Using ’omics technologies to understand pathogenesis and seek alternative therapies for otitis media in children

Rachael Joanna Lappan
BSc (Hons)

This thesis is presented in fulfilment of the requirements for the degree of Doctor of Philosophy (Microbiology)

Supervised by A/Prof Christopher Peacock, Dr Sarra Jamieson, A/Prof Timo Lassmann and Prof Manfred Beilharz

The Marshall Centre for Infectious Diseases Research and Training
The School of Biomedical Sciences
Faculty of Health and Medical Sciences
The University of Western Australia

2019
Thesis Declaration

I, Rachael Joanna Lappan, certify that:

- This thesis has been substantially accomplished during enrolment in the degree.

- This thesis does not contain material which has been submitted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

- No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

- This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text and, where relevant, in the Authorship Declaration that follows.

- The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

- The research involving human data reported in this thesis was assessed and approved by The University of Western Australia Human Research Ethics Committee (approval #: RA/4/1/6839 and RA/4/1/7870) and all relevant hospital governance committees. Written patient consent has been received and archived for the research involving patient data reported in this thesis.

- The following additional ethics approvals were obtained prior to commencing the relevant work described in this thesis: Princess Margaret Hospital for Children (2013119EP/RG5000002432, 2046EP/RG5000002534) and St John of God Health Care (#708).

- The work described in this thesis was funded by a Telethon New Children’s Hospital research grant and a Wesfarmers Centre of Vaccines and Infectious Diseases seed funding grant.
This thesis contains published work and/or work prepared for publication, some of which has been co-authored.

Signature: ________________________________  Date: 18/12/2018
Abstract

Recurrent acute otitis media (rAOM, recurrent ear infection) is a common childhood disease caused by bacteria known as otopathogens. Current treatments for rAOM involve antibiotic prescription for individual episodes of infection, and after several recurrent infections, surgical ventilation of the middle ear. However, antibiotic treatment is not effective at preventing the recurrence of infection and some children continue to experience ear problems after surgical intervention. Alternative therapies to break the cycle of reinfection or to prevent the onset of rAOM are required.

This thesis describes the use of multiple next-generation sequencing ’omics approaches to characterise the microbiome of the upper respiratory tract in children. Two avenues towards the development of alternative therapies for rAOM are explored; 1) the investigation of potential novel otopathogens that may be difficult to detect by culture and thus may be new targets for therapy; and 2) the identification of protective microbes or microbial factors that are suitable for translation into a probiotic therapy.

This thesis provides novel insight into the bacterial and viral communities of the upper respiratory tract in rAOM-resistant children, who have been exposed to major environmental risk factors for rAOM and have not developed the disease, and in children undergoing surgery for rAOM.

First, this thesis describes the use of 16S rRNA gene sequencing to determine the taxonomic composition of the bacterial communities in the nasopharynx of rAOM-resistant children, and in the middle ear, nasopharynx and external auditory canal of children with rAOM. *Alloiococcus* and *Turicella* were dominant organisms in the middle ear, and were thus possible novel otopathogens; however they were also abundant in the ear canal. The commensal genera *Corynebacterium* and *Dolosigranulum* were enriched in the nasopharynx of rAOM-resistant children compared to those with rAOM, and may have a protective role.

To characterise the roles of these genera in rAOM, a subset of samples underwent shotgun metagenomics sequencing to identify the organisms at species level, detect DNA respiratory viruses and to explore the genomic functions of the upper respiratory
tract microbiota. It was determined that an overwhelming proportion of human DNA is present in middle ear fluid and nasopharyngeal specimens, and it is challenging to remove. This affects the characterisation of functional differences in the microbial communities of children with and without rAOM, though taxonomic analysis is still possible. *Alloiococcus otitidis* and *Turicella otitidis* were confirmed to be abundant in the middle ear of children with rAOM, and *Corynebacterium pseudodiphtheriticum* and *Dolosigranulum pigrum* were identified as the dominant commensal organisms enriched in the rAOM-resistant nasopharynx.

To determine if the presence of commensal species *C. pseudodiphtheriticum* and *D. pigrum* in the microbiota is protective against the development of rAOM, these organisms were assessed for bacterial interference on agar. *D. pigrum* did not display antagonistic activity against the known otopathogen species, but *C. pseudodiphtheriticum* reduced the growth of multiple *Moraxella catarrhalis* strains. This inhibitory activity appears to require live *C. pseudodiphtheriticum*, as extracellular and intracellular supernatants derived from *C. pseudodiphtheriticum* did not replicate the effect.

Finally, a pilot study was carried out to explore the role of the potential novel otopathogens *A. otitidis* and *T. otitidis* in the middle ear via metatranscriptomics. Dual RNA-seq, a new technique, was applied for the first time to human middle ear fluid specimens to concurrently profile the microbial metatranscriptome and the host transcriptome. Methodological issues were characterised and recommendations are given for the optimisation of this technique in middle ear fluid specimens, as well as alternative methods to study the activity of these organisms in the otitis-prone middle ear with the goal of understanding whether they play a role in the pathogenesis of rAOM.

This thesis has presented novel contributions to the understanding of the microbial communities in the upper respiratory tract of children with rAOM. This includes the identification of two organisms that may be involved in the pathogenesis of rAOM, and two organisms strongly associated with resistance to rAOM. The thesis has highlighted important avenues for further work leading to the potential translation of the commensal organisms into probiotic therapies. This knowledge, including the characterisation of methodological challenges associated with undertaking next-generation sequencing on these specimen types, is essential for further 'omics-based studies in the upper respiratory tract.
## Table of Contents

### Chapter 1: Literature review

1.1 Otitis media: a common childhood disease ........................................... 2  
  1.1.1 Acute otitis media (AOM) and recurrent AOM (rAOM) .................... 2  
  1.1.2 Otitis media with effusion (OME) and chronic OME (COME) ............ 3  
  1.1.3 Chronic suppurative otitis media (CSOM) ....................................... 5  
1.2 Factors influencing susceptibility to OM ............................................. 5  
  1.2.1 Host-related risk factors ............................................................. 5  
  1.2.2 Environmental risk factors .......................................................... 6  
1.3 Pathogenesis of acute otitis media .................................................. 7  
  1.3.1 Major bacterial otopathogens ...................................................... 7  
  1.3.2 The role of viruses in otitis media .............................................. 8  
  1.3.3 Novel bacterial otopathogens ..................................................... 9  
  1.3.4 Mechanisms of bacterial resistance ............................................ 11  
1.4 The efficacy of current treatments for AOM ....................................... 12  
  1.4.1 Antibiotic treatment ................................................................. 12  
  1.4.2 Surgical ventilation ..................................................................... 13  
  1.4.3 Prevention of AOM by vaccination ............................................. 13  
1.5 Next-generation sequencing approaches for understanding complex  
  microbial diseases .............................................................................. 14  
  1.5.1 The concept of ‘microbiome’ ....................................................... 14  
  1.5.2 Metataxonomics (who is there?) ................................................... 15  
  1.5.3 Metagenomics (what are they capable of doing?) ............................ 15  
  1.5.4 Metatranscriptomics (what are they doing?) .................................. 16  
  1.5.5 The microbiome of the upper respiratory tract in otitis-prone  
    children ......................................................................................... 16  
1.6 Probiotic therapy: an option for otitis media? .................................... 22  
  1.6.1 Probiotics .................................................................................... 22  
  1.6.2 Probiotics derived from the microbiota ....................................... 23  
  1.6.3 Probiotics in the treatment of AOM ........................................... 24  
1.7 Aims of this thesis ............................................................................. 27
Chapter 2: Characterising the microbiota with 16S rRNA gene sequencing

2.1 Abstract .................................................. 30
2.2 Background .................................................. 31
2.3 Methods .................................................. 32
   2.3.1 Patient recruitment .................................. 32
   2.3.2 Sample collection and storage ....................... 33
   2.3.3 DNA extraction and sequencing preparation ........ 34
   2.3.4 Viral detection ...................................... 34
   2.3.5 Positive sequencing control generation ............. 36
   2.3.6 Amplicon sequencing ................................ 36
   2.3.7 Preprocessing of sequence data ...................... 38
   2.3.8 Covariate and diversity analyses .................... 39
   2.3.9 Co-occurrence analyses ............................. 40
   2.3.10 Differential abundance analyses ................... 40
2.4 Results .................................................. 42
   2.4.1 Study population characteristics ..................... 42
   2.4.2 Bacterial taxa identified across the samples ........ 42
   2.4.3 Diversity within the nasopharynx and middle ear ..... 44
   2.4.4 Comparing the microbiome of the nasopharynx between rAOM-prone and rAOM-resistant children .... 44
   2.4.5 Comparing the microbiome of different sample types within children with rAOM ......................... 49
   2.4.6 Patterns of bacterial co-occurrence .................. 54
   2.4.7 Detection of respiratory viruses ...................... 56
2.5 Discussion ............................................... 57
2.6 Conclusions ............................................ 66

Chapter 3: Exploration of the microbiome with shotgun metagenomics sequencing

3.1 Introduction ............................................. 68
3.2 Methods .................................................. 70
   3.2.1 Patient recruitment and sample collection ........... 70
   3.2.2 Selection of samples for metagenomic shotgun sequencing ... 71
   3.2.3 Removal of human DNA ................................ 72
   3.2.4 Metagenomic shotgun sequencing ..................... 73
   3.2.5 Data pre-processing .................................. 73
   3.2.6 Taxonomy analysis ................................... 74
3.2.7 Functional analysis .................................................. 74
3.3 Results ................................................................. 76
  3.3.1 Characteristics of participants .................................... 76
  3.3.2 Removing contaminating host DNA and sequences from nasopharyngeal swabs and middle ear fluids .......... 78
  3.3.3 Species level taxonomy of bacteria and viruses .............. 78
  3.3.4 Comparisons between the A. otitidis and T. otitidis sequences in the MEF and ECS sample pools ............... 81
  3.3.5 The functions of the nasopharyngeal and middle ear microbiome .............................................. 81
3.4 Discussion ............................................................. 93
  3.4.1 Removal of human DNA from specimens with high host contamination: a persistent challenge ................. 93
  3.4.2 Bacterial taxonomy – shotgun metagenomics compared to 16S rRNA profiling ................................... 94
  3.4.3 Identifying strain variation in Alloiococcus and Turicella ................................................................. 96
  3.4.4 Viral taxonomy ....................................................... 96
  3.4.5 Characterising microbiome function with shotgun metagenomics .................................................... 98
  3.4.6 Predicting microbiome function from 16S rRNA amplicon data ....................................................... 99
  3.4.7 Towards further characterisation of the upper respiratory tract microbiome ........................................ 100
3.5 Conclusion ............................................................. 102

Chapter 4: In vitro assessment of inhibitory activity by the commensal species C. pseudodiphtheriticum and D. pigrum 103
4.1 Introduction ........................................................... 104
4.2 Methods .................................................................. 111
  4.2.1 Bacterial strains and growth conditions ....................... 111
  4.2.2 Bacterial growth curves ........................................... 113
  4.2.3 Agar drop assay ...................................................... 113
  4.2.4 Agar overlay assay .................................................. 114
  4.2.5 Disc diffusion assay ............................................... 114
4.3 Results ................................................................. 115
  4.3.1 Growth of C. pseudodiphtheriticum and D. pigrum in broth ............................................................... 115
  4.3.2 Detecting inhibition via an agar drop assay .................. 117
  4.3.3 Testing for inhibitory compounds via diffusion through agar ........................................................... 117
  4.3.4 Testing for inhibitory compounds in the extracellular and intracellular supernatants of C. pseudodiphtheriticum ................................................................. 120
4.4 Discussion ............................................................. 122
4.4.1 The inhibition of *M. catarrhalis* ........................................ 122
4.4.2 Understanding the mechanism of inhibition .......................... 123
4.4.3 Antimicrobial effects may be strain-specific ......................... 124
4.4.4 Further research to determine how *Corynebacterium* and
*Dolosigranulum* contribute to health in the URT ..................... 125

