The microbiome and Tuberculosis - filling the gaps

B Mann

orcid.org / 0000-0002-9658-7868

Thesis submitted in fulfilment of the requirements for the degree Doctor of Philosophy in Pharmaceutical Science at the North-West University

Promoter: Dr JJ Bezuidenhout

Co-promoter: Prof AF Grobler

Co-promoter: Dr D Swanevelder

Examination: March 2021

Student number: 22787984

Abbreviations

A

AIDS	Acquired Immune Deficiency Syndrome
ANOVA	Analysis of Variance
ANOSIM	Analysis of Similarities
ARC	Agricultural Research Council
ARGs	Antibiotic Resistance Genes

B

Bp	Base pair

C

CF	Cystic Fibrosis

D

DHFR	Dihydrofolate Reductase
DHPS	Dihydropteroate Synthase
DNA	Deoxyribonucleic acid
DSI	Department of Science and Innovation

E

EMB	Ethambutol

F

G

H

HGT	Horizontal Gene Transfer
HIV	Human Immunodeficiency Virus
HREC	Health Research Ethics Committee
HRZ	Isoniazid-rifampin-pyrazinamide

I

IMPC	International Mouse Phenotyping Consortium
INH	Isoniazid

J

K

L

LMT	Lysis Micro Tube

M

MAF	*Mycobacterium Africanum*
MAIT	Mucosal-Associated Invariant T
MDR	Multi Drug Resistant
MIC	Minimum Inhibitory Concentration
MGE	Mobile Genetic Element
MTB	*Mycobacterium Tuberculosis*
MTBC	*Mycobacterium Tuberculosis* Complex

N

NCBI	National Center for Biotechnology Information
NDH-2	Type 2 NADH-quinone Oxidoreductase
NGS	Next Generation Sequencing
NTM	Non Tuberculous Mycobacteria
NWU	North-West University

O

ONT	Oxford Nanopore Technologies
OTU	Operational Taxonomic Unit

P

PAS	Para-aminosalicylic Acid
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCDDP	Preclinical Drug Development Platform
PCoA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
PZA	Pyrazinamide

Q

QC	Quality Control

R

RR	Rifampicin Resistance

RIF	Rifampicin
S	
SA	South Africa
SCFA	Short Chain Fatty Acid
SNP	Single Nucleotide Polymorphism
T	
Tregs	Regulatory T-Cells
TB	Tuberculosis
TSS	Total Sum Scaling
U	
V	
W	
WHO	World Health Organisation
X	
XDR	Extensively Drug Resistant
Y	
Z	

Definitions

Bioinformatics	The science of collecting and analysing complex biological data such as genetic codes
Dysbiosis	Dysbiosis is a term for a microbial imbalance on or inside the body
Metagenomics	Metagenomics is the study of the collective genome of microorganisms from a sample to provide information on the microbial diversity and ecology of a specific environment
Microbiome	The microorganisms in a particular environment including parts of the body
Polymerase chain reaction	The method of making multiple copies of a Deoxyribonucleic acid (DNA) sequence, involving repeated reactions with a polymerase
Resistome	The collection of all antibiotic resistance genes and their precursors in pathogenic and non-pathogenic bacteria
Sequencing	The process of determining the precise order of nucleotides within a DNA molecule
Shotgun metagenomics	Shotgun metagenomics refers to the clipping of DNA extracted from a sample and sequencing the small fragments

Declaration

I declare that the dissertation submitted by me for the degree PhD in Pharmaceutical sciences at the North-West University (Potchefstroom Campus), Potchefstroom, North-West, South Africa, is my own independent work and has not previously been submitted by me at another university.

Signed in Potchefstroom, South Africa

Signature: BC Mann

Date: 17/03/2021

Brendon Mann

Abstract

Globally tuberculosis remains one of the leading causes of mortality from any infectious agent. Evidence has shown that there are links between TB and the human microbiome and that alterations to the host micro flora in response to TB infection and subsequent treatment may have a powerful impact on both treatment outcome and patient health. Research regrading several factors surrounding TB and the microbiome, for example the resistome require urgent attention. Along with this there is also a constant need for new methods to study TB and the microbiome as the research field progresses. The first objective of this study was to determine the suitability of DNA extracted by means of the NWU lysis method for bacterial microbiome analysis; and amplicon-based next generation sequencing (NGS) in comparison to DNA obtained by various commercial kits; and the second objective was to examine changes in the bacterial microbiome of the respiratory tract during the process of TB therapy over time in human participants (sputum) by whole genome shotgun sequencing in comparison to the respiratory microbiomes of healthy controls. Applying the NWU lysis method to a mock microbial community, did not reveal any conclusive results. Following analyses on more complex samples revealed that the NWU lysis method has potential for downstream microbiome analyses but further optimisation will be required. The second objective was to examine changes in the respiratory tract microbiome during the process of TB therapy over time in the sputum human participants. Results revealed significant changes in both taxonomic composition and within the associated resistome. Overall microbiome diversity was increased in response to treatment, suggesting that microbial perturbation in response to treatment may give foreign or less abundant, potentially more resilient bacteria a chance to proliferate. Results also revealed a increase in the presence of acquired aminoglycoside resistance genes (*aadA, aadE, aac(2')-I, aph(3')-I, aph(3")-I, aph(3"')-III*, and *aph(6)-I*) linked with patient samples. The results obtained could thus be a cause for concern, especially considering the high prevalence of HIV/AIDS amongst individuals infected with TB, but further studies will be required to determine if there is any physiological or clinical significance to these observations.

Keywords: Microbiome, Antibiotic resistance, Metagenomics, Tuberculosis

Acknowledgements

To my Heavenly Father, thank you for giving me the strength and ability to complete this dissertation.

To my friends and family for their support and encouragement.

To Dr Jaco Bezuidenhout, my project supervisor for the past few years, thank you for your guidance and for keeping me motivated throughout this project.

To Prof Anne Grobler, the principal investigator of the project. Thank you for your guidance and believing in me and the idea behind this project.

Last but not least I would also like to thank Dr Z.H Swanevelder and Mr U. Vermeulen, for guidance regarding several aspects of this project.

List of figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Table of contents

Chapter 1 – Problem statement, scope of the study, aims and objectives

This is an introductory chapter to the thesis, it defines the scope of the study as it relates to current challenges/problems associated with tuberculosis and the microbiome and outlines the aims and objectives of the study in order to address some of the challenges that have been highlighted.

1. Background

Tuberculosis (TB) is recognised as one of the oldest known human diseases and occurs as a result of infection by *Mycobacterium tuberculosis* (MTB) and/or any other Mycobacterium forming part of the Mycobacterium tuberculosis complex (MTBC). Today TB remains a major global concern affecting up to 10 million people each year and for the past couple of years has been the leading cause of death from any single infectious agent, even surpassing the human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) (Corbett *et al*., 2003; World Health Organisation, 2020). Human immunodeficiency virus infection is the strongest recognised risk factor for TB, and estimates indicate that South Africa (SA), with 0,7% of the world's population contributes up to 17% of the global burden of HIV (Karim *et al*., 2009). The increasing global burden of TB is directly linked to HIV infection; weakened immunity due to HIV increases the risk of activating or reactivating latent TB infection, while also increasing the risk of infection and reinfection in general (Corbett *et al*., 2003; Getahun *et al*., 2010). According to the World Health Organisation (WHO), South Africa ranks as one of the highest in the world in terms of TB burden, while O'Donnel *et al*. (2015), states that the majority of reported cases of extensively drug resistant tuberculosis (XDR-TB) is from South Africa (World Health Organization, 2017; World Health Organization, 2020).

Tuberculosis therapy is a long and tedious process: standard treatment regimens consists of an intensive phase of two months, followed by a continuation phase of four months. This may be extended by up to seven months in instances of severe infection (Department of Health: Republic of South Africa, 2014; World Health Organisation, 2017; World Health Organization, 2020). The WHO guidelines recommend rifampicin, ethambutol and pyrazinamide for up to nine months with the addition of fluoroquinolones in the instances where the specific Mycobacteria strain has concomitant resistance to ethambutol or pyrazinamide (Gegia *et al*., 2017; World Health Organisation, 2017. Anti-microbial treatment commonly used during first line TB therapy includes isoniazid (INH), pyrazinamide (PZA), ethambutol (EMB) and rifampicin (RIF). Narrow spectrum antibiotics such as INH, PZA and EMB, are said to have very little effect outside the *Mycobacterial* genus, but rifampicin, fluoroquinolones

as well as aminoglycosides, such as streptomycin, are all broad spectrum antimicrobial agents affecting various bacteria (Nahid *et al*., 2016; Melander *et al*., 2018; Wang *et al*., 2020; Hu *et al.,* 2020).

Intensive use of antibiotics for several purposes has over time caused severe antibiotic pollution, subsequently promoting dissemination and accumulation of antibiotic resistant bacteria and antibiotic resistance genes (ARGs) into the natural, and in some instances, clinical environment (Canton, 2009; Jia *et al*., 2017; Hu *et al.,* 2020). The human gut microbiome/resistome has been described as a reservoir of antimicrobial resistance by several studies but very little attention has been paid to the rest of the human microbiome (Salyers *et al*., 2004; Penders *et al*., 2013; Duranti *et al*., 2017). In studies involving the sputum microbiome of cystic fibroses (CF) patients, the sputum microbiome/resistome as a potential reservoir for ARGs has been described. This occurrence is attributed to the frequent use of antibiotics during acute exacerbations associated with CF, leading to the emergence and selection of multidrug-resistant (MDR) bacterial species (Fancello *et al*., 2011). Other disease conditions, such as TB, has not been as extensively studied in this regard (Fancello *et al*., 2011; Hong *et al*., 2016; Wang *et al*., 2020; Hu *et al.,* 2020). A review of literature suggests that the microbiome of TB patients has only very recently received attention and evidence suggests that the gut and lung microbiome are associated with TB infection and disease (Salyers *et al*., 2004; Wood *et al*., 2017; Wang *et al*., 2020). There is a rapidly growing body of evidence linking the TB microbiome to TB disease, progression and clinical outcome; but more importantly, the severe impact of TB therapy causing extensive dysbacteriosis has also been highlighted. To date, the resistome of TB patients that have undergone extensive anti-microbial therapy remains completely unexplored. It is thus of great import to study these effects to determine whether there is the potential for the proliferation of pathogens and resistant strains of bacteria carrying ARGs (Wu et al., 2013; Wipperman *et al*., 2017; Hu *et al*., 2019; Wang *et al*., 2020; Hu *et al.,* 2020).

Another problem identified when attempting to study TB and the microbiome includesis variation, due to a lack of standardised methods. Sample collection, sequencing platforms and particularly deoxyribonucleic acid (DNA) extraction methods may all

cause variation in the final results (Mackenzie *et al*., 2015; Hong *et al*., 2016). Some bacterial cells such as those of Mycobacteria are extremely hard to lyse, and some commercially available DNA extraction kits do not sufficiently lyse all bacterial cells in a sample to give an accurate representation of the microbial community. The use of more robust methods containing bead beating steps have thus become favoured for these types of studies (Mackenzie *et al*., 2015).

There are currently several methods available for the diagnosis of TB but the sole purpose of many of these tests are identification of TB and determining its antibiotic resistance profile. All these methods though are designed to either chemically or mechanically lyse difficult to lyse Mycobacteria, but not all are equally effective at lysing these cells and in some instances DNA yields are still low (Sulis *et al*., 2016; Mohammadi *et al*., 2017). The NWU TB diagnostic method (NWU-TB test) has a very robust cell lysis step consisting of both chemical and mechanical lysis, which may make it ideal for DNA extraction to evaluate microbial communities (Mutingwende *et al.*, 2015). It is thus important to start adapting and evaluating current available methods such as the NWU TB diagnostic method (NWU-TB test) to carry out evaluations on other aspects relating to TB and disease as well, for e.g. the microbiome (Mutingwende *et al*., 2015; Sulis *et al*., 2016).

2. Research problems

The research problems identified to be addressed include the following:

- Evidence suggests that the microbiomes of the gut and respiratory tract are associated with TB infection, but more etiological and longitudinal research is required before any true conclusions can be made. This study will be, as far as we could ascertain, the first focussing on longitudinal change in the sputum microbiome in clinical sputum samples, during TB therapy; it will also be the first to exclusively attempt to characterise the resistome of patients undergoing anti-TB therapy.

- The true clinical relevance of the respiratory tract microbiome during TB infection is still unknown. The study results will thus be used to contribute to the debate of whether the respiratory tract microbiome has any true clinical relevance in regards to the potential development and dissemination of antibiotic resistance determinants within the respiratory resistome.

Parts of the human microbiome has been described as a reservoir of antimicrobial resistance. In view of the air borne infectious route as well as the drug resistant profile of TB patients, the following questions need to be addressed:

- Does exposure to TB drugs select for resistant strains of bacteria that form part of the respiratory tract microbiome?
- Is it possible for the respiratory tract microbiome to act as a reservoir for ARGs during and after TB therapy?
- Does the respiratory tract microbiome have an impact on the dissemination of ARGs and airborne infection spread?

3. Scope of the study, aims and objectives

The above questions delineated the scope of the study and resulted in formulation of the following aims and objectives to address the some of the various research questions described above:

3.1 Aims

The following aims have been identified:

- To evaluate the NWU TB diagnostic method as preferential tool for downstream microbiome analyses;
- To examine shifts in the respiratory tract microbiome of TB positive patients, during the course of tuberculosis treatment;

- To determine the presence, and if present, the abundance of clinically relevant ARGs in clinical respiratory samples during the treatment process.

3.2 Objectives

The aims of this study will be achieved by accomplishing the following objectives:

- Comparative methods: To determine the suitability of DNA extracted by means of the NWU lysis method for bacterial microbiome analysis; and amplicon-based next generation sequencing (NGS) in comparison to DNA obtained by various commercial kits;
- To examine changes in the bacterial microbiome of the respiratory tract during the process of TB therapy over time in human participants (sputum) by whole genome shotgun sequencing in comparison to the respiratory microbiomes of healthy controls.

4. Structure of the thesis

This thesis was submitted in full fulfilment of the requirements of a Doctor of Philosophy degree in Pharmaceutics at North-West University, Potchefstroom Campus, South Africa.The study was supported, and the manuscript compiled with the support of funding provided by the South African Department of Science and Innovation (DSI) in conjunction with the International Mouse Phenotyping Consortium (IMPC). Opinions expressed, findings and concluding statements are those of the authors and are not necessarily attributed to the South African Department of Science and Innovation or the International Mouse Phenotyping Consortium (IMPC). This thesis is submitted in a manuscript format in accordance with the General Academic Rules (A.7.5.7.4) of the North-West University. Each chapter is written in accordance with specific guidelines as stipulated by the journals intended for publication.

Manuscript 1 has been accepted for publication in *Data in Brief* and, and is in the final stages of review and minor amendments; manuscript 2 has been provisionally

accepted pending review and a preprint has been published, pending finalisation of the review process and final decision. Manuscript 3 has been submitted to *Nature Communications* and currently with the editor to determine the suitability of the manuscript for the journal.

5. Dissemination of findings

The manuscripts emanating from this project were written to distribute the findings of the study. The articles will mainly cater for the academic community, laboratory technologists and clinicians.

6. Declarations

The contribution for each author in the manuscripts is as follows:

B.C. Mann (PhD candidate)

- Conceptualisation of the research project
- Planning and design of the studies
- Experimental work
- Interpretation of results
- Writing of the manuscripts and final thesis.

Prof. A.F. Grobler (Principle investigator/co-supervisor)

- Refinement of the concept
- Supervision of the planning and design of the studies
- Assistance in the interpretation of results
- Supervision and critical review in the writing of the manuscripts and thesis
- Sourcing of funding for the study.

Dr. J.J. Bezuidenhout (Supervisor)

- Supervision of the planning and design of the studies
- Assistance in the interpretation of results
- Supervision and critical review in the writing of the manuscripts and thesis
- Sourcing of funding for the study.

DR. Z.H. Swanevelder

- Assistance with the planning and design of the studies
- Critical review in the writing of some of the manuscripts.

Mr. U. Vermeulen

- Design and manufacturing of some of the special equipment used in this study
- Critical review in the writing of some of the manuscripts.

Mrs. C. Wepener

- Experimental work
- Interpretation of results.

References

Canton, R. 2009. Antibiotic resistance genes from the environment: A perspective through newly identified antibiotic resistance mechanisms in the clinical setting. *Clinical Microbiology and Infection*, 15(1):20-25.

Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C. & Dye, C. 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Archives of Internal Medicine*, 163(9):1009-21.

Department of Health: Republic of South Africa. 2014. National Tuberculosis Management Guidelines: South Africa. [http://www.nicd.ac.za/assets/files/Acrobat%20Document2.pdf] Date of access 17 February 2018.

Duranti, S., Lugli, G.A., Mancabelli, L., Turroni, F., Milani, C., Manggifesta, M., Ferrario, C., Anzalone, R., Viapiani, A., van Sinderen, D. & Ventura, M. 2017. Prevalence of antibiotic resistance genes among human gut-derived *Bifidobacteria*. *Applied Environmental Microbiology*, 83(3):1-14.

Fancello, L., Desnues, C., Raoult, D. & Rolaine, J.M. 2011. Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota. *Journal of Antimicrobial Chemotherapy*, 66:2448-2454.

Gegia, M., Winters, N., Benedetti, A., van Soolingen, D. & Menzies, D. 2017. Treatment of isoniazid-resistant tuberculosis with first-line drugs: A systematic review and meta-analysis. *The Lancet, Infectious Diseases*, 17(2):223-234.

Getahun, H., Gunneberg, C., Granich, R. & Nunn, P. 2010. HIV infection associated tuberculosis: the epidemiology and the response. *Clinical Infectious Diseases*, 50(3):201-207.

Hansen, L.M., Talat, N., Hill, P.C., Hussain, R., Adegbola, R.A., Flynn, J., Canfield, D. & Parsonnet, J. 2010. Infection with *Helicobacter pylori* is associated with protection against tuberculosis. *PLoS One,* 5(1):1-9.

Hong, B., Maulen, N.P., Adami, A.J., Granados, H., Balcells, M.E. & Cervantes, J. 2016. Microbiome changes during tuberculosis and antituberculous therapy. *Clinical Microbiology Reviews*, 29:914-925.

Hu, Y., Feng, Y., Wu, J., Liu, F., Zhang, Z., Hao, Y., Liang, S., Li, B., Li, J., Lv, N., Xu, Y., Zhu, B., & Sun, Z. 2019. The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. *Frontiers in cellular and infection microbiology*, 9, 90.

Hu, Y., Cheng, M., Liu, B., Dong, J., Sun, L., Yang, J., Chen, X., & Jin, Q. 2020. Metagenomic analysis of the lung microbiome in pulmonary tuberculosis - a pilot study. *Emerging Microbes and Infections,* 9(1):1444-1452.

Jia, S., Zhang, X.X, Miao, Y., Zhao, Y., Ye, L., Li, B. & Zhang, T. 2017. Fate of antibiotic resistance genes and their associations with bacterial community in livestock breeding wastewater and its receiving river water. *Water Research*, 124:259-268.

Karim, S.S.A., Churchyard, G.J., Karim, Q.A. & Lawn, S.D. 2009. HIV infection and tuberculosis in South Africa: An urgent need to escalate the public health response. *Lancet*, 374(9693):921-933.

Mackenzie, B. W., Waite, D.W. & Taylor, M.W. 2015. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Frontiers in Microbiology*, 6:130.

Mohammadi, S., Esfahani, B.N., Moghim, S., Mirhendi, H., Riyahi, F., Safei, Z.H.G., Fazeli, H. & Mahshid Saleh. 2017. Optimal DNA isolation method for detection of nontuberculous *Mycobacteria* by polymerase chain reaction. *Advanced Biomedical Research*, 6:133.

Mutingwende, I., Vermeulen, U., Steyn, F., Viljoen, H. & Grobler, A. 2015. Development and evaluation of a rapid multiplex-PCR based system for *Mycobacterium tuberculosis* diagnosis using sputum samples. *Journal of Microbiological Methods*, 116: 37-43.

Nahid, P., Dorman, S.E., Alipanah, N., Barry, P.M., Brozek, J.L., Cattamanchi, A., Chaisson, L.H., Chaisson, R.E., Daley, C.L., Grzemska, M., Higashi, J.M., Ho, C.S., Hopewell, P.C., Keshavjee, S.A., Lienhardt, C., Menzies, R., Merrifield, C., Narita, M., O'Brien, R., Peloquin, C.A., Raftery, A., Saukkonen, J., Schaaf, H.S., Sotgiu, G., Starke, J.R., Migliori, G.B. & Vernon, A. 2016. Executive summary: Official American thoracic society/centers for disease control and prevention/infectious diseases society of America clinical practice guidelines: Treatment of drug susceptible tuberculosis. *Clinical infectious diseases*, 63(7):853-67.

O'Donnell, M.R., Wolf, A., Werner, L., Horsburgh, C.R. & Padayatchi, N. 2015. Adherence in the treatment of patients with extensively drug-resistant tuberculosis and HIV in South Africa: A prospective cohort study. *Journal of Acquired Immune Deficiency Syndrome*, 67(1)22-2.

Penders, J., Stobberingh, E.E., Savelkoul, P.H.M. & Wolf, P.F.G. 2013. The human microbiome as a reservoir of antimicrobial resistance. *Frontiers in Microbiology*, 4:1-8.

Salyers, A.A., Gupta, A. & Wang, Y. 2004. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology*, 12(9):412-416.

Sulis, G., Centis, R., Sotgui, G., D'Ambrosio, L., Pontali, E., Spanevello, A., Matteeli, A., Zumla, A. & Migliori, G.B. 2016. Recent developments in the diagnosis and management of tuberculosis. *npj Primary Care Respiratory Medicine*, 26.

Ventola, C.L. 2015. The antibiotic resistance crisis. *Pharmacy and Therapeutics*, 40(4):277-283.

Wang, J., Xiong, K., Zhao, S., Zhang, C., Zhang, J., & Xu, L. 2020. Long-term effects of multi-drug-resistant tuberculosis treatment on gut microbiota and its health consequences. *Frontiers in Microbiology,* 11:53.

Wipperman, M. F., Fitzgerald, D. W., Juste, M., Taur, Y., Namasivayam, S., Sher, A., Bean, J. M., Bucci, V., & Glickman, M. S. 2017. Antibiotic treatment for tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. *Scientific reports*, 7(1), 10767.

Wood, M.R., Yu, E.A. & Mehta, S. 2017. The human microbiome in the fight against tuberculosis. *The American Journal of Tropical Medicine and Hygiene*, 96(6):1274-1284.

World Health Organization. 2020. Global tuberculosis report 2020. [Web: https://www.who.int/teams/global-tuberculosis-programme/tb-reports] Date of access: 12 January 2020.

World Health Organization 2017. Guidelines for treatment of drug-susceptible tuberculosis and patient care. [Web: http://apps.who.int/iris/bitstream/10665/255052/1/9789241550000-eng.pdf] Date of access 21 February 2018

Wu, J., Liu, W., He, L., Huang, F., Chen, J., Cui, P., Shen, Y., Zhao, J., Wang, W., Zhang, Y., Zhu, M., Zhang, W. & Zhang, Y. 2013. Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PLoS One*, 8(12):1-10.

Chapter 2 – Literature review

This chapter reviews all available relevant literature on tuberculosis and the microbiome. It includes all available information on both animal and human studies that have been published to date with an emphasis on the potential for the development and dissemination of anti-microbial resistant bacteria and their associated genes.

1. Tuberculosis infection and prevalence

Tuberculosis (TB) caused by bacteria the *Mycobacterium tuberculosis* complex (MTBC) is a potentially deadly disease currently listed amongst the top 10 causes of death worldwide. The MTBC complex consists of *Mycobacterium tuberculosis* (MTB), *M. africanum , M. bovis, M. caprae , M. canettii, , M. microti, M. mungi, M. pinnipedii, M. suricattae* (Sinha *et al.*, 2016; World Health Organisation, 2020a). Non-tuberculous mycobacterial (NTM) infections caused by organisms such as *M. avium complex (MAC), M. kansasii,* and *M. abscessus and M. intracellulare* have also been on the rise (Johnson & Odell, 2014; Sinha *et al.*, 2016). Mycobacteria are aerobic non-motile, non-spore producing, slow growing acid fast bacteria characterised by an exceptionally strong cell wall described as being hydrophobic, waxy and rich in mycolic acids, providing an almost impermeable barrier to compounds such as antimicrobials (Delogu *et al.*, 2013). Besides being the causative agent of TB, there is a growing body of evidence suggesting that *M. tuberculosis* may also be associated with various other human conditions and diseases, such as autoimmune diseases, metabolic syndromes and pulmonary complications. The *M. tuberculosis* bacterium can also interact with and effect the human microbiome which in recent times has been described as playing an integral role in immune balance and human health (Hu *et al.,* 2019; Namasivayam *et al.,* 2020).

The bacilli are generally inhaled in aerosol droplets and enter into the lungs, here the *Mycobacterium* cells encounter a first line of defence; consisting of airway epithelial cells and alveolar macrophages. If this first line of defence is successful in eliminating the *Mycobacterium* cells the infection is aborted; but if this first line fails the bacterium starts to multiply and spread while evading the immune system by hiding inside alveolar macrophages (Figure 1) (Martino *et al.,* 2019). As seen in figure 1, TB can be present in an active or latent state, active TB can be classified as either pulmonary TB or extra pulmonary TB, with pulmonary TB being the most prevalent. Symptoms of pulmonary TB include a cough with sputum and blood at times, chest pains, weight loss, weakness and fever (Delogu *et al.*, 2013; Sinha *et al.*, 2016). One of the hallmarks of a TB infection is the development of a granuloma and its subsequent degeneration and

necrosis (Orme *et al.,* 2014). Granulomas consist of organised collections of macrophages, often with distinctive morphological variations; along with various other and other immune cells they act as a defensive mechanism triggered by the host immune response to wall off an invading pathogen. This inflammatory mononuclear cell infiltrate is capable of restricting the growth of *Mycobacterium tuberculosis;* but simultaneously also provides a unique survival niche in which the bacteria may replicate, and from which they may disseminate (Ehlers *et al.,* 2012; Orme *et al.,* 2014).

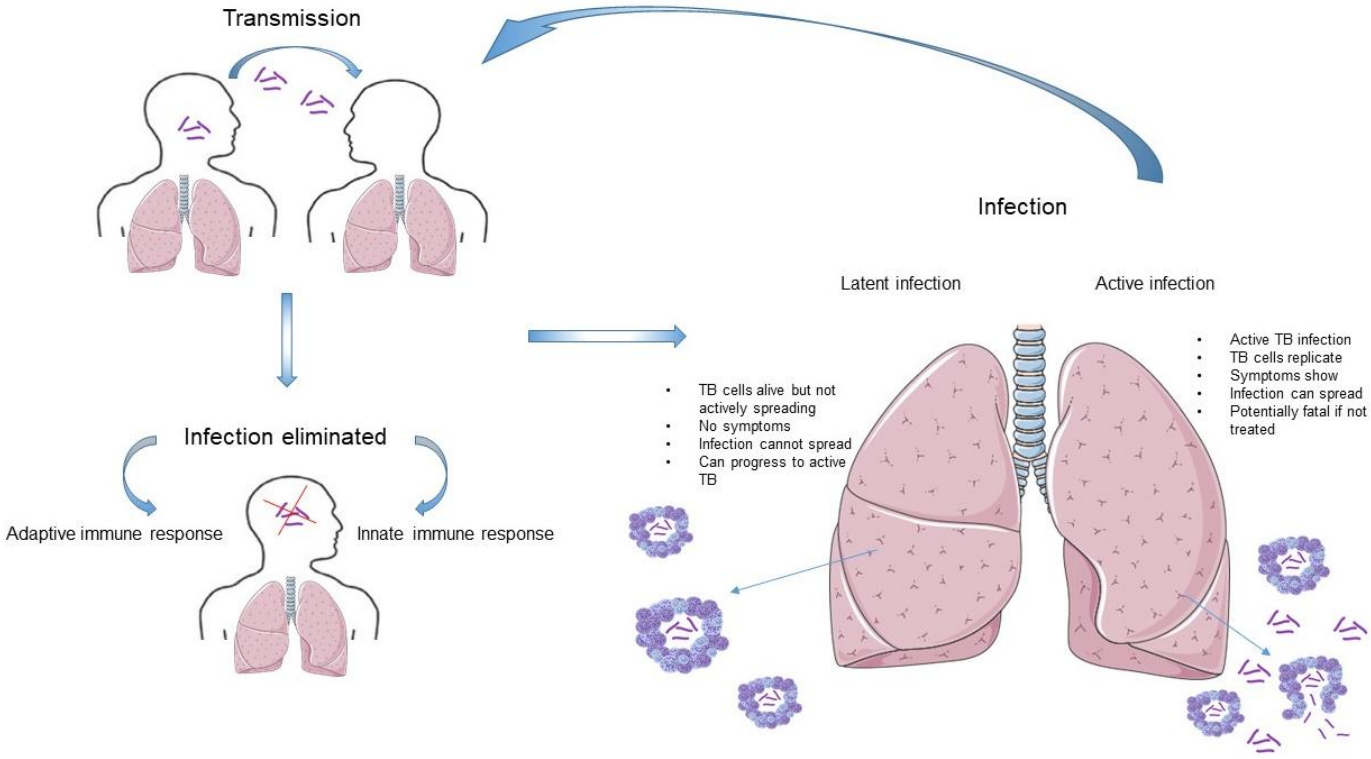


Figure 1: Illustration of TB transmission and infection of pulmonary TB (Image generated with the aid of SMART - Servier Medical ART and biorender).

In the majority of cases though the immune response either eradicates the pathogen or manage to contain it leading to latent tuberculosis infection. Someone can be classified as having latent TB if they are infected by the bacterium but do not exhibit any clinical manifestation, however there is a possibility that they may develop active TB later in life (Sinha *et al.*, 2016; Ilievska-Poposka *et al.*, 2018; World Health Organisation, 2018).

In recent years the majority of new cases were registered in the South-East Asia and Western pacific regions at around 60%, followed by Africa at roughly 25%. A smaller proportion of cases occurred in the Eastern Mediterranean regions at 8.1%, regions of the Americas at 2.9% and the European region at 2.6%. A total of 87% of all estimated incident cases worldwide were attributed to the 30 high TB burden countries (Figure 2); with India, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa accounting for two thirds of all new TB cases worldwide (World Health Organisation, 2019).

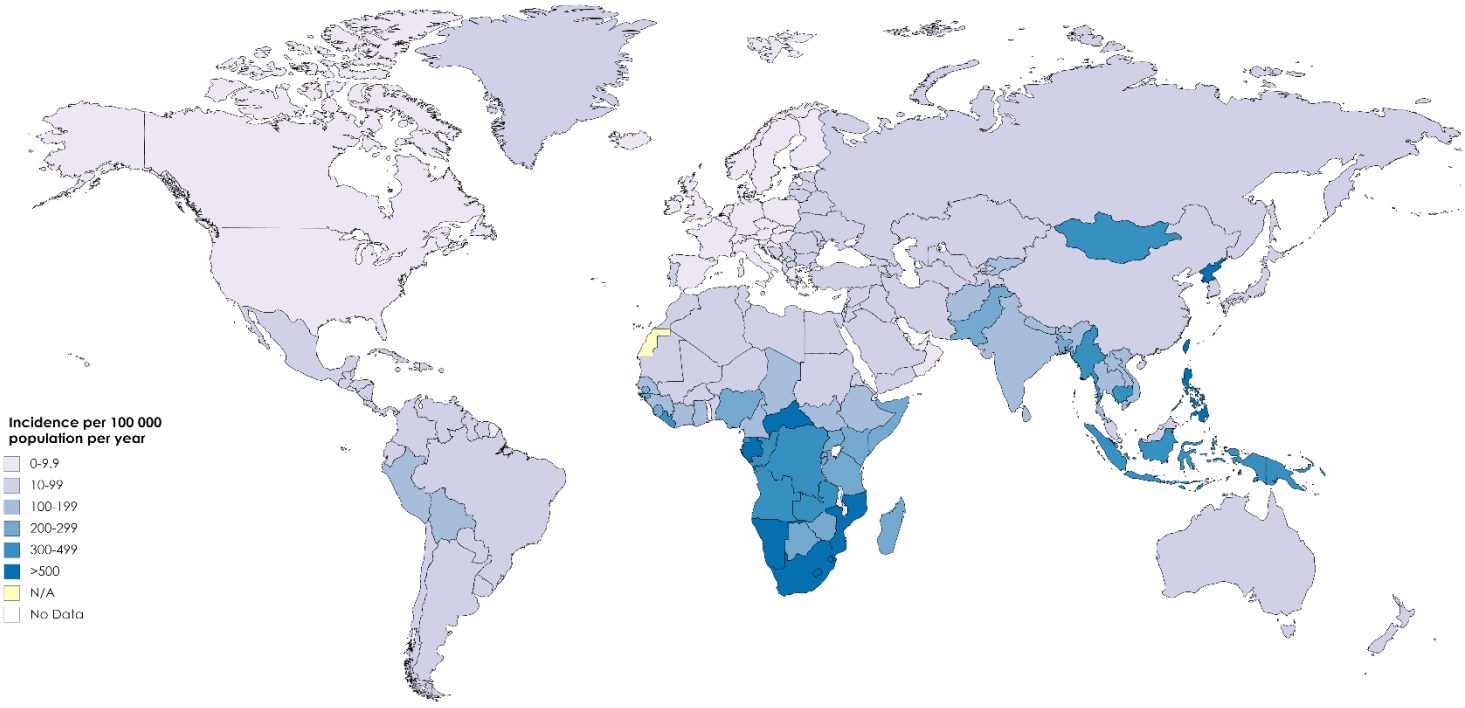


Figure 2: Incidence per 100 000 population per year (Image generated with the aid of mapchart.net and numbers drawn from the Global tuberculosis report 2019 - World Health Organisation).

According to the World Health Organisation (WHO) Global Tuberculosis Report (2018), around 10 million people fell ill with TB in 2017, and around 1.6 million died from the disease. In addition to active infection, current statistics suggests that potentially up to one third of the global population amounting to 2.3 billion individuals are carrying a latent TB infection (Houben & Dodd, 2016). TB is currently the leading cause of death among persons living with HIV, especially in low and middle-income countries such as South Africa (Houben & Dodd, 2016; World Health Organisation, 2019). Along with the

high risk associated with Human Immunodeficiency Virus (HIV), the continued development and spread of susceptible, multi and extensively drug resistant TB remains one of the greatest public health crises of our time (Maningi *et al*., 2018; World Health Organisation, 2019).

2. Tuberculosis in South Africa

Previous statistics suggest that despite being home to just about 0.7% of the world's population South Africa nevertheless accounted to around 17% of the global burden of HIV. South Africa is home to a blend of several social economic and other risk factors such as overcrowded living or working conditions, poor nutrition, smoking, alcoholism, diabetes and exposure to indoor air pollution. These factors along with the high burden of HIV all contribute to the current dire TB crisis (Abdool Karim *et al*., 2009; Naidoo *et al*., 2017). Human Immunodeficiency Virus infection is the most influential risk factor associated with both latent TB and rapid progression of active TB infection. Due to all these factors the WHO has currently ranked South Africa as one of the world's top 8 highest contributors to the global TB burden, contributing to roughly 3% of the total burden of tuberculosis worldwide (World Health Organisation, 2019). According to Naidoo *et al*. (2017), there were an estimated 454 000 incident cases, at a rate of approximately 834 cases per 100 000 persons with around 50% of new cases being in HIV-co-infected patients, suggesting that HIV is a major contributing factor to the TB problem in SA (Abdool Karim *et al*., 2009).

Along with the overall prevalence of TB in South Africa the prevalence of multi-drug resistant (MDR)-TB has also reached alarming proportions. Current statistics suggest that up to 9.6% of all TB cases reported in South Africa can be classified as multi-drug resistant, thus ranking South Africa as one of the highest MDR-TB burden countries in the world (Streicher *et al.*, 2012). To further exacerbate the problem data would suggest that up to 10.5% of MDR cases in South Africa are in fact extensively drug resistant (XDR)-TB. Another problem faced in combating TB in South Africa is adherence to the prescribed treatment regimens (O' Donnell *et al*., 2014). Adherence to the prescribed treatment is critical for positive treatment outcomes, and failure to do

so could lead to the development of resistance not only among mycobacteria but several other bacterial species as well (Axelsson, 2013). Treatment for TB is very readily available in South Africa but poor adherence to treatment is a very common occurrence due to the aforementioned social economic environment and various other risk factors present in South Africa, leading to the rapid development of both MDR- and XDR-TB (Streicher *et al.*, 2012; O' Donnell *et al*., 2014).

3. The human microbiome in health and disease

Millions of people get sick with or acquire latent TB infection every year without the presence of any of the traditionally known risk factors suggesting that there may be an additional risk factor such as the microbiome playing a key role (Hu *et al.,* 2020; Namasivayam *et al*., 2020).The microbiota or microbiome can be defined as a community of commensal, symbiotic and pathogenic microorganisms, and generally includes archaea, protists, fungi, viruses, bacteria and the genes they carry (Ursell *et al*., 2012). Since the onset of the human microbiome project in 2007 it has been found that the human microbiome consist of trillions of microorganisms including archaea, protists, and especially viruses, fungi and bacteria. These organisms and their associated genes provide us with several traits we have not had to evolve on our own and contributes to the notion that we are supraorganisms composed of both human and microbial components (Turnbaugh *et al.*, 2007). The microorganism composing the human microbiome are found at various sites on and inside the human body including the nasal passage, oral cavity, skin, urogenital tract and gastrointestinal tract with the gut microbiome being the primary focus of most studies (Barko *et al*., 2018).

The human microbiome is seen as an integral component in the general health of the host, and various studies have found links between the microbiome and various states of human health and disease (Lloyd-Price *et al*., 2017). Microorganisms forming part of the human microbiome play a role in host energy metabolism and immunity, and thus significantly influence development of a variety of human diseases (Liang *et al.*, 2018). The microbiome has been described as a defining factor influencing the host's energy balance, increasing the host's ability to produce energy from digested food, while also producing various microbial products such as metabolites,

lipopolysaccharides, short-chain fatty and secondary bile acids. These microbial products in turn act as signalling molecules modulating energy uptake, storage and expenditure as well as gut motility and appetite (Heiss & Olofsson, 2018). Commensal microorganisms have also been said to play an integral role in the host immune system. Evidence suggests that microbes not only play a role in the development of the immune system but also continue to play a role in the regulation there off by the production of various metabolites that regulate the immune system via host receptors and other target molecules (Nash *et al.*, 2017; Kim, 2018).

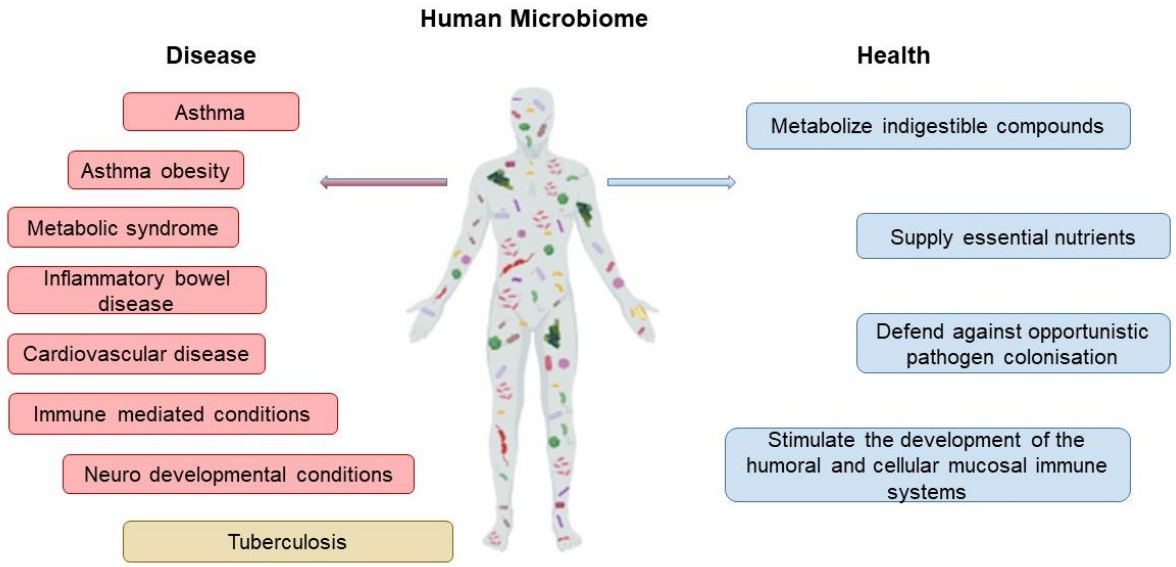


Figure 3: The human microbiome in health and disease (Image generated with the aid of SMART - Servier Medical ART and biorender).

The gut microbiome has in the past been the primary focus of microbiome based studies and has been associated with various diseases such as asthma, obesity, metabolic syndrome, inflammatory bowel disease, cardiovascular disease, immune-mediated conditions, cardiovascular disease, neurodevelopmental conditions and even TB (Figure 3) (Hong *et al.*, 2016; Barko *et al*., 2018). In spite of the gut being the primary focus in recent times, more and more attention is now being directed towards other parts of the human microbiome such as the respiratory tract as well, where it has been implicated to play a role in various diseases such as pneumonia, cystic fibrosis and even TB (Krishna *et al.*, 2016; Dickson *et al*., 2016).

There is a growing body of evidence suggesting the microbiome is of great relevance in the clinical outcome of tuberculosis (Wu *et al.,* 2013; Wipperman *et al.,* 2017; Hu *et al.,* 2019). In addition to this, anti-tuberculosis therapy is a long and tedious process, which often requires frequent or long term use of various antibiotics that frequently induce severe dysbacteriosis of the human microbiome (Wipperman *et al.,* 2017; Hu *et al.,* 2019). Beyond the initial antibiotic induced dysbactyeriosis, in many instances the necessary therapeutic dose of an antibiotic can be present at sub-inhibitory concentrations for host-associated bacteria. This presents another unique problem, as the presence of these sub-inhibitory concentrations may in turn increase the selection for resistant strains of bacteria harbouring antibiotic resistance genes (ARGs), and increase the potential for horizontal transfer of these genes (Sultan *et al*., 2018; Ghanbari *et al*., 2019).

4. The human microbiome as a reservoir for antimicrobial resistance

The human microbiome consists of complex, densely populated microbial communities associated with various sites in the body, such as the gut, skin and respiratory tract (Casals-Pascual *et al.*, 2018; MacLean & San Millan, 2019). The majority of bacteria found over various sites in the body are harmless commensals that do not cause disease and exists in a mutually beneficial relationship with the host, often outcompeting or directly inhibiting invasive pathogens by various means. It has, however, become evident that these commensal microbial communities can act as a reservoir for ARGs, due to the various means by which antibiotic resistance determinants can disseminate and spread within microbial communities (Casals-Pascual *et al.*, 2018). Drug resistant bacteria and the associated genes can be acquired by several different mechanisms: First of all, exogenous drug resistant bacteria can be attained by the host from the surrounding environment or by ingestion and secondly by the development of resistance through selection and/or induction of antibiotic-resistant mutants mediated by exposure to antibiotics at sub inhibitory concentrations or by gene transfer events (Sommer & Dantas, 2011; Casals-Pascual *et al*., 2018). Furthermore, several studies have indicated that numerous antibiotic

resistance genes present in infants may have been acquired from maternal sources (Moore *et al.,* 2013; Fouhy *et al.,* 2014; Casals-Pascual *et al*., 2018).

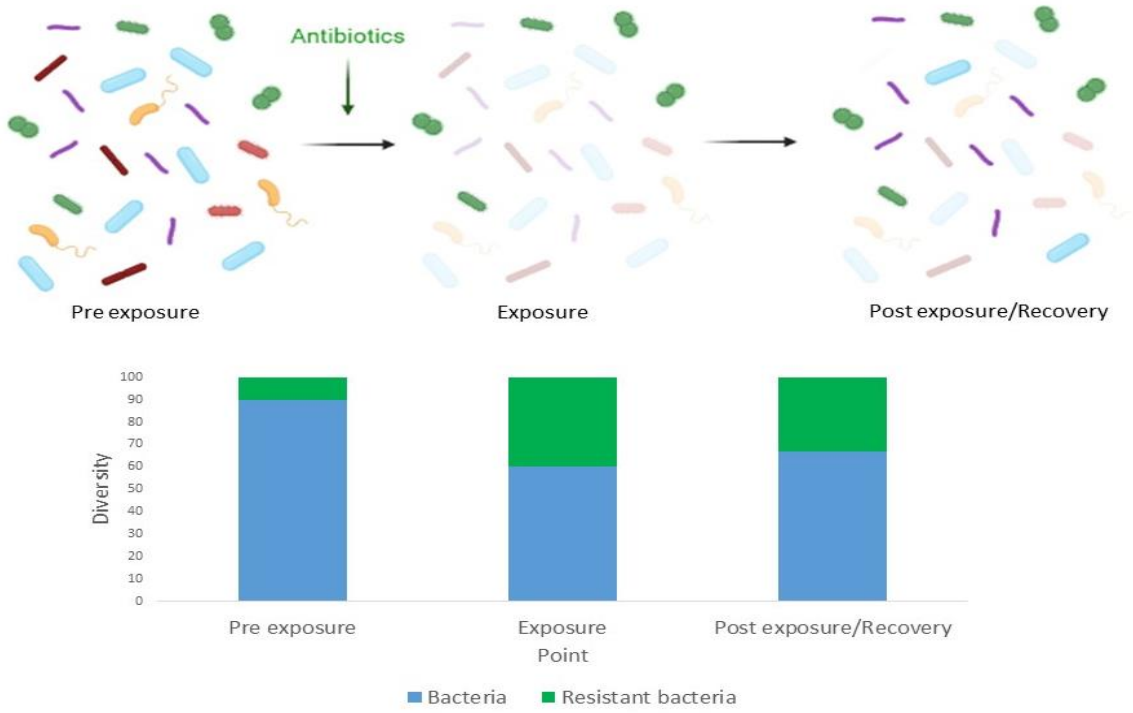


Figure 4: Illustration of the impact of antimicrobials on the microbiota and the subsequent selection and proliferation of drug resistant strains (Image generated with the aid of SMART - Servier Medical ART and biorender).

The extensive and irresponsible use of antibiotics has led to the frequent exposure of the human microbiome to various antimicrobial substances (Figure 4), greatly contributing to the development and spread of antibiotic-resistant bacterial species carrying an ever increasing abundance of antibiotic resistance genes. Antibiotic therapy thus positively selects for those microorganisms that harbour such genetic features; increasing the chances that both these organisms, and their associated antibiotic resistance determinants can continue to spread (Casals-Pascual *et al*., 2018). Similar to intrinsic and acquired resistance as previously discussed; the resistome can also be sub-divided into the intrinsic and the mobile resistome. Intrinsic ARGs are generally immobile, but can be captured by mobile genetic elements (MGEs) and eventually become mobile. Once mobile, these genes can easily disseminate by means of horizontal gene transfer (HGT), along with various other MGEs that may already be present among the microbial population (Jeong *et al.,* 2019). In many

instances, these genes are harboured by harmless commensal bacteria, which can under the correct circumstances become opportunistic pathogens; or mediate the transfer these genes to invading pathogens (Casals-Pascual *et al*., 2018; Relman & Lipsitch, 2018).

As previously mentioned Tuberculosis is one of the leading disease-based causes of morbidity and mortality worldwide; and treatment often involves the co-administration of various antimicrobials for extremely long period's time. To date, several studies have started to explore the link between the microbiome and tuberculosis but the impact of many of these antimicrobials on the microbiome and the associated resistome to date has not been extensively researched (Wu *et al.,* 2013; Wipperman *et al.,* 2017; Hu *et al.,* 2019; Namasivayam *et al.,* 2020).

5. Existing in use anti-tuberculosis agents that may impact the microbiome and resistome

Treatment regimens for TB differ depending on the target population and whether the infection is classified as susceptible, rifampicin resistant (RR-TB) or multidrug resistant – TB (MDR-TB), but are in general modelled on the WHO guidelines for treatment of tuberculosis and the WHO consolidated guidelines on drug-resistant tuberculosis treatment (World Health Organisation, 2010; World Health Organisation, 2020b). A summary of these drugs and their current accepted spectrum of activity can be found in Table 1.

Table 1: Summary of common anti-tuberculosis agents

Drug	Mechanism	Spectrum	TB Specific
Susceptible TB			
Isoniazid	Mycolic acid synthesis inhibitor	Narrow spectrum	Yes[?]
Ethambutol	Arabinosyl transferase inhibitor	Narrow spectrum	Yes[?]
Pyrazinamide	Enzyme fatty acid synthase I inhibitor	Narrow spectrum	Yes[?]
Rifampicin	DNA-dependent RNA polymerase inhibitor	Broad spectrum	No
RR and MDR - TB			
Streptomycin, kanamycin and amikacin	Protein synthesis inhibitors	Broad spectrum	No
Moxifloxacin and levofloxacin	Inhibiting DNA gyrase and topoisomerase IV	Broad spectrum	No
Ethionamide	Mycolic acid synthesis inhibitor	Narrow spectrum	Yes[?]
Tetrizidone and cycloserine	D-alanyl-D-alanine synthetase, alanine racemase, and alanine permease inhibitor	Broad spectrum	No
Bedaquiline	Mycobacterial ATP synthase inhibition	Narrow spectrum	Yes[?]
Linezolid	Protein synthesis inhibitors	Broad spectrum	No
Clofazimine	Debatable	Broad spectrum	Yes[?]
Para-aminosalicylic acid	Thought to limits folic acid production	Uncertain	Uncertain

5.1 Drug susceptible TB

First line anti-tuberculosis drugs for drug susceptible TB generally include isoniazid (INH), pyrazinamide (PZA), ethambutol (EMB) and rifampicin (RIF), with the addition of aminoglycosides or fluoroquinolones in some instances. Treatment includes an intensive phase lasting 2 months in which the combination of INH, RIF, EMB and PZA are used to quickly kill the TB cells. During the 4 month continuation phase, treatment with INH and RIF continues to eliminate any remaining TB cells. For the treatment of extra-pulmonary TB, the standard six month treatment is as effective as it would be in pulmonary disease, but in instances of complicated disease, such as TB associated meningitis, the continuation phase is generally prolonged to 7 months from the initial 4 (McIlleron & Chirehwa, 2018; World Health Organisation, 2019).

5.1.1 Isoniazid, pyrazinamide and ethambutol

Isoniazid, also known as isonicotinylhydrazide, is a narrow spectrum prodrug that is bactericidal among rapidly dividing bacteria and bacteriostatic in slower dividing bacteria. Isoniazid needs to be activated by a bacterial catalase-peroxidase enzyme found in *Mycobacterium tuberculosis* called KatG. This drug works by the formation of a nicotinoyl-NAD adduct that binds to the enoyl-acyl carrier protein reductase InhA. This in turn blocks the action of fatty acid synthase, thus inhibiting the synthesis of mycolic acids, which are essential components of the mycobacterial cell wall (Stagg *et al*., 2017).

