Association of *Faecalibacterium, Lachnospira*, *Veillonella*, and *Rothia* with childhood wheezing

By

Saara Kanyemba

KNYSAA001

A dissertation submitted to the University of Cape Town
in fulfilment of the requirements for the degree of

Master of Science
in Medicine
(Medical Microbiology)

February 2019

Division of Medical Microbiology
Department of Pathology
Faculty of Health Sciences
University of Cape Town
South Africa

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Declaration

I, ***Saara Kanyemba***, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

This thesis/dissertation has been submitted to the Turnitin module and I confirm that my supervisors have seen my report and any concerns revealed by such have been resolved with my supervisors.

Signature: Signed by candidate

Date: 08/02/2019

Certification by the supervisors

"In terms of the general rules for the degree of MSc at the University of Cape Town ", paragraph GM5, we, the supervisors of the MSc candidate Ms Saara Kanyemba, supports the submission of the thesis for examination.

Primary supervisor:

Dr. Mamadou Kaba (MD, MSc, PhD)[1,2]

[1]Division of Medical Microbiology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa.

[2]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.

Co-supervisor:

Dr Lemese Ah Tow (PhD)[1]

[1]Division of Medical Microbiology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa.

Preface

This thesis generated knowledge to better understand the contribution of *Faecalibacterium, Lachnospira, Veillonella* and *Rothia* (FLVR) bacteria in the development of infant wheezing. Following the completion of my BSc in Microbiology with the University of Namibia (UNAM), I joined the Division of Medical Microbiology (Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa) to undertake a Master of Science (MSc) in Medical Microbiology. I had an opportunity to be trained in laboratory techniques as well as data analysis relevant to my research project.

I am submitting this dissertation in fulfilment of the requirements for the degree of MSc in Medicine (MSc [Med]) in Medical Microbiology offered by the Division of Medical Microbiology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa.

This thesis (approved by Human Research Ethics Committee [HREC]) of the University of Cape Town, South Africa, HREC 650/2018), was conducted under the supervision of Dr Mamadou Kaba and Dr Lemese Ah Tow. The study is part of the Drakenstein Child Health Study (DCHS) and was supported by the Bill and Melinda Gates Foundation Global Health Grant (OPP1017641), the National Institutes of Health (NIH) of the United States of America (USA) through the H3Africa initiative (1U01AI110466–01A1), the National Research Foundation (NRF) of South Africa, the Carnegie Corporation of New York (USA), the Wellcome Trust (102429/Z/13/Z) of the United Kingdom (UK) and the CTN-CIHR Canadian HIV Trials Network (Canada).

The work presented in this thesis result from a collaborative effort between the Division of Medical Microbiology and the Division of Human Genetics, University of Cape Town, South Africa; as well as the Department of Health Research Methods, Evidence and Impact (HEI), McMaster University, Canada.

The samples used in this study were collected by the personnel of the DCHS, South Africa. Dr. Mamadou Kaba, Dr. Lemese Ah Tow and the MSc candidate were involved in the formulation of the research study. Dr. Ah Tow supervised the wet-laboratory work, as well as the bioinformatics analysis of the bacterial sequencing results of this study. Ms. Michelle Ngwarai performed the Deoxyribonucleic acid (DNA) extraction, and initial database management. Dr. Veronica Allen assisted with the real-time polymerase chain reaction (PCR) work as well as the analysis of bacterial sequences. Ms Faith M Lutomia assisted with the systematic review process. Dr. Kaba and Dr. Ah Tow supervised and contributed to the manuscript edits of the MSc thesis. The MSc candidate performed all the wet and dry laboratory experiments. In addition, the MSc candidate carried out the data analysis with the guidance from Dr. Gaston Kuzamunu and Mr Denis Awany (Division of Human Genetics, University of Cape Town), as well as Prof. Lehana Thabane (Department of HEI, McMaster University, Canada).

The manuscript was written by the MSc candidate. Unless stated by the acknowledgement, all the research work described in this thesis is the work of the MSc candidate.

Acknowledgements

I would like to significantly appreciate the support, advice, positive attitude and guidance from my supervisors Dr. Mamadou Kaba and Dr. Lemese Ah Tow. Their continuous input helped me through the entire course of my MSc studies.

I also wish to extend my sincere gratitude towards Prof. Mark P Nicol for his endless inputs and guidance throughout all my MSc work.

Many thanks to Prof. Heather Zar for granting me an opportunity to work with the Drakenstein Child Health Study (DCHS) and for sponsoring me through the National Research Foundation (NRF), South Africa.

I am extremely thankful to Prof. Lehana Thabane, Dr. Gaston Kuzamunu and Mr Denis Awany for their help and guidance through the statistical data analysis component of my thesis.

I am very thankful to my funders the German Academic Exchange (DAAD) and the Namibia Student Financial Assistance Fund (NSFAF), as well as the Division of Medical Microbiology for financial assistance.

I highly appreciate the advice and courageous words from Dr. Veronica Allen, Dr Bosco Kalulelo, Ms Fadheela Patel, Ms Faith M Lutomia, aunt Pamela Jordan and aunt Doris Moiloa. Also, thanks to my office colleagues and the entire Medical Microbiology family for the care and friendliness they showed me.

My outmost gratitude goes to my elder brother, my parents and the entire family for always believing in me as well as for their unconditional love, encouragements and prayers throughout all my studies. Finally, thanks to all my friends and everyone who was there for me during this whole time.

Above all I thank the Almighty Father for everything. It was all Him!

Dedication

This piece of work is dedicated to my dear Grandmother. The love and compassion she raised me with shaped me into what I am today.

Table of Contents

Summary 10
Part 1: General introduction 11
Chapter 1: 12
General Introduction 12
References cited 12
Part 2: Systematic review 13
Chapter 2: 14
Contribution of *Faecalibacterium*, *Lachnospira*, *Veillonella,* and *Rothia* in human respiratory diseases: a systematic review 14
Summary 14
1. Introduction 15
1.1 Background and rationale 15
2. Materials and methods 15
2.1 Definition of key terms in this review 15
2.2 Inclusion and exclusion criteria 15
2.3 Search strategy 16
2.4 Outcome measures 16
2.5 Study selection and quality assessment 16
2.6 Data extraction and management 16
3. Results 18
3.1 Study selection and quality assessment 18
3.2 Characteristics of the included studies 19
3.3 FLVR as potential pathogens in respiratory diseases 19
3.4 FLVR as potential protector against respiratory diseases 19
3.5 Opposing shift of FLVR in respiratory diseases 20
Discussion 24
References cited 26
Part 3: FLVR bacteria and infant wheezing 29
Chapter 3: 30
Optimization of molecular detection of *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* 30
Summary 30
1. Introduction 30
1.1 Background and rationale 30
2. Materials and methods 31
2.1 Template DNA and PCR primers 32
2.2 In-silico analysis of the primers 33
2.3 Annealing temperature optimization and PCR conditions 34
2.3.1 Real-time PCR (qPCR) 34

2.3.2 Conventional PCR 34
2.4 Sensitivity assessment of the PCR assays 35
2.5 Repeatability and reproducibility of the qPCR assays 36
2.6 Statistical analysis of the data 36
3. Results 37
3.1 qPCR assays 37
3.1.1 In-silico analysis of qPCR primers 37
3.1.2 Annealing temperature and optimal PCR conditions 37
3.1.3 In-vitro specificity of qPCR primers 37
3.1.4 Sensitivity of the qPCR assays 38
3.1.5 Repeatability and Reproducibility of the qPCR assays 38
3.2 Conventional PCR assays 40
3.2.1 In-silico analysis of the primers 40
3.2.2 Annealing temperature and optimal PCR conditions 40
3.2.3 In-vitro specificity of the primers 40
3.2.4 Sensitivity of conventional PCR assays 40
Discussion 42
References cited 43
Chapter 4: 45
Association of *Faecalibacterium, Lachnospira*, *Veillonella*, and *Rothia* with childhood wheezing in the Drakenstein Child Health Study, South Africa 45
Summary 45
1. Introduction 46
1.1 Background and rationale 46
1.2 Hypothesis 48
1.3 Specific objectives 48
2. Materials and methods 48
2.1 Ethical consideration 49
2.2 Study design and population 49
2.3 Participants selection 50
2.4 Infant stool samples and metadata variables 50
2.5 Study outcomes 50
2.6 Application of optimized PCR conditions for the detection of FLVR bacteria in human stool samples from our cohort 50
2.6.1 Stool samples for analysis 50
2.6.2 Detection of FLVR bacteria from stool samples using qPCR and conventional PCR 51
2.6.3 Sequencing and bioinformatics analysis of conventional PCR products 51
2.7 Statistical analysis of the data 51
2.7.1 Characteristics of the cohort 51
2.7.2 Factors associated with gut colonisation by FLVR bacteria 52
2.7.3 Association of FLVR bacteria with wheezing and recurrent-wheezing outcomes..52

2.7.4 Sensitivity of FLVR bacteria in predicting the wheezing outcome 52
3. Results 53
3.1 Characteristics of the study participants 53
3.2 Detection of FLVR bacteria in human stool samples 55
3.2.1 Performance of qPCR primers in the detection of FLVR from stool samples in our cohort 55
3.2.2 Longitudinal detection of FLVR bacteria 55
3.2.3 FLVR prevalence in human stool samples 58
3.2.4 FLVR species in birth and six weeks samples of wheezing and non-wheezing infants 58
3.2.5 Factors associated with gut colonisation by FLVR bacteria 61
3.3 Association of FLVR bacteria with wheezing and recurrent-wheezing outcomes 63
3.3.1 Wheezing outcome 63
3.3.2 Recurrent-wheezing outcome 67
3.4 Sensitivity of FLVR bacterial load within the first four days of life in predicting the wheezing outcome 70
3.4.1 Each FLVR bacterium alone 70
3.4.2 Two by two interactions of FLVR bacteria 70
3.4.3 Three by three interactions of FLVR bacteria 70
3.4.4 Four by four interactions of FLVR bacteria 71
3.5 Sensitivity of FLVR bacteria at six weeks, six months and 12 months in predicting any wheeze 71
3.5.1 Each bacterium alone 71
3.5.2 Two by two interactions of FLVR bacteria 71
3.5.3 Three by three interactions of FLVR bacteria 71
3.5.4 Four by four interactions of FLVR bacteria 72
Discussion 78
References cited 83
Supplementary Tables 86
Supplementary figures 91
Part 4: General discussion and conclusion 94
Chapter 5: 95
General discussion 95
References cited 97
Chapter 6: 99
General conclusion 99
Part 5: Appendices 100
Appendix A: Curriculum vitae of Saara Kanyemba 101
Appendix B: Ethical approval of the study 102
Appendix C: Amendment of the study into the DCHS 103

Summary

Wheezing symptoms among children, present major health and economic problems globally. A recent study conducted in Canada observed a reduction in stool bacterial genera *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* (FLVR) in three-month old infants with atopy-wheeze symptoms. It is not known whether this is true in different human populations worldwide. The overall aim of this dissertation was to investigate the contribution of any of the FLVR bacteria or their combination in the occurrence of infant wheezing within the Drakenstein Child Health Study (DCHS), South Africa.

To address this aim, I began my thesis's project by conducting a systematic literature review which investigated the association of FLVR bacteria with the occurrence of different respiratory diseases in humans. My review provided evidence for the possible involvement of FLVR bacteria in human respiratory diseases, including asthma, pulmonary tuberculosis and pneumonia. Furthermore, this review highlighted the need for a well-designed and large study to investigate the contribution of the FLVR bacteria in respiratory diseases, in an African setting.

Secondly, I optimized SYBR Green based real-time quantitative polymerase chain reaction (qPCR) as well as conventional PCR assays for the detection of FLVR bacteria. Using the optimized assays, I screened 533 stool samples collected from 140 wheezing and 140 non-wheezing infants. The optimized assays demonstrated good performance in the detection of FLVR bacteria from human stool samples. Using qPCR, *Rothia*, *Veillonella, Faecalibacterium* and *Lachnospira* were detected in 90% (479/533), 73% (388/533), 51% (274/533) and 14% (77/533) of the samples, respectively. Conventional PCR permitted the detection of *Rothia*, *Veillonella, Faecalibacterium* and *Lachnospira* in 55% (263/479), 74% (289/388), 53% (145/274) and 0% (0/77) of the qPCR positive samples, respectively. I also determined the factors associated with faecal colonization by FLVR bacteria in the first year of life. I showed that reduced colonisation by *Faecalibacterium* was associated with male gender (adjusted OR = 0.65, 95%CI: 0.42 - 0.98) and TC-Newman residence (adjusted OR = 0.52, 95%CI: 0.29 - 0.91). Breastfeeding was associated with less colonisation by both *Lachnospira*, (adjusted OR = 0.17, 95%CI: 0.05 - 0.49) and *Veilonella* (adjusted OR = 0.32, 95%CI: 0.10 - 0.91). Mother's tertiary education was significantly associated with high *Rothia* colonisation (adjusted OR = 11.73, 95%CI: 1.36 - 2.58).

In the last section of my thesis, I assessed the association of FLVR bacteria with infant wheezing using logistic regression models. I found a significant association of *Rothia* with reduced risk of infant wheezing (adjusted odds ratio (aOR)=0.54, 95%CI: 0.28-0.93) and recurrent wheezing (aOR=0.29, 95%CI: 0.05-0.88). Using receiver operating characteristic curves (ROC), I showed that among all FLVR bacteria, *Lachnospira* (AUROC = 0.833, 95%CI: 0.64-1.00) and *Rothia* (AUC=0.707, 95%CI: 0.62-0.79) could serve as biomarkers for early prediction of infant wheezing.

Overall, this is the first study on FLVR bacteria and infant wheezing to be conducted in Africa. Its findings encourage more research to be conducted in order to elucidate the potential protective role of *Rothia* against childhood wheeze and asthma, as well as the contribution of *Lachnospira* in asthma development.

Part 1: General introduction

Chapter 1:

General Introduction

Childhood wheeze is a burden in most populations worldwide [1-3], often associated with respiratory tract infections' organisms [4, 5]. An emerging factor associated with its occurrence include the contribution of the gastrointestinal microbiome, in particular the FLVR (*Faecalibacterium, Lachnospira, Veillonella* and *Rothia*) bacteria [6]. Only a single study conducted in Canada demonstrated the protective role of FLVR bacteria against childhood wheeze [6]. Authors reported reduced abundance of FLVR bacteria in stools of atopic-wheeze infants compared to controls [6]. It is not known whether these findings apply in all infants globally, as well as in non-atopic wheeze. The overall aim of this thesis was to elucidate the contribution of FLVR bacteria in childhood wheeze, using infants' stools samples from a South African birth cohort study.

This thesis begins with a systematic literature review (Part 2) which highlights findings on the association of FLVR bacteria with different respiratory diseases in humans. This is followed by a report of the findings of this MSc research project (Part 3), consisting of Chapter 3 and Chapter 4. The contents of these chapters were reported according to the STROBE (Strengthening the Reporting of Observational studies in Epidemiology) guideline for reporting observational studies [7]. Chapter 3 gives a detailed description of the optimization of the molecular based techniques undertaken to detect FLVR bacteria from human stool samples. Chapter 4 presents the major findings of this thesis on the association of FLVR bacteria with childhood wheezing. Part 4 consists of Chapter 5 which is the general discussion and Chapter 6 which gives the general conclusion of this study. The last part (Part 5) is the appendix. The references cited in each chapter are presented at the end of the chapter

References cited

1. Alvarez-Alvarez I, Niu H, Guillen-Grima F, Aguinaga-Ontoso I. Meta-analysis of prevalence of wheezing and recurrent wheezing in infants. Allergologia et immunopathologia. 2016.
2. Stevens C, Turner D, Kuehni C, Couriel J, Silverman M. The economic impact of preschool asthma and wheeze. European Respiratory Journal. 2003;21(6):1000-6.
3. Teijeiro A, Badellino H, Raiden M, Cuello M, Kevorkof G, Gatti C, et al. Risk factors for recurrent wheezing in the first year of life in the city of Córdoba, Argentina. Allergologia et immunopathologia. 2017;45(3):234-9.
4. McIntosh K, Ellis EF, Hoffman LS, Lybass TG, Eller JJ, Fulginiti VA. The association of viral and bacterial respiratory infections with exacerbations of wheezing in young asthmatic children. The Journal of pediatrics. 1973;82(4):578-90.
5. Daian CM, Wolff AH, Bielory L, editors. The role of atypical organisms in asthma. Allergy and asthma proceedings; 2000: OceanSide Publications.
6. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152.
7. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS medicine. 2007;4(10):e296.

Part 2: Systematic review

Chapter 2:

Contribution of *Faecalibacterium*, *Lachnospira*, *Veillonella,* and *Rothia* in human respiratory diseases: a systematic review

Summary

Introduction: Respiratory diseases are a major universal public health concern among humans. Alterations in the human microbiota has been linked to several diseases. In Canada, a reduced abundance of *Faecalibacterium, Lachnospira*, *Veillonella* and *Rothia* (FLVR) was associated with increased risk of developing childhood asthma. We conducted a systematic review to identify and examine studies associating FLVR bacteria with the occurrence of different respiratory diseases in humans.

Methods and analysis: We performed a comprehensive literature search on six biomedical electronic databases/platforms (PubMed, Scopus, Web of Science, CINAHL, Africa Wide Information and Academic search premier) to identify studies reporting on FLVR bacteria and respiratory diseases. The screening and data extraction were carried out by two independent reviewers.

Results: We identified 17 eligible studies. Nine studies (52.9%) reported on *Veillonella*, two (11.8%) on *Lachnospira*, two (11.8%) on *Faecalibacterium,* three (17.6%) on *Veillonella* and *Rothia,* and one (5.9%) on all four bacteria. All the studies were conducted outside Africa between 2012 and 2018. Next generation sequencing of the 16S rRNA gene was the main method used (76.5%; 13/17) to detect FLVR bacteria. The samples used to detect FLVR bacteria included stools (41.2%; 7/17), oropharyngeal swabs (29.4%; 5/17), sputa (11.8%; 2/17), throat swabs (5.9%; 1/17), bronchoalveolar lavage (5.9%; 1/17) and mucus from the middle meatus (5.9%; 1/17). No study investigated factors associated with the detection of FLVR bacteria. Seven studies (7/17;41.2%) identified a potential protective role, and seven other studies (7/17;41.2%) identified a potential pathogenic role of FLVR bacteria in different respiratory diseases. Two studies (2/17;11.8%) observed an opposing shift of *Veillonella* and *Rothia* in pneumonia. *Lachnospira* was protective against asthma and pulmonary tuberculosis. *Faecalibacterium* was protective against cystic fibrosis and pathogenic against pulmonary tuberculosis. The protective role of *Veillonella* was observed in non-infectious wheezing, bronchiolitis, asthma and chronic obstructive pulmonary disease (COPD). The pathogenic role of *Veillonella* was observed in chronic rhinosinusitis, cystic fibrosis, bronchial asthma, COPD and idiopathic pulmonary fibrosis (IPF). One study reported a potential protective role of all four bacteria against asthma.

Conclusion: There is evidence for the possible involvement of FLVR bacteria in human respiratory diseases. Research on FLVR and respiratory diseases is needed in Africa. Future studies should also identify factors associated with the detection of FLVR bacteria in respiratory diseases.

Key words: Asthma, *Faecalibacterium*, FLVR, *Lachnospira*, Respiratory diseases, *Rothia*, *Veillonella*

1. Introduction

1.1 Background and rationale

Respiratory diseases are one of the main causes of morbidity and mortality in humans world-wide [1-5]. In 2015 lower respiratory tract infections (LRTIs) were the third leading killer globally [6]. In the same year, chronic obstructive pulmonary disease (COPD) claimed approximately 3.2 million deaths [6].

Development and/or progression of different human diseases is increasingly linked with dysbiosis of the microbial communities [7-9]. A Canadian study reported that reduction in stool bacterial genera, *Faecalibacterium Lachnospira*, *Veillonella* and *Rothia* (FLVR) was associated with high risk of developing asthma (based on the asthma predictive index) in three-month old infants with atopy-wheeze symptoms [10]. Addition of these bacteria in a humanized mouse model showed a decrease in airway inflammation [10, 11]. This suggests that reduction in these bacteria could be associated with the development of asthma and airway inflammation.

There is no published systematic review assessing the contribution of FLVR bacteria in human diseases. We conducted this systematic review to investigate the association of FLVR bacteria with the development and/or progression of different respiratory diseases in humans.

2. Materials and methods

This systematic review is registered (registration number, CRD42018081039) with the International Prospective Register of Systematic Reviews (PROSPERO). We adopted the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guideline[12], and in combination with the MOOSE guideline for the reporting of systematic reviews of observational studies [13].

2.1 Definition of key terms in this review

Respiratory diseases - All disorders affecting the organs involved in gaseous exchange [14].

Positive association - Increased abundance of a bacterium in samples of respiratory disease patients as compared to the control group, hence positively associating with the disease.

Negative association - Decreased abundance of the bacterium in samples of respiratory disease patients as compared to the control group, hence negatively associating with the disease.

Potential pathogen - A bacterium that is abundant in patients compared to the control group. This bacterium is not expected to be mono-infecting the patient, hence the term potential.

Potential protector - A bacterium that is abundant in the control group compared to patients.

2.2 Inclusion and exclusion criteria

In this systematic review, we included studies comparing the microbiota of patients suffering from respiratory diseases to "healthy" or non-respiratory disease individuals. In addition, we considered studies which have used FLVR bacteria as a bio-therapeutic in respiratory diseases. We excluded studies reporting on diseases other than respiratory diseases as an outcome. Articles written in other languages besides English and French were excluded.

2.3 Search strategy

To identify potentially relevant studies, we performed a comprehensive literature search on biomedical electronic databases/platforms (PubMed, Scopus, Web of Science, CINAHL, Africa Wide Information and Academic search premier). The search was done using a combination of key terms (Table 2.1). In addition, we also searched google scholar (Table 2.1) for potentially relevant articles which would have been missed by database searching. All the studies were retrieved into an endnote library.

2.4 Outcome measures

The outcome of interest included the occurrence of a respiratory disease condition, associated with any or a combination of the FLVR bacteria.

2.5 Study selection and quality assessment

Two independent reviewers were involved in the study selection and quality assessment. All retrieved studies were initially screened based on their titles and abstracts to identify potentially eligible studies. Instances where the study design was not clearly mentioned in the abstracts, the full article was retrieved for a closer inspection. Each reviewer read through the full texts of potentially eligible studies, to identify eligible studies. Disagreements were resolved by a consensus between the two reviewers or a contribution of a third reviewer. The Joanna Briggs Institute (JBI) critical appraisal checklists [15] and the Newcastle-Ottawa scale (NOS) [16] were used to assess quality of eligible studies. The eligible studies were required to meet all components of the NOS and component 1-8 and 10 of the JBI checklist.

2.6 Data extraction and management

We used a questionnaire form generated in this study to extract relevant data from the eligible studies. The following information were independently extracted by two reviewers: study author, year of publication, sample size, detection method, detected FLVR bacterial genera, disease name, sample type, population age group, country, bacterial abundance, odds ratio if provided, and authors conclusion. All the information was saved in a table format (Table 2.2)

Table 2.1 Search strategy performed in electronic databases/platforms

Database/platforms	Search mode	Keywords
Medline via Pubmed	All fields	(*Faecalibacterium* OR *Lachnospira* OR *Veillonella* OR *Rothia*) AND ("lung disease" OR "lung diseases" OR "pulmonary disease" OR "pulmonary diseases" OR "respiratory diseases" OR "respiratory disease" OR "respiratory tract infections" OR "respiratory tract infection" OR "respiratory disorders" OR "respiratory disorder" OR asthma OR pneumonia OR "COPD" OR bronchitis OR emphysema OR bronchiectasis OR "pleural effusion" OR wheeze OR wheezing OR tuberculosis OR "chronic obstructive pulmonary disease" OR "chronic cough" OR "acute respiratory distress syndrome" OR "ARDS" OR bronchiolitis OR sarcoidosis)
Scopus via SciVerse*	Article title, abstract, keywords	
Academic Search Premier, Africa-Wide Information and CINAHL via EBSCOHost	Boolean/Phrase	
Web of Science via Web of Knowledge	Topic	
Google scholar		Same search terms as for the other databases (Bacterium name AND respiratory diseases), e.g. (*Faecalibacterium* AND respiratory diseases) (All bacterial names AND respiratory diseases), e.g. (*Faecalibacterium*, *Lachnospira*, *Veillonella*, *Rothia* AND respiratory diseases)

* We used the advanced search strategy

3. Results

3.1 Study selection and quality assessment

Using our search terms, 910 articles were retrieved from PubMed, Scopus, Web of Science, CINAHL, Africa Wide Information and Academic search premier, and 1500 from google scholar (Figure 2.1). From these, 1661 duplicate articles were removed. From the remaining 749 studies assessed based on their titles/abstracts, 712 studies were excluded as they did not meet the inclusion criteria. The remaining 37 potentially eligible studies underwent full text review and quality assessment using the JBI [15] and NOS [16] checklists. After quality assessment, we excluded 20 studies and 17 studies were eligible (Figure 2.1).

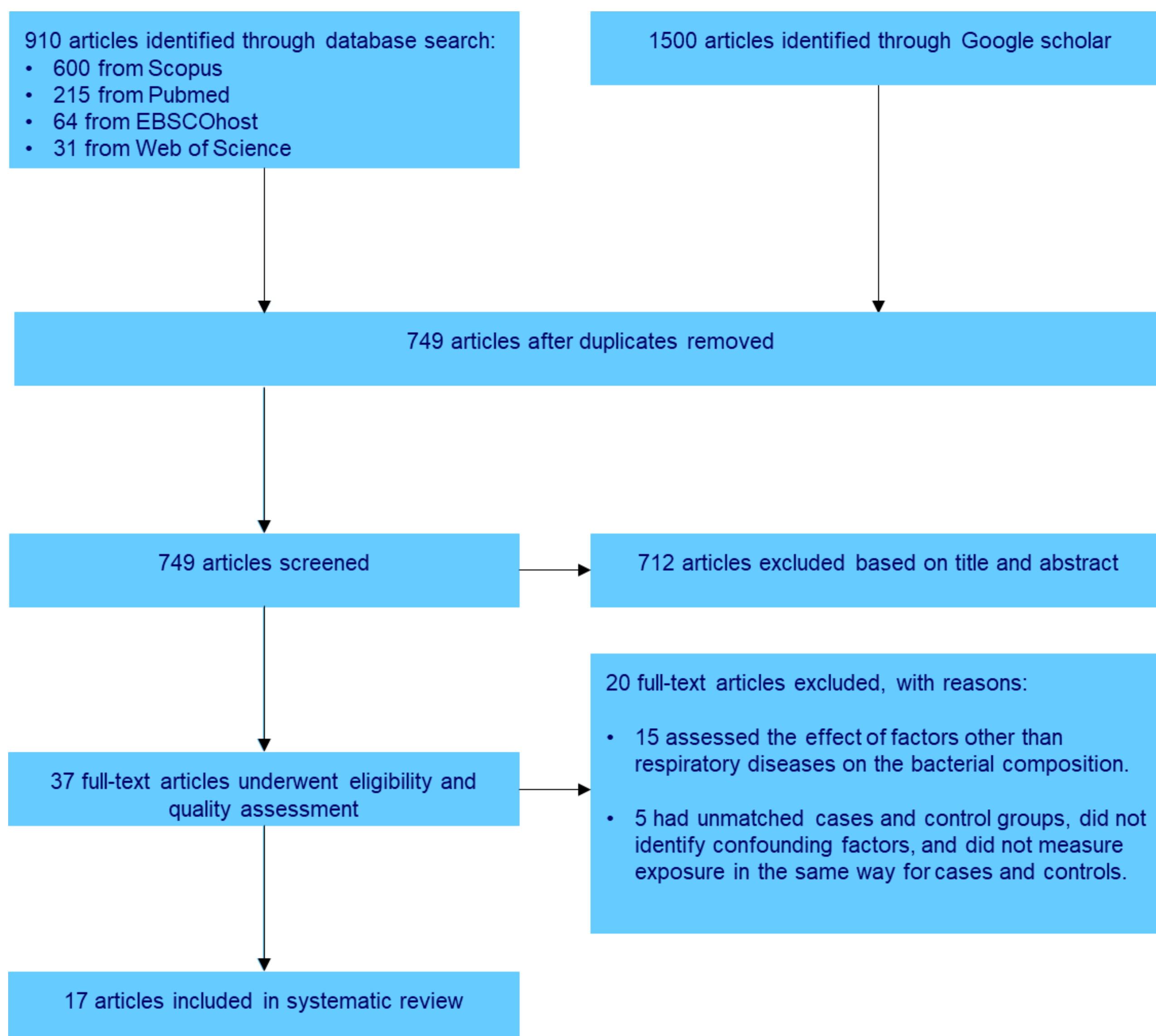


Figure 2.1 PRISMA flow diagram summarizing the study selection process

3.2 Characteristics of the included studies

Of the 17 eligible studies, 41.2% (7/17) were on childhood (<18 years) respiratory diseases. These diseases were, asthma (10, 17), cystic fibrosis (18, 19), non-infectious wheezing (20), atopic wheezing (21) and bronchiolitis (22). The remaining 58.8% (10/17) of the studies were focused on adults (≥18 years) respiratory diseases. These included chronic rhinosinusitis (23), pneumonia (24, 25), asthma (26, 27), chronic obstructive pulmonary disease (COPD) (27, 28), pulmonary tuberculosis (29-31) and idiopathic pulmonary fibrosis (IPF) (32). Most studies (76.5%; 13/17) (18-20, 22-25, 27-32) used 16S rRNA next generation sequencing as the detection method, 3/17 (11.8%) (17, 21, 33) studies used 16S rRNA qPCR and 1/17 (5.9%) (26) study used cloning combined with Sanger sequencing of the 16S rRNA gene. FLVR bacteria were detected in oropharyngeal swabs (29.4%; 5/17) (20, 25-28), sputa (11.8%; 2/17) (24, 29), throat swabs (5.9%; 1/17)(18), mucus from the middle meatus (5.9%; 1/17) (23), bronchoalveolar lavage (5.9%; 1/17) (32) and stool samples (41.2%; 7/17) (17, 19, 22, 30, 31, 33). From the 17 studies, *Veillonella* was detected in 52.9% (9/17) (18, 20-23, 26-28, 32), *Faecalibacterium* in 11.8% (2/17) (19, 31), *Lachnospira* in 11.8% (2/17) (17, 30), both *Veillonella* and *Rothia* in 17.6% (3/17) (24, 25, 29), and 5.9% (1/17) detected all four bacteria (33). None of the studies identified risk factors associated with the detection of FLVR bacteria. In addition, none of the included studies was conducted in Africa. No study gave the number of FLVR positive or negative participants. Furthermore, most studies had a very small sample size (Table 2.2).

3.3 FLVR as potential pathogens in respiratory diseases

A total of seven out of the 17 (41.2%) included studies reported a positive association of either one of the FLVR bacteria or more with respiratory diseases (Table 2.2). Six of these studies, using respiratory samples as the sample source, identified a potential pathogenic role of *Veillonella* in chronic rhinosinusitis (23), cystic fibrosis (18), bronchial asthma (26), COPD (28) and idiopathic pulmonary fibrosis (IPF) (32), as well as in atopic wheeze using stool as the sample source. One study identified *Faecalibacterium* in stools as a potential pathogen in pulmonary tuberculosis (31). One study associated both *Rothia* and *Veillonella* in sputum as potential pathogens in pulmonary tuberculosis (29).

3.4 FLVR as potential protector against respiratory diseases

We identified 41.2% (7/17) studies reporting the protective role of FLVR bacteria against the occurrence of respiratory diseases (Table 2.2). *Veillonella* was associated with protection against non-infectious wheezing (1/7 studies using oropharyngeal swabs) (20), bronchiolitis (1/7 studies using stool samples) (22) as well as asthma and COPD (1/7 studies using oropharyngeal swabs) (27). Stools' *Lachnospira* was associated with protection against pulmonary tuberculosis (1/7 studies) (30) and asthma (1/7 studies) (17). *Faecalibacterium* in stools was protective

against cystic fibrosis in 1/7 studies [19]. We identified 1/7 studies reporting the protective role of all four bacteria, detected in stools, against asthma [33].

3.5 Opposing shift of FLVR in respiratory diseases

We observed an opposing shift between *Veillonella* and *Rothia* in 11.8% (2/17) of the included studies. A decrease in *Veillonella* and an increase in *Rothia* abundance was observed in sputum and oropharyngeal swabs of pneumonia patients [24, 25] (Table 2.2).