4.5 Conclusions ................................................................. 127

Chapter 5: Profiling the microbial metatranscriptome and host tran-
scriptome of middle ear fluid using dual RNA-seq ........................ 129
5.1 Introduction ................................................................... 130
5.2 Methods ...................................................................... 132
5.2.1 Patient recruitment, sample collection and storage .............. 132
5.2.2 RNA extraction ......................................................... 133
5.2.3 Dual RNA sequencing .................................................. 134
5.2.4 Data processing and analysis ......................................... 135
5.2.5 Metatranscriptomics pipeline ......................................... 135
5.2.6 Human transcriptomics pipeline ..................................... 135
5.3 Results ........................................................................ 136
5.3.1 Isolation and sequencing of total RNA from middle ear effusion
and blood specimens .......................................................... 136
5.3.2 Differences in the host expression profiles: rAOM compared to
COME (Aim 1) .................................................................. 137
5.3.3 Characterisation of the microbial metatranscriptome of MEE
(Aim 2) ........................................................................ 140
5.3.4 The role of hypoxia in the otitis-prone middle ear (Aim 3) ...... 145
5.4 Discussion .................................................................... 147
5.4.1 Insights from the pilot data .......................................... 147
5.4.2 Preparation of RNA from MEE specimens ....................... 150
5.4.3 Achieving appropriate depth of sequencing for the dual analysis
of host and microbiome transcriptomes from MEE specimens .... 151
5.4.4 Improved methods for studying microbial activity in the middle
ear ............................................................................. 153
5.5 Conclusions .................................................................. 154

Chapter 6: General discussion ...................................................... 155

References ........................................................................ 165
Appendix A: Published manuscript and supplementary material
   (Chapter 2)  A1

Appendix B: Sequence read counts in the shotgun metagenomics
data (Chapter 3)  A51

Appendix C: List of hypoxia-related genes (Chapter 5)  A53
List of Tables

Table 1.1  Studies utilising next-generation sequencing methods to characterise the microbiome of the upper respiratory tract in otitis-prone children. ......................................................... 18
Table 1.2  Randomised controlled trials investigating the efficacy of probiotic therapies for otitis media in children. ................................................................. 25

Table 2.1  Gene targets for multiplex respiratory virus PCR. ................. 35
Table 2.2  Species included in the positive sequencing control. ............... 37
Table 2.3  Demographic characteristics of children recruited to the study. 43
Table 2.4  Genus-level community composition by sample type. ............ 45
Table 2.5  Viral identification rates in the nasopharynx of cases and controls. 58
Table 2.6  Viral concordance rates between the middle ear and nasopharynx of cases. ................................................................. 59

Table 3.1  Demographic characteristics of participants. ....................... 77
Table 3.2  DNA viruses detected by shotgun metagenomics sequencing. 82

Table 4.1  Studies observing the presence of Corynebacterium and Dolosigranulum in the upper respiratory tract microbiota and their association with a healthy state. ................................................. 107
Table 4.2  Otopathogen strains used in the interference assays. ............ 112

Table 5.1  RefSeq functions detected in the middle ear metatranscriptome. 143
Table 5.2  Functional content of the middle ear metatranscriptome categorised by the SEED subsystems database. 144
Table 5.3  Activity of Alloiococcus and Turicella in the middle ear. ....... 146

Table A.1  Taxonomic composition of the positive control (mock commu- nity) replicates. ................................................................. A33
Table A.2  Taxonomic composition of the negative extraction and PCR controls. ................................................................. A34

Table B.1  Summary of usable metagenomics reads per sample. .......... A52
Table C.1  Human genes analysed in Chapter 5 as part of the hypoxia pathway (n = 296).
List of Figures

Figure 1.1 Otoscopic images of OM phenotypes. ...................................... 4

Figure 2.1 Alpha diversity measured by a) Faith’s Phylogenetic Diversity
   and b) Inverse Simpson metrics grouped by sample type. ..................... 46

Figure 2.2 Principal coordinates analysis (PCoA) on nasopharyngeal sam-
   ples from cases and controls. ...................................................... 48

Figure 2.3 Log CSS normalised counts of differentially abundant OTUs
   between rAOM-prone and rAOM-resistant children. ............................. 50

Figure 2.4 Principal coordinates analysis (PCoA) on samples from children
   with rAOM. .............................................................................. 51

Figure 2.5 Log CSS normalised counts of differentially abundant OTUs
   amongst sample types within the cases. ............................................. 53

Figure 2.6 Correlations between OTUs. ................................................... 55

Figure 3.1 Comparison of taxonomic summaries by 16S rRNA amplicon
   sequencing and shotgun metagenomics sequencing. .......................... 79

Figure 3.2 Enriched functional categories in the nasopharyngeal micro-
   biomes of cases and controls. ...................................................... 84

Figure 3.3 Enriched functional categories in the middle ear and nasophar-
   ynx of children with rAOM. ............................................................ 89

Figure 4.1 Growth curve results for C. pseudodiphtheriticum and D. pigrum.116

Figure 4.2 Agar drop assay results for M. catarrhalis. ............................... 118

Figure 4.3 Overlay assay results. .......................................................... 119

Figure 4.4 Disc diffusion assay results for M. catarrhalis against C. pseu-
   dodiphtheriticum extracts. .............................................................. 121

Figure 5.1 Description of samples selected for RNA-seq. .......................... 138

Figure 5.2 Bioanalyzer gel images and peaks for the sequenced samples. 139

Figure 5.3 Abundance of the differentially expressed DGAT2 transcript. 141

Figure 5.4 Organism composition of the middle ear metatranscriptome. 142
Figure A.1  Diagrammatic overview of the 16S rRNA gene data analysis pipeline. .......................................................... A28

Figure A.2  Procrustes analysis of raw and rarefied datasets. .................. A36

Figure A.3  Beta diversity PCoA in the nasopharynx of cases and controls, sorted by other covariates. .............................................. A38

Figure A.4  Beta diversity PCoA in the samples from the cases, sorted by other covariates. .......................................................... A42

Figure A.5  Procrustes analysis of left and right ear samples. .................. A46

Figure A.6  Procrustes analysis of MEF and MER samples. ................. A48
Acknowledgements

This research was supported by an Australian Government Research Training Program (RTP) Scholarship and a Wesfarmers Centre of Vaccines and Infectious Diseases top-up scholarship. I would also like to thank the Friends of the Institute (Telethon Kids) and the Centre for Research Excellence in Ear and Hearing Health for their generous travel award support that allowed me to communicate this work to the otitis media and microbiome communities. This research was funded by a Telethon New Children’s Hospital research grant and a Wesfarmers Centre of Vaccines and Infectious Diseases seed funding grant. I appreciate the support of the Telethon Kids Institute, inside which I spent the majority of my PhD. The Institute is a welcoming place with many talented researchers and I am grateful to have been part of it.

There are a lot of people whose efforts and support have contributed to the completion of this thesis. I would first like to thank my supervisors; Associate Professor Christopher Peacock, Dr Sarra Jamieson, Associate Professor Timo Lassmann and Professor Manfred Beilharz for their expertise, support, mentorship, feedback and encouragement. To Chris and Sarra, thank you for giving me autonomy, for the opportunity to work on this project and to gain all of the skills that came with it; it has been a valuable and enjoyable experience.

I am also grateful for the support of many other people. Harvey Coates, Shyan Vijayasekaran, Paul Bumbak and all of the operating theatre staff for their assistance with the recruitment of patients and collection of samples; also Kara Imbrogno and Anke Hoskins for collecting many of them. Thank you to Denise Anderson for your solid statistical advice, and to Ghislaine Platell for helpful pipeline discussions. I benefited from the advice and expertise of Janessa Pickering, Barbara Chang, Kyra Cottrell and Hanna Sidjabat in planning my bacterial interference assays. I thank Kim Carter and Tim Finney for the management of the bioinformatics servers at Telethon Kids, and the Pawsey Supercomputing Centre for the use of their computing cloud. I am grateful to the members of AusMOM for supporting my effort to create a more formal group for sharing ideas and discussing important issues in OM microbiome research, particularly to Robyn Marsh for her sound advice.
and mentoring. To Professor Jennie Blackwell, thank you for your strong mentorship
and helping me to discover what I enjoy most.

Thank you to those who have ensured I remain human, providing mentorship,
friendship and food at the lab; especially Gen, Kara, Calila, Ben, Buzz, Dave, Alexia
and Demi. Dave, thank you for introducing me to blogging, Twitter and R Markdown.
Gen, Alexia, Ben and Philipp, thank you for casting an eye of improvement over my
chapters. To anyone who has assisted me in this journey whom I have inadvertently
missed - I am sorry, and I appreciate you.

Finally, I have deep gratitude for those closest to me who have always remained
proud of what I have achieved. To my parents, for their unconditional support and
encouragement as well as their inexpensive proof-reading, advisory and consultancy
services. To my partner, who has always supported my study and career choices
with enthusiasm, kept me properly fed and made sure that I never give up on that
which I want to achieve.

This thesis was written using Bookdown, created by Yihui Xie; using a template
created by Chester Ismay and modified by James Goldie.
Authorship Declaration

This thesis contains work that has been published.


**Location in thesis:** Chapter 2

**Student contribution to the work:** Assisted with study design, patient questionnaire interviews and collection of specimens; conducted sample processing for 16S rRNA gene sequencing, performed all bioinformatics and statistics analysis, contributed substantially to data interpretation and wrote the manuscript.

**Co-author signatures and dates:**

Kara Imbrogno
Date: 03/12/2018

Ch
Date: 03/12/2018

Denise Anderson
Date: 29/11/2018

Danny Mok
Date: 07/12/2018
Student signature: ___________________________ Date: 18/12/2018

I, Associate Professor Christopher Peacock certify that the student’s statements regarding their contribution to the work listed above is correct.