Pyrazinamide is a fairly narrow spectrum antimicrobial, and like INH only exhibits activity against *Mycobacteria*. Pyrazinamide diffuses into *M. tuberculosis* cells where pyrazinamide is converted to pyrazinoic acid by the enzyme pyrazinamidase. Pyrazinoic acid slowly leaks from the cells and under acidic conditions converts to the protonated conjugate acid, which easily diffuses back into the bacilli and accumulates there. The effect of this process is that more pyrazinoic acid accumulates inside the cells at the acid pH generated than at a neutral pH (Zhang & Mitchison, 2003). PZA then acts to exert various effects including disruption of membrane energy function, acidification of cytoplasm and inhibition of the trans-translation process. The main role of PZA during the treatment process is to kill off any non-replicating persisters that the other drugs in the TB regimen fail to kill (Chang *et al.*, 2011; Zhang *et al*., 2013). A recent study on the effect of anti-tuberculosis treatment on a murine model indicated statistically significant changes in certain species present in the gut microbiome of mice treated only with INH or PZA, suggesting that the drugs may have a broader effect on the microbiome than previously thought. Further study is needed to elucidate if this could potentially play any role in the selection for resistant strains among bacterial groups other than MTB (Namasivayam *et al.,* 2017).

The exact mechanisms of ethambutol is said to be unknown, but evidence suggests that the drug acts by inhibiting arabinotransferases involved in the biosynthetic pathway of mycobacterial cell wall (Dookie *et al.*, 2018). In addition to this, recent studies have demonstrated a synergistic effect with isoniazid (INH). EMB is said to bind

to a possible transcription factor encoded by the Rv0273c gene, leading to the repression of inhA and thereby an enchanced mycobactericidal effect of INH (Zhu *et al*., 2018). As with INH and PZA, EMB is a narrow spectrum antimicrobial, which studies suggest is in-fact specific to only the genus *Mycobacterium*, but no NGS based studies have been done to evaluate if this drug has any individual effect on the microbiome (Dookie *et al.*, 2018).

5.1.2 Rifampicin

Rifampicin is a broad spectrum antimicrobial that forms a stable drug enzyme complex that in turn inhibits bacterial RNA polymerase, the enzyme responsible for the transcription of DNA (Campbell *et al.*, 2001). Unlike other first line anti-tuberculosis drugs, rifampicin exhibits activity against several other bacterial species such as *Staphylococcus* spp., *Streptococcus* spp., *Legionella* spp., *Neisseria* spp., *Bacteroides fragilis, Yersinia pestis, Coxiella burnetii* and *Haemophilus influenza*. The development of resistance to RIF in several of these organisms has previously been reported but there are very few studies in which the extent to which the broad use of RIF contributes to the selection of resistant strains in both the clinical and natural environment were assessed (Campbell *et al.*, 2001; Durão *et al.*, 2016).

5.2 RR- and MDR-TB

As with other forms of drug resistance, RR- and MDR-TB is very difficult to treat, and treatment varies based on the drug resistance profile of the TB isolate responsible for the infection. According to the World Health Organisation, treatment for MDR-TB should include at least four possibly effective drugs, but preferably as many as can be tolerated for treatment durations of up to 24 months or longer based on clinical outcomes (WHO, 2010; WHO, 2019; WHO, 2020b). XDR-TB is a subgroup under MDR-TB that includes isolates that are resistant to isoniazid, RIF, any fluoroquinolone and at least one of three injectable second-line drugs. Treatment for XDR-TB is extremely difficult and usually requires at least 18–24 months of treatment with up to six second-line anti-TB drugs, or other new and repurposed drug alternatives (Hughes

& Osman, 2014; Silva *et al*., 2018). In the case of XDR-TB third line drugs such as bedaquiline, linezolid, clofazimine and para-aminosalicylic acid (PAS) are added to standard MDR-TB treatment regimens, but treatment regimens are highly personalised (Hughes & Osman, 2014; WHO, 2019).

5.2.1 Aminoglycosides and fluoroquinolones

Streptomycin, kanamycin and amikacin are broad spectrum aminoglycosides, primarily forming part of the second line anti-tuberculosis drugs, but are often also included alongside first line treatment drugs (WHO, 2019). Both kanamycin and streptomycin are protein synthesis inhibitors. Kanamycin binds to the 30S subunit of the bacterial ribosome, resulting in incorrect alignment with the mRNA, resulting in non-functional peptide chains due to the wrong amino acids being built into/incorporated into peptides. Streptomycin on the other hand, binds to the small 16S rRNA of the 30S subunit, thereby interfering with the binding of formyl-methionyl-tRNA to the 30S subunit, ultimately resulting in codon misreading and inhibition of protein synthesis (Krause *et al*., 2016). Resistance to aminoglycosides arises through a variety of intrinsic and acquired mechanisms such as target site gene mutations and/or multidrug efflux pumps (Garneau-tsodikova & Labby, 2016). The global spread of aminoglycoside resistance has been well documented, but to date there is very little information available as to what extent this drug may contribute to the spread and dissemination of antimicrobial resistance within the context of anti-tuberculosis therapy (Yamane *et al.*, 2005; Garneau-tsodikova & Labby, 2016).

Fluoroquinolones on the other hand act by inhibiting DNA gyrase and topoisomerase IV, which is necessary to separate replicated DNA before bacterial cell division. Moxifloxacin and levofloxacin are two broad spectrum antimicrobials forming part of the second line TB drug treatment regimen (Hooper, 2001). Fluoroquinolones have been extensively used in both human and veterinary settings and resistance to these agents is on the rise (Redgrave *et al.*, 2014). Resistance to these antimicrobials is multifactorial and can be by means of one or more target site gene mutations, modifying enzymes, target-protection proteins and/or multidrug efflux pumps. Fluoroquinolones are primarily used for the treatment of MDR-TB but their use for the

treatment of drug susceptible TB is increasing. As with aminoglycosides, there is very little information available as to what extent this drug may contribute to the spread and dissemination of antimicrobial resistance within the context of anti-tuberculosis therapy (Redgrave *et al.*, 2014; Jabeen *et al.*, 2015)

5.2.2 Ethionamide

Similar to INH, ethionamide is a prodrug activated by the enzyme ethA found in *M. tuberculosis* (Vannelli *et al.*, 2002). Ethionamide acts by the formation of a NAD^+ adduct which inhibits InhA in the same way as isoniazid, thus disrupting the synthesis of mycolic acids, which are essential components of the mycobacterial cell wall (Vannelli *et al.*, 2002; Jeeves *at al*., 2015; Stagg *et al*., 2017). Ethionamide is a very narrow spectrum antimicrobial with action specific to *M. tuberculosis*; as with INH very little has been done to evaluate its potential effect on the broader human microbiome (Vannelli *et al.*, 2002; Jin *et al.*, 2018).

5.2.3 Tetrizidone and cycloserine

Cycloserine and a better tolerated analogue tetrizidone are both broad spectrum antimicrobials often forming part of second line TB treatment regimens. The drug acts by inhibiting the enzymes D-alanyl-D-alanine synthetase, alanine racemase, and alanine permease. Studies on tetrizidone are scarce and its safety and efficacy are still debateable. The drug is thus not widely used for the treatment bacterial infections other than those associated with MDR-TB; as such information regarding its effect on other microorganisms or the microbiome in general is extremely limited (Mulubwa & Mugabo, 2018).

5.2.4 Bedaquiline

Bedaquiline, belonging to the diarylquinoline group, is a novel compound approved during the last 5 years for the treatment of XDR-TB. It is the only anti-tuberculosis drug

known to target the metabolism of mycobacteria by inhibiting mycobacterial ATP synthase (Pontali *et al*., 2018). According to Andries *et al.* (2005), the drug appears to have a narrow spectrum of activity and exhibits little activity beyond the mycobacterial species. The study by Andries *et al.* (2005), has found that bedaquiline had much higher minimum inhibitory concentrations (MIC's) for various other bacterial pathogens but none the less a measurable MIC was obtained using the standard NCCLS methods of susceptibility testing. To date there is no acceptable protocol to test bedaquiline susceptibility and more studies are thus required to determine if the drug could potentially have an impact on the broader microbiome (Pontali *et al*., 2018).

5.2.5 Linezolid

Linezolid is the first of a new class of antibiotics referred to as oxazolidinones. Oxazolidinones act by inhibiting the synthesis of bacterial proteins: The drug acts by joining itself to the bacterial ribosome preventing the formation of the functional initiation complex 70S and thus inhibits translation (Pérez-Cebrián *et al.*, 2015). Linezolid is a broad spectrum antimicrobial with activity against several bacterial species other than mycobacteria, but very little research has been done to elucidate the impact of this drug on various other microbes (Maartens & Benson, 2015).

5.2.6 Clofazimine

Clofazimine is classified as a lipophilic riminophenazine antibiotic considered to have both antimicrobial and anti-inflammatory activities (Swanson *et al*., 2015). The main clinical application of clofazimine since its discovery has been in the treatment of multibacillary leprosy. The drug exhibits impressive activity against *Mycobacteria* and most gram positive bacteria as well as yeasts and certain parasites (McGuffin *et al.*, 2017). Clofazimine's mechanism of action has not yet been fully elucidated, but current evidence has demonstrated that clofazimine is prodrug which releases reactive oxygen species upon natural reoxidation by O_2 after it has been reduced by type 2 NADH-quinone oxidoreductase (NDH-2). It is said that clofazimine competes with menaquinone, the only quinone present in mycobacteria, and an essential cofactor in the mycobacterial electron transfer chain, for reduction by NDH-2 (Lechartier & Cole,

2015). Although recent studies have shown that clofazimine is safe, effective and does not increase the prevalence of drug resistance among *Mycobacterial* species, the potential impact of this drug on other microbes is still poorly understood (Swanson *et al*., 2015; Swanson *et al.,* 2018).

5.2.7 Para-aminosalicylic acid

Para-aminosalicylic Acid (PAS), also known as 4-aminosalicylic acid is one of the last drug options available to treat XDR-TB. Until recently the mechanism of action of PAS was still unknown, but recent studies have indicated that PAS is a prodrug targeting Dihydrofolate Reductase (DHFR). The drug acts as a metabolic precursor that is incorporated into the folate pathway by Dihydropteroate Synthase (DHPS) and DHFS, thus generating a toxic dihydrofolate analogue inhibiting DHFR activity (Zheng *et al*., 2013). Original studies on PAS conducted in the 1940's, described that the antimicrobial activity of PAS has been evaluated on several pathogenic and non-pathogenic bacterial species. Most of these test organisms exhibited very high minimum inhibitory concentrations with only *Mycobacterium* spp. being extremely susceptible (O' Connor, 1948). In general PAS is considered to be a TB specific pro-drug, but some studies have indicated activity against a broader range of bacterial species (Zheng *et al*., 2013; Saifullah *et al.,* 2014; Maier *et al.,* 2018). From this summary it is clear that several of the commonly used anti-tuberculosis drugs have a great potential to impact the host microbiome and possibly also the resistome during the course of MTB infection and subsequent treatment, in particular some of the pro-drugs previously thought to be specific only to MTB (Namasivayam *et al,* 2017).

6. Impact of MTB *infection* and anti-tuberculosis therapy on the microbiome

6.1 Animal studies

Evidence suggests that both the gut and lung microbiome are associated with TB infection and disease, with the respiratory tract receiving most of the attention as the

primary site of infection (Wipperman et al., 2017; Hu et al., 2020). Studies evaluating the effect of TB infection and concurrent treatment on the microbiome are still fairly limited; to date there are only 7 published animal studies and 14 human studies; with only 7 of these focussing on the respiratory tract. The animal studies published to date are summarised in Table 1 below. The study by Winglee *et al.* (2014), is the earliest to have evaluated the effect of MTB infection on the gut microbiome of an animal model. They found that infection lead to rapid gut dysbacteriosis, followed by recovery to a significantly different composition in response to MTB infection. The study goes on to mention that there was a significant decrease in members of the orders *Clostridiales* and *Bacteroidales;* both of which may play a role in the immune response to MTB infection, and that their change in abundance may be due to immune system activation, followed by recovery as soon as bacterial load and immune activity have reached equilibrium. This study only evaluated the effects of infection and not of the effect of subsequent treatments on the microbiota.

During a study by Khan *et al*. (2016), the authors examined if antibiotic driven dysbacteriosis provoked host susceptibility to enteric infection. The study was conducted by using a pre- and post-antibiotic model. In the pre-antibiotics model, animals were treated with broad spectrum antibiotics prior to MTB infection, imitating conditions where the individuals undergo treatment with antibiotics prior to being exposed to MTB, which may have some impact on the progression of TB. In the post-antibiotics model, animals were treated with antibiotics after MTB infection. They found that antibiotic induced dysbacteriosis led to a significant increase in MTB burden in the lungs and additional dissemination of MTB to the spleen and liver. They also noted an elevated number of regulatory T-cells (Tregs) and a decline of IFN-γ- and TNF-α-releasing CD4 T-cells. Following a faecal transplant, the affected animals displayed improved Th1 immunity and a lower regulatory T-cell number. The authors concluded that these findings may indicate that the change in the gut microbiome can potentially facilitate the progression of MTB infection. A very important note in regard is that the results of this study also suggest the elimination of antibiotic sensitive bacteria, thus leading to the proliferation of antibiotics resistant microbes (Khan *et al*., 2016).

Table 2 : Animal studies

Study	Specimen/Animal	Time points	Antimicrobials administered	Key findings
Winglee *et al*., 2014	Stool/mouse	Prior to infection and throughout infection	Untreated	Mouse gut microbiota changes in response to MTB infection. Decrease in *Clostridiales* and *Bacteroidales*, changes in abundance may be due to immune activation.
Khan *et al.,* 2016	Stool/mouse	Single time points	Vancomycin, polymixinB, carbenicillin, trimethoprim, and amphotericin B	Change in the gut microbiome can potentially facilitate the progression of MTB infection, and the prevention and/or treatment of dysbacteriosis may improve treatment outcomes.
Namasivayam *et al*., 2017	Stool/mouse	Several time points	RIF+ INH + PZA, RIF + PZA, RIF + INH, RIF alone, INH alone, PZA alone	Significant decrease in diversity (mainly class *Clostridia* in phyla Firmicutes). Dysbiosis occurs mainly because of antimicrobial treatment and not due to infection by MTB. The main driver for dysbiosis is rifampin. Treatment results in long lasting dysbiosis.
Cadena *et al.,* 2018	BAL fluid/primate	2 weeks prior to infection, and 1, 4, and 5 months post-infection	Untreated	MTB-driven microbial reorganization occurs within the lung. Large degree of variability among subjects.
Dumas *et al*.,2018	Lung homogenate and stool/mouse	Single time points	Ampicillin, vancomycin, neomycin sulphate and metronidazole.	The host microbiome contributes to early protection of lung colonization by MTB, potentially through sustaining MAIT cell function.
Namasivayam *et al*., 2019	BAL fluid/primate	Pre- and post- infection	Untreated	Large degree of inter-individual variability in microbiomes of Rhesus Macaques. *Clostridiaceae* and *Lachnospiraceae* increased while *Streptococcaceae* decreased in monkeys who were more susceptible to disease.
Khan *et al.,* 2019	Stool/mouse	Single time points	RIF or INH+PZA	Increases in Bacteroidetes and Verrucomicrobia were observed with a decrease in Firmicutes (family level *Lachnospiraceae*). Host resistance to MTB infection is compromised by pre-treatment with INH/PZA, but MTB-specific T-cell responses were not affected. Macrophage bactericidal activity (innate immune response) in the lungs are impacted due to INH & PZA-induced dysbiosis of the gastrointestinal microbiome.

Abbreviations: Rifampicin (RIF), Isoniazid (INH), Pyrazinamide (PZA), Broncho alveolar lavage (BAL), *Mycobacterium tuberculosis* (MTB), Mucosal-Associated Invariant T (MAIT).

A more frequent study by Namasivayam *et al.* (2017), evaluated the longitudinal outcome of a conventional anti-tuberculosis treatment regimen consisting of isoniazid-rifampin-pyrazinamide (HRZ), as well as the effects of each of the individual antibiotics alone and in different combinations. The study found a decrease in *Clostridiales* following infection, in agreement with the previous study conducted by Winglee *et al.* (2014), but noted that these changes were minor when compared to the dysbacteriosis caused by the concurrent anti-tuberculosis therapy. The standard HRZ treatment regimen was found to result in significant decreases in the following genera: *Acetivibrio*, *Robinsoniella*, *Alkaliphilus*, *Stomatobaculum*, *Butyricicoccus*, *Acetanaero bacterium*, *Tyzzerella*, *Ruminococcus*, and *Peptococcus*; while also noting decreases in genera primarily belonging to the class Clostridia. Increases were limited to the genera *Erysipelatoclostridium* and *Eggerthia*.

The study goes on to mention that anti-tuberculosis therapy leads to rapid and long lasting (3 months) dysbacteriosis of the gut microbiome, but more importantly found that certain pro-drugs such as INH and PZA triggered statistically significant decreases in the relative abundance of various species with no obvious relationship to the mycobacteria for which they should be specific. The authors conclude that in addition to the possible presence of hidden mycobacterial species or unrelated bacteria sharing similar drug target and/or activating enzymes, the reason for this impact of isoniazid and pyrazinamide on the microbiome remains undetermined. Another frequently used pro-drug, ethambutol, has not yet been evaluated in this regard (Namasivayam *et al*., 2017). In a follow-up study the same authors employed a nonhuman primate *rhesus* macaque model to continue examining the possible relationship of the microbiome with TB disease and outcomes. During the course of the study it was evident that alterations observed in the gut microbiota due to MTB infection was lower in magnitude when compared to the variability seen between individual monkeys. However, the gut microbes of the monkeys who developed severe disease was distinct from those with less severe disease, noting that bacterial families *Clostridiaceae* and *Lachnospiraceae* were increased while *Streptococcaceae* were decreased in monkeys who were more susceptible. Similar to some previous studies the aim was to assess the impact of MTB infection and no evaluation of the impact of TB treatment on this animal model was included during the course of the study (Namasivayam *et al.,* 2019).

The most recently published mouse study, also by Khan *et al*. (2019), demonstrated a similar trend as that observed by Namasivayam *et al.* (2017), namely that mice treated with either RIF or the combination of INH and PZA presented with significantly altered gut microbiota following treatment. RIF pre-treatment resulted in an increase in *Bacteroidetes* and *Verrucomicrobia* with a decrease in *Firmicutes*; while treatment with INH/PZA increased the representation of Bacteroidetes. At the family level similar changes were observed with an increase in *Bacteroides, Verrucomicrobiaceae* and a decrease in *Lachnospiraceae* for mice treated with RIF and an increase in *Clostridiaceae* for mice treated with INH/PZA. The authors go on to report an effect on the innate immune response with an impact on alveolar macrophages. Results of the study demonstrated that macrophages had decreased expression levels of MHCII, while also noting a decrease in the production of TNFα and IL-1β. Additionally the authors also demonstrated that macrophages from INH/PYZ-treated mice exhibited a decrease in basal respiration, spare respiratory capacity, maximal respiration and ATP production; suggesting that the TB therapy promotes the oxidative phosphorylation of alveolar macrophages leading to an environment more permissive for MTB growth. Interestingly the study also demonstrated that the treatment of mice with the INH/PZA led to an increased MTB bacterial burden, which was reversible by faecal transplantation from untreated animals. This indicates that a specific gut microbial composition may be critical for preserving the integrity of the immune response to MTB infection, further strengthening the importance of the gut–lung axis in host defence. The authors conclude that for the time being it is unclear how anti-microbial induced dysbiosis led to an altered metabolic state in alveolar macrophages, but mention an exciting prospect. It could be that metabolites generated by the gut microbiota and accumulate in the blood are altered following INH/PYZ treatment, inducing changes in the metabolic state of alveolar macrophages, which is critical in determining the function of a macrophage in response to MTB (Khan *et al*., 2019).Similar to the previous study by Namasivayam *et al* (2017), the study also demonstrates that the mechanisms involved with the INH/PZA induction of dysbacteriosis has not been determined yet (Namasivayam *et al*., 2017; Khan *et al*., 2019).

During the course of this review only two studies were found that have focussed on the respiratory tract microbes of an animal model during MTB infection (Cadena *et al*., 2018; Dumas *et al*., 2018). The first involved the use of a non-human primate model. Several macaques were infected with MTB, following which infected and uninfected lung lobe washes were collected by bronchoalveolar lavage during the course of infection. As with the previously mentioned study, there was no monitoring of the impact of treatment on the microbiome, the main focus of the study being the impact of infection itself on the host microbes. At one month post infection, the authors found that an increase in microbial diversity that normalized by 5 months post infection in all subjects; in addition to this they also found that numerous taxa usually related to the oral microbiome increased in relative abundance. In conclusion they found that although there were various changes in the respiratory microbiome the effect was somewhat unassertive, and the effects observed were also very variable among subjects (Cadena *et al.,* 2018).

Similar to previous studies by Khan *et al.* (2016) and (2019), the study by Dumas *et al.* (2018), investigated whether the host microbiome contributes to any form of protection against colonisation by MTB in a C57BL/6 mouse model. Although the study did not make use of NGS technology like many of the others, it did establish that Mucosal-Associated Invariant T (MAIT) cells may be involved in the early microbiota-mediated control of *M. tuberculosis* infection for example through the production of IL-17A. To demonstrate this effect, the authors made use of mice that have been depleted of microbes using broad spectrum anti-microbials in comparison to controls not receiving any treatment prior to MTB infection. A reduced number of MAIT cells were observed in mice treated with antibiotics prior to TB infection when compared to untreated animals. It is stated that these cells preferentially reside in the mucosal tissues of the lungs and gut and their development can be influenced by on the host microbiota. In addition to the lowered count of MAIT cells, the study also indicated that the correction of microbial dysbiosis through faecal transplant reversed these phenotypes within the antibiotics-treated animals, improving the ability of MAIT cells to proliferate. This study is complementary to the results obtained by Khan *et al*., (2016) and (2019), enforcing the concept of the gut-lung-axis and its role in response to TB by demonstrating that

the microbiome and its functional capacity has a protective effect against MTB infection and disease progression (Khan *et al.,* 2016; Dumas *et al.,* 2018; Khan *et al.,* 2019).

Overall results obtained from studies performed in mouse models seem to indicate that MTB infection in itself does cause a shift in the gut microbiome, but the magnitude of the shift is negligible in comparison to the more profound impact that can be observed with the administration of general TB drugs (Winglee *et al.,* 2014; Khan *et al.,* 2016; Namasivayam *et al.,* 2017). A common trend emerged from several of these studies, suggesting that the gut microbiome may play an important role in immune-modulation during the course of MTB infection (Winglee *et al.,* 2014; Khan *et al.,* 2016; Namasivayam *et al.,* 2017; Khan *et al*., 2019). In addition to this it is also evident that severe dysbacteriosis induced by TB therapy may further negatively affect the immune response, as demonstrated in the studies described by Khan *et al*. (2016), and (2019), where an improved response to infection was observed among animals that received faecal transplants and thus maintained a healthier microbiome.

6.2 Human studies – The gut microbiome

Studies focussing on MTB infection and the effect of treatment on the human microbiome are extremely limited; a summary of published studies on the gut microbiome during TB disease and treatment can be found in Table 3. The earliest such study involved a single case study of a patient in France. Although the study makes no significant mention of the importance or in fact relevance of the microbiome during TB progression and treatment; it very importantly states that caution should be applied during treatment of TB with broad-spectrum antibiotics and preventative measures need to be taken to avoid invasive infection by bacterial pathogens other than MTB, as the patient eventually succumbed to a *Streptococcus pneumoniae* infection (Dubourg *et al.*, 2013).

Like in the animal studies, the authors also mention that the gut microbiota is said to be able to regulate immune defences against infection, and that severe perturbation of the gut microbiome may have contributed to the outcome (Winglee *et al.,* 2014; Khan *et al.,* 2016; Namasivayam *et al.,* 2017). The study makes no mention of susceptibility

testing for the *Streptococcus pneumoniae* isolate in question, but several of the antibiotics prescribed to the patient (Amikacin, adiazine and cylcoserine) should have successfully treated the infection unless the isolate was resistant. Whether the *Streptococcus pneumoniae* isolate was present and merely proliferated due to the elimination of antibiotic sensitive bacteria or whether it was introduced as a nosocomial infection remains undetermined. There should thus be a cause for concern that this potentially multidrug resistant infection may have been a direct consequence of the long term antibiotic treatment administered to treat the MTB infection (Dubourg *et al.*, 2013; Baron *et al.,* 2018).

During a study by Wipperman *et al.* (2017), a total of four groups were enrolled in the study. The study included individuals with no MTB infection, latently infected individuals, individuals receiving TB therapy (INH, RIF, PZA and EMB) for drug sensitive tuberculosis and several cases cured of active MTB infection. Overall they found that there is a depletion of multiple species of *Bacteroides, Eubacterium, Lactobacillus* and *Ruminococcus*, among patients receiving TB therapy, accompanied by a simultaneous increase in *Erysipeloclostridium* and *Prevotella*. The authors also note that with the change in abundance of various species, there was also a significant shift in microbiome function. Many of these bacteria have been associated with host immunity and as such HRZE induced dysbacteriosis may thus have a significant effect on the host immune response (Wipperman *et al*., 2017; Lazar *et al*., 2018). The findings of this study also agree with the findings of the previous animal studies by Namasivayam *et al.* (2017) and Khan *et al.* (2019), noting that:

(i) overall diversity is preserved;
(ii) there is a significant decrease in the abundance of several bacterial genera in response to TB therapy;
(iii) the induced dysbacteriosis may have a significant effect on the host immune response;
(iv) it is clear that in both humans and mice there is a persistent and long lasting dysbacteriosis after the completion of TB therapy (Namasivayam *et al.* (2017; Wipperman *et al*., 2017).

Table 3: Studies involving TB and the gut microbiome

Study	Country	Specimen	Population group	Sample size	Time points	Key findings
Dubourg *et al.,* 2013	France	Stool	A case of MDR - TB with prior treatment failure	One subject, no controls	Single time point before onset of treatment	Gut flora significantly depleted by long term antibiotic treatment. Preventative measures need to be taken to avoid invasive infection
Luo *et al.,* 2017	China	Stool	New pulmonary tuberculosis subjects (NTB), recurrent pulmonary tuberculosis subjects (RTB) and healthy controls	37 subjects (19 NTB and 18 RTB) and 20 controls	Single time point, one week into treatment	*Prevotella* and *Lachnospira* significantly decreased in new and recurrent TB patient groups. CD4 T-cell counts positively correlated with these genera in new subjects, and *Escherichia* in recurrent subjects
Wipperman *et al.,* 2017	Haiti	Stool	Individuals with no or latent MTB infection, individuals currently on treatment and those previously cured of infection	38 subjects (19 currently on standard treatment and 19 that have previously been cured) and 75 controls (50 with no MTB infection and 25 with latent MTB infection).	Single time point during or after treatment	Overall diversity is not perturbed by treatment, but several immunologically significant commensals are depleted. This perturbation can persist for long after the end of therapy potentially at least 1-2 years
Maji *et al.*, 2018	India	Stool	Individuals identified as TB positive and one blood relative acting as a healthy control	6 subjects and 6 controls	Three time points. Before the onset of treatment, one week after treatment has begun and then one month after the start of treatment	Significant enrichment of propionate-producing bacteria

Hu *et al.,* 2019	China	Stool	TB patients (TB), newly diagnosed pulmonary TB patients (NTB) and healthy controls	46 subjects (30 TB, 16 NTB) and 31 controls	Single time point, during treatment or before the onset of treatment	Significant decrease in diversity. Decrease in short-chain fatty acids producing bacteria. *Haemophilus parainfluenzae*, *Roseburia inulinivorans*, and *Roseburia hominis* classification model and SNPs of *B. vulgatus* can assist with discrimination between healthy and diseased states
Wang *et al*., 2020	China	Stool	Untreated controls diagnosed with pulmonary TB, MDR-TB group receiving treatment and a MDR-TB group that have recovered.	6 MDR subjects receiving treatment, 18 subjects that have recovered and two untreated groups of 24 and 28 respectively	Single time point	MDR-TB treatment induced lasting gut dysbacteriosis, leading to adverse changes in patient lipid profiles. Changes associated with the following genera: *Adlercreutzia, Akkermansia, Butyricicoccus, Coprococcus, Clostridioides, Eubacterium, Erysipelatoclostridium, Fusicatenibacter, Klebsiella, Psychrobacter, Streptococcus.*
Namasivayam *et al*., 2020	Mali	Stool	Healthy controls, patients diagnosed with *Mycobacterium africanum* (MAF) infection and individuals diagnosed with MTB.	10 controls, 21 MTB subjects and 20 MAF subjects	Two time points, one prior to treatment and one two months after	MAF patients have microbiomes distinct from those of MTB patients, elevation Enterobacteriaceae that positively correlates with enhanced inflammatory gene expression in peripheral blood, reversed after initiation of TB therapy. MAF and MTB microbiomes may be reflective of the degree of difference in susceptibility of Africans to these infections

Abbreviations: New pulmonary tuberculosis subjects (NTB), recurrent pulmonary tuberculosis subjects (RTB), TB patients (TB), *M.tuberculosis* (MTB), *Mycobacterium africanum* (MAF), Single Nucleotide Polymorphism (SNP)

Although the Wipperman study did make use of Shotgun Metagenomic Sequencing for a sample group to evaluate functional changes in the gut microbial community, there is no mention of changes in the resistome associated with the taxonomic variations mentioned. Nevertheless, within the scope of this review the long lasting dysbiotic effect is significant since the study demonstrates both a shift in abundance and in functionality that persists long after treatment has ended (Wipperman *et al*., 2017). This would suggest that any perturbation of the resistome, or the proliferation of resistant genotypes that occurs during this period of time could potentially also persist long after treatment has ended (Wipperman *et al*., 2017; Gasparinni *et* al., 2019; Schwartz *et al*., 2020). In addition to the potential impact on the host immune response, it is clear from several of these studies that the abundance of healthy commensal bacteria in the gut tends to decrease during the course of TB therapy, leading to an increased chance for the proliferation of potentially resistant strains of bacteria, whether they be now resistant commensals or potential pathogens (Namasivayam *et al.,* 2017; Wipperman *et al*., 2017; Khan *et al.,* 2019).

This notion is further supported by a study conducted by Luo *et al*. (2017). The study shows that in patients with recurrent tuberculosis i.e. that have already undergone treatment and/or are still undergoing extended treatment had a clear increase of Actinobacteria and Proteobacteria known to contain many pathogenic species; while the phylum Bacteroidetes, which contains many healthy commensals, was reduced when comparing the recurrent TB patients with healthy controls. Luo *et al*. (2017), also demonstrated that there is a positive correlation between the gut microbiota and peripheral CD4+ T cell counts in the patients, and demonstrates that the relative abundance of the genera *Prevotella* and *Lachnospira* were associated with peripheral CD4+ T cell count in patients. Additionally, the study indicates that along with the reduction in *Lachnospira* there was also a reduction in the genus *Roseburia* in patients, which may impair short chain fatty acids (SCFAs) production. SCFAs such as butyrate are well known for playing an integral role in gut health, first by maintaining intestinal mucosa integrity and secondly by playing a role in the regulation of the host immune response. Similar to previous animal studies the authors conclude that specific gut microbes may be associated with the hosts immune status and could thus play a role in the prognosis and outcome of TB infection, but conclude that further identification

and characterization of the microbes involved is urgently required before any further conclusions can be made (Khan *et al.,* 2016; Luo *et al*., 2017; Dumas *et al.,* 2018; Khan *et al.,* 2019).

Another study evaluating the gut microbiota during MTB infection by Maji *et al.* (2018), found that there was a significant enrichment of propionate-producing bacteria like *Faecalibacterium*, *Roseburia*, *Eubacterium* and *Phascolarctobacterium* in TB patients. Both butyrate and propionate play an intricate role in host immunity and nutrient metabolism, thus strongly implicating the gut microbiome in the pathophysiology of MTB infection. The study found an increase in SCFA producers like *Faecalibacterium*, *Coprococcus*, *Phascolarctobacterium* and *Pseudobutyrivibrio*, in the TB infected group, and also demonstrated a trend of increased mean abundance of butyrate producers *Roseburia inulinivorans* and *Faecalibacterium prausnitzii* among the infected patient group. The authors conclude that this in turn leads to a decline in the biosynthesis of vitamins and amino acids in favour of increased metabolism of butyrate and propionate, but state that more elaborate studies are required to determine if changes SCFA production has any implications regarding TB. These finding are contradictory to the findings of Luo *et al.* (2017), who described a decrease in SCFA producers for example *Roseburia*.

A study by Hu *et al.* (2019), evaluating the gut microbiome during MTB infection investigated changes in the gut microbiota, focussing primarily on functional changes, by comparing newly identified tuberculosis patients in comparison to healthy controls by shotgun sequencing. The study reported significant decreases in microbial diversity; characterised by a great decrease in SCFA-producing bacteria as well as associated metabolic pathways. The study indicated that a classification model based on *Haemophilus parainfluenzae*, *Roseburia inulinivorans*, and *Roseburia hominis* allowed for the discrimination between healthy and diseased patients, and that discrimination between healthy and diseased states can also be conducted by assessing single nucleotide polymorphisms (SNPs) in *B. vulgatus* (Hu *et al.,* 2019). The noted decrease in SCFA producers is in agreement with the results obtained by Luo *et al*. (2017), and contradictory to the results of Maji *et al*. (2018), who reported an increase in SCFA producers. Taking this into account all three the aforementioned studies indicate that

the microbiome and its functional pathways play an intricate role in TB disease, and that any level of dysbiosis that affect these pathways may further contribute to TB infection, progression and outcome (Luo *et al.,* 2017; Maji *et* al., 2018; Hu *et al.,* 2019).

The contradictory results warrants further investigation before any conclusions can be drawn. One discernible difference observed between the two studies when doing an overall comparison of the methodology and study design, was that the two complimentary studies were conducted one Chinese population group, while the contradictory study was conducted on an Indian population group, which may indicate population specific response to TB infection and microbiome response in regards to TB infection (Luo *et al.,* 2017; Maji *et* al., 2018; Hu *et al.,* 2019). Similar to the study conducted by Wipperman *et al*. (2017) and Maji *et al.* (2018), Hu *et al.* (2019), also made use of Shotgun Metagenomic Sequencing for a sample group to evaluate functional changes in the gut microbial community, but similarly did not evaluate or mention any changes in the resistome at the time of writing (Wipperman *et al*., 2017; Maji *et al.,* 2018; Hu *et al*., 2019).

One of the more recent studies conducted by Namasivayam *et al*. (2020), evaluated changes in the gut microbiome in response to MAF and MTB infection. In addition to this a blood transcriptome analysis was also done, the results of which were cross-compared with the same individual's microbiota. In agreement with the study by Maji *et al*. (2018), lower levels of Prevotella were observed in response to MTB infection, an effect that was not seen in MAF infected patients. Additionally other impacted taxa that were observed in studies by Maji *et al*. (2018) and Luo *et al*. (2017), were not recognised as differentially enriched when comparing MTB patients with healthy controls during the course of this study. The authors note that these discrepancies may be due to differences in environmental, geographical and dietary induced differences of the populations included in the studies. One of the key findings revealed during the course of this study was that MAF patients have microbiomes distinct from those of MTB patients, and that an increase in Enterobacteriaceae positively correlated with increased inflammatory gene expression, an effect later reversed by the commencement of TB therapy. The study concluded that MAF and MTB microbiomes may indeed be reflective of the degree of difference in susceptibility of Africans to these

infections (Namasivayam *et al*., 2020). Additionally, the results of this study again confirmed that MTB infection causes only minor alterations in the intestinal microbiome in comparison to the alterations induced by TB therapy (Khan *et* al., 2019; Namasivayam *et al*., 2019; Namasivayam *et al*., 2020).

The most recently published study involving the gut microbiome and TB, and also the only one focussing primarily on MDR-TB, found that MDR-TB treatment induced long lasting gut dysbacteriosis, associated with diverse changes in patient lipid profiles. The alterations in lipid profiles were associated with several changes in the gut microbe composition, primarily involving the phylum Firmicutes and Verrucomicrobia and the genera: *Adlercreutzia, Akkermansia, Butyricicoccus, Coprococcus, Clostridioides, Eubacterium, Erysipelatoclostridium, Fusicatenibacter, Klebsiella, Psychrobacter*, and *Streptococcus.* A previous study by Wipperman *et al*. (2017), noted that dysbacteriosis of the gut microbiome can persist for up to 1.2 years after the cessation of standard TB treatment. In comparison the current study by Wang *et al*. (2020), found that as one would expect with the longer and more intensive treatment period required for MDR-TB, gut dysbacteriosis can last for much longer, in this instance between 3 and 8 years after recovery. The study by Wipperman *et al*. (2017) also demonstrated that changes in the functional profile also persist long after treatment has ended, and it seems in the case of MDR-TB treatment this also holds true, as the gut dysbacteriosis was associated with an adverse lipid profile including increased LDLC and TC long after treatment has ended (Wang *et al.,* 2020). This is especially concerning, as previously mentioned, this could mean that if any perturbations to the resistome were to occur during the course of MDR-TB treatment, these may persist for years after treatment has ended (Wipperman et al., 2017; Gasparinni et al., 2019; Schwartz et al., 2020; Wang *et al*., 2020).

6.3 Human studies – The respiratory tract microbiome

To date only 6 studies that focus on the microbiome and the respiratory tract during the course of TB disease progression and subsequent treatment have been found (Table 4). Studies by Cui *et al.* (2012) and Cheung *et al.* (2013), revealed that the microbiota in the sputum of pulmonary tuberculosis patients were more diverse than those of

healthy participants with the presence of several foreign bacteria. Both studies concluded that the foreign bacterial inhabitants may play a role in the onset or development of pulmonary tuberculosis. The first study by Cui *et al.* (2012), found several foreign bacteria such as *Comamonas, Cupriavidus, Diaphorobacter, Methylobacterium, Mobilicoccus, Pseudomonas, Stenotrophomonas, Sphingomonas* and *Thermus* to be present in patients with pulmonary tuberculosis. Contrary to the first study, the latter demonstrated that *Bacteroidetes* were enriched in individuals infected with MTB. Other phyla that were also found to be enriched included *Proteobacteria* in MTB patients, while *Firmicutes* were more predominant in controls. The study by Cheung *et al.* (2013), additionally identified 16 major genera, of which *Actinomyces, Fusobacterium, Leptotrichia, Prevotella, Streptococcus,* and *Veillonellaz* may represent the core genera of the TB sputum microbiome. Although the study found that there was a change in relative abundance, the diversity of microbiota was similar when comparing TB and control samples (Cheung *et al*., 2013). Neither of these studies continued monitoring the sputum microbiome of the enrolled subjects during and after treatment but do demonstrate that the sputum of pulmonary tuberculosis patients is an ecological niche that can support a high variety of bacteria, some of which may potentially be pathogenic (Cui *et al.*, 2012; Cheung *et al.*, 2013).

This phenomenon may be further exacerbated by treatment with antimicrobials, especially if healthy commensals are more susceptible to decline during the treatment process by giving these foreign potentially more resilient bacteria a chance to proliferate (Khan *et* al., 2016; Namasivayam *et al*., 2017; Khan *et al.,* 2019). Although the first two studies by Cui *et al.* (2012) and Cheung *et al.* (2013), laid the ground work to characterise the genera present in the sputum samples when comparing newly infected TP patients with healthy controls, unlike previous gut microbiome studies no possible functional changes were mentioned. When attempting to compare these studies to those done in animal models it becomes evident that there is a lack of focus on the respiratory tract among the animal studies and thus we do not yet know if these changes observed occur in animal models frequently used to study TB and if this could have any impact on future findings (Winglee *et al*. 2014; Khan *et al.* 2016; Namasivayam *et al*. 2017; Cadena *et al.* 2018; Dumas *et al*. 2018; Khan *et al.* 2019; Namasivayam *et al*. 2020).

Wu *et al*. (2013), described the first study to include a more comprehensive cohort of patients, and sought to evaluate the sputum microbiota associated with *M. tuberculosis* infection in newly infected, recurrent-, and treatment failure TB cases (Cui *et al.*, 2012; Cheung *et al.*, 2013; Wu *et al*., 2013). The study involved 75 cases divided into several categories including: patients with newly developed TB, pulmonary TB patients who had been treated, cured and relapsed and a treatment failure group. The study thus represents examples of the sputum microbiome in various conditions. However, the studies did include any attempt to monitor these changes during the course of treatment, *i.e.* to obtain several samples from the same patient over a period of time. The Wu *et al.* (2013), study did find reduced frequency and abundance of some genera in recurrent TB patients. In addition, they found that the ratio *Pseudomonas/Mycobacterium* was higher in recurrent TB than that in new TB patients, while *Treponema/Mycobacterium* was lower in recurrent TB patients when compared to that of new TB patients. Disruption of these bacteria due to antibiotic therapy for example may thus be considered as a risk factor for recurrence (Wu *et al*., 2013). Similar to the results obtained by Cui *et al*. (2012), the authors also note that the microbiome of TB patients was distinct from that of healthy controls, but unlike the previous study they found that the healthy microbiome is more diverse. This also differs with the results obtained by Cheung *et al*. (2013), who found no significant differences in diversity between controls and TB sputum microbiomes. The authors elaborate on several potential reasons for this including increased sequencing depth, differences in sample size and the individuals used as controls. These differences demonstrate the urgent need for standardised methods and control groups when studying TB and the microbiome (Cui *et al*., 2012; Cheung *et al*., 2013; Wu *et al.,* 2013).

Table 4: Studies involving TB and the respiratory tract microbiome

Study	Country	Specimen	Population group	Sample size	Time points	Key findings
Cui *et al.,* 2012	China	Sputum and oropharyngeal pharyngeal secretions collected by deep coughing	New pulmonary tuberculosis subjects (NTB) and healthy controls	31 subjects and 24 controls	Single time point, before onset of treatment	Respiratory tract microbiome of pulmonary tuberculosis subjects is more complicated than that of healthy controls.
Wu *et al.,* 2013	China	Sputum and oropharyngeal swabs	New pulmonary tuberculosis subjects (NTB), recurrent pulmonary tuberculosis subjects (RTB), treatment failure subjects (TF) and healthy controls	75 subjects (25 NTB, 30 RTB and 20 TF) and 20 controls	Before or after treatment	*Streptococcus*, *Gramulicatella* and *Pseudomonas* enriched in TB subjects. *Pseudomonas* more abundant and more commonly present in TF subjects. Certain bacteria and the dysbiosis of lung microbiota may be related to both the onset of TB but also its recurrence.
Cheung *et al.,* 2013	China	Sputum	New pulmonary tuberculosis subjects and healthy controls	22 subjects and 14 controls	Single time point, before onset of treatment	*Actinomyces, Fusobacterium, Leptotrichia, Prevotella, Streptococcus*, and *Veillonella* found in all TB samples. Microbiome of TB subjects enriched in *Moryella, Mogibacterium*, and *Oribacterium.*
Botero *et al.,* 2014	Colombia	Sputum, oropharyngeal and nasal specimens	Pulmonary TB subjects and healthy controls	6 subjects and 6 controls	Single time point, before onset of treatment	Respiratory tract microbial communities were comparable in terms of the core phyla identified, but varied in terms of relative abundance and diversity. Oropharynx samples can be utilized to study community structure changes associated with TB.
Zhou *et al.,* 2015	China	Bronchoalveolar lavage fluid and saliva specimens	Pulmonary TB subjects and healthy controls	32 subjects and 24 controls	Two weeks after the onset of treatment at a single time point	More *Cupriavidus* in subjects; *Mycobacteria* and *Porphyromonas* inside TB lesions may be a co-factor in lesion formation.
Krishna *et al.,* 2016	India	Sputum	Pulmonary TB subjects and healthy controls	23 subjects and 16 controls	Single time point, before onset of treatment	Presence of diverse opportunistic pathogens in TB subjects increases microbiome complexity. More Firmicutes and Actinobacteria in subjects vs controls; *Neisseria, Veillonella* and *Streptococcus* the most dominant genera in subjects.

Hu *et al.* 2020	China	Bronchoalveolar lavage fluid	Pulmonary TB patients (MTB+) and patients representing with respiratory symptoms but testing TB negative (MTB-)	30 MTB+ subjects and 30 MTB- subjects	Single time point	MTB+ and MTB- lung microbial communities have distinct diversity and functional signatures. 16S rDNA sequencing underrepresents Mycobacterium.

Abbreviations: New pulmonary tuberculosis subjects (NTB), recurrent pulmonary tuberculosis subjects (RTB), treatment failure subjects (TF), TB patients (TB)

Regardless of the conflicting results, the increased presence of some specific bacteria, such as *Pseudomonas*, which may be associated with treatment outcome especially in recurrent and treatment failure patients, is particularly noteworthy. Although bacterial co-infection with MTB is not a common occurrence, a recent study has demonstrated that co-infection does occur (Attia *et al.,* 2019). Attia *et al.* (2019), found that around 33% of the tuberculosis patients that formed part of their study had a bacterial co-infection and that *Klebsiella* and *Pseudomonas* spp. were commonly isolated. *Pseudomonas aeruginosa* is one of the leading causes of morbidity and mortality in both chronic obstructive pulmonary disease and cystic fibrosis, and although the study by Wu *et al.* (2013) did not conclusively detect these species in samples, it remains a point of concern. *Pseudomonas* spp. such as *Pseudomonas aeruginosa* are known to be susceptible to RIF, especially in combination with other antimicrobials, but within the context of standard TB therapy, RIF and various other non-*Pseudomonas* specific antimicrobials may be employed at sub-inhibitory concentrations, which may lead to the selection of resistant mutants, with future consequences (Boyer *et* al., 2011; Hu *et al*., 2019)

The origin of collected samples, as well as the sampling procedure (see Chapter 3), are factors to consider when analysing the microbiome. Botero *et al.* (2014), examined changes in the microbiota of the sputum, oropharynx, and nasal respiratory tract, in patients with pulmonary TB in comparison to healthy controls. Although the study did not mention any major findings relating to TB and the microbiome, they did suggest that both oropharyngeal and sputum specimens had similar bacterial and fungal taxa, and that these were distinct from that of the nasal specimens. The study concludes that even though there are differences in abundance when comparing oropharyngeal and sputum specimens, the diversity of microbes remained the same; suggesting that oropharynx samples can thus be utilized to study community structure changes associated with TB (Botero *et al.,* 2014). During a more recent study utilising Broncho alveolar lavage fluid instead of sputum, significant changes were also observed when comparing the microbiome of the respiratory tract of TB patients when compared to healthy controls. The patients in this study were placed on a two week long TB treatment regiment, consisting of INH, RIF, PZY and EMB. The authors found a significant change when comparing the microbiota of the lower respiratory tract of TB patients in comparison to healthy controls (Zhou *et al.,* 2015).

In addition to this they also evaluated the microbiota of intra/extra-TB lesions. The results from this study revealed *Cupriavidus* as the most dominant bacterial genus, and furthermore found that the difference in abundance of *Mycobacteria* and *Porphyromonas* inside TB lesions may also be a co-factor in lesion formation. Although the study by Zhou *et al*. (2015) focussed on the microbiota in intra- or extra-TB lesion areas, it again highlights that it is difficult to draw any direct comparison to the previous studies due to different sampling methods as well as the addition of TB treatment, especially since the majority of studies have taken samples from individuals in various disease states at single time points. This draws attention to the fact that there is a urgent need for longitudinal studies examining the microbiomes of the same patients starting from infection all the way through treatment until cured (Cui *et al*., 2012; Cheung *et al*., 2013; Wu *et al.,* 2013; Zhou *et al.,* 2015).

The most recently published study focussing on the respiratory tract microbiome of an Indian population group, reported significantly different core microbiome when compared to those identified by previous studies focussing on a Chinese population (Cheung *et al*., 2013; Krishna *et al*., 2016). It is important to note that similar to studies by Cui *et al.* (2012) and Cheung *et al.* (2013), patients did not receive any TB drugs prior to sampling, and changes were thus primarily attributed to the effects of MTB infection (Krishna *et al*., 2016). The study demonstrates that although communities are similar at the phylum level, there is a difference in representation, for example, there was an increase in Firmicutes in Indian TB patients contrary to the Chinese population where Firmicutes were more dominant in control samples. Upon comparison of the genus level, the authors also found a different core group including *Actinomyces, Rothia, Granulicatella, Lactobacillus, Streptococcus, Veillonella, Leptotrichia*, and *Neisseria* as opposed to those found in previous studies (Cheung *et al*., 2013; Krishna *et al*., 2016).

Most importantly and as highlighted by Khan *et al.* (2016) as well; the presence of diverse opportunistic pathogens in TB patients further increases the complexity of the sputum microbiome. As such the characterisation of the sputum microbiome is likely to provide important pathogenic insights into pulmonary TB. The presence, emergence

and potential proliferation of opportunistic pathogens is of particular concern as it has not yet been elucidated whether their proliferation is due to the elimination of antibiotic sensitive bacteria, while the resistant microbes flourish (Khan *et al*., 2016; Krishna *et al*., 2016). The study by Krishna *et al*. (2016) also highlights the importance of taking population groups into account when attempting to elucidate the role of the microbiome in TB since the microbiome, although similar at the phylum level, differs quite significantly at the genus level based on the specific population group included in the study (Cheung *et al*., 2013; Krishna *et al*., 2016).

The most recent pilot study, and the only one thus far to investigate the respiratory tract by means of shotgun metagenomics, revealed a lower alpha diversity in the MTB+ group and a significant difference based on beta diversity when comparing the MTB+ and MTB- patients. Employing shotgun metagenomics the authors were also able to do a functional comparison of the microbiome between MTB+ and MTB- patients, discovering a significantly reduced gene abundance for PRPP biosynthesis in MTB+ patients, corroborated by a decreased abundance of genes associated with pentose phosphate and purine metabolism; additionally one pathway for aerobic respiration was also found to be more active in the microbiome of MTB+ patients (Hu *et al*., 2020). The authors simultaneously also evaluated the fungal microbiome associated with the respiratory tract and found that fungal sequences were primarily classified as belonging to the phylum Ascomycota and Basidiomycota, a result in agreement with previous finding by means of ITS sequencing by Botero *et al*. (2014).

Very importantly the study demonstrated that MTB+ lung samples were dominated by MTB, a phenomenon not previously demonstrated by any previous studies that made use of a 16S sequencing approach (Cui *et* al., 2012; Cheung *et* al., 2013; Wu *et* al., 2013; Hu *et al*., 2020). Mycobacteria are notoriously difficult to lyse, and as such he authors suggest that low abundances may be observed due to insufficient lysis; or more likely due to the fact that MTB tends to have only one or two 16S copies per genome, and are thus expected to be under-represented by 16S rDNA amplicon sequencing (Hu *et al*., 2020). Surprisingly several of the previous studies evaluating the respiratory microbiome and TB, have made use of kits that only incorporate chemical lysis without any mechanical disruption, and no mention is made of additional steps to ensure

efficient lysis of MTB cells. This may thus very well be one of the main reasons for underrepresentation of *Mycobacteria* in 16S sequencing data along with the fact that MTB have very few 16S copies (Cui *et* al., 2012; Cheung *et* al., 2013; Wu *et* al., 2013; Botero *et al*., 2014; Zhou *et al*., 2015; Hu *et al*., 2020).