Table 2.2 A summary of studies associating FLVR bacteria with respiratory diseases in humans

Study author/year	Study design	Number of participants	Age	Gender	Country	Disease	Sample type	Detection method	Detected bacteria	Relative abundance	Odds ratio (CI)	Association with disease	Authors conclusion
Biswas *et al.* 2017	Case control	Cases= 10 Controls= 9	≥18	Cases: 5M, 5F Controls: 6M, 3F	New Zealand	Chronic rhinosinusitis	Mucus from the middle meatus	16S rRNA NGS	*Veillonella*	_	_	Positive	Strongly associated with the disease (absolute effect size < -3)
Boutin *et al.* 2017	Case control	CF= 57 Asthma= 27 Controls= 62	Average age: 10.2 years	Cases: 46M, 6F Controls: 28M, 34F	Germany	Cystic fibrosis Asthma	Throat swabs	16S rRNA NGS	*Veillonella*	_	_	Positive	High abundance in CF as compared to asthmatic and healthy controls
Dang *et al.* 2013	Case control	Cases= 4 Controls= 4	≥18 years	Cases: 1M, 3F Controls: 3M, 1F	South Korea	Asthma	Oropharyngeal swabs	16S rRNA cloning and	*Veillonella*	Cases= 14% Controls= 5%	_	Positive	Dominant in asthmatic
Diao *et al.* 2017	Case control	Cases= 45 Controls= 20	Cases: 62.3±7.0 Controls: 58.7±6.4 years	Male	China	COPD	Oropharyngeal swabs	16S rRNA NGS	*Veillonella*	_	8.8 (4.6-13.4)	Positive	Significantly high abundance in cases compared to controls.
Krishna *et al.* 2016	Case control	Cases= 25 Controls= 16	Cases= 23-67 years Controls= 21-70 years	Cases: 13M, 12F Control: 10M, 6F	Germany	Pulmonary Tuberculosis	Sputum	16S rRNA NGS	*Veillonella* *Rothia*	*V. dispar* Cases= 3.3%, Controls=0.8%*Rothia mucilaginosa* Cases= 4%, Controls= 0%	_	Positive	*Veillonella* was highly dominant in cases (*P*= 0.023) *Rothia* (*P*= 0.047) was detected only in cases
Maji *et al.* 2018	Case control	Cases= 6 Controls= 6	Cases= 14-47 years Controls= 19-35 years	Not reported	India	Pulmonary Tuberculosis	Stool	16S rRNA NGS	*Faecalibacterium*	Cases= 0.03 Controls= 0.006	_	Positive	Significantly enriched in TB cases at day zero (*P*= 0.016) compared to controls

Table 2.2 continues

Study author/year	Study design	Number of participants	Age	Gender	Country	Disease	Sample type	Detection method	Detected bacteria	Relative abundance	Odds ratio (CI)	Association with disease	Authors conclusion
Molyneaux *et al.* 2014	Case control	Cases= 65 Controls= 44	Average age: Cases= 68 Controls= 58.2	Cases: 50M, 15F Controls: 27M, 17F	United Kingdom	IPF	Bronchoalveolar lavage	16S rRNA NGS	*Veillonella*	Cases= 84.8±5.7 Controls= 56.6±4.5	_	Positive	Highly detected in IPF, 1.5-fold increase in cases compared to controls
Chen *et al.* 2013	Case control	Cases= 45 Controls= 18	Cases= 42-78 years Controls= Not reported	Cases: 24M, 21F Controls: not reported	China	Pneumonia	Sputum	16S rRNA NGS	*Rothia*	CAP= 11.35% Controls= 1.04%	_	Positive	*Rothia* more prevalent in CAP cases. Might be an endogenous-disease causing factor.
									Veillonella	HAP= 0.86% Controls= 3.62%	_	Negative	*Veillonella* more abundant in controls compared to HAP cases
Piters *et al.* 2016	Case control	Cases= 127 Controls= 278	Cases= 46-76 years Controls= 34-75 years	Cases: 76M, 51F Controls: 68M, 210F	Netherlands	Pneumonia	Oropharyngeal swabs	16S rRNA NGS	*Rothia*	Cases= 26.7% Controls= 7.3%	_	Positive	*Rothia* more abundant in cases compared to controls.
									Veillonella dispar	_	_	Negative	*V. dispar* was more prevalent in controls,
Arrieta *et al.* 2015	Case control	Initial 319 (Cases=245, Controls=74)	3 months and 1 year	Cases: 14M, 7F Controls: 36M, 38F	Canada	Asthma Preceded by atopic-wheezing	Stool	Initial 16S rRNA NGS	*Faecalibacterium* *Lachnospira* *Veillonella* *Rothia*	_	_	Negative	Low abundance at 3 months is associated with high risk of asthma diagnosis by 3 years
		Subsequent: Cases= 21 Controls= 20 at 3months, 19 at 1 year						Subsequent 16S rRNA qPCR					
Cardenas *et al.* 2012	Case control	Cases= 24 Controls= 24	Average: Cases: 9.9 months Controls: 10.5 months	Cases: 12M, 12F Controls: 10M, 14F	United Kingdom	Non-infectious wheezing	Oropharyngeal swabs	16S rRNA NGS sequencing	*Veillonella*	Cases= 2623 sequences Controls= 4117 sequences	0.59 (0.56-0.62)	Negative	Significantly high abundance in controls (P= 8.06×10^{-86})

Table 2.2 continues

Study author/year	Study design	Number of participants	Age	Gender	Country	Disease	Sample type	Detection method	Detected bacteria	Relative abundance	Odds ratio (CI)	Association with disease	Authors conclusion
Hasegawa *et al.* 2016	Case control	Cases= 40 Controls= 115	Cases: 4 years Controls: 4 years	Cases: 18M, 22F Controls: 64M, 51F	United States	Bronchiolitis	Stool	16S rRNA NGS	*Veillonella*	_	_	Negative	Overrepresented in healthy infants (absolute effect size > 3.6)
Luo *et al.* 2017	Case control	NTB= 19 RTB= 18 Controls= 20	NTB= 14-75 years RTB= 15-77 years Controls= 25-66 years	NTB: 10M, 9F RTB: 15M, 3F Controls: 12M, 8F	China	Pulmonary Tuberculosis	Stool	16S rRNA NGS	*Lachnospira*	NTB= 0.9% RTB= 0.6% Controls= 2.7%	_	Negative	Significantly decreased in NTB and RTB compared to healthy controls
Park *et al.* 2014	Case control	Asthma= 18 COPD= 17 Controls= 12	Asthma= 53.4±17.1 COPD= 68.9±7.2 Controls= not reported	Asthma:10M,8F COPD: 15M,2F Controls: 6M, 6F	South Korea	Asthma, COPD	Oropharyngeal swabs	16S rRNA NGS	*Veillonella*	Asthma: 3.46% COPD: 3.98% Controls: 8.01%	_	Negative	Mostly abundant in healthy controls
Stiemsma *et al.* 2016	Case control	Cases= 39 Controls= 37	3 months and 1 year	Cases:21M,18F Controls: 20M, 17F	Canada	Asthma	Stool	16S rRNA qPCR	*Lachnospira*	_	_	Negative	Significantly reduced in 3months faecal microbiota of asthmatic (*P*= 0.008)
Vernocchi *et al.* 2017	Case control	Cases= 28 Controls= 31	Average: Cases= 3.5 years Controls= 3.06 years	Cases: 11M, 17F Controls: 20M, 11F	Italy	Cystic fibrosis	Stool	16S rRNA NGS	*Faecalibacterium*	_	_	Negative	Significantly abundant in controls ($P \leq 0.1$)
Arrieta *et al.* 2018	Case control	Cases=27 Controls=70	3 months	Cases= 13M, 14F Controls= 32M, 38F	Rural Ecuador	Atopic wheeze	Stool	16S rRNA NGS 16S rRNA qPCR	*Veillonella*	_	_	Positive	Significantly associated with atopic-wheeze

Abbreviations: rRNA- ribosomal RNA; NGS- Next generation sequencing; COPD- Chronic obstructive pulmonary disease; IPF- Idiopathic pulmonary fibrosis; CAP- Community acquired pneumonia; quantitative polymerase chain reaction; HAP- Hospital acquired pneumonia; NTB- New TB cases; RTB- Recurrent TB cases; M- Male; F- Female

Discussion

This is the first systematic review on FLVR bacteria and human respiratory diseases. We observed that most studies lack statistical evidence to support their findings. However, the studies which provided statistical evidence differ in terms of the detected bacteria, type of samples used, and the diseases investigated. As a result, we were unable to compare the findings between studies. We also could not calculate measures of associations such as odds ratio as none of the studies provided the number of participants which were positive or negative with FLVR bacteria. Because of the heterogeneity of the studies, we were unable to conduct a meta-analysis. We saw a need for larger studies, as the smaller sample sizes of most studies [17-20, 23, 24, 26-33] in this review may have biased the findings observed. The widespread use of NGS as the detection method observed in this review, indicates the need for a target-based approach in the detection of FLVR bacteria. The universal primers used in NGS may fail to capture some bacteria, mostly those which exist in low densities. Therefore, targeting FLVR bacteria using specific primers may help overcome this limitation. This will allow us to better understand the contribution of these bacteria in human diseases.

This review strongly suggests a potential protective role of *Lachnospira* as this bacterium was less abundant in stool samples of patients compared to controls. *Lachnospira* and other Firmicutes are well known producers of short-chain fatty acids (SCFAs) which have been shown to play a key role in immunity [34-36]. Hence, reduced amount of *Lachnospira* in the patients could mean an impaired immune response too. The impaired immune response could also mean poor health. This potential protective role strengthens the gut-lung axis hypothesis [37, 38], emphasizing the possible influence of gut *Lachnospira* on respiratory tract health. Findings on *Veillonella* and respiratory diseases are conflicting, this could be due to the use of different samples which could not permit comparability of results. *Faecalibacterium* has been associated with health in various diseases [39, 40]. We observed similar results with cystic fibrosis in this review. Therefore, the contradicting findings of *Faecalibacterium* with pulmonary tuberculosis also observed in this study require further investigation using bigger sample sizes for confirmation. In addition, we suggest the use of target approaches as these findings are based on microbiome profiling. An opposing shift of *Veillonella* and *Rothia* observed in sputa and oropharyngeal swabs of pneumonia patients in this review is interesting. Since these two sample types are shown to exhibit a similar bacterial composition [41], it is very likely that these bacteria are playing an interactive role in pneumonia. *Rothia* has been reported as a potential pathogen for pneumonia in another review [42], hence supporting our findings. Co-colonisation with FLVR bacteria could be therapeutic against asthma, given the simultaneous reduction of all four bacteria observed in one of the studies included in this review. This suggests a possible synergistic role played by these bacteria in asthma.

The main limitation of our study was exclusion of studies in languages other than French and English, which might have biased our findings. Furthermore, the data generated in microbiome studies could vary greatly due to various factors such as extraction protocols, the primers used, the sequencing platforms as well as bioinformatics pipelines used [(43)]. Therefore, comparison of studies with similar objectives and data derived from similar platforms could provide accurate information as it is very difficult to make final conclusions using data generated differently.

In conclusion, our review demonstrates evidence for the possible contribution of FLVR bacteria in different respiratory diseases. The gut *Lachnospira* could potentially be protective against asthma and pulmonary tuberculosis, thus strengthening the gut-lung axis hypothesis. Future studies on *Veillonella* using the same sample type are needed to help resolve the conflicting findings observed in this study. *Faecalibacterium* could be a potential pathogen in pulmonary tuberculosis, although studies have mostly associated this bacterium with health. We encourage future studies to investigate the role of *Faecalibacterium* in childhood asthma and cystic fibrosis. The opposing shift of *Rothia* and *Veillonella* in pneumonia should be studied further both in children and adults. It will also be interesting to look at the interaction among FLVR bacteria with other organisms in the progression of respiratory diseases in future.

Having explored literature for the possible contribution of the FLVR bacteria in different respiratory diseases, we see a gap for lack of studies on FLVR and respiratory diseases in an African setting. We also noticed a need for target-based detection of FLVR bacteria. We therefore encourage future target-specific studies based in Africa with larger sample sizes, to see if these findings are applicable worldwide. We further observed that none of the included studies identified the risk factors associated with FLVR detection, which is something worth looking into in the future. We noted that studies are reporting only on bacterial abundance and failing to mention how many participants had the bacteria. We thus suggest that future studies should put this into consideration, as it is very important for comparability purposes. Lastly, we encourage species level investigation to help identify the exact FLVR bacterial species which are involved in human diseases.

References cited

1. Ferkol T, Schraufnagel D. The global burden of respiratory disease. Annals of the American Thoracic Society. 2014;11(3):404-6.
2. Zar HJ, Ferkol TW. The global burden of respiratory disease—impact on child health. Pediatric pulmonology. 2014;49(5):430-4.
3. Cruz AA. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach: World Health Organization; 2007.
4. Mathers C. The global burden of disease: 2004 update: World Health Organization; 2008.
5. Mizgerd JP. Acute lower respiratory tract infection. New England Journal of Medicine. 2008;358(7):716-27.
6. World Health Organization. The top 10 causes of death. 2017.
7. Huang YJ, Marsland BJ, Bunyavanich S, O'mahony L, Leung DY, Muraro A, et al. The microbiome in allergic disease: Current understanding and future opportunities—2017 PRACTALL document of the American Academy of Allergy, Asthma & Immunology and the European Academy of Allergy and Clinical Immunology. Journal of Allergy and Clinical Immunology. 2017;139(4):1099-110.
8. Zhang Y-J, Li S, Gan R-Y, Zhou T, Xu D-P, Li H-B. Impacts of gut bacteria on human health and diseases. International journal of molecular sciences. 2015;16(4):7493-519.
9. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. Cell. 2012;148(6):1258-70.
10. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152-307ra152.
11. Arrieta MC, Sadarangani M, Brown EM, Russell SL, Nimmo M, Dean J, et al. A humanized microbiota mouse model of ovalbumin-induced lung inflammation. Gut microbes. 2016;7(4):342-52.
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009;151(4):264-9.
13. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Jama. 2000;283(15):2008-12.
14. Alvarado A, Arce I. Metabolic functions of the lung, disorders and associated pathologies. Journal of clinical medicine research. 2016;8(10):689.
15. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, et al. Chapter 7: Systematic reviews of etiology and risk. Joanna Briggs Institute Reviewer's Manual The Joanna Briggs Institute. 2017.
16. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. European journal of epidemiology. 2010;25(9):603-5.
17. Stiemsma LT, Arrieta MC, Dimitriu PA, Cheng J, Thorson L, Lefebvre DL, et al. Shifts in *Lachnospira* and *Clostridium sp.* in the 3-month stool microbiome are associated with preschool age asthma. Clinical science (London, England : 1979). 2016;130(23):2199-207.

18. Boutin S, Depner M, Stahl M, Graeber SY, Dittrich SA, Legatzki A, et al. Comparison of Oropharyngeal Microbiota from Children with Asthma and Cystic Fibrosis. Mediators of inflammation. 2017;2017:5047403.
19. Vernocchi P, Del Chierico F, Quagliariello A, Ercolini D, Lucidi V, Putignani L. A Metagenomic and in Silico Functional Prediction of Gut Microbiota Profiles May Concur in Discovering New Cystic Fibrosis Patient-Targeted Probiotics. Nutrients. 2017;9(12).
20. Cardenas PA, Cooper PJ, Cox MJ, Chico M, Arias C, Moffatt MF, et al. Upper airways microbiota in antibiotic-naive wheezing and healthy infants from the tropics of rural Ecuador. PLoS One. 2012;7(10):e46803.
21. Arrieta M-C, Arévalo A, Stiemsma L, Dimitriu P, Chico ME, Loor S, et al. Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. Journal of Allergy and Clinical Immunology. 2018;142(2):424-34. e10.
22. Hasegawa K, Linnemann RW, Mansbach JM, Ajami NJ, Espinola JA, Petrosino JF, et al. The Fecal Microbiota Profile and Bronchiolitis in Infants. Pediatrics. 2016;138(1).
23. Biswas K, Mackenzie BW, Waldvogel-Thurlow S, Middleditch M, Jullig M, Zoing M, et al. Differentially regulated host proteins associated with chronic rhinosinusitis are correlated with the sinonasal microbiome. Front Cell Infect Microbiol. 2017;7(DEC).
24. Chen C, Shen T, Tian F, Lin P, Li Q, Cui Z, et al. New microbiota found in sputum from patients with community-acquired pneumonia. Acta biochimica et biophysica Sinica. 2013;45(12):1039-48.
25. De Steenhuijsen Piters WAA, Huijskens EGW, Wyllie AL, Biesbroek G, Van Den Bergh MR, Veenhoven RH, et al. Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients. ISME J. 2016;10(1):97-108.
26. Dang HT, Kim SA, Park HK, Shin JW, Park SG, Kim W. Analysis of oropharyngeal microbiota between the patients with bronchial asthma and the non-asthmatic persons. J Bacteriol Virol. 2013;43(4):270-8.
27. Park H, Shin JW, Park SG, Kim W. Microbial communities in the upper respiratory tract of patients with asthma and chronic obstructive pulmonary disease. PLoS One. 2014;9(10):e109710.
28. Diao W, Shen N, Du Y, Qian K, He B. Characterization of throat microbial flora in smokers with or without COPD. International journal of chronic obstructive pulmonary disease. 2017;12:1933-46.
29. Krishna P, Jain A, Bisen P. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. European Journal of Clinical Microbiology & Infectious Diseases. 2016;35(7):1205-10.
30. Luo M, Liu Y, Wu P, Luo DX, Sun Q, Zheng H, et al. Alternation of Gut Microbiota in Patients with Pulmonary Tuberculosis. Frontiers in physiology. 2017;8:822.
31. Maji A, Misra R, Dhakan DB, Gupta V, Mahato NK, Saxena R, et al. Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. Environmental microbiology. 2018;20(1):402-19.
32. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2014;190(8):906-13.
33. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152.

34. Canani RB, Di Costanzo M, Leone L. The epigenetic effects of butyrate: potential therapeutic implications for clinical practice. Clinical epigenetics. 2012;4(1):4.
35. Duncan SH, Belenguer A, Holtrop G, Johnstone AM, Flint HJ, Lobley GE. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. Applied and environmental microbiology. 2007;73(4):1073-8.
36. Brestoff JR, Artis D. Commensal bacteria at the interface of host metabolism and the immune system. Nature immunology. 2013;14(7):676.
37. Budden KF, Gellatly SL, Wood DL, Cooper MA, Morrison M, Hugenholtz P, et al. Emerging pathogenic links between microbiota and the gut–lung axis. Nat Rev Microbiol. 2017;15(1):55.
38. Marsland BJ, Trompette A, Gollwitzer ES. The gut–lung axis in respiratory disease. Annals of the American Thoracic Society. 2015;12(Supplement 2):S150-S6.
39. Miquel S, Martin R, Rossi O, Bermudez-Humaran L, Chatel J, Sokol H, et al. Faecalibacterium prausnitzii and human intestinal health. Current opinion in microbiology. 2013;16(3):255-61.
40. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux J-J, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proceedings of the National Academy of Sciences. 2008;105(43):16731-6.
41. Liu H-Y, Zhang S-Y, Yang W-Y, Su X-F, He Y, Zhou H-W, et al. Oropharyngeal and sputum microbiomes are similar following exacerbation of chronic obstructive pulmonary disease. Frontiers in microbiology. 2017;8:1163.
42. Maraki S, Papadakis IS. *Rothia mucilaginosa* pneumonia: a literature review. Infectious Diseases. 2015;47(3):125-9.
43. Chatellier S, Mugnier N, Allard F, Bonnaud B, Collin V, van Belkum A, et al. Comparison of two approaches for the classification of 16S rRNA gene sequences. Journal of medical microbiology. 2014;63(10):1311-5.

Part 3: FLVR bacteria and infant wheezing

Chapter 3:

Optimization of molecular detection of ***Faecalibacterium***, ***Lachnospira***, ***Veillonella*** and ***Rothia***

Summary

Background: There is rising interests in stool bacterial genera *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* (FLVR) as potential biomarkers for infant wheezing. Isolation of FLVR by culture-based methods is a challenge. Polymerase Chain Reaction (PCR) is a very sensitive, culture-independent method which is widely used for microbial detection in clinical samples. However, no single study has described optimal PCR conditions for FLVR detection thus far. This study aimed to optimize SYBR Green based real-time PCR (qPCR) assays (targeting the 16S rRNA), as well as conventional PCR assays (targeting the 16S rRNA and *rpoB* genes) for the detection of FLVR bacteria.

Method: We assessed the overall performance of previously published primers as well as newly designed primers and determined their optimal PCR conditions. Pure genomic DNA of FLVR bacteria was used for the optimisation of SYBR Green qPCR and conventional PCR assays. A thermal gradient PCR approach was used to determine the optimal annealing temperature for all primer pairs. We determined the primers' limit of detection (LOD), the inter-assay and intra-assay variation.

Results: All optimized conditions demonstrated good performance for the detection of FLVR bacteria. The optimal annealing temperature was 65 ^{0}C for all qPCR primers, and in the range 45 – 65 ^{0}C for conventional PCR primers. The qPCR assays were more sensitive (detection limit: 1.28 x 10^{-5}- 3.2 x 10^{-4} ng/µl) compared to conventional PCR assays (detection limit: 8 x 10^{-3} - 1 x 10^{0} ng/µl). The intra-assay %CV were in the range 0 - 1.63%, 0.24 - 1.25%, 0.34 - 2.67% and 0.21 - 6.78% for *Faecalibacterium, Lachnospira, Veillonella* and *Rothia* respectively. The inter-assay %CV for *Faecalibacterium, Lachnospira, Veillonella* and *Rothia* ranged between 0 - 3.73%, 1.21 - 5.54%, 0.73 - 3.1% and 0 - 7.06% respectively.

Conclusion: We consider our PCR assays suitable for the detection of FLVR bacteria.

Keywords: *Faecalibacterium*, *Lachnospira*, *Veillonella, Rothia,* FLVR, quantitative real-time PCR, optimization, SYBR green, conventional PCR

1. Introduction

1.1 Background and rationale

The stool bacterial genera *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* (FLVR) have been identified as potential diagnostic markers for infant atopic-wheeze [1]. A single study has

looked at these bacteria in infants with wheezing [1]. Therefore, there is a need for further research to help strengthen these findings. Detection of FLVR bacteria by traditional culture-based methods is challenging because of their strict growth requirements [2-5]. *Faecalibacterium* [3], *Lachnospira* [2] and *Veillonella* [4] are extremely oxygen sensitive (EOS) and *Rothia* [5] grows very slowly. To overcome the drawback of culture-based microbial detection, clinical research has shifted towards the use of culture independent molecular-based methods [6]. Polymerase chain reaction (PCR) is one of the widely used culture-independent methods for microbial detection with high sensitivity and specificity [7, 8]. Quantitative real-time PCR (qPCR) allows both the detection and quantification of the target nucleic acid [9], whereas conventional (end-point) PCR only allows the detection [10]. In qPCR, detection is facilitated by the use of fluorescent reporters such as hybridization probes [11, 12] and the DNA minor-groove intercalating dye SYBR green I [13]. These fluorescent reporters also monitor the amplification process [14, 15]. In contrast, detection by conventional PCR relies on visualization of a fraction of the amplicons under ultraviolet (UV) light after gel electrophoresis (that is, post PCR amplification) [10]. In addition, conventional PCR allows amplification of a larger DNA fragment for sequence identification.

There is no study which has described PCR conditions for the detection of FLVR bacteria. Establishment of successful PCR assays is an essential requirement, in ensuring reproducible and valid results [16]. The objective of this study was to establish optimal SYBR Green based qPCR as well as conventional PCR assays for the detection of FLVR bacteria.

2. Materials and methods

Figure 3.1 outlines the workflow of this study.

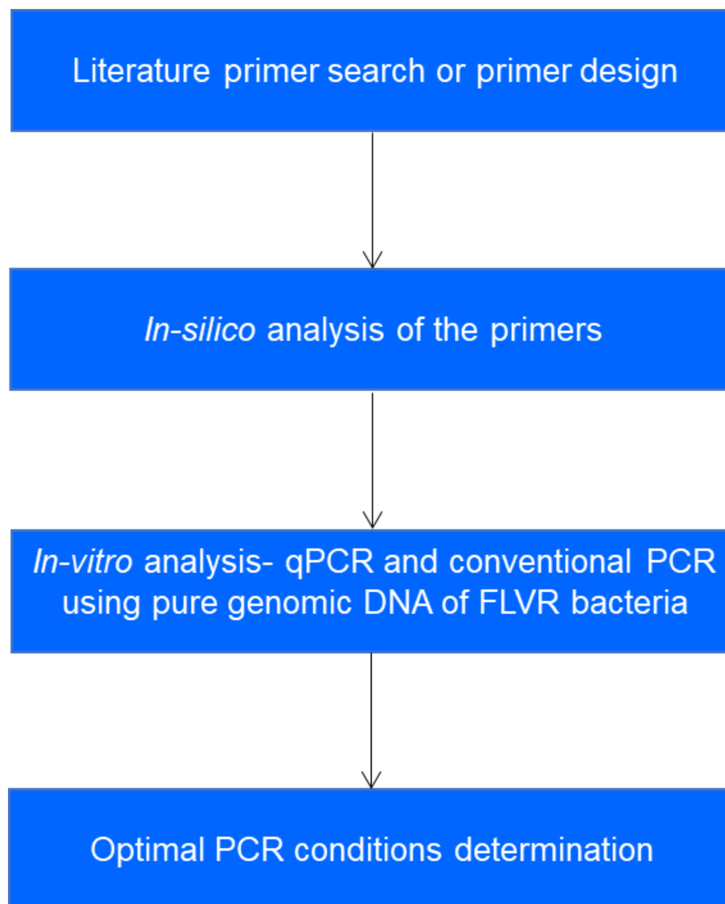


Figure 3.1 Study workflow showing the optimisation experiments for the molecular detection of FLVR bacteria. F- *Faecalibacterium*, L- *Lachnospira*, V- *Veillonella,* R- *Rothia;* DNA- Deoxyribonucleic acid; qPCR- Quantitative Polymerase Chain Reaction

2.1 Template DNA and PCR primers

We used pure genomic DNA from *Faecalibacterium prausnitzii* (DSM 17677), *Lachnospira multipara* (DSM 3073), *Veillonella parvula* (DSM 2008) and *Rothia mucilaginosa* (DSM 20746), as template DNA in this analysis. This DNA was purchased from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures.

The primers used for the optimization of the SYBR green-based qPCR assays target various regions of the 16S rRNA gene in the genera *Faecalibacterium* [17], *Veillonella* [1, 16], *Lachnospira* and *Rothia* [18] (Table 3.1). For conventional PCR assays, we used primers targeting the 16S rRNA gene of *Faecalibacterium* [19, 20] and *Rothia*, and the *rpoB* gene of *Veillonella* [21, 22] and *Lachnospira* (Table 3.2). We designed conventional PCR primers for *Lachnospira* using the National Centre for Biotechnology Information (NCBI) primer-Basic Local Alignment Search Tool (primer-BLAST) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). This was performed by inserting the *rpoB* sequence of *L. multipara* (accession no. NZ_AUJG01000003.1) into primer BLAST option and allowing the programme (default) to design primers for this sequence. The resulting primers were assessed *in-silico* (as described in section 2. 2 below) to

determine the suitable primer pair. We also designed conventional PCR primers for the detection of *Rothia* in this study (Table 3.2).

Table 3.1 List of primers targeting the 16S ribosomal RNA gene for the detection of FLVR bacteria using a SYBR green based quantitative real-time PCR method

Bacterial genus	Primer name	Primer sequences (5' → 3')	Amplicon size	Reference
Faecalibacterium	Faec-qFor	CTCAAAGAGGGGGACAACAGTT	101	(19)
	Faec-qRev	GCCATCTCAAAGCGGATTGCTC		(19) *
Lachnospira	Lach_qFor	GTAAAGGGAGTGTAGGTGGCA	78	(18)
	Lach-qRev	AACAGTTTCAATAGCAGTTCCGA		(18) *
Veillonella	Veil-qFor	AAGCTATCACTGAAGGAGGG	238	(1)
	Veil-qRev	CGTCCCGATTAACAGAGCTT		(16)
Rothia	Roth-qFor	AGGCTTGACATATACTGGACCG	67	(18)
	Roth-qRev	GCACCACCTGTATACCAGCC		(18) *

*- Modified (underlined nucleotides); qFor- qPCR forward; qRev- qPCR reverse

Table 3.2 List of primers for the detection of FLVR bacteria using conventional PCR method

Bacterial genera	Primer name	Primer sequences (5' → 3')	Target gene	Product size (bp)	Reference
Faecalibacterium	Fprau136F	CTCAAAGAGGGGGACAACAGTT	16S rRNA	601	(19)
	Fprau645R	AATTCCGCCTACCTCTGCACT			(20)
Lachnospira	Lach-cFor	GGTCCTGGTGGTCTTTC	*rpoB*	446	This study
	Lach-cRev	GTACAGTCATCATTCTGAAG			This study
Veillonella	VeillrpoBF	GTAACAAAGGTGTCGTTTCTCGC	*rpoB*	896	(21)
	VrpoBR	GTGTAACAAGGGAGTACGGACC			(22)
Rothia	Roth-cFor	GGTGTTGGTGGAAAGCGTTATG	16S rRNA	797	This study
	Roth-cRev	CGGTCCAGTATATGTCAAGCCT			This study

cFor- conventional PCR forward; cRev- conventional PCR reverse

2.2 *In-silico analysis of the primers*

Computational analysis was performed to assess each primer specificity, melting temperature, secondary structure formation, primer dimer formation and GC content, using DNAMAN v4.13 (Lynnon Biosoft, Lynnon Corporation, Quebec, Canada) and NCBI BLAST [23]. Species-specific 16S rRNA sequences of FLVR bacteria and closely related bacteria were retrieved from NCBI GenBank (http://www.ncbi.nlm.nih.gov/) and aligned using ClustalW2 (EMBL-EBI) [24]. Each 16S rRNA primer was then mapped onto the specific sequences, aligned for each FLVR bacterium and its close relatives. *Lachnospira* and *Veillonella rpoB* primers were mapped onto their respective species-specific *rpoB* sequences aligned with closely related bacterial sequences. We used the reference 16S rRNA sequences of *F. prausnitzii* (accession no. LN907769.1), *L. multipara* (accession no. NR_104758.1), *V. parvula* (accession no. NR_074980.1) and *R. mucilaginosa* (accession no. NR_044873.1). The *rpoB* sequences of *L.*

multipara (accession no. AUJG01000003.1) and *V. parvula* (accession no. EF185160.1) were included as reference sequences. We considered primer pairs which anneal to each other by 5 continuous bases or less [25]. We also made sure that all our primers have a GC content (50-55%) that fell within the recommended range of 50-60% [26]. To enhance the specificity of our primers, slight modifications were made in some primer sequences (Table 3.1). The selected conventional PCR primers were further mapped onto the multiple sequence alignments of various species for the respective genus. This was done to ensure that the region to be amplified will have enough variations to distinguish between the different species of that specific genus.

The *in-silico* specificity of each primer was assessed based on the ability of the primer to align 100% to their target sequences but not to the close relatives. This specificity was further validated *in-vitro* by qPCR and conventional PCR, using pure genomic DNA of FLVR bacteria. The *in-vitro* specificity of the primers was determined using a melt curve analysis for qPCR primers, and gel electrophoresis for conventional PCR primers as described below.

2.3 Annealing temperature optimization and PCR conditions

We determined the optimal annealing temperatures for both the qPCR and conventional PCR primers by performing gradient PCRs using the C1000^{TM} Thermal cycler CFX96^{TM} Real-Time system (Bio-Rad Laboratories). The lowest temperature without non-specific amplifications in both the no template DNA controls (NTC) and the positive controls was considered as the optimal annealing temperature for each primer set.

2.3.1 Real-time PCR (qPCR)

We assessed and optimized four singleplex qPCR protocols. A thermal gradient PCR spanning a temperature range of 54ºC - 68ºC was performed for each primer set. This gradient range was decided upon, based on the *in-silico* melting temperatures of the various primers. The amplification reactions contained 1x IQ SYBR Green supermix (Bio-Rad Laboratories), 0.5µM of each forward and reverse primer, 2µl of template DNA, and nuclease free water to a total volume of 20µl. In each qPCR run we included a No Template Control (NTC). The cycling conditions included an initial denaturation at 95ºC for 5 minutes, 40 cycles of denaturation at 95ºC for 15 seconds, annealing at 54ºC - 68ºC for 30 seconds, and extension at 72ºC for 30 seconds. A melt curve analysis was included at the end of each run with 0.5 ºC temperature increments, every 5 seconds, from 60ºC to 95ºC.

2.3.2 Conventional PCR

We tested various temperature ranges to determine the optimal annealing temperature for each primer set. As for qPCR, this temperature ranges were based on the *in-silico* predicted melting temperatures for the various primers. Each reaction contained 1x GoTaq buffer

(Promega), 1.5 mM $MgCl_2$ (Promega), 0.2 mM of each dNTP (Thermoscientific), 1-unit GoTaq DNA polymerase (Promega), 0.5 µM of each forward and reverse primer, 2 µl of template DNA and sterile water to a total volume of 20 µl. The PCR cycling conditions included an initial DNA denaturation step at 95 ºC for 5 minutes, followed by 30 cycles of denaturation at 95 ºC for 30 seconds, annealing at specific temperatures for 30 seconds, extension at 72 ºC for 30 seconds, and a final extension step at 72 ºC for 5 minutes. For each bacterium in question, we included the other three bacteria as negative controls as well as a NTC to correct for contaminations. A PCR positive control (pure DNA template of the respective bacterial target) was included in each run.

Ten (10) µl of the resulting amplicons were run on a 1.5% agarose gel, at 90 volts for 2 hours. Molecular weight marker (HyperLadder IV, Bioline) was included on all agarose gels. *Veillonella* and *Lachnospira* amplicons were electrophoresed in agarose gels containing ethidium bromide (10 mg/ml in water), in tris-acetate-ethylenediaminetetraacetic acid buffer (1x TAE). *Faecalibacterium* and *Rothia* amplicons were electrophoresed in agarose gels containing SYBR safe (10 000x in DMSO), in Tris-Borate-EDTA buffer (1x TBE). To ensure that a single fragment of correct size was amplified, the resulting bands were compared to the molecular weight marker as well as the PCR positive control. Visualisation was performed under short wavelength Ultraviolet (UV) light using GeneSnap software, version 7.12.06.

2.4 Sensitivity assessment of the PCR assays

To determine the sensitivity of the PCR assays, we assessed five-fold serial dilutions of each of the pure genomic DNA samples in the range $5x10^0$ ng/µl - $2.56x10^{-6}$ ng/µl using the optimized annealing temperature.

For qPCR assays, a standard curve was generated for each bacterium. The standard curve slope, coefficient of determination (R^2) and efficiency were calculated automatically by the qPCR instrument (Thermal cycler CFX96™). The limit of detection (LOD) at 95% confidence interval (CI) is the minimum concentration of the target that is uniformly detectable in at least 95% of the tested samples [(27)]. This was estimated as previously described [(27)]. First, we calculated the percentage number of times that an amplification occurred (n) for each standard dilution over 10 repeats (%n = n/10 x 100). We then plotted %n (Y-axis) against the standard dilutions (X-axis) and determined the dilution at which the curve cuts Y = 95% (Figure 3.2).