Coordinating supervisor signature: ___________________________

Date: 13/12/2018
List of Abbreviations

AOM  Acute otitis media
ATCC  American Type Culture Collection
ATP  Adenosine triphosphate
AUD  Australian dollars
BAM  Binary alignment map
BLASTN  Basic local alignment search tool (nucleotide)
bp  Base pairs
BP  Biological process
CAPD  Continuous ambulatory peritoneal dialysis
CC  Cellular component
CFU  Colony forming unit
CI  Confidence interval
COME  Chronic otitis media with effusion
COPD  Chronic obstructive pulmonary disease
Cp1  Corynebacterium pseudodiphtheriticum strain 1
Cp2  Corynebacterium pseudodiphtheriticum strain 2
CPM  Count per million
CSOM  Chronic suppurative otitis media
CSS  Cumulative sum scaling
DNA  Deoxyribonucleic acid
Dp1  Dolosigranulum pigrum strain 1
**Dp2** *Dolosigranulum pigrum* strain 2

dsDNA  Double-stranded DNA

DTT  Dithiothreitol

DUF  Domain of unknown function

EAC  External auditory canal

ECS  Ear canal swab

EDTA  Ethylenediaminetetraacetic acid

ENT  Ear, nose and throat

FC  Fold change

FDR  False discovery rate

Gb  Gigabases

GWAS  Genome-wide association study

HAdV  Human adenovirus

HBoV  Human bocavirus

HCoV  Human coronavirus

HIV  Human immunodeficiency virus

HMPV  Human metapneumovirus

HPIV  Human parainfluenza virus

HREC  Human research ethics committee

HS  High sensitivity

ID  Identification (number)

IFV  Influenza virus

IGV  Integrative genomics viewer

IL  Interleukin

IQR  Inter-quartile range

IS  Inverse Simpson

ITS  Internal transcribed spacer
RNA  Ribonucleic acid
rRNA  Ribosomal RNA
RPK  Reads per kilobase
RSV  Respiratory syncytial virus
RTP  Research Training Program
RV  Rhinovirus
SAM  Sequence alignment map
SNP  Single nucleotide polymorphism
STGGB  Skim milk tryptone glucose glycerol broth
TE  Tris-EDTA
TGF  Transforming growth factor
TNF  Tumor necrosis factor
TTV  Torque teno virus
URT  Upper respiratory tract
URTI  Upper respiratory tract infection
US  United States
UTR  Untranslated region
UV  Ultraviolet
WGS  Whole genome shotgun
x g  Times gravitational force
µg  Microgram
µl  Microlitre
Chapter 1

Literature review

This chapter contains an introduction to the disease *otitis media* in children, including the risk factors, efficacy of current available treatments and the pathogenic organisms responsible for the disease. Next-generation sequencing techniques including metataxonomics, metagenomics and metatranscriptomics are introduced as tools for the exploration and characterisation of the microbial communities in the upper respiratory tract. A summary of otitis media microbiome work to date is provided. Probiotics are introduced as a promising alternative therapy for otitis media; the development of which relies on improved knowledge of these microbial communities. The chapter concludes with a summary of the aims addressed by this thesis.
1.1 Otitis media: a common childhood disease

Otitis media (OM) describes a group of inflammatory conditions of the middle ear, characterised by the presence of fluid behind the tympanic membrane (the eardrum).\textsuperscript{1,2} It occurs most commonly in young children due to their underdeveloped immune system, and the shorter length and more horizontal angle of their Eustachian tubes compared to adults through which movement of air and fluid is difficult.\textsuperscript{3-5} OM consists of a spectrum of interrelated disease phenotypes,\textsuperscript{1,6} though children with OM may be broadly diagnosed with acute otitis media (AOM) or otitis media with effusion (OME). OM is a very common childhood disease; an estimated 73\% of Australian children experience an episode of OM before 12 months of age.\textsuperscript{7} The economic burden of OM is substantial, with the total treatment costs of OM in Australia estimated at $100 million to $400 million in 2008.\textsuperscript{7}

In Australia, OM and its complications are disproportionately higher in Indigenous Australians, particularly those living in remote communities. Australian Indigenous children experience far more severe and frequent OM than is experienced by non-Indigenous children,\textsuperscript{8,9} and have one of the highest rates of severe OM worldwide.\textsuperscript{10} In a study in remote Northern and Central Australia, only 8\% of the 709 Indigenous children examined had two normal middle ears; 91\% of the children had an OM diagnosis.\textsuperscript{11}

1.1.1 Acute otitis media (AOM) and recurrent AOM (rAOM)

AOM describes a short-term infection of the middle ear and is characterised by redness, pain and bulging of the eardrum, often with the presence of a fever and the collection of purulent fluid within the middle ear\textsuperscript{1,2} (see Figure 1.1). AOM is a common childhood illness, with around 60\% of children experiencing AOM at least once by 3 years of age.\textsuperscript{12,13} Infection of the middle ear typically follows a viral upper respiratory tract infection (URTI),\textsuperscript{6,14,15} which causes congestion and swelling of the tissues surrounding the Eustachian tube. This impairs the drainage of fluid and regulation of middle ear pressure, which may result in the aspiration of bacteria and viruses from the nasopharynx to the middle ear, where they provoke an immune response and can cause an opportunistic infection.\textsuperscript{5,15,16} Episodes of AOM are usually of short duration and are commonly self-resolving; in randomised controlled trials, around 60\% of children recover by 24 hours from diagnosis or the
1.1. Otitis media: a common childhood disease

The most commonly implicated pathogens in AOM (known as otopathogens) are *Streptococcus pneumoniae*, non-typeable *Haemophilus influenzae* and *Moraxella catarrhalis*; however, AOM is a complex microbial infection and may involve other bacteria and respiratory viruses. Some children experience frequent, recurring episodes of AOM. Recurrence of 3 episodes or more in 6 months, or 4 or more in 12 months is diagnosed as recurrent AOM (rAOM). A substantial proportion of children experience rAOM, with one large study reporting 17% of children experiencing at least 3 episodes during their first year of life, and 1% with at least 6 episodes. Children with rAOM are commonly treated with antibiotics by their general practitioner, and when infections persist are referred to an otolaryngologist for grommet surgery. This is typically a day surgery procedure in which a grommet (a small plastic or stainless steel tube) is inserted into an incision in the eardrum made by the surgeon. Ventilation of the middle ear allows for drainage of middle ear fluid, and bypasses the dysfunctional Eustachian tube to normalise the pressure within the middle ear; preventing the accumulation of more fluid.

1.1.2 Otitis media with effusion (OME) and chronic OME (COME)

OME, also known as glue ear, describes the presence of middle ear fluid without signs of an acute infection and so is relatively asymptomatic (see Figure 1.1). OME can occur spontaneously as a result of Eustachian tube dysfunction. Inflammation of the mucosa, enlarged adenoids or inability of the surrounding muscles to open the tube can result in negative pressure in the middle ear and the transudation of fluid into the middle ear space. It is also common for OME to follow an episode of AOM, hence the interrelated nature of OM phenotypes. The prevalence of OME is more challenging to estimate due to its asymptomatic nature, but it remains high. Based on the data from large cohort studies involving a total of over 5000 children, Bhutta (2014) predicted via modelling that the overall prevalence of OME amongst 2 year old children is 74.5%, where 15.6% of children with OME are predicted to have no prior history of AOM.

When OME persists for more than 3 months, it is considered chronic OME (COME); like rAOM, COME is also managed by ventilation of the middle ear cavity with grommets. As the presence of middle ear fluid reduces the conduction
Figure 1.1: Otoscopic images of OM phenotypes. From left to right, otoscopic images of (top): a normal tympanic membrane, acute otitis media, otitis media with effusion; (bottom): a stainless steel grommet in place and acute otorrhoea through membrane perforation (CSOM). Images provided courtesy of Michael Hawke MD (otitisme-media.hawkelibrary.com).
of sound, persistent OME can lead to hearing loss which can subsequently affect a child’s development. Children with an early history of OM tend to have poorer phonological awareness, reading and spelling skills, demonstrating inattention and distractibility. In a Western Australian cohort of children, significant associations were found between recurrent OM and internalising behaviours and withdrawal.

1.1.3 Chronic suppurative otitis media (CSOM)

CSOM describes a severe state of OM where the tympanic membrane has spontaneously perforated, the middle ear is inflamed and there may be a persistent discharge from the ear (otorrhoea) (see Figure 1.1). CSOM usually occurs after repeated episodes of AOM, but can also follow COME in populations where it is highly prevalent, such as Indigenous Australians. Perforation occurs when the pressure in the middle ear increases due to acute inflammation, and the membrane ruptures. In Australia, CSOM is more common in Indigenous children than in non-Indigenous children, and Australian Indigenous children have one of the highest rates of CSOM in the world. The disease can lead to further complications like mastoiditis (inflammation of the mastoid bone) and meningitis, leading to death in some patients in developing countries. Following perforation of the tympanic membrane, the middle ear can be secondarily infected by *Pseudomonas aeruginosa* and *Staphylococcus aureus*, resulting in chronic otorrhoea.

1.2 Factors influencing susceptibility to OM

1.2.1 Host-related risk factors

OM is a multifactorial disease, with several host-related and environmental risk factors contributing to its pathogenesis which have been well-reviewed. An important component of susceptibility to OM is genetic. OM is heritable in some populations, with twin and triplet studies in predominantly Caucasian children reporting a heritability estimate (proportion of phenotypic variance attributable to genetics) of 0.57 for symptoms of acute infection and 0.73 for duration of middle ear effusion. Genome wide association studies (GWAS) have aimed to identify specific genomic variants associated with OM, and have reported associations with variants in the *CAPN14* and *GALTN14* genes, involved in the TGFβ pathway.
in an intergenic region;\textsuperscript{39} and variants downstream of the gene \textit{FNDC1}, which is thought to have a role in inflammation.\textsuperscript{40} Variants within other genes encoding proteins involved in immune regulation and inflammation such as IL-1, IL-10, TNF$\alpha$ and mucin genes have also shown significant association with OM in candidate gene studies;\textsuperscript{41,42} however these studies have had small sample sizes and the results have not necessarily been replicated.

Other host-related factors strongly influencing the risk of OM are age and ethnicity. The incidence of OM is higher in very young children, peaking in the first year of life.\textsuperscript{2,43} Australian Indigenous children experience the disease as early as 6 weeks of age,\textsuperscript{44} with incidence peaking at 5-9 months compared to 10-14 months in non-Indigenous children.\textsuperscript{45} Australian Indigenous children are at increased risk of severe OM compared to non-Indigenous children;\textsuperscript{2,9} a large-scale study in remote Far North Queensland, Australia reported a 24.6% prevalence of CSOM in children within Aboriginal communities compared to <5% in predominantly non-Indigenous communities.\textsuperscript{46} Part of this burden is due to the poor infrastructure and overcrowding experienced by Indigenous people in rural and remote areas.\textsuperscript{9,11,47} Additionally, differences in craniofacial morphology may influence a child’s risk of OM, via alteration of the shape and function of the Eustachian tube. Differences in craniofacial morphology have been reported for children with OME\textsuperscript{48} and children with rAOM\textsuperscript{49} compared to children without OM, and these differences disappear with age.\textsuperscript{48} Children with immunodeficiencies are also susceptible to OM; a study in Sub-Saharan Africa reported the prevalence of tympanic membrane perforations due to OM at 17% in children infected with HIV, compared to 3% in HIV-uninfected children.\textsuperscript{50} Australian Indigenous children with OM have been reported to display decreased antibody responses to non-typeable \textit{Haemophilus influenzae}, one of the major otopathogens, compared to non-Aboriginal children with OM and non-Aboriginal healthy controls.\textsuperscript{51}

\subsection*{1.2.2 Environmental risk factors}

There are also several environmental factors contributing to a child’s risk of developing OM. Amongst the most substantial observed in many studies is attendance at day care (odds ratio, OR, > 1.4),\textsuperscript{2,52-54} where contact with other children facilitates the transmission of respiratory viruses and otopathogen colonisation of the nasopharynx. Introduction to day care at an earlier age has also been associated with a risk of OM,\textsuperscript{55} and living with siblings similarly increases this risk (OR > 1.2).\textsuperscript{2,52,53,55} Breastfeeding is a significant protective factor; children who are exclusively breastfed for longer (at
1.3 Pathogenesis of acute otitis media