However, care also needs to be taken with the results of the study by Hu *et al*. (2020), as the study in itself is not without its limitations. The study does not mention if samples were stored or processed immediately after collection. In regards to DNA extraction the study incorporated the QIAamp DNA Microbiome Kit (Catalogue 51704, Qiagen, Hilden, Germany), a kit that includes both mechanical and chemical lysis steps, but also includes a host DNA depletion step. This method greatly reduces the sequencing depth required to evaluate the respiratory microbiome by removing host contaminating DNA, but may in some instances lead to biased results. As stated in the Kit Handbook "Freeze–thaw cycles can compromise bacterial integrity, so the Benzonase treatment for the degradation of host DNA might lead to loss of exposed bacterial DNA." If the integrity of any of the easier to lyse Gram-negative microbes have been impacted by pre-treatment or storage at low temperatures prior to the host DNA removal steps, the DNA of these organisms may also be removed, and as such results could be skewed towards the more difficult to lyse bacteria such as Gram-positive and/or *Mycobacteria* (Stinson *et al*., 2018; Hu *et al*., 2020). The study also pooled samples prior to sequencing, although previous studies have demonstrated this to be a feasible strategy it may be problematic once it comes to the evaluation of host microbiomes in response to treatment (Ray *et al*., 2019; Hu *et al*., 2020). A recent study by Namasivayam *et al*. (2020), observed a large degree of inter-individual variability in antibiotic induced microbiome perturbations, which may impact results if samples are pooled (Namasivayam *et al*., 2020). Nevertheless, the study is the first of its kind to investigate the airway microbiome using a shotgun sequencing approach, delivering a lot of meaningful results, highlighting the need for both 16S and/or follow-up shotgun sequencing in order to truly evaluate TB and the microbiome. The study did not go as far as to evaluate the resistome, but very importantly it highlighted the urgent need for standardization of sample collection and sequencing methods when investigating TB microbiome interactions (Hu *et al*., 2020).

At the time of writing there were no studies available on the respiratory tract microbiota and TB in regards to the impact of long term TB therapy on changes in the respiratory microbiome; or any studies focussing on functional changes or impacts of treatment on the resistome. None the less the data presented by the above mentioned studies suggests that the presence of certain bacteria and dysbiosis of the respiratory tract microbiota may be associated with the onset of TB, its recurrence and treatment failure. The lung microbiota thus potentially plays a role in pathogenesis and treatment outcome of TB and should be considered during treatment and control of TB (Cui *et al*., 2012; Cheung *et al*., 2013; Wu *et al.,* 2013; Zhou *et al.,* 2015; Krishna *et al*., 2016)

7. *Mycobacterium tuberculosis*, the microbiome, resistome and TB therapy: key messages, research opportunities, and concluding remarks

Upon examination of the various studies conducted so far, various factors requiring attention can be identified. With regards to technical variation, consideration needs to be given to the origin of sample, the sampling method, the DNA extraction and sequencing methodology employed as well as population groups included in these studies (Fiedorova *et al.,* 2019; Sui *et al*., 2020). Several bacteria present in the human microbiome, like MTB*,* are incredibly hard to lyse and the employment of a cell lysis method that does not adequately lyse these cells may lead to a skewed representation. Care thus needs to be taken to ensure adequate lysis of the entire microbial community (Fiedorova *et al.,* 2019; Greathouse *et al*., 2019; Sui *et al*., 2020). A large degree of variation can also be introduced by the sequencing methodology employed. For example, long *vs*. short reads or 16S *vs.* whole metagenome. The majority of the studies reviewed relied on 16S metagenomic sequencing and although cost effective, it can only provide taxonomic information down to the genus level, and provide a prediction of functionality. Whole genome shotgun sequencing on the other hand, which is employed by some of the above mentioned studies, can provide species/strain level resolution and provide a much more comprehensive picture of functionality. The latter method is more costly and computationally intensive but has several advantages over standard 16S metagenomic sequencing, allowing the user to comprehensively

evaluate all genes present in the microbial community, including those of a functional capacity (Jovel *et al.,* 2016; Brumfield *et al.,* 2020). Additional consideration needs to be given to sequencing depth, as some of the studies reviewed relied on relatively low sequencing depth, which may lead to rare genera or species being overlooked (Cui *et al.,* 2012; Wu *et al.,* 2013).

The sample types and population groups used also need consideration; stool samples have their drawbacks but remain one of the most frequently used non-invasive sample types used to evaluate changes in the gut microbiome (Liang *et al*., 2020). Obtaining the ideal respiratory sample remains a more complicated affair. Broncho alveolar lavage would be ideal but the procedure is invasive and requires specialist personnel to obtain the sample. More research is required as to which less invasive sample type, e.g. sputum, is best suited for the continued study of the TB microbiome. As with patients, similar care and standardisation are required with regards to controls (Cui *et al.*, 2012; Cheung *et al.*, 2013; Hu *et al.,* 2019; Hu *et al.,* 2020). Recent studies have shown that there are significant variations in the microbiome composition of healthy individuals from different races and ethnicities, often making it difficult to infer direct relationships between the microbiome and various states of health and disease (Fettweis *et al.,* 2014; Kemppainen *et al.,* 2015; Gupta *et al.,* 2017). The human microbiome composition from various body sites has been characterised over the past several years, and it has been shown that the skin, vaginal, gut, oral and even respiratory tract microbiomes differ depending on various factors including race, geography, ethnicity, subsistence or exposure to environmental variables (Fettweis *et al.,* 2014; Kemppainen *et al.,* 2015; Leung *et al.,* 2015; Li *et al*., 2014; Gupta 2017).

A comparison of the studies by Cheung *et al.* (2013) and Krishna *et al*. (2016), highlights the importance of inter-population variation; the authors noted that although similar at the phylum level, the respiratory tract microbiome composition differs quite significantly at the genus level. However, inclusion of vulnerable subpopulations such as HIV positive individuals in any of the studies was lacking or scarce. Eight countries account for the majority of new TB cases, with India leading the count, followed by, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa. Several of these countries, such as Nigeria and South Africa also have a relatively

large HIV burden, necessitating that TB and microbiome research be expanded to include these high risk population groups (Bruchfeld *et al.,* 2015; World Health Organisation, 2019).

More longitudinal research on the microbiome is required to evaluate the impact of disease progression as well as the impact of treatment in animal models. No study other than that of Namasivayam *et al.* (2017), monitored the microbiome throughout the entire process of both infection and TB treatment. A similar trend was observed with studies that included human participants. There is also little evidence of a concurrent evaluation of the respiratory tract and the gut microbiome. As a matter of fact, several studies have emphasized the importance of the gut lung axis especially in a primarily pulmonary disease like tuberculosis (Cervantes *et al*., 2017; Dumas *et al*., 2018). Current evidence suggest that there is bidirectional immunological interaction between the gut and the lung microbiome during MTB infection, disease progression and possibly during the course of treatment, and that dysbacteriosis of the microbiome may impact disease outcomes. The study by Dumas *et al*. (2018), was the only one to simultaneously evaluate both the gut and lung during MTB infection and found that treatment induced dysbacteriosis may lead to impaired metabolism of alveolar macrophages. Several studies have also highlighted the importance of SCFAs, and that the microbiomes functional pathways, especially that of the gut play an intricate role in TB disease. At this point in time, contradictory results between studies suggest that more research is required before any conclusions can be drawn (Luo *et al.,* 2017; Maji *et al.,* 2018; Hu *et al.,* 2019).

This review highlights a major gap in research – that of the resistome and its involvement with long term TB treatment, as well as the potential for the proliferation of resistant bacteria during the course of TB therapy. The human gut has been shown to provide an ideal environment for the occurrence of horizontal gene transfer between commensal flora as well as to pathogenic bacteria that pass through the intestine, and has been very well characterised as a reservoir for antimicrobial resistance determinants (van Schaik, 2015; Casals-Pascual *et al.*, 2018; Li *et al.,* 2019; Willmann *et al.,* 2019). Even though the gut resistome has been very well characterised, studies

have begun to assess the impact of generally used TB drugs on the microbiome only recently.

The presence of ARG's in the respiratory microbiome has not been comprehensively studied in any regard, but various studies focussing on patients with cystic fibrosis (CF) have indicated that the respiratory tract microbiome may act as a reservoir for antimicrobial resistant bacteria and their associated genes (Dickson *et al*., 2016; Taylor *et al.*, 2018). Cystic fibrosis is a chronic pulmonary disease caused by mutations in the cystic fibrosis transmembrane conductance regulator gene, leading to thickened mucus secretions, which cannot be easily cleared. The mucus is colonised by bacteria and enables the development and perseverance of pulmonary bacterial infection (Sherrard *et al.*, 2014). Antibiotic therapy has contributed tremendously to increased survival rates of people with CF, but in many instances chronic use of multiple antibiotics is required to improve life expectancy (Chmiel *et al*., 2014). The chronic use of multiple antibiotics has increased the potential for the development of multidrug resistance and several studies have indicated that the respiratory tract of CF patients receiving regular antimicrobial therapy do carry an unappreciated diversity of ARG's (Chmiel *et al*., 2014; Sherrard *et al.*, 2014). Upon review of previous studies on CF, it becomes clear that the respiratory tract can, as is the case with the gut microbiome, act as a reservoir for ARG's under the right conditions, especially where there is chronic or long term infection and thus also frequent exposure to antimicrobials. Again, as with the gut microbiome, other pulmonary conditions requiring long term antimicrobial therapy with several antimicrobials, like TB infection has not previously been evaluated in this regard (Hong *et al.*, 2016; Namasivayam *et al*., 2017; Kidd *et al.*, 2018; Hu *et al.,* 2019; Hu *et al.,* 2020).

A comprehensive search of literature did not reveal any studies dedicated to the impact of TB therapy on the resistome but several studies reviewed did indicate long lasting perturbation of the microbiome in terms of both diversity and functionality following anti-tuberculosis therapy (Namasivayam *et* al., 2017; Wipperman *et al*., 2017). Additionally, there is some evidence that this long term treatment does potentially generate a unique ecological niche where commensals are depleted, giving rise to the possible enrichment of more resilient bacteria, which may be pathogens or opportunistic

pathogens (Khan *et al.,* 2016; Namasivayam *et al.,* 2017; Khan *et al.,* 2019). It is thus imperative that we continue to evaluate the potential for the development and progression of antimicrobial resistance amongst bacteria of the various human microbiomes, especially where there is poor regulation of, poor adherence to and a high prevalence in the use of antibiotics (Llor & Bjerrum, 2014; Ventola, 2015; Martens & Demain, 2017).

References

Abdool Karim, S.S., Churchyard, G.J., Karim, Q.A. & Lawn, S. D. 2009. HIV infection and tuberculosis in South Africa: An urgent need to escalate the public health response. *Lancet*, 374(9693):921–933.

Andries, K., Verhasselt, P., Guillemont, J., Neefs, J., Winkler, H., Gestel, J. Van, Timmerman, P., Zhu, M., Lee, E., Williams, P., Chaffoy, D. De, Huitric, E., Hoffner, S., Cambau, E., Truffot-pernot, C., Lounis, N. & Jarlier, V. 2005. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science*, 307(57070):223–228.

Attia, E.F., Pho, Y., Nhem, S., Sok, C., By, B., Phann, D., Nob, H., Thann, S., Yin, S., Noce, R., Kim, C., Letchford, J., Fassier, T., Chan, S. & West, T.E. 2019. Tuberculosis and other bacterial co-infection in Cambodia: a single center retrospective cross-sectional study. *BMC Pulmonary Medicine,* 19(60).

Axelsson, M. 2013. Report on personality and adherence to antibiotic therapy : a population-based study. *BMC Psychology,* 1(1): 24.

Barko, P.C., McMichael, M.A., Swanson, K.S. & Williams, D.A. 2018. The Gastrointestinal Microbiome: A Review. *Journal of Veterinary Internal Medicine*, 32(1), 9–25.

Baron, S.A., Diene, S.M. & Rolain, J. 2018. Human microbiomes and antibiotic resistance. *Human Microbiome Journal,* 10:43-52.

Botero, L.E., Delgado-serrano, L., Cepeda, M.L., Bustos, J.R., Anzola, J.M., Portillo, P. Del, Robledo, J. & Zambrano, M.M. 2014. Respiratory tract clinical sample

selection for microbiota analysis in patients with pulmonary tuberculosis. *BMC Microbiome*, 2:29.

Boyer, A., Doussau, A., Thiébault, R., Venier, A. G., Tran, V., Boulestreau, H., Bébéar, C., Vargas, F., Hilbert, G., Gruson, D. & Rogues, A.M. 2011. *Pseudomonas aeruginosa* acquisition on an intensive care unit: relationship between antibiotic selective pressure and patients' environment. *Critical Care,* 15(1).

Bruchfeld, J., Correia-Neves, M. & Källenius, G. 2015. Tuberculosis and HIV coinfection. *Cold Spring Harbor perspectives in medicine*, 5(7):a017871.

Brumfield, K.D., Huq, A., Colwell, R.R., Olds, J.L. & Leddy, M.B. 2020. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PloS One*, 15(2):e0228899.

Krishna, P., Jain, A., & Bisen, P.S. 2016. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. *European Journal of Clinical Microbiology & Infectious Diseases*, 35(7):1205–1210.

Cadena, A.M., Ma, Y., Ding, T., Bryant, M., Maiello, P., Geber, A., Lin, P.L., Flynn, J.L. & Ghedin, E. 2018. Profiling the airway in the macaque model of tuberculosis reveals variable microbial dysbiosis and alteration of community structure. *Microbiome*, 6(1):180.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. & Darst, S.A. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, 104(6):901–912.

Casals-Pascual, C., Vergara, A. & Vila, J. 2018. Intestinal microbiota and antibiotic resistance : Perspectives and solutions. *Human Microbiome Journal*, 9:11–15.

Centis, R., Caminero, J.A. & Migliori, G.B. 2017. Effectiveness and safety of clofazimine in multidrug-resistant tuberculosis : a nationwide report from Brazil. *European Respiratory Journal,* 1:9–13.

Cervantes, J. & Hong, B. 2017. The gut–lung axis in tuberculosis. *Pathogens and Disease,* 75(8).

Chang, K.C., Yew, W.W. & Zhang, Y. 2011. Pyrazinamide susceptibility testing in *Mycobacterium tuberculosis* : a Systematic review with meta-analyses. *Antimicrobial Agents and Chemotherapy,* 55(10):4499–4505.

Cheung, M.K., Lam, W.Y., Fung, W.Y.W., Law, P.T.W., Au, C.H., Nong, W., Kam, K.M., Kwan, H.S. & Tsui, S.K.W. 2013. Sputum microbiota in tuberculosis as revealed by 16S rRNA Pyrosequencing. *PLoS One*, 8(1).

Chmiel, J.F., Aksamit, T.R., Chotirmall, S.H., Dasenbrook, E. C., Elborn, J. S., LiPuma, J.J., Ranganathan, S.C., Waters, V.J. & Ratjen, F.A. 2014. Antibiotic management of lung infections in cystic fibrosis. II. Nontuberculous mycobacteria, anaerobic bacteria, and fungi. *Annals of the American Thoracic Society*, 11(8):1298–1306.

Cui, Z., Zhou, Y., Li, H., Zhang, Y., Zhang, S., Tang, S. & Guo, X. 2012. Complex sputum microbial composition in patients with pulmonary tuberculosis. *BMC Microbiology,* 12:276–283.

Delogu, G., Sali, M. & Fadda, G. 2013. The biology of *Mycobacterium Tuberculosis* infection. *Mediterranean Journal of Haematology and Infectious Diseases,* 5(1):2013070.

Dickson, R.P., Erb-Downward, J.R., Martinez, F.J. & Huffnagle, G.B. 2016. The microbiome and the respiratory tract. *Annual Review of Physiology*, 78:481–504.

Dookie, N., Rambaran, S., Padayatchi, N., Mahomed, S. & Naidoo, K. 2018. Evolution of drug resistance in *Mycobacterium tuberculosis*: a review on the molecular determinants of resistance and implications for personalized care. *Journal of Antimicrobial Chemotherapy,* 73(5):1138-1151.

Dubourg, G., Lagier, J.C., Armougom, F. & Robert, C. 2013. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *European Journal of Clinical Microbiology and Infectious Diseases,* 32(5):637-45.

Dumas, A., Corral, D., Colom, A., Levillain, F., Peixoto, A., Hudrisier, D., Poquet, Y. & Neyrolles, O. 2018. The host microbiota contributes to early protection against lung colonization by *Mycobacterium tuberculosis*. *Frontiers in immunology*, 9:2656.

Durão, P., Güleres, D., Proença, J. & Gordo, I. 2016. Enhanced survival of rifampin- and streptomycin-Resistant *Escherichia coli* inside macrophages. 60(7):4324–4332.

Ehlers, S. & Schaible, U.E. 2013. The granuloma in tuberculosis : dynamics of a host – pathogen collusion. *Frontiers in Immunology,* 3:411.

Fettweis, J.M., Brooks, J.P., Serrano, M.G., Sheth, N.U., Girerd, P.H., Edwards, D.J., Strauss, J.F., Consortium, M., Jefferson, K.K., Buck, G.A. & Buck, G.A. 2014. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology (Reading, England),* 160(10):2272-2282.

Fiedorová, K., Radvanský, M., Němcová, E., Grombiříková, H., Bosák, J., Černochová, M., Lexa, M., Šmajs, D. & Freiberger, T. 2019. The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery. *Frontiers in microbiology*, 10:821.

Fouhy, F., Ogilvie, L.A., Jones, B. V, Ross, R.P., Ryan, A.C., Dempsey, E.M., Fitzgerald, G.F., Stanton, C. & Cotter, P.D. 2014. Identification of aminoglycoside and b-lactam resistance genes from within an infant gut functional metagenomic library. *PLoS One,* 9(9):1–10.

Garneau-tsodikova, S. & Labby, K.J. 2016. Mechanisms of resistance to aminoglycoside antibiotics: overview and perspectives. *Medchemcomm,* 7(1):11–27.

Gasparrini, A.J., Wang, B., Sun, X., Kennedy, E.A., Hernandez-Leyva, A., Ndao, I.M., Tarr, P.I., Warner, B.B. & Dantas, G. 2019. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nature Microbiology,* 4:2285–2297.

Ghanbari, M., Klose, V., Crispie, F. & Cotter, P.D. 2019. The dynamics of the antibiotic resistome in the feces of freshly weaned pigs following therapeutic administration of oxytetracycline. *Scientific Reports,* 9(4062):1-11.

Greathouse, K.L., Sinha, R. & Vogtmann, E. 2019. DNA extraction for human microbiome studies: the issue of standardization. *BMC Genome Biology,* 20(212).

Gupta, V.K., Paul, S. & Dutta, C. 2017. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Frontiers in Microbiology*, 8(1162).

Heiss, C.N. & Olofsson, L.E. 2018. Gut microbiota-dependent modulation of energy metabolism. *Journal of Innate Immunity,* 10(3):163–171.

Hong, B.Y., Maulén, N.P., Adami, A.J., Granados, H., Balcells, M.E. & Cervantes, J. 2016. Microbiome Changes during tuberculosis and antituberculous therapy. *Clinical microbiology reviews*, 29(4):915–926.

Hooper, D.C. 2001. Emerging mechanisms of fluoroquinolone resistance. *Emerging Infectious Diseases,* 7(2):337–341.

Houben, R.M.G.J. & Dodd, P.J. 2016. The global burden of latent tuberculosis infection: A re-estimation using mathematical modelling. *PLoS Medicine,* 13(10):1–13.

Hu, Y., Feng, Y., Wu, J., Liu, F., Zhang, Z. & Hao, Y. 2019. The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. *Frontiers in Cellular and Infection Microbiology,* 9(90):1–8.

Hu, Y., Cheng, M., Liu, B., Dong, J., Sun, L., Yang, J., Chen, X. & Jin, Q. 2020. Metagenomic analysis of the lung microbiome in pulmonary tuberculosis - a pilot study. *Emerging Microbes and Infections,* 9(1):1444-1452.

Hughes, J. & Osman, M. 2014. Diagnosis and management of drug-resistant tuberculosis in South African adults. *South African Medical Journal,* 104(12):894.

Ilievska-Poposka, B., Metodieva, M., Zakoska, M., Vragoterova, C. & Trajkov, D. 2018. Latent tuberculosis infection - diagnosis and treatment. *Open Access Macedonian Journal of Medical Sciences,* 6(4):651–655.

Jabeen, K., Shakoor, S. & Hasan, R. 2015. Fluoroquinolone-resistant tuberculosis: implications in settings with weak healthcare systems. *International Journal of Infectious Diseases,* 32:118–123.

Jeeves, R.E., Marriott, A.A.N., Pullan, S.T., Hatch, K.A., Allnutt, J.C., Freire-Martin, I., Hendon-Dunn, C.L., Watson, R., Witney, A.A., Tyler, R.H., Arnold, C., Marsh, P.D., McHugh, T.D. & Bacon, J. 2015. *Mycobacterium tuberculosis* is resistant to isoniazid at a slow growth rate by single nucleotide polymorphisms in katg codon ser315. *PLoS One*, 10(9):1–21.

Jeong, H., Arif, B., Caetano-anollés, G., Kim, K.M. & Nasir, A. 2019. Horizontal gene transfer in human- associated microorganisms inferred by phylogenetic reconstruction and reconciliation. *Scientific Reports,* 9:5953.

Jin, M., Lu, J., Chen, Z., Nguyen, S.H., Mao, L., Li, J., Yuan, Z. & Guo, J. 2018. Antidepressant fluoxetine induces multiple antibiotics resistance in *Escherichia coli* via ROS-mediated mutagenesis. *Environment International,* 120:421–430.

Johnson, M.M. & Odell, J.A. 2014. A review article: Nontuberculous mycobacterial pulmonary infections. *Journal of Thoracic Disease,* 6(3):210–220.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., & Wong, G.K. 2016. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, 7(459).

Kemppainen, K.M., Ardissone, A.N., Davis-Richardson, A.G., Fagen, J.R., Gano, K. A., León-Novelo, L.G., Vehik, K., Casella, G., Simell, O., Ziegler, A.G., Rewers, M.J., Lernmark, Å., Hagopian, W., She, J.X., Krischer, J.P., Akolkar, B., Schatz, D.A., Atkinson, M.A., Triplett, E.W. & TEDDY Study Group. 2015. Early childhood gut

microbiomes show strong geographic differences among subjects at high risk for type 1 diabetes. *Diabetes Care*, 38(2):329–332.

Khan, N., Vidyarthi, A., Nadeem, S., Negi, S., Nair, G. & Agrewala, J. N. 2016. Alteration in the gut microbiota provokes susceptibility to tuberculosis. *Frontiers in Immunology*, 7(529).

Khan, N., Mendonca, L., Dhariwal, A., Fontes, G., Menzies, D., Xia, J. & Divangahi, M. 2019. *Mycobacterium tuberculosis* intestinal dysbiosis compromises alveolar macrophage immunity to *Mycobacterium tuberculosis*. *Mucosal Immunology,* 12:772–783.

Kidd, T.J., Canton, R., Ekkelenkamp, M., Johansen, H.K., Gilligan, P., LiPuma, J.J., Bell, S.C., Elborn, J.S., Flume, P.A., VanDevanter, D.R. & Waters, V.J. 2018. Defining antimicrobial resistance in cystic fibrosis. *Journal of Cystic Fibrosis*, 17(6):696–704.

Kim, C.H. 2018. Immune regulation by microbiome metabolites. *Immunology*. 154(2):220–229.

Krause, K.M., Serio, A.W., Kane, T.R. & Connolly, L.E. 2016. Aminoglycosides: An overview. *Cold Spring Harbor Perspectives in Medicine*, 6(6):a027029.

Lazar, V., Ditu, L.M., Pircalabioru, G.G., Gheorghe, I., Curutiu, C., Holban, A.M., Picu, A., Petcu, L. & Chifiriuc, M.C. 2018. Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. *Frontiers in Immunology*, 9:1830.

Lechartier, B. & Cole, S.T. 2015. Mode of action of clofazimine and combination therapy with benzothiazinones against *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy,* 59(8):4457–4463.

Leung, M.H.Y., Wilkins, D. & Lee, P.K.H. 2015. Insights into the pan-microbiome : skin microbial communities of Chinese individuals differ from other racial groups. *Scientific Reports,* 5:11845.

Li, J., Quinque, D., Horz, H., Li, M., Rzhetskaya, M., Raff, J.A., Hayes, M.G. & Stoneking, M. 2014. Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC Microbiology,* 14:316.

Li, B., Qiu, Y., Song, Y., Lin, H. & Yin, H. 2019. Dissecting horizontal and vertical gene transfer of antibiotic resistance plasmid in bacterial community using microfluidics. *Environment International*, 131:105007.

Liang, D., Ka, R., Leung, K., Guan, W. & Au, W.W. 2018. Involvement of gut microbiome in human health and disease : brief overview, knowledge gaps and research opportunities. *Gut Pathogens,* 10:3.

Liang, Y., Dong, T., Chen, M., He, L., Wang, T., Liu, X., Chang, H., Mao, J.H., Hang, B., Snijders, A.M. & Xia, Y. 2020. Systematic analysis of impact of sampling regions and storage methods on fecal gut microbiome and metabolome profiles. *mSphere,* 5:e00763-19.

Lloyd-price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., Mccracken, C., Giglio, M.G., Mcdonald, D. & Franzosa, E.A. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature,* 550:61–66.

Llor, C. & Bjerrum, L. 2014. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic advances in drug safety*, 5(6):229–241.

Luo, M., Liu, Y., Wu, P., Luo, D.X., Sun, Q., Zheng, H., Hu, R., Pandol, S.J., Li, Q.F., Han, Y.P. & Zeng, Y. 2017. Alternation of gut microbiota in patients with pulmonary tuberculosis. *Frontiers in Physiology,* 8:822.

Maartens, G. & Benson, C.A. 2015. Linezolid for treating tuberculosis: A delicate balancing act. *EBioMedicine*, 2(11):1568–1569.

MacLean, R.C. & San Millan, A. 2019. The evolution of antibiotic resistance. *Science*, 365(6458):1082-1083.

Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E. & Brochado, A.R. 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*, 555:623–628.

Maji, A., Misra, R., Dhakan, D.B., Gupta, V., Mahato, N.K., Saxena, R., Mittal, P., Thukral, N., Sharma, E., Singh, A., Virmani, R., Gaur, M., Singh, H., Hasija, Y., Arora, G., Agrawal, A., Chaudhry, A., Khurana, J.P., Sharma, V.K., Lal, R. & Singh, Y. 2018. Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. *Environmental Microbiology,* 12:772–783.

Maningi, N.E., Daum, L.T., Rodriguez, J.D., Said, H.M., Peters, R.P.H., Sekyere, J.O., Fischer, G.W., Chambers, J.P. & Fourie, P.B. 2018. Multi- and extensively drug

resistant *Mycobacterium tuberculosis* in South Africa: a molecular analysis of historical isolates. *Journal of Clinical Microbiology*, 56:e01214-17.

Martens, E. & Demain, A.L. 2017. The antibiotic resistance crisis, with a focus on the United States. *The Journal of Antibiotics,* 70(5):520-526.

Martino, M. De, Lodi, L., Galli, L. & Chiappini, E. 2019. Immune response to *Mycobacterium tuberculosis*: A narrative review. *Frontiers in Paediatrics,* 7:350.

McGuffin, S.A., Pottinger, P.S. & Harnisch, J.P. 2017. Clofazimine in nontuberculous mycobacterial infections: A growing niche. *Open Forum Infectious Diseases,* 4(3):1–7.

Mcilleron, H. & Chirehwa, M.T. 2019. Current research toward optimizing dosing of first-line antituberculosis treatment. *Expert Review of Anti-infective Therapy,* 17(1):27–38.

Moore, M., Patel, S., Forsberg, K.J., Wang, B., Bentley, G., Razia, Y., Qin, X., Tarr, P.I. & Dantas, G. 2013. Paediatric faecal microbiota harbour diverse and novel antibiotic resistance genes. *PLoS One,* 8(11).

Mulubwa, M. & Mugabo, P. 2019. Steady-state population pharmacokinetics of terizidone and its metabolite cycloserine in patients with drug-resistant tuberculosis. *British Journal of Clinical Pharmacology,* 85(9):1946-1956.

Naidoo, P., Theron, G., Rangaka, M.X., Chihota, V.N., Vaughan, L., Brey, Z.O., & Pillay, Y. 2017. The South African tuberculosis care cascade: Estimated losses and methodological challenges. *The Journal of Infectious Diseases,* 216(7):702–713.

Namasivayam, S., Maiga, M., Yuan, W., Thovarai, V., Costa, D.L., Mittereder, L.R., Wipperman, M.F., Glickman, M.S., Dzutsev, A., Trinchieri, G. & Sher, A. 2017. Longitudinal profiling reveals a persistent intestinal dysbiosis triggered by conventional anti-tuberculosis therapy. *Microbiome*, 5(1):71.

Namasivayam, S., Kauffman, K.D., McCulloch, J.A., Yuan, W., Thovarai, V., Mittereder, L.R., Trinchieri, G., Barber, D.L. & Sher, A. 2019. Correlation between disease severity and the intestinal microbiome in *Mycobacterium tuberculosis*-infected rhesus macaques. *mBio*, 10(3):e01018-19.

Namasivayam, S., Diarra, B., Diabate, S., Sarro, S., Kone, A., Kone, B., Tolofoudie, M., Baya, B., Diakite, T., Kodio, O., Cohen, K., Id, J.H., Achenbach, C.J., Chatterjee, S., Murphy, R.L., Bishai, W., Diallo, S., Sher, A. & Id, M.M. 2020. Patients infected with *Mycobacterium africanum* versus *Mycobacterium tuberculosis* possess distinct intestinal microbiota. *PLoS Neglected Tropical Diseases,* 14(5):1–20.

Nash, A.K., Auchtung, T.A., Wong, M.C., Smith, D.P., Gesell, J.R., Ross, M.C., Stewart, C.J., Metcalf, G.A., Muzny, D.M., Gibbs, R.A., Ajami, N.J. & Petrosino, J.F. 2017. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome*, 5(1):153.

O'Connor, J.A. 1948. Para-aminosalicylic acid in the treatment of tuberculosis. *Postgraduate Medical Journal*, 24(275): 455–462.

O'Donnell, M. R., Daftary, A., Frick, M., Hirsch-Moverman, Y., Amico, K. R., Senthilingam, M., Wolf, A., Metcalfe, J. Z., Isaakidis, P., Davis, J. L., Zelnick, J. R., Brust, J. C., Naidu, N., Garretson, M., Bangsberg, D. R., Padayatchi, N. & Friedland, G. 2016. Re-inventing adherence: toward a patient-centered model of care for drug-resistant tuberculosis and HIV. *The international journal of tuberculosis and lung*

disease: the official journal of the International Union against Tuberculosis and Lung Disease, 20(4):430–434.

Orme, I.M. & Basaraba, R.J. 2014. The formation of the granuloma in tuberculosis infection. *Seminars in Immunology,* 26(6):601–609.

Pérez-Cebrián, M., Suárez-Varela, M.M.M., Font-Noguera, I., Monte-Boquet, E., Poveda-Andrés, J.L., Martín-Moreno, J.M., Rubio-López, N., Ruiz-Rojo, E. & Llopis-González, A. 2015. Study on the linezolid prescription according to the approval of indication in a university hospital. *Iranian Journal of Pharmaceutical Research,* 14(3):857–864.

Pontali, E., Sotgiu, G., Tiberi, S., Tadolini, M., Visca, D., Ambrosia, L., Centis, R., Spanevello, A. & Migliori, G.B. 2018. Combined treatment of drug-resistant tuberculosis with bedaquiline and delamanid: A systematic review. *European Respiratory Journal,* 52:1800934.

Ray, K.J., Cotter, S.Y., Arzika, A.M., Kim, J., Boubacar, N., Zhou, Z., Zhong, L., Porco, T.C., Keenan, J.D., Lietman, T.M. & Doan, T. 2019. High-throughput sequencing of pooled samples to determine community-level microbiome diversity. *Annals of Epidemiology*, 39:63–68.

Redgrave, L.S., Sutton, S.B., Webber, M.A. & Piddock, L.J. V. 2014. Fluoroquinolone resistance: Mechanisms, impact on bacteria, and role in evolutionary success. *Trends in Microbiology,* 22(8):438–445.

Relman, D.A. & Lipsitch, M. 2018. Microbiome as a tool and a target in the effort to address antimicrobial resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 115(51):12902–12910.

Saifullah, B., Arulselvan, P., Ezzat, M., Zowalaty, E., Fakurazi, S., Webster, T.J., Geilich, B. & Hussein, M.Z. 2014. Development of a highly biocompatible antituberculosis nanodelivery formulation based on para-aminosalicylic acid — zinc layered hydroxide nanocomposites. *The Scientific World Journal,* 2014:401460.

Schwartz, D.J., Langdon, A.E. & Dantas, G. 2020. Understanding the impact of antibiotic perturbation on the human microbiome. *Genome Medicine*, 12(1):82.

Sherrard, L.J., Tunney, M.M. & Elborn, J.S. 2014. Antimicrobial resistance in the respiratory microbiota of people with cystic fibrosis. *Lancet*, 384(9944):703–713.

Silva, D.R., Dalcolmo, M., Tiberi, S., Arbex, M.A., Munoz-Torrico, M., Duarte, R., D'Ambrosio, L., Visca, D., Rendon, A., Gaga, M., Zumla, A. & Migliori, G.B. 2018. New and repurposed drugs to treat multidrug- and extensively drug-resistant tuberculosis. *Jornal brasileiro de pneumologia: publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia*, 44(2):153–160.

Sinha, P., Gupta, A., Prakash, P., Anupurba, S., Tripathi, R. & Srivastava, G.N. 2016. Differentiation of *Mycobacterium tuberculosis* complex from non-tubercular mycobacteria by nested multiplex PCR targeting IS6110, MTP40 and 32kD alpha antigen encoding gene fragments. *BMC Infectious Diseases,* 16(1):1–10.

Sommer, M.O.A. & Dantas, G. 2011. Antibiotics and the resistant microbiome. *Current Opinion in Microbiology*, 14(5):556–563.

Stagg, H.R., Lipman, M.C., McHugh, T.D. & Jenkins, H.E. 2017. Isoniazid-resistant tuberculosis: a cause for concern?. *The international journal of tuberculosis and lung*

disease: the official journal of the International Union against Tuberculosis and Lung Disease, 21(2):129–139.

Stinson, L.F., Keelan, J.A. & Payne, M.S. 2018. Comparison of meconium DNA extraction methods for use in microbiome studies. *Frontiers in microbiology*, 9(270).

Streicher, E.M., Bergval, I., Dheda, K., Böttger, E.C., Pittius, N.C.G. Van, Bosman, M., Coetzee, G. & Anthony, R.M. 2012. *Mycobacterium tuberculosis* population structure determines the outcome of genetics-based second-line drug resistance testing. *Antimicrobial Agents and Chemotherapy,* 6(5):2420-7.

Sui, H.Y., Weil, A.A., Nuwagira, E., Qadri, F., Ryan, E.T., Mezzari, M. P., Phipatanakul, W. & Lai, P.S. 2020. Impact of DNA extraction method on variation in human and built environment microbial community and functional profiles assessed by shotgun metagenomics sequencing. *Frontiers in Microbiology*, 11:953.

Sultan, I., Rahman, S., Jan, A.T. & Siddiqui, M.T. 2018. Antibiotics, resistome and resistance mechanisms : A bacterial perspective. *Frontiers in Microbiology,* 9: 2066.

Swanson, R. V, Ammerman, N.C., Ngcobo, B., Adamson, J., Moodley, C., Dorasamy, A., Moodley, S., Mgaga, Z., Bester, L.A., Singh, S.D., Almeida, D.V. & Grosset, J.H. 2016. Clofazimine contributes sustained antimicrobial activity after treatment cessation in a mouse model of tuberculosis chemotherapy. *Antimicrobial Agents and Chemotherapy,* 60(5):2864–2869.

Swanson, K.S. & Williams, D.A. 2018. The gastrointestinal microbiome: A review. *Journal of Veterinary Internal Medicine,* 32(1): 9–25.

Taylor, S.L., Leong, L.E.X., Mobegi, F.M., Choo, J.M., Burr, L.D., Wesselingh, S. & Rogers, G.B. 2018. Understanding the impact of antibiotic therapies on the respiratory

tract resistome: A novel pooled-template metagenomic sequencing strategy. *Multidisciplinary Respiratory Medicine,* 13(1).

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-liggett, C.M., Knight, R. & Gordon, J.I. 2007. The Human Microbiome Project. *Nature,* 449:804–810.

Ursell, L.K., Metcalf, J.L., Parfrey, L.W. & Knight, R. (2012). Defining the human microbiome. *Nutrition Reviews*, 70(1):38–44.

Vannelli, T.A., Dykman, A. & Ortiz De Montellano, P.R. 2002. The antituberculosis drug ethionamide is activated by a flavoprotein monooxygenase. *Journal of Biological Chemistry,* 277(15):12824–12829.

van Schaik, W. 2015. The human gut resistome. *Philosophical Transactions of the Royal Society of London Series B Biological sciences*, 370(1670):20140087.

Wang, J., Xiong, K., Zhao, S., Zhang, C., Zhang, J. & Xu, L. 2020. Long-term effects of multi-drug-resistant tuberculosis treatment on gut microbiota and its health consequences. *Frontiers in Microbiology,* 11:53.

Willmann, M., Vehreschild, M.J.G.T., Biehl, L.M., Vogel, W., Dörfel, D., Hamprecht, A., Seifert, H., Autenrieth, I.B. & Peter, S. 2019. Distinct impact of antibiotics on the gut microbiome and resistome: a longitudinal multicenter cohort study. *BMC Biology,* 17(76).

Winglee, K., Eloe-fadrosh, E., Gupta, S., Guo, H. & Fraser, C. 2014. Aerosol *Mycobacterium tuberculosis* infection causes rapid loss of diversity in gut microbiota. *PLoS One,* 9(5):1–9.

Wipperman, M.F., Fitzgerald, D.W., Juste, M.A.J., Taur, Y., Namasivayam, S., Sher, A., Bean, J.M., Bucci, V. & Glickman, M.S. 2017. Antibiotic treatment for tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. *Scientific Reports,* 7(1):1–11.

World Health Organization. 2010. Guidelines for treatment of tuberculosis. [Web: https://apps.who.int/iris/bitstream/handle/10665/44165/9789241547833_eng.pdf?sequence=1]. Date of access: 23 March 2018.

World Health Organization. 2018. Global tuberculosis report 2020. [Web: https://www.World Health Organisation.int/teams/global-tuberculosis-programme/tb-reports]. Date of access: 24 April 2019.

World Health Organization. 2019. Global tuberculosis report 2020. [Web: https://www.World Health Organisation.int/teams/global-tuberculosis-programme/tb-reports]. Date of access: 10 May 2019.

World Health Organization. 2020a. Global tuberculosis report 2020. [Web: https://www.World Health Organisation.int/teams/global-tuberculosis-programme/tb-reports]. Date of access: 20 July 2020.

World Health Organization. 2020b. WHO Consolidated Guidelines on Tuberculosis, Module 4: Treatment - Drug-Resistant Tuberculosis Treatment. [Web: https://www.who.int/publications/i/item/9789240007048]. Date of access: 12 January 2020.

Wu, J., Liu, W., He, L., Huang, F., Chen, J., Cui, P., Shen, Y., Zhao, J., Wang, W., Zhang, Y., Zhu, M., Zhang, W. & Zhang, Y. 2013. Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PLoS One*. 8(12):1–11.

Yamane, K., Wachino, J.I., Doi, Y., Kurokawa, H. & Arakawa, Y. 2005. Global spread of multiple aminoglycoside resistance genes. *Emerging Infectious Diseases,* 11(6):951–953.

Zhang, Y., Wade, M.M., Scorpio, A., Zhang, H. & Sun, Z. 2003. Mode of action of pyrazinamide : disruption of *Mycobacterium tuberculosis* membrane transport and energetics by pyrazinoic acid. *Journal of Antimicrobial Chemotherapy,* 52(5):790–795.

Zheng, J., Rubin, E.J., Bifani, P., Mathys, V., Lim, V., Au, M., Jang, J., Nam, J., Dick, T., Walker, J.R., Pethe, K. & Camacho, L.R. 2013. para-Aminosalicylic acid is a prodrug targeting dihydrofolate reductase in Mycobacterium tuberculosis. *The Journal of biological chemistry*, 288(32):23447–23456.

Zhou, Y., Lin, F., Cui, Z., Zhang, X., Hu, C., Shen, T., Chen, C., Zhang, X. & Guo, X. 2015. Correlation between either Cupriavidus or Porphyromonas and primary pulmonary tuberculosis found by analysing the microbiota in patients' bronchoalveolar lavage fluid. *PLoS One*, 10(5):1–18.

Zhu, C., Liu, Y., Hu, L., Yang, X.M. & He, Z. 2018. Molecular mechanism of the synergistic activity of ethambutol and isoniazid against *Mycobacterium tuberculosis*. *Journal of Biological Chemistry,* 293(43):16741–16750.

Chapter 3 – Initial evaluations of the NWU lysis method for downstream microbiome analyses

This chapter consists of a data article accepted for publication in *Data in Brief* pending final minor revisions. Supplementary material mentioned in the manuscript can be found at the link to Mendeley Data mentioned in the “Data Accessibility” section of the specifications table within the manuscript. The author guidelines followed in submitting to the Journal can be found in Annexure A. The manuscript is presented in this thesis and has been published. Since the journal format allows only for the presentation of data sets generated during the experimental process, the remainder of the information relating to the initial evaluations is represented in article format, following the *Data in Brief* manuscript. The chapter addresses the first objective of the study as outlined in Chapter 1. Personal contributions to the below manuscript included the following: Initial conceptualization, preforming all wet laboratory work associated with the manuscript, assisting with the generation of the below mentioned workflow, data generation, testing, application and data analyses with the presented workflow and also personally writing the final manuscript.

1. Manuscript 1

Article Title

MinION 16S datasets of a commercially available microbial community enables the evaluation of DNA extractions and data analyses

Authors

Mann, B.C.[1], Bezuidenhout, J.J.[2*], Swanevelder Z.H.[3] & Grobler, A.F. [4]

Affiliations

[1] DSI/NWU Preclinical Drug Development Platform, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa.

[2] Unit for Environmental Science and Management: Microbiology, North-West University, Potchefstroom Campus, South Africa.

[3] Biotechnology Platform, Agricultural Research Council, Pretoria, South Africa.
[4]Austell Pharmaceuticals, Sherborne Road, Parktown, South Africa.

***Corresponding author(s)**

jaco.bezuidenhout@nwu.ac.za

Abstract

New advances in sequencing technology and bioinformatics analysis tools have significantly supported the culture-independent analysis of complex microbial communities associated with environmental, plant, animal and human samples. However, previous work has shown that DNA extraction can have a major influence in the community profile. As such there is a constant need for new methods to efficiently and rapidly prepare and analyze DNA for microbiome research, especially in the case new and emerging technology like the Oxford Nanopore Technologies (ONT) MinION. A commercial standard was used, in triplicate, to evaluate three DNA extraction

protocols, including two commercially available and one ”in-house” DNA extraction method. All DNA extractions were done as per manufacturer’s instructions and prepared with the same commercial ONT 16S sample preparation kit, prior to being analysed using MinION sequencing. Eight MinION 16S datasets of this microbial reference community were obtained. Reads were initially base called and demultiplexed using ONT’s Guppy™ sequencing software (version 3.2.4), filtered using NanoFilt and then classified using Usearch. A set of R scripts are presented to process sintax files generated from Usearch and produce an OTU table that can be used for further analyses. All datasets were deposited into the SRA (NCBI) database. These datasets will allow future extraction kit comparisons using MinION sequencing since a standardize laboratory process using commercially available components, such as the MinION 16S sample preparation kit, microbial reference community and extraction kits, were used. The current ONT 16S workflow making use of the Epi2me agent only provides QC metrics and the ID’s of the main genera identified and does not provide any tools currently for further downstream community comparison. The analyses scripts provided in the supplementary material will thus further enable the testing of new datasets against these reference sets and provide users the ability to compare their workflows with ours, thus standardising comparisons and workflows.

Keywords

Full length 16S meta-barcoding microbiome, reference community standard, long read sequencing

Specifications Table

Subject	Biological sciences
Specific subject area	Targeted reference metagenomic comparison of DNA extractions

Type of data	Raw fastq files Filtered and merged fastq files Table Figure Diagram R scripts
How data were acquired	ONT MinION platform was used for sequencing of eight (9) 16S amplicon libraries
Data format	Raw and analysed
Parameters for data collection	DNA from a single ZymoBIOMICS™ Microbial Community Standard (Zymo, USA) was extracted with two commercially available kits, and one in-house developed method. All 16S amplicon libraries were prepared with the same sequencing preparation kit prior to sequencing according to the recommended ONT 16S Barcoding protocol. All reactions were performed in triplicate.
Description of data collection	DNA from the ZymoBIOMICS™ Microbial Community Standard was extracted using three methods. The two commercially available kits included a standard beat beating kit (GenElute™ Stool DNA Isolation Kit, Sigma, USA) and a kit with a host DNA removal step prior to DNA extraction (QIAamp DNA microbiome kit, Qiagen, Germany). The in-house kit used a “lyses micro tube” to extract DNA. 16S libraries were generated and multiplexed using same the ONT 16S Barcoding Kit (SQK-RAB204, ONT, UK). MinION sequencing was carried out with the aid of the MinKNOW software

	(ONT, UK), with the fast5 files obtained converted to fastq with the ONT's Guppy™ sequencing software (version 3.2.4. Resulting fastq sequences were analysed with available opensource software and a set of R scripts included as part of the dataset.
Data source location	Institution: North-West University City/Town/Region: North-West, Potchefstroom Country: South Africa GPS coordinates: Latitude: -26.7167 Longitude: 27.1000
Data accessibility	Raw fastq and, filtered and merged MinION sequence data is available at NCBI under the BioProject No. PRJNA675451. SRA accession numbers: SRR13011317 - SRR13011324 and SRR13632466 - SRR13632459. SRA link: https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=search_obj&m=&s=&term=PRJNA675451&go=Search Supplementary material has been deposited on Mendeley data at the following link: https://data.mendeley.com/datasets/yhv4rsr426/

Value of the Data

- The work presented provides raw fastq files obtained following sequencing of a standard microbial community on the MinION nanopore sequencer along with the already filtered and merged fastq files. Additionally, a simplified, rapid workflow for the analyses of ONT 16S reads is provided with the MinION dataset to act as a standard workflow for comparison.
- The data sets provided generates useful information for researchers involved in the application of long read 16S metagenomics especially those working with 16S data obtained by MinION sequencing.
- The 16S sequencing data provided is of a commercially available microbial community standard of known composition, and as such, can be used to test newly developed laboratory workflows by comparing datasets with these already existing workflows needed to process the 16S data produced by the MinION platform. The workflow provided with the R scripts will enable the testing of new workflows and datasets against the data sets provided as a reference.

1. Data Description

Samples for the sequencing dataset were based on a comparison study of various DNA extraction methods using a reference microbial community (ZymoBIOMICS™ Microbial Community Standard, Zymo, USA). Dataset 1 represents the obtained raw fastq data after sequencing by the ONT MinION 16S Barcoding Kit, as well as the filtered and merged fastq files that can be used directly with the supplied workflow of dataset 2. Dataset 2 represent a series of text files (Supplementary material) containing instructions and R script for processing of the data once quality control and demultiplexing have been done. The set of files included in the supplementary files include the R scripts (Database and sintax generation.txt, Script 1 – Multiple files.txt, Script 2 – Final version.txt, Script 3 – Final version.txt, Script 4 – Final version.txt, Script 5 – Final version.txt) applied in the workflow as depicted in Table 1. Additionally the supplementary files also contain the expected outputs (st_pre.txt, presum.txt, for_xls.txt, MA_OTU.txt, Summary_out.xlsx) when processing the supplied fastq files for each sample with the provided workflow from Table 1 as well as an example mapping file (Mappingfile_eg.txt). The MinION 16S sequencing dataset contains a total of 5, 427, 602 reads. Table 1 provides a summary of the supplied pipeline that was established for bioinformatics analysis of the long

sequencing reads obtained from a single sequencing run on a MinION™. Table 2 contains a summary of the observed, and thus expected DNA quality metrics seen after extraction of the ZymoBIOMICS™ Microbial Community Standard with the 3 applied methods, while Table 3 contains a summary of the subsamples following sequencing – sample Q1 failed at PCR and thus no values were recorded. Figure 1 represents an example of the abundance graphs that can be drawn using the excel output from step 5 the workflow described in Table 1. Figures 2, 3 and 4 represents the expected outputs such as alpha (chao1, observed and Shannon index), beta diversity and heatmap examples generated as part of the workflow.

Table 1: Summarised workflow steps for the simple and rapid analyses of 16S reads produced by the ONT MinION 16S barcoding

Step	Description	Software	Input file(s)	Output files and data
1	Install Usearch	Usearch	NA	NA
2	Database generation	Usearch	Downloaded database file (RDP 16S training set v16 (RTS))	Database ready for classification
3	Classification	Usearch	Merged and filtered fastq files for each demultiplexed sample	.SINTAX file ready for processing using R
4	Script 1 – Pre-processing	R	.SINTAX file/files	Temporary output file for Script 2 (st_pre.txt)
	Script 2 – Pre-processing	R	Temporary file for Script 2 (st_pre.txt)	Temporary output file for Script 3 (presum.txt)
	Script 3 – Pre-processing	R	Temporary file for Script 3 (presum.txt)	Temporary output file for Script 4 (for_xls.txt)
5	Script 4 – Generate Excel summary	R	Temporary file for Script 4 (for_xls.txt)	Excel summary
6	Script 5 – Generate OUT table	R	Temporary file for Script 4 (for_xls.txt)	MA_OTU.txt
7	Import into Microbiome analyst	Web application	MA_OTU.txt Mapping file (generated by user, see example supplied in supplementary material)	

Table 2: Summary of expected DNA quality following extraction of the ZymoBIOMICS™ Microbial Community Standard. Sample H refers to the average values obtained for the in-house kit, S refers to the GenElute™ Stool DNA Isolation Kit and Q to the QIAamp DNA microbiome kit

Method	Nanodrop Concentration (ng/µL)	Qubit Concentration (ng/µL)	A260/A280	A260/A280
H	27.73 ± 0.87 [c]	15.23 ± 3.72 [b]	1.9 ± 0.01 [a]	1.72 ± 0.08 [a,b]
S	20.46 ± 2.02 [b]	1.33 ± 0.21 [a]	1.82 ± 0.03 [a]	1.76 ± 0.02 [b]
Q	12.8 ± 1.23 [a]	1.97 ± 0.09 [a]	1.81 ± 0.02 [a]	1.4 ± 0.1 [a]

[a,b,c] Methods sharing a common letter belong to the same group according to one way ANOVA with Tukey post-hoc test for multiple pairwise comparisons. These values do not differ significantly according to the test ($p > 0.05$). Values not sharing a common letter indicate a statistically significant difference in comparison to a value belonging to other groups ($p < 0.05$).

n = 3

Table 3. SRA dataset matrices, including each triplicate repeat's, raw reads, filtered reads, and percentage of surviving reads after filtering that should be expected after running the presented data analyses pipeline. Sample H refers to the in-house kit repeats, S refers to the repeats for the GenElute™ Stool DNA Isolation Kit and Q to the repeats for the QIAamp DNA microbiome kit

Sample	Barcode	Reads prior to processing	Reads post filtering	Percentage retained
H1	1	219,020	170,943	78.05%
H2	2	113,7251	759,743	66.81%
H3	3	365,529	279,836	76.56%
S1	4	151,1825	993,650	65.73%
S2	5	443,793	351,756	79.26%
S3	6	973,672	648,025	66.55%
Q1*	7	NA	NA	NA
Q2	8	523,226	370,027	70.72%
Q3	9	253,286	166,910	65.90%

* Sample Q1 failed at PCR

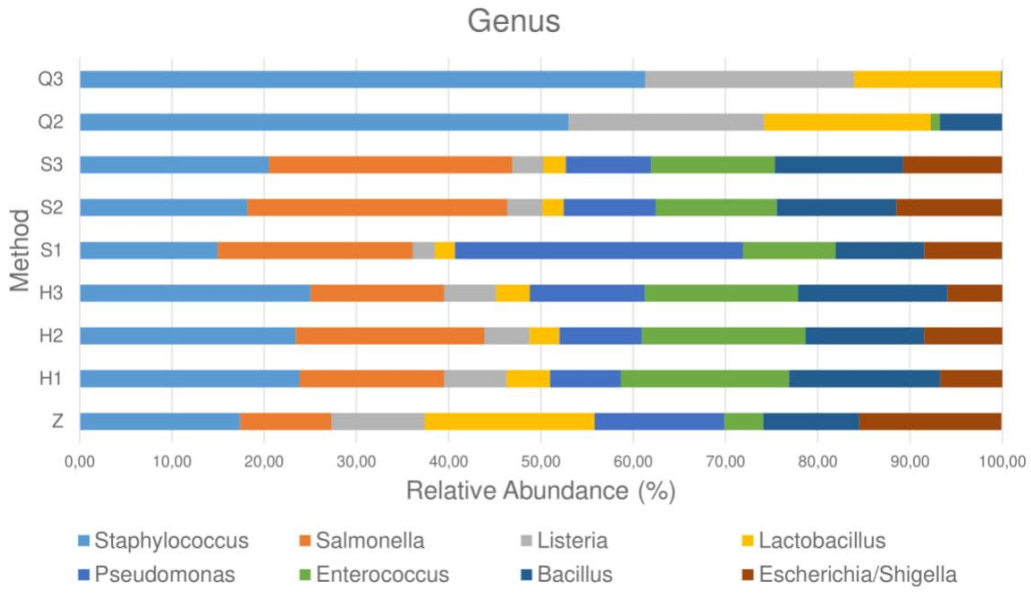


Figure 1: Relative abundance chart of the bacterial genera from each sample above a 0.01% abundance cut-off. Subsamples extracted with the GenElute™ Stool DNA Isolation Kit, QIAamp DNA microbiome kit and the in-house methods have been labelled S, Q and H respectively. The expected relative abundance from the published reference community standard is labelled as Z in the figure.