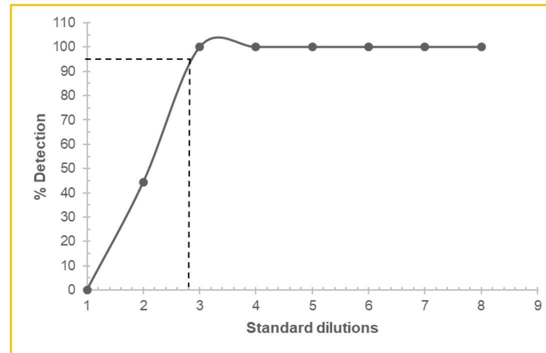


Figure 3.2 An example of how the limit of detection was determined at 95% confidence interval. The dotted line indicates 95%CI (Y-axis) and the corresponding dilution (X-axis). The dilutions are represented as units (1 – 9, represents 2.56×10^{-6} ng/µl – 2×10^{0} ng/µl), the 10th dilution would be 5×10^{0}. ng/µl.

For conventional PCR assays, the amplification was performed on the 2720 Thermal Cycler version 2.08 (Applied Biosystems). The amplicons resulting from the different dilutions were run on agarose gel using the same conditions described in section 2.3.2 above. The lowest dilution showing the right sized band was considered the LOD for the specific primer set.

2.5 Repeatability and reproducibility of the qPCR assays

To measure the repeatability (intra-assay variation), we performed a 10 replicates qPCR assay of four pure DNA dilutions for each bacterium. To assess the reproducibility (inter-assay variation) of the qPCR assays, we tested four different dilutions of pure genomic DNA for each bacterium repeatedly in 10 assays. The dilutions used for the intra and inter assay variations were selected to represent high (1 ng/µl), middle (3.2×10^{-4} ng/µl, 6.4×10^{-5} ng/µl) and low (2.56×10^{-6} ng/µl) DNA concentrations of the assay.

2.6 Statistical analysis of the data

The average quantification cycle number (Cq), standard deviation (SD) and the percentage coefficient of variation (%CV) for the inter- and intra-assay variation assays were calculated using Microsoft excel 2010. The DNA copy number was calculated using the formula [28]:

$$Copy\ number = \frac{(amplicon\ amount\ (ng)\ *\ 6.0221 \times 10^{23}\ molecules/mole)}{(dsDNA\ amplicon\ length\ *\ 660g/mole)\ *\ (1 \times 10^{9}\ ng/g)}$$

Abbreviations: ng- nanograms; dsDNA- double stranded deoxyribonucleic acid; g- grams

3. Results

3.1 qPCR assays

3.1.1 In-silico analysis of qPCR primers

All primers aligned 100% to their targeted bacterial genera. The qPCR primers' melting temperature was in the range of 59 ºC – 69 ºC. None of the primers showed secondary structure formation. All the primers showed a GC content of 50-55%.

3.1.2 Annealing temperature and optimal PCR conditions

The optimal annealing temperature of all qPCR primer sets was 65ºC. All the other PCR conditions tested, including the cycling conditions and the specified reagents concentrations showed optimal performance.

3.1.3 In-vitro specificity of qPCR primers

All the qPCR primers could specifically amplify their target from the pure genomic DNA as shown in Figure 3.3. The melt peaks were produced at 80 ºC, 77.50 ºC, 84.9 ºC and 77.50 ºC for *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* respectively.

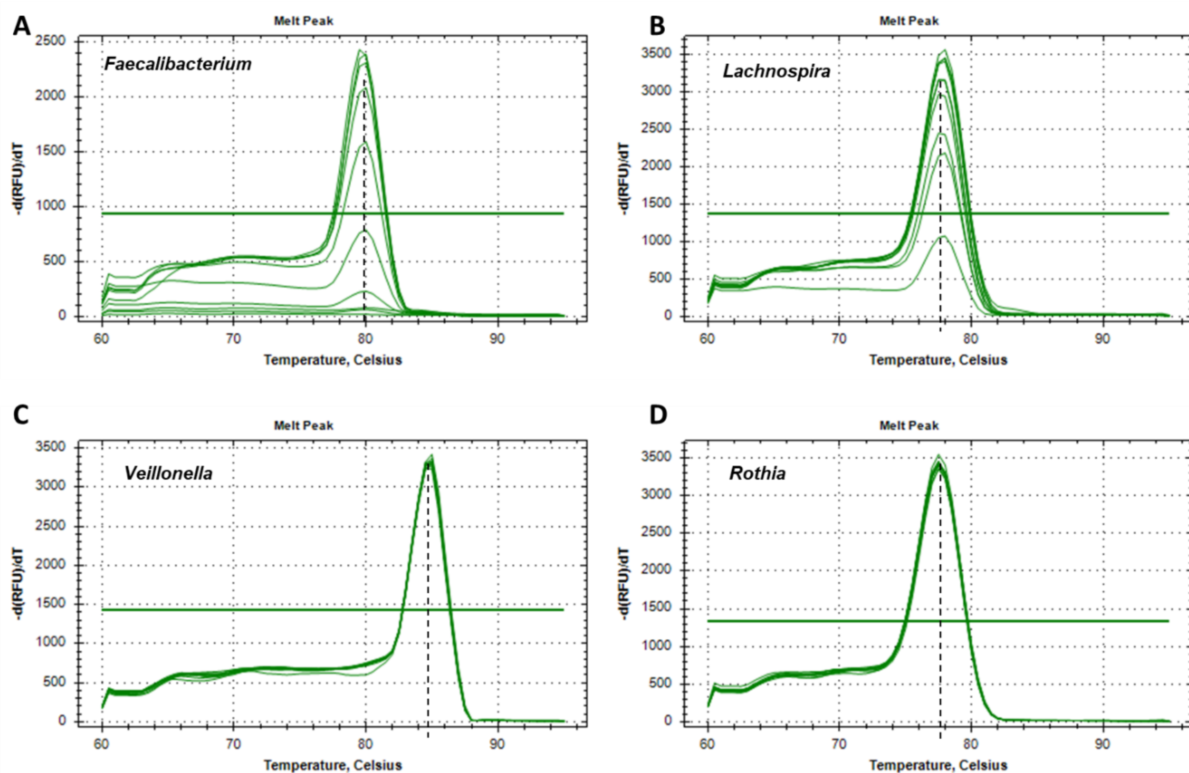


Figure 3.3 Melt curves of FLVR bacterial pure genomic DNA (positive controls). **A.** depicts melt curves for *Faecalibacterium*, **B.** depicts melt curves of *Lachnospira*, **C.** depicts melt curves of *Veillonella*, and **D.** depicts melt curves of *Rothia*. The dotted lines show the melting temperatures of the amplicons.

3.1.4 Sensitivity of the qPCR assays

The reaction sensitivity results for qPCR are shown in Table 3.3. All the assays demonstrated a high efficiency. *Veillonella* primers showed the lowest LOD followed by *Lachnospira* (Table 3.3). *Faecalibacterium* and *Rothia* showed the same highest LOD at 95% CI (Table 3.3).

Table 3.3. Sensitivity of the SYBR green qPCR assays for the detection of FLVR bacteria

	Faecalibacterium	***Lachnospira***	***Veillonella***	***Rothia***
Efficiency	91.2%	88.9%	89.2%	82.7%
R^2	0.879	0.998	0.999	0.959
Slope	-3.554	-3.620	-3.611	-3.820
Estimated LOD 95% CI (ng/µl)	3.2×10^{-4}	6.4×10^{-5}	1.28×10^{-5}	3.2×10^{-4}
LOD genomic DNA copies/ µl	2.891×10^{6}	7.487×10^{5}	4.907×10^{4}	4.358×10^{6}

LOD- limit of detection, CI- confidence interval, R^2- coefficient of determination

3.1.5 Repeatability and Reproducibility of the qPCR assays

The difference between the intra (repeatability) and inter (reproducibility) assays were minimal for all bacteria. The intra-assay %CV for *Faecalibacterium, Lachnospira, Veillonella* and *Rothia* were in the range 0 - 1.63%, 0.24 - 1.25%, 0.34 - 2.67% and 0.21 - 6.78% respectively (Table 3.4). While the inter-assay %CV ranged between 0 - 3.73%, 1.21 - 5.54%, 0.73 - 3.1% and 0 - 7.06% for *Faecalibacterium, Lachnospira, Veillonella* and *Rothia* respectively (Table 3.5).

Table 3.4 Intra-assay variation (repeatability) of FLVR bacteria SYBR green qPCR optimization based on 10 replicates of each dilution

DNA dilution (ng/rxn)	*Faecalibacterium*				*Lachnospira*				*Veillonella*				*Rothia*			
	Copies/rxn	**Cq mean**	**SD**	**CV (%)**	**Copies/rxn**	**Cq mean**	**SD**	**CV (%)**	**Copies/rxn**	**Cq mean**	**SD**	**CV (%)**	**Copies/rxn**	**Cq mean**	**SD**	**CV (%)**
2×10^{0}	1.807×10^{10}	21.50	0.35	1.63	2.34×10^{10}	17.41	0.12	0.67	7.667×10^{9}	15.99	0.05	0.34	2.724×10^{10}	20.25	0.04	0.21
3.2×10^{-3}	2.891×10^{7}	35.75	0.21	0.58	3.743×10^{7}	27.23	0.07	0.24	1.227×10^{7}	26.01	0.13	0.49	4.358×10^{7}	32.60	2.21	6.78
6.4×10^{-4}	5.782×10^{6}	39.32	0.28	0.72	7.487×10^{6}	29.71	0.17	0.58	2.454×10^{6}	28.65	0.17	0.61	8.716×10^{6}	32.60	0.31	0.95
2.56×10^{-5}	2.313×10^{5}	0.00	0	0	2.995×10^{5}	34.84	0.44	1.25	9.814×10^{4}	34.38	0.92	2.67	3.486×10^{5}	36.82	1.02	2.78

Table 3.5 Inter-assay variation (reproducibility) of FLVR bacteria SYBR green qPCR optimization based on 10 separate experiments

DNA dilution (ng/rxn)	*Faecalibacterium*				*Lachnospira*				*Veillonella*				*Rothia*			
	Copies/rxn	**Cq mean**	**SD**	**CV (%)**	**Copies/rxn**	**Cq mean**	**SD**	**CV (%)**	**Copies/rxn**	**Cq mean**	**SD**	**CV (%)**	**Copies/rxn**	**Cq mean**	**SD**	**CV (%)**
2×10^{0}	1.807×10^{10}	20.19	0.59	2.95	2.34×10^{10}	16.61	0.45	2.73	7.667×10^{9}	15.68	0.35	2.25	2.724×10^{10}	20.81	1.47	7.06
3.2×10^{-3}	2.891×10^{7}	34.14	0.42	1.23	3.743×10^{7}	28.28	1.57	5.54	1.227×10^{7}	26.72	0.30	1.14	4.358×10^{7}	32.62	1.06	3.25
6.4×10^{-4}	5.782×10^{6}	38.34	1.43	3.73	7.487×10^{6}	30.05	0.36	1.21	2.454×10^{6}	29.36	0.21	0.73	8.716×10^{6}	36.15	0.681	1.88
2.56×10^{-5}	2.313×10^{5}	0	0	0	2.995×10^{5}	35.62	0.67	1.87	9.814×10^{4}	34.80	1.19	3.1	3.486×10^{5}	0	0	0

Abbreviations: ng/rxn- DNA amount in nanograms per reaction, Copies/rxn- Genomic DNA copies per reaction, Cq- quantification cycle, SD- standard deviation, CV- coefficient of variation

3.2 Conventional PCR assays

3.2.1 In-silico analysis of the primers

As with qPCR, all conventional PCR primers aligned 100% to their specific bacterial genera. No primer showed formation of secondary structures and all primers had a GC content of 50-55%. The *in-silico* melting temperatures for conventional PCR primers are shown in Table 3.6.

3.2.2 Annealing temperature and optimal PCR conditions

Table 3.6 includes optimal annealing temperatures of conventional PCR primers for the detection of the bacteria under investigation. The tested cycling conditions and reagents concentrations showed good performance.

3.2.3 In-vitro specificity of the primers

All the primers demonstrated ability to amplify their targeted amplicon sizes (specified in Table 3.2 and shown by the molecular weight marker in Figure 3.4).

3.2.4 Sensitivity of conventional PCR assays

The conventional PCR reaction sensitivity results are shown in Table 3.6 and Figure 3.4. *Veillonella* primers demonstrated the highest sensitivity (lowest LOD) and *Lachnospira* primers were the least sensitive (highest LOD) (Table 3.6).

Table 3.6 Optimal annealing temperatures and sensitivity of FLVR conventional PCR primers

Bacteria	Target gene	*In-silico* T_m	Gradient PCR range	Annealing temperature	LOD	Genomic DNA copies/ µl
Faecalibacterium	16S rRNA	59 ºC	57ºC - 67ºC	61ºC	2×10^{-1} ng/µl	3.036×10^{8}
Lachnospira	*rpoB*	54 ºC	52ºC - 65ºC	45ºC*	1×10^{0} ng/µl	2.046×10^{9}
Veillonella	*rpoB*	61 ºC	59ºC - 67ºC	65ºC	8×10^{-3} ng/µl	8.147×10^{6}
Rothia	16S rRNA	59 ºC	57ºC - 67ºC	63ºC	8×10^{-3} ng/µl	9.159×10^{6}

T_m- melting temperature; PCR- Polymerase Chain Reaction; *- outside the gradient PCR range; LOD- limit of detection; DNA- Deoxyribonucleic acid.

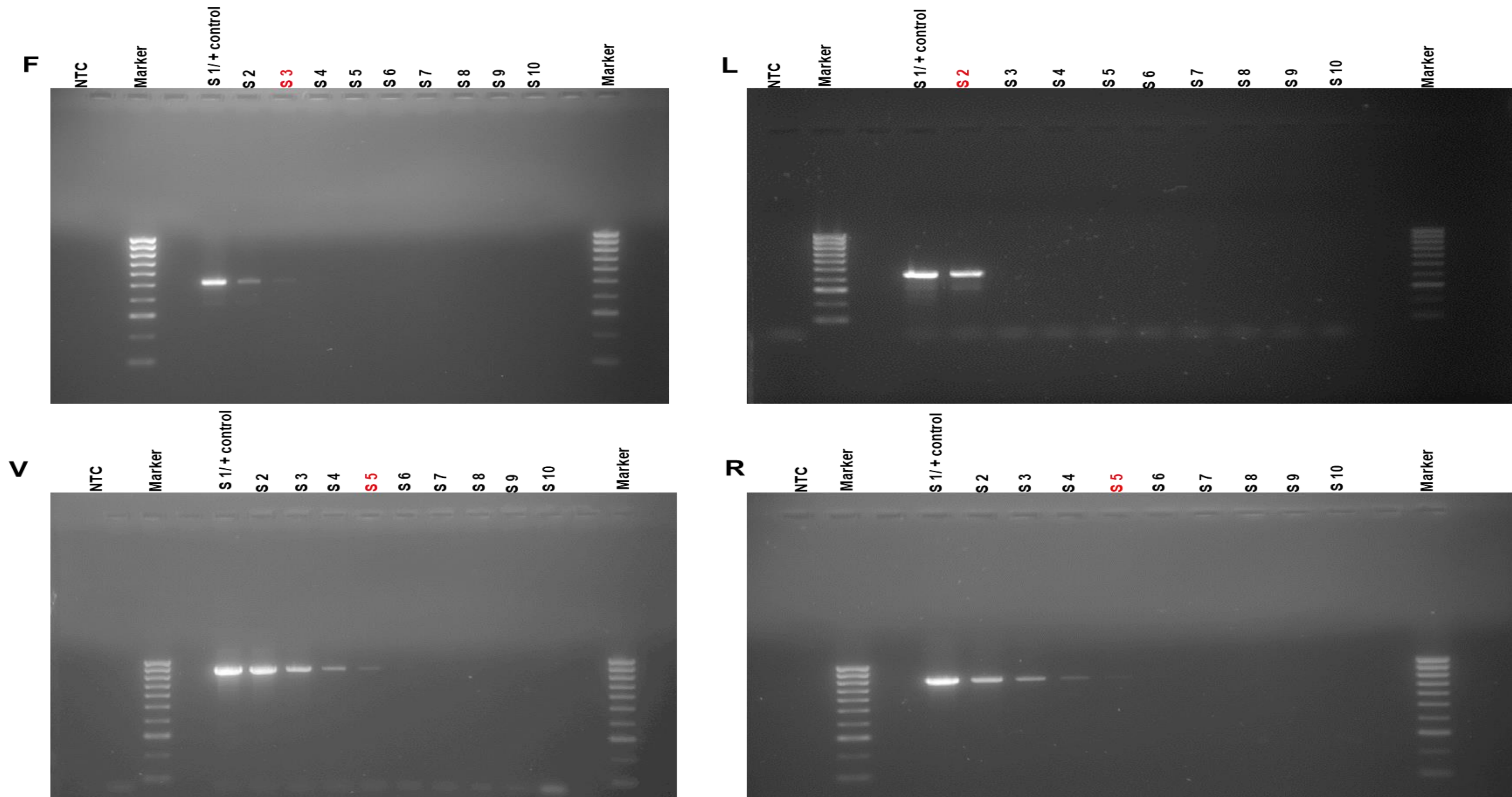


Figure 3.4 Limit of detection (in red) for each of the conventional polymerase chain reaction (PCR) primers determined at the specific annealing temperatures and run on a 1.5% agarose gel. F- *Faecalibacterium* (DSM 17677); L- *Lachnospira* (DSM 3073); V- *Veillonella* (DSM 2008); R- *Rothia* (DSM 20746); NTC- No template control; Marker- 100 lanes; S1 - S10 depicts five-fold serial dilutions in the range $5x10^{0}$ ng/µl - $2.56x10^{-6}$ ng/µl. S1=$5x10^{0}$ ng/µl; S2=$1x10^{0}$ ng/µl; S3=$2x10^{-1}$ ng/µl; S4=$4x10^{-2}$ ng/µl; S5=$8x10^{-3}$ ng/µl; S6=$1.6x10^{-3}$ ng/µl; S7=$3.2x10^{-4}$ ng/µl; S8=$6.4x10^{-5}$ ng/µl; S9=$1.28x10^{-5}$ ng/µl; S10=$2.56x10^{-6}$ ng/µl. S1 was used as the positive control (+ control). *Faecalibacterium* and *Rothia* primers targeted 16S rRNA, while *Lachnospira* and *Veillonella* primers targeted the *rpoB* gene.

Discussion

In this study, we demonstrated the importance of PCR reactions' optimisation, for applicability in future experiments. We showed that at 65 ^{0}C annealing temperature, our qPCR primers could amplify from their respective templates, resulting in the generation of a single melt peak. This melt peak indicates the production of multiple copies of the same amplicon for each of these bacteria, meaning that the primers did not amplify any other gene target from the pure genomic DNA that was used as template. The ability to amplify only the gene of interest is also the reason we obtained a single gel band for each of our bacteria, in the conventional PCR experiments. Moreover, the variations within and between our qPCR assays were very small. This suggests that future application of these optimised assays in the detection of FLVR, should be able to generate reproducible results. Studies have demonstrated the importance of reproducible PCR experiments, as these usually permit generation of reliable results [(29)].

Similar to previous studies [(30)], we observed a high sensitivity of real-time qPCR over conventional PCR. The high sensitivity (lowest LOD) of *Veillonella* primers compared to other FLVR bacteria was demonstrated in both qPCR and conventional PCR assays. Therefore, studies aiming to detect *Veillonella* DNA copies of at least 4.907×10^4 in qPCR and 8.147×10^6 in conventional PCR, can successfully use these primers. However, for *Faecalibacterium*, *Lachnospira* and *Rothia*, the DNA copy number in the sample of interests may need to be higher than this amount in order to be detected. In an attempt to increase these primers' sensitivity (especially for *Lachnospira*), we lowered the annealing temperature to as low as 45 ^{0}C. This was done to ensure that the observed low sensitivity was not caused by the primers being unable to anneal to the template. The reason behind this is that at low annealing temperature, all the primers should be able to bind to their respective templates properly. We therefore, encourage future studies targeting *Lachnospira* at low concentration, to design very sensitive primers as our primers may cause false negative results. In conclusion, we consider our optimized SYBR Green based qPCR and conventional PCR assays suitable for the detection of FLVR bacteria. Our conventional PCR assays may fail to detect very low DNA concentrations thus giving false negative results. As a result, future studies targeting FLVR bacteria should consider more sensitive approaches which can detect very low DNA concentrations.

References cited

1. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152-307ra152.
2. Cotta M, Forster R. The family *Lachnospiraceae*, including the genera *Butyrivibrio*, *Lachnospira* and *Roseburia*. The Prokaryotes: Springer; 2006. p. 1002-21.
3. Duncan SH, Hold GL, Harmsen HJ, Stewart CS, Flint HJ. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. International journal of systematic and evolutionary microbiology. 2002;52(6):2141-6.
4. MacFarlane T. A semiselective medium for the isolation of *Veillonella* species from the mouth. Journal of clinical pathology. 1977;30(2):191.
5. Paściak M, Mordarska H, Szponar B, Gamian A. Chemotaxonomical methods in the diagnostics of clinical strains causing actinobacterial infections. Postepy higieny i medycyny doswiadczalnej (Online). 2007;61:403-12.
6. Dekio I, Hayashi H, Sakamoto M, Kitahara M, Nishikawa T, Suematsu M, et al. Detection of potentially novel bacterial components of the human skin microbiota using culture-independent molecular profiling. Journal of medical microbiology. 2005;54(12):1231-8.
7. Schochetman G, Ou C-Y, Jones WK. Polymerase chain reaction. The Journal of infectious diseases. 1988;158(6):1154-7.
8. Garibyan L, Avashia N. Research techniques made simple: polymerase chain reaction (PCR). The Journal of investigative dermatology. 2013;133(3):e6.
9. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, et al. The real-time polymerase chain reaction. Molecular aspects of medicine. 2006;27(2-3):95-125.
10. Antiabong JF, Ngoepe MG, Abechi AS. Semi-quantitative digital analysis of polymerase chain reaction-electrophoresis gel: Potential applications in low-income veterinary laboratories. Veterinary world. 2016;9(9):935.
11. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Genome research. 1996;6(10):986-94.
12. McBride L, Livak K, Lucero M, Goodsaid F, Carlson D, Stevens J, et al. Quantitative PCR technology. Gene Quantification: Springer; 1998. p. 97-110.
13. Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP. Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques. 2013;54(6):314-20.
14. Higuchi R, Dollinger G, Walsh PS, Griffith R. Simultaneous amplification and detection of specific DNA sequences. Nature Biotechnology. 1992;10(4):413-7.
15. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Nature Biotechnology. 1993;11(9):1026-30.
16. Rinttilä T, Kassinen A, Malinen E, Krogius L, Palva A. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. Journal of applied microbiology. 2004;97(6):1166-77.
17. Lopez-Siles M, Martinez-Medina M, Surís-Valls R, Aldeguer X, Sabat-Mir M, Duncan SH, et al. Changes in the abundance of *Faecalibacterium prausnitzii* phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and patients with colorectal cancer. Inflammatory bowel diseases. 2016;22(1):28-41.

18. Stiemsma L, Arrieta M-C, Dimitriu P, Cheng J, Thorson L, Lefebvre D, et al. Shifts in *Lachnospira* and *Clostridium sp.* in the 3-month stool microbiome are associated with preschool-age asthma. Clinical Science. 2016:CS20160349.
19. Lopez-Siles M, Martinez-Medina M, Surís-Valls R, Aldeguer X, Sabat-Mir M, Duncan SH, et al. Changes in the abundance of *Faecalibacterium prausnitzii* phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and patients with colorectal cancer. Inflammatory bowel diseases. 2015;22(1):28-41.
20. Ramirez-Farias C, Slezak K, Fuller Z, Duncan A, Holtrop G, Louis P. Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. British Journal of Nutrition. 2008;101(4):541-50.
21. Arif N, Do T, Byun R, Sheehy E, Clark D, Gilbert SC, et al. *Veillonella rogosae sp. nov.*, an anaerobic, Gram-negative coccus isolated from dental plaque. International journal of systematic and evolutionary microbiology. 2008;58(3):581-4.
22. Igarashi E, Kamaguchi A, Fujita M, Miyakawa H, Nakazawa F. Identification of oral species of the genus *Veillonella* by polymerase chain reaction. Molecular Oral Microbiology. 2009;24(4):310-3.
23. Stover NA, Cavalcanti AR. Using NCBI BLAST. Current Protocols Essential Laboratory Techniques. 2009:11.1. 1-.1. 36.
24. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic acids research. 2013;41(W1):W597-W600.
25. Thornton B, Basu C. Real-time PCR (qPCR) primer design using free online software. Biochemistry and Molecular Biology Education. 2011;39(2):145-54.
26. Saiki RK. The design and optimization of the PCR. PCR technology: Springer; 1989. p. 7-16.
27. Jesse A. Canchola RMD, Pleasanton, California, USA, Pari Hemyari RMD, Pleasanton, California, USA. Limit of Detection (LoD) Estimation Using Parametric Curve Fitting to (Hit) Rate Data: The LoD_Est SAS® Macro. 2016.
28. Prediger E. Calculations: Converting from nanograms to copy number 2013 [updated 16 March 2017. Available from: https://www.idtdna.com/pages/education/decoded/article/calculations-converting-from-nanograms-to-copy-number.
29. Bustin SA. The reproducibility of biomedical research: sleepers awake! Biomolecular Detection and Quantification. 2014;2:35-42.
30. Schmittgen TD, Zakrajsek BA, Mills AG, Gorn V, Singer MJ, Reed MW. Quantitative reverse transcription–polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Analytical biochemistry. 2000;285(2):194-204.

Chapter 4:

Association of *Faecalibacterium, Lachnospira*, *Veillonella*, and *Rothia* with childhood wheezing in the Drakenstein Child Health Study, South Africa

Summary

Background and rationale: Wheezing is a prevalent respiratory tract symptom among children in different populations globally. There is increasing evidence on the contribution of faecal microbiome in the occurrence of allergic diseases. A single study has observed reduced abundance of bacterial genera *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* (FLVR) in stools of three-month old Canadian infants with atopy-wheeze symptoms. In this chapter, we aimed: (1) to evaluate the applicability of the optimized PCR assays in FLVR detection from human stool samples; (2) to identify FLVR bacterial species present in stool samples of wheezing and non-wheezing infants within the first three months of life; (3) to identify factors associated with gut colonization by FLVR bacteria in the first year of life;(4) to determine the association of FLVR bacteria with wheezing and recurrent-wheezing outcomes; and (5) to determine the sensitivity of FLVR bacterial load within the first four days of life (birth time-point) in the prediction of wheezing outcome.

Method: Using the optimized conditions, we screened 533 human stool samples for the presence of each bacterium by qPCR. The samples were collected at four time-points (birth, six weeks, six months and 12 months) from wheezing and non-wheezing infants. All samples positive by qPCR were subjected to conventional PCR. The birth and six weeks samples positive by conventional PCR were sequenced using Sanger's method and identified using the NCBI BLAST tool. Mann-Whitney test was used to determine the difference in FLVR bacterial load between wheezing and non-wheezing infants. We determined the association of FLVR bacteria with childhood wheeze and recurrent wheeze at two levels: first using bacterial detection and second using bacterial load. Conditional and generalised logistic regressions were used to determine the association of FLVR bacteria with wheezing and recurrent-wheezing respectively. Receiver operating characteristic (ROC) curves were used to determine the sensitivity of each FLVR bacteria as well as their combinations in predicting the wheezing outcome. We considered three wheezing phenotypes in the ROC analysis: 'any wheeze' (wheeze with or without lower respiratory tract infections (LRTIs)), 'no-LRTIs wheeze' (wheeze without LRTIs), and 'LRTIs wheeze' (wheeze with LRTIs). Data were statistically analysed using RStudio version 3.4.4 and, we considered a P-value less than 0.05 to be significant. The bacterial load was converted to log base 10 in all the analysis.

Results: Using qPCR, we detected *Rothia*, *Veillonella, Faecalibacterium* and *Lachnospira* in 90% (479/533), 73% (388/533), 51% (274/533) and 14% (77/533) of the samples, respectively. Conventional PCR allowed detection of *Rothia*, *Veillonella, Faecalibacterium* and *Lachnospira* in 55% (263/479), 74% (289/388), 53% (145/274) and 0% (0/77) of the qPCR positive samples, respectively. *Faecalibacterium prausnitzii* was identified (at ≥97% sequence similarity), in 100% (13/13) of the samples sequenced for the genus *Faecalibacterium*. Similarly, we identified *Rothia mucilaginosa* (at ≥97% sequence similarity) in 100% (80/80) of the samples sequenced for the genus *Rothia*. *Veillonella* sequences (65%; 60/93) identified at ≥97% sequence similarity, included *Veillonella rogosae* (1.7%; 1/60), *Veillonella parvula* (13.3%; 8/60), *Veillonella dispar* (23.3%; 14/60) and *Veillonella atypica* (28.3%; 17/60). The remaining 33.3% (20/60) of the *Veillonella* sequences were 92-94% similar to *V. atypica*. A high proportion of infants became colonized with FLVR bacteria as they age from birth towards 12 months. Using logistic regression, we observed that irrespective of the time points, male gender (adjusted OR = 0.65, 95%CI: 0.42 - 0.98) and TC-Newman residence (adjusted OR = 0.52, 95%CI: 0.29 - 0.91) were associated with less colonisation by *Faecalibacterium*.

Breastfeeding was associated with less colonisation by *Lachnospira*, (adjusted OR = 0.17, 95%CI: 0.05 - 0.49) and *Veilonella* (adjusted OR = 0.32, 95%CI: 0.10 - 0.91). Colonisation by *Rothia* was significantly associated with mother's tertiary education (adjusted OR = 11.73, 95%CI: 1.36 - 2.58). This study observed a significantly reduced *Rothia* load within the first four days of life in infants at risk of wheezing ($P<0.001$). A conditional logistic regression also verified that infants at low risk of wheezing (adjusted odds ratio (aOR)=0.54, 95%CI: 0.28-0.93) and recurrent-wheezing (aOR=0.29, 95%CI: 0.05-0.88) had high *Rothia* load very early in life. *Lachnospira* load within the first four days of life was significantly high in wheezing infants ($P=0.00562$). However, the regression analysis showed no significant association of *Lachnospira* with wheezing. The ROC curve analysis showed that quantification of *Lachnospira* (AUC=0.833, 95%CI: 0.64-1.00) and *Rothia* (AUC=0.707, 95%CI: 0.62-0.79) in the first four days of life performs best in wheeze prediction. *Lachnospira* alone was more sensitive in predicting wheeze, compared to other FLVR bacteria whose performances increased upon interaction with each other. Separating LRTIs from no-LRTIs wheeze increased the performance of FLVR bacteria in predicting wheeze. FLVR bacteria performance was highest in predicting no-LRTIs wheeze.

Conclusion: Colonisation with *Rothia* very early in life may potentially be protective against infant wheezing. *Lachnospira* and *Rothia* could serve as single biomarkers for early diagnosis of childhood wheezing. Interaction among FLVR bacteria is important in wheezing prediction.

Keywords: Infants, stool, FLVR, wheeze, recurrent-wheeze, regression analysis, ROC curve

1. Introduction

1.1 Background and rationale

Majority of children experience a wheezing symptom globally [1, 2]. This symptom is more common in young children (due to their smaller airways) and is characterized by a continuous whistling sound resulting from an obstruction in the lower respiratory tract [3, 4]. Wheezing prevalence rates of 31%, 41% and 16% have been reported in European countries, Latin America, and African countries, respectively [5]. Studies conducted in developed countries have shown prevalence rates of wheezing in infants, ranging between 20% and 30% [1]. In addition, it has been reported that in Latin America, a significant proportion of children who die of acute respiratory tract infections in their first year of life have a history of wheezing [1]. Wheezing-related illnesses in pre-school children is estimated to cost the health service about 53 million pounds in the United Kingdom (UK) [6], and billions of dollars annually in the United States of America (USA) [7]. Three or more wheezing episodes occurring in twelve consecutive months is referred to as recurrent wheezing, and is usually accompanied with frequent paediatric medical consultations [1]. Recurrent wheezing is one of the most prevalent chronic respiratory symptoms during childhood, affecting millions of infants and children in the United States of America [7].

Chronic respiratory diseases (such as bronchiolitis and asthma) have been reported as the most frequent cause of wheezing in young children [3]. These diseases usually develop later during childhood (13-14years), but present with wheezing as an early onset symptom [2, 8, 9].

Therefore, wheezing during early-life could mean increased risk of developing chronic respiratory diseases later in life [10, 11]. In adolescence with airway hyperresponsiveness, the odds of onset wheezing and recurrent asthma was reported at 3.91 (95% CI 1.21-12.66) [12]. Because of this connection, an efficient early diagnosis of wheezing followed by effective treatment options is needed to help alleviate these symptoms at an early onset, and therefore reducing the risk of developing chronic respiratory diseases later in life [3, 7]. Given the health burden and high financial costs associated with wheezing-associated illnesses [5-7], studies aiming to explore possible diagnostic and treatment options for these illnesses are therefore required.

Microbiome studies have demonstrated the association of gastrointestinal tract microbiota with the development and progression of respiratory diseases [13]. For example, colonisation with *Clostridium difficile* has been associated with high risk of childhood wheeze, eczema and asthma [14]. An increase in the relative abundance of *Clostridium neonatale* have been reported to increase the risk of developing childhood asthma [15]. An increased abundance of *Bacteroides* have been associated with high risk of bronchiolitis in infants [16]. In contrast, the prevalence of *Bifidobacterium*, *Lactobacilli* [17] or *Lachnospira* [15, 18] has been reported to protect against allergic diseases. Furthermore, the gastrointestinal microbiota has also been reported to protect the host during pneumococcal pneumonia in mice [19]. This evidence suggests that manipulating the gastrointestinal microbiota may provide potential for treating or alleviating symptoms of respiratory diseases.