1.3.1 Major bacterial otopathogens

There are three bacterial species which are most commonly implicated as the major otopathogens, or causative agents of AOM. These organisms, *Streptococcus pneumoniae*, non-typeable *Haemophilus influenzae* (NTHi) and *Moraxella catarrhalis* are each commensal inhabitants of the nasopharynx that can opportunistically infect the middle ear. They are the organisms most commonly isolated or detected by polymerase chain reaction (PCR) in the middle ear fluid of children with OM.\(^{20,60,61}\) In some developing countries and in Indigenous Australian children, *Staphylococcus aureus* is also a common pathogen.\(^{60,62}\) A recent systematic review assessed the global prevalence of the three major otopathogens, finding NTHi as the predominant otopathogen in children with rAOM.\(^{20}\) This represents a shift in recent years, where the introduction of the 7-valent and 13-valent pneumococcal vaccines (PCV7 and PCV13) in 2000 and 2010 respectively\(^{63}\) has reduced the prevalence of *S. pneumoniae* carriage and therefore *S. pneumoniae*-associated OM, with a subsequent proportional increase in NTHi-associated OM.\(^{13,64,65}\) A large epidemiological study in the US recently reported a substantial decrease in *S. pneumoniae* in the PCV7 and PCV13 eras, with an increase in *M. catarrhalis* OM in the PCV13 era and a consistently high prevalence of NTHi OM.\(^{13}\) *M. catarrhalis* is often the least commonly detected otopathogen of the three, however it has been thought that its role in OM has been underestimated due to its dormancy in biofilm structures and the fact that it appears to cause milder disease.\(^{66}\) *M. catarrhalis* does commonly co-occur with the other otopathogens and may also contribute to the pathogenesis of AOM by synergistically supporting NTHi and *S. pneumoniae*.\(^{67–70}\) All three major otopathogens can also
opportunistically invade the lower airways to cause severe disease.\textsuperscript{71–73}

At least one of the three major otopathogens are typically detectable by culture or PCR from the majority of children with rAOM, with some children harbouring two or all three species in the nasopharynx.\textsuperscript{64,68,74} However, there are approximately 12-35\% of children with AOM in which no known otopathogen can be detected in the middle ear fluid by culture or PCR.\textsuperscript{68,75–79} These cases may experience AOM due to viruses only;\textsuperscript{15} though it is also possible that there are other bacterial otopathogens which are not well-characterised or easy to detect that may be involved in the pathogenesis of AOM.

1.3.2 The role of viruses in otitis media

AOM is commonly preceded by a viral URTI. It is thought that the mechanism of virus-induced AOM involves the inflammatory response to the virus in the upper respiratory tract, which impairs the function of the Eustachian tube. Negative pressure in the middle ear results in aspiration of bacteria and viruses from the nasopharynx into the middle ear, where further inflammatory responses occur.\textsuperscript{15} Respiratory viruses that have most commonly been associated with AOM include influenza virus, rhinovirus, respiratory syncytial virus and adenovirus.\textsuperscript{80,81} Children with rAOM have significantly higher rates of nasopharyngeal colonisation with viruses compared to healthy children.\textsuperscript{81} Viruses are also commonly detected in middle ear fluid.\textsuperscript{81} The presence of adenovirus is correlated with \textit{S. pneumoniae} and \textit{M. catarrhalis} carriage in children with rAOM,\textsuperscript{67,81} and rhinovirus is correlated with all three major otopathogens.\textsuperscript{67} Viral load of respiratory syncytial virus has also been associated with increased risk of AOM.\textsuperscript{82} While it has been debated whether the presence of viral nucleic acids detected by PCR in middle ear fluid is indicative of a role for viruses in AOM or merely represents a ‘bystander’ role without evidence of pathogenicity,\textsuperscript{83} some studies have observed a mechanistic association between respiratory viruses and the major otopathogens. Firstly, it appears that asymptomatic upper respiratory viral infection does not lead to AOM; supporting the notion that the pathogenesis of viral-induced AOM is reliant on the inflammatory response and dysfunction of the Eustachian tube.\textsuperscript{84} \textit{In vitro} studies have reported increased adherence of \textit{S. pneumoniae} to human respiratory epithelial cells after infection with rhinovirus or respiratory syncytial virus.\textsuperscript{85,86} Within animal models, coinfection with influenza A virus or adenovirus has also been reported to enhance \textit{S. pneumoniae} colonisation.\textsuperscript{87,88} Promotion of \textit{M. catarrhalis} AOM by respiratory
syncytial virus has been observed in a chinchilla model, an animal which has also been useful for modelling NTHi AOM when coinfected with adenovirus to prolong the presence of NTHi in the nasopharynx and middle ear and increase the severity of disease. It is therefore becoming well-established that prior or concurrent infection with respiratory viruses predisposes children to AOM, most likely via the induction of inflammation in the upper respiratory tract.

1.3.3 Novel bacterial otopathogens

There are two main bacterial species whose roles as potential otopathogens have remained under debate since they were initially characterised; *Alloiococcus otitidis* and *Turicella otitidis*. Both *A. otitidis* and *T. otitidis* were first isolated by culture from the middle ear fluid of children with otitis media, prompting their association with the disease. *A. otitidis* appears to be more dominant than *T. otitidis* in the literature, likely because *A. otitidis* can be detected by both culture and by targeted PCR, whereas there are currently no published primer pairs for the detection of *T. otitidis* and it is not often distinguished from closely-related commensal *Corynebacterium* species. *A. otitidis* and *T. otitidis* are most commonly reported in the middle ear and external auditory canal (EAC) of children with OM. *A. otitidis*, as detected by species-specific PCR, is present in 19-61% of middle ear specimens from children with OME or COME, and 7-50% of those with AOM or rAOM. The organism is frequently detected at a higher rate than the three major otopathogens, though only in children with OME. Higher frequencies of *A. otitidis* have also been reported in non-purulent or OME middle ear fluid specimens (25-61%) compared to purulent or AOM specimens (18-50%) by PCR, though this has not reached statistical significance. These patterns suggest that *A. otitidis* may peak outside of acute infections.

One of the reasons that the pathogenicity of *A. otitidis* and *T. otitidis* in OM remains under debate is their prevalence in the EAC; though the numbers of *A. otitidis* or *T. otitidis*-positive specimens are often small, and there are few studies that have investigated their prevalence in the EAC. They have been reported at a higher frequency in the EAC of children with OM than in healthy children (*A. otitidis* 4% vs. 2% by PCR, *T. otitidis* 23% vs. 11% by culture). However, one study observed a higher frequency of *A. otitidis* in healthy children (29% vs 6%). Both organisms have also been reported amongst the most commonly cultured organisms in the EAC of healthy children and adults with *A. otitidis* reported in 6% of EAC
specimens, and \textit{T. otitidis} in 12\%\textsuperscript{108}. It has been suggested that the presence of \textit{T. otitidis} in middle ear fluid is a result of contamination from the EAC during sampling and it is not involved in OM\textsuperscript{105} though it has also been suggested that the EAC may be a reservoir for infection of the middle ear with \textit{A. otitidis} via a tympanic membrane perforation\textsuperscript{109}. The degree of contamination from the ear canal remains unclear as studies reporting the detection of \textit{A. otitidis} rarely specify a lack of tympanic membrane perforations in the study population and sterilisation of the EAC or measures taken to avoid contact with it during sampling. However, at least two studies have done this and have reported culture of \textit{A. otitidis} from 24-35\% of specimens from children with OME\textsuperscript{97,103}. Thus, there is still evidence to suggest that \textit{A. otitidis} does inhabit the middle ear in children with OM.

Both organisms are uncommonly found in the nasopharynx. \textit{A. otitidis} has previously been reported by PCR in 9\% of nasopharyngeal samples from children with culture-negative AOM – a lower detection rate than the otopathogens in this cohort (\textit{S. pneumoniae} 45\%, \textit{H. influenzae} 31\% and \textit{M. catarrhalis} 10\%)\textsuperscript{75} – and in 7\% of children with an URTI\textsuperscript{107}. In the nasopharynx of healthy children, it is almost absent; reported in 0.5\% of children in one study\textsuperscript{110} and not detected in others\textsuperscript{96,106} each of these studies screened for \textit{A. otitidis} using PCR. \textit{T. otitidis} has never been reported to inhabit the nasopharynx. The pattern exhibited by \textit{A. otitidis} and possibly by \textit{T. otitidis} – high prevalence in the ear canal and low prevalence in the nasopharynx – is the opposite to that exhibited by the known otopathogens.

Studies investigating the ability of \textit{A. otitidis} and \textit{T. otitidis} to cause disease are scarce, possibly due to their status as EAC commensals and the lack of sufficient experimental evidence for their involvement in OM beyond detection in the middle ear and EAC. Live, killed and cell-free filtrates of \textit{A. otitidis} have been demonstrated to elicit an immune response \textit{in vitro}\textsuperscript{111–114} and children with \textit{A. otitidis}-positive OM do produce antibodies specific to the organism\textsuperscript{115}. The bacterial load of \textit{A. otitidis} in middle ear fluid is also similar to that of NTHi\textsuperscript{101}. However, in the only study examining the pathogenicity of \textit{A. otitidis} in an animal model, reactions to the organism injected into the middle ear of rats at high concentration were mild and did not lead to OM, suggesting that it is not otopathogenic\textsuperscript{107}. No studies have explored the behaviour of \textit{T. otitidis} \textit{in vitro} or \textit{in vivo}.

The differences between observations of the three major otopathogens and these two ‘controversial’ otopathogens indicates that their role in OM may be complex; they do not appear to have the ability to independently cause acute infection and may instead involve synergism with the main otopathogens\textsuperscript{116} or perpetuation of the
1.3. Pathogenesis of acute otitis media

inflammatory response. However, the involvement of *A. otitidis* and *T. otitidis* in the pathogenesis of OM remains poorly studied and they may simply be members of the normal aural flora.

1.3.4 Mechanisms of bacterial resistance

Each of the three major otopathogen species exhibits resistance to treatment with antibiotics or assault by the immune response. The resistance mechanisms employed by the pathogens render them capable of persisting within the middle ear between episodes of AOM; this persistence can contribute to the recurrence of AOM. The otopathogens are capable of evading antibiotic molecules or immune factors by living intracellularly or within biofilm structures, and some strains also carry antibiotic resistance genes.

Biofilms are structures consisting of a matrix of extracellular substances, including polysaccharides, proteins and DNA; a ‘slime’ in which bacteria live. These substances contribute to the adhesion of the biofilm matrix to a surface as well as providing nutrients for the bacteria within. From the outermost layers, the biofilm sheds planktonic, metabolically active bacteria. The deeper layers of biofilm contain metabolically dormant bacteria. The biofilm matrix provides a physical barrier to antibiotic molecules and immune cells or factors that require contact with bacterial cells to initiate killing or to interfere with their growth. Biofilms are a persistent reservoir of pathogens, which are capable of reinfection upon release from the biofilm structure. Biofilms have been directly observed in the middle ear, on the inside of the tympanic membrane and on the surface of adenoids of children with OM; these studies did not observe biofilm at these sites in healthy children.