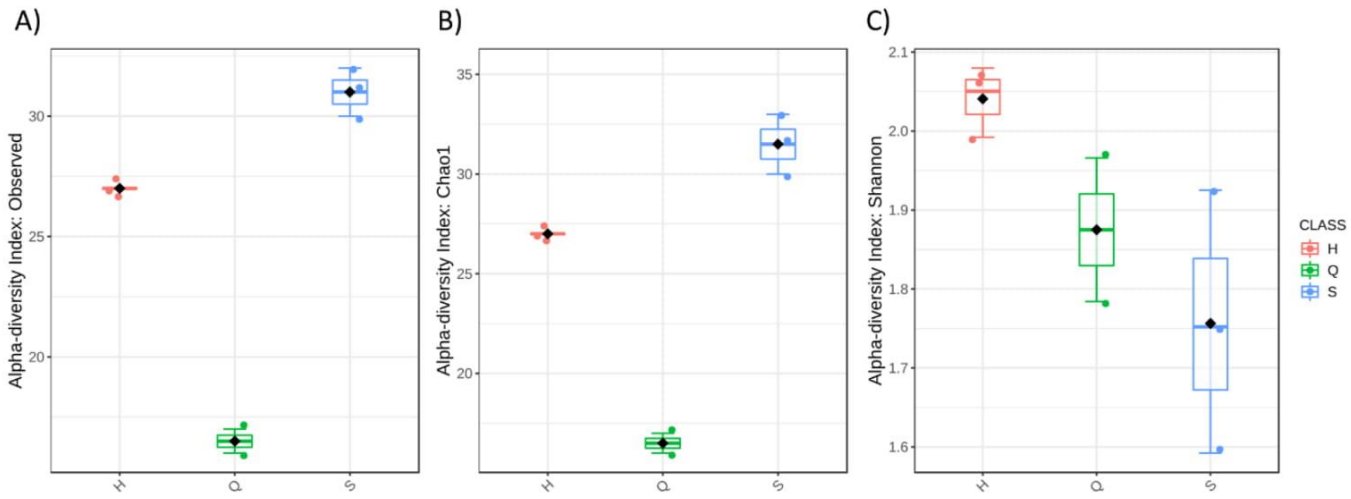


Figure 2: An example of Alpha-diversity, measured by Observed (A), Chao1 (B) and Shannon diversity (C) indexes for the ZymoBIOMICS™ Microbial Community Standard.

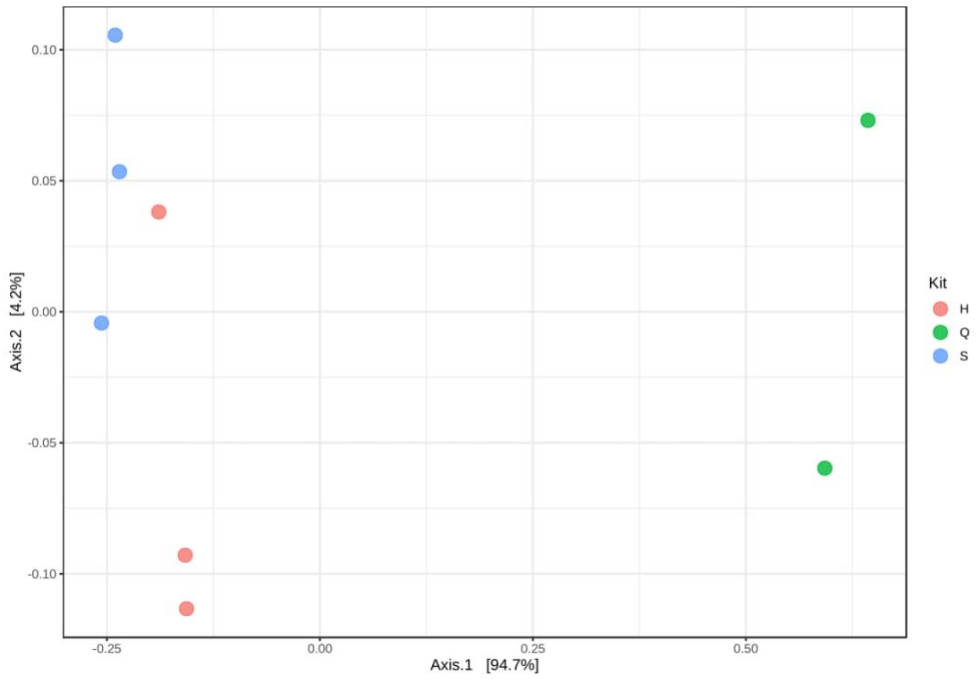


Figure 3: An example of Principal coordinates analysis based on Bray–Curtis distances of samples extracted by various methods for the ZymoBIOMICS™ Microbial Community Standard using the presented workflow

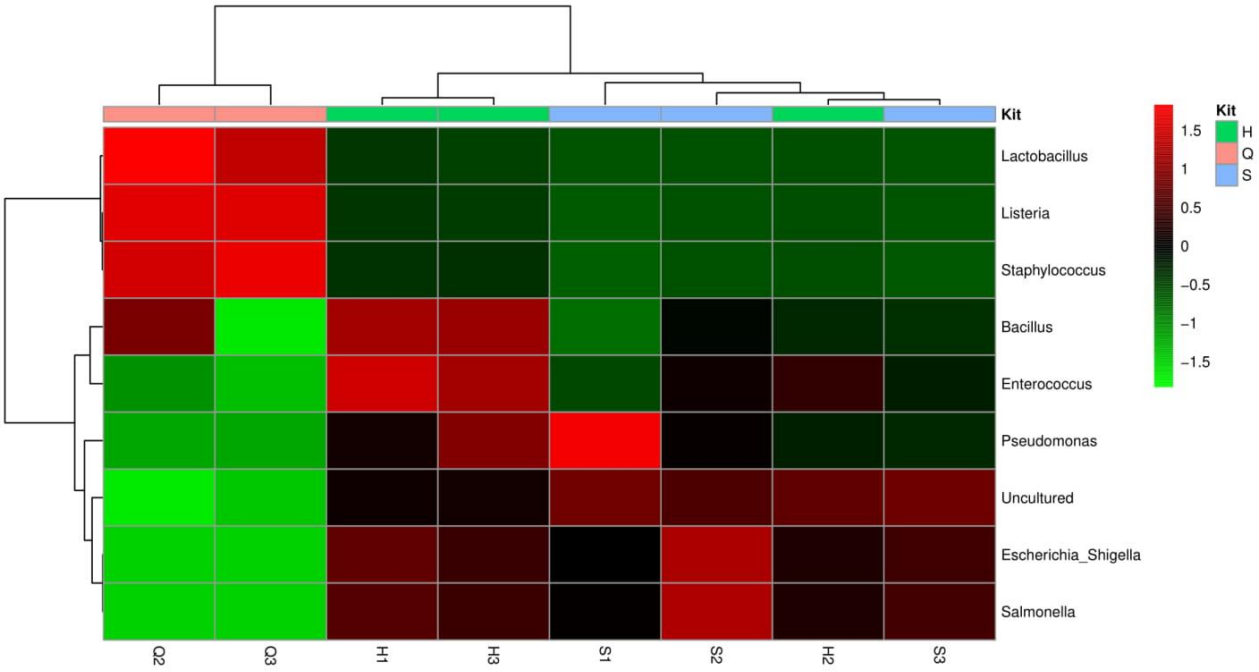


Figure 4: The hierarchical clustering heat map based on the relative abundance of the most abundant genera classified to the genus level for ZymoBIOMICS™ Microbial Community Standard generated with the presented workflow. Individual cells in the heat map are colour-coded according to the row Z-scores

2. Experimental Design, Materials and Methods

2.1 DNA extraction methods evaluated with a commercial standard

DNA from the same ZymoBIOMICS™ Microbial Community Standard (Zymo, USA) was prepared by 3 different DNA extraction methods, with the standard being divided into 9 aliquots (75 µl at 26.7 ng/µl) prior to DNA extractions in triplicate. Commercial kits selected included both chemical lyses and mechanical shearing to ensure that DNA extraction from difficult to lyse bacteria occurred. QIAamp DNA microbiome kit (Qiagen, Germany) and GenElute™ Stool DNA Isolation Kit (Sigma, USA) were used as the commercial kits that acted as “industry controls” to which we’ve compared a NWU in-house cell lysis method as previously described (Mutingwende *et* al., 2015). Briefly 250 µl of sample was mixed with 250 µl of a proprietary lysis buffer. A lyses micro tube (LMT) was placed on the pre-set (95 °C and 3600 rpm) lyser device for 7 min. Bacterial cell lysis was then concurrently achieved through chemical, thermal and mechanical means (Mutingwende *et al*., 2015). Samples were labelled as “S” for the GenElute™ Stool DNA Isolation Kit, “Q” for the QIAamp DNA microbiome kit and “H” for the NWU in-house cell lysis method. DNA concentration was assessed using the Qubit 4

Fluorometer (Thermofisher Scientific, USA) along with the Qubit BR assay kit (Thermofisher Scientific, USA), while quality was determined by nanodrop spectrophotometry on a Nanodrop One (Thermofisher Scientific, USA), summarised in Table 2. All isolated DNA samples were stored at -20 °C until further processing.

2.2 Amplicon library and flow cell preparation

Sequencing of the ZymoBIOMICS™ Microbial Community Standard was carried out at the Agricultural Research Council's (ARC) Biotechnology Platform using the ONT 16S Barcoding Kit (SQK-RAB204) according to the ONT protocol. Polymerase chain reaction (PCR) barcoding amplification was conducted in 50 µl reactions consisting of 1 µl of 16S barcode primer, 25 µl of LongAmp Taq 2X master mix (New England Biolabs, USA), 14 µl nuclease-free water and 1 µl of template DNA (10 ng). A total of 9 samples, 3 extracted by each kit were prepared for sequencing. PCR cycling conditions were set at 95 °C for 1 min followed by 25 cycles of denaturation at 95 °C for 20 s, annealing at 55 °C for 30 s, extension at 65 °C for 2 min and a final extension step of 65 °C for 5 min before holding at 4 °C. PCR products were cleaned using AMPure XP beads (Beckman Coulter, USA) and eluted in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. Following PCR, 1 µl of eluted sample was quantified using a Qubit fluorometer to pool the DNA barcoded libraries into an equal ratio. All barcoded libraries were pooled in the desired ratios to a total of 50-100 fmoles in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. Platform quality control (QC) was carried out using MinKNOW™ on a new R9.4.1 chemistry MinION™ flow cell before the flowcell was primed. In total 75 µl of sequencing mix, consisting of the DNA library, sequencing buffer and library loading beads, was prepared according to the manufacturer's instructions and added in a drop-wise fashion via the SpotOn sample port. The standard 48 h sequencing script was chosen with 1D live base calling.

2.3 Bioinformatics analysis

The raw data generated from the ZymoBIOMICS™ Microbial Community Standard was acquired as fast5 files using ONT's MinION™ sequencing software (version 19.06.8). The fast5 files were base called and de-multiplexed prior to adapter and primer sequence removal with ONT's Guppy™ sequencing software (version 3.2.4). These barcode-sorted fastq files were merged prior to further processing and renamed according to sample. Reads were filtered to remove reads with a Phred score below 7 and keep lengths between 1200 and 1500bp

NanoFilt (https://github.com/wdecoster/nanofilt) [5]. The passing reads were processed with the bioinformatics workflow described in Table 1. The first part of the workflow (steps 1 – 3) includes the installation of Usearch, database generation and classification. Usearch takes quality controlled and demultiplexed fastq files containing ONT generated 16S reads from each sample as input and assigns taxonomy to each read using the Usearch sintax command and the selected 16S database (RDP 16S training set v16, RTS) [2]. Detailed instructions can be found in the Database and sintax generation.txt file supplied in the supplementary information file.

Following classification the generated .SINTAX files are pre-processed in step 4 with a set of R scripts. Step 4 involves pre-processing and includes the use of scripts 1 to 3. The pre-processing scripts import and merge the .SINTAX files generated for each sample during step 3, remove any unnecessary information and generate a counts table (i.e. “for_xlx.txt”, supplementary files). Following pre-processing, the script for step 5 takes the output from the final script of step 4 (Table 1) and generates an excel sheet (i.e. “Summary_out.xlsx”, supplementary file). This excel output allows easy and simple generation of relative abundance graphs at several taxonomic levels as demonstrated by Figure 1. Finally, the scripts (step 6) take the output from step 4 (Table 1) and generate an OTU table (i.e.“MA_OTU.txt”, supplementary file). The OTU table, along with a mapping file (step 7, Table 1), can then be imported into the Microbiome analyst (https://www.microbiomeanalyst.ca/) online software suite for further evaluation which included alpha and beta diversity determination and clustering analyses [1, 3]. As an example, by importing of the output file from step 6 into microbiome analyst the following standard analyses were carried out in a demonstration: Data was normalized using total sum scaling (TSS) and the alpha diversity of samples was measured by observed, chao1 and Shannon diversity indexes as demonstrated with Figure 2. Additionally, beta-diversity was determined and visualized using principal coordinate analysis plots (PCoA) based on Bray–Curtis distances, and compared using the nonparametric analysis of similarities (ANOSIM) test (Figure 3). Finally, a heatmap (Figure 4) of the most abundant genera classified to the genus level was generated using complete hierarchical clustering by Euclidian distance [1, 3, 6]. Detailed descriptions and directions for the use of each script and links for all software used and packages required in R, are provided in the supplementary material within the R scripts.

2.4 Statistical analysis

The influences of each extraction method on the DNA quantity (yield) and quality (A260/A280 and A260/A230) was evaluated with ANOVA (one-way analysis of variance) and Tukey *post hoc* test for multiple pairwise comparisons [4]. This statistical testing was carried out using statistica (v13.1) (Statsoft, Inc, USA) and visualised in GraphPad Prism (v8) (GraphPad, Inc., USA).

Ethics Statement

This study was approved by the North-West University Health Research Ethics Committee (NWU-HREC) of the faculty of health sciences, Ethics number: NWU-00127-18-A1.

Author Contributions

Mann, B.C: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Bezuidenhout, J.J:** Conceptualization, Methodology, Software, Writing - Review & Editing, Supervision. **Grobler, A.F:** Conceptualization, Writing - Review & Editing, Resources, Project administration, funding acquisition. **Swanevelder, Z.H:** Conceptualization, Writing - Review & Editing.

Acknowledgments

This manuscript was written with the support of funding provided by the DSI/NWU Preclinical Drug Development Platform and HANKS TB diagnostics (Pty) Ltd.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. A. Dhariwal, J. Chong, S. Habib, I.L. King, L.B. Agellon, J. Xia. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic acids research*, 45 (2017), https://doi.org/10.1093/nar/gkx295

2. E.C. Edgar. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ,* 6 (2018), e4652 https://doi.org/10.7717/peerj.4652

3. J. Chong, P. Liu, G. Zhou, J. Xia. Using Microbiome Analyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols,* 15 (2020), pp. 799-821

4. N. Corcol, T. Österlund, L. Sinclair, A. Eiler, E. Kristiansson, T. Backhaus, K.M. Eriksson. Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing. *FEMS Microbiology Letters*, 364 (2017), https://doi.org/10.1093/femsle/fnx139

5. S. Kai, Y. Matsuo, S. Nakagawa, K. Kryukov, S. Matsukawa, H. Tanaka, T. Iwai, T. Imanishi, K. Hirota. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS open bio*, 9 (2019), pp. 548–557

6. Y. Xia, J. Sun. Hypothesis Testing and Statistical Analysis of Microbiome *Genes & diseases*, 4 (2017), pp. 138–148

2. Additional Background

It is a well-known fact that the human microbiome, consisting of millions of microbes living on and within human beings, is a complex assemblage of organisms that have been demonstrated to play an intricate role during various states of human health and disease (Eloe-Fadrosh & Rasko, 2013; Wang *et al.*, 2017). Advances in next generation sequencing (NGS) technology have shown that the microbiome plays a role in various biological processes, including modulating the immune system (Claesson *et al.*, 2010; Wang *et al.*, 2017). Other than immune regulation the human microbiome has been linked to various diseases such as inflammatory bowel disease, chirrosis, metabolic syndrome, *Clostridium difficile* and potentially even *Mycobacterium tuberculosis* infection (Wang *et al.*, 2017; Wood *et al.*, 2017; Samarkos *et al*., 2018). NGS technology has seen major development in the field of diagnostics in the past few years, but several barriers and limitations slowed progress in the field. These include, but are not limited to, assay validation, data management, interpretation of results, choice of sequencing technologies, IT requirements and most importantly, very high initial infrastructure costs. However, new and emerging technology like the MinION developed by Oxford Nanopore Technologies (ONT), have slowly begun to overcome many of these limitations (Singh *et al.*, 2016; Quainoo *et al.*, 2017; Besser *et al.*, 2018).

Sequencing of the 16S gene is a method often employed to evaluate changes in the bacterial microbial community in clinical samples, and several studies have stressed that effective bacterial lysis is required in order to properly profile the microbiome (Lim *et al.*, 2018; Pollock *et al.*, 2018). Several commercially available DNA extraction kits are freely available for several applications, and methods often vary in how the lysis of bacterial cells is achieved. The process depends on the kit employed and usually consists of a combination of heat, chemical, enzymatic, or mechanical lysis steps (Lim *et al.*, 2018). Previous studies have highlighted that in order to prevent bias when profiling microbiome samples, a combination of lysis methods is ideally required to ensure that more resistant microbial cells are effectively lysed (Lim *et al.*, 2018; Pollock *et al.*, 2018).

The NWU-TB test is a highly sensitive and specific current TB diagnostic tool making use of an automated Lyser device (NWU lysis system). The lysis portion of the NWU-TB test is a highly efficient lysis process making use of heat, chemical and mechanical means to accomplish cellular lysis. In addition to this, the process is extremely safe with the entire process taking place within a lysis micro tube (LMT) (Mutingwende *et al.*, 2015). The decision was thus made to evaluate DNA obtained by means of the NWU lysis method as preferential tool for downstream microbiome analyses in comparison to two commercially available DNA extraction kits: the QIAamp DNA microbiome kit (Qiagen, Germany) and the GenElute™ Stool DNA Isolation Kit (Sigma, USA), using the ZymoBIOMICS™ Microbial Community Standard (Zymo, USA).

3. Additional experimental design, materials and methods

3.1 Epi2me workflow

In addition to the experimental design, materials and methods described in the Data in Brief manuscript, the raw sequencing data was also analysed with the Oxford nanopore native bioinformatics pipeline with the aid of the Epi2me agent for a comparison to the workflow described in the manuscript (Kai *et al.,* 2019; Winand *et al*., 2020). The raw data generated from the ZymoBIOMICS™ Microbial Community Standard sequencing run on the MinION™ was obtained as fast5 files; using ONT's MinION™ sequencing software (version 19.06.8). The fast5 files were base called using ONT's Guppy™ sequencing software (version 3.2.4). The fastq files obtained were then uploaded to the ONT Epi2me agent (version 2019.7.9), and run through the 16S workflow according to the ONT16S protocol for the ONT 16S Barcoding Kit (SQK-RAB204). The ONT workflow is designed to quality control reads, sort them according to barcode; and then BLAST base called sequences against the NCBI 16S bacterial database containing 18,927 16S sequences from different organisms. Each read is classified based on the % coverage and identity and the output is visualised in the form of a taxonomy tree. Quality control (QC) metrics and read assignments were accessed using the Epi2me agent (Kai *et al.,* 2019; Winand *et al*., 2020). The current ONT 16S workflow making use of the Epi2me agent only provides QC metrics and the IDs' of the main genera identified; it does not currently provide any tools to further explore the data generated.

The bioinformatics pipeline described in the manuscript was thus designed to perform additional analyses of the 16S reads generated.

4. Results

4.1 Yield and quality of extracted nucleic acids

The concentration and quality of the extracted nucleic acids from the ZymoBIOMICS™ Microbial Community Standard and the selected sputum sample, extracted by various means were compared. Extraction by the various methods yielded different nucleic acid concentrations, based on Nanodrop and Qubit quantification as seen in Table 2 of the manuscript. The nucleic acid concentration measured by nanodrop was significantly higher with method H (NWU lysis and extraction) when compared to both methods S and Q ($p < 0.05$); while there was also a significant difference between methods S and Q ($p < 0.05$). Upon examination of the qubit measurements a similar pattern was observed: the average nucleic acid concentration obtained by means of method H was again significantly higher compared to both methods S and Q ($p < 0.05$); but there was no significant difference in nucleic acid concentration between methods S and Q ($p > 0.05$). All methods had an A260/A280 absorbance ratio between 1.7-2.0, suggesting the extraction of pure nucleic acids (Table 2). There was no significant difference concerning the methods as far as the A260/A280 ratio is concerned, but there was a significant difference between method H and Q when examining the A260/A230 ratios. All methods had A260/A230 value below 2. The integrity of the extracted DNA was assessed by agarose gel electrophoresis. Both method S and Q yielded high molecular weight intact DNA without smearing, while method H produced a smear evident of shearing of the DNA.

4.2 NGS analysis of bacterial diversity and composition from various DNA extraction methods

A single sequencing run on a MinION™ flow cell of R9.4 chemistry, produced 6 016 210 reads from 9 DNA barcodes. Platform QC analysis preceding the sequencing run revealed a total of 1356 available pores, split into 4 groups. The mean

read length was 1424 base pairs (bp), with an average Phred score of 9.83. Individual sample metrics can be found in Table 3. Sample Q1 failed to amplify and was not included in any further analyses. Using the Epi2me platform, a total of 5 850 982 reads were classified with an average accuracy of 87%. Taxonomy tree outputs from the Epi2me platform were summarised in Table 4 and 5. Figure 5 includes an example of the Epi2me taxonomy tree output.

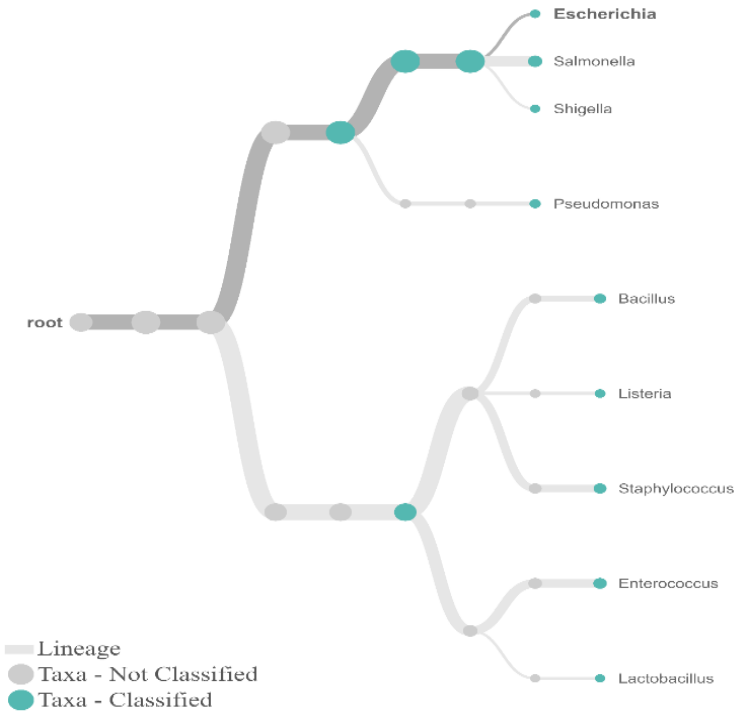


Figure 5: Taxonomy tree generated with the Oxford nanopore native bioinformatics pipeline for 16S microbial composition analyses.

Analysis with the Epi2me workflow at a 1% abundance cut-off (Table 4), yielded varying results between the different extraction methods applied. The Epi2me workflow, in conjunction with method H and S, yielded the expected results at the phylum level with both a 1% and 0.1% identity cut-off; identifying the phyla Firmicutes and Proteobacteria, method Q on the other hand only identified the phylum Firmicutes at the 0.1% abundance cut-off. Upon examination of the genus level identifications, method H managed to identify 8/8 of the organisms present in the sample, the use of method S led to the identification of 7/8 and that of method Q to 5/8. The use of method S did not lead to the identification of the genus *Lactobacillus* at the set 1% abundance cut-off, which is one of the hard to lyse Gram-positive genera present in the sample, while method Q did not identify any of the Gram–negative genera that should be present in the sample.

Table 4: Genera identified at a 1% abundance cut-off (Epi2me workflow)

Genus	Gram	H	S	Q
Salmonella	-	X	X	□
Escherichia/shigella	-	X	X	□
Pseudomonas	-	X	X	□
Bacillus	+	X	X	X
Staphylococcus	+	X	X	X
Listeria	+	X	X	X
Enterococcus	+	X	X	X
Lactobacillus	+	X	□	X

When reducing the abundance cut-off to 0.1%, the use of method S allowed the identification of 8/8 organisms, the same as method H; but method Q again only allowed identification of 5/8. An additional problem encountered at a cut-off of 0.1% is the identification of several genera that are not present in the mock community highlighted in Table 5.

Table 5: Genera identified at a 0.1% abundance cutoff (Epi2me workflow)

Genus	Gram	H	S	Q
Salmonella	-	X	X	□
Escherichia/shigella	-	X	X	□
Citrobacter	-	X	X	□
Enterobacter	-	X	X	□
Klebsiella	-	X	X	□
Vibrio	-	X	X	□
Pseudomonas	-	X	X	□
Bacillus	+	X	X	X
Staphylococcus	+	X	X	X
Listeria	+	X	X	X
Enterococcus	+	X	X	X
Lactobacillus	+	X	X	X

At the time of writing there was no way to set a more stringent identity cut-off, which is problematic due to the error prone nature of ONT reads. The Epi2me workflow does provide a csv file that could be manually formatted into a counts table but there is currently no way to do any downstream statistical analyses to compare samples using the Epi2me 16S workflow; thus the bioinformatics pipeline described in the manuscript

was tested for further downstream analyses. When evaluating the in house workflow, 3 740 890 reads were left for classification after initial demultiplexing and filtering; a total of 3 659 253 reads were classified to the phylum level, and 1 138 882 reads were classified to the genus level.

Table 6: Genera identified at a 0.1% abundance cutoff (In house workflow)

Genus	Gram	H	S	Q
Salmonella	-	X	X	□
Escherichia/shigella	-	X	X	□
Pseudomonas	-	X	X	□
Bacillus	+	X	X	X
Staphylococcus	+	X	X	X
Listeria	+	X	X	X
Enterococcus	+	X	X	X
Lactobacillus	+	X	X	X

The in-house workflow classified fewer reads than the Epi2me workflow, but managed to identify the organisms known to be present in the sample at a 0.1% abundance cut-off without the addition of several reads matched to wrong identities. The decision was thus made to use outputs from the in-house workflow for all further downstream analyses of the sequencing reads generated for the initial evaluations. Rarefaction curves were generated (Figure 6) and the alpha diversity based on the number of Observed species, Chao1 and Shannon diversity indices, was compared according to the DNA extraction method (Figure 2).

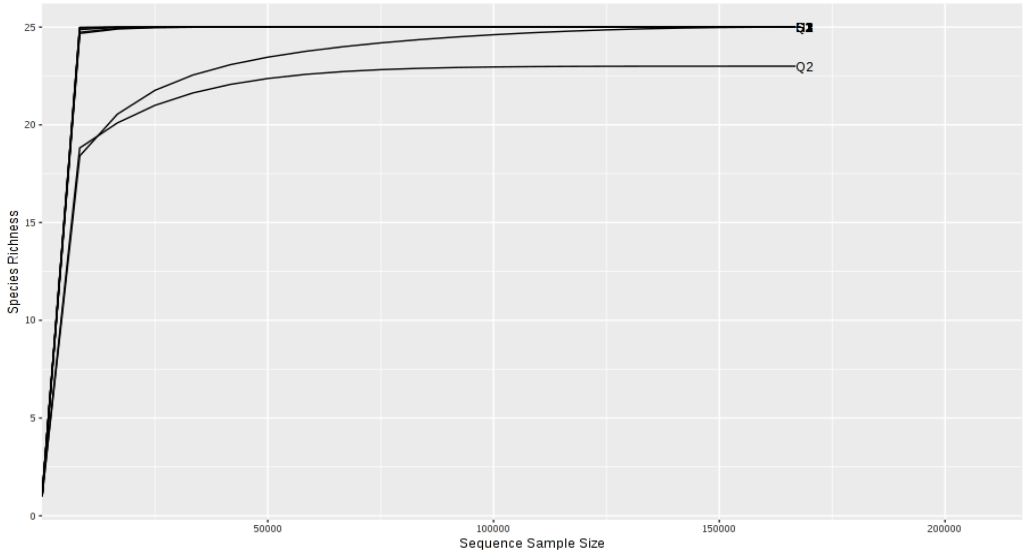


Figure 6: Rarefaction analysis of 16S libraries from each sample of the ZymoBIOMICS™ Microbial Community Standard.

Statistical testing between methods showed no significant difference in alpha diversity observed for the Shannon diversity index, the Chao1 or observed indices ($p > 0.05$). Beta diversity principal coordinates analysis, based on Bray–Curtis distances (Figure 3), show samples subjected to various extraction methods differed significantly (ANOSIM R: 0.78231; p-value = 0.012). Samples extracted with various methods formed 2 distinct clusters. Cluster 1 consisted of the samples extracted with method H and the second cluster consisted of samples extracted with method Q. Samples from method S ordinated closely to those of H but no distinct clustering for this method was observed.

A second clustering analyses was performed by means of a hierarchical clustering heat map, which revealed two clusters (Figure 4). Cluster one, consisting of method S and H, indicated a slightly higher relative abundance of Gram-negative genera for method S; while method H had a higher abundance of Gram-positive genera. With method Q on the other hand, complete absence of the DNA from all Gram-negative organisms during the extraction process was observed. None of the samples extracted with any of the methods matched the theoretical abundance (Z) of the ZymoBIOMICS™ Microbial Community Standard (Figure 1).

5. Discussion

The yield and quality of the DNA extracted by various methods were assessed and compared between the various extraction methods. In this study, 2 commercial mechanical lysis kits were compared with an automated in-house procedure in terms of DNA recovery from a mock microbial community. The DNA extracted by these different methods was assessed based on the quantity and quality of nucleic acids, as well as the microbial community composition. ZymoBIOMICS™ Microbial Community Standard based on Qubit quantification did not differ significantly between methods Q and S, but both differed significantly when compared to method H, from which the highest quantity of DNA was recovered. The ZymoBIOMICS™ Microbial Community Standard comes prepared in a storage solution (DNA/RNA Shield, Zymo, USA), which

according to the manufacturer is inhibitory and as such the lysate produced with the NWU lysis method had to undergo an additional clean up prior to PCR. The increased DNA yield may be due to the use of magnetic beads in comparison to the other 2 methods, which both employ filtration through a silica membrane. Magnetic bead based systems provide more surface area for the binding of DNA when compared to the column based methods, but there is currently no consensus as to which provides a higher yield of nucleic acids (Peng *et al.,* 2013 ; Ali *et al.,* 2017).

The A260/A280 ratio is viewed as the primary measure of DNA purity, and DNA with a ratio between 1.70 and 2.00 is generally accepted as pure (Lucena-Aguilar *et al.,* 2016; Teng *et al.,* 2018). All methods evaluated produced pure DNA when examining the A260/A280 absorbance ratio; although the A260/A230 ratio was low. The expected value for the A260/A230 absorbance ratio is between 2.0 and 2.22. The low A260/A230 may be due to several factors, such as carbohydrate carryover, proteins, lipids, EDTA, phenol or salts such as guanidine HCL, which is a common substance found in various preparations used in commercial DNA extraction kits. Nonetheless, the A260/A230 ratio is viewed as a secondary measure for DNA quality due to the instability of this value when certain elution buffers are used to elute the DNA (Lucena-Aguilar *et al.,* 2016).

PCR amplicons were successfully obtained from all samples, thus results indicate that all methods yielded nucleic acid of sufficient quality and quantity for sequencing of 16S rRNA gene (Pollock *et al.,* 2018). Fragmentation of DNA was assessed by agarose gel electrophoresis. The DNA extracted with methods S and Q yielded high molecular weight DNA, but method H produced a smear evident of shearing. Method H was originally designed to form part of a TB diagnostic system and as such the lysis procedure is incredibly robust. The method involves heat, chemical and mechanical lysis of samples as described under the materials and methods section in the above manuscript, producing damaged and sheared DNA by the end of the process (Mutingwende *et al.,* 2015).

NGS analysis of bacterial diversity and composition was compared between different extraction methods. A single sequencing run on a MinION™, produced several high

quality reads for all samples with the exception of sample Q1. Analyses with the Epi2me workflow revealed several wrongly identified organisms when the abundance cut-off was reduced to a point (0.1 %) where all 8 of the bacteria present in the Mock microbial community could be identified; the decision was thus made to base the remainder of the discussion on the results obtained with the-in house work flow, which correctly identified all organisms present in the sample without any wrongful identification at an abundance cut off of 0.01% (Table 6). There was no significant differences between extraction methods when examining alpha diversity, while beta diversity did show significant clustering based on extraction method.

None of the samples matched the theoretical composition of the mock microbial community (Figure 1) but this may be due to several factors such PCR bias or bias introduced by the sequencing technology itself that could contribute to this. According to Pollock *et al.* (2018), while bias can be introduced at any methodological stage of a microbiome study, the main factor that needs to be kept in mind is consistency from start to end of a study. What was evident is that the pre-processing involved with method Q removed all Gram-negative organisms from the sample prior to DNA capture and elution (Table 4). Method Q was designed to remove host DNA, and the kit has a very specific warning that it may impact samples that are compromised due to freezing, storage buffers etc. The ZymoBIOMICS™ Microbial Community Standard comes frozen and stored in the DNA/RNA shield buffer. According to the manufacturers, the buffer does act as an inactivation buffer; the Gram-negative organisms present may already have been compromised, resulting in the removal of their DNA during the host DNA removal pre-processing with method Q (Stinson *et al*., 2018; Hu *et al*., 2020). This indicates another limitation of the use of the ZymoBIOMICS™ Microbial Community Standard for these evaluations, as the cells are already compromised prior to extraction; additionally, this also points to a major limitation of method Q, which is thus only compatible with freshly collected samples. It has been proven in the past that long-term stool storage at −80 °C had only limited effects on the microbiota, and as such samples are frequently collected and stored for future analyses (Tap *et al.,* 2018; Stinson *et al*., 2018; Hu *et al*., 2020). Current results did not allow drawing a final conclusion on which method can be deemed more appropriate, due to the limited complexity of the sample; thus further evaluation of a more complex samples is

required. Although no clear conclusion could be drawn the current results revealed several important variables to consider:

(i) There are limitations on the use of mock microbial communities for assessments based on the principle of the DNA extractions kits used. There are also limitations regarding these products, in relation to how they are stored. The storage media and methods of storage (e.g. freezing), of these products prior to shipment may have an impact on both cell and DNA integrity which could influence downstream results.
(ii) DNA extraction kits, especially the next generation, incorporating methods for host DNA removal can have limitations, due to the fact that they may induced bias if samples are in any way compromised, and this is not clearly stated when purchasing the kits or in the manuals accompanying the kits.
(iii) The study also revealed that more tools are urgently needed to assess long read 16S sequences generated by the ONT MinION, as the current native workflow offered by ONT, although user friendly to use at the time of this study, is still very limited in the downstream assessments that can be done.

In conclusion, the study revealed several shortfalls and will aid in the experimental design and implementation to assess the NWU lysis method on more complex clinical samples. In addition, the study also served as an opportunity to develop and test a custom workflow for the downstream analyses of ONT generated 16S reads.

References

Ali, N., Rampazzo, R., Costa, A. & Krieger, M. A. 2017. Current nucleic acid extraction methods and their implications to point-of-care diagnostics. *BioMed Research international*, 9306564.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L. & Trees, E. 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical microbiology and infection: *Clinical Microbiology and Infectious Diseases*, 24(4):335–341.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P. & O'Toole, P.W. 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research*, 38(22):e200.

Eloe-Fadrosh, E.A. & Rasko, D.A. 2013. The human microbiome: from symbiosis to pathogenesis. *Annual Review of Medicine*, 64:145–163.

Hu, Y., Cheng, M., Liu, B., Dong, J., Sun, L., Yang, J., Chen, X. & Jin, Q. 2020. Metagenomic analysis of the lung microbiome in pulmonary tuberculosis - a pilot study. *Emerging Microbes and Infections,* 9(1):1444-1452.

Kai, S., Matsuo, Y., Nakagawa, S., Kryukov, K., Matsukawa, S., Tanaka, H., Iwai, T., Imanishi, T. & Hirota, K. 2019. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS open bio*, 9(3):548–557.

Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Ávila, J.A., López-Guerrero, J.A. & Aguilar-Quesada, R. 2016). DNA source selection for downstream applications based on dna quality indicators analysis. *Biopreservation and Biobanking*, 14(4):264–270.

Lim, Y., Totsika, M., Morrison, M. & Punyadeera, C. 2017. The saliva microbiome profiles are minimally affected by collection method or DNA extraction protocols. *Scientific Reports*, 7(1):8523.

Mutingwende, I., Vermeulen, U., Steyn, F., Viljoen, H. & Grobler, A. 2015. Development and evaluation of a rapid multiplex-PCR based system for *Mycobacterium tuberculosis* diagnosis using sputum samples. *Journal of Microbiological Methods*, 116: 37-43.

Peng, X., Yu, K.Q., Deng, G.H., Jiang, Y.X., Wang, Y., Zhang, G.X. & Zhou, H.W. 2013. Comparison of direct boiling method with commercial kits for extracting fecal microbiome DNA by Illumina sequencing of 16S rRNA tags. *Journal of Microbiological Methods*, 95(3):455-62.

Pollock, J., Glendinning, L., Wisedchanwet, T. & Watson, M. 2018. The madness of microbiome: Attempting to find consensus "best practice" for 16s microbiome studies. *Applied and Environmental Microbiology*, 84(7).

Quainoo, S., Coolen, J., van Hijum, S., Huynen, M.A., Melchers, W., van Schaik, W. & Wertheim, H. 2017. Whole-genome sequencing of bacterial pathogens: The future of nosocomial outbreak analysis. *Clinical Microbiology Reviews*, 31(1):e00082-17.

Samarkos, M., Mastrogianni, E. & Kampouropoulou, O. 2018. The role of gut microbiota in *Clostridium difficile* infection. *European Journal of Internal Medicine*, 50:28-32.

Singh, J., Mishra, A., Pandian, A.J., Mallipatna, A.C., Khetan, V., Sripriya, S., Kapoor, S., Agarwal, S., Sankaran, S., Katragadda, S., Veeramachaneni, V., Hariharan, R., Subramanian, K. & Mannan, A.U. 2016. Next-generation sequencing-based method shows increased mutation detection sensitivity in an Indian retinoblastoma cohort. *Molecular Vision*, 22:1036–1047.

Stinson, L.F., Keelan, J.A. & Payne, M.S. 2018. Comparison of meconium DNA extraction methods for use in microbiome studies. *Frontiers in Microbiology*, 9(270).

Tap, J., Cools-Portier, S., Pavan, S., Druesne, A., Öhman, L., Törnblom, H., Simren, M., & Derrien, M. 2019. Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. *Scientific Reports,* 9(601).

Teng, F., Nair, S.S.D., Zhu, P., Li, S., Huang, S., Li, X., Xu, J. & Yang, J. 2018. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Scientific Reports,* 8:16321.

Wang, B., Yao, M., Lv, Longxian, Ling, Z. & Li, L. 2017. The Human Microbiota in Health and Disease. *Engineering,* 3(1):71-82.

Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., Fu, Q., Roosens, N.H., De Keersmaecker, S.C. & Vanneste, K. Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. *International Journal of Molecular Sciences*, 21(1):298.

Wood, M.R., Yu, E.A. & Mehta, S. 2017. The human microbiome in the Fight against Tuberculosis. *The American Journal of Tropical Medicine and Hygiene*, 96(6):1274-1284.

Chapter 4 – Evaluations of the NWU lysis method for downstream microbiome analyses on more complex samples

This chapter consists of a method article submitted and under the review in *Molecular Biology Reports.* The author guidelines followed in submitting to the Journal and the supplementary material can be found in Annexure B. The manuscript is presented in this thesis as it appears online as a preprint prepared by the journal. The publicly available preprint is available with the figures in a separate folder instead of the main article body. Figures thus referred to in text and supplementary information can be found directly following the manuscript. The chapter addresses the continuation of the first objective of the study as outlined in Chapter 1. Personal contributions to the below manuscript included the following: Initial conceptualization, all wet laboratory work associated with the sputum samples mentioned in the below manuscript, data generation, data analyses, assisting and supervising the student working on the gut samples mentioned in the below manusrcipt as part of their M.Sc., reviewing of data generated by the student prior to addition to this manuscript and also personally writing the final manuscript.

1. Manuscript 2

Comparative Investigation of the Suitability of DNA Preparation Methods for Microbiome Analysis of Complex Clinical Samples

Brendon Coenraad Mann
North-West University Potchefstroom Campus: North-West University
Cara Wepener
North-West University Potchefstroom Campus: North-West University
Johannes Jacobus Bezuidenhout (10926542@nwu.ac.za)
North-West University https://orcid.org/0000-0003-3798-2848
Urban Vermeulen
Austell Pharmaceuticals
Anne Grobler
Austell Pharmaceuticals

Research Article

Keywords: DNA preparation, Mechanical lysis, Bead beating, 16S rRNA gene sequencing, Microbiome

DOI: https://doi.org/10.21203/rs.3.rs-207952/v1

Abstract

Efficient DNA preparation is essential for accurate and reproducible microbial data acquisition. In this study an in-house method (NWU lysis system), involving a lysis micro tube (LMT) was employed to determine its suitability for safe, rapid and accurate characterisation of the bacterial microbiome associated with clinical samples in comparison to two available commercial DNA extraction kits. During the experimental setup, it was confirmed that the LMT is suitable for downstream bacterial microbiome analyses with the incorporation of short read sequencing; but due to the fragmented nature of the DNA its suitability is highly dependent on the sequencing technology applied downstream. The study also served to demonstrate that shearing of DNA can have a significant impact on downstream microbiome analyses, based on the sequencing technology used; and that care must be taken especially if long read sequencing is to be used employed.

Introduction

Since the onset of next generations sequencing (NGS) the field of clinical genetics has been rapidly changing (Vrijenhoek et al., 2015). The swift progression of NGS technology and its use in clinical laboratories has allowed for remarkable progress in the genetic diagnosis of both inherited disorders and infectious diseases, as well as the provision of methods for studying accompanying risk factors such as the microbiome (Di Reste *et al.*, 2018). Novel interactions between the human microbiome and health and disease are continuously emerging, and important host–microbiome interactions are increasingly used as targets for both diagnostics and personalised therapeutics (Behrouzi et al., 2019). The majority of NGS platforms still carry high initial infrastructure costs, have extensive computing requirements and are not of a compact and portable nature. These limitations have resulted in the introduction of platforms such as the MinION developed by Oxford nanopore technologies (ONT), as well as rapidly reducing the cost associated with current technologies such as the Illumina Miseq (Besser et al., 2018). Taking the rapid development, improvement and reduction in costs of various NGS platforms into account, many clinical laboratories and research entities are preparing and/or are already in the progress of adapting their diagnostic process to incorporate or focus mainly on NGS approaches Vrijenhoek et al., 2015.

Sequencing of the 16S gene is a commonly employed method to evaluate the microbial community in clinical samples (Pollock et al., 2018). Previous studies have emphasized that effective DNA preparation is crucial for the accurate characterisation of the complex bacterial populations often present in clinical samples, and that lysis efficiency is a major limitation, especially with samples containing hard to lyse bacteria (Lim et al., 2018). Several methods are currently available for effective DNA preparation prior to 16S sequencing, in the form of various commercial kits. These approaches to cellular lysis of available commercial DNA preparation methods differ, and the lysis method may depend on heat, chemical, enzymatic, or mechanical lysis. Additionally, the specific DNA isolation techniques used also differ, such as spin column-based methods or the use of magnetic beads; while others have opted for direct DNA amplification of crude lysates, forgoing conventional DNA purification (Videvall et al., 2017; Pollock et al., 2018; Fiedorová et al., 2019)

The NWU-TB test is a molecular-based sensitive and specific TB diagnostic tool. The NWU-TB system consists of three sequential steps: cell lysis in a LMT (Supplementary Fig. 2), using an automated lyser device (NWU lysis system: Supplementary Fig. 3) completely inactivating TB within 7 minutes, followed by multiplex-PCR within 25 min. The lysis step of the NWU-TB test is a highly efficient and cost effective process that is based on the use of heat, chemical and mechanical means to achieve cellular lysis. In addition to its lysis efficacy, the lysis process is also extremely safe, with lysis occurring within the confines of a lysis micro tube, offering the potential for this system to be used with decreased biosafety measures (Mutingwende et al., 2015).

Authors of recent studies have emphasized that the inclusion of mechanical lysis in conjunction with other lysis methods is essential to minimize possible biases due to some microbial cells being more resistant to lysis than others (Lim et al., 2018; Pollock et al., 2018)

The decision was thus made to evaluate DNA obtained by means of the NWU lysis method as preferential tool for downstream microbiome analyses on two sample types: a clinical sputum sample evaluated with long read sequencing technology (ONT MinION), representing a lower complexity sample and a more complex gut sample evaluated with short read sequencing technology (Illumina Miseq), focussing on 16S amplicon-based sequencing in comparison to DNA obtained with the use of various commercially available kits.

Materials And Methods

Samples and DNA preparation

DNA obtained by means of the NWU lysis method (method H) was compared to DNA obtained by means of two available commercial kits on both the *caecum* contents from an available mouse study (representing a gut sample), and a previously collected less complex clinical sputum sample; thus representing two samples types of varying complexity. The two commercial kits selected for comparison included the QIAamp DNA microbiome kit (Qiagen, Germany) and the GenElute™ Stool DNA Isolation Kit (Sigma, USA), and were labelled as method Q and S respectively. These kits were selected on the basis that both were bead-beating kits, and thus included both chemical and mechanical lysis to ensure that difficult to lyse bacteria are properly lysed during the DNA preparation process.

Sputum sample

For the evaluation of a low complexity clinical sample, a sputum sample of sufficient volume was collected from a concurrently running study. The sample was initially collected, snap frozen and stored at -80°C until processing. The selected sample was divided into 12 x 250 µl aliquots before further processing. Aliquots were extracted in triplicate using either method Q or S; while lysate was prepared with the NWU in-house cell lysis method (H) with and without additional purification (carried out with the spin column technology of method Q and labelled as HQ).

Extractions with method Q and S were done according to the manufacturers' instructions. Freeze–thaw cycles may compromise bacterial integrity, and the benzonase treatment used during the host DNA removal Protocol of method Q may lead to a loss of exposed bacterial DNA. The decision was thus made to omit the host DNA removal step from method Q considering the samples have been previously frozen. Preparation of the crude lysate using the NWU in-house cell lysis method was done according to previously described methods (Mutingwende et al., 2015). For the NWU cell lysis method, lysis was carried out using the NWU lyser device: 250 µl of sample was mixed with 250 µl of a proprietary lysis buffer. The LMT was placed on the pre-set (95°C and 3600 rpm) lyser device for 7 min. Bacterial cell lysis was concurrently achieved through chemical, thermal and mechanical means (Mutingwende et al., 2015).

DNA concentration was assessed using the Qubit 4 Fluorometer (Thermofisher Scientific, USA) along with the Qubit BR assay kit (Thermofisher Scientific, USA), while quality was determined by nanodrop spectrophotometry on a Nanodrop One (Thermofisher Scientific, USA). The integrity of extracted genomic DNA was evaluated by visualization of a 1.5% (w/v) agarose gel using GelRed® dye (Biotium, USA), after electrophoresis in the presence of a 1kb ladder as size reference standard.

Representative gut sample

The *caecum* content from a concurrently running mouse study was investigated as a complex sample. As part of the initial study the *caecums* were dissected, snap frozen and stored at -80°C prior to further analysis. To generate sufficient sample volume, three *caecums* from C3HeB/JeF mice were physically cut open and the content was added to phosphate buffered saline (PBS). The mixture was vortexed at high speed for 2 minutes to generate a homogenized sample; 250 µl of this mix was then used for DNA extraction and lysate preparation in triplicate. DNA was extracted and/or prepared, subjected to quality control and samples were labelled as described under Sect. 2.1.1

Amplicon library and flow cell preparation

Two sequencing runs were carried out; one for the sputum sample and one for the gut sample. For the sputum sample a 16S rRNA sequencing library was constructed according to the 16S Barcoding Kit (SQK-RAB204) Protocol (Oxford Nanopore Technologies, Oxford, UK) for sequencing on the ONT MinION platform. Library construction for the gut sample was performed according to the 16S metagenomics sequencing library preparation protocol (Illumina, San Diego, CA, USA) for sequencing on the Illumina MiSeq platform.