In Canada, 16S rRNA sequencing followed by target-specific real-time polymerase chain reaction (PCR) demonstrated a significant reduction in bacterial genera, *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* (FLVR) in stools of three-month old infants with atopy-wheeze as compared to controls [18]. This study further showed causal evidence of these bacterial genera in a mice model of airway inflammation with human microbiota. Germ-free mice inoculated with faecal transplant from three-month old infants with atopy-wheeze and supplemented with FLVR bacteria showed reduced airway inflammation compared to the controls (no FLVR bacteria supplement). Since airway inflammation and obstruction cause wheezing [3, 4], this reduction in airway inflammation could also mean reduction in wheezing symptoms as well.

Because only a single study has associated FLVR bacteria with wheezing, further research is needed to strengthen this finding [18]. In addition, it is also not known whether this is true in different human populations worldwide. The overall aim of this study was to determine the contribution of any of the FLVR bacteria or their combination in the development of infant

wheezing using stool samples from the Drakenstein Child Health Study (DCHS), Cape Town, South Africa.

1.2 Hypothesis

The abundance of FLVR bacteria is reduced in wheezing children compared to non-wheezing children in the Drakenstein Child Health Study (DCHS).

1.3 Specific objectives

- To evaluate the applicability of the optimized PCR assays in the longitudinal detection of FLVR bacteria from human stool samples.
- To identify the species of FLVR bacteria present in the first three months stool samples of wheezing and non-wheezing infants.
- To identify factors associated with gastrointestinal colonization by FLVR bacteria during the first year of life.
- To determine the association of FLVR bacteria with wheezing and recurrent wheezing outcomes.
- To determine the sensitivity of FLVR bacterial load early in life in predicting the wheezing outcome.

2. Materials and methods

Figure 4.1 outlines the study workflow.

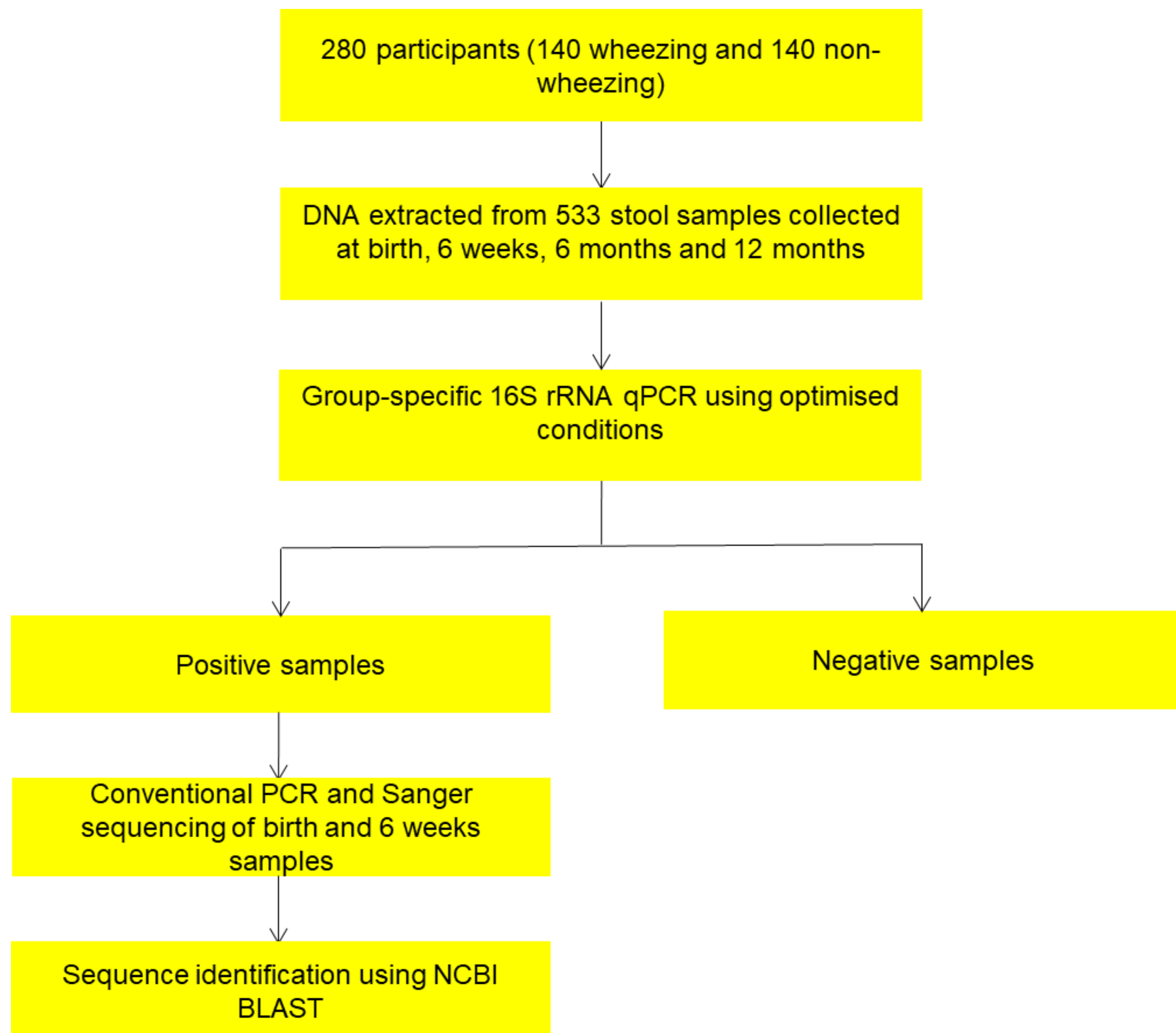


Figure 4.1 Study workflow showing the molecular detection of FLVR bacteria from stools, using optimal PCR conditions. F- *Faecalibacterium*, L- *Lachnospira*, V- *Veillonella,* R- *Rothia;* DNA- Deoxyribonucleic acid; qPCR- Quantitative Polymerase Chain Reaction; rRNA- ribosomal Ribonucleic Acid; PCR- Polymerase Chain Reaction; NCBI BLAST; National Centre for Biotechnology Information Basic Local Alignment Search Tool

2.1 Ethical consideration

Ethical clearance was provided by the Faculty of Health Sciences, Human Research Ethics Committee (HREC) of the University of Cape Town, South Africa (HREC REF 650/2018).

2.2 Study design and population

This study (HREC 650/2018) is a longitudinal case-control study, nested within the DCHS (HREC 401/2009) which is primarily funded by the Bill and Melinda Gates Foundation [(20)]. The DCHS aims to investigate the epidemiology, hazardous factors, causation and extended impact of early lower respiratory tract infections (LRTIs) on child health in a low and middle-income country (LMIC) [(20)]. The DCHS is situated approximately 60 kilometres from Cape Town, in the sub-district called Drakenstein. It has enrolled approximately 1000 mother/infant pairs and is following the children from birth up to 5 years [(20)]. Detailed characteristics of the DCHS have been previously described [(20-22)].

2.3 Participants selection

Our study included 280 participants, of which 140 are wheezers (cases) and 140 are non-wheezers (controls). We define a wheezing case, as a child who has wheezed at least once during his/her first year of life. Non-wheezing controls were age and site matched infants who never wheezed during their first year of life. These infants were selected from the main study database based on whether they have stool samples available, to represent at least one of the time points, birth, six weeks, six months and twelve months (Table 4.1).

2.4 Infant stool samples and metadata variables

In the DCHS, infant stool samples were collected at birth, 1, 2, 6, 9, 12, 18 and 24 months and stored at -80°C until use [(20)]. Comprehensive data were collected, which include biomedical, psychosocial, environmental, socio-demographic, physical health of the parents and infant, and inter current morbidity [(20)].

2.5 Study outcomes

The primary outcome was child wheezing, defined as at least one wheezing episode in the first year of life. The secondary outcome was recurrent wheezing, defined as two or more wheezing episodes occurring in a single year. Wheezing in the DCHS is measured as maternally reported wheeze during clinical visits, as well as episodes identified by nurses during the surveillance for lower respiratory tract infection (LRTI) symptoms [(20, 23)].

2.6 Application of optimized PCR conditions for the detection of FLVR bacteria in human stool samples from our cohort

2.6.1 Stool samples for analysis

Frozen (-80 ºC) deoxyribonucleic acid (DNA) was used for this analysis. This DNA was previously extracted from approximately 50 mg of passed stool or 100 µl of aspirate stool using the ZR Faecal DNA MiniPrep™ (Catalog No. D6010, Zymo Research, CA, USA), with minor changes. A maximum of 600 µl of the supernatant (from centrifuging with ZR BashingBead) was transferred to the Zymo-Spin™ IV Spin Filter, instead of the recommended 400 µl (step 4 of the protocol). In addition, our DNA was eluted in 60 µl of the DNA elution buffer instead of the 100 µl recommended by the manufacturer (step 10 of the protocol). Table 4.1 shows the number of samples screened at the four different time-points included in this study.

Table 4.1 Definitions of the time points and number of samples screened for each time-point

Time point	Sample collection time range	Number of samples
Birth	Within the first four days	170
6 weeks	Between four and eight weeks	118
6 months	Between five and seven months	147
12 months	Between 11 and 13 months	98

2.6.2 Detection of FLVR bacteria from stool samples using qPCR and conventional PCR

To evaluate the diagnostic ability of the optimized PCR conditions, we screened 533 stool samples for the presence of each FLVR bacterium by qPCR. Samples were considered to be positive for the targeted bacteria only if the resulting amplicon had a melt temperature differing from the positive control by less than 2ºC, as previously reported [24]. The positive samples identified by qPCR were subjected to a round of conventional PCR followed by gel electrophoresis. Positive controls of pure genomic DNA as well as negative controls (NTC) were included in all PCR assays. The C1000™ Thermal cycler CFX96™ Real-Time system (Bio-Rad Laboratories) was used for qPCR screening, and conventional PCR was performed using the 2720 Thermal Cycler version 2.08 (Applied Biosystems).

2.6.3 Sequencing and bioinformatics analysis of conventional PCR products

Subsequent to gel electrophoresis, amplicons of birth and six weeks samples displaying the correct sizes were sent for Sanger sequencing to Inqaba Biotechnical industries (Muckleneuk, Pretoria, South Africa). For samples showing non-specific amplifications, bands of correct sizes were excised from the agarose gels with sterile surgical blades and purified using the QIAquick Gel Extraction Kit (Whitehead Scientific (Pty) Ltd) as per the manufacturer's instructions. However, we eluted our DNA in 25µl of the elution buffer instead of the recommended 30µl to increase the concentration (step 9 of the protocol). The purified products were sent to Inqaba Biotec (Muckleneuk, Pretoria, South Africa) for sequencing. Each sample was sequenced in both forward and reverse directions using the ABI V3.1 Big dye kit according to the manufacturer's instructions and an ABI3500XL genetic analyser (Applied Biosystems). The chromatograms were visualized with chromas software version 2.6, and the sequences were aligned using DNAMAN version 4.13. The consensus region between the forward and reverse sequences for each sample was compared against the annotated sequences in NCBI GenBank using the BLAST search tool. This was done to identify the bacterial species to which the sequences belong to.

2.7 Statistical analysis of the data

All tests were performed in RStudio version 3.4.4, and a *P*-value of less than 0.05 was considered significant.

2.7.1 Characteristics of the cohort

Descriptive continuous data are reported as medians, lower (Q1) and upper (Q3) quartiles, while categorical data reported as proportions. Since our wheezing and non-wheezing groups were matched, we used univariate and multivariate conditional logistic regression to test the association of each covariable with the wheezing outcome. We employed conditional logistic regression as it takes into account matching in observational cases-control studies [25]. Since

only a maximum of 14 covariables can be adjusted for in a regression model [26], we excluded residential area and mother's education. The 14 covariates included in the analysis were: baby gender, baby length, birth weight expressed as z-score, breastfeeding, delivery mode, ever had LRTIs, monthly household income, pets' exposure (dog and/or cat), smoking during pregnancy, HIV exposure, full term gestational age, tobacco exposure after birth, supplements use (zinc and/or multivitamins and/or folic acid) and cotinine exposure.

2.7.2 Factors associated with gut colonisation by FLVR bacteria

Univariate and multivariate generalized logistic regression was used to identify factors associated with gut colonisation by FLVR bacteria based on qPCR detection results. The co-variates adjusted for in the multivariate analysis were: wheezing, baby gender, birth weight (z-score), breastfeeding, delivery mode, ever had LRTIs, monthly household income, pets' exposure (dog and/or cat), smoking during pregnancy, HIV exposure, gestational age, tobacco exposure after birth, supplements use (zinc and/or multivitamins and/or folic acid) and cotinine exposure.

2.7.3 Association of FLVR bacteria with wheezing and recurrent-wheezing outcomes

Univariate and multivariate conditional logistic regression analysis was used to test the association of FLVR bacteria with infant wheezing. Since we did not have matched recurrent and non-recurrent wheezing groups, we assessed the association of FLVR bacteria with recurrent wheezing using generalized logistic regression. Mann-Whitney rank sum test was used to compare the bacterial load (DNA copies) of each FLVR bacterium between the wheezing and non-wheezing groups at each time-point. The bacterial load was converted to logarithmic scale (log base 10).

2.7.4 Sensitivity of FLVR bacteria in predicting the wheezing outcome

Receiver Operating Characteristic (ROC) curves were used to test the sensitivity of FLVR bacterial load at the birth time-point (first four days of life) in predicting the wheezing outcome. Because no study has determined the sensitivity of FLVR bacteria in wheezing prediction, we performed the ROC test at four levels. These were: a) Each bacterium alone, b) all the six possible two by two interactions (FL, FV, FR, LV, LR and VR), c) all the four possible three by three interactions (FLV, FLR, FVR and LVR), and d) the four by four (FLVR) interaction. The interaction analysis was performed to see how sensitive FLVR bacteria perform when they are in combinations compared to when they are alone. We first performed ROC analysis of FLVR bacteria in predicting 'any wheeze' (main cohort of 140 wheezing and 140 non-wheezing infants). Secondly, we performed FLVR bacterial ROC analysis to predict 'no-LRTIs wheeze' (sub-cohort of 50 wheezing and 124 non-wheezing infants) which excludes wheezing resulting from LRTIs. Lastly, we tested the performance of FLVR bacteria in predicting 'LRTIs wheeze'

(sub-cohort of 90 wheezing and 17 non-wheezing infants), which includes only wheezing resulting from LRTIs. In addition to the ROC curves at birth, we also determined the performance of FLVR bacteria in predicting ‘any wheeze’ at six weeks, six months and 12 months. The pROC package in RStudio was used to plot ROC curves as well as calculating the area under ROC curve (AUC) and the confident interval (CI) of the AUC. All this analysis was based on the bacterial load transformed into log base 10.

3. Results

3.1 Characteristics of the study participants

Table 4.2 summarized characteristics of the study participants. After adjusting for potential confounders (baby gender, baby length, birth weight expressed as z-score, breastfeeding, delivery mode, ever had LRTIs, monthly household income, pets’ exposure (dog and/or cat), smoking during pregnancy, HIV exposure, full term gestational age, tobacco exposure after birth, supplements use (zinc and/or multivitamins and/or folic acid) and cotinine exposure), most variables were evenly distributed among wheezing and non-wheezing infants. However, non-wheezing infants had a high birthweight compared to wheezing infants (z-score: -0.0809 wheeze vs 0.294 non-wheeze). We observed a high proportion of breastfed non-wheezing infants (90.7% non-wheezing vs 83.7% wheezing) as well as vaginally delivered non-wheezing infants (81.4% non-wheezing vs 80% wheezing). Most wheezing infants had LRTIs (64.3% vs 11.4%) and were exposed to cats (37.1% vs 32.1%) compared to non-wheezing infants.

Table 4.2 Characteristics of the study participants

Covariates	Wheeze (n=140)	Non-wheeze (n=140)	cOR (95% CI)	aOR (95% CI)
Baby gender (male)	77 (55%)	76 (54.2%)	1.05 (0.82-1.33)	0.65 (0.45-0.94) *
Birth length (cm), Median (Q1 - Q3)	50 (48 - 52)	50.5 (49 - 52)	0.98 (0.95-1.02)	1.06 (0.98-1.14)
Birth weight (z-score), Median (Q1 - Q3)	-0.08 (-0.79 – 0.56)	0.29 (-0.44 – 0.84)	0.76 (0.67-0.86) ***	0.59 (0.45-0.79) ***
Breastfeeding (Yes)	117 (83.6%)	127 (90.7%)	0.29 (0.19-0.45) ***	0.25 (0.12-0.53) ***
Delivery mode (vaginal)	112 (80%)	114 (81.4%)	1.08 (0.82-1.41)	0.49 (0.31-0.77) **
Ever had LRTIs (Yes)	90 (64.3%)	16 (11.4%)	13.95 (10.27-19.19) ***	24.13 (16.16-37.46) ***
Monthly household income				
<R1000	50 (35.7%)	48 (34.3%)	1.41 (0.99-1.99)	1.63 (0.96-2.79)
R1000-5000	60 (42.9%)	57 (40.7%)	0.87 (0.73-1.03)	1.15 (0.69-1.91)
Pets exposure (Cat and/or Dog)	67 (47.9%)	61 (43.6%)	1.23 (0.97-1.56)	1.67 (1.13-2.49) **
Cat (Yes)	52 (37.1%)	45 (32.1%)	1.28 (0.99-1.65)	2.44 (1.43-4.23) **
Dog (Yes)	61 (43.6%)	57 (40.7%)	1.16 (0.91-1.47)	0.95 (0.57-1.57)
Residential area (TC-Newman)	82 (58.6%)	84 (60%)	0.98 (0.76-1.25)	1.16 (0.70-1.92)
Smoking during pregnancy (Yes)	47 (33.6)	37 (26.4)	1.33 (1.02-1.73) *	1.29 (0.84-1.99)
HIV exposure	27 (19.3%)	18 (12.9%)	1.62 (1.17-2.25) **	0.59 (0.32-1.09)
Mother's education				
Secondary	122 (87.1%)	121 (86.4%)	0.60 (0.28-1.23)	0.49 (0.16-1.62)
Tertiary	7 (5%)	12 (8.6%)	0.35 (0.15-0.81) *	0.57 (0.16-2.10)
Full term gestation (Yes)	129 (92.1%)	130 (92.9%)	1.19 (0.65-2.20)	2.90 (0.94-9.70)
Tobacco exposure after birth (Yes)	116 (82.9%)	117 (83.6)	0.91 (0.66-1.25)	0.77 (0.44-1.34)
Supplements use (Yes)	132 (94.3%)	134 (95.7%)	0.62 (0.34-1.08)	0.43 (0.16-1.11)
Cotinine exposure (Yes)	110 (78.6%)	107 (76.4%)	1.16 (0.85-1.57)	0.90 (0.55-1.49)
Recurrent wheeze				
Yes	36 (25.7%)	-	-	-
No	98 (70%)	-	-	-
NA	6 (4.3%)	-	-	-

n- number of participants; NA- missing data; Q1- 25th percentile; Q3- 75th percentile; *- significant (**- *P* < 0.01, ***- *P* < 0.001)

3.2 Detection of FLVR bacteria in human stool samples

3.2.1 Performance of qPCR primers in the detection of FLVR from stool samples in our cohort

When used in the analysis of stool samples, the qPCR primers of *Veillonella* (Veil-qFor/ Veil-qRev) and *Rothia* (Roth-qFor/ Roth-qRev) were specific for their targets. A single melt peak was produced for *Veillonella* and *Rothia* at 84 ºC and 77.50 ºC, respectively (Figure 4.2 C and D, respectively). However, *Lachnospira* primers (Lach-qFor/ Lach-qRev) produced multiple peaks in addition to the positive control at 77 ºC (Figure 4.2 B). The primer set for *Faecalibacterium* (Faec-qFor/ Faec-qRev) produced two closer peaks at 81 ºC (samples) and 80 ºC (positive controls) (Figure 4.2 A).

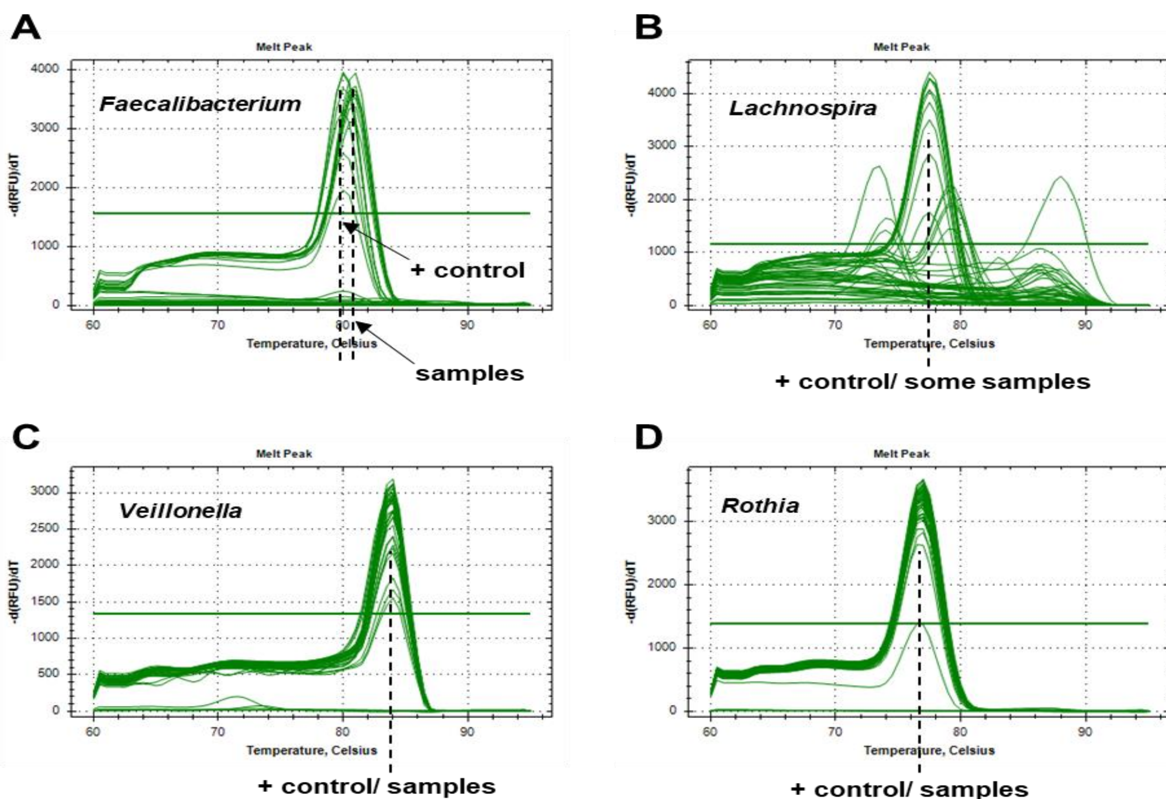


Figure 4.2 Melt curves of FLVR bacteria detection in stool samples including the melt curves of the PCR positive controls (+ control). **A.** depicts melt curves for *Faecalibacterium*, **B.** depicts melt curves of *Lachnospira*, **C.** depicts melt curves of *Veillonella*, and **D.** depicts melt curves of *Rothia*. The dotted lines show the melting temperatures of the amplicons. The extra melt curves for *Lachnospira* (**B**) indicates non-specific amplifications.

3.2.2 Longitudinal detection of FLVR bacteria

Table 4.3 reports the longitudinal detection of FLVR bacteria from the participants in our cohort using qPCR. Out of 280 participants included in this study, only 8.6% (24/280) had samples collected for all the four time-points (Table 4.3). Similar proportions of wheezing and non-wheezing participants had samples collected at the different time-points. Compared to other

time-points, the participants positive for *Faecalibacterium, Veillonella* and *Rothia* were less at birth. We observed that the proportions of wheezing participants positive for *Faecalibacterium,* and *Rothia* increases, with an increase in age. In the non-wheeze group, *Faecalibacterium* and *Veillonella* were detected in most participants as they age. *Lachnospira* was frequently detected at birth in all groups. Its detection dropped at six weeks and started to increase through six months to 12 twelve months. Of the 24 participants from whom samples were collected for all the four time-points (Table 4.3), *Faecalibacterium, Veillonella* and *Rothia* were detected in 8.3% (2/24), 25% (6/24) and 70.8% (17/24) participants, respectively. None of the participants had *Lachnospira* present at all the four time-points under investigation in this study.

Table 4.3 Longitudinal detection of FLVR bacteria from participants in our cohort using quantitative polymerase chain reaction

All	Time-points																		
	B	**6W**	**6M**	**12M**	**B only**	**6W_only**	**6M_only**	**12M_only**	**B+6W**	**B+6M**	**B+12M**	**6W+6M**	**6W+12M**	**6M+12M**	**B+6W+6M**	**B+6W+12M**	**B+6M+12M**	**6W+6M+12M**	**All**
Total number of participants with a sample	170	118	147	98	53	16	31	14	69	70	57	71	46	54	40	34	29	34	24
Participants positive for **F**	20.5% (35/170)	46.6% (55/118)	63.3% (93/147)	92.9% (91/98)	34% (18/53)	75% (12/16)	61.3% (19/31)	92.9% (13/14)	4.3% (3/69)	11.4% (8/70)	12.3% (7/57)	12.7% (9/71)	56.5% (26/46)	66.7% (36/54)	–	–	17.2% (5/29)	38.2% (13/34)	8.3% (2/24)
Participants positive for **L**	13.5% (23/170)	3.4% (4/118)	17.7% (26/147)	24.5% (24/98)	13.2% (7/53)	–	9.7% (3/31)	35.7% (5/14)	–	4.3% (3/70)	3.5% (2/57)	1.4% (1/71)	2.2% (1/46)	1.9% (1/54)	–	–	–	2.9% (1/34)	–
Participants positive for **V**	22.4% (38/170)	94.9% (112/118)	95.2% (140/147)	100% (98/98)	17% (9/53)	87.5% (14/16)	96.8% (30/31)	100% (14/14)	23.2% (16/69)	15.7% (11/70)	26.3% (15/57)	91.5% (65/71)	21.7% (10/46)	100% (54/54)	17.5% (7/40)	20.6% (7/34)	24.1% (7/29)	100% (34/34)	25% (6/24)
Participants positive for **R**	78.2% (133/170)	94.9% (112/118)	96.6% (142/147)	93.9% (92/98)	67.9% (36/53)	100% (16/16)	93.5% (29/31)	100% (14/14)	78.3% (54/69)	78.6% (55/70)	78.9% (45/57)	28.2% (20/71)	6.5% (3/46)	96.3% (52/54)	75% (30/40)	73.5% (25/34)	72% (21/29)	38.2% (13/34)	70.8% (17/24)
Wheeze																			
Total number of participants with a sample	50.6% (86/170)	50% (59/118)	49.7% (73/147)	50% (49/98)	49.1% (26/53)	50% (8/16)	48.4% (15/31)	50% (7/14)	50.7% (35/69)	51.4% (36/70)	50.9% (29/57)	49.3% (35/71)	50% (23/46)	50% (27/54)	50% (20/40)	50% (17/34)	51.7% (15/29)	50% (17/34)	50% (12/24)
Participants positive for **F**	13.9% (12/86)	45.8% (27/59)	60.3% (44/73)	91.8% (45/49)	3.8% (1/26)	50% (4/8)	60% (9/15)	85.7% (6/7)	5.7% (2/35)	8.3% (3/36)	17.2% (5/29)	14.3% (5/35)	52.2% (12/23)	55.6% (15/27)	–	–	20% (3/15)	29.4% (5/17)	8.3% (1/12)
Participants positive for **L**	13.9% (12/86)	5.1% (3/59)	17.8% (13/73)	22.4% (11/49)	19.2% (5/26)	–	13.3% (2/15)	7.4% (2/7)	–	2.8% (1/36)	3.4% (1/29)	2.9% (1/35)	4.3% (1/23)	–	–	–	–	–	–
Participants positive for **V**	27.9% (24/86)	98.3% (58/59)	94.5% (69/73)	100% (49/49)	19.2% (5/26)	100% (8/8)	100% (15/15)	100% (7/7)	28.6% (10/35)	13.9% (5/36)	34.5% (10/29)	94.3% (33/35)	21.7% (5/23)	100% (27/27)	20% (4/20)	23.5% (4/17)	33.3% (5/15)	100% (17/17)	33.3% (4/12)
Participants positive for **R**	74.4% (64/86)	91.5% (54/59)	95.9% (70/73)	100% (4949)	73.1% (19/26)	100% (8/8)	93.3% (14/15)	100% (7/7)	65.7% (23/35)	69.4% (25/36)	79.3% (23/29)	31.4% (11/35)	8.7% (2/23)	100% (27/27)	65% (13/20)	64.7% (11/17)	66.7% (10/15)	35.3% (6/17)	66.7% (8/12)
Non-wheeze																			
Total number of participants with a sample	49.4% (84/170)	50% (59/118)	50.3% (74/147)	50% (49/98)	50.9% (27/53)	50% (8/16)	51.6% (16/31)	50% (7/14)	49.3% (34/69)	48.6% (34/70)	49.1% (28/57)	50.7% (36/71)	50% (23/46)	50% (27/54)	50% (20/40)	50% (17/34)	48.3% (14/29)	50% (17/34)	50% (12/24)
Participants positive for **F**	27.4% (23/84)	47.5% (28/59)	66.2% (49/74)	93.9% (46/49)	62.9% (17/27)	100% (8/8)	62.5% (10/16)	100% (7/7)	2.9% (1/34)	14.7% (5/34)	7.1% (2/28)	11.1% (4/36)	60.9% (14/23)	77.8% (21/27)	–	–	14.3% (2/14)	47.1% (8/17)	8.3% (1/12)
Participants positive for **L**	13.1% (11/84)	1.7% (1/59)	17.6% (13/74)	26.5% (13/49)	7.4% (2/27)	–	6.3% (1/16)	42.9% (3/7)	–	5.9% (2/34)	3.6% (1/28)	–	–	3.7% (1/27)	–	–	–	5.9% (1/17)	–
Participants positive for **V**	16.7% (14/84)	91.5% (54/59)	95.9% (71/74)	100% (49/49)	14.8% (4/27)	75% (6/8)	93.8% (15/16)	100% (7/7)	17.6% (6/34)	17.6% (6/34)	17.9% (5/28)	88.9% (32/36)	21.7% (5/23)	100% (27/27)	15% (3/20)	17.6% (3/17)	14.3% (2/14)	100% (17/17)	16.7% (2/12)
Participants positive for **R**	82.1% (69/84)	98.3% (58/59)	97.3% (72/74)	87.8% (43/49)	62.9% (17/27)	100% (8/8)	93.8% (15/16)	100% (7/7)	91.2% (31/34)	88.2% (30/34)	78.6% (22/28)	25% (9/36)	4.3% (1/23)	92.6% (25/27)	85% (17/20)	82.3% (14/17)	78.6% (11/14)	41.2% (7/17)	75% (9/12)

F- *Faecalibacterium*; L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; B- Birth (first four days), 6W- six weeks (first four to eight weeks), 6M- six months (first five to seven months) and 12M- 12 months (first 11 to 13 months of an infant life) time-point.

3.2.3 FLVR prevalence in human stool samples

Using SYBR green qPCR, we detected *Rothia*, *Veillonella, Faecalibacterium* and *Lachnospira* in 90% (479/533), 73% (388/533), 51% (274/533) and 14% (77/533) of the samples, respectively (Table 4.4). The qPCR median bacterial load was $6.41{\times}10^9$, $3.39{\times}10^8$, $1.57{\times}10^9$ and $5.58{\times}10^9$ for *Faecalibacterium*, *Lachnospira, Veillonella* and *Rothia*, respectively (Table 4.4). Conventional PCR detected, *Rothia*, *Veillonella, Faecalibacterium* and *Lachnospira* in 55% (263/479), 74% (289/388), 53% (145/274) and 0% (0/77) of the qPCR positive samples, respectively (Table 4.4).

3.2.4 FLVR species in birth and six weeks samples of wheezing and non-wheezing infants

The birth and six weeks samples represent a time period within the first 100 days of an infant's life. *Faecalibacterium* conventional PCR primers allowed amplification from 14% (13/90) of the qPCR birth and six weeks positive samples (Table 4.4). These 13 samples belonged to eight wheezing and five non-wheezing infants. All 13 sequences from *Faecalibacterium* positive samples where 99-100% identical to the species *F. prausnitzii*.

Rothia conventional PCR primers showed amplification from 32.7% (80/245) of the qPCR birth and six weeks positive samples (Table 4.4). The 80 samples belonged to 31 wheezing infants (one infant had samples at both birth and six weeks) and 47 non-wheezing infants (one infant had samples at both birth and six weeks). All 80 sequences were 98-100% identical to *R. mucilaginosa*.

For the genus *Veillonella*, only 65% (60/93) of the conventional PCR positive samples (32 wheezing and 28 non-wheezing infants) yielded non-redundant sequences. Of the 60 sequences, 1/60 was 99% identical to *V. rogosae*, 8/60 were 99% identical to *V. parvula*, 14/60 were 97-99% identical to *V. dispar*, 17/60 were 97-99% identical to *V. atypica* and 20/60 were 92-94% identical to *V. atypica*. The remaining 33 sequences (15 wheezing and 17 non-wheezing infants) contained many redundant bases and we could not identify which species they belonged to. The identified *Veillonella* species were similarly distributed between wheeze and non-wheezing infants (Figure 4.3). However, *V. dispar* was more common in the wheezing group and *V. rogosae* was not detected in the non-wheezing group at all (Figure 4.3).