All three otopathogens have been observed within biofilm, and often coexist in a multispecies biofilm. *A. otitidis* is also capable of forming biofilms as a single species as well as enhancing the growth and resistance of NTHi in a polymicrobial biofilm, compared to NTHi biofilm alone. While oral antibiotics do appear to reach the middle ear in sufficient concentrations to have an effect on planktonic bacteria, they are not capable of eradicating biofilms. Dornase alfa, an enzyme utilised to deplete biofilm in cystic fibrosis, has been trialled as a novel treatment to reduce biofilm in OM. This showed promise *in vitro*, fragmenting the DNA matrix of biofilms in clinical middle ear fluid samples. A clinical trial on children with clogged grommets demonstrated dornase alfa’s ability to unclog the grommets of 59% of the patients, though this outcome was not significantly different from treatment
with antibiotic drops.\textsuperscript{125}

Otopathogens also have other mechanisms for evading the immune system. \textit{S. pneumoniae}, NTHi and \textit{M. catarrhalis} are all capable of surviving intracellularly, where persistence within host cells shields them from antibiotic molecules and immune attacks.\textsuperscript{118,126} NTHi expresses IgA1 protease, which cleaves and inactivates IgA1; NTHi utilises this to increase its ability to invade and persist inside cells.\textsuperscript{127} \textit{M. catarrhalis} produces outer membrane vesicles; these vesicles can contain β-lactamase, an enzyme that degrades penicillins, shielding it from antibodies.\textsuperscript{128} \textit{M. catarrhalis} outer membrane vesicles with ubiquitous surface proteins UspA1 and UspA2 have also been demonstrated to interfere with the killing of NTHi by complement cascade.\textsuperscript{129}

Otopathogen strains whose genomes encode resistance to penicillins or macrolides may also contribute to the recalcitrance of rAOM.\textsuperscript{64,130–132} There is also established evidence that \textit{M. catarrhalis} can confer passive protection to \textit{S. pneumoniae} and NTHi against β-lactam antibiotics via the β-lactamases it expresses in outer membrane vesicles.\textsuperscript{66} Antibiotic resistance is an increasing global problem, however it is becoming apparent that vaccination is an effective way to reduce the carriage of antimicrobial resistant strains by enhancing the specific immune response against these pathogens with additional reduction of carriage via herd immunity.\textsuperscript{133} Additionally, the prevalence of penicillin-resistant otopathogen strains appears to have reduced following the introduction of the pneumococcal vaccines.\textsuperscript{65,134}

1.4 The efficacy of current treatments for AOM

1.4.1 Antibiotic treatment

While episodes of AOM commonly resolve spontaneously, antibiotic prescription rates for ear infections remain high; accounting for 23\% of all antibiotic usage in Australian children.\textsuperscript{8} The standard antibiotic prescription for AOM is amoxicillin.\textsuperscript{8,135} Antibiotics have been found to reduce the duration of middle ear fluid\textsuperscript{136} and the number of episodes of AOM.\textsuperscript{137} However, much of the recent literature reports a modest or marginal benefit of antibiotic use in most children with AOM.\textsuperscript{138,139} A recent Cochrane review determined that 60\% of children recover from an AOM episode within 24 hours from the start of treatment regardless of whether they have
taken antibiotics or placebo, and antibiotics are not effective at reducing pain in that time.\textsuperscript{18} Additionally, some children experience side effects with antibiotic use such as vomiting, diarrhoea and rash.\textsuperscript{18}

There are also some potential, long-term side effects of antibiotic use early in life. There is evidence to suggest that antibiotics significantly alter the microbiota in young children towards a less diverse state. Reduced richness of the gut microbiota has been shown to persist for up to 2 years after the use of macrolides,\textsuperscript{140} and a similar effect occurs in the respiratory microbiota of children with cystic fibrosis after long-term antibiotic use.\textsuperscript{141} Early antibiotic use has been associated with the development of asthma, obesity and autoimmune diseases later in life.\textsuperscript{140} Due to these marginal benefits and negative effects, antibiotics are currently not recommended for the treatment of AOM except in children under 2 years of age, Indigenous children and those with severe symptoms (high fever or vomiting).\textsuperscript{18,135,143,144}

\subsection{1.4.2 Surgical ventilation}

Surgical ventilation of the middle ear using grommets is the recommended treatment for children with rAOM and COME.\textsuperscript{22} However, the benefits of grommets in children with rAOM remain unclear. They have been shown to reduce the incidence of AOM after surgery and improve quality of life,\textsuperscript{145} however it is possible for children to have further infections while the grommet is in place. While grommets are designed to eventually be extruded, an estimated 15-20\% of children will require grommet reinsertion.\textsuperscript{146,147} A recent Cochrane systematic review found that there are no randomised controlled trials providing more than low-quality evidence on the effectiveness of grommets in children with rAOM, and these trials with low or very low-quality evidence were conducted prior to the introduction of the pneumococcal vaccine which has altered the epidemiology of AOM.\textsuperscript{23} Other reviews have also reported limited or unknown benefit for grommets in rAOM.\textsuperscript{135,148}

\subsection{1.4.3 Prevention of AOM by vaccination}

The introduction of PCV7 and PCV13 has substantially reduced the prevalence of \textit{S. pneumoniae} vaccine serotypes in children with OM.\textsuperscript{13,64,65,100,149,150} As a result of this reduction, the general trend in many countries involves a decrease in pneumococcal OM and an increase in NTHi-related OM.\textsuperscript{13,64,65} However, these vaccines cover only a small number of the >90 distinct \textit{S. pneumoniae} serotypes and the replacement of
vaccine serotypes with other serotypes in pneumococcal disease remains an issue.\textsuperscript{151} The vaccine against the encapsulated \textit{H. influenzae} b strain does not provide protection against NTHi, which is unencapsulated.\textsuperscript{71} There is currently no vaccine against NTHi and it remains challenging to develop one, with no capsular surface antigens to target and a large amount of variation in the genomes and outer membrane proteins of different strains.\textsuperscript{152} There are several promising vaccine candidates,\textsuperscript{153} and clinical trials in adults are underway.\textsuperscript{152,154} There are also no vaccines currently available for \textit{M. catarrhalis}. Also involved in invasive lung disease, \textit{M. catarrhalis} is a similarly challenging pathogen for vaccine development. This is partially due to the limitations of animal models of \textit{M. catarrhalis} infection as well as the lack of a targetable surface antigen; like NTHi, \textit{M. catarrhalis} lacks a capsule.\textsuperscript{66} As with NTHi, some promising vaccine targets have been identified\textsuperscript{155} and an ongoing clinical trial is assessing a vaccine containing antigens from both NTHi and \textit{M. catarrhalis} in adults with COPD.\textsuperscript{152} Additionally, there is evidence that influenza vaccines have been effective in reducing the burden of influenza-associated OM.\textsuperscript{156–158}

1.5 Next-generation sequencing approaches for understanding complex microbial diseases

1.5.1 The concept of ‘microbiome’

In microbiology, microorganisms are often sequenced individually from a pure culture. To fully understand the entire microbial community that lives in a particular environment, researchers may study the microbiome. The term ‘microbiome’ is used to refer to microorganisms, their genomes and the habitat in which they live; used similarly and derived from the word ‘biome’.\textsuperscript{159} The ‘microbiota’, like ‘flora’ or ‘fauna’, refers to the microbial organisms themselves which inhabit a particular niche.\textsuperscript{159} Microbiome studies, which commonly analyse the nucleic acids of microorganisms, have been made possible only in recent years by the rapid and substantial improvements in the output and cost of next-generation sequencing technologies. The microbiome of various human body sites has been extensively studied,\textsuperscript{160} primarily to understand the disruptions in the microbiota that occur as a consequence of, or lead to, human disease. This is a highly useful approach for studying rAOM, as it is a polymicrobial disease that involves both bacteria and viruses, where disruptions to the microbial community (e.g. colonisation with otopathogens, infection with viruses, treatment with antibiotics) play a major role. Studying the microbiome via next-generation
sequencing also has the substantial benefit of detecting and characterising organisms that are unknown, poorly studied, or challenging to grow in pure culture: research is otherwise largely limited to organisms that have been characterised and can be isolated and propagated in vitro. Exploring the microbiome of otitis-prone children enables the characterisation of potentially novel pathogens, the roles and interactions of different organisms in the disease, and an understanding of beneficial microbes that may reduce a child’s risk of developing rAOM. There are several approaches that exist for the characterisation of microbiomes.

1.5.2 Metataxonomics (who is there?)

The most commonly used high throughput method for identifying the microorganisms that are present in a sample and their estimated relative abundance is metataxonomics; the use of a conserved marker gene that is amplified with ‘universal’ primers. For bacteria, this can be done via 16S rRNA gene sequencing. This technique involves the use of PCR primers that anneal to evolutionarily conserved regions of the 16S rRNA gene, which is present in all bacteria. These conserved regions are highly similar across bacterial taxa; some primers have been experimentally demonstrated to cover up to 96% of all bacterial phyla. The PCR reaction amplifies the sequence in between, which is variable amongst different taxa; any of the 9 hypervariable regions of the gene may be sequenced, but human microbiome studies commonly utilise the first to the third (V1-V3) or the third and fourth (V3-V4) regions. The PCR amplicons are sequenced and analysed to identify the bacterial taxa that they belong to, generating a profile of the bacteria whose DNA is present in the sample. For eukaryotes and fungi, the 18S rRNA gene or internal-transcribed spacer (ITS) region may be sequenced instead, as these organisms do not have 16S rRNA. Metataxonomics generates a taxonomic profile, but does not generate information on the whole genomes of these organisms. Despite this, 16S rRNA gene sequencing is often erroneously referred to as ‘metagenomics’.

1.5.3 Metagenomics (what are they capable of doing?)

The metagenome refers to the “collection of genomes and genes from the members of a microbiota”. Sequencing of a metagenome involves fragmentation of all genomic DNA extracted from a specimen, next-generation sequencing of these fragments and the assembly or reference-based mapping of these sequence reads to provide complete
or partially complete genome sequences of each organism present. Importantly, as it captures sequence information from across the whole genomes of organisms rather than only a segment of a marker gene, metagenomics can provide information on the functional content of these genomes and therefore their possible functional role in the environment in which they live. Because this method sequences genomic DNA, it can describe the functions that organisms are capable of performing, but not necessarily those which are being used – it also cannot discriminate between intact, living microbes and dead microbes or extracellular DNA. It can however provide a more complete picture of the microbiome than metataxonomics, as bacteria, archaea, eukaryotes and DNA viruses can be detected; and by sequencing all genomic regions of an organism, allows species or strain-level classification. Metagenomics also overcomes some sources of bias inherent to metataxonomics. The relative abundance of organisms determined by 16S rRNA gene sequencing analysis can be skewed by variable efficiency of the PCR primers on different bacterial taxa, as well as variation in the number of copies of the multi-copy 16S rRNA gene.\textsuperscript{162,163}

1.5.4 Metatranscriptomics (what are they doing?)

Transcriptomics is the study of the transcripts (messenger RNAs) produced by an organism; the genes which are being actively expressed. Metatranscriptomics covers the analysis of all genes that are being actively expressed by the microbiota to provide a profile of the microbial functional activity at the time of sampling. Metatranscriptomics is a relatively new approach, though it is very powerful when used in conjunction with metagenomics data.\textsuperscript{164,165} A novel technique that has recently been developed is dual RNA-seq; an approach for the simultaneous sequencing of a host and microbial transcriptome as a way of studying host-pathogen interactions. Most of the dual RNA-seq studies conducted so far focus on \textit{in vitro} models of infection with a human cell line and a single pathogen,\textsuperscript{166} but it has also been applied to respiratory clinical samples for the profiling of the host transcriptome and microbial metatranscriptome.\textsuperscript{167}

