrichia*, *Prevotella*, *Streptococcus*, and *Veillonella* were found in all TB samples, and they may represent the core genera in the TB sputum microbiota. Similarly, with the exception of *Leptotrichia*, the remaining five genera were found in all 36 samples, and they may represent the core genera in the sputum microbiota in general.<sup>6</sup>

*Streptococcus*, *Neisseria*, and *Prevotella* were the three most dominant genera recovered in the sputum samples. They were also the major genera identified in the sputum of patients with chronic obstructive pulmonary disease, nosocomial pneumonia, and cystic fibrosis. This was in contrast to the normal lung microbiota, in which the genus *Pseudomonas* dominated. For *Streptococcus*, *Streptococcus pneumoniae* is a well-known pathogen associated with pneumonia. Pathogenic *Neisseria* species such as *Neisseria meningitidis* can cause pneumonia, and *Prevotella* can cause lower respiratory tract infection.<sup>6</sup>

Each of the three most dominant genera was dominated by two to three major OTUs. These OTUs comprised one-third to 90% of the corresponding groups. As the OTUs of genus showed high similarity to pathogenic strains causing a wide range of diseases, such as pneumonia, it is thus reasonable to speculate that the increased relative abundance of these opportunistic pathogens during TB infection could alter the microbial community in TB lung and affect disease progression.<sup>6</sup>

The less abundant taxa may also affect the dynamics of the microbial community and clinical outcomes. *Mogibacterium*, *Moryella*, and *Oribacterium* were the genera that were statistically abundant in the TB samples. For *Mogibacterium*, *M timidum* has been identified in cases of acute lung infection. *Moryella* and *Oribacterium* have been identified in recent decades although little information is available since their discovery.

## Conclusion

The microbial diversity was similar in TB and control sputum samples. No special relationship between the microbial diversity and the disease state was observed. This may be because different diseases may involve different sets of microbiota.

## Acknowledgements

This study was supported by the Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#09080442). We thank all the research workers involved for their hard work.

Results of this study have been published in full in: Cheung MK, Lam WY, Fung WY, et al. Sputum microbiota in tuberculosis as revealed by 16S rRNA pyrosequencing. PLoS One 2013;8:e54574.

## References

1. World Health Organization. Global tuberculosis report 2012. Geneva, Switzerland. Available at: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/)
2. Piccaro G, Giannoni F, Filippini P, Mustazzolu A, Fattorini L. Activities of drug combinations against Mycobacterium tuberculosis grown in aerobic and hypoxic acidic conditions. Antimicrob Agents Chemother 2013;57:1428-33.
3. Centre for Health Protection, Department of Health, The Government of HKSAR. 2012. Statistics on notifiable infectious diseases. Available at: <http://www.chp.gov.hk/en/data/1/10/26/43/590.htm>.
4. World Health Organization. The global MDR-TB and XDR-TB response plan 2007–2008. WHO/HTM/STB/2007. Geneva, Switzerland: 2007;1-51.
5. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet 2012;13:260-70.
6. Cheung MK, Lam WY, Fung WY, et al. Sputum microbiota in tuberculosis as revealed by 16S rRNA pyrosequencing. PLoS One 2013;8:e54574.