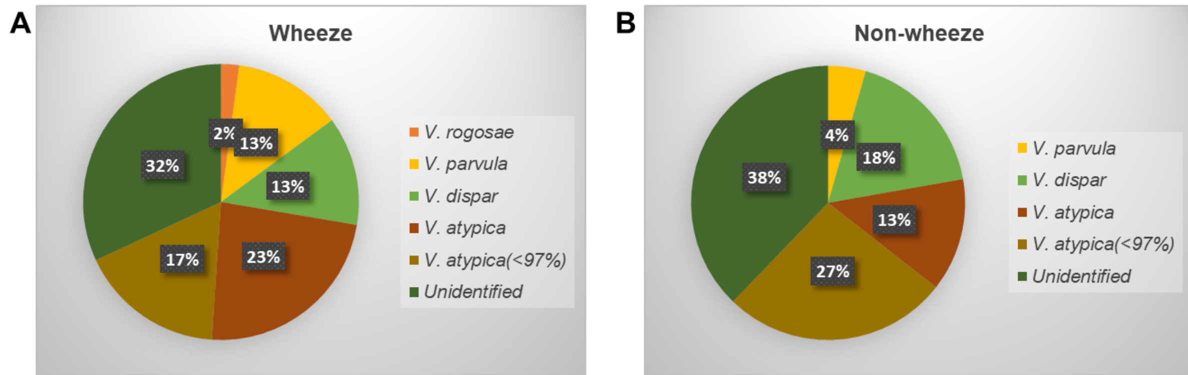


Figure 4.3 Destribution of *Veillonella* species (spp) in **A**. wheezing infants and **B.** non-wheezing infants. All sequences were at least 97% similar to the specified species, except for the category *V. atypica* (<97%) which contained sequences with percentage similarity less than 97%. The unidentified category contain sequences which could not be identified due to high redundancy

Table 4.4 FLVR positive samples identified by SYBR green qPCR, and conventional PCR

		All samples			Wheeze			Non-wheeze		
		Number of samples screened	**qPCR positive**	**Conventional PCR positive**	**Number of samples screened**	**qPCR positive**	**Conventional PCR positive**	**Number of samples screened**	**qPCR positive**	**Conventional PCR positive**
F	Birth	170	35 (20.6%)	3 (8.5%)	86 (50.6%)	12 (13.9%)	2 (16.7%)	84 (49.4%)	23 (27.4%)	1 (4.3%)
	6 weeks	118	55 (46.6%)	10 (18.2%)	59 (50%)	27 (45.8%)	6 (22.2%)	59 (50%)	28 (47.5%)	4 (14.3%)
	6 months	147	93 (63.3%)	52 (55.9%)	73 (49.7%)	44 (60.3%)	25 (56.8%)	74 (50.3%)	49 (66.2%)	27 (55.1%)
	12 months	98	91 (92.9%)	80 (87.9%)	49 (50%)	45 (91.8%)	38 (84.4%)	49 (50%)	46 (93.9%)	42 (91.3%)
	Total	533	**274 (51.4%)**	**145 (52.9%)**	**267 (50.1%)**	**128 (47.9%)**	**71 (55.5%)**	**266 (49.9%)**	**146 (54.9%)**	**74 (50.7%)**
	Median load (Q1 - Q3)		**6.41×10^{9} (2.89×10^{9} - 3.73×10^{10})**							
L	Birth	170	23 (13.5%)	-	86 (50.6%)	12 (13.9%)	-	84 (49.4%)	11 (13.1%)	-
	6 weeks	118	4 (3.4%)	-	59 (50%)	3 (5.1%)	-	59 (50%)	1 (1.7%)	-
	6 months	147	26 (17.7%)	-	73 (49.7%)	13 (17.8%)	-	74 (50.3%)	13 (17.6%)	-
	12 months	98	24 (24.5%)	-	49 (50%)	11 (22.4%)	-	49 (50%)	13 (26.5%)	-
	Total	533	**77 (78.6%)**	**-**	**267 (50.1%)**	**39 (14.6%)**	**-**	**266 (49.9%)**	**38 (14.3%)**	**-**
	Median load (Q1 - Q3)		**3.39×10^{8} (1.99×10^{8} – 5.264×10^{8})**							
V	Birth	170	38 (22.4%)	2 (5.3%)	86 (50.6%)	24 (27.9%)	1 (4.2%)	84 (49.4%)	14 (16.7%)	1 (7.1%)
	6 weeks	118	112 (94.9%)	91 (81.3%)	59 (50%)	58 (98.3%)	46 (79.3%)	59 (50%)	54 (91.5%)	45 (83.3%)
	6 months	147	140 (95.2%)	108 (77.1%)	73 (49.7%)	69 (94.5%)	51 (73.9%)	74 (50.3%)	71 (95.9%)	57 (80.3%)
	12 months	98	98 (100%)	88 (89.8%)	49 (50%)	49 (100%)	45 (91.8%)	49 (50%)	49 (100%)	43 (87.8%)
	Total	533	**388 (72.8%)**	**289 (74.5%)**	**267 (50.1%)**	**200 (74.9%)**	**143 (71.5%)**	**266 (49.9%)**	**188 (70.7)**	**146 (77.7%)**
	Median load (Q1 - Q3)		**1.57×10^{9} (7.28×10^{8} – 3.03×10^{9})**							
R	Birth	170	133 (78.2%)	7 (5.3%)	86 (50.6%)	64 (74.4%)	4 (6.3%)	84 (49.4%)	69 (82.1%)	3 (4.3%)
	6 weeks	118	112 (94.9%)	73 (65.2%)	59 (50%)	54 (91.5%)	29 (53.7%)	59 (50%)	58 (98.3%)	44 (75.9%)
	6 months	147	142 (96.6%)	126 (88.7%)	73 (49.7%)	70 (95.9%)	62 (88.6%)	74 (50.3%)	72 (97.3%)	64 (88.9%)
	12 months	98	92 (93.9%)	56 (60.9%)	49 (50%)	49 (100%)	27 (55.1%)	49 (50%)	43 (87.8%)	29 (67.4%)
	Total	533	**479 (89.9%)**	**263 (54.9%)**	**267 (50.1%)**	**237 (88.8%)**	**122 (51.5%)**	**266 (49.9%)**	**242 (90.9%)**	**141 (58.3%)**
	Median load (Q1 - Q3)		**5.58×10^{9} (2.59×10^{9} – 1.16×10^{10})**							

n= number of samples screened; qPCR- Quantitative Polymerase Chain Reaction; Q1- 25th percentile; Q3- 75th percentile; F- *Faecalibacterium*, L- *Lachnospira*, V- *Veillonella,* R- *Rothia*; Median bacterial load depicts the median DNA copy number calculated using the DNA amount that was obtained from the qPCR standard curve. Conventional PCR was performed only on qPCR positive samples. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant's life.

3.2.5 Factors associated with gut colonisation by FLVR bacteria

Irrespective of the time-points, the male gender (adjusted OR (aOR)=0.65, 95%CI: 0.42-0.98) and TC-Newman residence (aOR=0.52, 95%CI: 0.29 - 0.91) were associated with reduced colonisation by *Faecalibacterium* (Table 4.5). We observed that reduced stool colonisation by *Faecalibacterium* at six weeks (Table S4.1) was associated with male gender (aOR=0.19, 95%CI: 0.06-0.61), increase in birthweight (aOR=0.39, 95%CI: 0.15-0.91), TC-Newman residence (aOR=0.15, 95%CI: 0.03-0.78) and tobacco exposure after birth (aOR=0.20,95%CI: 0.04-0.88). At six months (Table S4.1), male gender (aOR=0.19, 95%CI: 0.06-0.61), TC-Newman residence (aOR=0.17, 95%CI: 0.04-0.69), vaginal delivery (aOR=0.15, 95%CI: 0.03-0.69) and smoking during pregnancy (aOR=0.21, 95%CI: 0.06-0.69) were associated with reduced stool colonisation by *Faecalibacterium*.

Breastfeeding (aOR=0.17, 95%CI: 0.05-0.49) was significantly associated with reduced stool colonisation by *Lachnospira* (Table 4.5), irrespective of the time-points. We observed that an increase in birth length (aOR=0.68, 95%CI: 0.44-0.96) as well as breastfeeding (aOR =0.01, 95%CI: 2.66e-04-0.23), were associated with reduced colonisation by *Lachnospira* at 12 months (Table S4.2). Lower colonisation by *Veillonella* was associated with TC-Newman residence (crude odds ratio (cOR)=0.52, 95%CI: 0.34-0.78) and breastfeeding (aOR=0.32, 95%CI: 0.10-0.91) in the overall univariate and multivariate models, respectively (Table 4.5). Irrespective of the time-points, mother's secondary education (crude OR=3.55, 95%CI: 1.10-9.78) and tertiary education (cOR=13.2, 95%CI: 1.91 - 265.35; aOR=11.73, 95%CI: 1.36-2.58) were significantly associated with *Rothia* colonisation (Table 4.5). However, *Rothia* colonisation at birth was associated with mother's secondary education (cOR=6.32, 95%CI: 1.00-49.73) in the univariate analysis (Table S4.4).

Table 4.5 Factors associated with the overall gut colonisation by FLVR bacteria within the first year of life

Co-variates	*Faecalibacterium*		*Lachnospira*		*Veillonella*		*Rothia*	
	cOR (95% CI)	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**
Wheezing	0.76 (0.54-1.06)	0.72 (0.43- 1.20)	1.02 (0.63-1.66)	1.01 (0.50-2.02)	1.24 (0.85-1.82)	1.11 (0.62-1.99)	0.78 (0.44-1.38)	0.80 (0.33-1.92)
Baby gender (male)	**0.61 (0.43-0.87)** **	**0.65 (0.42- 0.98)** *	0.81 (0.49-1.33)	0.87 (0.48-1.55)	1.03 (0.69-1.53)	1.19 (0.74-1.90)	0.62 (0.33-1.14)	0.83 (0.39-1.75)
Birth length (cm)	0.99 (0.94-1.04)	1.00 (0.92-1.09)	0.97 (0.90-1.04)	0.97 (0.86-1.09)	0.98 (0.93-1.04)	0.97 (0.88-1.07)	1.01 (0.93-1.09)	1.03 (0.88-1.19)
Birth weight (z-score)	0.92 (0.77-1.09)	0.82 (0.59-1.12)	0.88 (0.69-1.12)	0.98 (0.63-1.53)	0.95 (0.78-1.16)	1.08 (0.75-1.57)	1.02 (0.75-1.36)	0.90 (0.50-1.62)
Breastfeeding (Yes)	**0.54 (0.32-0.92)** *	0.47 (0.20-1.07)	**0.52 (0.28-1.01)** *	**0.17 (0.05-0.49)** **	0.56 (0.28-1.04)	**0.32 (0.10-0.91)** *	0.39 (0.09-1.09)	0.18 (0.01-1.15)
Delivery mode (vaginal)	1.00 (0.66-1.52)	0.92 (0.53-1.58)	0.77 (0.44-1.38)	0.79 (0.41-1.62)	1.11 (0.69-1.75)	1.01 (0.54-1.84)	1.09 (0.53-2.10)	1.08 (0.39-2.69)
Ever had pneumonia (Yes)	0.82 (0.58-1.16)	0.77 (0.45-1.30)	0.69 (0.41-1.15)	0.65 (0.31- 1.35)	1.24 (0.839-1.85)	0.92 (0.50-1.68)	0.95 (0.54-1.69)	1.21 (0.48-3.02)
Monthly household income								
<R1000	1.46 (0.87-2.48)	0.99 (0.52-1.89)	0.86 (0.44-1.75)	0.87 (0.38-2.07)	1.96 (1.09-3.47)	1.18 (0.58-2.38)	1.26 (0.52-2.84)	0.67 (0.21-1.94)
R1000-5000	1.07 (0.64-1.79)	0.93 (0.52-1.72)	0.65 (0.33-1.34)	0.49 (0.22-1.13)	1.42 (0.81-2.44)	1.05 (0.55-1.99)	0.95 (0.62-1.47)	1.03 (0.33-2.89)
Pets exposure (Cat and/or Dog)	0.84 (0.59-1.19)	1.09 (0.70-1.70)	0.77 (0.46-1.26)	0.75 (0.41-1.36)	0.83 (0.56-1.23)	0.94 (0.57-1.54)	0.75 (0.41-1.36)	0.73 (0.34-1.56)
Cat (Yes)	0.85 (0.59-1.22)	1.39 (0.77-2.52)	0.72 (0.41-1.22)	1.05 (0.46-2.42)	0.71 (0.47-1.06)	1.04 (0.54-1.99)	0.77 (0.42-1.44)	1.01 (0.37-2.69)
Dog (Yes)	0.76 (0.53-1.08)	0.74 (0.42-1.28)	0.72 (0.42-1.19)	0.64 (0.29-1.34	0.72 (0.48-1.07)	0.75 (0.41-1.39)	0.62 (0.34-1.13)	0.56 (0.22-1.43)
Residential area (TC-Newman)	**0.53 (0.37-0.76)** ***	**0.52 (0.29-0.91)** *	0.96 (0.59-1.58)	1.39 (0.62-3.15)	**0.52 (0.34-0.78)** **	0.54 (0.28-1.05)	0.69 (0.37-1.25)	0.96 (0.35-2.64)
Smoking during pregnancy (Yes)	**0.58 (0.39-0.86)** **	0.77 (0.46-1.29)	1.03 (0.59-1.75)	1.25 (0.61-2.54)	0.76 (0.49-1.18)	1.02 (0.58-1.82)	0.68 (0.37-1.29)	0.92 (0.39-2.23)
HIV exposure (Yes)	1.54 (0.99-2.40)	0.74 (0.37-1.45)	1.15 (0.61-2.05)	0.48 (0.15-1.34)	1.51 (0.91-2.60)	0.69 (0.32-1.50)	1.62 (0.75-4.02)	0.79 (0.27-2.64)
Mom education								
Secondary	0.97 (0.38-2.45)	0.98 (0.32-2.96)	1.59 (0.45-10.16)	1.80 (0.31-34.78)	1.30 (0.45-3.38)	1.28 (0.37-3.96)	**3.55 (1.10-9.78)** *	3.51 (0.83-1.30)
Tertiary	1.11 (0.36-3.37)	1.08 (0.29-4.09)	1.06 (0.19-8.16)	1.59 (0.19-34.95)	1.49 (0.42-5.04)	1.51 (0.35-6.35)	**13.2 (1.91-265.35)** *	**11.73 (1.36-2.58)** *
Full term gestation (Yes)	1.60 (0.68-3.95)	2.77 (0.84-9.43)	1.08 (0.36-4.67)	1.54 (0.30-10.64)	0.82 (0.27-2.12)	0.81 (0.16-3.15)	0.97 (0.15-3.48)	0.92 (0.04-7.07)
Tobacco exposure after birth (Yes)	0.70 (0.44-1.11)	0.77 (0.41-1.44)	0.62 (0.35-1.13)	0.60 (0.27-1.37)	0.91 (0.53-1.50)	1.34 (0.65-2.74)	0.71 (0.28-1.52)	1.35 (0.42-4.02)
Suppliment use (Yes)	1.33 (0.51-3.52)	1.16 (0.38-3.74)	1.42 (0.39-9.05)	2.28 (0.39-43.84)	1.74 (0.63-4.50)	0.97 (0.25-3.14)	0.51 (0.03-2.55)	0.65 (0.03-3.97)
Cotinine exposure (Yes)	0.87 (0.55-1.37)	1.47 (0.82-2.65)	0.94 (0.51-1.82)	1.20 (0.56-2.74)	0.85 (0.49-1.41)	1.05 (0.52-2.05)	0.67 (0.27-1.45)	0.61 (0.16-1.84)

- significance (- $P < 0.05$, **- $P < 0.01$, ***- $P < 0.001$); cOR- crude odds ratio; aOR- adjusted odds ratio; CI- Confidence Interval.

3.3 Association of FLVR bacteria with wheezing and recurrent-wheezing outcomes

3.3.1 Wheezing outcome

We observed a significantly high bacterial load (DNA copies) for *Lachnospira* at birth (Figure 4.4B, Table 4.6) in the wheezing group as compared to the non-wheezing group (P = 0.00562). In contrast, *Rothia* was significantly reduced in the birth samples of wheezing infants (P<0.001) (Figure 4.4D, Table 4.6). However, at 12 months (Figure 4.4D, Table 4.6) *Rothia* was significantly high in wheezing compared to non-wheezing infants (P = 0.0328).

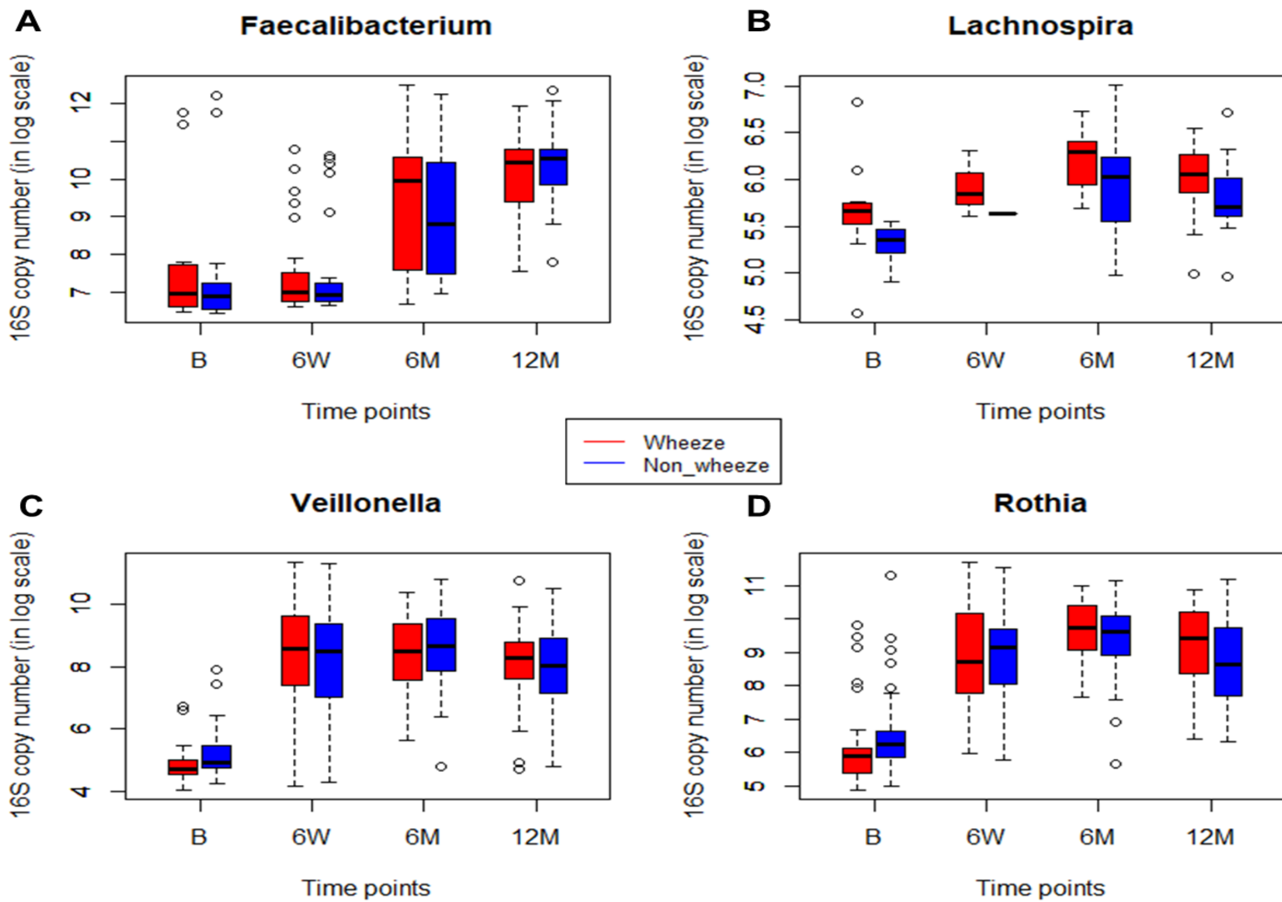


Figure 4.4 Difference in FLVR bacterial load (DNA copies) between wheezing (n=140) and non-wheezing (n=140) infants as determined by quantitative polymerase chain reaction (qPCR). The graphs are based on logarithmic scale (base 10) **A,** differences in *Faecalibacterium* load. **B,** difference in *Lachnospira* load. **C,** differences in *Veillonella* load. **D,** differences in *Rothia* load. B-Birth (representing the first four days), 6W- six weeks (first four to eight weeks), 6M- six months (first five to seven months) and 12M- 12 months (first 11 to 13 months of an infant's life) time-points.

Table 4.6 Mann-Whitney test *P*-values for the difference in FLVR bacterial load (DNA copies) between wheeze and non-wheezing infants at four time-points

	Mann-Whitney test, *P*-value			
	Birth	6 Weeks	6 Months	12 Months
Faecalibacterium	0.439	0.519	0.364	0.499
Lachnospira	0.00562 **	1.000	0.064	0.277
Veillonella	0.106	0.648	0.391	0.557
Rothia	<0.001***	0.937	0.127	0.0328*

*- significant (**- P < 0.01, ***- P < 0.001)

Table 4.7 shows the conditional logistic regression results based on FLVR bacterial detection. The univariate model (Table 4.7) showed a significantly low risk for developing wheezing in infants who had *Faecalibacterium* at birth (crude odds ratio (cOR)=0.43, 95%CI: 0.19-0.94), as well as at birth and/or six weeks (cOR=0.58, 95%CI: 0.33-0.99). In the multivariate model, we observed a significant association of *Rothia* at six weeks (adjusted odds ratio (aOR)=0.02, 95%CI: 0.0006-0.97) with low risk of wheezing (Table 4.7).

A significantly high risk of wheezing (aOR=2.20, 95%CI: 1.15-4.95) was observed in infants with high *Faecalibacterium* load at six months (Table 4.8). Both the univariate (cOR=0.59, 95%CI: 0.36-0.88) and multivariate (aOR=0.54, 95%CI: 0.28-0.93) conditional logistic regression, showed a significantly low risk of wheezing in infants with high *Rothia* load at birth (Table 4.8).

Table 4.7 Association of the detection of FLVR bacteria with the wheezing outcome determined by conditional logistic regression

		Wheeze			Non-wheeze				
		Number of participants screened	Positive	Negative	Number of participants screened	Positive	Negative	cOR (95% CI)	aOR (95% CI)
F	Birth (n= 170)	86 (50.6%)	12 (13.9%)	74 (86.1%)	84 (49.4%)	23 (27.4%)	61 (72.6%)	**0.43 (0.19-0.94)** *	0.49 (0.15-1.67)
	6 weeks (n = 118)	59 (50%)	27 (45.8%)	32 (54.2%)	59 (50%)	28 (47.5%)	31 (52.5%)	0.93 (0.45-1.92)	0.42 (0.09-1.75)
	6 months (n = 147)	73 (49.7%)	44 (60.3%)	29 (39.7%)	74 (50.3%)	49 (66.2%)	25 (33.8%)	0.78 (0.39-1.52)	0.78 (0.22-2.83)
	12 months (n = 98)	49 (50%)	45 (91.8%)	4 (8.2%)	49 (50%)	46 (93.9%)	3 (6.1%)	0.74 (0.16-3.45)	2.52 (0.16-38.86)
	≤ 3 months (n = 219) #	110 (50.2%)	36 (32.7%)	74 (67.3%)	109 (49.8%)	50 (45.9%)	59 (54.1%)	**0.58 (0.33-0.99)** *	0.46 (0.19-1.06)
	> 3 months (n = 191) ##	95 (49.7%)	73 (76.8%)	22 (23.2%)	96 (50.3%)	74 (77.1%)	22 (22.9%)	0.99 (0.50-1.93)	1.30 (0.40-4.22)
	Overall (n = 280) ###	140 (50%)	88 (62.9%)	52 (37.1%)	140 (50%)	101 (72.1%)	39 (27.9%)	0.65 (0.39-1.08)	0.53 (0.24-1.16)
L	Birth (n= 170)	86 (50.6%)	12 (13.9%)	74 (86.1%)	84 (49.4%)	11 (13.1%)	73 (86.9%)	1.08 (0.45-2.59)	0.77 (0.19-2.99)
	6 weeks (n = 118)	59 (50%)	3 (5.1%)	56 (94.9%)	59 (50%)	1 (1.7%)	58 (98.3%)	3.03 (0.31-29.74)	13.15 (0.59-288.69)
	6 months (n = 147)	73 (49.7%)	13 (17.8%)	60 (82.2%)	74 (50.3%)	13 (17.6%)	61 (82.4%)	1.02 (0.44-2.37)	1.03 (0.22-4.82)
	12 months (n = 98)	49 (50%)	11(22.4%)	38 (77.6%)	49 (50%)	13 (26.5%)	36 (73.5%)	0.80 (0.32-2.01)	0.65 (0.13-3.31)
	≤ 3 months (n = 219) #	110 (50.2%)	15 (13.6%)	95 (86.4%)	109 (49.8%)	12 (11%)	97 (89%)	1.29 (0.57-2.89)	1.14 (0.35-3.72)
	> 3 months (n = 191) ##	95 (49.7%)	24 (25.3%)	71 (74.7%)	96 (50.3%)	24 (25%)	72 (75%)	1.01 (0.53-1.95)	0.94 (0.32-2.73)
	Overall (n = 280) ###	140 (50%)	35 (25%)	105 (75%)	140 (50%)	32 (22.9%)	108 (77.1%)	1.12 (0.65-1.95)	1.01 (0.44-2.31)
V	Birth (n= 170)	86 (50.6%)	24 (27.9%)	62 (72.1%)	84 (49.4%)	14 (16.7%)	70 (83.3%)	1.93 (0.92-4.04)	1.51 (0.54-4.23)
	6 weeks (n = 118)	59 (50%)	58 (98.3%)	1 (1.7%)	59 (50%)	54 (91.5%)	5 (8.5%)	5.31 (0.60-46.62)	2.13 (0.09-45.26)
	6 months (n = 147)	73 (49.7%)	69 (94.5%)	4 (5.5%)	74 (50.3%)	71 (95.9%)	3 (4.1%)	0.73 (0.16-3.37)	1.04 (0.05-20.19)
	12 months (n = 98)	49 (50%)	49 (100%)	_	49 (50%)	49 (100%)	_	0.99 (0.94- 1.05)	0.97 (0.92- 1.02)
	≤ 3 months (n = 219) #	110 (50.2%)	72 (65.5%)	38 (34.5%)	109 (49.8%)	61 (55.9%)	48 (44%)	1.49 (0.86-2.56)	1.01 (0.43-2.38)
	> 3 months (n = 191) ##	95 (49.7%)	91 (95.8%)	4 (4.2%)	96 (50.3%)	93 (96.9%)	3 (3.1%)	0.73 (0.16-3.36)	1.07 (0.08 -13.91)
	Overall (n = 280) ###	140 (50%)	118 (84.3%)	22 (15.7%)	140 (50%)	113 (80.7%)	27 (19.3%)	1.29 (0.68-2.38)	1.02 (0.37-2.77)
R	Birth (n= 170)	86 (50.6%)	64 (74.4%)	22 (25.6%)	84 (49.4%)	69 (82.1%)	15 (17.9%)	0.63 (0.30-1.33)	0.77 (0.23-2.58)
	6 weeks (n = 118)	59 (50%)	54 (91.5%)	5 (8.5%)	59 (50%)	58 (98.3%)	1 (1.7%)	0.19 (0.02-1.66)	**0.02 (0.0006-0.97)** *
	6 months (n = 147)	73 (49.7%)	70 (95.9%)	3 (4.1%)	74 (50.3%)	72 (97.3%)	2 (2.7%)	0.65 (1.11-3.98)	0.50 (0.03-7.59)
	12 months (n = 98)	49 (50%)	49 (100%)	_	49 (50%)	43 (87.8%)	6 (12.2%)	_	_
	≤ 3 months (n = 219) #	110 (50.2%)	95 (86.4%)	15 (13.6%)	109 (49.8%)	96 (88.1%)	13 (11.9%)	0.86 (0.39-1.89)	0.74 (0.19-2.78)
	> 3 months (n = 191) ##	95 (49.7%)	92 (96.8%)	3 (3.2%)	96 (50.3%)	90 (93.8%)	6 (6.3%)	2.04 (0.49-8.37)	2.49 (0.34-18.00)
	Overall (n = 280) ###	140 (50%)	131 (93.6%)	9 (6.4%)	140 (50%)	129 (92.1%)	11 (7.9%)	1.24 (0.49-3.09)	1.64 (0.36-7.45)

F- *Faecalibacterium*; L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; *- significance (*- $P < 0.05$, **- $P < 0.01$, ***- $P < 0.001$); n- total number of participants with a sample; cOR- crude odds ratio; aOR- adjusted odds ratio; CI- confidence Interval; ≤ 3 months- Birth and/or 6 weeks; #- depicts the number of babies who had samples at birth and/or six weeks; overall- detection of the bacteria in an infant irrespective of the time-points; ###- depicts the number of infants who had at least one of the four time-points with a sample; >3 months- six and/or 12 months; ##- depicts the number of infants who had samples at six and/or 12 months. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant's life.

Table 4.8 Association of FLVR bacterial load (DNA copies) with the wheezing outcome determined by conditional logistic regression

		Wheeze Median load (Q1 – Q3)	Non-wheeze Median load (Q1 – Q3)	cOR (95% CI)	aOR (95% CI)
F	Birth	6.95 (6.62 – 7.68)	6.89 (6.54 – 7.23)	1.17 (0.75-1.82)	-
	6 weeks	7.01 (6.75 – 7.39)	6.92 (6.77 – 7.19)	1.03 (0.66-1.60)	0.41 (0.08- 1.15)
	6 months	9.95 (7.69 – 10.56)	8.81 (7.49 – 10.43)	1.17 (0.89-1.54)	2.20 (1.15- 4.95) *
	12 months	10.46 (9.40 – 10.77)	10.54 (9.88 – 10.78)	0.76 (0.47-1.18)	0.54 (0.18- 1.36)
L	Birth	5.66 (5.53 – 5.73)	5.35 (5.22 – 5.47)	23.03 (1.28- 2339.65)	-
	6 weeks	5.84 (5.73 – 5.96)	5.63 (5.63 – 5.63)	-	-
	6 months	6.29 (5.95 – 6.39)	6.03 (5.55 – 6.23)	5.47 (0.84- 57.96)	-
	12 months	6.06 (5.86 – 6.24)	5.71 (5.60 - 6.02)	2.38 (0.36- 20.75)	-
V	Birth	4.73 (4.58 – 4.89)	4.93 (4.77 – 5.39)	0.58 (0.24- 1.23)	-
	6 weeks	8.59 (7.43 – 9.64)	8.49 (7.03 – 9.37)	1.00 (0.99- 1.01)	1.26 (0.82- 1.96)
	6 months	8.49 (7.57 – 9.37)	8.66 (7.86 – 9.52)	0.88 (0.66- 1.17)	0.69 (0.36- 1.24)
	12 months	8.28 (7.59 – 8.79)	8.02 (7.16 – 8.92)	1.12 (0.81- 1.56)	1.13 (0.67- 1.94)
R	Birth	5.91 (5.43 – 6.13)	6.27 (5.86 – 6.65)	0.59 (0.36- 0.88) *	0.54 (0.28- 0.93) *
	6 weeks	8.72 (7.83 – 10.17)	9.14 (8.05 – 9.70)	1.01 (0.76- 1.34)	1.15 (0.72- 1.91)
	6 months	9.74 (9.09 – 10.37)	9.63 (8.92 – 10.09)	1.39 (0.96- 2.07)	1.06 (0.51- 2.41)
	12 months	9.43 (8.37 – 10.22)	8.64 (7.70 – 9.72)	1.47 (1.03- 1.14)	1.79 (0.94- 3.76)

F- *Faecalibacterium*; L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; *- significance (*- $P < 0.05$, **- $P < 0.01$, ***- $P < 0.001$); Q1- 25th percentile; Q3- 75th percentile; cOR- crude odds ratio; aOR- adjusted odds ratio; CI- confidence interval; Median bacterial load depicts the median DNA copy number (in log base 10 scale) calculated using the DNA amount that was obtained from the qPCR standard curve. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant's life.

3.3.2 *Recurrent-wheezing outcome*

Using generalized logistic regression model, we did not find any significant association between FLVR bacterial detection with recurrent-wheezing (Table 4.9).

Like the wheezing outcome, we observed a significantly low risk of recurrent-wheezing (cOR=0.32, 95%CI: 0.11-0.78; aOR=0.29, 95%CI: 0.05-0.88) in infants with high *Rothia* load at birth (Table 4.10).