Sputum sample

Sequencing of the sputum sample was carried out at the North-West University using the ONT 16S Barcoding Kit (SQK-RAB204) according to the ONT Protocol with the only difference being the use of an

inhibitor tolerant high-fidelity polymerase. Polymerase chain reaction barcoding amplification was conducted on a C1000™ Thermal Cycler (Bio-Rad, US). A 50 µl reaction volume consisting of: 1 µl (10 µM) of 16S barcode primer (Oxford Nanopore Technologies, Oxford, UK), 25 µl of Invitrogen Platinum SuperFi DNA Polymerase master mix (Thermo Fisher Scientific, USA), 10 µl of GC enhancer (Thermo Fisher Scientific, USA), 13 µl nuclease-free water (Sigma-Aldrich, USA) and 1 µl of template DNA (10 ng/µl) was prepared for each sample. In the case of the lysate produced by the NWU in-house cell lysis method, 1 µl of sample was added to the reaction mixture. PCR cycling conditions were set at 95°C for 1 min followed by 25 cycles of denaturation at 95°C for 20 s, annealing at 55°C for 30 s, extension at 65°C for 2 min and a final extension step of 65°C for 5 min before holding at 4°C. PCR products were cleaned using AMPure XP beads (Beckman Coulter, USA) and eluted in 10 µl of a buffer containing 10 mM Tris-HCl pH 8.0 and 50 mM NaCl. Following PCR, 1 µl of eluted sample was quantified using a Qubit fluorometer in order to pool the DNA barcoded libraries at an equal ratio. All barcoded libraries were pooled in the desired ratios to a total of 50–100 fmoles in 10 µl of 10 mM Tris-HCl pH 8.0 and/ 50 mM NaCl. Platform quality control (QC) was carried out using MinKNOW™ on 2 new R9.4.1 chemistry MinION™ flow cells before the flow cell was primed. In total 75 µl of sequencing mix consisting of the DNA library, sequencing buffer and library loading beads was prepared according to the ONT Protocol and added in a drop-wise fashion via the SpotOn sample port. The standard 48 h sequencing script was chosen with 1D live base calling.

Representative gut sample

The V3 and V4 regions of the 16S rRNA gene from the isolated microbiome were amplified on the C1000™ Thermal Cycler (Bio-Rad, US). A 25 µl reaction volume consisting of: 1 µl (5 µM) of forward primer, 1 µl (5 µM) of reverse primer, 12.5 µl of Invitrogen Platinum SuperFi DNA Polymerase master mix (Thermo Fisher Scientific, USA), 5 µl of GC enhancer (Thermo Fisher Scientific, USA), 4.5 µl nuclease-free water (Sigma-Aldrich, USA) and 1 µl of template DNA (20 ng/µl) was prepared for each sample. In the case of the lysate produced by the NWU in-house cell lysis method, 1 µl of sample was added to the reaction mixture. The following PCR conditions were used during amplification: initial denaturation at 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 55°C for 10 s and 72°C for 30 s and a final elongation step at 72°C for 300 s. After amplification, PCR products were purified using the Agencourt AMPure XP PCR Purification kit (Beckman Coulter, USA); and transported to the Agricultural Research Council Biotechnology Platform (ARC-LNR), Pretoria, South Africa, for sequencing on the Illumina MiSeq platform according to the standard protocol.

Bioinformatics analysis

Fast 5 files generated from the sputum sample sequencing runs on the MinION™ were base called and de-multiplexed, followed by removal of adapter and primer sequences using ONT's Guppy™ sequencing software (version 3.2.4). The resultant fastq files were then filtered to remove reads with a Phred score below 7; and lengths below 1200 and above 1500bp, using NanoFilt

(https://github.com/wdecoster/nanofilt) (De Coster et al., 2018). Taxonomy was assigned to sequences using the sintax command from Usearch (Edgar, 2018), with a sintax cut-off of 0.8. Sequences were classified using the RDP 16S training set v16 (RTS) database comprising 13,212 sequences belonging to 2,126 genera. Following classification, the output sintax files were processed with in-house R scripts to produce OTU tables and excel summary files. The OTU table along with a mapping file were then fed into the MicrobiomeAnalyst online software suite for further evaluation, which included alpha and beta diversity determination, rarefaction curve generation and clustering analyses (Dhariwa *et al.*, 2017; Chong et al., 2020).

Fastq files produced by the Illumina MiSeq platform were processed and analysed using the mothur (version 1.42.3) shell program (Schloss et al., 2009). The sequences were aligned with SILVA-based bacterial reference alignment (version July 2019). Chimeric reads were filtered out using the abundant sequences (dereplicate = t) as a reference with the VSEARCH command in the shell program. Host DNA sequences along with other undesirable sequences were removed with the remove.lineage command where the DNA of chloroplasts, mitochondria, archaea and eukaryotes were removed. Sequences that shared a minimum of 97% pairwise nucleotide identity were clustered into operational taxonomic units (OTUs). OTUs were classified up to species level using the 16S rRNA RDP reference (version 16) and the classify.otu command in the mothur shell program. The abundance file obtained using the make.shared command and the taxonomy file obtained using the classify.otu command was exported for further analyses (Schloss et al., 2009).

Statistical analysis

To evaluate the influence of each extraction method on the DNA quantity/yield and quality (A260/A280 and A260/A230); ANOVA (one-way analysis of variance) was employed with Tukey's post-hoc test for multiple pairwise comparisons [15]. Statistical testing was carried out using Statistica (v13.1) (Statsoft, Inc, USA) and visualised in GraphPad Prism (v8) (GraphPad, Inc., USA). Statistical analyses and visualisation of microbiome data were done with the assistance of MicrobiomeAnalyst on default settings unless specified otherwise (Dhariwa *et al.*, 2017; Chong et al., 2020). Data was normalised using total sum scaling (TSS). The alpha diversity of samples was measured by observed, Chao1 and Shannon diversity indexes (Corcoll et al., 2017). The observed and Chao1 indices act as measures of species richness. The observed species index measures the number of distinguishable taxa in every sample; whereas the Chao1 index is a qualitatively measure, which beside species richness also takes into account the ratio of singletons, and hence gives more weight to rare species. The Shannon diversity index on the other hand, is a measure of both richness and evenness of the microbes of the given sample; thus addressing the question of whether there is evenness and possible domination in the main genera/species found in the sample (Xia & Sun, 2017).

Beta-diversity calculations, which describes the diversity in a microbial community between different samples, were visualized using principal coordinate analysis plots (PCoA). Beta-diversity PCoA plots were based on Bray–Curtis distances, and compared using the nonparametric analysis of similarities

(ANOSIM) test. Heat maps of the most abundant genera classified to the genus level were generated using complete hierarchical clustering by Euclidian distance (Dhariwa *et al.*, 2017; Xia & Sun, 2017; Chong et al., 2020). These tests were performed at the feature level with only the higher abundance taxa (> 0.1% of total).

Ethics

This study was approved by the North-West University Research Ethics Committee under ethics numbers: NWU-00127-18-A1 and NWU-00584-19-A5.

Results

Yield and quality of prepared nucleic acids

The concentration and quality of the prepared nucleic acids were compared for the various methods used for both sample types as summarised in Table 1. Method H without the addition of a purification step yields a crude lysate and thus no quality assessments could be obtained.

Table 1
DNA Quality

Sample	Nanodrop Concentration (ng/µL)	A260/A280	A260/A230
Sputum sample			
H	-	-	-
HQ	15,03 ± 0,13 [a]	1,95 ± 0,03 [a]	1,96 ± 0,38 [b]
Q	186,53 ± 5,89 [b]	1,87 ± 0 [b]	2,15 ± 016 [b]
S	20,2 ± 0,31 [a]	2,12 ± 0,03 [c]	0,13 ± 0,12 [a]
Gut sample			
H	-	-	-
HQ	10.43 ± 2.74 [a]	1.77 ± 0.01 [b]	1.33 ± 0.27 [a,b]
Q	610.5 ± 111.49 [b]	2.06 ± 0.04 [a]	1.86 ± 0.41 [b]
S	79.87 ± 23.72 [a]	1.96 ± 0.05 [a]	0.67 ± 0.34 [a]

Values sharing a common letter belong to the same group according to one way ANOVA with Tukey's post-hoc test for multiple pairwise comparisons. These values do not differ significantly according to the test ($p > 0.05$). Values not sharing a common lower case letter indicate a statistically significant difference in comparison to a value belonging to other groups ($p < 0.05$).

Sputum sample

Initial nanodrop results (Table 1) indicate the nucleic acid concentration obtained with method Q was significantly higher ($p < 0.05$) when compare to HQ, and S. All methods with the exception of method S (2.1) had a A260/A280 absorbance ratio between 1.7-2.0, suggesting the preparation of pure nucleic acids. Both methods HQ and Q had A260/A230 ratios within the acceptable range, while method S had a significantly lower A260/A230 ratio. The integrity of the prepared DNA was assessed by agarose gel electrophoresis. Method S yielded high molecular weight intact DNA with minimal shearing, while method Q produced large amounts of high molecular weight DNA, but shearing was also evident. Method HQ on the other hand produced sheared low molecular weight DNA.

Representative gut sample

Upon examination of DNA obtained from the gut sample, a similar pattern in the average nucleic acid concentrations obtained can be seen (Table 1). Method Q again had a significantly higher nucleic acid concentration when compared to other methods.

Again all had acceptable A260/A280 absorbance ratios between 1.7-2.0, while only method Q had A260/A230 ratios within the acceptable range. DNA integrity was assessed by agarose gel electrophoresis and revealed similar results to those obtained with the selected sputum sample.

NGS analyses of bacterial diversity and composition of samples from various DNA preparation methods

NGS analyses of bacterial diversity and composition from various DNA preparation methods were carried out on sequencing reads produced for both sample types.

Sputum sample

A single sequencing run on a MinION™ flow cell of R9.4 chemistry, produced 5 169 844 reads from 12 DNA barcodes. Platform quality control (QC) analysis preceding the sequencing run revealed a total of 1356 available pores, split into 4 groups. The mean read length was 1352 base pairs (bp), with an average Phred score of 10.01. The alpha diversity based on the number of observed species, Chao1 and Shannon diversity indices, was compared according to the DNA preparation method (Fig. 1: A).
As seen in Fig. 1A, there was no significant difference between methods HQ, Q and H, based on the Shannon, Chao1 and observed indices, while method S ranked significantly higher than any of the other methods for all three alpha diversity measures. Clustering analyses were performed by means of a hierarchical clustering heat map (Fig. 2), which revealed three clusters: Cluster one, consisting of method H and HQ, and one cluster each for methods Q and S respectively. In samples extracted with method S, Gram-negative bacterial genera were more abundant, while samples extracted with method Q resulted in an increased abundance for Gram-positive bacterial genera.

Principal coordinate analysis of beta-diversity yielded similar results, showing significant clustering according to the method. Methods H and HQ clustered together, while methods Q and S formed unique clusters (Supplementary Fig. 1(A)). Analyses of similarity confirmed significant variability among the preparation methods ([ANOSIM] R: 0.74691; p-value < 0.001).

Gut sample

The resulting run on the Illumina Miseq platform produced 4 582 278 reads for 12 samples, with an average quality of 32.58 and an average length of 453 bp. Again the alpha diversity based on the number of observed species, Chao1 and Shannon diversity indices, was compared according to the DNA preparation method (Fig. 1: B). For the representative gut sample, there was no significant difference between methods H, HQ, Q and S based on the Shannon, Chao1 and observed indices. Clustering analyses performed by means of a hierarchical clustering heat map, revealed random clustering among samples, also suggesting no statistically significant difference based on the various preparation methods (Fig. 3).
This was again confirmed by principal coordinate analysis of beta-diversity (Supplementary Fig. 1(B)), which revealed no statistically significant difference between samples based on preparation method. Analyses of similarity confirmed no significant variability among the preparation methods ([ANOSIM] R: -0.11728; p-value < 0.759).

Discussion

The A260/A280 ratio is viewed as the primary measure of DNA purity, and DNA with a ratio between 1.70 and 2.00 is generally accepted as pure (Lucena-Aguilar et al., 2016; Teng et al., 2018). All methods evaluated produced relatively pure DNA when examining the A260/A280 absorbance ratio; although the A260/A230 ratio was generally lower for all methods evaluated, especially for method S. The expected value for the A260/A230 absorbance ratio is between 2.0 and 2.22. The low A260/A230 may be due to such factors as carbohydrate carryover, proteins, lipids, EDTA, phenol or salts such as guanidine HCL, a common substance found in various commercial DNA extraction kits (Lucena-Aguilar et al., 2016). None the less, the A260/A230 ratio is viewed as a secondary measure of DNA quality, due to the instability of this value when certain elution buffers are used to elute the DNA (Lucena-Aguilar et al., 2016). PCR amplicons were successfully obtained from all samples; thus the results indicate that all methods yielded nucleic acid of sufficient quality and quantity for sequencing of the 16S rRNA gene (Pollock et al., 2018). Fragmentation of DNA was assessed by agarose gel electrophoresis. The DNA extracted with methods S and Q yielded high molecular weight DNA. Method H was originally designed to form part of a TB diagnostic system and as such the lysis procedure is incredibly robust primarily producing sheared single stranded DNA. The method was originally incorporated with real time qPCR, involving the amplification of short fragments of DNA; and the potential impact of shearing on larger PCR products has not been previously investigated (Kennedy *et al.*, 2009).

NGS analyses of bacterial diversity and composition were compared between different DNA preparation methods and with two sequencing approaches. Evaluations of a representative gut sample with the incorporation of short read sequencing on the Illumina Miseq platform indicated no statistically significant difference between either DNA preparation methods at the end of analyses of the microbiome. Analysis of the sputum sample with the incorporation of long read sequencing, on the other hand showed a statistically significant difference when compared to one of the commercial methods (S). When comparing methods H and HQ it is evident that purification of the DNA obtained with method H did not have any significant impact on the end results and that the differences observed were the result of the level of integrity of the DNA and the sequencing method applied. The DNA obtained with methods H, HQ and Q was consistently sheared, which may not be a problem when sequencing short fragments; but for single molecule or long-read sequencing technologies, particularly those from Pacific Biosciences (PacBio) and ONT, this can be a major issue (Lim et al., 2018). The MinION 16S workflow incorporates full length 16S amplicons of around 1500bp, and it would thus follow that a workflow incorporating larger amplicons will fail to capture the true diversity of a sample if there is a low amount of intact DNA, as illustrated by the current results.

A limitation of the present study was the use of only one representative sample for each evaluation. It was therefore not possible to determine whether the variation between the preparation methods was less than the inter-subject variation. According to previous studies the inter-subject variation generally tends to be greater than the technical variation. Hence, it is expected that analysis of additional samples would be in accordance with those of prior accounts (Lim et al., 2018; Kennedy et al., 2014). This study focussed on an evaluation of a new method, producing a crude lysate and its efficiency in comparison to already existing validated commercial DNA extraction methods (Wagner Mackenzie et al., 2015; Lim et al., 2018). Various studies tend to use mock microbial samples of known composition to evaluate new methods, but these mock samples also have their limitations: They are often presented in a liquid or broth and as such lack the complexity of real biological samples such as sputum, in which the viscosity of the sample is a major stumbling block. Furthermore, mock microbial communities tend to have a lower degree of diversity when compared to actual clinical samples, which harbour a much more diverse bacterial population. Given the growing significance of the microbiome, new products and tools are consistently being developed to study the human microbiome and its various relations to health and disease (Mohajeri et al., 2018). Despite the mentioned limitations, the current study provides useful information, describing the potential of a LMT to carry out downstream microbiome analyses with improved speed and safety.

Conclusion

To conclude, the NWU lysis system is a promising prospect for clinical microbiome studies due to its lysis efficiency. In addition to lysis efficiency the system also provides improved safety due to the fact that lysis of clinical specimens take place within a closed container. In its current form though the lysis method is only compatible with short read sequencing because of the sheared nature of the DNA produced and further optimisation will be required before the lysis step can be incorporated into a

portable sequencing technology such as the ONT MinION. In addition, this study also highlights the importance of evaluating the integrity of DNA, as shearing may greatly impact the final composition of the microbiome depending on the sequencing technology employed.

Declarations

Funding

This study was performed, and the manuscript compiled with the support of funding provided by the South African Department of Science and Innovation in conjunction with the International *Mouse* Phenotyping Consortium (*IMPC*) and HANKS TB Diagnostics (Pty) Ltd. The authors also acknowledge financial support from the National Research Foundation of South Africa (DST-NRF Innovation Master's Scholarship Grant No.117255). Opinions expressed, findings and concluding statements are those of the authors and are not necessarily attributed to the NRF.

Nucleotide sequence data

Genbank All data is publicly available from Genbank and the NCBI SRA under Bio projects PRJNA675726 (SRR13014341-SRR13014338) and PRJNA675484 (SRR1301004 -SRR13010037).

References

1. Behrouzi A, Nafari AH, Siadat SD (2019) The significance of microbiome in personalized medicine. Clin Transl Med 13;8(1):16. doi: 10.1186/s40169-019-0232-y.
2. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect 24(4):335-341. doi: 10.1016/j.cmi.2017.10.013.
3. Lim MY, Song EJ, Kim SH, Lee J, Nam YD (2018) Comparison of DNA extraction methods for human gut microbial community profiling. Syst Appl Microbiol 41(2):151-157. doi: 10.1016/j.syapm.2017.11.008.
4. Chong J, Liu P, Zhou G, Xia J (2020) Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat Protoc 15(3):799-821. doi: 10.1038/s41596-019-0264-1.
5. Corcoll N, Österlund T, Sinclair L, Eiler A, Kristiansson E, Backhaus T, Eriksson KM (2017) Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing. FEMS Microbiol Lett 1;364(14). doi: 10.1093/femsle/fnx139.
6. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34(15):2666–2669.

https://doi.org/10.1093/bioinformatics/bty149.

7. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J (2017) MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res 3(45):180-188. doi: 10.1093/nar/gkx295.
8. Di Resta C, Galbiati S, Carrera P, Ferrari M (2018) Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. EJIFCC 29(1):4-14.
9. Edgar RC (2018) Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ 6. https://doi.org/10.7717/peerj.4652
10. Fiedorová K, Radvanský M, Němcová E, Grombiříková H, Bosák J, Černochová M, Lexa M, Šmajs D, Freiberger T (2019) The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. Front Microbiol 17;10:821. doi: 10.3389/fmicb.2019.00821.
11. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM; UK IBD Genetics Consortium, Satsangi J, Flint HJ, Parkhill J, Lees CW, Hold GL (2014) The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. PLoS One 24;9(2):e88982. doi: 10.1371/journal.pone.0088982.
12. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R (2016) DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreserv Biobank 14(4):264-70. doi: 10.1089/bio.2015.0064.
13. Mohajeri MH, Brummer RJM, Rastall RA, Weersma RK, Harmsen HJM, Faas M, Eggersdorfer M (2018) The role of the microbiome for human health: from basic science to clinical applications. Eur J Nutr 57:1-14. doi: 10.1007/s00394-018-1703-4.
14. Mutingwende I, Vermeulen U, Steyn F, Viljoen H, Grobler A (2015) Development and evaluation of a rapid multiplex-PCR based system for Mycobacterium tuberculosis diagnosis using sputum samples. J Microbiol Methods 116:37-43. doi: 10.1016/j.mimet.2015.06.007.
15. Pollock J, Glendinning L, Wisedchanwet T, Watson M (2018) The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. Appl Environ Microbiol 84(7). https://doi.org/10.1128/AEM.02627-17.
16. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537-41. doi: 10.1128/AEM.01541-09.
17. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang X, Li J, Xu J, Yang F (2018) Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. Sci Rep *8*(1):16321. https://doi.org/10.1038/s41598-018-34294-x.
18. Videvall E, Strandh M, Engelbrecht A, Cloete S, Cornwallis CK (2017) Direct PCR Offers a Fast and Reliable Alternative to Conventional DNA Isolation Methods for Gut Microbiomes. *mSystems*. 2(6).

https://doi.org/10.1128/mSystems.00132-17.

19. Vrijenhoek T, Kraaijeveld K, Elferink M, de Ligt J, Kranendonk E, Santen G, Nijman IJ, Butler D, Claes G, Costessi A, Dorlijn W, van Eyndhoven W, Halley DJ, van den Hout MC, van Hove S, Johansson LF, Jongbloed JD, Kamps R, Kockx CE, de Koning B, Kriek M, Lekanne Dit Deprez R, Lunstroo H, Mannens M, Mook OR, Nelen M, Ploem C, Rijnen M, Saris JJ, Sinke R, Sistermans E, van Slegtenhorst M, Sleutels F, van der Stoep N, van Tienhoven M, Vermaat M, Vogel M, Waisfisz Q, Marjan Weiss J, van den Wijngaard A, van Workum W, Ijntema H, van der Zwaag B, van IJcken WF, den Dunnen J, Veltman JA, Hennekam R, Cuppen E (2015) Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects. Eur J Hum Genet 23(9):1142-50. doi: 10.1038/ejhg.2014.279
20. Wagner Mackenzie B, Waite DW, Taylor MW (2015) Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. Front Microbiol 18(6):130. doi: 10.3389/fmicb.2015.00130.
21. Xia Y, Sun J (2017) Hypothesis Testing and Statistical Analysis of Microbiome. Genes Dis 4(3):138-148. doi: 10.1016/j.gendis.2017.06.001.

2. Manuscript 2 figures and supplementary figures

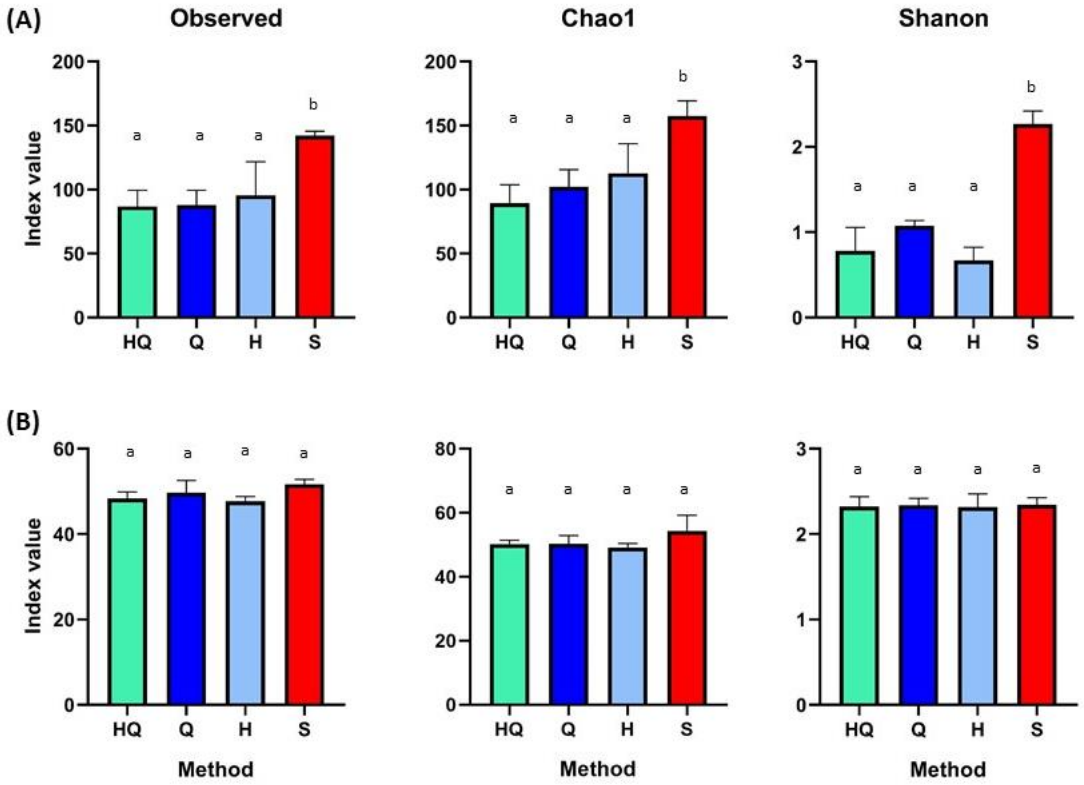


Figure 1: Alpha-diversity, measured by Shannon, Chao1 and observed diversity indices for the selected clinical sputum sample (A) and the representative gut sample (B). Error bars show the standard error of the mean (SEM). Methods sharing a common lower case letter belong to the same group according to one way ANOVA with Tukey's post-hoc test for multiple pairwise comparisons. These values do not differ significantly according to the test ($p > 0.05$). Values not sharing a common lower case letter indicate a statistically significant difference in comparison to a value belonging to other groups ($p < 0.05$).

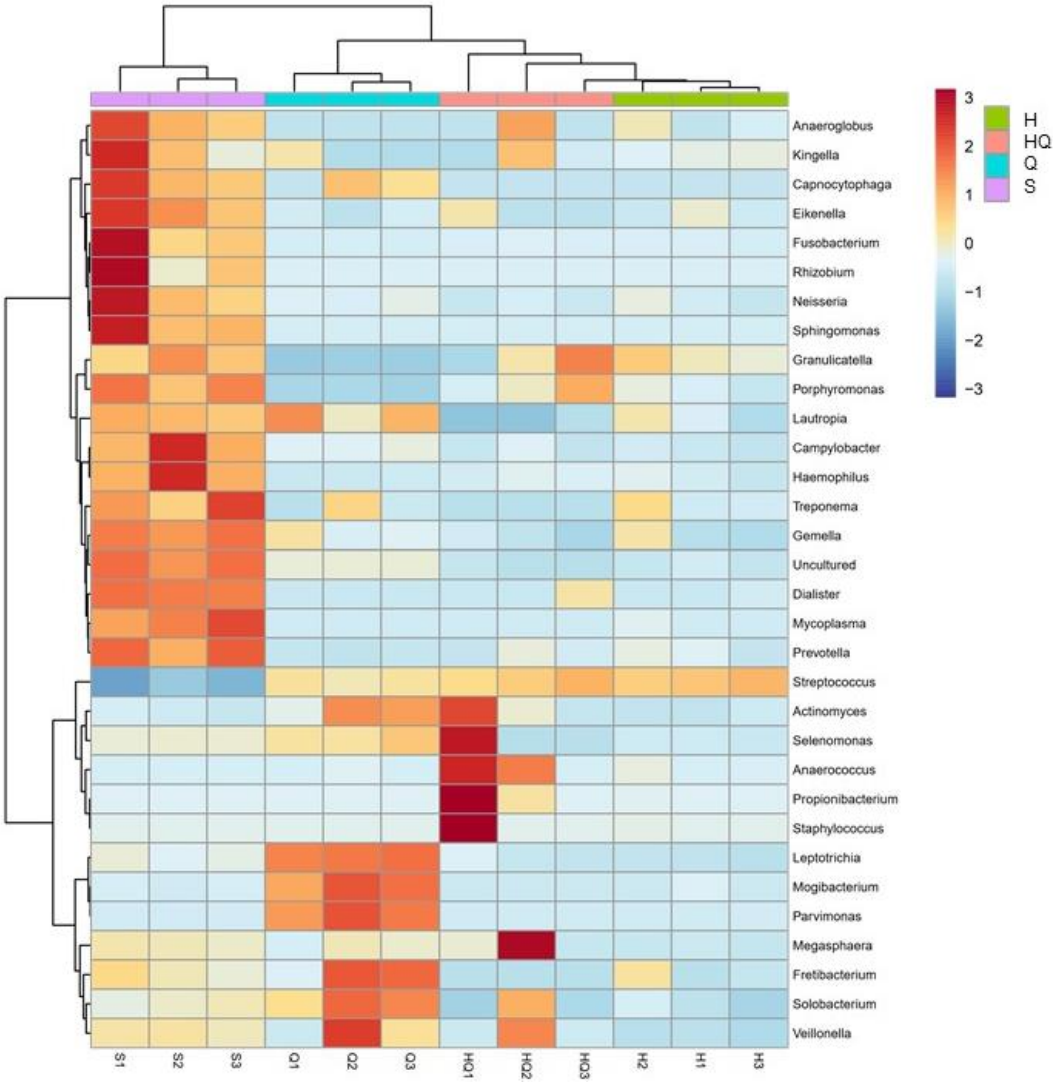


Figure 2: Hierarchical clustering heat map based on the relative abundance of the most abundant genera classified to the genus level for the selected clinical sputum sample. Individual cells in the heat map are colour-coded according to the row Z-scores.

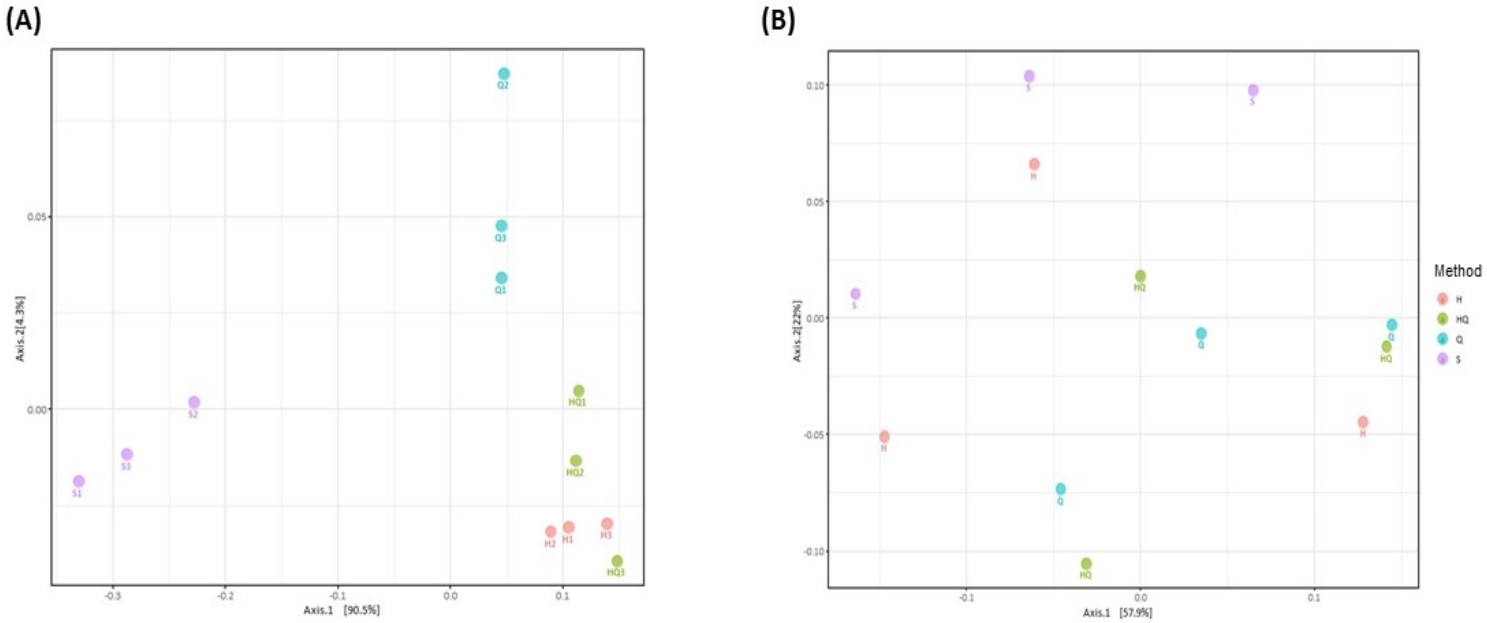
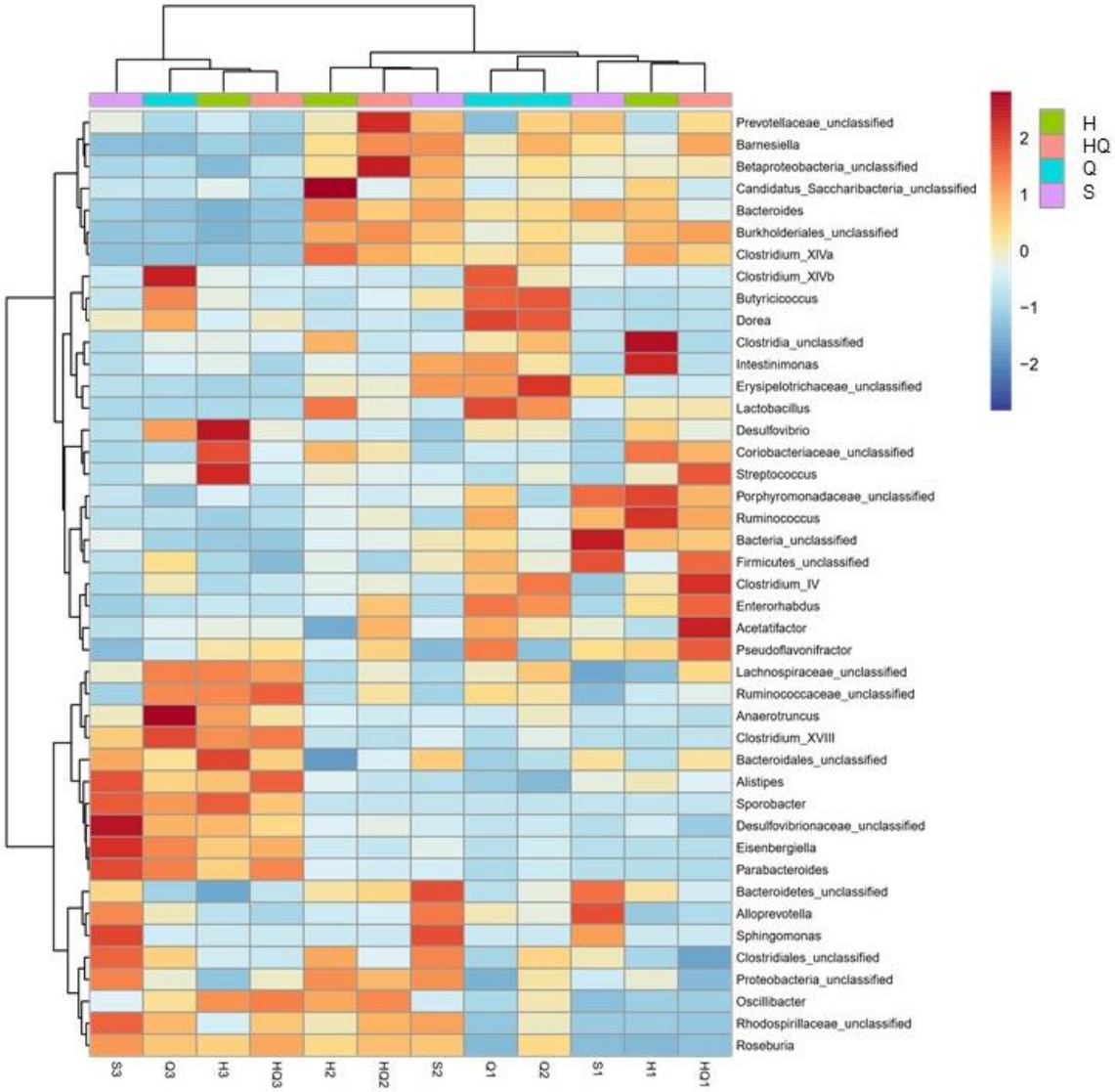
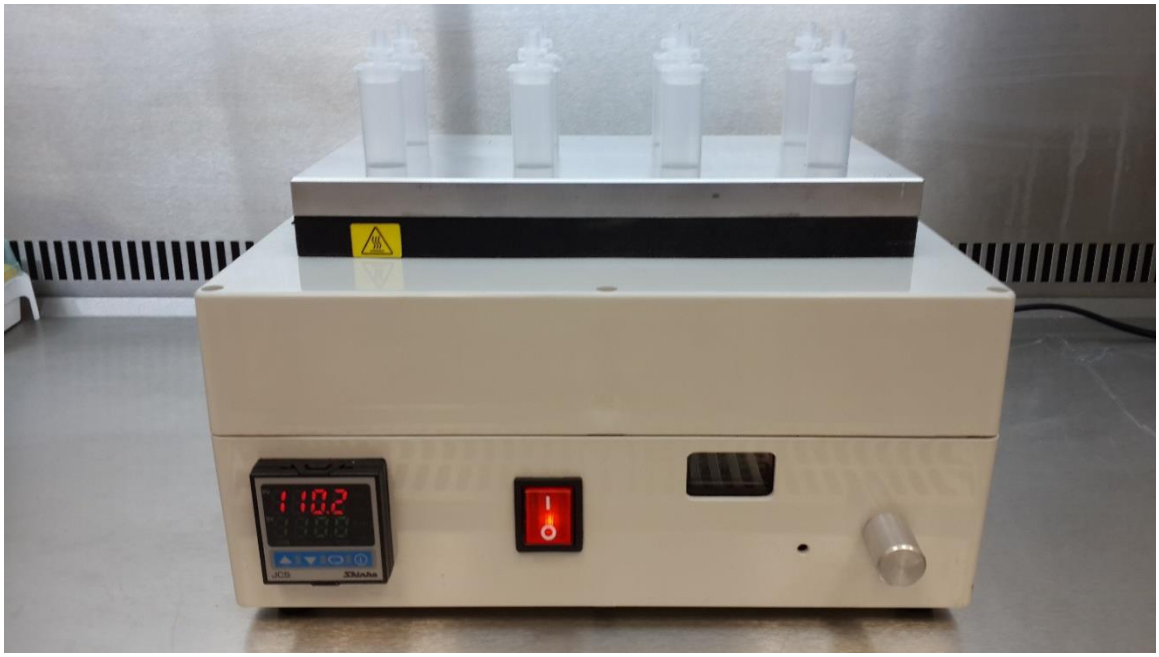


Figure 3: Hierarchical clustering heat map based on the relative abundance of the most abundant genera classified to the genus level from the representative gut sample. Individual cells in the heat map are colour-coded according to the row Z-scores,



Supplementary Figure 1: Principal coordinates analysis based on Bray–Curtis distances of samples extracted by various methods for selected clinical sputum sample (A) and representative gut sample (B). Samples that are more similar to one another are ordinated closer together.

Supplementary Figure 2: Lysis micro tube (LMT).



Supplementary Figure 3: NWU automated lyser device.

Chapter 5 – Insights into the TB resistome

This chapter consists of a full length article submitted to *Nature Communications* and is currently in the process of editorial decision on the articles suitability for the selected journal. The author guidelines followed in submitting to the Journal and the supplementary material can be found in Annexure C. The manuscript is presented in this thesis as it was submitted to the journal. Figures, referred to in text and supplementary information can be found beneath the manuscript. The chapter addresses the second objective of the study as outlined in Chapter 1. Personal contributions to the below manuscript included the following: Initial conceptualization, all wet laboratory work associated with the below manuscript, data generation, data analyses and also personally writing the final manuscript.

1. Manuscript 3

An attempt to characterise the TB resistome – a metagenomics approach

Mann, B.C.[1], Bezuidenhout, J.J.[2*], & Grobler, A.F. [1]

[1]DSI/NWU Preclinical Drug Development Platform, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa

[2]Unit for Environmental Science and Management: Microbiology, North-West University, Potchefstroom Campus, South Africa.

***Corresponding author(s)**

jaco.bezuidenhout@nwu.ac.za

Abstract

Microbiome dysbiosis has adverse health effects on the human body. Both drug susceptible and multi-drug resistant tuberculosis (TB) treatment require long term use of a variety of antibiotics, sometimes for more than 20 months. This continued use of antibiotics may induce severe microbiome dysbiosis. The aim of this study was to elucidate the impact of drug susceptible and drug resistant TB treatment on the respiratory tract microbiome and the associated resistome by a metagenomics approach. The study included 3 patients on treatment for drug susceptible TB treatment and 3 patients on multi-drug resistant TB treatment, as well as 2 healthy controls. An altered microbiome and an increase of richness were observed in both treatment groups. Significant changes in microbiome composition may give foreign and less abundant, bacteria a chance to proliferate, especially if the healthy commensals are depleted during the treatment process. In addition to this an altered resistome at the subtype level was observed, with a noticeable increase in the presence of detectable aminoglycoside resistance genes. Collectively, both drug susceptible and multi-drug resistant TB treatment induced dysbiosis of the respiratory microbiota, accompanied by changes in the resistome at the subtype level.

Keywords: Microbiome, Tuberculosis, Metagenomics, Antimicrobial Resistance, Bioinformatics

Introduction

Since its declaration as a global emergency by the World Health Organisation, tuberculosis (TB) has been, and still remains a major challenge around the globe. TB is now recognised as the leading cause of death by a one single infectious agent, and one of the top 10 causes of death worldwide [1]. Caused by the bacillus *Mycobacterium tuberculosis*, and other members of the *Mycobacterium tuberculosis* complex (MTBC), the disease is generally spread when those who are ill, expel bacteria into the air for example by coughing [1, 2]. TB is unusually difficult to treat; and as such tuberculosis therapy is an exceptionally long process, often requiring a minimum of 6 months of treatment with several first line antimicrobials to treat the disease. In the case of drug resistant and multidrug resistant TB, treatment can last up to 2 years with the daily administration of second line antimicrobials [1, 3, 4]. The fight against TB is furthermore hampered by poor adherence and non-compliance to treatment regimens, which often occur due to side effects, social stigmatization, illiteracy and insufficient information on the part of patients [5]. Poor adherence and non-compliance have been found to be the major causes of treatment failure and of the emergence of drug-resistant TB [6].

Beyond drug resistant tuberculosis, the vast and negligent use of the antibiotics has contributed significantly to the proliferation of both multidrug resistant (MDR) and extensively drug resistant (XDR) bacteria. Organisms such as the ESKAPE pathogens, consisting of *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* spp., are becoming increasingly difficult to treat [7,8]. The collection of organisms and factors associated with antimicrobial resistance in the microbiome can collectively be referred to as the resistome. Studies have shown that the administration of antibiotics can affect the human microbiome, and thus the resistome in various ways; including altering the diversity, increasing the abundance of drug resistant bacteria and increasing the overall load of antibiotic resistance genes (ARGs) [2, 9, 10].

Conventional first line anti-tuberculosis drug regimens generally consist of rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), and ethambutol (EMB). While it is known that RIF is a broad spectrum antimicrobial, the remainder are said to target only Mycobacteria; a claim recent studies have demonstrated to not necessarily be the case [3, 11]. This intense long

term antimicrobial exposure has been shown to have a severe dysbiotic effect on the microbiome of both the respiratory tract and gut in human and animal studies [3, 11, 12, 13]. Current literature suggests that this antibiotic-induced dysbiosis includes not only a shift in abundance and diversity, but also a shift in the functionality of the microbiome, which can persist for years after treatment has ended [11, 14]. Additionally, there is evidence that tuberculosis therapy may generate a distinctive ecological niche where commensals are depleted, giving rise to the potential enrichment and proliferation of resistant genotypes, some of which may be opportunistic pathogens [3, 4, 11, 15]. At the time of writing, several studies have drawn connections between the microbiome and tuberculosis, but characterization of the associated resistomes have not been described. The aim of this study was the evalaution the resistome of patients that are known to be on tuberculosis therapy at various time points in comparison to those of healthy controls.

Results

Demographic and treatment characteristics of the participants

The study included 2 healthy controls, 3 patients receiving treatment for drug susceptible TB and 3 patients receiving treatment for drug-resistant TB. All individuals were endemic to the North-West Province, South Africa. The TB status was determined by smear, culture and Xpert assay. Patients were already on treatment at the first selected time point and were all culture and smear negative by the second time point. The mean age of the TB patients was 47, all participants except one control were male and the time between the collected samples ranged from 4 to 11 months for TB patients and 6 months for controls. A table with detailed participant data can be found in Supplementary Table 1. Patients were treated according to the National Tuberculosis Management Guidelines [21], the Management of Drug Resistant TB Policy Guidelines [22] and the New drugs and regimens for the management of drug-resistant TB in South Africa framework [23]. Treatment for drug susceptible TB patients included treatment with isoniazid (INH), pyrazinamide (PZA), ethambutol (EMB) and rifampicin (RIF) with the addition of streptomycin (STREP) during the intensive phase followed by RIF and INH during the continuation phase. The MDR-TB treatment for the patients is highly individualized and changes according to the patients' tolerance and toxicity. It was therefore difficult to determine precisely which drugs these patients were exposed to throughout the treatment period. Nonetheless, according to current

guidelines, these patients were treated with a minimum of 4 drugs likely to be effective but most probably as many as were tolerated. A standardised approach to MDR-TB treatment is recommended for all newly diagnosed MDR- or XDR-TB patients. This includes an intensive phase with kanamycin or amikacin, moxifloxacin, ethionamide, terizidone or cycloserine and PZA, followed by a continuation phase of moxifloxacin/ levofloxacin, ethionamide, terizidone or cycloserine and PZA. In some instances clofazimine, bedaquiline and linezolid are also added at the clinician's discretion.

Changes in the relative abundance of the core microbiota at taxonomic level

At the phylum level, the core phyla of the controls were found to consist of Firmicutes and Proteobacteria with low abundances of Fusobacteria, Bacteroidetes and Actinobacteria. Patients on treatment for drug susceptible TB and drug resistant TB had a similar core microbiome, but with increased abundances of Actinobacteria, Bacteroidetes and Fusobacteria (Supplementary figure 1). At the genus, level the core genera associated with healthy controls consisted of *Streptococcus, Haemophilus, Veilonella, Gamella, Rothia, Neiseria and Fusobacterium*. The core genera associated with that of patients undergoing treatment showed a large degree of inter-individual variability regarding abundance, *Streptococcus* was present in all samples and was found to be the most abundant; *Actinomyces, Gemella, Haemophilus, Neisseria, Prevotella, Rothia and Veillonella* all featured prominently among all patient samples, but abundances varied between patients and time points (Supplementary Figure 2). Additionally the following genera were found in various patient samples but not in the healthy controls: *Treponema, Dialister, Klebsiella, Parvimonas, Peptostreptococcus, Corynebacterium, Capnocytophaga, Tannerella, Bacteroides, Actinobaculum, Kocuria, Tropheryma, Olsenella, Abiotrophia, Lautropia, Morococcus, Moraxella*.

At the species level, controls did show variation in abundance between time point one and two, although the species making up the majority of the microbiome remained relatively consistent. Control microbiomes were primarily dominated by streptococcal species including: *Streptococcus oralis, Streptococcus parasanguinis* and *Streptococcus salivarius.* Among patients the most abundant species varied between patients and time points, indicating a constantly shifting microbiome during the course of treatment. The over-all comparison revealed that patient samples were also dominated by streptococcal species, including: *Streptococcus oralis, Streptococcus parasanguinis* and *Streptococcus salivarius,*

but in addition exhibited an increased abundance of *Streptococcus mitis*. Several species including: *Prevotella melaninogenica, Gemella haemolysins, Streptococcus gordonii, Veillonella parvula, Actinomyces odontolyticus, Neisseria flavescens* and *Rothia mucilaginosa* were found to be differentially enriched among patient samples in comparison to samples from healthy controls as seen in Figure 1. The core genera making up the majority of the microbiome remained similar from time point one to two in both treatment groups with changes observed in the relative abundance of these genera present between time points. Comparison using the paired Wilcoxon signed-rank tests revealed that none of the major microbial taxa present in samples were found to be significantly altered at the species level when comparing samples from time point one and time point two, for either the susceptible or resistant treatment groups. Thus these samples were grouped for further downstream comparison.

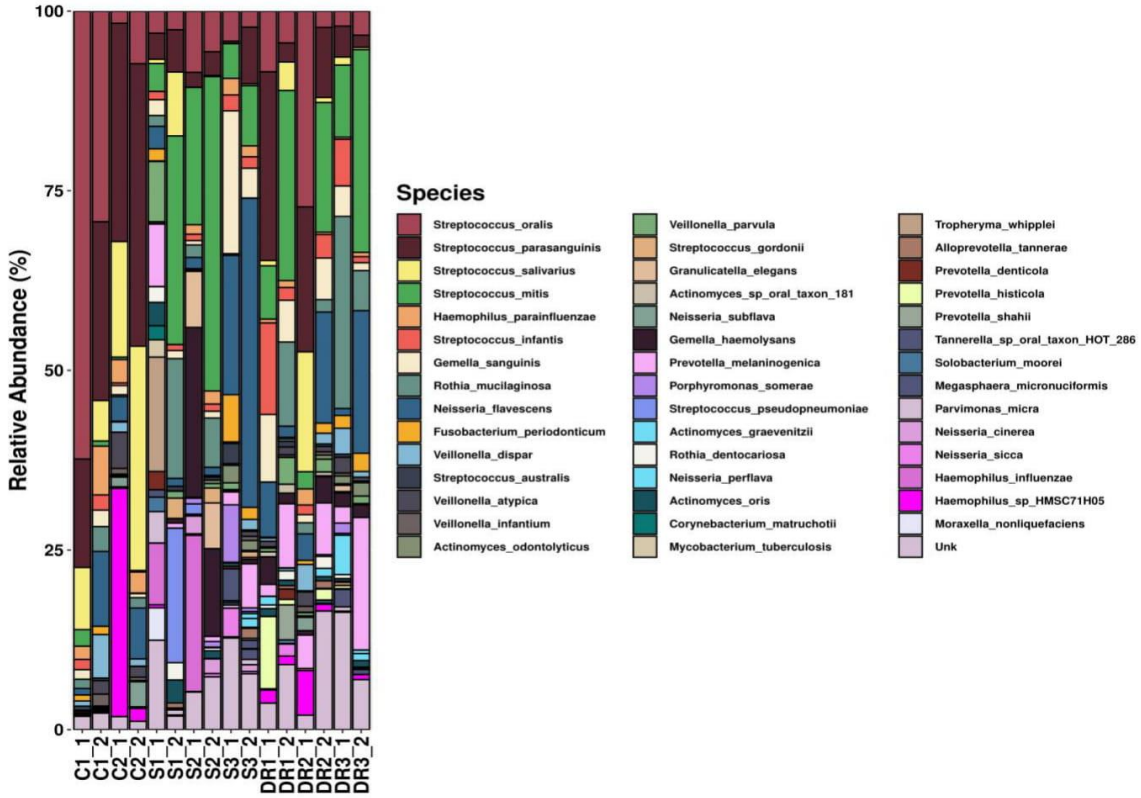


Figure 1: The relative abundance of major taxa at the species level for all samples in this study. The initial letter refers to the group: C = Control, S = Drug susceptible treatment group and DR = Drug resistant treatment group. The first number denotes the individual and the second the time point (e.g. C1_1 = control sample 1, time point 1; S2_2 = drug susceptible treatment group patient 2, time point 2).

The Wilcoxon rank sum test comparison of species found to be present in the majority of samples when comparing the control group to the susceptible treatment group and the susceptible treatment group to the resistant treatment group revealed significant increases and decreases of several of the most abundant taxa as depicted in Figure 2.

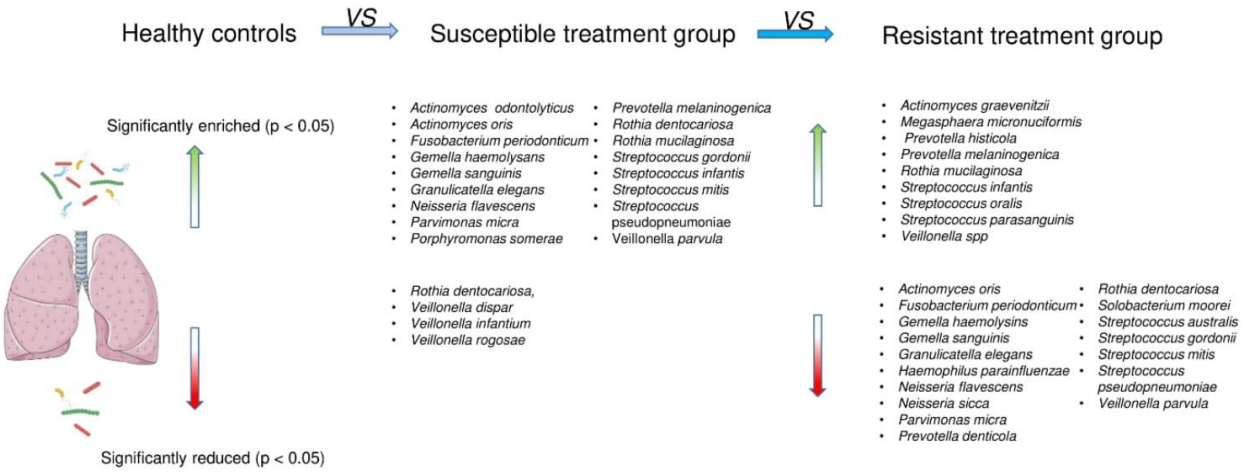


Figure 2: Significant increases ($p < 0.05$) and decreases ($p > 0.05$) observed between the major taxa present when first comparing the control group to the susceptible treatment group followed by comparison of the susceptible treatment group to the drug resistant treatment group.