1.5.5 The microbiome of the upper respiratory tract in otitis-prone children

Several studies have described the members of the upper respiratory tract microbiota in otitis-prone children, and are summarised in Table 1.1. Each of these studies
employ a metataxonomics approach, utilising 16S rRNA gene sequencing to describe the members of the microbiota in the nasopharynx and middle ear of children with OME as well as those with AOM. Microbiome studies can be difficult to compare due to differences in the populations studied, the DNA extraction method, the region of the 16S rRNA gene sequenced and the analysis pipeline, however there are some key patterns in what has been reported. Overall, these microbiota studies have observed that the otopathogens are more common in the nose or nasopharynx of children with OM than in healthy controls; this supports the notion that otopathogen colonisation is an important prerequisite for developing OM. Commensal genera, particularly Corynebacterium and Dolosigranulum, are more common in the nose or nasopharynx of healthy controls; being reported at reduced relative abundance in children with OM. This suggests that either the reduction in important nasopharyngeal commensals increases the risk for OM, or that these commensals are lost as a result of OM or antibiotic treatment. Most of these studies are cross-sectional and cannot distinguish between these possibilities. The recent longitudinal study by Chonmaitree et al. (2017) suggests that the relative abundance of commensals drops with the surge in otopathogens; however as microbiota profiles are compositional it is not known whether this corresponds to an actual reduction in the population of commensals, or an overall increase in bacterial load with proportionate overrepresentation of the otopathogens. These microbiota studies have also observed a reduced diversity in the nose or nasopharynx of children with OM compared to healthy controls, indicating that during OM some members of the healthy microbiota may be absent or below the level of detection. There is also increasing evidence that the diversity of the microbiota in middle ear fluid is lower than that of the nasopharynx.

In middle ear fluid, the otopathogen genera tend to be the major organisms observed; however in some studies A. otitidis is dominant instead. A. otitidis seems to be more prevalent in children with OME than in those with AOM, however relative abundances in the taxonomic profile can be skewed by the inherent biases in 16S rRNA gene sequencing described above, so one organism cannot be interpreted as more abundant than another. T. otitidis has also featured in some of these studies as an apparent major constituent of the middle ear microbiota.

These studies have provided a useful background on the bacterial organisms in the nasopharynx and middle ear which may not be detectable by culture or species-specific PCR. However, there is a limit to what can be learnt from these metataxonomics studies. 16S rRNA gene sequencing describes only bacteria, usually at genus level or
Table 1.1: Studies utilising next-generation sequencing methods to characterise the microbiome of the upper respiratory tract in otitis-prone children. All of these studies have used metataxonomics with 16S rRNA gene sequencing to describe the microbiota. Studies which include adult participants, do not use next-generation sequencing methods for describing the microbiota or do not distinguish between participants with OM and those with other conditions are not shown. MEF = middle ear fluid.

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of sample</th>
<th>Age of participants</th>
<th>Number and type of cases</th>
<th>Number and type of controls</th>
<th>Major findings</th>
</tr>
</thead>
</table>
| Laufer et al. (2011)
(2011)\(^{168}\) | Anterior nasal swabs | 6 months – 6 years | 25 children with OM | 83 children without OM but with URTI | • Corynebacterium and Dolosigranulum associated with decreased risk of OM  
• Haemophilus associated with increased risk of OM |
| Liu et al. (2011)
(2011)\(^{173}\) | MEF, adenoid, tonsil | 8 years | 1 child with COME | NA | • Pseudomonadaceae dominant in middle ear (intact tympanic membrane) |
| Hilty et al. (2012)
(2012)\(^{174}\) | Nasopharyngeal swabs | < 2 years | 153 infants with AOM | 10 infants without AOM | • Otopathogen families most frequent  
• Lower diversity and lower prevalence of commensal families in AOM |
| Pettigrew et al. (2012)
(2012)\(^{169}\) | Anterior nasal swabs | < 3 years | 72 children with AOM | 95 children without AOM but with URTI and 73 healthy children | • Lower diversity in children colonised with otopathogens  
• Children with Lactococcus, Propionibacterium, Corynebacterium and Dolosigranulum less likely to have AOM |
Table 1.1: (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of sample</th>
<th>Age of participants</th>
<th>Number and type of cases</th>
<th>Number and type of controls</th>
<th>Major findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lazarevic et al. (2013)</td>
<td>Saliva</td>
<td>1-6 years</td>
<td>18 children with AOM receiving amoxicillin</td>
<td>15 children with AOM not receiving amoxicillin</td>
<td>• Salivary microbiota recovers after 3 weeks, but with reduced diversity</td>
</tr>
<tr>
<td>Jervis-Bardy et al. (2015)</td>
<td>MEF, nasopharyngeal swabs, adenoid swabs</td>
<td>3-9 years</td>
<td>11 Australian Indigenous children with OME</td>
<td>NA</td>
<td>• <em>A. otitidis</em> dominant in MEF, followed by <em>H. influenzae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• <em>A. otitidis</em> not present in nasopharynx or adenoid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Diversity in MEF lower than in nasopharynx</td>
</tr>
<tr>
<td>Chan et al. (2016)</td>
<td>MEF, adenoid swabs</td>
<td>1-12 years</td>
<td>23 children with COME</td>
<td>10 children without COME undergoing other respiratory surgery</td>
<td>• <em>A. otitidis</em> dominant in MEF, followed by <em>Haemophilus</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• <em>A. otitidis</em> almost absent from adenoids</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Adenoid microbiota in cases and controls similar, case adenoid and case MEF microbiota dissimilar</td>
</tr>
<tr>
<td>Chan et al. (2017)</td>
<td>MEF and external auditory canal lavages</td>
<td>1-14 years</td>
<td>18 children with OME</td>
<td>NA</td>
<td>• <em>A. otitidis</em> dominant in MEF, followed by <em>Haemophilus</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• <em>A. otitidis</em> and <em>Staphylococcus</em> abundant and otopathogens rare in ear canal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Proposed ear canal as well as nasopharynx as reservoirs for middle ear microbiota</td>
</tr>
</tbody>
</table>
Table 1.1: (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of sample</th>
<th>Age of participants</th>
<th>Number and type of cases</th>
<th>Number and type of controls</th>
<th>Major findings</th>
</tr>
</thead>
</table>
| Sillanpaa et al. (2017)¹⁷⁷  | MEF                          | 5 months – 3.5 years | 79 children with AOM     | NA                         | • *S. pneumoniae* most prevalent organism  
                                |                              |                      |                          |                                                                                   | • *A. otitidis*, *T. otitidis* and *Staphylococcus auricularis* also observed in MEF, less commonly than otopathogens  
                                |                              |                      |                          |                                                                                   | • Suggested it is not likely that there are any other pathogens in MEF |
| Chonmaitree et al. (2017)¹⁷¹| Nasopharyngeal swabs         | < 12 months (longitudinal) | 65 children with AOM     | 74 children without AOM   | • Otopathogen colonisation associated with rates of URTI  
                                |                              |                      |                          |                                                                                   | • *Staphylococcus* and *Sphingobium* associated with decreased risk of URTI transitioning to AOM  
                                |                              |                      |                          |                                                                                   | • *Corynebacterium* reduced in infants with AOM but did not influence the URTI/AOM process |
| Krueger et al. (2017)¹⁷⁸     | MEF                          | 3 months – 14.6 years | 55 children with COME    | NA                         | • *Haemophilus* most abundant in MEF, followed by *Morazella* and *Turicella*  
                                |                              |                      |                          |                                                                                   | • *Turicella* positively associated with age > 2 years, negatively associated with hearing loss  
                                |                              |                      |                          |                                                                                   | • *Haemophilus* associated with increased mucin production |
| Boers et al. (2018)¹⁷⁹       | MEF and nasopharyngeal swabs | < 12 years          | 9 children with gastroesophageal reflux-associated OM | 21 children with OM only   | • *Alloiosoccus* and *Turicella* most commonly detected in MEF  
<pre><code>                            |                              |                      |                          |                                                                                   | • Gastroesophageal reflux had no obvious effect on microbiota |
</code></pre>
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of sample</th>
<th>Age of participants</th>
<th>Number and type of cases</th>
<th>Number and type of controls</th>
<th>Major findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man et al. (2018)¹7⁰</td>
<td>MEF (otorrhea) and nasopharyngeal swabs</td>
<td>&lt; 5 years</td>
<td>94 children with AOM + grommets</td>
<td>NA</td>
<td>• <em>M. catarrhais</em> may have a role as a commensal rather than pathogen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Higher diversity in nasopharynx than MEF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Abundance of <em>Corynebacterium</em> and <em>Dolosigranulum</em> associated with shorter duration of otorrhea</td>
</tr>
</tbody>
</table>
above. In the upper respiratory tract, genera such as *Streptococcus*, *Haemophilus*, *Moraxella* and *Staphylococcus* may contain pathogenic species as well as commensal species, so there is a depth to these microbial communities that remains poorly explored. Metataxonomics only provides an indication of the organisms from which DNA is present at the time of sampling, and it does not provide information on the role of organisms in their environment or their viability or activity. The cross-sectional nature of these studies also precludes the investigation of causal relationships between members of the microbiota and disease outcomes. Data on healthy children (without URTIs or other conditions) who have never experienced OM is also minimal in the studies that have been conducted thus far; it is difficult to establish the features of a ‘healthy’ microbiota in the context of OM without direct comparison between large groups of children with OM and healthy controls. Future research on the microbiota in children with OM requires a focus on establishing causal relationships rather than a description of relative abundance; particularly those exploring the role of commensal organisms associated with health as this may lead to novel avenues for therapy.

1.6 Probiotic therapy: an option for otitis media?

1.6.1 Probiotics

Probiotics are defined as “live microorganisms which when administered in adequate amounts confer a health benefit on the host.” While most probiotics are considered food supplements and are taken orally to benefit the gut, the definition extends to probiotics administered for other purposes and is not restricted to humans. Probiotic treatment has been reported to provide benefits to a wide range of human health problems including necrotizing enterocolitis in very low birth weight infants, antibiotic associated diarrhoea, periodontal disease, and infant sepsis.

There are several mechanisms by which probiotics can provide a beneficial effect, reviewed by Bermudez-Brito et al. (2012); though for many of the probiotic formulations commonly in use, the exact mechanism is not known. In the gut where probiotics are well-studied, probiotic bacterial species may enhance the epithelial barrier by stimulating the production of mucins and defensins; produce antimicrobial substances that inhibit pathogenic organisms; competitively exclude pathogens from
binding to host cells; compete for limited available nutrients or they may modulate
the immune system, enhancing its activity against pathogenic bacteria. The effect
of probiotics in the gut may also be enhanced by the co-administration of prebiotics;
compounds which support the growth of probiotic bacteria. Probiotic/prebiotic
combinations are known as synbiotics.

1.6.2 Probiotics derived from the microbiota

Oral probiotics containing commensal organisms that inhabit the gut have shown
substantial benefits for gut-related diseases. However, their beneficial effects may
not be apparent in the context of other diseases. Niche specificity of probiotics is
an important factor; for the probiotic bacteria to engage in some of the protective
mechanisms listed above, it is necessary for them to colonise the correct niche
to enable interference with the pathogenic organisms in that habitat. As different
species or strains of probiotic bacteria do not necessarily exhibit the same mechanisms
of action, selection of the correct organisms with regard to niche specificity and
disease specificity is vital for optimal efficacy. Additionally, a combination
of multiple organisms may be important for conferring a beneficial effect, which each
organism on its own may not be capable of producing.