Table 4.9 Association of the detection of FLVR bacteria with recurrent-wheezing determined by generalised logistic regression

		Recurrent-Wheeze			Non-recurrent wheeze				
		Number of participants screened	Positive	Negative	Number of participants screened	Positive	Negative	cOR (95% CI)	aOR (95% CI)
F	Birth (n= 80)	24 (30%)	2 (8.3%)	22 (91.7%)	56 (70%)	10 (17.9%)	46 (82.1%)	0.31 (0.05-1.14)	0.18 (0.01- 1.11)
	6 weeks (n = 58)	19 (32.8%)	6 (31.6%)	13 (68.4%)	39 (67.2%)	20 (51.3%)	19 (48.7%)	0.46 (0.15-1.27)	0.34 (0.04-2.18)
	6 months (n = 71)	22 (31%)	12 (54.5%)	10 (45.5%)	49 (69%)	31 (63.3%)	18 (36.7%)	0.62 (0.25-1.59)	0.59 (0.14-2.39)
	12 months (n = 49)	13 (26.5%)	13 (100%)	_	36 (73.5%)	32 (88.9%)	4 (11.1%)	-	-
	≤ 3 months (n = 104) #	30 (28.8%)	7 (23.3%)	23 (76.7%)	74 (71.2%)	28 (37.8%)	46 (62.2%)	0.41 (0.16-0.97)	0.39 (0.12- 1.14)
	> 3 months (n = 93) ##	26 (27.9%)	20 (76.9%)	6 (23.1%)	67 (72%)	52 (77.6%)	15 (22.3%)	0.93 (0.37-2.71)	1.06 (0.29-4.18)
	Overall (n = 134) ###	36 (26.9%)	22 (61.1%)	14 (38.9%)	98 (73.1)	64 (65.3%)	34 (34.7%)	0.68 (0.33-1.44)	0.66 (0.26-1.70)
L	Birth (n= 80)	24 (30%)	2 (8.3%)	22 (91.7%)	56 (70%)	8 (14.3%)	48 (85.7%)	0.56 (0.09- 2.14)	0.16 (0.02- 0.99)
	6 weeks (n = 58)	19 (32.8%)	_	19 (100%)	39 (67.2%)	3 (7.7%)	36 (92.3%)	-	-
	6 months (n = 71)	22 (31%)	6 (27.3%)	16 (72.7%)	49 (69%)	7 (14.3%)	42 (85.7%)	1.88 (0.61- 5.20)	2.06 (0.47- 8.56)
	12 months (n = 49)	13 (26.5%)	1 (7.7%)	12 (92.3%)	36 (73.5%)	10 (27.8%)	26 (72.2%)	1.24 (0.01- 1.24)	0.48 (0.02- 6.16)
	≤ 3 months (n = 104) #	30 (28.8%)	2 (6.7%)	28 (93.3%)	74 (71.2%)	11 (14.9%)	63 (85.1%)	0.50 (0.08-1.84)	0.22 (0.02-1.23)
	> 3 months (n = 93) ##	26 (27.9%)	7 (26.9%)	19 (73.1%)	67 (72%)	17 (25.4%)	50 (74.6%)	1.07 (0.39-2.63)	0.98 (0.26- 3.38)
	Overall (n = 134) ###	36 (26.9%)	9 (25%)	27 (75%)	98 (73.1%)	24 (24.5%)	74 (75.5%)	1.06 (0.44- 2.31)	0.82 (0.27- 2.25)
V	Birth (n= 80)	24 (30%)	6 (25%)	18 (75%)	56 (70%)	17 (30.4%)	39 (69.6%)	1.14 (0.39- 2.99)	0.56 (0.13- 1.99)
	6 weeks (n = 58)	19 (32.8%)	19 (100%)	_	39 (67.2%)	38 (97.4%)	1 (2.6%)	-	-
	6 months (n = 71)	22 (31%)	22 (100%)	_	49 (69%)	46 (93.9%)	3 (6.1%)	-	-
	12 months (n = 49)	13 (26.5%)	13 (100%)	_	36 (73.5%)	36 (100%)	_	-	-
	≤ 3 months (n = 104) #	30 (28.8%)	21 (70%)	9 (30%)	74 (71.2%)	49 (66.2%)	25 (33.8%)	1.49 (0.87- 2.58)	0.87 (0.29- 2.67)
	> 3 months (n = 93) ##	26 (27.9%)	26 (100%)	_	67 (72%)	64 (95.5%)	3 (4.5%)	0.73 (0.14- 3.42)	-
	Overall (n = 134) ###	36 (26.9%)	32 (88.9%)	4 (11.1%)	98 (73.1%)	83 (84.7%)	15 (15.3%)	1.28 (0.69- 2.39)	1.82 (0.48- 9.29)
R	Birth (n= 80)	24 (30%)	17 (70.8%)	7 (29.2%)	56 (70%)	45 (80.4%)	11 (19.6%)	0.54 (0.21- 1.53)	0.47 (0.12- 1.88)
	6 weeks (n = 58)	19 (32.8%)	17 (89.5%)	2 (10.5%)	39 (67.2%)	36 (92.3%)	3 (7.7%)	0.37 (0.07- 2.81)	2.16 (0.07-71.97)
	6 months (n = 71)	22 (31%)	22 (100%)	_	49 (69%)	47 (95.9%)	2 (4.1%)	-	-
	12 months (n = 49)	13 (26.5%)	13 (100%)	_	36 (73.5%)	36 (100%)	_	-	-
	≤ 3 months (n = 104) #	30 (28.8%)	27 (90%)	3 (10%)	74 (71.2%)	66 (89.2%)	8 (10.8%)	1.13 (0.36-5.03)	0.71 (0.15- 4.52)
	> 3 months (n = 93) ##	26 (27.9%)	26 (100%)	_	67 (72%)	65 (97%)	2 (3%)	-	-
	Overall (n = 134) ###	36 (26.9%)	35 (97.2%)	1 (2.8%)	98 (73.1%)	93 (94.9%)	5 (5.1%)	2.57 (0.49- 47.09)	1.10 (0.17- 21.88)

F- *Faecalibacterium*; L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; n- total number of wheezing participants with a sample; cOR- crude odds ratio; aOR- adjusted odds ratio; CI- confidence interval; ≤ 3 months- Birth and/or 6 weeks; #- depicts the number of wheezing infants who had samples at birth and/or six weeks; >3 months- six and/or 12 months; ##- depicts the number of infants who had samples at six and/or 12 months; overall- detection of the bacterium irrespective of the time-points; ###- depicts the number of wheezing infants who had at least one of the four time-points with a sample. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant's life.

Table 4.10 Association of FLVR bacterial load with recurrent-wheezing determined by generalised logistic regression

		Recurrent-wheeze Median load (Q1 – Q3)	Non-recurrent wheeze Median load (Q1 – Q3)	cOR (95% CI)	aOR (95% CI)
F	Birth	7.74 (7.71 – 7.77)	6.76 (6.59 – 7.39)	1.08 (0.29- 2.11)	-
	6 weeks	6.98 (6.78 – 7.09)	7.04 (6.86 – 7.83)	1.03 (0.44- 1.87)	2.90 (0.66- 34.14)
	6 months	9.87 (8.12 – 10.34)	9.91 (7.59 -10.62)	1.07 (0.72- 1.61)	1.05 (0.59- 1.90)
	12 months	10.37 (9.40 – 10.69)	10.48 (9.65 – 10.80)	0.74 (0.41- 1.36)	1.48 (0.52- 5.14)
L	Birth	5.66 (5.65 – 5.66)	5.66 (5.53 – 5.85)	2.29 (7.00- 50.16)	-
	6 weeks	-	5.84 (5.73 – 6.08)	-	-
	6 months	6.37 (6.29 – 6.52)	6.24 (5.80 – 6.35)	1.29 (1.08- 433.56)	-
	12 months	6.54 (6.54 – 6.54)	6.01 (5.85 – 6.19)	877.66 (0.63- 3.39e+10)	-
V	Birth	4.66 (4.47 – 4.78)	4.77 (4.66 – 5.11)	0.32 (0.02- 1.33)	-
	6 weeks	8.70 (7.58 – 9.78)	8.49 (7.14 – 9.53)	1.06 (0.80- 1.42)	1.04 (0.65- 1.65)
	6 months	8.64 (8.05 – 9.39)	8.54 (7.48 – 9.37)	1.01 (0.69- 1.52)	0.93 (0.55- 1.59)
	12 months	7.56 (7.24 – 8.14)	8.61 (8.08 – 8.85)	0.65 (0.41- 1.02)	0.62 (0.26- 1.41)
R	Birth	5.79 (5.62 – 5.99)	5.91 (5.28 – 6.22)	**0.32 (0.11- 0.78) ***	**0.29 (0.05- 0.88) ***
	6 weeks	8.73 (8.21 – 10.34)	8.58 (7.67 – 9.87)	1.21 (0.82- 1.82)	1.45 (0.72- 3.13)
	6 months	10.21 (9.49 – 10.53)	9.70 (9.05 – 10.29)	1.69 (0.95-3.22)	1.65 (0.74- 3.96)
	12 months	9.39 (8.37 – 10.18)	9.46 (8.44 – 10.25)	1.19 (0.73- 2.05)	1.01 (0.44- 2.26)

F- *Faecalibacterium*; L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; *- significant ($P < 0.05$); Q1- 25th percentile; Q3- 75th percentile; cOR- crude odds ratio; aOR- adjusted odds ratio; CI- confidence interval; Median bacterial load depicts the median DNA copy number (in log base 10 scale) calculated using the DNA amount that was obtained from the qPCR standard curve. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant's life.

3.4 Sensitivity of FLVR bacterial load within the first four days of life in predicting the wheezing outcome

Overall, FLVR bacteria performed the least in predicting any wheeze (140 wheezing and 140 non-wheezing infants) outcome as compared to predicting wheeze without LRTIs (50 wheeze and 124 non-wheeze infants) and wheeze with LRTIs (90 wheezing and 17 non-wheezing infants) (Table 4.11, Figure 4.4-4.7).

3.4.1 Each FLVR bacterium alone

In predicting any wheeze, *Lachnospira* and *Rothia* showed the highest performance among all FLVR bacteria. The area under receiver operating characteristics curve (AUC) for *Lachnospira* was 0.833 (95%CI: 0.64-1.00) and that of *Rothia* was 0.707 (95%CI: 0.62-0.79) (Table 4.11, Figure 4.4A). Except for *Faecalibacterium,* the performance of all other bacteria improved in predicting wheeze without LRTIs (No-LRTIs wheeze) compared to predicting any wheeze (Table 4.11, Figure 4.4B). *Lachnospira* still exhibited the highest AUC of 0.98 (95%CI: 0.93-1.00), followed by *Veillonella* (AUC=0.758, 95%CI: 0.50-1.00) and *Rothia* (AUC=0.757, 95%CI: 0.64-0.88). Among all FLVR bacteria, only the performance of *Rothia* dropped in predicting wheeze with LRTIs (LRTIs wheeze) when compared to predicting any wheeze (Table 4.11, Figure 4.4C). *Lachnospira* showed the highest AUC of 0.857 and *Rothia* (AUC=0.656, 95%CI: 0.46-0.85) showed the least performance (Table 4.11, Figure 4.4C). Overall, except for *Faecalibacterium* the performance of other FLVR bacteria was highest in predicting no-LRTIs wheeze compared to predicting any wheeze as well as LRTIs wheeze (Table 4.11, Figure 4.4).

3.4.2 Two by two interactions of FLVR bacteria

Except for *Lachnospira*, the two by two interactions improved the performance of other FLVR bacterium alone in predicting the wheezing outcome (Table 4.11, Figure 4.5). In predicting any wheeze, the highest performance was from *Faecalibacterium* and *Veillonella* (FV) (AUC=0.739, 95%CI: 0.61-0.87) as well as *Faecalibacterium* and *Rothia* (FR) (AUC=0.736, 95%CI: 0.65-0.82) (Table 4.11, Figure 4.5A). The performance increased in predicting no-LRTIs wheeze (Table 4.11, Figure 4.5B), with the interaction of *Faecalibacterium* and *Lachnospira* (FL) demonstrating the highest AUC of 0.785 (95%CI: 0.55-1.00). The performance of FLVR bacteria further increased in predicting LRTIs wheeze, with *Faecalibacterium* and *Veillonella* (FV) exhibiting the highest AUC of 0.864 (95%CI: 0.70-1.00) (Table 4.11, Figure 4.5C).

3.4.3 Three by three interactions of FLVR bacteria

Like the two by two interactions, the three by three interactions also improved the performance of each FLVR bacterium alone in predicting any wheeze, except for *Lachnospira* (Table 4.11,

Figure 4.6). The highest performance in predicting any wheeze (AUC=0.749, 95%CI: 0.67-0.83) as well as no-LRTIs wheeze (AUC=0.792, 95%CI: 0.69-0.89) was from the interaction of *Faecalibacterium, Veillonella* and *Rothia* (FVR) (Table 4.11, Figure 4.6A-B). In predicting LRTIs wheeze, the interaction of *Faecalibacterium, Lachnospira* and *Veillonella* (FLV) demonstrated the highest performance (AUC=0.857, 95%CI: 0.69-1.00) (Table 4.11, Figure 4.6C).

3.4.4 Four by four interactions of FLVR bacteria

FLVR bacterial four by four interactions demonstrated the highest performance in predicting LRTIs wheeze (AUC=0.839, 95%CI: 0.70-0.97), followed by no-LRTIs wheeze (AUC=0.797, 95%CI: 0.69-0.89) and then any wheeze (AUC=0.763, 95%CI: 0.68-0.84) (Table 4.11, Figure 4.7).

3.5 Sensitivity of FLVR bacteria at six weeks, six months and 12 months in predicting any wheeze

After comparing the performance of FLVR bacteria at all time-points in predicting any wheeze, the birth time-point (section 3.4) showed the highest performance (Table 4.11, Table S4.5).

3.5.1 Each bacterium alone

As for the birth time-point, *Lachnospira* still exhibited the highest performance at six weeks (AUC=0.667), six months (AUC=0.716, 95%CI: 0.50-0.93) and 12 months (AUC=0.636, 95%CI: 0.39-0.88) (Table S4.5, Figure S4.1A-C).

3.5.2 Two by two interactions of FLVR bacteria

Table S4.5 and Figure S4.2 display the two by two interactive performance of FLVR bacteria at six weeks, six months and 12 months in predicting any wheeze. The highest two by two performance at six weeks (Table S4.5, Figure S4.2A) was from *Veillonella* and *Rothia* (VR) (AUC = 0.60, 95%CI: 0.49-0.70). At six months (Table S4.5, Figure S4.2B), the interaction of *Faecalibacterium* and *Lachnospira* (FL) exhibited the highest performance (AUC=0.606, 95%CI: 0.49-0.72). The highest two by two interactive performance at 12 months (Table S4.5, Figure S4.2C) was that of *Faecalibacterium* and *Rothia* (FR) (AUC=0.687, 95%CI: 0.58-0.79).

3.5.3 Three by three interactions of FLVR bacteria

The three by three interactive performance of FLVR bacteria at six weeks, six months and 12 months in predicting any wheeze is shown in Table S4.5 and Figure S4.3A. At six weeks (Table S4.5, Figure S4.3A), the combination of *Lachnospira, Veillonella* and *Rothia* (LVR) showed the highest AUC of 0.637 (95%CI: 0.54-0.74). The highest performance at six months (AUC =0.647, 95%CI: 0.56-0.74) was from the combination of *Faecalibacterium*, *Lachnospira* and *Veillonella* (FLV) (Table S4.5, Figure S4.3B). The interactive performance of *Faecalibacterium,*

Veillonella and *Rothia* (FVR) with an AUC of 0.739 (95%CI: 0.64-0.84) was the highest at 12 months (Table S4.5, Figure S4.3C).

3.5.4 Four by four interactions of FLVR bacteria

Following the birth time-point, the next highest four by four performance was at 12 months (AUC=753, 95%CI: 0.66-0.85), six months (AUC=679, 95%CI: 0.59-0.77) and then six weeks (AUC=0.655, 95%CI: 0.56-0.75) (Table S4.5, Figure S4.3D).

Table 4.11 Area under the curve and corresponding confidence intervals of FLVR bacterial load at birth in predicting any wheeze, no-LRTIs wheeze as well as LRTIs wheeze

		Birth					
		Any wheeze		No-LRTIs wheeze		LRTIs wheeze	
		AUC	95%CI	AUC	95%CI	AUC	95%CI
Each bacterium	F	0.583	0.37- 0.79	0.50	0.028-0.97	0.756	0.48-1.00
	L	0.833	0.64-1.00	0.98	0.93-1.00	0.857	-
	V	0.661	0.48- 0.85	0.758	0.50-1.00	0.706	-
	R	0.707	0.62- 0.79	0.757	0.64-0.88	0.656	0.46-0.85
2by2 interactions	FL	0.644	0.48- 0.80	0.785	0.55-1.00	0.789	0.58-0.99
	FV	0.739	0.61- 0.87	0.735	0.50-0.97	0.864	0.70-1.00
	FR	0.736	0.65- 0.82	0.80	0.69-0.91	0.754	0.59-0.92
	LV	0.564	0.41- 0.72	0.514	0.30-0.73	0.786	0.35-1.00
	LR	0.707	0.62- 0.79	0.738	0.62-0.86	0.601	0.39-0.81
	VR	0.709	0.62- 0.79	0.733	0.61-0.85	0.753	0.57-0.93
3by3 interactions	FLV	0.694	0.58- 0.81	0.715	0.54-0.89	0.857	0.69-1.00
	FLR	0.732	0.65- 0.82	0.776	0.67-0.89	0.798	0.65-0.94
	FVR	0.749	0.67- 0.83	0.792	0.69-0.89	0.798	0.64-0.95
	LVR	0.702	0.62- 0.79	0.726	0.61-0.84	0.787	0.62-0.95
4by4 interactions	FLVR	0.763	0.68- 0.84	0.797	0.69-0.89	0.839	0.70-0.97

F- *Faecalibacterium,* L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; AUC- area under the curve; CI- confidence interval; LRTIs- lower respiratory tract infections. The AUC and CI are calculated from qPCR bacterial load (DNA copies) converted into log base 10 scale. Birth- first four days of an infant's life.

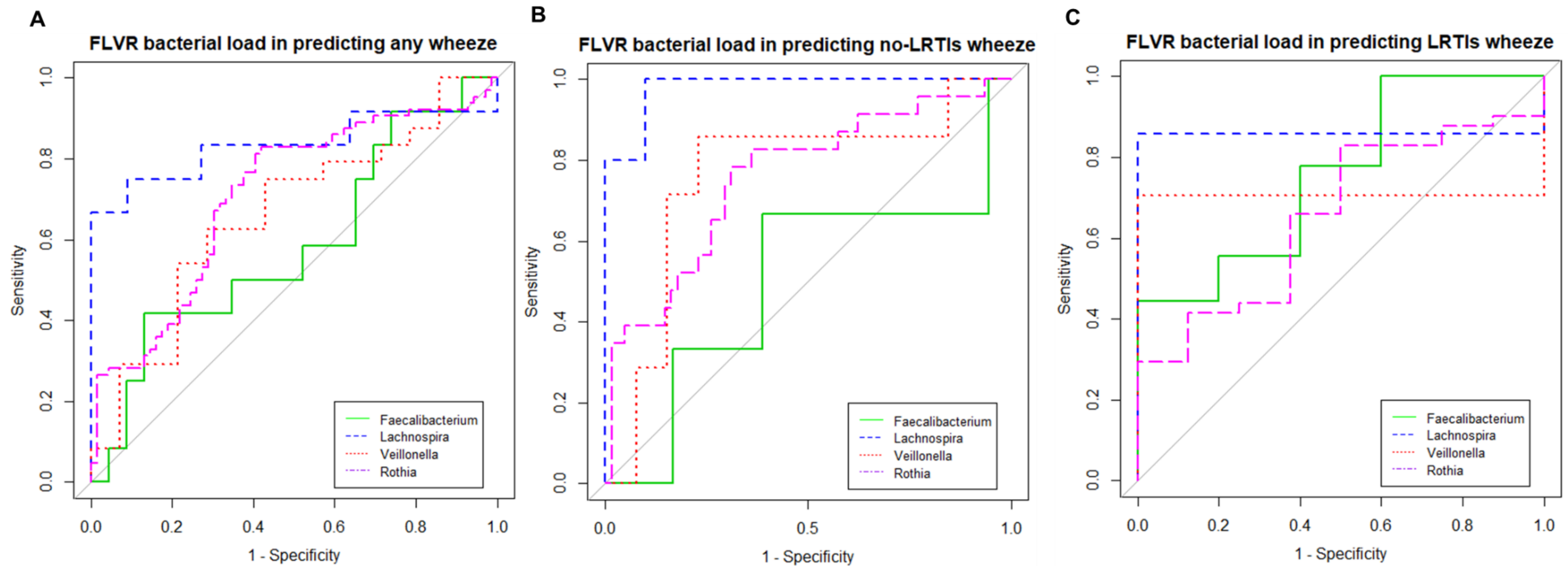


Figure 4.4 Receiver operating characteristic (ROC) curves of FLVR bacterial load (DNA copies) at birth (first four days of life), determined by quantitative polymerase chain reaction (qPCR). **A.** FLVR bacteria predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants). **B.** FLVR bacteria predicting no-LRTIs wheeze (sub-cohort of 50 wheezing and 124 non-wheezing infants). **C.** FLVR bacteria predicting LRTIs wheeze (sub-cohort of 90 wheezing and 17 non-wheezing infants). Plots are based on logarithmic scale (log base 10) of DNA copies.

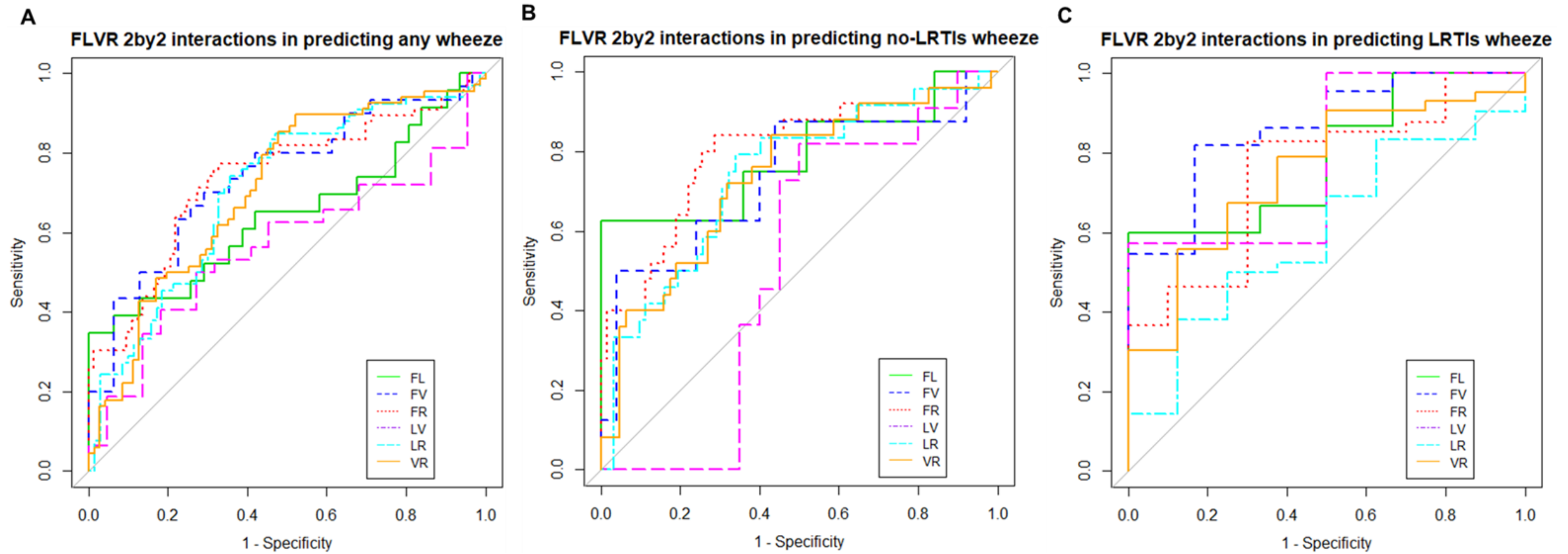


Figure 4.5 Receiver operating characteristic (ROC) curves for the two by two interactions of FLVR bacterial load (DNA copies) at birth (first four days of life), determined by quantitative polymerase chain reaction (qPCR). **A.** FLVR bacteria predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants). **B.** FLVR bacteria predicting no-LRTIs wheeze (sub-cohort of 50 wheezing and 124 non-wheezing infants). **C.** FLVR bacteria predicting LRTIs wheeze (sub-cohort of 90 wheezing and 17 non-wheezing infants). Plots are based on logarithmic scale (log base 10) of DNA copies.

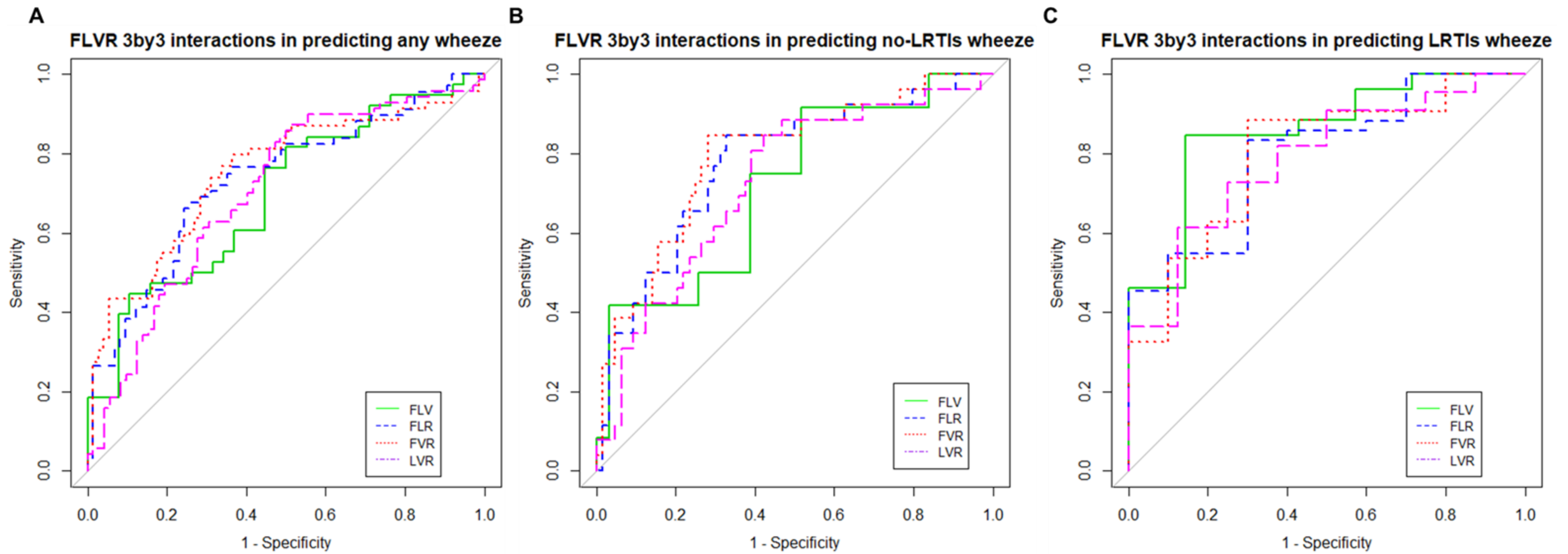


Figure 4.6 Receiver operating characteristic (ROC) curves for the three by three interactions of FLVR bacterial load (DNA copies) at birth (first four days of life), determined by quantitative polymerase chain reaction (qPCR). **A.** FLVR bacteria predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants). **B.** FLVR bacteria predicting no-LRTIs wheeze (sub-cohort of 50 wheezing and 124 non-wheezing infants). **C.** FLVR bacteria predicting LRTIs wheeze (sub-cohort of 90 wheezing and 17 non-wheezing infants). Plots are based on logarithmic scale (log base 10) of DNA copies.

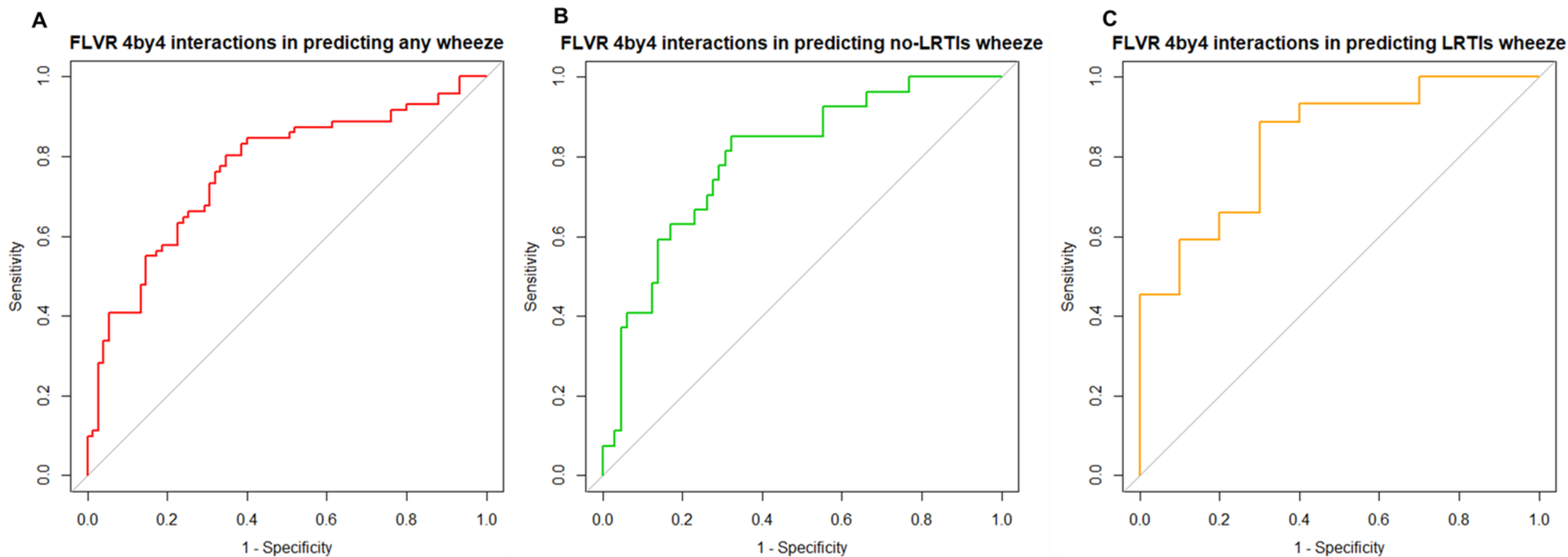


Figure 4.7 Receiver operating characteristic (ROC) curves for the four by four interactions of FLVR bacterial load (DNA copies) at birth (first four days of life), determined by quantitative polymerase chain reaction (qPCR). **A.** FLVR bacteria predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants). **B.** FLVR bacteria predicting no-LRTIs wheeze (sub-cohort of 50 wheezing and 124 non-wheezing infants). **C.** FLVR bacteria predicting LRTIs wheeze (sub-cohort of 90 wheezing and 17 non-wheezing infants). Plots are based on logarithmic scale (log base 10) of DNA copies.

Discussion

Subsequent to Arrieta et al. [18], this is the second study to find a significantly reduced *Rothia* load in the early life stool samples of wheezing infants as compared to non-wheezing infants. A conditional logistic regression adjusted for potential confounders further confirmed that infants at risk of wheezing in our study had a lower amount of *Rothia* within the first four days of life. We also noticed that the presence of this bacterium from the fourth to eighth week of age (six weeks' time-point), irrespective of its load is associated with low wheezing risk. Although the difference in terms of *Rothia* load was not significant at the six weeks' time-point, the median load was still lower in wheezing compared to non-wheezing infants. According to Arrieta et al., infants at risk of asthma (based on the asthma predictive index) with atopy and wheeze symptoms have a reduced relative abundance of *Rothia* in their guts at three months of age [18]. Though we did not have asthma predictive data for our infants, these similar findings provide strong evidence that a high presence of *Rothia* within the first three months may potentially be protective against wheezing and asthma. In addition, we observed similar findings for recurrent wheezing which is likely to indicate an asthma risk. Studies have shown that infants who experience recurrent-wheezing are more at risk of asthma [27-29]. The human gut is initially colonised by aerobes and facultative anaerobes [30-32]. These consume all the oxygen thus creating an oxygen free environment for the anaerobes [30-32]. *Rothia* is a facultative anaerobe and thus an early gut coloniser, which means that it should be present in the gut of every infant. Hence, the reduced amount of *Rothia* observed in these infants may be contributing to their wheezing symptoms. Arrieta et al. reported that *Rothia* was the second most abundant after *Lachnospira,* in baby mice showing reduced airway inflammation and which were born to mothers supplemented with FLVR bacteria [18]. This suggest that supplying wheezing infants with *Rothia* might alleviate the wheezing symptoms. However, findings on *Lachnospira* were contradictory in our study. As mentioned earlier (section 3.2.1), *Lachnospira* qPCR primers were less specific. It is possible that we missed out on some *Lachnospira* and this is why we got contradicting results. In addition, the findings on *Lachnospira* from Arrieta et al. could also be specific to atopic-wheeze only. However, we were unable to prove this scenario since the data on atopy was not collected in the Drakenstein study. Nevertheless, it is reported that atopy may be less associated with infant wheezing and asthma in low- and middle-income countries (LMIC) such as South Africa, compared to developed countries such as Canada [33]. Therefore, it is possible that our results were not affected by the lack of atopy data at all. Apart from *Rothia* and *Lachnospira*, we did not observe any association of *Faecalibacterium* and/or *Veillonella* with childhood wheezing. This agrees with the latest study by Arrieta et al.[34] which observed no significant association between FLVR bacteria and infant wheezing in rural Ecuador. This study demonstrated the impact of geographic location

on the human microbiota. Therefore, the association of FLVR bacteria observed in our study as well as the Canadian study could also be influenced by the geographical location.

This study further showed that the first four days of an infant's life represent a period at which FLVR bacteria may be used in predicting the likelihood of infants to wheeze later in life. Based on the ROC curve analysis, *Lachnospira* and *Rothia* may serve as early biomarkers for diagnosing infant wheezing. We also noticed that except for *Lachnospira,* interaction of FLVR bacteria increases their performance in predicting wheezing compared to each bacterium alone. This suggests that co-occurrence of FLVR bacteria may be more important in childhood wheeze and thus should be investigated further. However, we observed that using *Lachnospira* as a single factor performs much better compared to when it is combined with other FLVR bacteria. Our study also demonstrated the potential impact of LRTIs on the sensitivity of FLVR bacteria in predicting wheezing. This was shown when the predictive performance of FLVR bacteria increased after separating wheeze with LRTIs from wheeze without LRTIs, compared to when they were in combination. However, this separation has resulted in an unbalanced number of wheeze and non-wheezing infants. Hence, it could be possible that the increased sensitivity observed was due to this imbalance. Therefore, comprehensive study designs of wheeze with LRTIs as well as without LRTIs are needed to verify these findings as our study was designed only across general wheeze. Interestingly we observed that infants at risk of wheezing in our cohort had a high *Faecalibacterium* load at six months. This is intriguing because as we mentioned in chapter 2 (systematic review), *Faecalibacterium* is usually associated with health (35, 36). However, this is the first study to observe this in the stool samples of wheezing and non-wheezing infants at six months. Hence, it is important for future studies to confirm this finding. Furthermore, two strains of *Faecalibacterium* have been defined (37). Therefore, the strain of *Faecalibacterium* observed to be associated with the risk of wheezing in this study, may be different from that which is usually associated with health. This suggests a need for strain level analysis in future studies aiming to investigate the association of *Faecalibacterium* with human diseases.