Over-all initial comparisons of taxonomy revealed that the core microbiota remains present but there are significant changes in the abundance of several of the microbes present in response to treatment in the drug susceptible treatment group. Upon examination of the drug resistant treatment group, it is noticeable that there is further enrichment in the abundance of select species and the majority of the species found to be enriched among the susceptible treatment group are now significantly reduced in response to the drug resistant TB treatment.Additionally the following species: *Actinomyces oris, Alloprevotella tannerae, Corynebacterium matruchotii, Haemophilus influenzae, Moraxella nonliquefaciens, Neisseria cinerea, Neisseria sicca, Parvimonas micra, Prevotella denticola, Prevotella shahii, Solobacterium moorei, Tannerella sp oral taxon HOT 286* and *Tropheryma whipplei*, were found to be present in various patient samples, but not in samples from controls.

These observations reveal that although the majority of species remain present, there is a large degree of dysbiosis, while there is also an increased prevalence of species not found to be present in healthy controls. Several of the enriched taxa and those found present exclusively in patient groups are known to cause opportunistic infection; these include *Actinomyces oris. Moraxella nonliquefaciens, Neisseria cinerea, Neisseria sicca,*

Parvimonas micra, Prevotella denticola, Streptococcus mitis, Prevotella melaninogenica, Gemella haemolysans, Streptococcus gordonii, Veillonella parvula, Neisseria flavescens and *Rothia mucilaginosa. Haemophilus influenza* and *Tropheryma whipplei,* also found in patient samples are both well-known pathogens [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36]. A summary of potential/opportunistic pathogens and the samples they were identified in can be found in Supplementary Table 3. *Mycobacterium tuberculosis* was only identified in two of the samples from the drug susceptible group at time point 1 only, which is to be expected, as the patients were already on treatment at the time the samples were collected.

Alpha and beta diversity metrics

Next the microbiota richness, diversity and evenness were calculated at the species level. The microbiota richness was determined by the Observed and Chao1 indices. The microbiota diversity was measured by the Shannon index, which is based on both the number and the evenness of the observed features. For alpha diversity analyses, controls were grouped and patients' samples were grouped into drug susceptible and drug resistant treatment groups. The Wilcoxon rank sum test comparison indicated a significant increase in diversity among treatment groups according to the Observed and Chao1 indices (Figure 3 A and B) when comparing the healthy control group to the susceptible TB treatment group ($p < 0.05$) as well as the drug resistant treatment group ($p < 0.05$). Regarding the Shannon index (Figure 1C), there was a significant increase in diversity for both groups when compared to those of the healthy controls ($p < 0.05$).

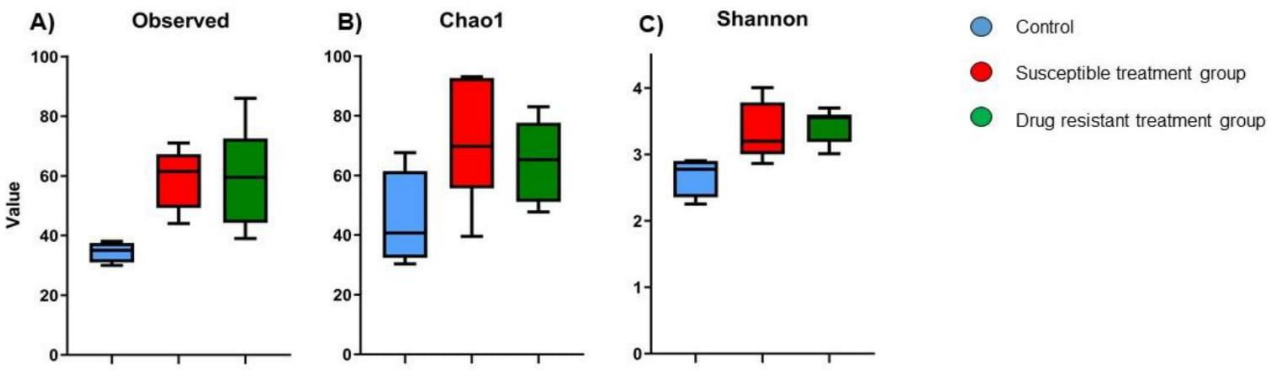


Figure 3: Alpha diversity of the control, drug susceptible and drug resistant treatment groups estimated using the Shannon (A), Chao1 (B) and Observed (C) indices. Error bars indicate minimum and maximum values.

Additionally, similarities and differences of the microbial communities among groups was assessed by pair-wise beta diversity clustering analyses using the Bray-Curtis dissimilarity matrix to compare the bacterial community structure between the three groups, as visualised by two-dimensional NMDS. The Analysis of Similarities statistic (ANOSIM) revealed a significant difference in microbial composition when comparing both the drug susceptible and drug resistant treatment groups to the samples obtained from healthy controls (Supplementary Figure 3, A and B). The microbiota of the drug resistant treatment group clustered separately from the healthy control group, while samples from the drug susceptible treatment group were more dispersed. Samples from the control group formed a close cluster, suggesting high sample similarity as opposed to widespread clustering of samples for the treatment groups, reflecting high cohort variability.

A degree of heterogeneity was observed among samples from the susceptible treatment group as well as the samples from time point 1 for the drug resistant treatment group. Samples from time point 2 tended to cluster closer together. The beta-diversity analyses of the bacterial communities comparing the two treatment groups indicated that these two groups were not significantly different from each other, although samples from the drug resistance treatment group did cluster closer together, indicating a higher degree of variability among samples from the susceptible treatment group (Supplementary Figure 3, C). Beta dispersion analyses revealed that the dispersion among the susceptible treatment group was significantly different when compared to that of the control group ($p < 0.05$), but no significant differences were observed when comparing the drug resistance group to either to susceptible or control group ($p > 0.05$).

Resistome

An investigation of samples for the presence of antibiotic resistance genes using DeepARG revealed that the resistome is rather similar for all samples at the Type level, with the exception of an influx of aminoglycoside, sulphonamide and nucleoside resistance genes among patient samples (Figure 4, A). An antibiotic resistant determinant was referred to as part of the core if found to be present in the majority of samples assessed within each group. Upon examination of control samples at the subtype level, a core resistome consisting of the following genes was identified: farB, MecA/PBP-1A, PBP-2x, MecB/PBP-1B, penA, patB, mel, mefA, RlmA(ii), lsaC, ermB, ermF, ermX, macA, macB, mdtG, mdtK, emrB, mtrC, mtrE, mtrD, mtrR, tetA(46), tetB(46), tetA(60), tetB(60), tetC, tetM, tetR, tetA, tetc, tetQ, tetW, tet32 and ugd (Figure 4, B).

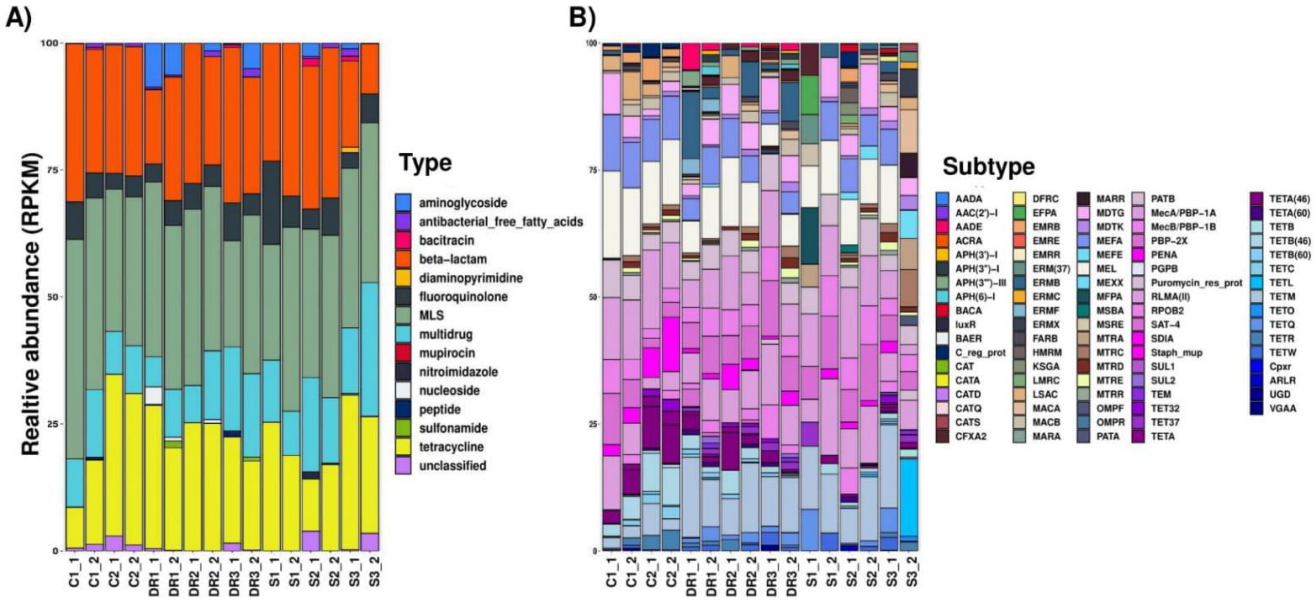


Figure 4: Relative abundance of resistance genes to different types (A) and subtypes (B) of antibiotics across all samples in this study, based on the relative abundance (RPKM) of ARGs.

Although differences in relative abundance of the antibiotic resistance genes present were noted, the genes present within controls were largely identical for both individuals at both time points (Figure 4, B, and Supplementary Table 2). Subtypes identified within patient samples exhibited a large degree of inter-individual variability regarding the presence/absence of genes and their relative abundance, both between patients from both groups and from time point 1 to 2 among all samples:

(a) For the drug susceptible treatment group, the following core group of resistance determinants were identified: MecA/PBP-1A, PBP-2x, MecB/PBP-1B, penA, patB, mel, ermB, mdtG, tetM.

(b) Within the drug resistant treatment group on the other hand, the following core genes were identified: aadE, farB, MecA/PBP-1A, PBP-2x, MecB/PBP-1B, penA, CfxA2, patB, mel, mefA, RlmA(ii), IsaC ermB, mdtG, rpoB2, mtrD, tetA(46), tetB(46), tetA(60), tetB(60), tetA, tetC, tetM, tetQ, tetR, tetW, tet32, tet37.

(c) The following genes were found to be conserved in both treatment groups in comparison to healthy controls: RlmA(ii), MecA/PBP-1A, PBP-2x, MecB/PBP-1B, penA, patB, mel, ermB, and mdtG.

(d) The genes mel, mefA, tetA(46), tetB(46), tetQ and mdtG were found to be conserved in above 90% of samples.

(e) In addition to this, several genes were found to be present in samples that were not present in healthy controls. These include several aminoglycoside, and multidrug resistance genes as summarised in Table 1. A detailed summary of presence and absence of genes found in control and both treatment groups can be found in Supplementary Table 2.

Table 1 : Antimicrobial resistance genes identified in patient samples, but not found in samples from healthy controls

Resistance type	Genes identified
Aminoglycoside	aadA, aadE, acra, AAC(2')-I, APH(3')-I, APH(3'')-I, APH(3'')-I, APH(3''')-III, APH(6)-I, ksgA
Fluoroquinolone	mfpA, patA
Phenicol	catQ
Sulphonamide	sul1, sul2
Nucleoside	SAT-4
MLS	lmrC, mefE
MDR	baeR, acrA, efpA, emrB, emrE, emrR, hmrM, marA, mexX, mtrA
Unclassified	luxR, camp-regulatory_protein, puromycin_resistance_protein

Differences regarding the resistome associated with the samples were assessed by pair-wise beta diversity clustering analyses using the Bray-Curtis dissimilarity matrix to compare the bacterial community structure between the three groups, visualised by two-dimensional NMDS. The non-metric multidimensional scaling (NMDS) ordination plot showing the Bray-Curtis dissimilarity matrices between all the samples based on the relative abundance of ARGs (RPKM) revealed that there was, however, a significant difference when comparing the healthy control group to the drug resistant treatment group (Supplementary Figure 3, E). The resistome of controls clustered closely, and remained relatively stable. No significant difference was observed when comparing the control and drug susceptible treatment groups or the drug susceptible and drug resistant treatment groups (Supplementary Figure 3, D and F)

Discussion

This study is to our knowledge only the second to explore the lung microbial communities associated with pulmonary tuberculosis using shotgun metagenomics sequencing techniques, and the first noting changes to the respiratory resistome of patients undergoing drug resistant tuberculosis therapy. In addition, at the time of writing we found no previous studies that attempted to elucidate changes in the respiratory tract resistome associated with pulmonary tuberculosis in response to treatment [37, 38, 16, 39, 40, 14, 41, 42, 13, 43, 44, 40]. In view of the fact that the study is the first to explore the respiratory microbiome in

response to drug resistant tuberculosis therapy and the TB associated resistome, and because of the limited number of participants included in the study, the results need to be interpreted with particular caution.

Several studies on the gut microbiome have indicated that significant shifts in microbiome composition occurs in response to tuberculosis drug administration, but there is a lack of similar studies focussing on shifts in the respiratory microbiome in response to treatment [13, 14, 40, 41, 43, 44]. Earlier studies characterising the TB respiratory microbiome have focused primarily on the changes in the microbiome in response to TB infection rather than evaluating changes in the respiratory microbiome in response to treatment [16, 37, 38, 39, 42]. Based on the results of this study, there is a clear shift in microbiome composition in response to treatment, and alpha diversity metrics did reveal an interesting trend. Whereas several gut studies found that microbial diversity was significantly reduced in response to treatment, this study found an overall increase in the respiratory tract microbiome diversity when comparing the two treatment groups to those of healthy controls [4, 43, 44]. Since some of the patients included in this study have been identified as HIV positive, the results should be interpreted within that context - previous studies have demonstrated that infection with HIV could lead to a more diverse lung microbiome. Additional studies will be needed to determine whether HIV, the TB treatment or a combination of the two are responsible for the observed increase in diversity [45].

At the time of writing, two studies have been found that have assessed changes in the microbiome in response to drug susceptible tuberculosis therapy, and no studies have been found that have attempted to assess the impact of multi-drug-resistant tuberculosis therapy on the respiratory microbiome [12, 46]. One of the two studies that have assessed the respiratory microbiome in response to treatment by Zhou *et al*. [46], found that *Cupriavidus* replaces *Streptococcus* as the most abundant genera when comparing TB patients with healthy controls. The second study by Wu *et al.* [12], where a recurrent TB and a treatment failure group were included, can be used as a point of comparison for changes induced by treatment. The study by Wu *et al.* [12], found that healthy controls were dominated by *Prevotella*, *Streptococcus* and *Neisseri*a. This finding is contradictory to the results of the current study where the composition of the microbiome samples from healthy controls at the genus level consisted primarily of *Streptococcus, Haemophilus* and *Veilonella*. When examining the groups that received treatment, the Wu *et al.* study found that *Pseudomonas* was more abundant and more commonly present in treatment failure and recurrent TB subjects, while also demonstrating that although there is a decrease in *Prevotella, Streptococcus* and *Neisseria* still remain the most dominant genera in the groups that have

received treatment. Similarly to the study by Wu and colleagues, this study also found *Streptococcus* be one of the most dominant genera among all patient samples, but beyond that there was a large degree of inter-individual variability regarding abundance of the main genera present in patient samples. *Actinomyces, Gemella, Haemophilus, Neisseria, Prevotella, Rothia and Veillonella* all featured prominently although at varying abundances. No clear correlations between the findings of these prior studies and the current study could be found, but this could be explained by the differences in methodology applied as well as the differences associated with the populations sampled, which makes it difficult to compare the results of the different studies. This is the first study to approach TB and the microbiome with a cohort of South African patients and controls, and previous studies on the topic have demonstrated that there can be differences based on population groups [40, 41, 42]. The study also demonstrates that, although there was no significant difference between the cohort of patients receiving drug resistant tuberculosis therapy and the group of patients receiving treatment for drug susceptible TB based on beta diversity testing, the microbiome was significantly altered when compared to that of healthy individuals. In addition to this, the drug resistant treatment group also showed a lower degree of variability.

A recent study by Wang *et al.* [44], has begun elucidating the long-term effects of multi-drug-resistant tuberculosis treatment on gut microbiota and its potential health consequences [44]. The authors found that MDR-TB treatment induced a lasting gut microbiota dysbiosis, with reduced richness and a radically altered taxonomic composition that still persists up to 8 years after recovery. The altered taxonomic composition may have health implications. In the current study, clustering of the samples for time point two (Supplementary Figure 3) suggests that the microbial communities might reach some point of homeostasis in response to MDR-TB treatment. If these changes in community structure persists similarly to that of the gut microbiome as demonstrated by Wang *et al.* [44], it remains to be seen whether the change has any severe health implications. Regardless, the overall evaluations of changes in taxonomic composition has highlighted that antimicrobial treatment may give foreign/less abundant, potentially more resilient bacteria a chance to proliferate, especially if the healthy commensals are more susceptible to decline during the treatment process [3, 4, 11]. The results of this investigation of the TB resistome and its response to treatment demonstrate that along with the significant changes in microbial composition there is also, as expected, a shift in the resistance profile of the patients undergoing treatment. The resistance gene profiles at both the type and subtype level for healthy controls remain fairly unperturbed from time point one to time point two.

A core resistome associated with resistance to beta-lactam, fluoroquinolone, tetracycline and macrolide, lincosamide and streptogramin (MLS) antibiotics was identified. Resistome studies regarding the respiratory tract are hard to come by, but a recent study by Mac Aogáin *et al*. [47], assessed the airway resistome associated with several chronic respiratory disease states (severe asthma, chronic obstructive pulmonary disease (COPD), and bronchiectasis) in comparison to non-diseased (healthy) individuals in order to characterize airway resistomes. Similar to this study, the authors demonstrated the presence of a core resistome, dominated by genes from common antibiotic classes including macrolides, β-lactams, and fluroquinolones unrelated to antibiotic exposure amongst healthy controls. In addition to this, the Aogáin *et al*. study [47], established that broad changes in the resistome is observed under disease conditions like COPD and bronchiectasis, where frequent use of antibiotics may be required [47]. This phenomenon is also seen in the current study, although a core resistome associated with resistance to beta-lactam, fluoroquinolone, tetracycline and macrolide, lincosamide and streptogramin (MLS) antibiotics remains largely maintained at the type level, when comparing healthy controls to the treatment groups; there is a clear influx in the relative abundance of genes conferring resistance to aminoglycosides, with a large shift in overall resistome composition at the subtype level (Figure 4). At the subtype level, significant changes in composition can be seen when comparing the control group to the susceptible treatment group and the drug resistant treatment group respectively (Supplementary Figure 3). The most noteworthy change is the trend of increasing aminoglycoside resistance genes present among various patient samples (Figure 4 and Supplementary Table 2).

Aminoglycosides are potent, highly bactericidal antibiotics and are commonly used to treat serious Gram-negative bacterial infections (*e.g.*, *Escherichia coli*, *Klebsiella pneumoniae*, *and Pseudomonas aeruginosa*), certain gram-positive pathogens (*e.g.* Staphylococci) and some mycobacterial infections, while often also being employed as prophylaxis in high risk cases [48. 49. 50]. With the advent of safer antimicrobials there has been a shift away from the prolonged administration of aminoglycosides, but growing instances of resistance among several pathogens to commonly used antibiotics has shown that several aminoglycosides still have clinical utility [49]. This is especially true in instances where pathogens and opportunistic pathogens exhibit resistance to other safer antimicrobials, for example the treatment of multidrug-resistant *Acinetobacter and Pseudomonas* infections or infections caused by extended-spectrum β-lactamase-bearing *Enterobacteriaeceae* species [49, 51]. Similar to other antibiotics, resistance to aminoglycosides is growing and consequently failure of treatments with aminoglycosides is becoming more common [49, 50].

During the course of this study several aminoglycoside resistance genes were identified within both the susceptible and drug resistant treatment groups of the genes identified (Table 1). Of the genes identified: *aadA, aadE, AAC(2')-I, APH(3')-I, APH(3'')-I, APH(3''')-III, and APH(6)-I* have all been described as acquired resistance genes, suggesting their increased presence may be directly related to exposure to this class of antimicrobial during the course of treatment [52]. *APH(3')-I, APH(3'')-I, APH(3''')-III,* and *APH(6)-I* as well as the sulphonamide resistance genes *sul1, sul2* have previously been described as plasmid mediated ARGs, and their presence should be a cause for concern, particularly if they become persistently maintained in the microbiomes of TB patients over the long term [44, 53].

Evidence suggests that the prolonged use of antibiotics is associated with alterations in the resident population with a possible increase in the occurrence of opportunistic microorganisms [54]. Previous studies, and the current one have demonstrated that foreign or less abundant, potentially more resilient bacteria may have a chance to proliferate among the TB infected population groups receiving tuberculosis therapy, while functional changes and changes in the presence and abundance of ARGs are also impacted [3, 4, 11]. Approximately 1.7 million people are newly infected with human immunodeficiency virus (HIV) every year, with a global estimate of around 38 million people living with the disease [55]. HIV infection is one the strongest risk factors for TB infection and subsequent progression to TB disease. This is a particularly concerning issue, especially in low and middle income countries. In addition to the increased risk for TB infection, HIV infected individuals are exceptionally prone to other opportunistic infections [56]. Considering the high prevalence of HIV among individuals receiving treatment for tuberculosis and drug resistant tuberculosis, and their increased risk for opportunistic infections, the perturbations of the microbiome as well as the shifts recorded in the resistome, especially an increased prevalence in aminoglycoside resistance genes, may be a cause for concern, especially if these genes are maintained in the resistome over time [9, 57, 58].

In conclusion the study highlights that similarly to treatment of drug susceptible TB treatment, drug resistant TB treatment also has a significant impact on the respiratory microbiome. Furthermore, the study demonstrates that the associated resistome changes with treatment, as characterised by an increase in aminoglycoside resistance genes. However, extensive studies on larger cohorts of patients will be central to determining the physiological and potential clinical significance of these initial observations.

Materials and methods

Sample collection and DNA extraction

Primary sputum samples were originally collected by staff at the Orkney AngloGold Ashanti Health Westvaal Hospital (the Westvaal Hospital for short) by qualified staff members of the TB Laboratory as a part of clinical work. The study population included newly and previously diagnosed TB individuals being treated at the health care facility. Sputum samples were coded by the Westvaal Hospital: Samples were anonymized and meta-data pertaining to each sample (collection date, age, sex, TB treatment status, HIV status, smear test results and culture results) was collected. A portion of each sample collected was refrigerated at 2-8 ºC, until collected from the Westvaal Hospital, after which they were frozen at -80 ºC. These samples formed part of a biobank, some of which were used for this study. Several patient samples from the biobank were identified for inclusion in this study. Inclusion criteria for patients were that the patients had to test positive for TB, the patient had to be on treatment for tuberculosis, and samples had to be available at, at least two different time points; one where the patient was on treatment, and one several months later or once the person has tested negative. A total of 12 samples from 6 patients were included in the study, 3 testing positive for drug susceptible TB (labelled as Susceptible or S) and 3 testing positive for drug resistant TB (labelled as Drug-resistant or DR). To generate a baseline for comparison, samples from individuals acting as healthy controls were taken in the form of oral-pharyngeal swabs [16]; all controls were non-smokers, and did not take any antibiotics for at least 2 months prior to the onset of the study. Swabs (Sigma Transwab®, MWE, UK) were taken at an initial time point and again 6 months later. Two individuals were included to act as controls (2 samples each at 2 time points). DNA was extracted with the use of the NucleoSpin® DNA Stool kit (Macherey-Nagel, Germany). This method incorporates chemical, enzymatic and mechanical lyses. For sputum samples, 1 mL of sputum was mixed with 1 mL of extraction buffer (ST1), with the addition of 20 µl Proteinase K (20 mg/mL) (Macherey-Nagel, Germany). The mixture was incubated at 56 °C for 20 minutes; 1 mL of this mixture was then used for extraction according to the manufacturer's protocol. In the case of the swab samples, samples were originally eluted in 1 mL of the provided transport media, as described above. DNA quality was assessed with the use of the Tapestation genomic DNA Assay (Agilent Technologies, CA, USA), and quantity was assessed with the Qubit 2.0 DNA HS Assay (Life Technologies, Grand Island, NY).

Metagenomic sequencing and data processing

To deplete host DNA, samples were treated with the NEBNext® Microbiome DNA Enrichment Kit (New England Biolabs, UK) prior to sequencing. Whole-genome shotgun sequencing of the extracted DNA was conducted by Admera Health (SouthPlainfield, USA). Library preparation was done using the KAPA Hyper Prep kit and sequencing was carried out on the Illumina HiSeq 2000/2500 (Illumina, San Diego, USA) platform. The samples were sequenced with a 150-bp read length for paired-end sequences at a sequencing depth of 7.5 GB. Initial quality assessment was done using FastQC, following which low quality reads were removed and adapters were trimmed using fastp. The subsequent cleaned FASTQ files were mapped to the GRCh38 assembly of the human genome using bowtie 2. Reads that passed quality control (QC) and did not map to the host genome were extracted and used for subsequent downstream analyses.

Microbial taxonomy, alpha and beta diversity metrics and resistome analyses

MetaPhlAn 3.0 was used for metagenomic profiling, to assign taxonomy and determine relative abundance. MetaPhlAn relies on a reference database containing unique clade-specific marker genes selected from 17,000 reference genomes in order to identify the putative bacterial taxonomies and their relative abundances. Species were filtered by 10% occurrence across all samples and the relative abundances of remaining species were used for further statistical comparisons. Differential abundances were evaluated by the Wilcoxon rank sum test and paired Wilcoxon signed-rank tests, respectively. The Benjamini-Hochberg correction was applied to adjust p-values for multiple tests in R [17].

Community analyses was carried out in R with the aid of the vegan package [18]. The within sample alpha diversity was calculated using the Observed, Chao1 and Shannon indices; for alpha diversity analyses, data was rarefied and normalised using total sum scaling (TSS). The alpha diversity of samples was measured by Observed, Chao1 and Shannon diversity indices. The Observed and Chao1 indices act as measures of species richness. The Observed species index measures the number of distinguishable taxa in every sample, whereas the Chao1 index is a qualitative measure; besides species richness it also takes into account the ratio of singletons, and hence gives more weight to rare species. The Shannon diversity index on the other hand, involves a measure of both richness and

evenness of the microbes of the given sample. Statistically significant differences between groups were evaluated by means of the Wilcoxon rank sum test. Beta diversity was estimated by Non-metric multidimensional scaling (NMDS) analysis based on the relative abundance of taxa identified. The Bray-Curtis dissimilarity index was utilized to estimate beta-diversity and this index was visualized on a two-dimensional NMDS plot generated using ggplot2 in R. Significant dissimilarities of ordination between groups were assessed using the Analysis of Similarities statistic (ANOSIM) with 999 permutations to test the statistical significance of differences between groups. To evaluate whether one group tends to vary more than another group, Vegan's betadisper() function was applied.

Resistome analyses

Following taxonomic evaluation, reads were then queried for antibiotic resistance genes using DeepARG (version 1) [19]. Considering the clinical nature of the samples, more stringent parameters than the default were selected, including 60% coverage with a minimum probability cut off of 0.8, E-value cut off of 1e-10 and an 80 percent (80%) minimum of identity. The reads per kilobase of read per million (RPKM) was calculated for each sample as the number of reads divided by the total number of library reads per million, then divided by the gene length in kilobases. This approach normalizes the datasets based on gene length and sample sequencing depth, making the abundance comparable between samples [10, 20]. Non-metric multidimensional scaling (NMDS), using Euclidean distances between transformed normalized counts (RPKM), was used to ordinate samples based on resistome composition. Significant dissimilarity of ordination between groups was assessed using ANOSIM at both the ARG type and ARG subtype level. Hierarchical clustering analyses of ARG profiles based on transformed normalized counts (RPKM) were based on the Euclidean distance using the Ward agglomeration method. The resistome was characterised at the Type level, referring to the class of antibiotic to which resistance is conferred and at the subtype level, referring to the actual genes involved with resistance to specific classes of antibiotics at the type level.

Ethics Statement

This study was approved by the North-West University Health Research Ethics Committee (HREC: NWU-00127-18-A1). Written informed consent was obtained from the individuals

recruited to act as healthy controls. In the case of patient samples, the patients did not sign any informed consent for the production of sputum specimens as it is a non-invasive procedure; for a patient who voluntarily presents to a health care facility there is implicit consent for non-invasive processes that only require verbal discussion and assent. All specimens used in the present study are anonymous. A study number (Laboratory number) was allocated to each specimen, and investigators involved in this study were, and still are blind to the identity of the individual patients with the identity key held by the clinician. The specimens were collected as a part of clinical work up, and a portion of each specimen was stored in the biobank.

Data availability

The raw sequence files are available under NCBI Bioproject PRJNA679716.

References

1. World Health Organization. 2020. Global tuberculosis report 2020. [Web: https://www.World Health Organisation.int/teams/global-tuberculosis-programme/tb-reports]. Date of access: 20 July 2020.
2. Smith, I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews*, 16(3):463–496.
3. Namasivayam, S., Maiga, M., Yuan, W., Thovarai, V., Costa, D.L., Mittereder, L.R., Wipperman, M.F., Glickman, M.S., Dzutsev, A., Trinchieri, G. & Sher, A. 2017. Longitudinal profiling reveals a persistent intestinal dysbiosis triggered by conventional anti-tuberculosis therapy. *Microbiome*, 5(1):71.
4. Khan, N., Vidyarthi, A., Nadeem, S., Negi, S., Nair, G. & Agrewala, J. N. 2016. Alteration in the gut microbiota provokes susceptibility to tuberculosis. *Frontiers in Immunology*, 7(529).
5. AlSahafi, A.J., Shah, H.B.U., AlSayali, M.M, Najlaa, M., Assiri, M., Almohammadi, E.L., Khalawi, A., AlGarni, A., Filemban, M.K., AlOtaibe, A.K., AlFaifi, A.W.A. & AlGarni, F. 2019. High non-compliance rate with anti-tuberculosis treatment: a need to shift facility-based directly observed therapy short course (DOTS) to community mobile outreach team supervision in Saudi Arabia. *BMC Public Health,* 19(1168).
6. Bagchi, S., Ambe, G. & Sathiakumar, N. 2010. Determinants of poor adherence to anti-tuberculosis treatment in mumbai, India. *International Journal of Preventive Medicine*, 1(4): 223–232.
7. Mulani, M. S., Kamble, E. E., Kumkar, S. N., Tawre, M. S. & Pardesi, K. R. 2019. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review. *Frontiers in microbiology*, 10(539).
8. Zaman, S.B., Hussain, M.A., Nye, R., Mehta, V., Mamun, K.T. & Hossain, N. 2017. A review on antibiotic resistance: Alarm bells are ringing. *Cureus*, 9(6):e1403.
9. Willmann, M. & Peter, S. 2017. Translational metagenomics and the human resistome: Confronting the menace of the new millennium. *Journal of Molecular Medicine,* 95(1):41–51.
10. Carr, V.R., Witherden, E.A., Lee, S., Shoaie, S., Mullany, P., Proctor, G.B., Gomez-cabrero, D. & Moyes, D.L. 2020. Abundance and diversity of resistomes differ between healthy human oral cavities and gut. *Nature Communications,* 11(639).
11. Khan, N., Mendonca, L., Dhariwal, A., Fontes, G., Menzies, D., Xia, J. & Divangahi, M. 2019. *Mycobacterium tuberculosis* intestinal dysbiosis compromises alveolar

macrophage immunity to *Mycobacterium tuberculosis*. *Mucosal Immunology,* 12:772–783.

12. Wu, J., Liu, W., He, L., Huang, F., Chen, J., Cui, P., Shen, Y., Zhao, J., Wang, W., Zhang, Y., Zhu, M., Zhang, W. & Zhang, Y. 2013. Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PLoS One*. 8(12):1–11.
13. Hu, Y., Feng, Y., Wu, J., Liu, F., Zhang, Z. & Hao, Y. 2019. The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. *Frontiers in Cellular and Infection Microbiology,* 9(90):1–8.
14. Wipperman, M.F., Fitzgerald, D.W., Juste, M.A.J., Taur, Y., Namasivayam, S., Sher, A., Bean, J.M., Bucci, V. & Glickman, M.S. 2017. Antibiotic treatment for tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. *Scientific Reports,* 7(1):1–11.
15. Schwartz, D.J., Langdon, A.E. & Dantas, G. 2020. Understanding the impact of antibiotic perturbation on the human microbiome. *Genome Medicine*, 12(1):82.
16. Botero, L.E., Delgado-serrano, L., Cepeda, M.L., Bustos, J.R., Anzola, J.M., Portillo, P. Del, Robledo, J. & Zambrano, M.M. 2014. Respiratory tract clinical sample selection for microbiota analysis in patients with pulmonary tuberculosis. *BMC Microbiome*, 2:29.
17. Nie, X., Chen, J., Ma, X., Ni, Y., Shen, Y., Yu, H., Panagiotou, G. & Bao, Y. 2020. A metagenome-wide association study of gut microbiome and visceral fat accumulation. *Computational and structural biotechnology journal*, 18:2596–2609.
18. Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P., O'Hara, R.B., Simpson, G., Solymos, P., Stevens, M.H.H. & Wagner, H. 2015. Vegan: community ecology package. R package vegan, vers. 2.2-1.
19. Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P. & Zhang, L. 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, 6(1):23.
20. Cheng, X., Xu, J., Smith, G. & Zhang, Y. 2021. Metagenomic insights into dissemination of antibiotic resistance across bacterial genera in wastewater treatment. *Chemosphere*. 271:129563.
21. Department of Health. 2014. National Tuberculosis Management Guidelines. [Web: https://www.tbonline.info/media/uploads/documents/national_tuberculosis_management_guidelines_%282014%29.pdf]. Date of access: 18 June 2020.
22. Department of Health. 2013. Management of Drug Resistant TB Policy Guidelines lines. [Web: https://www.nicd.ac.za/assets/files/MANAGEMENT%20of%20drug%20resistant%20TB%20guidelines.pdf]. Date of access: 23 July 2020.

23. Department of Health. 2015. New drugs and regimens for the management of drug-resistant TB in South Africa framework. [Web: https://www.nicd.ac.za/assets/files/Introduction%20of%20new%20drugs%20and%20drug%20regimens%20for%20the%20management%20of%20drug%20resistant%20TB%20in%20SA%20-%202015.pdf]. Date of access: 10 May 2020.
24. La Scola, B. & Raoult, D. 1998. Molecular identification of *Gemella* species from three patients with endocarditis. *Journal of Clinical Microbiology*, 36(4): 866–871.
25. Bhatti, M.A. & Frank, M.O. 2000. *Veillonella parvula* Meningitis: Case report and review of Veillonella infections. *Clinical Infectious Diseases,* 31:839–840.
26. Huang, L., Ma, L., Fan, K., Li, Y., Xie, L., Xia, W., Gu, B. & Liu, G. 2014. Necrotizing pneumonia and empyema caused by *Neisseria flavescens* infection. *Journal of Thoracic Disease*, 6(5), 553–557.
27. Ramanan, P., Barreto, J.N., Osmon, D.R. & Tosh, P.K. 2014. Rothia bacteraemia: A 10-Year Experience at Mayo Clinic, Rochester, Minnesota, *Journal of Clinical Microbiology,* 52(9):3184–3189.
28. Kietzell, M. Von, Richter, H., Bruderer, T., Goldenberger, D., Emonet, S. & Strahm, C. 2016. Meningitis and bacteremia due to *Neisseria cinerea* following a percutaneous rhizotomy of the trigeminal ganglion. *Journal of Clinical Microbiology*, 54(1):233-5.
29. Adams, H. M., Joyce, L. R., Guan, Z., Akins, R. L. & Palmer, K. L. 2017. *Streptococcus mitis* and *S. oralis* lack a requirement for cdsA, the enzyme required for synthesis of major membrane phospholipids in bacteria. *Antimicrobial Agents and Chemotherapy*, 61(5): e02552-16.
30. Kondo, Y., Sato, K., Nagano, K., Nishiguchi, M., Hoshino, T. & Fujiwara, T. 2018. Involvement of PorK, a component of the type IX secretion system, in *Prevotella melaninogenica* pathogenicity. *Microbiology and Immunology*, 62(9):554-566.
31. Correa-martínez, C.L., Rauwolf, K.K., Schuler, F., Füller, M., Kampmeier, S. & Groll, A.H. 2019. *Moraxella nonliquefaciens* bloodstream infection and sepsis in a paediatric cancer patient: Case report and literature review. *BMC Infectious Diseases*, 19(1):836.
32. Mosailova, N., Truong, J., Dietrich, T. & Ashurst, J. 2019. *Streptococcus gordonii*: A Rare Cause of Infective Endocarditis. *Case Reports in Infectious Diseases*, 2019:7127848.
33. Wu, J., Kuo, C. & Wu, J. 2019. *Prevotella denticola* septic embolic cerebral infarction after difficult lower wisdom tooth extraction. *Journal of Dental Science,* 14(4):426–427.

34. Kozlova, A., Palazzolo, L. & Michael, A. 2020. *Neisseria sicca*: A rare cause of bacterial conjunctivitis. *The American Journal of Case Reports*, 21: e923135.
35. Phichaphop, C., Apiwattanakul, N., Wanitkun, S. & Boonsathorn, S. 2020. Bacterial endocarditis caused by *Actinomyces oris*: First reported case and literature review. *Journal of Investigative Medicine High Impact Case Reports*, 8.
36. Watanabe, T., Hara, Y., Yoshimi, Y., Fujita, Y., Yokoe, M. & Noguchi, Y. 2020. Clinical characteristics of bloodstream infection by *Parvimonas micra*: Retrospective case series and literature review. *BMC Infectious Diseases*, 20(578).
37. Cui, Z., Zhou, Y., Li, H., Zhang, Y., Zhang, S., Tang, S. & Guo, X. 2012. Complex sputum microbial composition in patients with pulmonary tuberculosis. *BMC Microbiology,* 12:276–283.
38. Cheung, M.K., Lam, W.Y., Fung, W.Y.W., Law, P.T.W., Au, C.H., Nong, W., Kam, K.M., Kwan, H.S. & Tsui, S.K.W. 2013. Sputum microbiota in tuberculosis as revealed by 16S rRNA Pyrosequencing. *PLoS One*, 8(1).
39. Krishna, P., Jain, A., & Bisen, P.S. 2016. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. *European Journal of Clinical Microbiology & Infectious Diseases*, 35(7):1205–1210.
40. Luo, M., Liu, Y., Wu, P., Luo, D.X., Sun, Q., Zheng, H., Hu, R., Pandol, S.J., Li, Q.F., Han, Y.P. & Zeng, Y. 2017. Alternation of gut microbiota in patients with pulmonary tuberculosis. *Frontiers in Physiology,* 8:822.
41. Maji, A., Misra, R., Dhakan, D.B., Gupta, V., Mahato, N.K., Saxena, R., Mittal, P., Thukral, N., Sharma, E., Singh, A., Virmani, R., Gaur, M., Singh, H., Hasija, Y., Arora, G., Agrawal, A., Chaudhry, A., Khurana, J.P., Sharma, V.K., Lal, R. & Singh, Y. 2018. Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. *Environmental Microbiology,* 12:772–783.
42. Hu, Y., Cheng, M., Liu, B., Dong, J., Sun, L., Yang, J., Chen, X. & Jin, Q. 2020. Metagenomic analysis of the lung microbiome in pulmonary tuberculosis - a pilot study. *Emerging Microbes and Infections,* 9(1):1444-1452.
43. Namasivayam, S., Diarra, B., Diabate, S., Sarro, S., Kone, A., Kone, B., Tolofoudie, M., Baya, B., Diakite, T., Kodio, O., Cohen, K., Id, J.H., Achenbach, C.J., Chatterjee, S., Murphy, R.L., Bishai, W., Diallo, S., Sher, A. & Id, M.M. 2020. Patients infected with *Mycobacterium africanum* versus *Mycobacterium tuberculosis* possess distinct intestinal microbiota. *PLoS Neglected Tropical Diseases,* 14(5):1–20.
44. Wang, J., Xiong, K., Zhao, S., Zhang, C., Zhang, J. & Xu, L. 2020. Long-term effects of multi-drug-resistant tuberculosis treatment on gut microbiota and its health consequences. *Frontiers in Microbiology,* 11:53.

45. Salas, J. T. & Chang, T. L. 2014. Microbiome in human immunodeficiency virus infection. *Clinics in Laboratory Medicine*, 34(4):733–745.
46. Zhou, Y., Lin, F., Cui, Z., Zhang, X., Hu, C., Shen, T., Chen, C., Zhang, X. & Guo, X. 2015. Correlation between either Cupriavidus or Porphyromonas and primary pulmonary tuberculosis found by analysing the microbiota in patients' bronchoalveolar lavage fluid. *PLoS One*, 10(5):1–18.
47. Mac Aogáin, M., Lau, K., Cai, Z., Kumar Narayana, J., Purbojati, R. W., Drautz-Moses, D. I., Gaultier, N. E., Jaggi, T. K., Tiew, P. Y., Ong, T. H., Siyue Koh, M., Lim Yick Hou, A., Abisheganaden, J. A., Tsaneva-Atanasova, K., Schuster, S. C. & Chotirmall, S. H. 2020. Metagenomics reveals a core macrolide resistome related to microbiota in chronic respiratory disease. *American Journal of Respiratory and Critical Care Medicine*, 202(3):433–447.
48. Vakulenko, S. B. & Mobashery, S. 2003. Versatility of aminoglycosides and prospects for their future. *Clinical Microbiology Reviews*, 16(3): 430–450.
49. Avent, M.L., Rogers, B.A., Cheng, A.C. & Paterson. 2011. Current use of aminoglycosides: indications, pharmacokinetics and monitoring for toxicity. *Internal Medicine Journal,* 41(6): 441-449.
50. Ramirez, M. S., Nikolaidis, N., & Tolmasky, M. E. 2013. Rise and dissemination of aminoglycoside resistance: The aac(6')-Ib paradigm. *Frontiers in Microbiology*, 4:121.
51. Eliopoulos, G.M., Drusano, G.L., Ambrose, P.G., Bhavnani, S.M., Bertino, J.S., Nafziger, A.N. & Louie. A. 2007. Back to the future: Using aminoglycosides again and how to dose them optimally. *Clinical Infectious Diseases,* 45(6):753-60.
52. van Hoek, A. H., Mevius, D., Guerra, B., Mullany, P., Roberts, A. P. & Aarts, H. J. 2011. Acquired antibiotic resistance genes: an overview. *Frontiers in Microbiology*, 2(203).
53. Khezri, A. & Avershina, E. 2021. Plasmid identification and plasmid-mediated antimicrobial gene detection in Norwegian isolates. *Microorganisms,* 9(1):52.
54. Querido, S. M., Back-Brito, G. N., Dos Santos, S. S., Leão, M. V., Koga-Ito, C. Y. & Jorge, A. O. 2011. Opportunistic microorganisms in patients undergoing antibiotic therapy for pulmonary tuberculosis. *Brazilian Journal of Microbiology*, 42(4):1321–1328.
55. Dybul, M., Attoye, T., Baptiste, S., Cherutich, P., Dabis, F., Deeks, S. G., Dieffenbach, C., Doehle, B., Goodenow, M. M., Jiang, A., Kemps, D., Lewin, S. R., Lumpkin, M. M., Mathae, L., McCune, J. M., Ndung'u, T., Nsubuga, M., Peay, H. L., Pottage, J., Warren, M. & Izukanji Sikazwe. 2021. The case for an HIV cure and how to get there. *The Lancet. HIV*, 8(1):e51–e58.

56. O'Connor, D. E., Frederix, K., Saito, S., Maama, L. B., Hirsch-Moverman, Y., Pitt, B., Hayes-Larson, E., Lebelo, L., Shale, M. & Howard, A. A. 2017. Pulmonary tuberculosis diagnostic practices among people living with the human immunodeficiency virus in Lesotho. *The International Journal of Tuberculosis and Lung Disease,* 21(10):1133–1138.
57. Relman, D.A. & Lipsitch, M. 2018. Microbiome as a tool and a target in the effort to address antimicrobial resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 115(51):12902–12910.
58. Casals-pascual, C., Vergara, A. & Vila, J. 2018. Intestinal microbiota and antibiotic resistance : Perspectives and solutions. *Human Microbiome Journal*, 9:11–15.

Acknowledgments

This study was performed, and the manuscript compiled with the support of funding provided by the South African Department of Science and Innovation in conjunction with the International Mouse Phenotyping Consortium (IMPC). Opinions expressed, findings and concluding statements are those of the authors and are not necessarily attributed to the South African Department of Science and Innovation or the International Mouse Phenotyping Consortium (IMPC).

Author information

Affiliations

DSI/NWU Preclinical Drug Development Platform, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa

Mann, B.C & Grobler, A.F

Unit for Environmental Science and Management: Microbiology, North-West University, Potchefstroom Campus, South Africa.

Contributions

All authors read and approved the final manuscript and contributions (CRediT) were as follows:

- **Mann, B.C:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization

- **Bezuidenhout, J.J:** Conceptualization, Methodology, Software, Writing - Review & Editing, Resources, Supervision, Project administration

- **Grobler, A.F:** Conceptualization, Methodology, Writing - Review & Editing, Resources, Supervision, Project administration, funding acquisition

Corresponding author

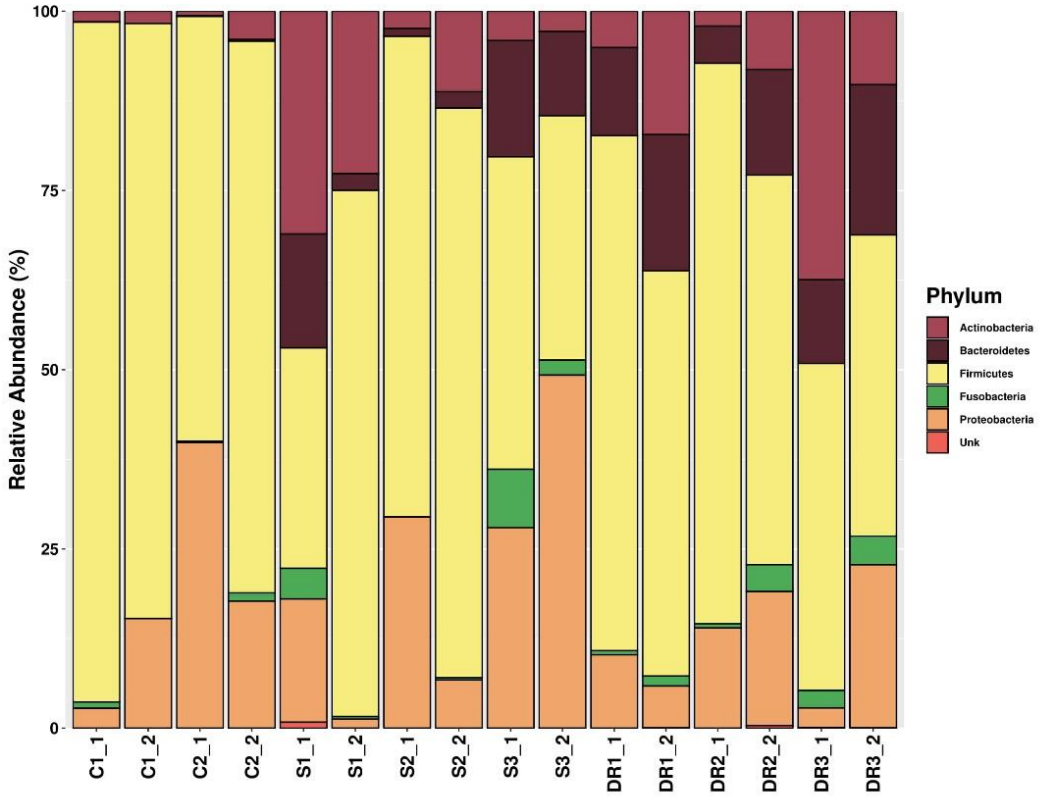
Bezuidenhout, J.J

Ethics declarations

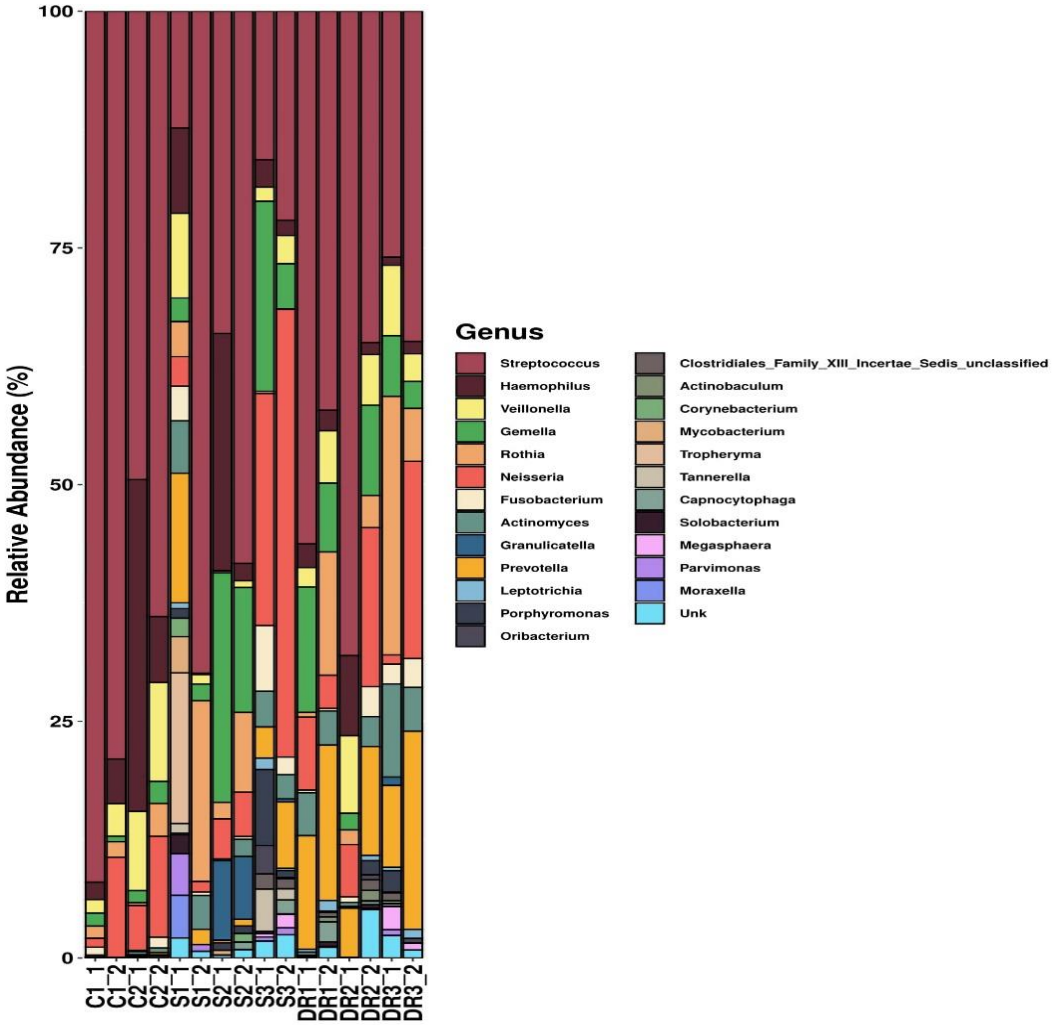
Competing interests

The authors declare no competing interests.