Specificity may be achieved by ‘mining the microbiota for therapeutics’; the approach
where the normal, ‘healthy’ microbiota within a particular niche is explored for
microbial factors that are associated with health, providing candidate organisms for
use in probiotic therapy. This involves establishing a causal relationship between
commensals and health rather than pathogens and disease and there are several
examples of studies that have demonstrated the efficacy of this approach. Casey et al.
(2007) identified a mixture of five commensal intestinal strains from pigs which served
to prevent severe infection in the animals upon challenge with Salmonella enterica
serovar Typhimurium. Lawley et al. (2012) similarly developed a therapy for a
mouse model using six commensal intestinal strains derived from healthy mice, which
when administered to mice with experimentally-induced Clostridium difficile chronic
intestinal disease, displaced the pathogen and resolved disease. In humans, Iwase et al.
(2010) demonstrated that Staphylococcus epidermidis sourced from healthy adult
volunteers was capable of inhibiting the colonisation and formation of biofilm by S.
aureus in the nose of S. aureus-colonised volunteers. For children with OM, Roos et al.
(2001) reported the competitive activity of alpha-haemolytic streptococci isolated
from the opening of the Eustachian tube in healthy children against the otopathogen
species in vitro, and developed a nasal spray with five strains that demonstrated the ability to reduce the recurrence of AOM episodes. These examples demonstrate that niche-specific probiotic therapies have been successful in preventing or treating diseases that involve one or more pathogens causing a disruption to the local microbial community. This approach may therefore be suitable for the prevention or treatment of rAOM.

1.6.3 Probiotics in the treatment of AOM

During an episode of AOM, the otopathogens are present in the nasopharynx at high bacterial load and antibiotics are acting against any organisms that are susceptible to them, including commensal bacteria. These factors can disrupt the normal nasopharyngeal microbiota; resulting in a state of dysbiosis, or imbalance. Probiotic therapy may therefore be an effective alternative treatment for OM, and has been reported as the most promising complementary or alternative treatment for the disease. The use of probiotics to either restore the healthy microbiota that is lost through antibiotic treatment for AOM, or to colonise the nasopharynx with important nasopharyngeal commensals that may be absent prior to AOM could be an effective strategy to interrupt the recurrent nature of rAOM or to prevent its onset.

Several clinical trials have investigated the efficacy of various probiotic therapies in children with OM, summarised in Table 1.2. Some of these have utilised generic oral probiotics that are designed to influence the gut microbiota. Most of these studies administered the probiotic to healthy children and reported the incidence of AOM during follow-up; many of these studies reported no substantial reduction in the incidence of AOM. There are four randomised trials that have utilised a probiotic nasal spray for the treatment of OM, also summarised in Table 1.2. Each of these was carried out in a cohort of otitis-prone children and investigated the recurrence of AOM after the local probiotic therapy. Three of these studies reported a significant beneficial effect, with one noting that successful colonisation of the nasopharynx with the probiotic strain was essential for the reduction of AOM. In the study where a significant effect was not observed, it was suggested that pre-treatment with antibiotics prior to probiotic administration may assist in the re-establishment of a healthy microbiota by impeding the recolonization or re-emergence of the otopathogen species. Overall, these studies suggest that oral administration of gastrointestinal
Table 1.2: Randomised controlled trials investigating the efficacy of probiotic therapies for otitis media in children.

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects completing study</th>
<th>Age of subjects</th>
<th>Probiotic treatment</th>
<th>Outcome of trial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oral administration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hatakka et al. (2001)</td>
<td>513 healthy children</td>
<td>1-6 years</td>
<td>Lactobacillus rhamnosus GG</td>
<td>Non-significant effect on incidence of AOM.</td>
</tr>
<tr>
<td>Hatakka et al. (2007)</td>
<td>269 otitis-prone children</td>
<td>10 months – 6 years</td>
<td>L. rhamnosus GG, L. rhamnosus LC 705, Bifidobacterium breve 99 and Propionibacterium freudenreichii JS</td>
<td>Non-significant effect on incidence of AOM, recurrence of AOM, or nasopharyngeal carriage of otopathogen species.</td>
</tr>
<tr>
<td>Rautava et al. (2008)</td>
<td>72 healthy infants</td>
<td>&lt; 2 months</td>
<td>L. rhamnosus GG and Bifidobacterium lactis Bb-12</td>
<td>Significant reduction in incidence of AOM.</td>
</tr>
<tr>
<td>Stecksén Blicks et al. (2009)</td>
<td>186 healthy children attending day care</td>
<td>1-5 years</td>
<td>L. rhamnosus LB21 with fluoride</td>
<td>Significant reduction in days with OM and days on antibiotics.</td>
</tr>
<tr>
<td>Taipale et al. (2011)</td>
<td>69 healthy infants</td>
<td>1-2 months</td>
<td>B. lactis Bb-12</td>
<td>Non-significant effect on incidence of AOM.</td>
</tr>
<tr>
<td>Cohen et al. (2013)</td>
<td>166 healthy infants</td>
<td>7-13 months</td>
<td>Streptococcus thermophilus, Streptococcus salivarius, and L. rhamnosus with prebiotic</td>
<td>Non-significant effect on incidence of AOM.</td>
</tr>
<tr>
<td>Tapiovaara et al. (2014)</td>
<td>31 children undergoing tympanostomy for OM</td>
<td>1-5 years</td>
<td>L. rhamnosus GG</td>
<td>Non-significant effect on the presence of otopathogens in the middle ear.</td>
</tr>
<tr>
<td>Di Pierro et al. (2016)</td>
<td>222 healthy children</td>
<td>3 years</td>
<td>S. salivarius K12</td>
<td>Significant reduction in the incidence of AOM.</td>
</tr>
<tr>
<td><strong>Nasal administration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roos et al. (2001)</td>
<td>108 otitis-prone children</td>
<td>6 months – 6 years</td>
<td>Streptococcus mitis and Streptococcus sanguinis (five strains derived from healthy children)</td>
<td>Significant reduction in the recurrence of AOM.</td>
</tr>
<tr>
<td>Tano et al. (2002)</td>
<td>36 children with rAOM</td>
<td>&lt; 4 years</td>
<td>S. sanguinis, S. mitis and Streptococcus oralis (five strains derived from healthy children)</td>
<td>Non-significant effect on the recurrence of AOM.</td>
</tr>
</tbody>
</table>
### Table 1.2: (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects completing study</th>
<th>Age of subjects</th>
<th>Probiotic treatment</th>
<th>Outcome of trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skovbjerg <em>et al.</em> (2009)&lt;sup&gt;210&lt;/sup&gt;</td>
<td>54 children with OME</td>
<td>1-8 years</td>
<td><em>S. sanguinis</em> or <em>L. rhamnosus</em></td>
<td><strong>Significant</strong> increase in the resolution of OME for both probiotic species; effect more substantial with <em>S. sanguinis</em>. \</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Significant</strong> reduction in the recurrence of AOM and antibiotic use; effect more substantial in children where <em>S. salivarius</em> successfully colonised the nasopharynx.</td>
</tr>
<tr>
<td>Marchisio <em>et al.</em> (2015)&lt;sup&gt;211&lt;/sup&gt;</td>
<td>97 children with rAOM</td>
<td>1-5 years</td>
<td><em>S. salivarius</em> 24 SMB</td>
<td></td>
</tr>
</tbody>
</table>
probiotic strains is not as effective as nasal administration of upper respiratory tract strains, and that probiotics may be more effective in preventing the recurrence of AOM in otitis-prone children than preventing initial AOM episodes in healthy children. Reviews summarising the use of probiotic therapy in children with OM have generally concluded that while the treatment is promising, further data is required before it can be recommended in practice. Further development of niche-specific, disease-specific probiotic therapy for rAOM requires thorough knowledge on the nasopharyngeal microbiota of healthy, rAOM-resistant children and the members of the microbiota which are contributing to health in the upper respiratory tract.

1.7 Aims of this thesis

To facilitate the development of novel therapies for the treatment of rAOM in children, it is necessary to gain a more detailed understanding of the microbial communities inhabiting the upper respiratory tract in children with rAOM, as well as those in children that appear to be resistant to rAOM. ‘Resistant’ healthy children are defined in this thesis as those who have been exposed to major environmental risk factors for rAOM (i.e. attendance at day care or living with siblings) but have experienced no more than one episode of AOM in their lifetime. It is important to identify and define the role of other organisms from the middle ear in the pathogenesis of rAOM, as they may be future targets for therapy. It is also important to identify the key members of the rAOM-resistant nasopharyngeal microbiota and investigate their role in contributing to health, as these organisms or their products may be candidates for the development of niche-specific and disease-specific probiotic therapies. Advances in next-generation sequencing techniques allow these avenues to be explored at a greater scale and in greater detail than has previously been achieved.

This thesis describes the work undertaken in the Perth Otitis Media Microbiome (biOMe) Study and a subsequent pilot study. The main aims of this thesis are as follows:

1. To use 16S rRNA gene sequencing (metataxonomics) to characterise the genus-level differences between the nasopharyngeal microbiota of children with rAOM compared to those that are resistant to developing rAOM. This methodology will additionally be used to explore the middle ear microbiota of children with rAOM for novel otopathogens. (Chapter 2)

2. To use whole genome shotgun sequencing (metagenomics) to identify the
bacterial species contributing to the differences between microbial communities determined in Aim 1, and to explore whether there are differences in the functional capability of these microbiota. (Chapter 3)

3. Utilising the information from Aims 1 and 2, to identify commensal bacterial species that may contribute to resistance to rAOM and test these species in vitro to assess their capacity to interfere with the survival of otopathogen species. (Chapter 4)

4. To conduct a pilot study using dual RNA-seq to characterise the host transcriptome in the middle ear in OM and the microbial metatranscriptome, to investigate the viability and activity of potential novel otopathogens in the middle ear. (Chapter 5)
Chapter 2

Characterising the microbiota with 16S rRNA gene sequencing

This chapter describes the first aspect of the Perth Otitis Media Microbiome (biOMe) study, which aimed to describe the microbial communities of the nasopharynx in children with and without rAOM, and of the middle ear in children with rAOM. The nasopharyngeal microbiomes of the two groups of children were compared to identify candidate organisms associated with health that could be useful for development into a probiotic therapy for rAOM. The middle ear and nasopharyngeal microbiomes of children with rAOM were compared to identify potential novel otopathogens.

This work was published in BMC Microbiology, volume 18, issue 13 (online, https://doi.org/10.1186/s12866-018-1154-3). A copy of the published article can be found in Appendix A.1 with the associated supplementary information. The content of this chapter is identical to the published paper with the exception of some minor revisions.