The single melt peak for *Rothia* observed in this study could be the reason why only the species *R. mucilaginosa* was identified from all the sequenced samples. Studies have shown that most species of this genus are predominant in the oral cavity (38), therefore it is very likely that *R. mucilaginosa* is the only species which inhabits the gut. In contrast, *Lachnospira* qPCR primers produced multiple peaks in addition to that of the positive control which is a signal of multiple amplicons being produced. This generation of different amplicons is not surprising given that we are using SYBR green which produces fluorescence upon intercalating into any double stranded DNA (39, 40). Since our qPCR primers target the 16S rRNA which is common to all bacteria (41), it is very likely that these primers bind to a similar region in bacteria other

than our target. However, the generation of more than one peak can also be explained by the fact that our qPCR primers are genus specific. Therefore, distinct species or strains of the same genus which differ slightly in their amplified regions could have been detected thus resulting in amplicons with different melting temperatures. This scenario is more likely with *Faecalibacterium* for which the two peaks are much closer in melting temperatures. This could be the reason why only one species of *Faecalibacterium* was detected in the sequenced samples. As a result, the two melt peaks observed are most likely to be due to the presence of two distinct strains rather than species of *Faecalibacterium*. In support of our findings, two strains of *F. prausnitzii* have been characterized because of the differences in their 16S rRNA sequences [37], as we mentioned above.

The decline in the number of qPCR positive samples after conventional PCR analysis, could be attributed to the low sensitivity of conventional compared to qPCR primers. The observed high sensitivity of *Veillonella* primers may be the reason for the high number of samples positive for this genus, as this means that the primers can detect low DNA copy number of this bacterium from samples. For both *Rothia* and *Veillonella*, the median qPCR bacterial load (5.58×10^9 and 1.57×10^9) was higher than the conventional PCR primers LOD (8.147×10^6 and 9.159×10^6), which explains the high number of positive samples for these two bacteria observed in our study. This means that the load of *Rothia* and *Veillonella* in most samples was high enough to be detected by the primers. Although *Lachnospira* qPCR primers demonstrated much sensitivity, very few samples were positive for this bacterium as compared to others. This could mean that *Lachnospira* was generally not present in most of these samples, or it could be present in very low concentrations. This is further verified by the inability to detect *Lachnospira* at all by conventional PCR. We see that the median bacterial load (3.39×10^8) detected by qPCR is below the LOD (2.046×10^9) for conventional PCR primers. As mentioned in chapter 3, the inability of these primers to detect low DNA concentrations was confirmed by lowering the annealing temperature. In addition, the conventional PCR primers of *Lachnospira* are targeting a single copy *rpoB* gene unlike the qPCR primers which target the multi-copy 16S rRNA. This could also be the reason for not detecting this bacterium, as there may be less bacterial DNA copies for these primers to pick up. However, the reduced number of positive samples for *Rothia* and *Faecalibacterium* is very likely due to the lower sensitivity of the conventional PCR primers. This is because the primers of *Rothia* and *Faecalibacterium* were all targeting the 16S rRNA gene which exist in multiple copies, hence it should be easy to detect this gene from sample. The high sensitivity of real-time qPCR over conventional PCR has been reported previously [42], thus it is not surprising that we found the same results in our experiments.

We found that 23% more wheezing infants had *F. prausnitzii* compared to non-wheezing infants at or before six weeks of age. In contrast, we observed a 17.5% increase in the number of non-wheezing infants who had *R. mucilaginosa* compared to the wheezing group. We also observed a 10% increase in the proportion of wheezing infants with *V. parvula* and *V. atypica* as compared to the non-wheezing group, while *V. dispar* was nearly equal between the two groups. However, the observed results might have been affected by the unidentified high redundant *Veillonella* sequences. The high redundancy means that so many *Veillonella* species were present in these samples, such that the sequencing instrument could not distinguish them based on the region targeted by our primers. Alternative methods such as designing of species-specific primers, cloning [43], and droplet PCR [44] could have allowed us to distinguish these species. However, we could not proceed further due to budget and time constraints. For all bacteria, the proportion of infants with the identified species might have been affected by the reduced number of samples sequenced. The use of more sensitive conventional PCR primers to increase the number of sequenced samples, would have allowed us to make better comparison between the wheezing and non-wheezing infants based on the identified species. Furthermore, sequencing of all other positive samples in addition to the birth and six weeks samples could also have increased our comparison strength. Nevertheless, due to financial constraints we sequenced only samples representing the first 100 days of an infant's life as it has been shown that wheezing infants exhibit stool bacterial alterations during this period [18].

Compared to other bacteria, the facultative aerobe *Rothia* was detected in more infants at birth and continued to colonize more infants until six weeks before it drops at 12 months. However, the anaerobes *Faecalibacterium, Lachnospira* and *Veillonella* were detected in few infants at birth and the number rose with increasing age. Our study is the first of its kind to identify factors associated with gut colonisation by FLVR bacteria. We showed that male infants, and those from TC-Newman had lower odds of being colonised with *Faecalibacterium* during the first year of life. In the Drakenstein study, the TC-Newman resident serves a mixed-race population whereas Mbekweni consist of a black-African population both of which belong to a low socio-economic status [20]. However, approximately 69.3% of TC-Newman residents in our cohort have a monthly household income of at least R1000, as compared to Mbekweni with only about 46.2% households earning such amount. Therefore, less colonisation with *Faecalibacterium* in TC-Newman may be associated with the high-income status in this area compared to Mbekweni. *Lachnospira* [45] and *Veillonella* [46] proliferate very slowly in the presence of lactose. This could explain the reduced colonisation with these two bacteria observed in breastfed infants in our study. In support of our findings, another study has found lower abundance of *Veillonella* and *Lachnospira* species in breastfed infants [47]. In contrast, colonisation with *Rothia* was very high for infants whose mothers have tertiary education. *Rothia* is known

for degrading gluten, a protein component of cereal grains [48]. It could be that infants of mothers with tertiary education were fed with cereals during their first year of life, thus facilitating colonisation with *Rothia*. This study showed that FLVR bacteria are detectable in human stool samples using molecular-based methods. This is in line with another study which has shown that among other bacteria, FLVR bacteria are present in the human gastrointestinal tract [49]. In support of this, findings from our review (Chapter 2) also show that most of the included studies (38%) detected FLVR bacteria in human stool samples [50-55].

In conclusion, this study strengthens the evidence that *Rothia* may potentially be protective against childhood wheezing and recurrent-wheezing. Stool-based quantification of *Lachnospira* and *Rothia* within the first four days of an infant's life may serve as an early diagnostic marker for childhood wheezing. The association of *Faecalibacterium* with high risk of infant wheezing observed in this study needs to be investigated further. Future studies using very specific *Lachnospira* primers are needed as our primers may have missed to detect some members of this genus. Since, this is the first study investigating FLVR bacteria and their combinations in predicting wheezing, we encourage more studies to elucidate our findings. Furthermore, due to the small recurrent wheezing sample size, we could not determine the sensitivity of FLVR bacteria in predicting recurrent-wheezing. We therefore encourage future studies with much bigger sample sizes to investigate the predictive ability of FLVR bacteria in recurrent wheezing. Moreover, the 16S rRNA primers may produce non-specific amplification; however, production of melt curves in SYBR green qPCR allows identification of positive samples. There are many species of *Veillonella* in human stool samples which could not be distinguished by our conventional PCR primers. Hence, a more sensitive detection method should be applied in future studies aiming to identify *Veillonella* species from stool samples. A more sensitive method able to detect very low DNA concentrations should also be acquired for the detection of *Lachnospira* from human stool samples. Although the multi-copy feature of the 16S rRNA has proven to increase PCR amplifications' sensitivity, existence of many copy variants within one bacterium may cause difficulties in identifying the bacterial species [56]. However, the *rpoB* gene is a valuable alternative gene target in this case as it exists only in one copy [56]. In addition, through *in-silico* analysis (chapter 3) we ensured that our resulting amplicons will have enough variation so as to distinguish the various species within the respective genus. Nevertheless, the multicopy feature of the 16S rRNA may have hindered the species level identification of *Faecalibacterium* and *Rothia* in our study and should therefore be considered in future studies. We further suggest future studies to determine the species differentiating capacity of other genes such as *tuf, gyrA, gyrB, sodA* as well as heat shock proteins which have proven ability to separate certain bacterial species [56]. Finally, this is the first study to

identify factors associated with gut colonisation by FLVR bacteria. Therefore, more studies are needed to further elucidate our findings.

References cited

1. Teijeiro A, Badellino H, Raiden M, Cuello M, Kevorkof G, Gatti C, et al. Risk factors for recurrent wheezing in the first year of life in the city of Córdoba, Argentina. Allergologia et immunopathologia. 2017;45(3):234-9.
2. Bacharier LB. Viral-induced wheezing episodes in preschool children: approaches to therapy. Curr Opin Pulm Med. 2010;16(1):31-5.
3. Ducharme FM, Sze MT, Chauhan B. Diagnosis, management, and prognosis of preschool wheeze. The Lancet. 2014;383(9928):1593-604.
4. Muglia C, Oppenheimer J. Wheezing in Infancy: An Overview of Recent Literature. Current Allergy and Asthma Reports. 2017;17(10):67.
5. Alvarez-Alvarez I, Niu H, Guillen-Grima F, Aguinaga-Ontoso I. Meta-analysis of prevalence of wheezing and recurrent wheezing in infants. Allergologia et immunopathologia. 2016.
6. Stevens C, Turner D, Kuehni C, Couriel J, Silverman M. The economic impact of preschool asthma and wheeze. European Respiratory Journal. 2003;21(6):1000-6.
7. Hoch HE, Collaco JM. Recurrent Wheezing in Childhood and Palivizumab. Am Thoracic Soc; 2017.
8. Pearce N, Aït-Khaled N, Beasley R, Mallol J, Keil U, Mitchell E, et al. Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). Thorax. 2007;62(9):758-66.
9. Asher M, Keil U, Anderson H, Beasley R, Crane J, Martinez F, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. European respiratory journal. 1995;8(3):483-91.
10. Stocks J, Sonnappa S. Early life influences on the development of chronic obstructive pulmonary disease. Therapeutic advances in respiratory disease. 2013;7(3):161-73.
11. Martinez FD. The origins of asthma and chronic obstructive pulmonary disease in early life. Proceedings of the American Thoracic Society. 2009;6(3):272-7.
12. Litonjua AA, Weiss ST. Risk factors for asthma. UpToDate. 2015.
13. Budden KF, Gellatly SL, Wood DL, Cooper MA, Morrison M, Hugenholtz P, et al. Emerging pathogenic links between microbiota and the gut-lung axis. Nat Rev Microbiol. 2017;15(1):55-63.
14. van Nimwegen FA, Penders J, Stobberingh EE, Postma DS, Koppelman GH, Kerkhof M, et al. Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. Journal of Allergy and Clinical Immunology. 2011;128(5):948-55. e3.
15. Stiemsma L, Arrieta M-C, Dimitriu P, Cheng J, Thorson L, Lefebvre D, et al. Shifts in *Lachnospira* and *Clostridium sp.* in the 3-month stool microbiome are associated with preschool-age asthma. Clinical Science. 2016:CS20160349.
16. Hasegawa K, Linnemann RW, Mansbach JM, Ajami NJ, Espinola JA, Petrosino JF, et al. The fecal microbiota profile and bronchiolitis in infants. Pediatrics. 2016:e20160218.
17. Raghunath P. Role of Gut Microbiota and Infectious Burden in the Development of Autoimmune and Allergic Diseases. Iranian Journal of Allergy, Asthma and Immunology. 2017;16(1):77.
18. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152-307ra152.
19. Schuijt TJ, Lankelma JM, Scicluna BP, e Melo FdS, Roelofs JJ, de Boer JD, et al. The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. Gut. 2015:gutjnl-2015-309728.

20. Zar H, Barnett W, Myer L, Stein D, Nicol M. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. Thorax. 2014:thoraxjnl-2014-206242.
21. Gray D, Willemse L, Visagie A, Czövek D, Nduru P, Vanker A, et al. Determinants of early-life lung function in African infants. Thorax. 2016:thoraxjnl-2015-207401.
22. Stein D, Koen N, Donald K, Adnams C, Koopowitz S, Lund C, et al. Investigating the psychosocial determinants of child health in Africa: The Drakenstein Child Health Study. Journal of neuroscience methods. 2015;252:27-35.
23. le Roux DM, Myer L, Nicol MP, Zar HJ. Incidence of childhood pneumonia: facility-based surveillance estimate compared to measured incidence in a South African birth cohort study. BMJ open. 2015;5(12):e009111.
24. Ririe KM, Rasmussen RP, Wittwer CT. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Analytical biochemistry. 1997;245(2):154-60.
25. Connolly MA, Liang K-Y. Conditional logistic regression models for correlated binary data. Biometrika. 1988;75(3):501-6.
26. McDonald JH. Handbook of biological statistics: sparky house publishing Baltimore, MD; 2009.
27. Chong Neto HJ, Rosário NA, Grasselli EA, Carnieli e Silva F, Bojarski LdFM, Rosário CS, et al. Recurrent wheezing in infants: epidemiological changes. Jornal de pediatria. 2011;87(6):547-50.
28. Castro-Rodríguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. American journal of respiratory and critical care medicine. 2000;162(4):1403-6.
29. Guilbert TW, Morgan WJ, Krawiec M, Lemanske Jr RF, Sorkness C, Szefler SJ, et al. The Prevention of Early Asthma in Kids study: design, rationale and methods for the Childhood Asthma Research and Education network. Controlled clinical trials. 2004;25(3):286-310.
30. Adlerberth I, Wold A. Establishment of the gut microbiota in Western infants. Acta paediatrica. 2009;98(2):229-38.
31. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006;118(2):511-21.
32. Bezirtzoglou E. The intestinal microflora during the first weeks of life. Anaerobe. 1997;3(2-3):173-7.
33. Claassen-Weitz S, Wiysonge CS, Machingaidze S, Thabane L, Horsnell WG, Zar HJ, et al. Current knowledge and future research directions on fecal bacterial patterns and their association with asthma. Frontiers in microbiology. 2016;7:838.
34. Arrieta M-C, Arévalo A, Stiemsma L, Dimitriu P, Chico ME, Loor S, et al. Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. Journal of Allergy and Clinical Immunology. 2018;142(2):424-34. e10.
35. Miquel S, Martin R, Rossi O, Bermudez-Humaran L, Chatel J, Sokol H, et al. *Faecalibacterium prausnitzii* and human intestinal health. Current opinion in microbiology. 2013;16(3):255-61.
36. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux J-J, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proceedings of the National Academy of Sciences. 2008;105(43):16731-6.
37. Martín R, Miquel S, Benevides L, Bridonneau C, Robert V, Hudault S, et al. Functional characterization of novel *Faecalibacterium prausnitzii* strains isolated from healthy volunteers: a step forward in the use of *F. prausnitzii* as a next-generation probiotic. Frontiers in microbiology. 2017;8:1226.

38. Tsuzukibashi O, Uchibori S, Kobayashi T, Umezawa K, Mashimo C, Nambu T, et al. Isolation and identification methods of *Rothia* species in oral cavities. Journal of microbiological methods. 2017;134:21-6.
39. Cao H, Shockey JM. Comparison of TaqMan and SYBR Green qPCR methods for quantitative gene expression in tung tree tissues. Journal of agricultural and food chemistry. 2012;60(50):12296-303.
40. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, et al. The real-time polymerase chain reaction. Molecular aspects of medicine. 2006;27(2):95-125.
41. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences. 1985;82(20):6955-9.
42. Schmittgen TD, Zakrajsek BA, Mills AG, Gorn V, Singer MJ, Reed MW. Quantitative reverse transcription–polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Analytical biochemistry. 2000;285(2):194-204.
43. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual: Cold spring harbor laboratory press; 1989.
44. Markey AL, Mohr S, Day PJ. High-throughput droplet PCR. Methods. 2010;50(4):277-81.
45. Cornick NA, Jensen N, Stahl D, Hartman P, Allison M. Lachnospira pectinoschiza sp. nov., an anaerobic pectinophile from the pig intestine. International Journal of Systematic and Evolutionary Microbiology. 1994;44(1):87-93.
46. Marcobal A, Barboza M, Froehlich JW, Block DE, German JB, Lebrilla CB, et al. Consumption of human milk oligosaccharides by gut-related microbes. Journal of agricultural and food chemistry. 2010;58(9):5334-40.
47. Wang M, Li M, Wu S, Lebrilla CB, Chapkin RS, Ivanov I, et al. Fecal microbiota composition of breast-fed infants is correlated with human milk oligosaccharides consumed. Journal of pediatric gastroenterology and nutrition. 2015;60(6):825.
48. Zamakhchari M, Wei G, Dewhirst F, Lee J, Schuppan D, Oppenheim FG, et al. Identification of *Rothia* bacteria as gluten-degrading natural colonizers of the upper gastro-intestinal tract. PloS one. 2011;6(9):e24455.
49. Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Reddy DN. Role of the normal gut microbiota. World journal of gastroenterology: WJG. 2015;21(29):8787.
50. Maji A, Misra R, Dhakan DB, Gupta V, Mahato NK, Saxena R, et al. Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. Environmental microbiology. 2018;20(1):402-19.
51. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152.
52. Hasegawa K, Linnemann RW, Mansbach JM, Ajami NJ, Espinola JA, Petrosino JF, et al. The Fecal Microbiota Profile and Bronchiolitis in Infants. Pediatrics. 2016;138(1).
53. Luo M, Liu Y, Wu P, Luo DX, Sun Q, Zheng H, et al. Alternation of Gut Microbiota in Patients with Pulmonary Tuberculosis. Frontiers in physiology. 2017;8:822.
54. Stiemsma LT, Arrieta MC, Dimitriu PA, Cheng J, Thorson L, Lefebvre DL, et al. Shifts in *Lachnospira* and *Clostridium sp.* in the 3-month stool microbiome are associated with preschool age asthma. Clinical science (London, England : 1979). 2016;130(23):2199-207.
55. Vernocchi P, Del Chierico F, Quagliariello A, Ercolini D, Lucidi V, Putignani L. A Metagenomic and in Silico Functional Prediction of Gut Microbiota Profiles May Concur in Discovering New Cystic Fibrosis Patient-Targeted Probiotics. Nutrients. 2017;9(12).
56. Reller LB, Weinstein MP, Petti CA. Detection and identification of microorganisms by gene amplification and sequencing. Clin Infect Dis. 2007;44(8):1108-14.

Supplementary Tables

Table S4.1 Factors associated with gut colonisation by *Faecalibacterium* at birth, six weeks, six months and twelve months

Covariates	*Faecalibacterium* Birth		Six weeks		Six months		Twelve months	
	cOR (95% CI)	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**
Wheezing	0.43 (0.19 - 0.92)	0.49 (0.13 - 1.62)	0.93 (0.45 - 1.93)	0.41 (0.09 - 1.78)	0.77 (0.39 - 1.51)	0.45 (0.12 - 1.66)	0.73 (0.14 - 3.51)	0.39 (0.01 - 11.9)
Baby gender (male)	0.84 (0.39 - 1.82)	0.86 (0.29 - 2.48)	**0.39 (0.18 - 0.83)** *	**0.19 (0.06 - 0.61)** **	0.52 (0.25 - 1.05)	**0.23 (0.07 - 0.69)** *	0.70 (0.13 - 3.36)	0.29 (0.004 - 6.78)
Birth length (cm)	1.07 (0.95 - 1.21)	1.06 (0.83 - 1.35)	1.04 (0.93 - 1.16)	1.07 (0.85 - 1.37)	0.96 (0.87 - 1.06)	1.12 (0.92 - 1.39)	1.06 (0.86 1.31)	1.26 (0.69 - 3.48)
Birth weight (z-score)	1.11 (0.74 - 1.69)	1.06 (0.42 - 2.66)	0.79 (0.53 - 1.18)	**0.39 (0.15- 0.91)** *	0.89 (0.64 - 1.25)	0.47 (0.20 - 1.04)	1.17 (0.55 - 2.42)	2.08 (0.22 - 31.9)
Breastfeeding (Yes)	1.19 (0.36 - 5.43)	0.59 (0.06 - 7.07)	**0.27 (0.08 - 0.78)** *	0.23 (0.02 - 1.89)	0.63 (0.19 - 1.77)	10.6 (0.92 - 166.31)	_	_
Delivery mode (vaginal)	2.61 (0.95 - 9.24)	4.69 (0.99 - 36.3)	1.04 (0.42 - 2.59)	0.52 (0.09 - 2.56)	0.48 (0.18 - 1.17)	**0.15 (0.03- 0.64)** *	_	_
Ever had LRTIs (Yes)	1.13 (0.52 - 2.41)	1.01 (0.28 - 3.62)	0.73 (0.35 - 1.52)	1.09 (0.23 - 5.12)	0.59 (0.29 - 1.17)	0.64 (0.19 - 2.14)	0.76 (0.16 - 4.05)	1.00 (0.02 - 94.9)
Monthly household income								
<R1000	0.44 (0.14 - 1.36)	0.63 (0.11 - 3.54)	2.97 (0.86 - 12.14)	1.48 (0.23 - 10.4)	1.17 (0.36 - 3.49)	0.55 (0.09 - 2.75)	2.22 (0.27 - 14.98)	4.34 (0.05 - 1102.1)
R1000-5000	0.82 (0.31 - 2.27)	1.66 (0.39 - 8.12)	2.05 (0.59 - 8.28)	2.67 (0.48 - 17.3)	0.56 (0.18 - 1.61)	0.39 (0.08 - 1.76)	3.08 (0.34 - 28.07)	1.84 (0.005 - 6301.7)
Pets exposure (Cat and/or Dog)	0.55 (0.25 - 1.20)	0.48 (0.16 - 1.40)	0.73 (0.34 - 1.55)	0.99 (0.28 - 3.57)	0.92 (0.46 - 1.84)	1.92 (0.62- 6.56)	1.30 (0.27 - 6.93)	3.41 (0.23 - 133.3)
Cat (Yes)	0.68 (0.29 - 1.51)	0.56 (0.11 - 2.67)	0.70 (0.30 - 1.59)	0.98 (0.19 - 4.81)	0.86 (0.42 - 1.78)	1.80 (0.44 - 7.96)	1.56 (0.32 - 11.32)	276.2 (2.58 - 8.56e+05)
Dog (Yes)	0.63 (0.28 - 1.36)	0.93 (0.21 - 3.78)	0.76 (0.35 - 1.64)	0.82 (0.18 - 3.69)	0.74 (0.37 - 1.47)	0.85 (0.23 - 3.24)	0.96 (0.19 - 5.09)	0.073 (0.0003 - 4.99)
Residential area (TC-Newman)	2.79 (1.14 - 7.91)	3.23 (0.66 - 19.2)	**0.33 (0.15 - 0.71)** **	**0.15 (0.03 - 0.78)** *	**0.25 (0.12 - 0.51)** ***	**0.17 (0.04 - 0.69)** *	0.47 (0.06 - 2.29)	1.59 (0.04 - 234.7)
Smoking during pregnancy (Yes)	1.22 (0.53 - 2.71)	0.61 (0.16 - 2.14)	0.55 (0.23 - 1.29)	2.90 (0.56 - 16.6)	**0.31 (0.14 - 0.68)** **	**0.21 (0.06 - 0.69)** *	0.67 (0.12 - 5.05)	0.61 (0.001 - 64.0)
HIV exposure (Yes)	0.79 (0.22 - 2.28)	0.61 (0.07 - 3.71)	1.58 (0.68 - 3.72)	0.53 (0.09 - 2.78)	2.21 (0.91 - 5.93)	6.87 (0.76 - 104.11)	1.38 (0.22 - 26.93)	0.29 (0.002 - 16.9)
Mom education								
Secondary	1.03 (0.15 - 20.63)	0.69 (0.06 - 16.9	0.89 (0.10 - 7.63)	0.51 (0.02 - 20.5)	1.37 (0.26 - 6.46)	0.57 (0.05 - 5.46)	_	_
Tertiary	1.00 (0.07 - 25.45)	0.69 (0.003 - 23.2)	0.83 (0.08 - 9.16)	0.96 (0.02 - 55.5)	1.25 (0.15 - 10.59)	0.09 (0.004 - 2.05)	_	_
Full term gestation (Yes)	-	-	-	7.69e+08 (1.23e-77)	0.70 (0.09 - 3.39)	0.57 (0.02 - 10.69)	3.63 (0.17 - 29.99)	_
Tobacco exposure after birth (Yes)	3.42 (0.94 - 22.00)	1.26 (0.17 - 12.1)	**0.24 (0.08 - 0.63)** **	**0.20 (0.04- 0.88)** *	0.62 (0.23 - 1.54)	1.17 (0.19 - 6.39)	0.78 (0.04 - 5.02)	_
Supplements use (Yes)	0.24 (0.04 - 1.37)	0.33 (0.01 - 10.4)	1.74 (0.16 - 38.01)	2.07 (0.13 - 60.9)	4.59 (0.95 - 32.89)	10.19 (1.14 - 138.14)	_	_
Cotinine exposure (Yes)	1.09 (0.43 - 3.16)	0.78 (0.17 - 3.90)	0.44 (0.16 - 1.13)	0.37 (0.07- 1.82)	0.63 (0.21 - 1.66)	2.00 (0.38 - 10.47)	3.91 (0.71 - 19.69)	209.8 (3.09 - 2.29e+05)

- significance (- $P < 0.05$, **- $P < 0.01$, ***- $P < 0.001$); cOR- crude odds ratio; aOR- adjusted odds ratio; CI- Confidence Interval. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant life.

Table S4.2 Factors associated with gut colonisation by *Lachnospira* at birth, six weeks, six months and twelve months

	Lachnospira							
Covariates	**Birth**		**Six weeks**		**Six months**		**Twelve months**	
	cOR (95% CI)	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**
Wheezing	1.08 (0.44 - 2.63)	0.57 (0.14 -2.19)	3.05 (0.38 - 62.71)	_	1.02 (0.43 - 2.39)	0.80 (0.15 - 3.78)	0.80 (0.31 - 2.02)	0.56 (0.11 - 2.56)
Baby gender (male)	1.39 (0.55 - 3.71)	1.86 (0.58 - 6.50)	_	_	0.72 (0.30 - 1.70)	0.68 (0.22 - 2.08)	1.06 (0.42 - 2.67)	1.25 (0.32 - 5.08)
Birth length (cm)	1.06 (0.92 - 1.21)	1.09 (0.84- 1.42)	0.89 (0.70 - 1.17)	_	0.97 (0.86 - 1.09)	1.03 (0.85 - 1.26)	0.93 (0.81 - 1.06)	**0.68 (0.44 - 0.96)** *
Birth weight (z-score)	0.94 (0.57 - 1.54)	0.67 (0.23 - 1.79)	0.52 (0.19 - 1.42)	_	1.01 (0.68 - 1.55)	1.18 (0.51 - 2.65)	0.84 (0.53 - 1.31)	2.09 (0.62 - 8.70)
Breastfeeding (Yes)	0.73 (0.22 - 3.39)	0.27 (0.01 - 4.28)	0.18 (0.02 - 1.57)	_	0.46 (0.15 - 1.57)	0.40 (0.55 - 2.66)	0.47 (0.14 - 1.69)	**0.01 (2.66e-04 - 0.23)** **
Delivery mode (vaginal)	0.79 (0.29 - 2.33)	0.39 (0.09 - 1.59)	0.75 (0.09 - 15.54)	_	0.78 (0.29 - 2.33)	0.43 (0.10 - 1.89)	0.70 (0.24 - 2.21)	1.77 (0.35 - 10.5)
Ever had LRTIs (Yes)	0.87 (0.33 - 2.13)	0.98 (0.24 - 4.02)	_	_	0.92 (0.37 - 2.17)	1.71 (0.38 - 8.57)	0.64 (0.22 - 1.68)	0.35 (0.07 - 1.55)
Monthly household income				_				
<R1000	1.77 (0.54 - 6.89)	1.73 (0.33 – 10.6)	_	_	0.63 (0.19 - 2.26)	0.56 (0.12 - 2.75)	0.48 (0.13 - 1.86)	0.49 (0.06 - 3.57)
R1000-5000	0.61 (0.16 - 2.56)	0.43 (0.08 - 2.52)	_	_	0.54 (0.16 - 1.96)	0.32 (0.06 - 1.57)	0.54 (0.14 - 2.13)	0.47 (0.07 - 3.16)
Pets exposure (Cat and/or Dog)	0.39 (0.13 - 1.02)	0.32 0.09 - 1.03)	1.60 (0.19 - 13.78)	_	1.23 (0.52 - 2.91)	1.47 (0.46 - 4.76)	0.54 (0.20 - 1.36)	0.63 (0.17 - 2.26)
Cat (Yes)	0.35 (0.09 - 1.01)	0.79 (0.14 - 4.34)	0.73 (0.32 - 23.62)	_	1.33 (0.32 - 23.62)	2.34 (0.47 - 1.29)	0.35 (0.11 - 0.96)	0.39 (0.06 - 2.27)
Dog (Yes)	0.44 (0.15 - 1.14)	0.44 (0.09 - 1.84)	1.95 (0.23 - 16.75)	_	1.00 (0.42 - 2.36)	0.63 (0.14 - 2.52)	0.48 (0.17 - 1.27)	0.89 (0.17 - 4.39)
Residential area (TC-Newman)	1.04 (0.41 - 2.91)	1.44 (0.28 - 7.88)	1.33 (0.16 - 11.44)	_	0.71 (0.29 - 1.66)	0.64 (0.13 - 3.09)	0.95 (0.38 - 2.43)	3.10 (0.48 - 25.2)
Smoking during pregnancy (Yes)	0.89 (0.32 - 2.28)	1.31 (0.34 - 5.06)	0.89 (0.04 - 7.32)	_	1.06 (0.38 - 2.73)	1.63 (0.43 - 6.42)	1.29 (0.43 - 3.59)	2.89 (0.52 - 19.0)
HIV exposure (Yes)	0.95 (0.21 - 3.11)	0.45 (0.02 - 4.22)	_	_	2.00 (0.74 - 5.09)	1.86 (0.27 – 11.5)	1.23 (0.36 - 3.76)	0.17 (0.005 - 1.96)
Mom education				_				
Secondary	0.59 (0.08 - 12.04)	0.14 (0.009 - 4.45)	_	_	_	_	0.69 (0.06 - 15.22)	_
Tertiary	0.44 (0.01 - 13.18)	0.12 (0.003 - 5.69)	_	_	_	_	0.25 (0.01 - 8.19)	_
Full term gestation (Yes)	_	_	_	_	0.54 (0.11 - 3.93)	0.35 (0.03 - 5.33)	1.31 (0.18 - 26.41)	3.51 (0.009 - 8791.2)
Tobacco exposure after birth (Yes)	0.79 (0.27 - 2.95)	0.91 (0.15 - 6.41)	0.72 (0.09 - 14.88)	_	0.62 (0.23 - 1.88)	0.67 (0.14 - 3.45)	0.38 (0.13 - 1.18)	0.12 (0.01 - 0.89)
Supplements use (Yes)	0.78 (0.12 - 15.32)	_	_	_	1.25 (0.20 - 24.22)	0.97 (0.10 - 22.6)	_	_
Cotinine exposure (Yes)	0.75 (0.27 - 2.44)	0.42 (0.09 - 1.96)	0.70 (0.08 - 14.53)	_	0.97 (0.32 - 3.63)	2.22 (0.49- 13.0)	1.14 (0.35 - 4.41)	2.53 (0.39 - 21.9)

- significance (- $P < 0.05$, **- $P < 0.01$); cOR- crude odds ratio; aOR- adjusted odds ratio; CI- Confidence Interval. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant life.