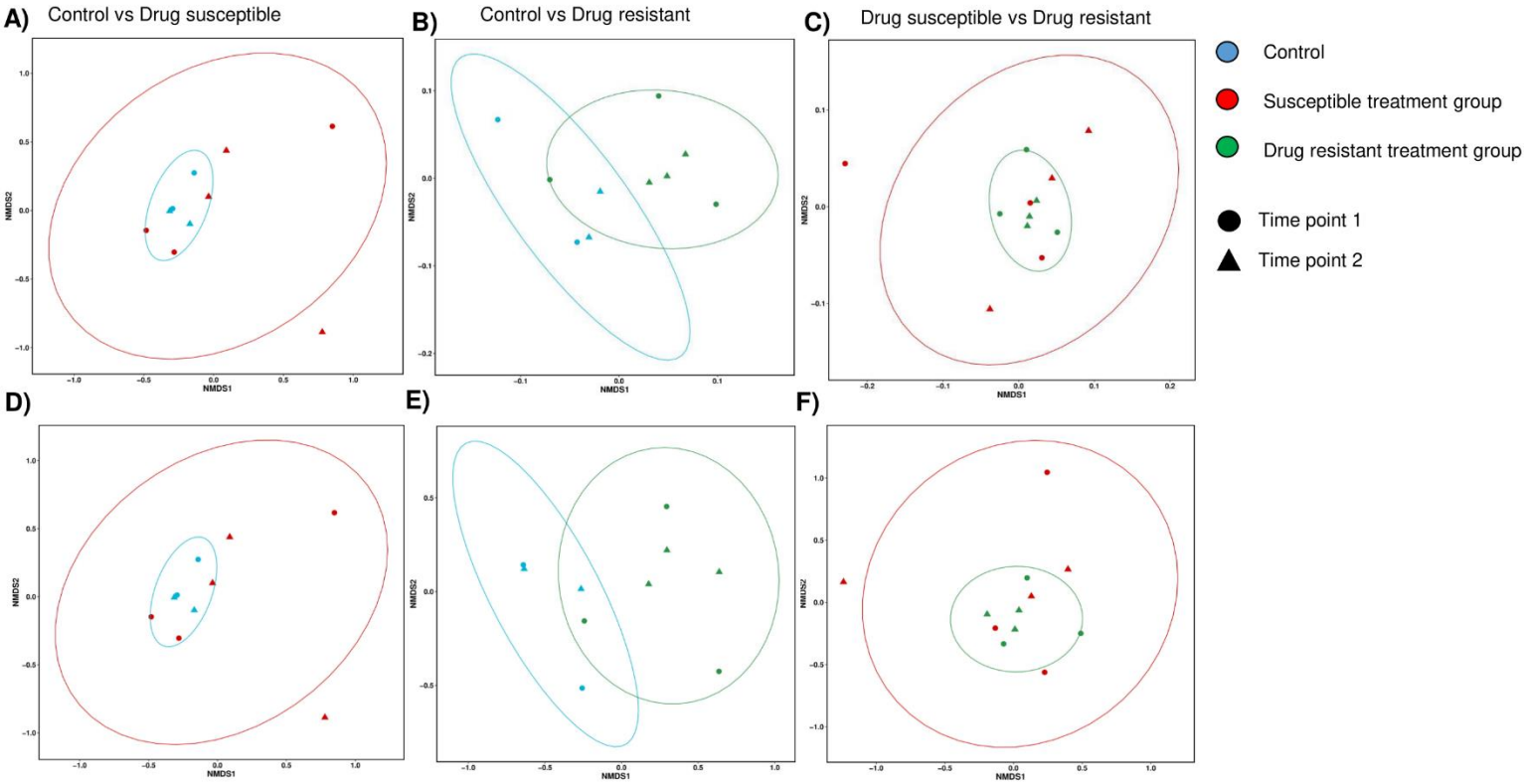
2. Manuscript 3 supplementary figures and tables



Supplementary figure 1: The relative abundance of major taxa at the phylum level for all samples in this study. The initial letter refers to the group: C = Control, S = Drug susceptible treatment group and DR = Drug resistant treatment group. The first number denotes the individual and the second the time point (e.g. C1_1 = control sample 1, time point 1; S2_2 = drug susceptible treatment group patient 2, time point 2).



Supplementary figure 2: The relative abundance of major taxa at the genus level for all samples in this study. The initial letter refers to the group: C = Control, S = Drug susceptible treatment group and DR = Drug resistant treatment group. The first number denotes the individual and the second the time point (e.g. C1_1 = control sample 1, time point 1; S2_2 = drug susceptible treatment group patient 2, time point 2).



Supplementary Figure 3: Beta-diversity estimates were calculated for the Control vs Drug susceptible group (A) ($p < 0.05$), Control vs Drug resistant group (B) ($p < 0.05$) and the Drug susceptible vs Drug resistant group (C) (n.s), using the Bray-Curtis dissimilarity index and represented here on a two-dimensional NMDS plots. Circles denote samples from time point 1 and triangles denote samples from time point two. Statistical significance was calculated using Analysis of Similarities statistic (ANOSIM) with 999 permutations. The non-metric multidimensional scaling (NMDS) ordination plot showing the Bray-Curtis dissimilarity matrices between all the samples based on the relative abundance (RPKM) of ARGs. (D) (n.s), Control vs Drug resistant group (D) ($p < 0.05$) and the Drug susceptible vs Drug resistant group (E) (n.s), using the Bray-Curtis dissimilarity index and represented here on a two-dimensional NMDS plots. Circles denote samples from time point 1 and triangles denote samples from time point two. Statistical significance was calculated using Analysis of Similarities statistic (ANOSIM) with 999 permutations and is indicated for each comparison (n.s. = not significant).

Supplementary table 1: Meta data table of patients and controls selected for inclusion in the study

Participant	Sample	Time point	Time between samples (Months)	Age	Sex	HIV status	Culture results	Smear results
Susceptible 1	S1_1	1	9	45	M	Reactive	MTBC - Susceptible	Positive
	S1_2	2					Negative	Negative
Susceptible 2	S2_1	1	9	45	M	Reactive	MTBC - Susceptible	Positive
	S2_2	2					Negative	Negative
Susceptible 3	S3_1	1	7	51	M	Non-reactive	MTBC - Susceptible	Positive
	S3_2	2					Negative	Negative
Drug-resistant 1	DR1_1	1	13	56	M	Reactive	MTBC - ISO-R,RIF - R,ETHA - R,STREP - R,KAN - R	Negative
	DR1_2	2					Negative	Negative
Drug-resistant 2	DR2_1	1	4	36	M	Non-reactive	MTBC - R, - ISO-R, RIF - R	Positive
	DR2_2	2					Negative	Negative
Drug-resistant 3	DR3_1	1	11	51	M	Reactive	MTBC ISO - R,RIF - R,ETHA - R,STREP - R Resistant	Positive
	DR3_2	2					Negative	NEgative
Control 1	M1	1	6	23	M	Negative	NA	NA
	M2	2				Negative	NA	NA
Control 2	F1	1	6	21	F	Negative	NA	NA
	F2	2				Negative	NA	NA

Isoniazid (INH), pyrazinamide (PZA), ethambutol (EMB) and rifampicin (RIF) with the addition of streptomycin (STREP)

Supplementary table 2: Antimicrobial resistance genes present in each sample

Resistance type	C1_1	C1_2	C2_1	C2_2	DR1_1	DR1_2	DR2_1	DR2_2	DR3_1	DR3_2	S1_1	S1_2	S2_1	S2_2	S3_1	S3_2
Antibacterial free fatty acids	FARB	FARB	FARB	FARB	FARB	FARB	-	FARB	FARB	FARB	FARB	FARB	FARB	FARB	FARB	FARB
Aminoglycoside	-	-	-	-	AADE, APH(3")-I, APH(3'')-III, APH(6)-I	AADE, APH(3")-I, APH(3'')-III, APH(6)-I	-	AADE, APH(3'')-III, APH(6)-I	AADE	AADE, APH(3")-I, APH(3'')-I, APH(3'')-III, APH(6)-I	-	-	AAC(2')-I, KSGA	-	APH(3")-I, APH(6)-I, KSGA	AADA, AADE, APH(6)-I
Bacitracin	BACA	BACA	-	-	BACA	BACA	-	-	BACA	BACA	-	-	BACA	-	BACA	-
Beta-lactam	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2, TEM	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, TEM	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2, TEM	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2, TEM	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2, TEM	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2, TEM	MecA/PBP-1A, PBP-2X, CFXA2	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, PENA, CFXA2, TEM	MecA/PBP-1A, PBP-2X, MecB/PBP-1B, TEM
Diaminopyrimidine	-	DFRC	-	-	-	-	-	-	-	-	-	-	-	-	-	DFRC
Fluoroquinolone	PATB	PATB	PATB	PATB	PATB	PATB	PATB	PATB	PATB, PATA	PATB	PATB, MFPA	PATB	PATB	PATB	PATB	PATB, PATA
MLS	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMX, ERMF, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMX, ERMF, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMX, ERMF, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMX, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMF, MEFE, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMF, MEFE, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, MEFE, RLMA(II)	MEL, MACB, ERM(37), RLMA(II)	MEL, MEFA, ERMB, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, LMRC, MEFE, MSRE, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, MEFE, RLMA(II)	MEL, MEFA, LSAC, ERMB, MACB, MACA, ERMC, RLMA(II)	MEFA, LSAC, MACA, ERMX, ERM(37), ERMC, LMRC, MEFE, RLMA(II)
Multidrug	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, MTRD, MDTK	MDTG, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, EMRB, MTRD, MDTK,	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	-	MDTG, MDTK	MDTG, EMRB, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, MTRC, MTRD, MDTK, MTRE ,MTRR	MDTG, MTRC, MTRD, MDTK, MTRE
Nitroimidazole	-	-	-	-	-	-	-	-	-	-	-	-	MSBA	-	-	-
Nucleoside	-	-	-	-	SAT-4	SAT-4	-	SAT-4	-	-	-	-	-	-	SAT-4	-
Peptide	UGD	UGD	-	-	-	-	-	-	UGD, PGPB	-	-	-	-	-	-	-
Phenicol	CATA	-	CATD, CATS	CATD, CATS	CATQ	CATQ	-	-	-	CATQ	-	-	-	-	CATA	-
Sulfonamide	-	-	-	-	SUL2	SUL1, SUL2	-	-	-	SUL2	-	-	-	SUL2	SUL2	SUL1, SUL2
Tetracycline	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TETO, TET32, TETQ	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TET32, TETQ, TET37, TETB	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TET32, TETQ, TET37, TETB	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TET32, TETQ, TETB	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60) ,TETB(60) ,TETA ,TETC ,TETO ,TET32 ,TETQ ,TET37	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TETO, TET32, TETQ , TET37	TETA(46), TETB(46), TETM, TETR, TETA, TETC, TETO, TET32 ,TETQ	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TETO, TET32, TETQ, TET37	TETA(46), TETB(46), TETM, TETW, TETA(60) ,TETB(60), TETO, TET32, TETQ, TET37	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TETO, TET32, TETQ, TET37	TETM, TETA(60) ,TETB(60), TETQ, TET37	TETA(46), TETB(46), TETM, TETW	TETA(46), TETB(46), TETM, TETW, TETA(60), TETB(60), TETQ	TETA(46), TETB(46), TETM, TETW, TETB(60), TETQ, TETL	TETA(46), TETB(46), TETM, TETR, TETW, TETA(60), TETB(60), TETA, TETC, TET32, TETQ, TET37, TETL	TETA(46), TETB(46), TETM, TETR, TETB(60), TETA, TETO, TET32, TETQ, TET37, TETL
Unclassified	CAMP-REGULATORY_PROTEIN, ARLR	CAMP-REGULATORY_PROTEIN, ARLR	CAMP-REGULATORY_PROTEIN	CAMP-REGULATORY_PROTEIN	-	-	-	-	LUXR, PUROMYCIN-RESISTANCE_PROTEIN	-	-	-	ARLR	-	CPXR	CAMP-REGULATORY_PROTEIN, PUROMYCIN-RESISTANCE_PROTEIN

Supplementary table 3: Organisms know to be opportunistic or potential pathogens and the samples they were present in

Opportunistic/potential pathogens	Sample
Actinomyces oris	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_2, DR3_1, DR3_2
Gemella haemolysans	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_1, DR2_2, DR3_1, DR3_2
Haemophilus influenza	S1_1, S2_1
Moraxella nonliquefaciens	S1_1
Neisseria cinerea	S2_1, S2_2, S3_2
Neisseria flavescens	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_1, DR2_2, DR3_1, DR3_2
Neisseria sicca	S2_1, S2_2, S3_1, S3_2, DR1_2, DR2_2, DR3_1
Rothia mucilaginosa	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_1, DR2_2, DR3_1, DR3_2
Parvimonas micra	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR2_2, DR3_1
Prevotella denticola	S1_1, S1_2, S3_1, S3_2, DR1_1, DR1_2, DR2_2, DR3_2
Prevotella melaninogenica	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_1, DR2_2, DR3_1, DR3_2
Streptococcus gordonii	S1_2, S2_2, S3_1, S3_2, DR2_1, DR3_1
Streptococcus mitis	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_1, DR2_2, DR3_1, DR3_2
Tropheryma whipplei	S1_1
Veillonella parvula	S1_1, S1_2, S2_1, S2_2, S3_1, S3_2, DR1_1, DR1_2, DR2_1, DR2_2, DR3_1, DR3_2

Chapter 6 – Summary and future prospects

This chapter starts with a review of the objectives of the studies combined in this thesis, followed by the central findings in each of the studies, and finally highlights potential areas for future research.

1. Introduction

Globally tuberculosis remains one of the leading causes of death from any infectious disease among adults, with millions becoming newly infected each year (WHO, 2020). In recent times, evidence has begun to surface surrounding the link between TB and the microbiome and how shifts in the microbiome in response to TB infection and subsequent treatment may have a profound impact on both treatment outcome, and the long term health of patients. (Cui *et al.,* 2012; Cheung *et al.,* 2013; Botero *et al.,* 2014; Krishna *et al.,* 2016; Luo *et al.,* 2017; Wipperman *et al.,* 2017; Maji *et al.,* 2018; Hu *et al.,* 2019; Hu *et al.,* 2020; Namasivayam *et al*., 2020; Wang *et al*., 2020).

To date only 21 studies have been published regarding TB and the resistome, 7 of which were in animal models, 7 focussed on the gut microbiome of human participants and the final 7 focussed on the respiratory tract of human participants. However none of the respiratory tract studies have elucidated the impact of drug resistant TB treatment on the microbiome, and neither have any of the studies attempted to evaluate the overall changes in the accompanying resistome (Luo *et al.,* 2017; Wipperman *et al.,* 2017; Maji *et al.,* 2018; Hu *et al.,* 2019; Namasivayam *et al*., 2020; Wang *et al*., 2020).

In addition to the gaps in research regarding TB and the microbiome, it is a well-known fact that the mycobacterial bacillus is encompassed by an extraordinarily elaborate cell wall structure. This makes the organisms incredibly hard to lyse and several studies have demonstrated that optimal DNA extraction methods are required to carry out studies with regards to TB (de Almeida *et al.,* 2013; Kolia-Diafouka *et al.,* 2018). As such, the research community is constantly in need of new and more robust methods to conduct studies on both TB and the associated microbiome. The NWU TB diagnostic method (NWU-TB test) has a very robust cell lysis step and was thus selected for evaluation during the course of this study to perform TB and microbiome associated studies (Mutingwende *et al*., 2015; Sulis *et al.,* 2016).

The aims and objectives for the study were as follows:

- Comparative methods: To determine the suitability of DNA extracted by means of the NWU lysis method for bacterial microbiome analysis and amplicon-based next generation sequencing (NGS) in comparison to DNA obtained by various commercial kits;

- To examine changes in the bacterial microbiome of the respiratory tract during the process of TB therapy over time in human participants (sputum) by whole genome shotgun sequencing.

2. Summary of study design, execution and results

2.1 Objective 1

Applying the NWU lysis method to a mock microbial community, followed by sequencing on the Oxford Nanopore Technologies (ONT) MinION Platform did not reveal any conclusive results, but did lay the ground work for a proper evaluation of the technology. The works described in Chapter 3 revealed that the use of mock microbial communities for assessments of lysis efficiency has several limitations, while also revealing some limitations regarding the extraction kits themselves. Mock microbial communities like the ZymoBIOMICS™ Microbial Community Standard (Zymo, USA) often come stored in a special protective inactivation media for ease of use and transport. Unfortunately this means cell integrity is not always maintained. This is especially problematic when incorporated with some DNA kits that include an additional step for the selective lysis of host cells. As demonstrated in Chapter 3, the host DNA removal steps of the QIAamp DNA microbiome kit (Qiagen, Germany), completely removed all Gram-negative bacterial DNA from the samples during the extraction process. In addition to this the works done during Chapter 3 revealed that more tools are urgently needed to assess long read 16S sequences generated by the ONT MinION. At the time of writing the current native workflow, although easy to use, only allows the identification of the bacterial genera present in a sample, but does not

offer any means by which to estimate relative abundance or any means to do further downstream statistical analyses. Thus as part of this Chapter a bioinformatics workflow was designed and validated, using the ZymoBIOMICS™ Microbial Community Standard, as it is a sample of known composition. The study revealed that the customized workflow turned out to be more accurate in identifying the bacterial species present and had fewer miss-classifications, due to the fact that a stricter identity cut-off could be applied, an option missing from the current native ONT workflow. This workflow would then be applied in the following Chapter 4, for the assessment of sequencing reads generated from a more complex clinical sputum sample sequenced on the ONT MinION.

Chapter 4 continued with evaluations of the NWU lysis method, this time with application to more complex samples, and included the results from two studies: one data set originating from this PhD and the other dataset from a concurrently running MSc, co-supervised as part of this PhD and focussing on TB/microbiome related research. Samples included clinical sputum samples and gut samples from a concurrently running study, representing less and more complex microbiome samples respectively. The final results were compounded into the manuscript presented in Chapter 4. Results from Chapter 4 revealed that the NWU lysis system is a promising prospect for microbiome studies due to its lysis efficiency, but only if incorporated with short read sequencing. The lysis method is very robust, as its initial purpose was the complete lysis of hard to lyse mycobacterial cells. Due to the robust lysis process, the DNA produced is of a sheared nature. This does not seem to impact shorter sequencing reads of the 16S gene used for sequencing on platforms like the Illumina Miseq, but it is troublesome when attempting to incorporate it with technology like the ONT MinION. The primers targeting the 16S gene in this instance flank the entire gene, and if the DNA is of a highly sheared nature, it is very likely that bias may be introduced during PCR prior to sequencing if full length DNA fragments are to be amplified. To conclude, Chapters 3 and 4 addresses the first objective successfully. The work done demonstrates that the NWU lysis method has potential for downstream microbiome analyses but further optimisation will be required.

2.2 Objective 2

The second objective was to examine changes in the bacterial microbiome of the respiratory tract during the process of TB therapy over time in the sputum human participants by whole genome shotgun sequencing. Initial plans included the recruitment of patients from the Potchefstroom Hospital, North-West Province, South Africa. Unfortunately, due several delays relating to the initial ethics applications, the Covid-19 crisis and other difficulties in obtaining samples, a different approach had to be undertaken to obtain samples. Luckily, with permission and aid from HANKS TB Diagnostics (Pty) Ltd, samples, were identified from their TB biobank that would be suitable to complete the second objective. A total of 12 samples from 6 patients were included in the study, 3 testing positive for drug susceptible TB (labelled as Susceptible or S) and 3 testing positive for drug resistant TB (labelled as Drug-resistant or DR). To generate a baseline for comparison, samples from individuals acting as healthy controls were taken in the form of oropharyngeal swabs. The study, a first of its kind, aimed to characterise changes in the respiratory microbiome in response to drug resistant tuberculosis therapy, and also to attempt to characterise the associated resistome. The results as shown in Chapter 5 revealed significant changes in both taxonomic composition and within the accompanying resistome. Alpha diversity metrics did reveal a trend where overall microbiome diversity was increased in response to treatment. This is in contradiction to several gut studies where microbial diversity was found to be significantly reduced, but additional studies on a larger and more diverse cohort of patients will be needed to determine whether the TB treatment is the main driving force behind this phenomenon. Regardless, overall evaluations of taxonomic composition agrees with previous studies that have noted that antimicrobial treatment may give foreign or less abundant, potentially more resilient bacteria a chance to proliferate. This occurrence is even more likely if the drugs in question deplete the more susceptible healthy commensals. Additionally the study also provided the first evidence of a complex resistome associated with TB patients, which is in some instances significantly different from that of healthy controls. The most important observation regarding the resistome was the clear influx of acquired aminoglycoside resistance genes (*aadA, aadE, aac(2')-I, aph(3')-I, aph(3'')-I, aph(3''')-III*, and *aph(6)-I*) associated with patient samples. Considering the high prevalence of HIV among

individuals receiving treatment for tuberculosis and drug resistant tuberculosis, these findings may be a cause for concern, but further studies are required to determine if there is any physiological or clinical significance to these initial observations. Nonetheless the results aided in the completion of objective two and in the investigation of changes in the bacterial microbiome of the respiratory tract during the process of TB therapy over time in human participants (sputum) by whole genome shotgun sequencing.

3. Future prospects

In relation to the completed objectives, several future prospects have been identified. The areas for further research include the following:

1. In its current form, the DNA produced with the NWU lysis method is too sheared, and further optimisation will be required before it can be incorporated with portable long read sequencing devices like the ONT MinION. Such research would include:
    a. Optimisation of the lysis conditions to produce intact high molecular weight DNA; and
    b. Optimisation of the lysis buffer to provide additional protection and stabilisation of the DNA during the lysis process.

2. In relation to TB and the microbiome within the context of available literature (Chapter 2), as well as the results from the current study, several key points have been identified that require further attention:
    a. Evaluation on a larger, more diverse sample cohort to properly determine whether the increase in diversity observed is primarily related to the administration of TB drugs or also connected to the HIV status of individuals; and;
    b. to determine whether the changes in the resistome observed, especially the influx of aminoglycoside resistance genes, is a frequent occurrence across a larger sample group; and finally;

c. to determine whether changes in the resistome persist long after treatment and whether there are any long term health implications;
d. the results also open the door to future prospective studies, to determine if treatment regimens designed with the aim of causing a lesser degree of dysbiosis, or treatment regimens supplemented with probiotic beneficial bacteria, could lead to improved treatment outcomes by maintaining the balance and presence of beneficial host flora.

Annexure - A

DATA IN BRIEF

AUTHOR INFORMATION PACK

TABLE OF CONTENTS

ISSN: 2352-3409

DESCRIPTION

Data in Brief provides a way for researchers to easily share and reuse each other's datasets by publishing data articles that:

Thoroughly describe your data, facilitating reproducibility. Make your data, which is often buried in supplementary material, easier to find. Increase traffic towards associated research articles and data, leading to more citations. Open up doors for new collaborations.

Because you never know what data will be useful to someone else, *Data in Brief* welcomes submissions that describe data from all research areas.Please note: almost any piece of information can be defined as data. However, to merit publication in *Data in Brief* data should be a set of information that are acquired/collected with a scientific method and have value to the research community. Please see 'What data are suitable for Data in Brief?' for more details.

Authors submitting to *Data in Brief* should use this template to write their manuscript.

Data in Brief has a CiteScore of 1.5 and is ranked 32nd of 111 (Q2) in the 'Multidisciplinary' subject category.

ABSTRACTING AND INDEXING

PubMed Central
PubMed/Medline
Scopus
Emerging Sources Citation Index (ESCI)
Directory of Open Access Journals (DOAJ)

EDITORIAL BOARD

Section Editors

Agricultural Science
Magda Pál, Hungarian Academy of Sciences Centre for Agricultural Research, Martonvasar, Hungary

Artificial Intelligence
Benjamin Quost, University of Applied Sciences for Technology Compiegne, Compiègne, France

Biochemistry and Immunology
Nicholas A. Pullen, University of Northern Colorado, Greeley, Colorado, United States

Bioinformatics, Data Science
Vivek Kumar, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States

Business, Management and Accounting
Noemi Sinkovics, The University of Auckland Business School, Auckland, New Zealand

Chemistry
Paolo Bertoncello, Swansea University, Swansea, United Kingdom

Civil Engineering
Stamatis Zoras, University of Derby, Derby, United Kingdom

Computer Science
Niko Lukač, University of Maribor, Maribor, Slovenia

Earth and Planetary Sciences
Chaosheng Zhang, National University of Ireland Galway, Galway, Ireland

Engineering
Maria Grazia De Giorgi, University of Salento, Lecce, Italy

Economics
Larisa Yarovaya, University of Southampton, Southampton, United Kingdom

Energy
Giacomo Salvadori, University of Pisa, Department of Energy, Systems, Territory and Construction Engineering, Pisa, Italy

Environmental Science: Environmental Chemistry and Hydrology
Georgios Bartzas, National Technical University of Athens - Zografou Campus, Zografos, Greece

Environmental Science: Waste Management and Pollution
Yolanda Pico, University of Valencia, Valencia, Spain

Genetics, Genomics and Molecular Biology
Giulio Chiesa, Boston University, Boston, Massachusetts, United States of America

Materials Science: Chemistry
Yibin Xu, National Institute for Materials Science, Tsukuba-Shi, Japan

Materials Science: Materials Informatics
Taylor D. Sparks, The University of Utah, Salt Lake City, Utah, United States

Medicine and Pharmacology
Nektarios Barabutis, The University of Louisiana Monroe College of Pharmacy, Monroe, Louisiana, United States

Medicine and Public Health
Iain Brownlee, Northumbria University Faculty of Health and Life Sciences, Newcastle Upon Tyne, United Kingdom

Microbiology
Rasha Maal-Bared, EPCOR, Edmonton, Alberta, Canada

Neuroscience
Joseph D. Zak, Harvard University, Cambridge, Massachusetts, United States

Pharmacology, Toxicology, Pharmaceutical Sciences
Luisa Camacho, US Food and Drug Administration, Jefferson, Arkansas, United States

Archan Ganguly, University of California San Diego Department of Cellular and Molecular Medicine, La Jolla, California, United States
Vicente García Díaz, University of Oviedo, Oviedo, Spain
Maria Garcia-Dominguez, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, United States
Giulia Giubertoni, University of Amsterdam, Amsterdam, Netherlands
Daren Gooddy, British Geological Survey - Wallingford Office, Wallingford, United Kingdom
Stella Gryllia, Leiden University Medical Center, Leiden, Netherlands
Jyotasana Gulati, Alkahest Inc, San Carlos, California, United States
Hamad K. Hadrawi, University of Kufa, Kufa, Iraq
Fei Han, Shenyang Pharmaceutical University, Shenyang, China
Sherif Hassan, Czech University of Life Sciences Prague, Praha, Czech Republic
Andres Sandoval Hernandez, University of Bath, Bath, United Kingdom
David Hernandez, NEW YORK UNIVERSITY SCHOOL OF MEDICINE, New York, New York, United States
Tao Huan, The University of British Columbia, Vancouver, British Columbia, Canada
Tanya Hundal, Mayo Foundation for Medical Education and Research, Rochester, Minnesota, United States of America
José Francisco Islas Cisneros, Autonomous University of Nuevo Leon, San Nicolas De Los Garza, Mexico
Ravirajsinh Jadeja, The Medical College of Georgia, Georgia, United States of America
Begoña Jiménez, Spanish Scientific Research Council, Madrid, Spain
Erin Kane, Boston University, Boston, Massachusetts, United States of America
Kurunthachalam Kannan, New York University Department of Pediatrics, New York, New York, United States
Ramakrishnan Kannan, Yale University, New Haven, Connecticut, United States of America
Charalambos Karagiannidis, University of Thessaly, Volos, Greece
Nerantzis Kazakis, Aristotle University of Thessaloniki, Thessaloniki, Greece
John F. Kennedy, Chembiotech Laboratories Ltd, Tenbury Wells, United Kingdom
Nima Khakzad, Ryerson University, Toronto, Ontario, Canada
Kaido Kikkas, Tallinn University of Technology, Tallinn, Estonia
Ki-Hyun Kim, Hanyang University College of Engineering Department of Civil and Environmental Engineering, Seongdong-gu, Korea, Republic of
Giorgia La Barbera, University of Copenhagen, Copenhagen, Denmark
Joseph Yui-yip Lau, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
Francisco Javier Lena Acebo, University of Cantabria, Department of Business Administration, Santander, Spain
Bo Li, Tsinghua University, Beijing, China
Yu Li, South China Normal University, Guangzhou, China
Heng Liang, Harbin Institute of Technology, School of Environment, Harbin, China
Rock Keey Liew, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia
Fuchen Liu, Yale University School of Medicine, New Haven, Connecticut, United States
Ronald Machaka, CSIR South Africa, Mining, Manufacturing, Defence & Security Division, Pretoria, South Africa
Howard I. Maibach, University of California San Francisco Department of Dermatology, San Francisco, California, United States of America
Manousos Makridakis, Biomedical Research Foundation of the Academy of Athens, Athens, Greece
Krystyna Malińska, Czestochowa University of Technology, Czestochowa, Poland
Maria Malvoni, National Technical University of Athens School of Electrical and Computer Engineering, ATHINA, Greece
Valeria Marasco, University of Vienna, Wien, Austria
Lorenzo Mari, Polytechnic of Milan, Milano, Italy
Francesco Marinello, University of Padua Department of Land Environment Agriculture and Forestry, Legnaro PD, Italy
Mauro Masiol, Ca' Foscari University of Venice, Venezia, Italy
Khairul Mastor, National University of Malaysia, Bangi, Malaysia
Janine McCartney, HHC Services Inc, Lester, Pennsylvania, United States
Jun Mei, Queensland University of Technology, Brisbane, Queensland, Australia
Weizhi Meng, Technical University of Denmark Department of Applied Mathematics and Computer Science, Kgs Lyngby, Denmark
Lukasz Migas, Delft University of Technology Delft Centre for Systems and Control, Delft, Netherlands
Eiko N. Minakawa, National Center of Neurology and Psychiatry National Institute of Neuroscience, Tokyo, Japan
Hyo-Bang Moon, Hanyang University Department of Marine Science and Convergence Engineering, Ansan, Korea, Republic of
Sébastien Mouchet, University of Exeter, Exeter, United Kingdom
Subhadip Mukhopadhyay, NYU LANGONE MEDICAL CENTER, New York, New York, United States
Rabindra Nepal, University of Wollongong School of Accounting Economics and Finance, Wollongong, Australia
Georgios Nikolopoulos, National Public Health Organization, Athens, Greece
Lokman Nor Hakim Norazmi, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia
Michael D. O'Toole, The University of Manchester, Manchester, United Kingdom
Shany Ofaim, Northeastern University, Boston, Massachusetts, United States

GUIDE FOR AUTHORS

INTRODUCTION

Data in Brief only publishes **Data in Brief** data articles that describe research data you've made publicly available either through a repository or directly in your Data in Brief article. These articles ensure that your data, which is often buried in supplmentary material, is actively peer-reviewed, curated, formatted, indexed, given a DOI and publicly available to all upon publication. All Data in Brief submissions must contain data produced and owned by either the author or the author's institution. Data in Brief will not accept any submissions that contain datasets with scant variables or samples.

Article Structure
We ask all authors to use our Data in Brief template when preparing their manuscript. All Data in Brief articles must contain the following sections:

Title, authors and author affiliations Abstract A completed Specifications Table 3-5 bullet points that explain why the data is valuable to other researchers A direct link to the deposited data (when applicable) A complete description of the experimental design and materials and methods References

For a quick, easy guide see our Data in Brief at a Glance

Submit your own Data in Brief
To submit a Data in Brief article directly to Data in Brief: Fill in this template. Submit the document along with any computer code used to analyze the data through the journal's online submission system. Important: choose 'Data Article' as the article type when submitting directly, even if your data article is linked to another publication.

OR to submit via another Elsevier journal (a Co-submission):

Fill in this template Submit your Data in Brief article, alongside your original research article to your Elsevier journal of choice. Upon acceptance of your article, your Data in Brief article will be sent to Data in Brief for editorial review and publication.

This step by step video guide will also tell you how to complete the template correctly to maximise your chances of acceptance.

Algal Announcements Article Collection

Data in Brief is partnering with *Algal Research* and *Harmful Algae* for data announcements to the algal research community including **announcements of Harmful Algae Blooms (HABs)** as well as **announcements of new algal strains**. The objective is to provide freely available information that will benefit researchers, policymakers, entrepreneurs, water managers, engineers and other readers. As this is a different type of data article, a new article type has been created called "Algal Announcement" and it is intended to communicate the following: New algae strains (including macroalgae, microalgae, and cyanobacteria) and their biological and biochemical characteristics, as well as their isolation or collection, identification, culturing, and molecular profiles. These may include new isolates from prospecting studies or from environments where the algae were found/isolated in conjunction with other organisms. Harmful algal blooms announcements, including location, the number of samplings, the date(s) of collection, water quality data and description of toxins.

Authors who wish to submit an Algal Announcement article should use this template to prepare their manuscript and select the article type "Algal Announcement" article type when submitting their manuscript.

Submission Checklist
You can use this list to carry out a final check of your submission before you send it to the journal for review. Please check the relevant section in this Guide for Authors for more details.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:

• E-mail address
• Full postal address

All necessary files have been uploaded:
Manuscript:
• Include keywords
• All figures (include relevant captions)
• All tables (including titles, description, footnotes)
• Ensure all figure and table citations in the text match the files provided
• Indicate clearly if color should be used for any figures in print
where applicable)
Supplemental files (where applicable)

Further considerations
• Manuscript has been 'spell checked' and 'grammar checked'
• All references mentioned in the Reference List are cited in the text, and vice versa
• Permission has been obtained for use of copyrighted material from other sources (including the Internet)
• A competing interests statement is provided, even if the authors have no competing interests to declare
• Journal policies detailed in this guide have been reviewed
• Referee suggestions and contact details provided, based on journal requirements

For further information, visit our Support Center.

BEFORE YOU BEGIN

The *Data in Brief* peer review process

Because Data in Brief articles are pure descriptions of data they are reviewed differently than a typical research article. The Data in Brief peer review process focuses on data transparency.

Reviewers review manuscripts based on the following criteria:
Do the description and data make sense? Do the authors adequately explain its utility to the community? Are the protocol/references for generating data adequate? Data format (is it standard? potentially re-usable?) Does the article follow the Data in Brief template? Is the data well documented?

Data in Brief that are converted supplementary files submitted alongside a research article via another Elsevier journal are editorially reviewed by the Data in Brief editor. In these cases, the supporting data has already been peer reviewed as a part of the research article peer review process. The Data in Brief editor then performs a final review on the Data in Brief article using the same criteria mentioned above before the Data in Brief may be accepted for publication.

Ethics in publishing

Please see our information pages on Ethics in publishing and Ethical guidelines for journal publication.

Studies in humans and animals

If the work involves the use of human subjects, the author should ensure that the work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The manuscript should be in line with the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals and aim for the inclusion of representative human populations (sex, age and ethnicity) as per those recommendations. The terms sex and gender should be used correctly.

Authors should include a statement in the manuscript that informed consent was obtained for experimentation with human subjects. The privacy rights of human subjects must always be observed.

All animal experiments should comply with the ARRIVE guidelines and should be carried out in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, or the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978) and the authors should clearly indicate in the manuscript that such guidelines have been followed. The sex of animals must be indicated, and where appropriate, the influence (or association) of sex on the results of the study.

Data gathered using social media

Using data from social media platforms (e.g. Twitter, Facebook, etc.) poises ethical concerns regarding consent. **Firstly**, authors need to determine if the data collection needs ethical approval from appropriate institutional review boards and ethics committees. Particular attention is needed to determine if informed consent from possible identifiable participants is needed or to ensure that these are properly anonymized. **Secondly**, authors need to consider the **data redistribution policies** from the platform used to gather information and ensure they have the right to share these data. Policies vary among platforms and are updated regularly therefore it is the author?s responsibility to ensure that their work complies with platforms? policies.

Declaration of interest

All authors must disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work. Examples of potential competing interests include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding. Authors should complete the declaration of interest statement during the submission process under Additional Questionnaire. This statement will be published within the article if accepted.

Submission declaration and verification

Submission of an article implies that the work described has not been published previously (except in the form of an abstract, a published lecture or academic thesis, see 'Multiple, redundant or concurrent publication' for more information), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. To verify originality, your article may be checked by the originality detection service Crossref Similarity Check.

Preprints

Please note that preprints can be shared anywhere at any time, in line with Elsevier's sharing policy. Sharing your preprints e.g. on a preprint server will not count as prior publication (see 'Multiple, redundant or concurrent publication' for more information).

Use of inclusive language

Inclusive language acknowledges diversity, conveys respect to all people, is sensitive to differences, and promotes equal opportunities. Content should make no assumptions about the beliefs or commitments of any reader; contain nothing which might imply that one individual is superior to another on the grounds of age, gender, race, ethnicity, culture, sexual orientation, disability or health condition; and use inclusive language throughout. Authors should ensure that writing is free from bias, stereotypes, slang, reference to dominant culture and/or cultural assumptions. We advise to seek gender neutrality by using plural nouns ("clinicians, patients/clients") as default/wherever possible to avoid using "he, she," or "he/she." We recommend avoiding the use of descriptors that refer to personal attributes such as age, gender, race, ethnicity, culture, sexual orientation, disability or health condition unless they are relevant and valid. These guidelines are meant as a point of reference to help identify appropriate language but are by no means exhaustive or definitive.

Author Contributions

For transparency, authors need to submit an author statement outlining their individual contributions to the paper using the relevant CRediT roles: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing. Authorship statements should be formatted with the names of authors first and CRediT role(s) following. More details and an example

Authorship

All authors should have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition of data, or analysis of data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version to be submitted.

Changes to authorship

Authors are expected to consider carefully the list and order of authors **before** submitting their manuscript and provide the definitive list of authors at the time of the original submission. Any addition, deletion or rearrangement of author names in the authorship list should be made only **before** the manuscript has been accepted and only if approved by the journal Editor. To request such a change, the Editor must receive the following from the **corresponding author**: (a) the reason for the change in author list and (b) written confirmation (e-mail, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed.

Only in exceptional circumstances will the Editor consider the addition, deletion or rearrangement of authors **after** the manuscript has been accepted. While the Editor considers the request, publication of the manuscript will be suspended. If the manuscript has already been published in an online issue, any requests approved by the Editor will result in a corrigendum.

Copyright

Upon acceptance of an article, authors will be asked to complete a 'License Agreement' (see more information on this). Permitted third party reuse of open access articles is determined by the author's choice of user license.

Author rights

As an author you (or your employer or institution) have certain rights to reuse your work. More information.

Elsevier supports responsible sharing

Find out how you can share your research published in Elsevier journals.

Role of the funding source

You are requested to identify who provided financial support for the conduct of the research and/or preparation of the article and to briefly describe the role of the sponsor(s), if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. If the funding source(s) had no such involvement then this should be stated.

Open access

Please visit our Open Access page for more information.

Elsevier Researcher Academy

Researcher Academy is a free e-learning platform designed to support early and mid-career researchers throughout their research journey. The "Learn" environment at Researcher Academy offers several interactive modules, webinars, downloadable guides and resources to guide you through the process of writing for research and going through peer review. Feel free to use these free resources to improve your submission and navigate the publication process with ease.

Language (usage and editing services)

Please write your text in good English (American or British usage is accepted, but not a mixture of these). Authors who feel their English language manuscript may require editing to eliminate possible grammatical or spelling errors and to conform to correct scientific English may wish to use the English Language Editing service available from Elsevier's Author Services.

Submission

Our online submission system guides you stepwise through the process of entering your article details and uploading your files. The system converts your article files to a single PDF file used in the peer-review process. Editable files (e.g., Word, LaTeX) are required to typeset your article for final publication. All correspondence, including notification of the Editor's decision and requests for revision, is sent by e-mail.

Submit your article

Please submit your article via https://www.editorialmanager.com/dib.

Referees

Please submit the names and institutional e-mail addresses of several potential referees. For more details, visit our Support site. Note that the editor retains the sole right to decide whether or not the suggested reviewers are used.

PREPARATION

Peer review

This journal operates a single anonymized review process. All contributions are typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. Editors are not involved in decisions about papers which they have written themselves or have been written by family members or colleagues or which relate to products or services in which the editor has an interest. Any such submission is subject to all of the journal's usual procedures, with peer review handled independently of the relevant editor and their research groups. More information on types of peer review.

Use of word processing software

It is important that the file be saved in the native format of the word processor used. The text should be in single-column format. Keep the layout of the text as simple as possible. Most formatting codes will be removed and replaced on processing the article. In particular, do not use the word processor's options to justify text or to hyphenate words. However, do use bold face, italics, subscripts, superscripts etc. When preparing tables, if you are using a table grid, use only one grid for each individual table and not a grid for each row. If no grid is used, use tabs, not spaces, to align columns. The electronic text should be prepared in a way very similar to that of conventional manuscripts (see also the Guide to Publishing with Elsevier). Note that source files of figures, tables and text graphics will be required whether or not you embed your figures in the text. See also the section on Electronic artwork.

To avoid unnecessary errors you are strongly advised to use the 'spell-check' and 'grammar-check' functions of your word processor.

Graphical abstract

Although a graphical abstract is optional, its use is encouraged as it draws more attention to the online article. The graphical abstract should summarize the contents of the article in a concise, pictorial form designed to capture the attention of a wide readership. Graphical abstracts should be submitted as a separate file in the online submission system. Image size: Please provide an image with a minimum of 531 × 1328 pixels (h × w) or proportionally more. The image should be readable at a size of 5 × 13 cm using a regular screen resolution of 96 dpi. Preferred file types: TIFF, EPS, PDF or MS Office files. You can view Example Graphical Abstracts on our information site.

Authors can make use of Elsevier's Illustration Services to ensure the best presentation of their images and in accordance with all technical requirements.

Abbreviations

Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention there, as well as in the footnote. Ensure consistency of abbreviations throughout the article.

Acknowledgements

Collate acknowledgements in a separate section at the end of the article before the references and do not, therefore, include them on the title page, as a footnote to the title or otherwise. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

Formatting of funding sources

List funding sources in this standard way to facilitate compliance to funder's requirements:

Funding: This work was supported by the National Institutes of Health [grant numbers xxxx, yyyy]; the Bill & Melinda Gates Foundation, Seattle, WA [grant number zzzz]; and the United States Institutes of Peace [grant number aaaa].

It is not necessary to include detailed descriptions on the program or type of grants and awards. When funding is from a block grant or other resources available to a university, college, or other research institution, submit the name of the institute or organization that provided the funding.

If no funding has been provided for the research, please include the following sentence:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Units
Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI.

Math formulae
Please submit math equations as editable text and not as images. Present simple formulae in line with normal text where possible and use the solidus (/) instead of a horizontal line for small fractional terms, e.g., X/Y. In principle, variables are to be presented in italics. Powers of e are often more conveniently denoted by exp. Number consecutively any equations that have to be displayed separately from the text (if referred to explicitly in the text).

Footnotes
Footnotes should be used sparingly. Number them consecutively throughout the article. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes in the text and present the footnotes themselves separately at the end of the article.

Artwork

Electronic artwork
General points
• Make sure you use uniform lettering and sizing of your original artwork.
• Embed the used fonts if the application provides that option.
• Aim to use the following fonts in your illustrations: Arial, Courier, Times New Roman, Symbol, or use fonts that look similar.
• Number the illustrations according to their sequence in the text.
• Use a logical naming convention for your artwork files.
• Provide captions to illustrations separately.
• Size the illustrations close to the desired dimensions of the published version.
• Submit each illustration as a separate file.
• Ensure that color images are accessible to all, including those with impaired color vision.

A detailed guide on electronic artwork is available.
You are urged to visit this site; some excerpts from the detailed information are given here.
Formats
If your electronic artwork is created in a Microsoft Office application (Word, PowerPoint, Excel) then please supply 'as is' in the native document format.
Regardless of the application used other than Microsoft Office, when your electronic artwork is finalized, please 'Save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):
EPS (or PDF): Vector drawings, embed all used fonts.
TIFF (or JPEG): Color or grayscale photographs (halftones), keep to a minimum of 300 dpi.
TIFF (or JPEG): Bitmapped (pure black & white pixels) line drawings, keep to a minimum of 1000 dpi.
TIFF (or JPEG): Combinations bitmapped line/half-tone (color or grayscale), keep to a minimum of 500 dpi.
Please do not:
• Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); these typically have a low number of pixels and limited set of colors;
• Supply files that are too low in resolution;
• Submit graphics that are disproportionately large for the content.

Color artwork
Please make sure that artwork files are in an acceptable format (TIFF (or JPEG), EPS (or PDF) or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color online (e.g., ScienceDirect and other sites). Further information on the preparation of electronic artwork.

Illustration services
Elsevier's Author Services offers Illustration Services to authors preparing to submit a manuscript but concerned about the quality of the images accompanying their article. Elsevier's expert illustrators can produce scientific, technical and medical-style images, as well as a full range of charts, tables and graphs. Image 'polishing' is also available, where our illustrators take your image(s) and improve them to a professional standard. Please visit the website to find out more.

Figure captions
Ensure that each illustration has a caption. Supply captions separately, not attached to the figure. A caption should comprise a brief title (**not** on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

Tables

Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place any table notes below the table body. Be sparing in the use of tables and ensure that the data presented in them do not duplicate results described elsewhere in the article. Please avoid using vertical rules and shading in table cells.

References

Citation in text
Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Reference links
Increased discoverability of research and high quality peer review are ensured by online links to the sources cited. In order to allow us to create links to abstracting and indexing services, such as Scopus, CrossRef and PubMed, please ensure that data provided in the references are correct. Please note that incorrect surnames, journal/book titles, publication year and pagination may prevent link creation. When copying references, please be careful as they may already contain errors. Use of the DOI is highly encouraged.

A DOI is guaranteed never to change, so you can use it as a permanent link to any electronic article. An example of a citation using DOI for an article not yet in an issue is: VanDecar J.C., Russo R.M., James D.E., Ambeh W.B., Franke M. (2003). Aseismic continuation of the Lesser Antilles slab beneath northeastern Venezuela. Journal of Geophysical Research, https://doi.org/10.1029/2001JB000884. Please note the format of such citations should be in the same style as all other references in the paper.

Web references
As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

Data references
This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

References in a special issue
Please ensure that the words 'this issue' are added to any references in the list (and any citations in the text) to other articles in the same Special Issue.

Reference management software
Most Elsevier journals have their reference template available in many of the most popular reference management software products. These include all products that support Citation Style Language styles, such as Mendeley. Using citation plug-ins from these products, authors only need to select the appropriate journal template when preparing their article, after which citations and bibliographies will be automatically formatted in the journal's style. If no template is yet available for this journal, please follow the format of the sample references and citations as shown in this Guide. If you use reference management software, please ensure that you remove all field codes before submitting the electronic manuscript. More information on how to remove field codes from different reference management software.

Users of Mendeley Desktop can easily install the reference style for this journal by clicking the following link:
http://open.mendeley.com/use-citation-style/data-in-brief
When preparing your manuscript, you will then be able to select this style using the Mendeley plug-ins for Microsoft Word or LibreOffice.

Reference style
Text: Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given.
Example: '..... as demonstrated [3,6]. Barnaby and Jones [8] obtained a different result ....'
List: Number the references (numbers in square brackets) in the list in the order in which they appear in the text.
Examples:
Reference to a journal publication:
[1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2010) 51–59. https://doi.org/10.1016/j.Sc.2010.00372.
Reference to a journal publication with an article number:
[2] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, 2018. The art of writing a scientific article. Heliyon. 19, e00205. https://doi.org/10.1016/j.heliyon.2018.e00205.
Reference to a book:
[3] W. Strunk Jr., E.B. White, The Elements of Style, fourth ed., Longman, New York, 2000.
Reference to a chapter in an edited book:
[4] G.R. Mettam, L.B. Adams, How to prepare an electronic version of your article, in: B.S. Jones, R.Z. Smith (Eds.), Introduction to the Electronic Age, E-Publishing Inc., New York, 2009, pp. 281–304.
Reference to a website:
[5] Cancer Research UK, Cancer statistics reports for the UK. http://www.cancerresearchuk.org/aboutcancer/statistics/cancerstatsreport/, 2003 (accessed 13 March 2003).
Reference to a dataset:
[dataset] [6] M. Oguro, S. Imahiro, S. Saito, T. Nakashizuka, Mortality data for Japanese oak wilt disease and surrounding forest compositions, Mendeley Data, v1, 2015. https://doi.org/10.17632/xwj98nb39r.1.

Journal abbreviations source
Journal names should be abbreviated according to the List of Title Word Abbreviations.

Video data

Elsevier accepts video material and animation sequences as a part of your data article. Authors who would like to describe video or animation files in their data article are strongly encouraged to include links to these within the body of the article. This can be done in the same way as a figure or table by referring to the video or animation content and noting in the body text where it should be placed. All submitted files should be properly labeled so that they directly relate to the video file's content. In order to ensure that your video or animation material is directly usable, please provide the files in one of our recommended file formats with a preferred maximum size of 150 MB. Video and animation files supplied will be published online in the electronic version of your article in Elsevier Web products, including ScienceDirect http://www.sciencedirect.com. Please supply 'stills' with your files: you can choose any frame from the video or animation or make a separate image. These will be used instead of standard icons and will personalize the link to your video data. For more detailed instructions please visit our video instruction pages at https://www.elsevier.com/artworkinstructions. Note: since video and animation cannot be embedded in the .pdf version of the article, please provide text for both the electronic (.html) and the .pdf version for the portions of the article that refer to this content.

Supplementary material

Authors may either upload their data into an appropriate data repository, or publish it as supplementary material alongside their data article. Supplementary files supplied will be published online alongside the electronic version of your article in Elsevier Web products, including ScienceDirect: http://www.sciencedirect.com. In order to ensure that your submitted material is directly usable, please provide the data in one of our recommended file formats, or ensure that, if it is in another format, the file(s) is/are bundled into a .zip file. Authors should submit the material in electronic format together with the article and supply a concise and descriptive caption for each file. For more detailed instructions please visit our artwork instruction pages at https://www.elsevier.com/artworkinstructions.

For data submitted as supplementary material, there is a limit 500MB. Special accommodations can be made for larger files, on a case-by-case basis.

Data linking
If you have made your research data available in a data repository, you can link your article directly to the dataset. Elsevier collaborates with a number of repositories to link articles on ScienceDirect with relevant repositories, giving readers access to underlying data that gives them a better understanding of the research described.

There are different ways to link your datasets to your article. When available, you can directly link your dataset to your article by providing the relevant information in the submission system. For more information, visit the database linking page.

For supported data repositories a repository banner will automatically appear next to your published article on ScienceDirect.

In addition, you can link to relevant data or entities through identifiers within the text of your manuscript, using the following format: Database: xxxx (e.g., TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN).

AFTER ACCEPTANCE

Online proof correction

To ensure a fast publication process of the article, we kindly ask authors to provide us with their proof corrections within two days. Corresponding authors will receive an e-mail with a link to our online proofing system, allowing annotation and correction of proofs online. The environment is similar to MS Word: in addition to editing text, you can also comment on figures/tables and answer questions from the Copy Editor. Web-based proofing provides a faster and less error-prone process by allowing you to directly type your corrections, eliminating the potential introduction of errors.
If preferred, you can still choose to annotate and upload your edits on the PDF version. All instructions for proofing will be given in the e-mail we send to authors, including alternative methods to the online version and PDF.
We will do everything possible to get your article published quickly and accurately. Please use this proof only for checking the typesetting, editing, completeness and correctness of the text, tables and figures. Significant changes to the article as accepted for publication will only be considered at this stage with permission from the Editor. It is important to ensure that all corrections are sent back to us in one communication. Please check carefully before replying, as inclusion of any subsequent corrections cannot be guaranteed. Proofreading is solely your responsibility.

Offprints

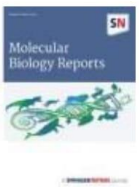
The corresponding author will be notified and receive a link to the published version of the open access article on ScienceDirect. This link is in the form of an article DOI link which can be shared via email and social networks. For an extra charge, paper offprints can be ordered via the offprint order form which is sent once the article is accepted for publication. Both corresponding and co-authors may order offprints at any time via Elsevier's Author Services.

Additional information
Questions, ask the managing editor, at dib-me@elsevier.com

AUTHOR INQUIRIES

Visit the Elsevier Support Center to find the answers you need. Here you will find everything from Frequently Asked Questions to ways to get in touch.
You can also check the status of your submitted article or find out when your accepted article will be published.

Annexure - B



Molecular Biology Reports
An International Journal on Molecular and Cellular Biology

- Journal home
- Submission guidelines

Submission guidelines

Contents

- Open access publishing

Instructions for Authors

General

The cover letter must include a valid institutional email address for each author. If this is not

possible, the cover letter must include a means of verifying each author – such as an ORCID ID (that contains information on publications, do not include empty IDs) or Scopus ID or a link to their publication record(s).

Authors are encouraged to suggest suitable reviewers and/or request the exclusion of certain individuals when they submit their manuscripts. When suggesting reviewers, authors should make sure they are totally independent and not connected to the work in any way. It is strongly recommended to have a mix of reviewers from different countries and different institutions. When suggesting reviewers, the Corresponding Author must provide an institutional email address for each suggested reviewer, or, if this is not possible to include other means of verifying the identity such as a link to a personal homepage, a link to the publication record or a researcher or author ID in the submission letter. Please note that the Journal may not use the suggestions, but suggestions are appreciated and may help facilitate the peer review process.

Please be sure to check the Compliance with Ethical Standards section. Conflict of Interest statement, ethical approval for animal studies and ethical approval together with informed consent for human studies are mandatory.

Back to top

Original Research Article Guidelines

1. Maximum Word Count: 5,000 words

2. References: 50 or less

3. Results and Discussion: sections should be separate, not combined,

4. Figures and/or Tables: maximum 5 separate figures/tables. A figure should have a maximum of 4 panels.

Back to top

Guidelines for Short Communication Articles

1. Maximum Word Count: 2,500 words

2. References: 25 or less

3. Results and Discussion: sections may be combined

4. Figures and/or Tables: 3 or less (4 panels maximum per figure)

Back to top

(Mini-) Review Article Guidelines

1. Maximum Word Count: 10,000 words (5,000 words for mini-review)

2. References: 100 or less (60 or less for mini-review)

3. Figures and/or Tables: 4 or less

4. Content/Format suggestion: A review article should focus on a single topic of a field and should discuss the most recent findings. They should not be focused on the author's own work but should try to incorporate the work of several different researchers. Language should be simple, new concepts should be defined and specialized terminology must be explained.

The main text of the review article should be divided into sections.

Back to top

Abstract Format

Max Word Count: 250 words

Format: The abstract should be presented divided into subheadings as follows (unless it is a mini or full review article):

I. Background: Brief summary of basic, relevant background info (2-3 sentences). Rationale and

purpose of the study.

II. Methods and Results: Brief explanation of experimental procedure and presentation of

significant results. Include sample sizes as well as animal species if applicable.

III. Conclusions: Succinct interpretation of results as well as significance of findings. Statement

of the main conclusion of the study. Emphasis should be on new information found during the study.

Back to top

Manuscript Submission

Manuscript Submission

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

Permissions

Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain permission from the copyright owner(s) for both the print and online format and to include evidence that such permission has been granted when submitting their

papers. Any material received without such evidence will be assumed to originate from the authors.

Online Submission

Please follow the hyperlink "Submit manuscript" on the right and upload all of your manuscript files following the instructions given on the screen.

Please ensure you provide all relevant editable source files. Failing to submit these source files might cause unnecessary delays in the review and production process.

Back to top

Title page

Title Page

Please make sure your title page contains the following information.

Title

The title should be concise and informative.

Author information

- The name(s) of the author(s)
- The affiliation(s) of the author(s), i.e. institution, (department), city, (state), country
- A clear indication and an active e-mail address of the corresponding author
- If available, the 16-digit ORCID of the author(s)

If address information is provided with the affiliation(s) it will also be published.

For authors that are (temporarily) unaffiliated we will only capture their city and country of residence, not their e-mail address unless specifically requested.

Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

For life science journals only (when applicable)

Trial registration number and date of registration

Trial registration number, date of registration followed by "retrospectively registered"

Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

Declarations

All manuscripts must contain the following sections under the heading 'Declarations'.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

To be used for all articles, including articles with biological applications

Funding (information that explains whether and by whom the research was supported)

Conflicts of interest/Competing interests (include appropriate disclosures)

Availability of data and material (data transparency)

Code availability (software application or custom code)

Authors' contributions (optional: please review the submission guidelines from the journal whether statements are mandatory)

Additional declarations for articles in life science journals that report the results of studies involving humans and/or animals

Ethics approval (include appropriate approvals or waivers)

Consent to participate (include appropriate statements)

Consent for publication (include appropriate statements)

Please see the relevant sections in the submission guidelines for further information as well as various examples of wording. Please revise/customize the sample statements according to your own needs.

Back to top

Text

Text Formatting

Manuscripts should be submitted in Word.

- Use a normal, plain font (e.g., 10-point Times Roman) for text.
- Use italics for emphasis.
- Use the automatic page numbering function to number the pages.
- Do not use field functions.
- Use tab stops or other commands for indents, not the space bar.
- Use the table function, not spreadsheets, to make tables.
- Use the equation editor or MathType for equations.
- Save your file in docx format (Word 2007 or higher) or doc format (older Word versions).

Manuscripts with mathematical content can also be submitted in LaTeX.

LaTeX macro package (Download zip, 188 kB)

Headings

Please use no more than three levels of displayed headings.

Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section on the title page. The names of funding organizations should be written in full.

Back to top

References

Citation

Reference citations in the text should be identified by numbers in square brackets. Some examples:

1. Negotiation research spans many disciplines [3].

2. This result was later contradicted by Becker and Seligman [5].

3. This effect has been widely studied [1-3, 7].

Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text. Do not use footnotes or endnotes as a substitute for a reference list.

The entries in the list should be numbered consecutively.

If available, please always include DOIs as full DOI links in your reference list (e.g. "https://doi.org/abc").

- Journal article

  Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. Eur J Appl Physiol 105:731-738. https://doi.org/10.1007/s00421-008-0955-8

  Ideally, the names of all authors should be provided, but the usage of "et al" in long author lists will also be accepted:

  Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. N Engl J Med 965:325–329

- Article by DOI

  Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. J Mol Med. https://doi.org/10.1007/s001090000086

- Book

  South J, Blass B (2001) The future of modern genomics. Blackwell, London

- Book chapter

  Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) The rise of modern genomics, 3rd edn. Wiley, New York, pp 230-257

- Online document

  Cartwright J (2007) Big stars have weather too. IOP Publishing PhysicsWeb. http://physicsweb.org/articles/news/11/6/16/1. Accessed 26 June 2007

- Dissertation

  Trent JW (1975) Experimental acute renal failure. Dissertation, University of California

Always use the standard abbreviation of a journal's name according to the ISSN List of Title Word Abbreviations, see

ISSN.org LTWA

If you are unsure, please use the full journal title.

Authors preparing their manuscript in LaTeX can use the bibtex file spbasic.bst which is included in Springer's LaTeX macro package.

Back to top

Tables

- All tables are to be numbered using Arabic numerals.
- Tables should always be cited in text in consecutive numerical order.
- For each table, please supply a table caption (title) explaining the components of the table.
- Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.
- Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

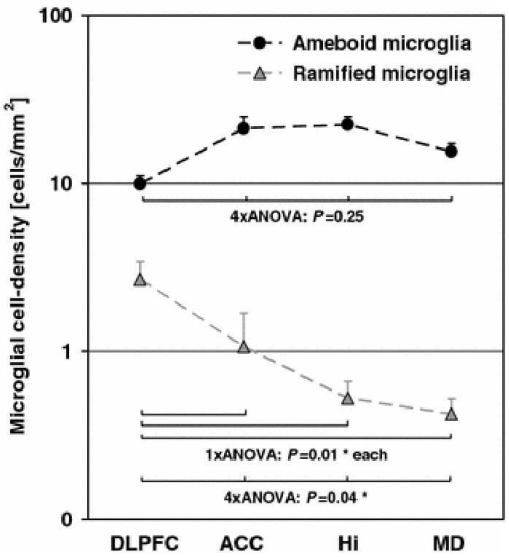
Back to top

Artwork and Illustrations Guidelines

Electronic Figure Submission

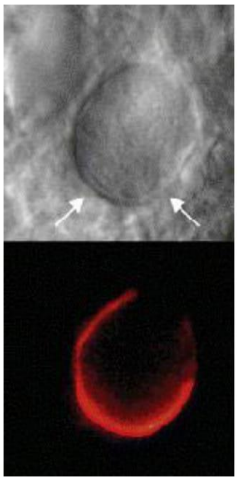
- Supply all figures electronically.
- Indicate what graphics program was used to create the artwork.
- For vector graphics, the preferred format is EPS; for halftones, please use TIFF format. MSOffice files are also acceptable.
- Vector graphics containing fonts must have the fonts embedded in the files.
- Name your figure files with "Fig" and the figure number, e.g., Fig1.eps.

Line Art



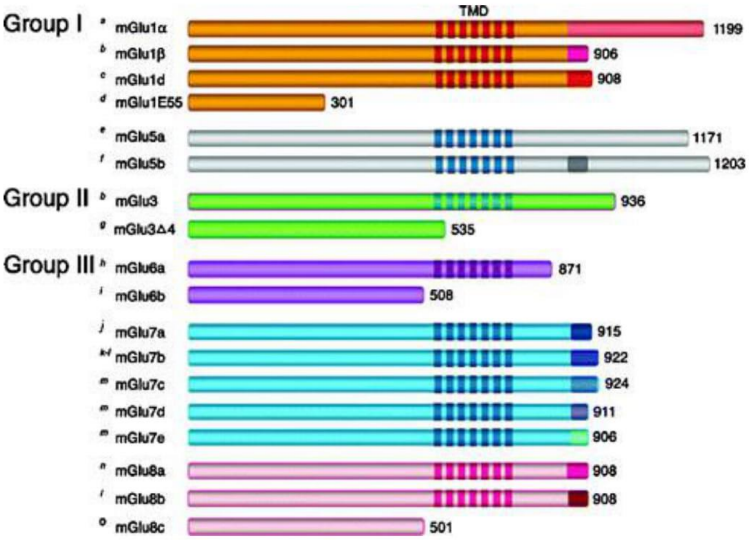
- Definition: Black and white graphic with no shading.
- Do not use faint lines and/or lettering and check that all lines and lettering within the figures are legible at final size.
- All lines should be at least 0.1 mm (0.3 pt) wide.
- Scanned line drawings and line drawings in bitmap format should have a minimum resolution of 1200 dpi.
- Vector graphics containing fonts must have the fonts embedded in the files.

Halftone Art



- Definition: Photographs, drawings, or paintings with fine shading, etc.
- If any magnification is used in the photographs, indicate this by using scale bars within the figures themselves.
- Halftones should have a minimum resolution of 300 dpi.

Combination Art



- Definition: a combination of halftone and line art, e.g., halftones containing line drawing, extensive lettering, color diagrams, etc.
- Combination artwork should have a minimum resolution of 600 dpi.

Color Art

- Color art is free of charge for online publication.
- If black and white will be shown in the print version, make sure that the main information will still be visible. Many colors are not distinguishable from one another when converted to black and white. A simple way to check this is to make a xerographic copy to see if the necessary distinctions between the different colors are still apparent.
- If the figures will be printed in black and white, do not refer to color in the captions.
- Color illustrations should be submitted as RGB (8 bits per channel).

Figure Lettering

- To add lettering, it is best to use Helvetica or Arial (sans serif fonts).
- Keep lettering consistently sized throughout your final-sized artwork, usually about 2–3 mm (8–12 pt).
- Variance of type size within an illustration should be minimal, e.g., do not use 8-pt type on an axis and 20-pt type for the axis label.
- Avoid effects such as shading, outline letters, etc.
- Do not include titles or captions within your illustrations.

Figure Numbering

- All figures are to be numbered using Arabic numerals.
- Figures should always be cited in text in consecutive numerical order.
- Figure parts should be denoted by lowercase letters (a, b, c, etc.).
- If an appendix appears in your article and it contains one or more figures, continue the consecutive numbering of the main text. Do not number the appendix figures,"A1, A2, A3, etc." Figures in online appendices [Supplementary Information (SI)] should, however, be numbered separately.

Figure Captions

- Each figure should have a concise caption describing accurately what the figure depicts. Include the captions in the text file of the manuscript, not in the figure file.
- Figure captions begin with the term Fig. in bold type, followed by the figure number, also in bold type.

- No punctuation is to be included after the number, nor is any punctuation to be placed at the end of the caption.
- Identify all elements found in the figure in the figure caption; and use boxes, circles, etc., as coordinate points in graphs.
- Identify previously published material by giving the original source in the form of a reference citation at the end of the figure caption.

Figure Placement and Size

- Figures should be submitted separately from the text, if possible.
- When preparing your figures, size figures to fit in the column width.
- For large-sized journals the figures should be 84 mm (for double-column text areas), or 174 mm (for single-column text areas) wide and not higher than 234 mm.
- For small-sized journals, the figures should be 119 mm wide and not higher than 195 mm.

Permissions

If you include figures that have already been published elsewhere, you must obtain permission from the copyright owner(s) for both the print and online format. Please be aware that some publishers do not grant electronic rights for free and that Springer will not be able to refund any costs that may have occurred to receive these permissions. In such cases, material from other sources should be used.

Accessibility

In order to give people of all abilities and disabilities access to the content of your figures, please make sure that

- All figures have descriptive captions (blind users could then use a text-to-speech software or a text-to-Braille hardware)
- Patterns are used instead of or in addition to colors for conveying information (colorblind users would then be able to distinguish the visual elements)
- Any figure lettering has a contrast ratio of at least 4.5:1

Back to top

Supplementary Information (SI)

Springer accepts electronic multimedia files (animations, movies, audio, etc.) and other supplementary files to be published online along with an article or a book chapter. This feature can add dimension to the author's article, as certain information cannot be printed or is more convenient in electronic form.

Before submitting research datasets as Supplementary Information, authors should read the journal's Research data policy. We encourage research data to be archived in data repositories wherever possible.

Submission

- Supply all supplementary material in standard file formats.
- Please include in each file the following information: article title, journal name, author names; affiliation and e-mail address of the corresponding author.
- To accommodate user downloads, please keep in mind that larger-sized files may require very long download times and that some users may experience other problems during downloading.

Audio, Video, and Animations

- Aspect ratio: 16:9 or 4:3
- Maximum file size: 25 GB
- Minimum video duration: 1 sec
- Supported file formats: avi, wmv, mp4, mov, m2p, mp2, mpg, mpeg, flv, mxf, mts, m4v, 3gp

Text and Presentations

- Submit your material in PDF format; .doc or .ppt files are not suitable for long-term viability.
- A collection of figures may also be combined in a PDF file.

Spreadsheets

- Spreadsheets should be submitted as .csv or .xlsx files (MS Excel).

Specialized Formats

- Specialized format such as .pdb (chemical), .wrl (VRML), .nb (Mathematica notebook), and .tex can also be supplied.

Collecting Multiple Files

- It is possible to collect multiple files in a .zip or .gz file.

Numbering

- If supplying any supplementary material, the text must make specific mention of the material as a citation, similar to that of figures and tables.
- Refer to the supplementary files as “Online Resource”, e.g., "... as shown in the animation (Online Resource 3)", “... additional data are given in Online Resource 4”.
- Name the files consecutively, e.g. “ESM_3.mpg”, “ESM_4.pdf”.

Captions

- For each supplementary material, please supply a concise caption describing the content of the file.

Processing of supplementary files

- Supplementary Information (SI) will be published as received from the author without any conversion, editing, or reformatting.

Accessibility

In order to give people of all abilities and disabilities access to the content of your supplementary files, please make sure that

- The manuscript contains a descriptive caption for each supplementary material
- Video files do not contain anything that flashes more than three times per second (so that users prone to seizures caused by such effects are not put at risk)

Back to top

English Language Editing

For editors and reviewers to accurately assess the work presented in your manuscript you need to ensure the English language is of sufficient quality to be understood. If you need help with writing in English you should consider:

- Asking a colleague who is a native English speaker to review your manuscript for clarity.
- Visiting the English language tutorial which covers the common mistakes when writing in English.
- Using a professional language editing service where editors will improve the English to ensure that your meaning is clear and identify problems that require your review. Two such services are provided by our affiliates Nature Research Editing Service and American Journal Experts. Springer authors are entitled to a 10% discount on their first submission to either of these services, simply follow the links below.

English language tutorial

Nature Research Editing Service

American Journal Experts

Please note that the use of a language editing service is not a requirement for publication in this journal and does not imply or guarantee that the article will be selected for peer review or accepted.

If your manuscript is accepted it will be checked by our copyeditors for spelling and formal style before publication.

.

为便于编辑和评审专家准确评估您稿件中陈述的研究工作，您需要确保您的英语语言质量足以令人理解。如果您需要英文写作方面的帮助，您可以考虑：

● 请一位以英语为母语的同事审核您**的稿件是否表意清晰**。

● 查看一些有关英语写作中常见语言错误的教程。

● 使用专业语言编辑服务，编辑人员会对英语进行润色，以确保您的意思表达清晰，并识别需要您复核的问题。我们的附属机构 Nature Research Editing Service 和合作伙伴 American Journal Experts 即可提供此类服务。

教程

Nature Research Editing Service

American Journal Experts

请注意，使用语言编辑服务并非在期刊上发表文章的必要条件，同时也并不意味或保证文章将被选中进行同行评议或被接受。

如果您的稿件被接受，在发表之前，我们的文字编辑会检查您的文稿拼写是否规范以及文体是否正式。

.

エディターと査読者があなたの論文を正しく評価するには、使用されている英語の質が十分に高いことが必要とされます。英語での論文執筆に際してサポートが必要な場合には、次のオプションがあります：

・英語を母国語とする同僚に、原稿で使用されている英語が明確であるかをチェックしてもらう。

・英語で執筆する際のよくある間違いに関する英語のチュートリアルを参照する。

・プロの英文校正サービスを利用する。校正者が原稿の意味を明確にしたり、問題点を指摘し、英語の質を向上させます。Nature Research Editing Service とAmerican Journal Experts の2つは弊社と提携しているサービスです。Springer の著者は、いずれのサービスも初めて利用する際には10%の割引を受けることができます。以下のリンクを参照ください。

英語のチュートリアル

Nature Research Editing Service

American Journal Experts

英文校正サービスの利用は、投稿先のジャーナルに掲載されるための条件ではないこと、また論文審査や受理を保証するものではないことに留意してください。

原稿が受理されると、出版前に弊社のコピーエディターがスペルと体裁のチェックを行います。

.

영어 원고의 경우, 에디터 및 리뷰어들이 귀하의 원고에 실린 결과물을 정확하게 평가할 수 있도록, 그들이 충분히 이해할 수 있을 만한 수준으로 작성되어야 합니다. 만약 영작문과 관련하여 도움을 받기를 원하신다면 다음의 사항들을 고려하여 주십시오:

• 귀하의 원고의 표현을 명확히 해줄 영어 원어민 동료를 찾아서 리뷰를 의뢰합니다.

• 영어 튜토리얼 페이지에 방문하여 영어로 글을 쓸 때 자주하는 실수들을 확인합니다.

• 리뷰에 대비하여, 원고의 의미를 명확하게 해주고 리뷰에서 요구하는 문제점들을 식별해서 영문 수준을 향상시켜주는 전문 영문 교정 서비스를 이용합니다. Nature Research Editing Service와 American Journal Experts에서 저희와 협약을 통해 서비스를 제공하고 있습니다. Springer 저자들이 본 교정 서비스를 첫 논문 투고를 위해 사용하시는 경우 10%의 할인이 적용되며, 아래의 링크를 통하여 확인이 가능합니다.

영어 튜토리얼 페이지

Nature Research Editing Service

American Journal Experts

영문 교정 서비스는 게재를 위한 요구사항은 아니며, 해당 서비스의 이용이 피어 리뷰에 논문이 선택되거나 게재가 수락되는 것을 의미하거나 보장하지 않습니다.

원고가 수락될 경우, 출판 전 저희측 편집자에 의해 원고의 철자 및 문체를 검수하는 과정을 거치게 됩니다.

Back to top

Additional Information

Authors should suggest six to eight individuals with position, affiliation, country, email address and expertise. To avoid possible bias and conflicts of interest, authors are not allowed to suggest reviewers from the same institute. Non-American authors must list at least four reviewers from outside their country of origin. Papers may be returned without review if authors do not adhere to the above rules.

Back to top

Ethical Responsibilities of Authors

This journal is committed to upholding the integrity of the scientific record. As a member of the Committee on Publication Ethics (COPE) the journal will follow the COPE guidelines on how to deal with potential acts of misconduct.

Authors should refrain from misrepresenting research results which could damage the trust in the journal, the professionalism of scientific authorship, and ultimately the entire scientific endeavour. Maintaining integrity of the research and its presentation is helped by following the rules of good scientific practice, which include*:

- The manuscript should not be submitted to more than one journal for simultaneous consideration.
- The submitted work should be original and should not have been published elsewhere in any form or language (partially or in full), unless the new work concerns an expansion of previous work. (Please

provide transparency on the re-use of material to avoid the concerns about text-recycling ('self-plagiarism').

- A single study should not be split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time (i.e. 'salami-slicing/publishing').
- Concurrent or secondary publication is sometimes justifiable, provided certain conditions are met. Examples include: translations or a manuscript that is intended for a different group of readers.
- Results should be presented clearly, honestly, and without fabrication, falsification or inappropriate data manipulation (including image based manipulation). Authors should adhere to discipline-specific rules for acquiring, selecting and processing data.
- No data, text, or theories by others are presented as if they were the author's own ('plagiarism'). Proper acknowledgements to other works must be given (this includes material that is closely copied (near verbatim), summarized and/or paraphrased), quotation marks (to indicate words taken from another source) are used for verbatim copying of material, and permissions secured for material that is copyrighted.

Important note: the journal may use software to screen for plagiarism.

- Authors should make sure they have permissions for the use of software, questionnaires/(web) surveys and scales in their studies (if appropriate).
- Research articles and non-research articles (e.g. Opinion, Review, and Commentary articles) must cite appropriate and relevant literature in support of the claims made. Excessive and inappropriate self-citation or coordinated efforts among several authors to collectively self-cite is strongly discouraged.
- Authors should avoid untrue statements about an entity (who can be an individual person or a company) or descriptions of their behavior or actions that could potentially be seen as personal attacks or allegations about that person.
- Research that may be misapplied to pose a threat to public health or national security should be clearly identified in the manuscript (e.g. dual use of research). Examples include creation of harmful consequences of biological agents or toxins, disruption of immunity

of vaccines, unusual hazards in the use of chemicals, weaponization of research/technology (amongst others).

- Authors are strongly advised to ensure the author group, the Corresponding Author, and the order of authors are all correct at submission. Adding and/or deleting authors during the revision stages is generally not permitted, but in some cases may be warranted. Reasons for changes in authorship should be explained in detail. Please note that changes to authorship cannot be made after acceptance of a manuscript.

*All of the above are guidelines and authors need to make sure to respect third parties rights such as copyright and/or moral rights.

Upon request authors should be prepared to send relevant documentation or data in order to verify the validity of the results presented. This could be in the form of raw data, samples, records, etc. Sensitive information in the form of confidential or proprietary data is excluded.

If there is suspicion of misbehavior or alleged fraud the Journal and/or Publisher will carry out an investigation following COPE guidelines. If, after investigation, there are valid concerns, the author(s) concerned will be contacted under their given e-mail address and given an opportunity to address the issue. Depending on the situation, this may result in the Journal's and/or Publisher's implementation of the following measures, including, but not limited to:

- If the manuscript is still under consideration, it may be rejected and returned to the author.
- If the article has already been published online, depending on the nature and severity of the infraction:

  - an erratum/correction may be placed with the article

  - an expression of concern may be placed with the article

  - or in severe cases retraction of the article may occur.

The reason will be given in the published erratum/correction, expression of concern or retraction note. Please note that retraction means that the

article is **maintained on the platform**, watermarked "retracted" and the explanation for the retraction is provided in a note linked to the watermarked article.

- The author's institution may be informed
- A notice of suspected transgression of ethical standards in the peer review system may be included as part of the author's and article's bibliographic record.

Fundamental errors

Authors have an obligation to correct mistakes once they discover a significant error or inaccuracy in their published article. The author(s) is/are requested to contact the journal and explain in what sense the error is impacting the article. A decision on how to correct the literature will depend on the nature of the error. This may be a correction or retraction. The retraction note should provide transparency which parts of the article are impacted by the error.

Suggesting / excluding reviewers

Authors are welcome to suggest suitable reviewers and/or request the exclusion of certain individuals when they submit their manuscripts. When suggesting reviewers, authors should make sure they are totally independent and not connected to the work in any way. It is strongly recommended to suggest a mix of reviewers from different countries and different institutions. When suggesting reviewers, the Corresponding Author must provide an institutional email address for each suggested reviewer, or, if this is not possible to include other means of verifying the identity such as a link to a personal homepage, a link to the publication record or a researcher or author ID in the submission letter. Please note that the Journal may not use the suggestions, but suggestions are appreciated and may help facilitate the peer review process.

Back to top

Authorship principles

These guidelines describe authorship principles and good authorship practices to which prospective authors should adhere to.

Authorship clarified

The Journal and Publisher assume all authors agreed with the content and that all gave explicit consent to submit and that they obtained consent from the responsible authorities at the institute/organization where the work has been carried out, **before** the work is submitted.

The Publisher does not prescribe the kinds of contributions that warrant authorship. It is recommended that authors adhere to the guidelines for authorship that are applicable in their specific research field. In absence of specific guidelines it is recommended to adhere to the following guidelines*:

All authors whose names appear on the submission

1) made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work;

2) drafted the work or revised it critically for important intellectual content;

3) approved the version to be published; and

4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

* Based on/adapted from:

ICMJE, Defining the Role of Authors and Contributors,

Transparency in authors' contributions and responsibilities to promote integrity in scientific publication, McNutt at all, PNAS February 27, 2018

Disclosures and declarations

All authors are requested to include information regarding sources of funding, financial or non-financial interests, study-specific approval by the appropriate ethics committee for research involving humans and/or

animals, informed consent if the research involved human participants, and a statement on welfare of animals if the research involved animals (as appropriate).

The decision whether such information should be included is not only dependent on the scope of the journal, but also the scope of the article. Work submitted for publication may have implications for public health or general welfare and in those cases it is the responsibility of all authors to include the appropriate disclosures and declarations.

Data transparency

All authors are requested to make sure that all data and materials as well as software application or custom code support their published claims and comply with field standards. Please note that journals may have individual policies on (sharing) research data in concordance with disciplinary norms and expectations.

Role of the Corresponding Author

One author is assigned as Corresponding Author and acts on behalf of all co-authors and ensures that questions related to the accuracy or integrity of any part of the work are appropriately addressed.

The Corresponding Author is responsible for the following requirements:

- ensuring that all listed authors have approved the manuscript before submission, including the names and order of authors;
- managing all communication between the Journal and all co-authors, before and after publication;*
- providing transparency on re-use of material and mention any unpublished material (for example manuscripts in press) included in the manuscript in a cover letter to the Editor;
- making sure disclosures, declarations and transparency on data statements from all authors are included in the manuscript as appropriate (see above).

* The requirement of managing all communication between the journal and all co-authors during submission and proofing may be delegated to a

Contact or Submitting Author. In this case please make sure the Corresponding Author is clearly indicated in the manuscript.

Author contributions

In absence of specific instructions and in research fields where it is possible to describe discrete efforts, the Publisher recommends authors to include contribution statements in the work that specifies the contribution of every author in order to promote transparency. These contributions should be listed at the separate title page.

Examples of such statement(s) are shown below:

• Free text:

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [full name], [full name] and [full name]. The first draft of the manuscript was written by [full name] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Example: CRediT taxonomy:

• Conceptualization: [full name], …; Methodology: [full name], …; Formal analysis and investigation: [full name], …; Writing - original draft preparation: [full name, …]; Writing - review and editing: [full name], …; Funding acquisition: [full name], …; Resources: [full name], …; Supervision: [full name],….

For **review articles** where discrete statements are less applicable a statement should be included who had the idea for the article, who performed the literature search and data analysis, and who drafted and/or critically revised the work.

For articles that are based primarily on the **student's dissertation or thesis**, it is recommended that the student is usually listed as principal author:

A Graduate Student's Guide to Determining Authorship Credit and Authorship Order, APA Science Student Council 2006

Affiliation

The primary affiliation for each author should be the institution where the majority of their work was done. If an author has subsequently moved, the current address may additionally be stated. Addresses will not be updated or changed after publication of the article.

Changes to authorship

Authors are strongly advised to ensure the correct author group, the Corresponding Author, and the order of authors at submission. Changes of authorship by adding or deleting authors, and/or changes in Corresponding Author, and/or changes in the sequence of authors are **not** accepted **after acceptance** of a manuscript.

- **Please note that author names will be published exactly as they appear on the accepted submission!**

Please make sure that the names of all authors are present and correctly spelled, and that addresses and affiliations are current.

Adding and/or deleting authors at revision stage are generally not permitted, but in some cases it may be warranted. Reasons for these changes in authorship should be explained. Approval of the change during revision is at the discretion of the Editor-in-Chief. Please note that journals may have individual policies on adding and/or deleting authors during revision stage.

Author identification

Authors are recommended to use their ORCID ID when submitting an article for consideration or acquire an ORCID ID via the submission process.

Deceased or incapacitated authors

For cases in which a co-author dies or is incapacitated during the writing, submission, or peer-review process, and the co-authors feel it is

appropriate to include the author, co-authors should obtain approval from a (legal) representative which could be a direct relative.

Authorship issues or disputes

In the case of an authorship dispute during peer review or after acceptance and publication, the Journal will not be in a position to investigate or adjudicate. Authors will be asked to resolve the dispute themselves. If they are unable the Journal reserves the right to withdraw a manuscript from the editorial process or in case of a published paper raise the issue with the authors' institution(s) and abide by its guidelines.

Confidentiality

Authors should treat all communication with the Journal as confidential which includes correspondence with direct representatives from the Journal such as Editors-in-Chief and/or Handling Editors and reviewers' reports unless explicit consent has been received to share information.

Back to top

Compliance with Ethical Standards

To ensure objectivity and transparency in research and to ensure that accepted principles of ethical and professional conduct have been followed, authors should include information regarding sources of funding, potential conflicts of interest (financial or non-financial), informed consent if the research involved human participants, and a statement on welfare of animals if the research involved animals.

Authors should include the following statements (if applicable) in a separate section entitled "Compliance with Ethical Standards" when submitting a paper:

- Disclosure of potential conflicts of interest
- Research involving Human Participants and/or Animals
- Informed consent

Please note that standards could vary slightly per journal dependent on their peer review policies (i.e. single or double blind peer review) as well as

per journal subject discipline. Before submitting your article check the instructions following this section carefully.

The corresponding author should be prepared to collect documentation of compliance with ethical standards and send if requested during peer review or after publication.

The Editors reserve the right to reject manuscripts that do not comply with the above-mentioned guidelines. The author will be held responsible for false statements or failure to fulfill the above-mentioned guidelines.

Back to top

Conflicts of Interest / Competing Interests

Authors are requested to disclose interests *that are directly or indirectly related to the work submitted for publication*. Interests within the last 3 years of beginning the work (conducting the research and preparing the work for submission) should be reported. Interests outside the 3-year time frame must be disclosed if they could reasonably be perceived as influencing the submitted work. Disclosure of interests provides a complete and transparent process and helps readers form their own judgments of potential bias. This is not meant to imply that a financial relationship with an organization that sponsored the research or compensation received for consultancy work is inappropriate.

Interests that should be considered and disclosed but are not limited to the following:

Funding: Research grants from funding agencies (please give the research funder and the grant number) and/or research support (including salaries, equipment, supplies, reimbursement for attending symposia, and other expenses) by organizations that may gain or lose financially through publication of this manuscript.

Employment: Recent (while engaged in the research project), present or anticipated employment by any organization that may gain or lose financially through publication of this manuscript. This includes multiple affiliations (if applicable).

Financial interests: Stocks or shares in companies (including holdings of spouse and/or children) that may gain or lose financially through publication of this manuscript; consultation fees or other forms of remuneration from organizations that may gain or lose financially; patents or patent applications whose value may be affected by publication of this manuscript.

It is difficult to specify a threshold at which a financial interest becomes significant, any such figure is necessarily arbitrary, so one possible practical guideline is the following: "Any undeclared financial interest that could embarrass the author were it to become publicly known after the work was published."

Non-financial interests: In addition, authors are requested to disclose interests that go beyond financial interests that could impart bias on the work submitted for publication such as professional interests, personal relationships or personal beliefs (amongst others). Examples include, but are not limited to: position on editorial board, advisory board or board of directors or other type of management relationships; writing and/or consulting for educational purposes; expert witness; mentoring relations; and so forth.

Primary research articles require a disclosure statement. Review articles present an expert synthesis of evidence and may be treated as an authoritative work on a subject. Review articles therefore require a disclosure statement.Other article types such as editorials, book reviews, comments (amongst others) may, dependent on their content, require a disclosure statement. If you are unclear whether your article type requires a disclosure statement, please contact the Editor-in-Chief.

Please note that, in addition to the above requirements, funding information (given that funding is a potential conflict of interest (as mentioned above)) needs to be disclosed upon submission of the manuscript in the peer review system. This information will automatically be added to the Record of CrossMark, however it is **not added** to the manuscript itself. Under 'summary of requirements' (see below) funding information should be included in the '**Declarations**' section.

Summary of requirements

The above should be summarized in a statement and placed in a 'Declarations' section before the reference list under a heading of 'Funding' and/or 'Conflicts of interests'/'Competing interests'. Other declarations include Ethics approval, Consent, Data, Material and/or Code availability and Authors' contribution statements.

Please see the various examples of wording below and revise/customize the sample statements according to your own needs.

When all authors have the same (or no) conflicts and/or funding it is sufficient to use one blanket statement.

Examples of statements to be used when funding has been received:

- Partial financial support was received from [...]
- The research leading to these results received funding from [...] under Grant Agreement No[...].
- This study was funded by [...]
- This work was supported by [...] (Grant numbers [...] and [...]

Examples of statements to be used when there is no funding:

- The authors did not receive support from any organization for the submitted work.
- No funding was received to assist with the preparation of this manuscript.
- No funding was received for conducting this study.
- No funds, grants, or other support was received.

Examples of statements to be used when there are interests to declare:

- **Financial interests:** Author A has received research support from Company A. Author B has received a speaker honorarium from Company Wand owns stock in Company X. Author C is consultant to company Y.

Non-financial interests: Author C is an unpaid member of committee Z.

- **Financial interests:** The authors declare they have no financial interests.

  Non-financial interests: Author A is on the board of directors of Y and receives no compensation as member of the board of directors.

- **Financial interests:** Author A received a speaking fee from Y for Z. Author B receives a salary from association X. X where s/he is the Executive Director.

  Non-financial interests: none.

- **Financial interests:** Author A and B declare they have no financial interests. Author C has received speaker and consultant honoraria from Company M and Company N. Dr. C has received speaker honorarium and research funding from Company M and Company O. Author D has received travel support from Company O.

  Non-financial interests: Author D has served on advisory boards for Company M, Company N and Company O.

Examples of statements to be used when authors have nothing to declare:

- The authors have no relevant financial or non-financial interests to disclose.
- The authors have no conflicts of interest to declare that are relevant to the content of this article.
- All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.
- The authors have no financial or proprietary interests in any material discussed in this article.

Authors are responsible for correctness of the statements provided in the manuscript. See also Authorship Principles. The Editor-in-Chief reserves the right to reject submissions that do not meet the guidelines described in this section.

Back to top

After acceptance

Upon acceptance, your article will be exported to Production to undergo typesetting. Once typesetting is complete, you will receive a link asking you to confirm your affiliation, choose the publishing model for your article as well as arrange rights and payment of any associated publication cost.

Once you have completed this, your article will be processed and you will receive the proofs.

Article publishing agreement

Depending on the ownership of the journal and its policies, you will either grant the Publisher an exclusive licence to publish the article or will be asked to transfer copyright of the article to the Publisher.

Offprints

Offprints can be ordered by the corresponding author.

Color illustrations

Online publication of color illustrations is free of charge. For color in the print version, authors will be expected to make a contribution towards the extra costs.

Proof reading

The purpose of the proof is to check for typesetting or conversion errors and the completeness and accuracy of the text, tables and figures. Substantial changes in content, e.g., new results, corrected values, title and authorship, are not allowed without the approval of the Editor.

After online publication, further changes can only be made in the form of an Erratum, which will be hyperlinked to the article.

Online First

The article will be published online after receipt of the corrected proofs. This is the official first publication citable with the DOI. After release of the printed version, the paper can also be cited by issue and page numbers.

Back to top

Open Choice

Open Choice allows you to publish open access in more than 1850 Springer Nature journals, making your research more visible and accessible immediately on publication.

Article processing charges (APCs) vary by journal – view the full list

Benefits:

- Increased researcher engagement: Open Choice enables access by anyone with an internet connection, immediately on publication.
- Higher visibility and impact: In Springer hybrid journals, OA articles are accessed 4 times more often on average, and cited 1.7 more times on average*.
- Easy compliance with funder and institutional mandates: Many funders require open access publishing, and some take compliance into account when assessing future grant applications.

It is easy to find funding to support open access – please see our funding and support pages for more information.

*) Within the first three years of publication. Springer Nature hybrid journal OA impact analysis, 2018.

Open Choice

Funding and Support pages

Copyright and license term – CC BY

Open Choice articles do not require transfer of copyright as the copyright remains with the author. In opting for open access, the author(s) agree to publish the article under the Creative Commons Attribution License.

Find more about the license agreement

Back to top

Open access publishing

Molecular Biology Reports publishes open access articles. Authors of open access articles published in this journal retain the copyright of their articles and are free to reproduce and disseminate their work.

Visit our Open access publishing page to learn more.

Back to top

For authors

Submission guidelinesEthics & disclosuresFees and fundingContact the journalSubmit manuscript

Explore

Online first articlesVolumes and issues Sign up for alerts

Publish with us

Authors & Editors
Journal authors
Publishing ethics
Open Access & Springer

Discover content

SpringerLink
Books A-Z
Journals A-Z
Video

Other services

Annexure - C

Revised May 3rd, 2019

Brief guide for submission to *Nature Communications*

This guide outlines key points for preparing primary research manuscripts for submission to *Nature Communications*.

The corresponding author should be familiar with the *Nature* journals' editorial policies and is solely responsible for communicating with the journal and managing communication between coauthors. Before submission, the corresponding author ensures that all authors are included in the author list and agree with its order, and that they are aware the manuscript is to be submitted. For more information on editorial and authorship policies please review our Guide to Authors.

Cover letter

Although optional, the cover letter is an excellent opportunity to briefly discuss the context and importance of the submitted work and why it is appropriate for the journal. Please avoid repeating information that is already present in the abstract and introduction. The cover letter is not shared with the referees, and should be used to provide confidential information, such as conflicts of interest, and to declare any related work that is in press or submitted elsewhere.

Main manuscript

The *Nature* journals are flexible with regard to the format of initial submissions. Within reason, style and length will not influence consideration of a manuscript. If revisions are requested, the editor will provide detailed formatting instructions at that time. For information on length and formatting consult *Nature Communications*' content types.

Title. If possible, this should be 15 words or fewer and should not contain technical terms, abbreviations, punctuation and active verbs.

Authors. Corresponding author(s) should be identified with an asterisk.

Abstract. Provide a general introduction to the topic and a brief non-technical summary of your main results and their implication.

Text length and formatting. Attention to the following details can help expedite publication if we invite a revision after external review.

- **Articles:** an abstract of approximately 150 words, unreferenced; main text of no more than 5,000 words and 10 display items (figures, tables). As a guideline, Articles allow up to 70 references. Section headings should be used and subheadings should appear in 'Results' and 'Methods'.

Methods. The Methods section appears in all online original research articles and should contain all elements necessary for interpretation and replication of the results. Methods should be written as concisely as possible and typically do not exceed 3,000 words but may be longer if necessary. We encourage you to deposit any step-by-step protocols used in your study in Protocol Exchange, an open resource maintained by NPG. These protocols are linked to the Methods section upon publication.

References. These may only contain citations and should list only one publication with each number. Include the title of the cited article or dataset.

Acknowledgements (optional). Keep acknowledgements brief and do not include thanks to anonymous referees or editors, or effusive comments. Grant or contribution numbers may be acknowledged.

Author contributions. You must include a statement that specifies the individual contributions of each co-author. For example: "A.P.M. 'contributed' Y and Z; B.T.R. 'contributed' Y," etc. See our authorship policies for more details.

Competing interests. Submission of a competing interests statement is required for all content of the journal.

Materials & Correspondence. Indicate the author(s) to whom correspondence and material requests should be addressed.

Tables. Each table should be prepared using the Table menu in Word or the table environment in TeX/LaTeX and accompanied by a short title sentence describing what the table shows. Further details can be included as footnotes to the table.

Figures

High-resolution image files are not required at initial submission, but please ensure that images are of sufficient resolution for referees to properly assess the data.

Should your manuscript be accepted, you will receive more extensive instructions for final submission of display items. However, some guidelines for final figure preparation are included below and here if you wish to minimize later revisions and possible delays.

- Provide images in RGB color and at 300 dpi or higher resolution.
- Use the same typeface (Arial or Helvetica) for all figures. Use symbol font for Greek letters.
- Use distinct colors with comparable visibility and avoid the use of red and green for contrast. Recoloring primary data, such as fluorescence images, to color-safe combinations such as green and magenta or other accessible color palettes is strongly encouraged. Use of the rainbow color scale should be avoided.
- Figures are best prepared at the size you would expect them to appear in print. At this size, the optimum font size is between 5pt and 8pt.
- We prefer vector files with editable layers. Acceptable formats are: .ai, .eps, .pdf, .ps and .svg for fully editable vector-based art; layered .psd and .tif for editable layered art; .psd, .tif, .png and .jpg for bitmap images; .ppt if fully editable and without styling effects; ChemDraw (.cdx) for chemical structures.
- Please use the Nature Research Chemical Structures Guide and ChemDraw template for formatting of chemical structures.

Figure legends should be <350 words each. They should begin with a brief title sentence for the whole figure and continue with a short statement of what is depicted in the figure, not the results (or data) of the experiment or the methods used. Legends should be detailed enough so that each figure and caption can, as far as possible, be understood in isolation from the main text.

Statistical information

Comprehensive information on the statistical analyses used must be included in the paper. The Methods must include a statistics section where you describe the statistical tests used and whether they were one- or two-tailed. Please ensure that the error bars are defined throughout the figures. For all statistics (including error bars), provide the EXACT n values used to calculate the statistics (reporting individual values rather than a range if n varied among experiments). For representative results, report the number of times that the measurements were repeated. Where relevant, provide exact values for both significant and non-significant P values. For ANOVAs, provide F values and degrees of freedom. For t-tests, provide t-values and degrees of freedom. Please specifically define the replicates.

Supplementary information

Please submit supplementary figures, small tables and text as a single combined PDF document. Tables longer than one page should be provided as an Excel or similar file type. For optimal quality video files please use H.264 encoding, the standard aspect ratio of 16:9 (4:3 is second best) and do not compress the video. We encourage submission of step-by-step synthesis procedures for chemical compounds and data on compound characterization. Supplementary information is not copy-edited, so please ensure that it is clearly and succinctly presented, and that the style and terminology conform to the rest of the manuscript.

Data availability

Please provide a data availability statement as a separate section after the Methods section but before the References, under the heading "Data Availability"; detailed guidance can be found in our data availability and data citations policy. Certain data types must be deposited in an appropriate public structured data depository (details are available here) and the accession number(s) provided in the manuscript. Full access is required at publication. Should full access to data be required for peer review, authors must provide it.

We encourage provision of other source data in unstructured public depositories such as Dryad or figshare, or as supplementary information. To maximize data reuse, we encourage publication of detailed descriptions of datasets in *Scientific Data*.

Crystallographic data

Manuscripts reporting new crystallographic structures of small molecules must be accompanied by a standard .cif file. A structural figure with probability ellipsoids should be included in the main supplementary information file. The structure factors for each structure should also be submitted, preferably embedded in the main .cif file, although they may be provided as a separate .hkl and/or .fcf file. Use of the latest version of the program SHELXL, which embeds the structure factors information in the main .cif file, is encouraged. The structure factors and structural output must be checked using IUCr's CheckCIF routine and a pdf copy of the output supplied, explaining any A- or B-level alerts.

Computer code

Any previously unreported custom computer code used to generate results reported in the manuscript that are central to the main claims must be made available to editors and reviewers upon request. Any practical issues preventing code sharing will be evaluated by the editors who reserve the right to decline the manuscript if important code is unavailable. At publication, Nature journals consider it best practice to release custom computer code in a way that allows readers to repeat the published results.

Authors of such research papers will also be asked to fill out a code and software submission checklist that will be made available to editors and reviewers during manuscript assessment. The aim is to make studies that use such code more reliable by ensuring that all relevant documentation is available and by facilitating testing of software by the reviewers. Further detailed guidance and required documentation at submission and acceptance of the manuscript can be found here.

For all studies using custom code that is deemed central to the conclusions, a statement must be included under the heading "Code availability", indicating whether and how the code can be accessed, including any restrictions to access. Code availability statements should be provided as a separate section after the data availability statement but before the References.

Reporting guidelines

To improve the transparency of reporting and the reproducibility of published results, authors of life sciences, behavioural & social sciences and ecological, evolutionary & environmental sciences research articles must provide a completed reporting summary that will be made available to editors and reviewers during manuscript assessment. The reporting summary will be published with all accepted manuscripts. Guidance and resources related to the use and reporting of statistics are available here.

All authors must also complete an editorial policy checklist to ensure compliance with Nature Research editorial policies.

Please note: because of the advanced features used in these forms, you must use Adobe Reader to open the documents and fill them out, instead of opening them in a web browser.

Other reporting checklist

To improve the transparency of reporting and the reproducibility of published results in certain other research areas, you may be asked to complete a different checklist and provide the requested information prior to peer review.

Source data

For relevant manuscripts, we may request a source data file in Microsoft Excel format or a zipped folder. The source data file should, as a minimum, contain the raw data underlying any graphs and charts, and uncropped versions of any gels or blots presented in the figures. Within the source data file, each figure or table (in the main manuscript and in the Supplementary Information) containing relevant data should be represented by a single sheet in an Excel document, or a single .txt file or other file type in a zipped folder. Blot and gel images should be pasted in and labelled with the relevant panel and identifying information such as the antibody used. We also encourage authors to include any other types of raw data that may be appropriate. An example source data file is available demonstrating the correct format.

Human subject data

If you are reporting phase II or phase III randomized controlled trials you must refer to the CONSORT Statement for recommendations to facilitate the complete and transparent reporting of trial findings. Reports that do not conform to the CONSORT guidelines may need to be revised before peer review.

We encourage authors reporting prognostic studies with tumor markers to follow the REMARK reporting guidelines.

Before the start of patient enrollment prospective clinical trials must be registered in www.clinicaltrials.gov or a similar public repository that matches ISMJE criteria and the trial registration number reported in the manuscript. (Trials in which the primary goal is to determine pharmacokinetics are exempt.)

For describing human biospecimens, we recommend referring to the BRISQ reporting guidelines and ensuring at least Tier 1 characteristics are provided (doi: 10.1002/cncy.20147).

Related manuscripts

It is a requirement of submission that you alert us to any related manuscripts with overlapping authorship that are under consideration (including under appeal) or in press at other journals (see our editorial policies on duplicate submissions for details). Copies of these manuscripts should be clearly marked and included as separate files with your submission. Abstracts or other unrefereed preprints do not compromise novelty.

Preprint servers

Nature Research journals support posting of primary research manuscripts on community preprint servers such as arXiv and bioRxiv. We do, however, ask you to respect our policies on posting, citation and licensing of preprints.

Transparent peer review

For published original research articles, the reviewer comments to the authors and the author rebuttal letters of revised versions are published online as a supplementary "peer review file". Authors may opt out of this scheme at the completion of the peer review process, before the paper is accepted. If the manuscript was transferred to us from another Nature journal, we will not publish reviewer reports or author rebuttals of versions of the manuscript under consideration at the originating Nature journal. For more information, please refer to our FAQ page.

Reviewer information

In recognition of the time and expertise our reviewers provide to *Nature Communications'* editorial process, as of November, 2018, we formally acknowledge their contribution to the external peer review of articles published in the journal. All peer-reviewed content will carry an anonymous statement of peer reviewer acknowledgement, and for those reviewers who give their consent, we will publish their names alongside the published article. We will continue to publish peer reviewer reports where authors opt in to our separate transparent peer review scheme. In cases where authors opt in to publication of peer reviewer comments and reviewers opt in to being named, we will not link a reviewer's name to their report unless they choose to sign their comments to the author with their name. For more information, please refer to our FAQ page.

Double-blind peer review

To participate in double-blind peer review, please prepare your manuscript in a way that conceals the identities of all the authors (see checklist) and tick the appropriate box during online submission. Please note that editors do not ensure that the paper is properly anonymized; that is the responsibility of the authors.

Transferring your manuscript

If an editor is unable to offer publication of your manuscript, you have the opportunity to transfer all manuscript materials, the decision letter and any referee comments to a selection of Springer Nature journals without re-entering submission information. Use the link in your decision letter to explore suggested alternative journals. You may then initiate the transfer process to the journal of your choice or submit elsewhere. Please see this page for more information.

Appeals

Authors who feel that they have strong grounds for appealing a decision may contact the journal to request the opening of an appeal, after which they may upload a cogently argued rebuttal letter that addresses the referees' and/or editor's comments in a point-by-point manner. Decisions are reversed on appeal only if the editors are convinced that

the original decision was made in error, or if critical new information or data has been added.

Comments on published articles

Important scientific comments and clarifications on content published in *Nature Communications* may be submitted as Correspondence.

Questions and manuscript submission

General editorial enquiries should be addressed to the Editor at naturecommunications@nature.com. Manuscripts should be submitted through our online submission system. Further submission details are available here.

Open access

As of January 2016, the journal only publishes open access content, and legacy subscription content published in Nature Communications prior to 2016 has been made freely accessible. All articles are published under a CC BY license (Creative Commons Attribution 4.0 International License). We offer APC waivers for papers whose corresponding authors are based in the world's lowest income countries as defined by the World Bank. Discretionary APC waivers for authors will be considered on a case-by-case basis, and may be granted in cases of financial need. All applications for discretionary APC waivers should be made at the point of manuscript submission; requests made during the review process or after acceptance will not be considered. To request a waiver please contact apcwaivers@springernature.com. Full details of our APC waiver and discount policies can be found here.