Table S4.3 Factors associated with gut colonisation by *Veillonella* at birth, six weeks, six months and twelve months

	Veillonella							
Covariates	**Birth**		**Six weeks**		**Six months**		**Twelve months**	
	cOR (95% CI)	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**
Wheezing	1.94 (0.93 - 4.15)	1.89 (0.62 - 6.01)	5.37 (0.83 - 104.65)	0.08 (7.12e-07 - 1.91)	0.73 (0.14 - 3.42)	_	_	_
Baby gender (male)	1.10 (0.53 - 2.33)	0.84 (0.32 - 2.19)	0.68 (0.09 - 3.66)	1.47 (0.04 - 100.1)	2.01 (0.32 - 15.61)	_	_	_
Birth length (cm)	1.11 (0.99 - 1.25)	1.12 (0.91 - 1.38)	0.95 (0.73 - 1.19)	0.74 (0.25 - 1.69)	0.89 (0.68 - 1.16)	_	_	_
Birth weight (z-score)	1.23 (0.83 - 1.85)	1.03 (0.48 - 2.23)	0.52 (0.18 - 1.33)	0.06 (1.66e-04 - 2.45)	0.59 (0.21 - 1.46)	_	_	_
Breastfeeding (Yes)	0.51 (0.18 - 1.58)	0.19 (0.02 - 1.42)	_	_	1.18 (0.06 - 7.49)	_	_	_
Delivery mode (vaginal)	1.71 (0.69 - 4.85)	2.18 (0.67 - 8.33)	_	_	_	_	_	_
Ever had LRTIs (Yes)	1.68 (0.81 - 3.50)	0.89 (0.27 - 2.86)	1.61 (0.30 - 11.98)	0.02 (1.63e-10 - 6.88e+09)	0.49 (0.09 - 2.28)	_	_	_
Monthly household income							_	_
<R1000	1.35 (0.49 - 3.99)	1.09 (0.27 - 4.60)	2.39 (0.29 - 16.08)	13.4 (0.13 - 2.56e+04)	1.61 (0.07 - 17.79)	_	_	_
R1000-5000	0.83 (0.30 - 2.42)	0.57 (0.15 - 2.22)	8.33 (0.74 - 188.35)	_	1.67 (0.07 - 18.39)	_	_	_
Pets exposure (Cat and/or Dog)	0.91 (0.43 - 1.89)	0.52 (0.19 - 1.36)	0.63 (0.11 - 3.52)	0.47 (7.93e-03 - 6.97)	0.55 (0.07 - 3.43)	_	_	_
Cat (Yes)	0.65 (0.28 - 1.40)	0.60 (0.14 - 2.56)	0.76 (0.14 - 5.68)	37.1 0.006 - 2.87e+08)	0.32 (0.14 - 5.68)	_	_	_
Dog (Yes)	0.78 (0.36 - 1.62)	0.59 (0.15 - 2.15)	0.51 (0.09 - 2.89)	0.05 (6.395270e-11 352.6)	0.48 (0.06 - 2.95)	_	_	_
Residential area (TC-Newman)	1.11 (0.51 - 2.48)	3.38 (0.84 - 15.0)	0.76 (0.14 - 4.26)	0.56 (6.49e-05 - 161.8)	_	_	_	_
Smoking during pregnancy (Yes)	1.07 (0.48 - 2.31)	0.80 (0.25- 2.51)	0.74 (0.14 - 5.52)	2.69 (7.89e-03 - 1.06e+05)	0.77 (0.14 - 5.76)	_	_	_
HIV exposure (Yes)	1.27 (0.43 - 3.34)	0.51 (0.06 - 2.88)	0.30 (0.05 - 1.72)	0.07 (2.59e-05 - 9.34)	1.57 (0.25 - 30.22)	_	_	_
Mom education							_	_
Secondary	_	_	_	_	5.13 (2.42e-01 - 42.27)	_	_	_
Tertiary	_	_	_	_	_	_	_	_
Full term gestation (Yes)	1.51 (0.23 - 29.48)	0.27 (0.02 - 7.93)	_	_	_	_	_	_
Tobacco exposure after birth (Yes)	1.62 (0.57 - 5.85)	1.11 (0.22 - 6.48)	2.14 (0.2 - 11.76)	0.04 (4.48e-10 - 88.3)	_	_	_	_
Suppliment use (Yes)	_	_	10.9 (0.46 - 135.12)	_	_	_	_	_
Cotinine exposure (Yes)	0.64 (0.27 - 1.60)	0.44 (0.12 - 1.61)	_	_	0.89 (0.05 - 5.58)	_	_	_

cOR- crude odds ratio; aOR- adjusted odds ratio; CI- Confidence Interval. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant life.

Table S4.4 Factors associated with gut colonisation by *Rothia* at birth, six weeks, six months and twelve months

	Rothia							
Covariates	**Birth**		**Six weeks**		**Six months**		**Twelve months**	
	cOR (95% CI)	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**	**cOR (95% CI)**	**aOR (95% CI)**
Wheezing	0.63 (0.29 - 1.32)	0.73 (0.22 - 2.39)	0.19 (0.01 - 1.20)	_	0.65 (0.08 - 4.02)	_	_	_
Baby gender (male)	0.75 (0.34 - 1.62)	0.73 (0.24 - 2.15)	_	_	1.32 (0.15 - 11.23)	_	0.46 (0.06 - 2.47)	_
Birth length (cm)	1.07 (0.95 - 1.21)	1.03 (0.81 - 1.31)	1.04 (0.81 - 1.29)	_	0.86 (0.62 - 1.14)	_	0.98 (0.77 - 1.23)	_
Birth weight (z-score)	1.35 (0.89 - 2.07)	1.49 (0.59 - 3.82)	0.78 (0.29 - 1.87)	_	0.40 (0.11 - 1.20)	_	0.92 (0.39 - 2.01)	_
Breastfeeding (Yes)	0.20 (0.01 - 1.08)	0.53 (0.02 - 7.76)	_	_	1.79 (0.09 - 13.04)	_	1.18 (0.06 - 8.17)	_
Delivery mode (vaginal)	0.95 (0.37 - 2.21)	0.89 (0.19 - 3.54)	2.05 (0.27 - 11.19)	_	_	_	2.06 (0.27 - 11.42)	_
Ever had LRTIs (Yes)	0.86 (0.41 - 1.83)	1.16 (0.35 - 3.99)	0.38 (0.05 - 2.00)	_	0.43 (0.06 - 2.69)	_	_	_
Monthly household income								
<R1000	1.76 (0.62 - 5.00)	1.72 (0.37 - 8.12)	_	_	_	_	_	_
R1000-5000	2.27 (0.82 - 6.24)	2.97 (0.69 - 13.44)	_	_	_	_	_	_
Pets exposure (Cat and/or Dog)	0.77 (0.35 - 1.67)	0.62 (0.19 - 1.93)	0.63 (0.11 - 3.52)	_	0.27 (0.0 - 2.18)	_	2.0 (0.37 - 14.94)	_
Cat (Yes)	0.98 (0.45 - 2.25)	2.13 (0.44 - 10.84)	0.76 (0.14 - 5.68)	_	0.16 (0.14 - 5.68)	_	1.23 (0.23 - 9.19)	_
Dog (Yes)	0.68 (0.31 - 1.47)	0.32 (0.08 - 1.36)	0.51 (0.09 - 2.89)	_	0.23 (0.01 - 1.88)	_	1.47 (0.27 - 10.99)	_
Residential area (TC-Newman)	0.85 (0.36 - 1.91)	1.59 (0.37 - 6.99)	1.57 (0.29 - 11.69)	_	0.31 (0.02 - 2.52)	_	1.24 (0.22 - 7.03)	_
Smoking during pregnancy (Yes)	0.79 (0.36 - 1.77)	1.89 (0.52 - 7.58)	1.93 (0.29 - 37.88)	_	0.12 (0.01 - 0.98)	_	1.29 (0.19 - 28.11)	_
HIV exposure (Yes)	2.01 (0.64 - 8.81)	1.02 (0.17 - 8.28)	1.67 (0.25 - 32.65)	_	0.37 (0.06 - 2.89)	_	1.13 (0.17 - 22.46)	_
Mom education								
Secondary	**6.32 (1.00 - 49.73) ***	6.81 (0.62 - 167.36)	6.33 (0.29 - 61.32)	_	6.89 (0.32 - 64.17)	_	_	_
Tertiary	13.49 (1.10 - 378.19)	15.72 (0.74 - 814.42)	_	_	_	_	_	_
Full term gestation (Yes)	0.8 (0.04 - 5.19)	0.82 (0.02 - 16.87)	_	_	_	_	4.4 (0.20 - 37.92)	_
Tobacco exposure after birth (Yes)	0.27 (0.04 - 0.97)	0.63 (0.07 - 4.12)	4.55 (0.79 - 26.20)	_	_	_	2.57 (0.33 - 14.44)	_
Suppliment use (Yes)	0.71 (0.04 - 4.56)	2.51 (0.07 - 44.28)	_	_	_	_	_	_
Cotinine exposure (Yes)	0.33 (0.08 - 1.01)	0.13 (0.006 - 0.88)	2.2 (0.29 - 12.11)	_	_	_	2.43 (0.32 - 13.70)	_

- significance (- $P < 0.05$); cOR- crude odds ratio; aOR- adjusted odds ratio; CI- Confidence Interval. Birth- first four days, 6 weeks- first four to eight weeks, 6 months- first five to seven months, and 12 months- first 11 to 13 months of an infant life.

Table S4.5 Area under the curve and corresponding confidence interval of FLVR bacterial load at six weeks, six months and 12 months in predicting any wheeze outcome

		Any wheeze					
		Six weeks		Six months		Twelve months	
		AUC	95%CI	AUC	95%CI	AUC	95%CI
Each bacterium	F	0.552	0.39 - 0.71	0.555	0.44 - 0.68	0.542	0.42 - 0.66
	L	0.667	-	0.716	0.50 - 0.93	0.636	0.39 - 0.88
	V	0.525	0.42 - 0.63	0.542	0.45 - 0.64	0.535	0.42 - 0.65
	R	0.496	0.39 - 0.61	0.574	0.48 - 0.67	0.629	0.51 - 0.74
2by2 interactions	FL	0.565	0.41 - 0.72	0.606	0.49 - 0.72	0.544	0.42 - 0.66
	FV	0.547	0.44 - 0.65	0.604	0.51 - 0.69	0.606	0.49 - 0.72
	FR	0.539	0.43 - 0.65	0.529	0.43 - 0.62	0.687	0.58 - 0.79
	LV	0.568	0.46 - 0.68	0.549	0.45 - 0.64	0.551	0.44 - 0.67
	LR	0.476	0.37 - 0.59	0.573	0.48 - 0.67	0.647	0.53 - 0.76
	VR	0.6	0.49 - 0.70	0.602	0.51 - 0.69	0.679	0.57 - 0.79
3by3 interactions	FLV	0.611	0.51 - 0.72	0.647	0.56 - 0.74	0.61	0.49 - 0.72
	FLR	0.581	0.47 - 0.69	0.538	0.44 - 0.63	0.7	0.59 - 0.80
	FVR	0.631	0.53 - 0.73	0.632	0.54 - 0.72	0.739	0.64 - 0.84
	LVR	0.637	0.54 - 0.74	0.641	0.55 - 0.73	0.699	0.59 - 0.80
4by4 interactions	FLVR	0.655	0.56 - 0.75	0.679	0.59 - 0.77	0.753	0.66 - 0.85

F- *Faecalibacterium,* L- *Lachnospira*; V- *Veillonella*; R- *Rothia*; AUC- area under the curve; CI- confidence interval; LRTIs- lower respiratory tract infections. The AUC and CI are calculated from qPCR bacterial load (DNA copies) converted into log base 10 scale.

Supplementary figures

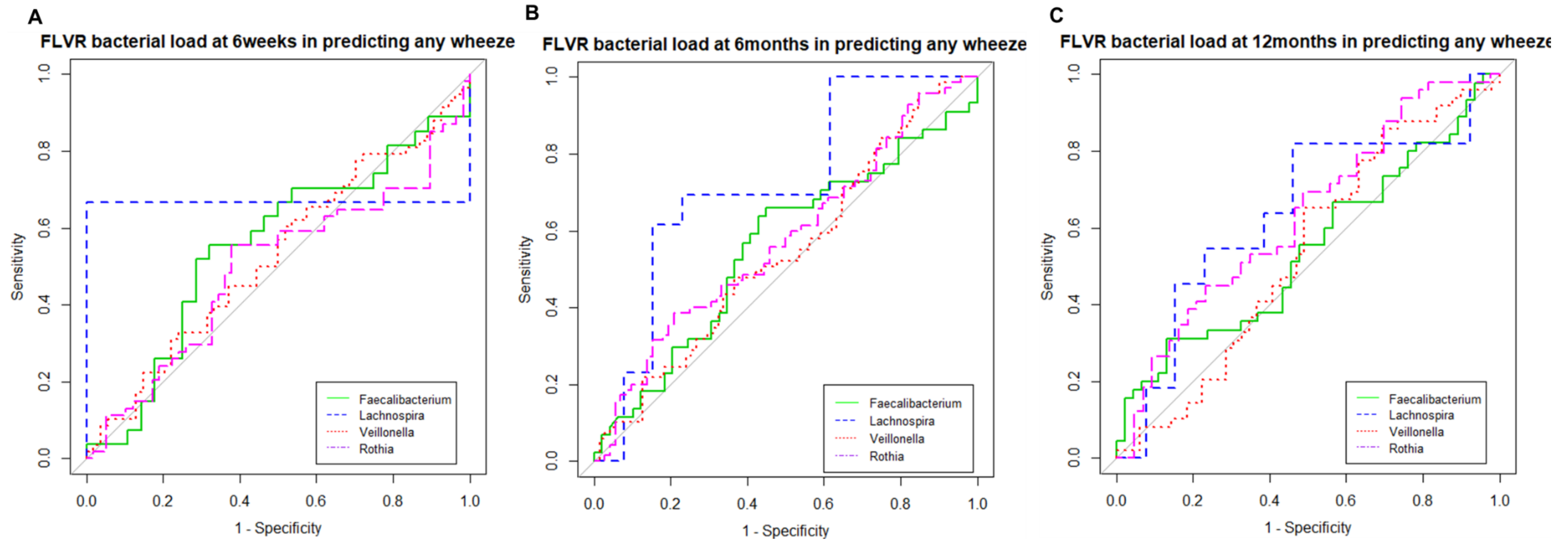


Figure S4.1 Receiver operating characteristic (ROC) curves predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants), using FLVR bacterial load (DNA copies) determined by quantitative polymerase chain reaction (qPCR). **A.** Prediction at six weeks (first four to eight weeks). **B.** Prediction at six months (first five to seven months). **C.** Prediction at 12 months (first 11 to 13 months of an infant's life). Plots are based on logarithmic scale (log base 10) of DNA copies.

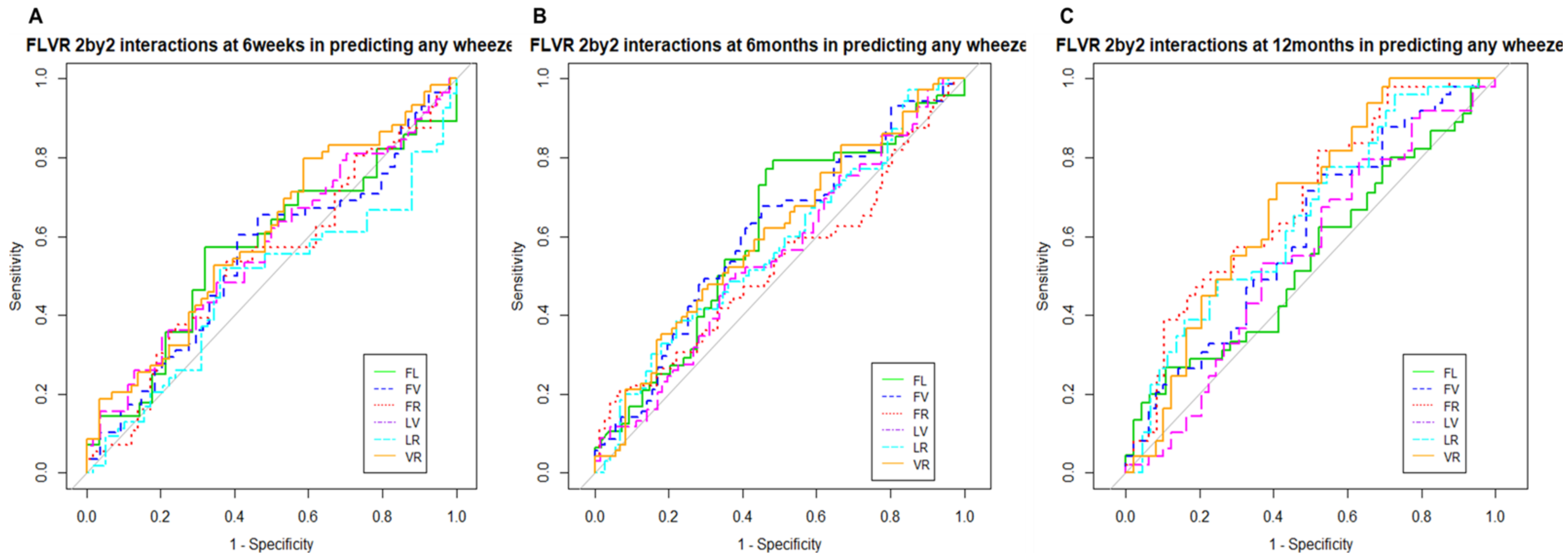


Figure S4.2 Receiver operating characteristic (ROC) curves predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants), using two by two interactions of FLVR bacterial load (DNA copies) determined by quantitative polymerase chain reaction (qPCR). **A.** Prediction at six weeks (first four to eight weeks). **B.** Prediction at six months (first five to seven months). **C.** Prediction at 12 months (first 11 to 13 months of an infant's life). Plots are based on logarithmic scale (log base 10) of DNA copies.

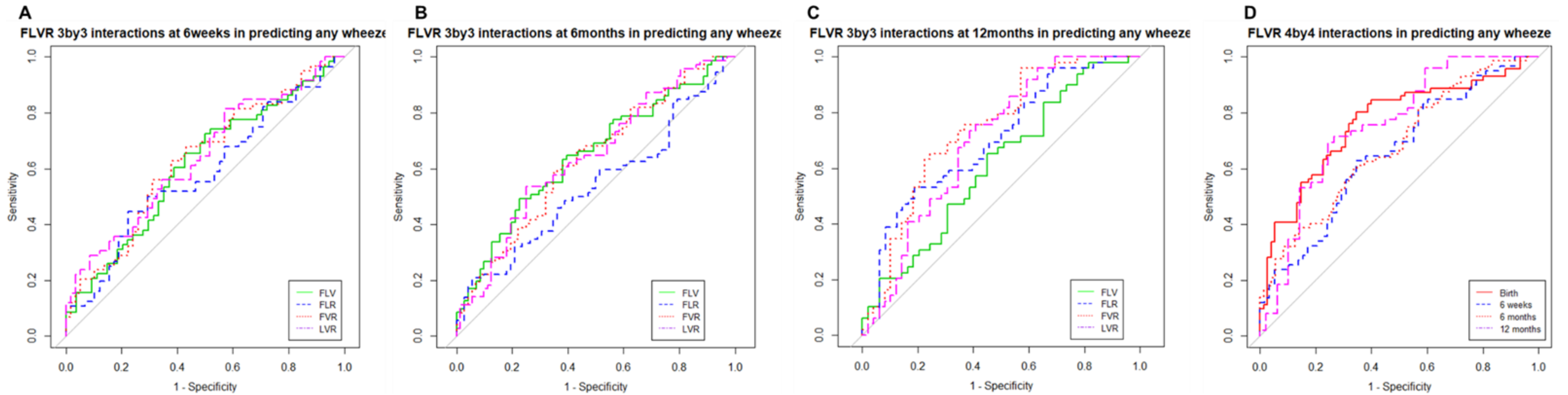


Figure S4.3 Receiver operating characteristic (ROC) curves predicting any wheeze (cohort of 140 wheezing and 140 non-wheezing infants), using FLVR bacterial load (DNA copies) determined by quantitative polymerase chain reaction (qPCR). **A.** Prediction based on three by three interactions at six weeks (first four to eight weeks). **B.** Prediction based on three by three interactions at six months (first five to seven months). **C.** Prediction based on three by three interactions at 12 months (first 11 to 13 months of an infant's life). **D.** Prediction based on four by four interactions. Plots are based on logarithmic scale (log base 10) of DNA copies.

Part 4: General discussion and conclusion

Chapter 5:

General discussion

The overall aim of this research project was to determine the contribution of FLVR bacteria in childhood wheezing within the DCHS, South Africa. Beginning with a systematic literature review (chapter 2), we observed that FLVR bacteria may be involved in the occurrence of different humans' respiratory diseases. However, we noticed a lack of studies on FLVR bacteria and respiratory diseases conducted in an African setting. This strongly encourages future studies with similar objectives to be conducted in Africa. Having seen this gap, we screened stool samples in our study using optimized PCR assays for the detection and quantification of FLVR bacteria. Using logistic regression analysis, we assessed whether the detection and quantification of FLVR bacteria was associated with wheezing and recurrent wheezing in our cohort. This study gives an overview of the prevalence of FLVR bacteria in stools samples of wheezing and non-wheezing infants in the DCHS. In addition, this study provides insights into the factors associated with faecal colonization by FLVR bacteria within the first year of life. The main finding from this study was the association of *Rothia* with low risk of wheezing, suggesting a potential protective role of this bacterium against infant wheezing. This has been previously reported (1). We further showed that quantification of FLVR bacteria within the first four days of life, especially *Rothia* and *Lachnospira* may serve as biomarkers for predicting the likelihood of infant wheezing. The four time-points investigated in this study were selected because wheezing is much more common in younger children (2). This can be explained by the very small airways during infancy, and therefore require less obstruction to produce a whistling sound (2). We also wanted to include the first 100 days of life because infants at risk of asthma (based on the asthma predictive index) have been reported to exhibit stool bacterial alterations during this life period (1). In support of this, it has been reported that there is widespread presence of recurrent wheezing in the first twelve months of an infants' life (3). The ROC analysis was mainly focused on the birth time-point (first four days of life) as we wanted to see whether the occurrence of FLVR bacteria early in life could predict later wheezing outcomes.

Because of the contamination risks involved, the conventional PCR experiments in this study were performed in two separate labs, hence the use of different stains and buffers. Briefly our lab is mainly used for 16S library preparation, therefore conventional PCR targeting the 16S rRNA gene is avoided to prevent contaminations. Nevertheless, both ethidium bromide and SYBR safe which were used for our gel electrophoresis are proven to have good DNA staining capacity (4). Moreover, all the analysis in this study were based on FLVR bacterial DNA copy number of the 16S rRNA gene. Studies have shown that quantification based on 16S rRNA

gene may give false results as this gene exists in multiple copies in some bacteria [5-7]. Hence, the load of FLVR bacteria in this study may not necessarily mean that the bacteria were in high abundance. This is because we could have detected multiple 16S rRNA copies from one bacterium. Furthermore, the use of antibiotic is known to impact wheezing [8-10] as well as microbial composition [11-13]. However, our regression analysis did not account for antibiotic use as this data was not captured in the DCHS database for the studied infants. We did not have control over data collection because our study is nested within the DCHS. If any of our infants had used antibiotics before or during sample collection, then this might have affected our findings. In addition, the influence of maternal asthma status has been shown to influence the gut microbiota and its association with wheezing risk in infants [14-16]. Similarly, this factor was not included in the regression analysis because only 1%; 3/280 infants (1 wheeze and 2 non-wheeze) had asthmatic mothers. We strongly encourage further studies to quantify FLVR bacteria based on a single copy gene such as the *rpoB* gene, and also to account for antibiotic use as well as parental asthma in the regression analysis. Moreover, other factors such as lung function [17], maternal stress [18] and alcohol exposure [19] have been linked to microbial alteration and wheezing. We strongly encourage future studies to control for these factors, and also to determine how they influence the proliferation of FLVR bacteria in both wheezing and non-wheezing infants. Knowing this will help us to take the necessary precautions which will help maintain the protective bacteria early in life, and thus reducing the risk of wheezing later in life. The systematic review (chapter 2) demonstrated the association of *Faecalibacterium* with pulmonary TB. This could mean that infants born to TB infected mothers will have high amount of *Faecalibacterium*, which may be the reason for the high risk of wheezing in infants with this bacterium at six months observed in our study. Due to this possible connection, we therefore encourage future studies to take into account the influence of maternal TB on FLVR bacteria and wheezing. In addition, it will be interesting to investigate the association of maternal FLVR bacteria during pregnancy with infant wheezing after birth. This will provide guidance to mothers regarding the food intake which will help reduce wheezing risk in children through the acquisition of the protective microbes. It has been shown that asthma development in offspring is associated with maternal diet and its influence on microbiome during pregnancy [20, 21]. The major limitation of this study is that we do not know whether the detected bacteria were alive or dead. We therefore encourage future studies to apply measures that can distinguish live bacterial DNA from dead ones such as the use of propidium monoazide [22]. Advances in culturomics have also shown the potential in distinguishing live from dead bacteria [23]. Other methods which can be used include transcriptomics [24] and flow cytometry [25].

References cited

1. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science translational medicine. 2015;7(307):307ra152-307ra152.
2. Muglia C, Oppenheimer J. Wheezing in Infancy: An Overview of Recent Literature. Current Allergy and Asthma Reports. 2017;17(10):67.
3. Mallol J, García-Marcos L, Solé D, Brand P, Group ES. International prevalence of recurrent wheezing during the first year of life: variability, treatment patterns and use of health resources. Thorax. 2010:thx. 2009.115188.
4. Guzaev M, Li X, Park C, Leung W-Y, Roberts L. Comparison of Nucleic Acid Gel Stains Cell permeability, safety, and sensitivity of ethidium bromide alternatives. 2017.
5. Vos M, Quince C, Pijl AS, de Hollander M, Kowalchuk GA. A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS One. 2012;7(2):e30600.
6. Crosby LD, Criddle CS. Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. Biotechniques. 2003;34(4):790-802.
7. Dahllöf I, Baillie H, Kjelleberg S. rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Applied and environmental microbiology. 2000;66(8):3376-80.
8. Alm B, Erdes L, Möllborg P, Pettersson R, Norvenius SG, Åberg N, et al. Neonatal antibiotic treatment is a risk factor for early wheezing. Pediatrics. 2008;121(4):697-702.
9. Penders J, Kummeling I, Thijs C. Infant antibiotic use and wheeze and asthma risk-a systematic review and meta-analysis. European Respiratory Journal. 2011:erj01050-2010.
10. Wickens K, Pearce N, Crane J, Beasley R. Antibiotic use in early childhood and the development of asthma. Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology. 1999;29(6):766-71.
11. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006;118(2):511-21.
12. Fjalstad JW, Esaiassen E, Juvet LK, van den Anker JN, Klingenberg C. Antibiotic therapy in neonates and impact on gut microbiota and antibiotic resistance development: a systematic review. J Antimicrob Chemother. 2017;73(3):569-80.
13. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term impacts of antibiotic exposure on the human intestinal microbiota. Microbiology. 2010;156(11):3216-23.
14. Strina A, Barreto ML, Cooper PJ, Rodrigues LC. Risk factors for non-atopic asthma/wheeze in children and adolescents: a systematic review. Emerging themes in epidemiology. 2014;11(1):5.
15. Azad MB, Vehling L, Lu Z, Dai D, Subbarao P, Becker AB, et al. Breastfeeding, maternal asthma and wheezing in the first year of life: a longitudinal birth cohort study. European Respiratory Journal. 2017;49(5):1602019.
16. Moraes LS, Takano OA, Mallol J, Solé D. Risk factors associated with wheezing in infants. Jornal de pediatria. 2013;89(6):559-66.
17. Gray DM, Czovek D, McMillan L, Turkovic L, Stadler JA, Vanker A, et al. Intra-breath measures of respiratory mechanics in healthy African infants detect risk of respiratory illness in early life. European Respiratory Journal. 2019;53(2):1800998.

18. MacGinty RP, Lesosky M, Barnett W, Stein DJ, Zar HJ. Associations between maternal mental health and early child wheezing in a South African birth cohort. Pediatric pulmonology. 2018;53(6):741-54.
19. Myers B, Koen N, Donald KA, Nhapi RT, Workman L, Barnett W, et al. Effect of hazardous alcohol use during pregnancy on growth outcomes at birth: findings from a South African Cohort Study. Alcohol Clin Exp Res. 2018;42(2):369-77.
20. Gray LE, O'Hely M, Ranganathan S, Sly PD, Vuillermin P. The maternal diet, gut bacteria, and bacterial metabolites during pregnancy influence offspring asthma. Front Immunol. 2017;8:365.
21. Thorburn AN, McKenzie CI, Shen S, Stanley D, Macia L, Mason LJ, et al. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. Nat Commun. 2015;6:7320.
22. Kim S, Ko G. Using propidium monoazide to distinguish between viable and nonviable bacteria, MS2 and murine norovirus. Letters in applied microbiology. 2012;55(3):182-8.
23. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clinical Microbiology and Infection. 2012;18(12):1185-93.
24. Simón-Soro A, Guillen-Navarro M, Mira A. Metatranscriptomics reveals overall active bacterial composition in caries lesions. J Oral Microbiol. 2014;6(1):25443.
25. Berney M, Hammes F, Bosshard F, Weilenmann H-U, Egli T. Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. Applied and environmental microbiology. 2007;73(10):3283-90.

Chapter 6:

General conclusion

This study adds insight into the contribution of FLVR bacteria in infant wheezing. Findings from the system review (part1, chapter 2) gives future direction to investigate the contribution of FLVR bacteria in human respiratory diseases, especially in Africa. We encourage future studies to further elucidate the contribution of these bacteria in childhood wheezing both in the absence and presence of atopic symptoms. Seeing that *Rothia* may potentially be protective against childhood wheeze, we suggest that future studies should aim to identify the exact species or strains of this genus which may be playing a protective role. In addition, it will be interesting to study the metabolic components of this bacterium which may be contributing to its protective role. This will help in taking feeding measures that will enhance the production of beneficial metabolites, should this bacterium be proven protective. Finally, though our study shows the potential protective role of *Rothia* against childhood wheeze, we do not encourage using this bacterium as a probiotic as this was just an observational study. We encourage mouse model experiments to study the probiotic capacity of this organism.

Part 5: Appendices

Appendix A: Curriculum vitae of Saara Kanyemba

Education

- MSc (Med). Medical Microbiology, University of Cape Town, Cape Town, South Africa 2017-2018
  Dissertation title: Association of *Faecalibacterium*, *Lachnospira*, *Veillonella*, and *Rothia* with childhood wheezing

- BSc. Hon. Microbiology, University of Namibia, Windhoek, Namibia 2013-2016
  Mini-thesis title: Anti-biofilm and antibacterial activity of *Diospyros lycioides* against *Staphylococcus aureus* and *Mycobacterium avium*

- NSSC- Namibia Senior Secondary Certificate (NSSCO/H), Mweshipandeka High School, Ongwediva, Namibia. **Subjects-** Mathematics, English as a 2nd Language, Physical Science, Biology (High level), Geography, Oshindonga (High level) as a 1st Language 2010-2011

Achievements

- Top achiever in the modules Biometrics II, and Biomolecules and Catalysis. University of Namibia 2014
- Top achiever in the Modules Physiology, and Biotechnology. University of Namibia 2015
- Top achiever in the modules: Immunology, Medical Bacteriology, Virology, Parasitology, Environmental and Industrial Microbiology, and Bioinformatics. University of Namibia 2016
- Best fourth year student, Department of Biological Sciences, University of Namibia 2016
- Best oral presenter at the EARTHS summer school, Hamburg, Germany 2018

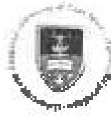
Technical skills

- Microbial plate culture
- Isolation of phytochemical extracts and disc diffusion assays for anti-microbial testing
- Microbial detection in biological samples using real time polymerase chain reaction (qPCR)
- Conventional Polymerase Chain Reaction and gel electrophoresis
- Purification of DNA amplicons from gel using commercialized kits
- DNA extraction from biological samples using commercialized kits
- Designing of primers using primer designing softwares
- Analysis of sequencing data (Sanger sequencing) using molecular biology-based softwares
- Biostatistical data analysis using STATA and R statistical softwares

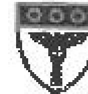
Presentations and workshops attended

- Attended H3ABioNet introduction to bioinformatics course, University of Cape Town, South Africa (May - August 2017).
- Attended EARTHS (Education and Research-oriented Training in Health Sciences) summer school, Hamburg University of Applied Sciences, Germany (01June - 30 June 2018).
- Presented during departmental meetings, as well as during the summer school.
- Demonstrated practical sessions of 2nd and 3rd year students, University of Namibia (2015-2016
- Did an internship with the Central Veterinary Laboratory (CVL), Windhoek (December 2015 & 2017)

Appendix B: Ethical approval of the study



UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee



Room E53-46 Old Main Buildin
Groote Schuur Hospit
Observatory 792
Telephone [021] 406 649
Email: sumayah.ariefdien@uct.ac.z
Website: www.health.uct.ac.za/fhs/research/humanethics/form

08 October 2018

HREC REF: 650/2018

Dr M Kaba
Division of Medical Microbiology
Falmouth Building
Level 5 Room 5.14
FHS

Dear Dr Kaba

PROJECT TITLE: ASSOCIATION OF FAECALIBACTERIUM, LACHNOSPIRA, VEILLONELLA AND ROTHIA WITH CHILDHOOD WHEEZING (Sub-study linked to 401/2009 & 748/2015) - MSc Candidate – Ms S Kanyemba

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

Approval is granted for one year until the 30 October 2019.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

We acknowledge that the student: Ms Saara Kanyemba will also be involved in this study.

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637

Appendix C: Amendment of the study into the DCHS

Form FHS007: Amendment – study staff

HREC office use only (FWA00001637; IRB00001938)		
☐ Approved		
This serves as notification that all changes to the study staff and documentation described below are approved.		
Chairperson of the HREC signature	Date	26/9/2018

Principal Investigator to complete the following:

1. Protocol information

Date (when submitting this form)	14 SEP 2018
HREC REF Number	748 / 2015
Protocol title	The Stool Microbiota and Infant Wheezing Illness – The Drakenstein Child Health Study, South Africa (Sub-study linked to 401/2009)
Protocol number (if applicable)	V1.0
Principal Investigator	Dr Mamadou Kaba
Department / Office Internal Mail Address	Pathology / Division of Medical Microbiology, Faculty of Health Sciences, Falmouth Building, Entrance 2, Level 5, Room 5.14, Anzio Road, Observatory mamadou.kaba@hotmail.com

1.1 Does this protocol receive US Federal funding?	☐ Yes	✓No

2.1 Staff changes (tick ✓)

Are new personnel being added to this research?	✓ Yes	☐ No
Are current personnel being removed from this research?	☐ Yes	✓ No
Is the principal investigator for this research being changed? If yes, please attach revised conflict of interest and PI declaration statements. (Refer: sections 7 and 8.3 in the New Protocol Application Form - FHS013)	☐ Yes	✓ No
Do the consent and assent forms need modification to reflect these staff changes? If yes, please attach copies of the revised forms, with all changes highlighted or tracked and listed in the documents for approval.	☐ Yes	✓ No

12 September 2016 | Page 1 of 2 | FHS007