Authors: Rachael Lappan, Kara Imbrogno, Chisha Sikazwe, Denise Anderson, Danny Mok, Harvey Coates, Shyan Vijayasekaran, Paul Bumbak, Christopher C. Blyth, Sarra E. Jamieson and Christopher S. Peacock.
Chapter 2. Characterising the microbiota with 16S rRNA gene sequencing

2.1 Abstract

Background

Recurrent acute otitis media (rAOM, recurrent ear infection) is a common childhood disease caused by bacteria termed otopathogens, for which current treatments have limited effectiveness. Generic probiotic therapies have shown promise, but seem to lack specificity. We hypothesised that healthy children with no history of AOM carry protective commensal bacteria that could be translated into a specific probiotic therapy to break the cycle of re-infection. We characterised the nasopharyngeal microbiome of these children (controls) in comparison to children with rAOM (cases) to identify potentially protective bacteria. As some children with rAOM do not appear to carry any of the known otopathogens, we also hypothesised that characterisation of the middle ear microbiome could identify novel otopathogens, which may also guide the development of more effective therapies.

Results

Middle ear fluids, middle ear rinses and ear canal swabs from the cases and nasopharyngeal swabs from both groups underwent 16S rRNA gene sequencing. The nasopharyngeal microbiomes of cases and controls were distinct. We observed a significantly higher abundance of *Corynebacterium* and *Dolosigranulum* in the nasopharynx of controls. *Alloiococcus, Staphylococcus* and *Turicella* were abundant in the middle ear and ear canal of cases, but were uncommon in the nasopharynx of both groups. *Gemella* and *Neisseria* were characteristic of the case nasopharynx, but were not prevalent in the middle ear.

Conclusions

*Corynebacterium* and *Dolosigranulum* are characteristic of a healthy nasopharyngeal microbiome. *Alloiococcus, Staphylococcus* and *Turicella* are possible novel otopathogens, though their rarity in the nasopharynx and prevalence in the ear canal means that their role as normal aural flora cannot be ruled out. *Gemella* and *Neisseria* are unlikely to be novel otopathogens as they do not appear to colonise the middle ear in children with rAOM.
2.2 Background

Otitis media (OM) refers to a group of inflammatory conditions of the middle ear and is commonly seen in young children. The disease can be divided into two broad categories; acute OM (AOM) and OM with effusion (OME). AOM involves signs of active infection including fever and irritability, and the middle ear contains purulent fluid with a bulging tympanic membrane. OME is characterised by non-purulent effusion and no signs of acute infection. Recurrent AOM (rAOM), defined as 3 or more episodes of AOM within 6 months; or 4 or more in 12 months is also common, with 17% of children experiencing at least 3 episodes before the age of 1 year. Children with rAOM are commonly prescribed repeated courses of antibiotics, and are often referred to an ear, nose and throat (ENT) surgeon for insertion of ventilation tubes (grommets) to prevent rupturing of the tympanic membrane. In Australia, 73% of children under 12 months of age will have experienced OM at least once, with costs to the health care system estimated at $100 to $400 million AUD.

Bacterial pathogens that cause AOM are referred to as otopathogens. The three bacterial species widely recognised as otopathogens are *Streptococcus pneumoniae*, non-typeable *Haemophilus influenzae* and *Moraxella catarrhalis*, which are thought to originate from the nasopharynx and are capable of migrating to the middle ear and persisting within biofilms. Children with AOM are commonly colonised with multiple otopathogens, and coinfection with respiratory viruses is common. However, a proportion of children with AOM do not appear to be colonised with any of the known otopathogens, implying that there may be other bacteria involved in AOM. *Alloiococcus otitidis* and *Turicella otitidis* have been associated with OM, but their role in the pathogenesis of AOM remains unknown. Antibiotic treatment has been shown to be of limited benefit for AOM, and all three otopathogens have exhibited resistance to the antibiotics commonly used to treat it. Additionally, one in five children fitted with grommets will require reinsertion in the future. Considering the limited effectiveness of current treatments, there is a need for alternative therapies for rAOM. Probiotic treatment for rAOM is one alternative that has been investigated in several clinical trials. Probiotics are defined as “live microorganisms, which when administered in adequate amounts, confer a health benefit on the host” and act via mechanisms such as competition for nutrients, stimulation of the immune system and direct inhibition with antibacterial molecules. The effect of probiotics on the recurrence of AOM has been
variable, with some studies showing a significant improvement\textsuperscript{198,205,206,210,211,219} and others reporting no change\textsuperscript{201–204,209,220} with nasally delivered probiotics appearing to perform better than orally delivered probiotics. The contradictory results of these studies highlight the importance for the development of probiotics containing bacterial strains that are relevant to the upper respiratory tract environment, active against otopathogens and able to colonise the nasopharynx.

A relatively new area of microbiome research involves the identification of commensal bacteria or bacterial products from a ‘healthy’ microbiome for therapeutic use. Therapies identified by this approach have been successful in resolving relapsing \textit{Clostridium difficile} disease in mice,\textsuperscript{191} resolving \textit{Salmonella enterica} serovar Typhimurium disease in pigs\textsuperscript{196} and inhibiting colonisation and biofilm formation of \textit{Staphylococcus aureus} in the nasopharynx of adult humans.\textsuperscript{197} These studies demonstrate that the microbiome can be a source for effective probiotic treatments, which may be a single species that acts specifically against the pathogen of interest, or a combination of multiple commensal species.

We had two hypotheses in this study. Firstly, we hypothesised that there are bacterial pathogens involved in rAOM other than the known otopathogens. We sought to identify potential novel otopathogens by characterising for the first time the microbiome of the middle ear in children with rAOM and comparing it to their nasopharyngeal microbiome. Secondly, understanding that exposure to other children via attendance at day care or the presence of multiple children in the home is a major risk factor for rAOM,\textsuperscript{52,221} we hypothesised that children exposed to this risk factor but who have not developed rAOM are carrying nasopharyngeal bacteria that provide protection against the disease. We aimed to identify potentially protective commensal bacteria, which may be of use as a specific probiotic therapy for rAOM, by comparing the nasopharyngeal microbiomes of these rAOM-resistant children with those of children with rAOM.

\section{2.3 Methods}

\subsection{2.3.1 Patient recruitment}

Children under the age of 5 years were recruited into either the case (rAOM-prone) or control (rAOM-resistant) group of the Perth Otitis Media Microbiome (biOMe) study in the Perth metropolitan area of Western Australia from December 2013 to
2.3. Methods

December 2015. Cases were undergoing grommet insertion for physician-diagnosed rAOM and were identified as eligible for inclusion by their ENT surgeon. Children undergoing grommet insertion for OME were excluded from recruitment. Middle ear fluid (MEF) from each ear and a nasopharyngeal swab (NPS) were collected at the time of surgery. A saline middle ear rinse (MER) from each ear and one ear canal swab (ECS) were also collected from a subset of patients (56/93). Healthy controls with no history of rAOM were recruited from a community immunisation clinic. Controls were attending day care or had a sibling up to 5 years of age at the time of collection (i.e. were exposed to previously described major risk factors for rAOM\textsuperscript{52,221}). A NPS sample was collected from controls. A questionnaire on demographics, risk factors and recent antibiotic use was completed for all case and control subjects (see Appendix A.2). Controls were matched to cases by age (within 3 months if case less than 1 year of age; within 6 months if case between 1-2 years of age; within 12 months if case between 2-5 years of age) and season (within 2 weeks of collection time; to the previous year if no match found in recruitment year). Subjects were matched by sex where possible. Exclusion criteria for both groups included diagnosis of cleft lip or palate, immune deficiency or genetic syndrome. All specimens and questionnaire data were obtained with informed written consent from a parent or guardian. Recruitment to the study was approved by the Human Research Ethics Committees (HREC) at Princess Margaret Hospital for Children (2013119/EP), St John of God Health Care (#708) and the University of Western Australia (RA/4/1/6839) as well as by all relevant hospital governance committees.

2.3.2 Sample collection and storage

All specimens obtained from the cases were collected at the time of grommet surgery by the performing surgeon. ECS specimens were taken from one ear prior to myringotomy (incision of the tympanic membrane) with a sterile FLOQswab (Copan) and were placed in 1 ml skim milk tryptone glucose glycerol broth (STGGB, PathWest). Following myringotomy, MEF specimens from each ear were aspirated into a sterile Argyle\textsuperscript{TM} specimen trap (Covidien) with 2 ml of sterile saline used to flush out the tubing. MER specimens were collected from each ear after aspiration of MEF whereby 2 ml of sterile saline was injected into the middle ear, then aspirated into an Argyle\textsuperscript{TM} trap. A NPS was taken with a sterile FLOQswab, rotating for at least 3 seconds at the nasopharynx before transferral into 1 ml of STGGB. Specimens from the cases were immediately frozen on dry ice or kept on wet ice and transported to the laboratory on the same day. NPS specimens were collected from controls in the
same manner and kept on wet ice until transport to the lab. All specimens were frozen at -80°C until DNA was extracted.

2.3.3 DNA extraction and sequencing preparation

Swab samples (NPS and ECS) were first prepared by vortexing followed by transferral of the swab, inverted, to a new sterile tube with sterile forceps. This was centrifuged to collect mucus attached to the swab which was then transferred back into the milk broth. All samples were then aliquoted for DNA extraction (500 µl for ECS and NPS, 750 µl for MEF and MER) and each MEF and NPS specimen was also aliquoted for viral typing (200 µl). The remainder of all samples were archived at -80°C. DNA was extracted with the Wizard SV Genomic DNA Purification System (Promega) and FastPrep Lysing Matrix B tubes (MP Biomedicals) as described in Teo et al. (2015) with some modifications. In brief, extractions were carried out inside a class II biohazard hood with UV-sterilised plastics and pipettes wiped with DNA Away (Molecular BioProducts) to minimise contamination. A negative extraction control (reagents with no specimen) was included in each extraction batch and each batch included samples of each type. DNA extraction aliquots were then processed as previously described, with the final purified genomic DNA stored in DNA Lo-Bind tubes (Eppendorf). Samples were quantified by the Qubit 2.0 fluorometer (dsDNA HS assay, Invitrogen) and diluted to 5 ng/µl with low TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, Fisher Biotec). All genomic DNA was frozen at -80°C until sequencing preparation.

2.3.4 Viral detection

DNA from the viral typing aliquots for all MEF and NPS specimens (cases and controls) was extracted using an automatic extraction platform (MagMAX express 96) according to manufacturer’s instructions. These samples were screened for 19 common respiratory viruses (see Table 2.1) using a routine multiplex PCR at PathWest (Perth, Western Australia). This method was developed and described by Chidlow et al. (2009).
### Table 2.1: Gene targets for multiplex respiratory virus PCR.
Where multiple strains of a virus were detected, results were combined.

<table>
<thead>
<tr>
<th>Virus name (strains targeted)</th>
<th>Abbreviation</th>
<th>PCR target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Adenovirus</td>
<td>HAdV</td>
<td>Hexon gene</td>
</tr>
<tr>
<td>Human Bocavirus</td>
<td>HBoV</td>
<td>VP1 gene</td>
</tr>
<tr>
<td>Influenza virus (A/B/C)</td>
<td>IFV</td>
<td>Haemagglutinin &amp; Matrix gene</td>
</tr>
<tr>
<td>Respiratory Syncytial Virus (Type A/Type B)</td>
<td>RSV</td>
<td>Nucleoprotein</td>
</tr>
<tr>
<td>Human Metapneumovirus</td>
<td>HMPV</td>
<td>Nucleoprotein</td>
</tr>
<tr>
<td>Human Coronavirus (OC43/229E/HKU1/NL63)</td>
<td>HCoV</td>
<td>Nucleocapsid (OC43, 229E, NL63)</td>
</tr>
<tr>
<td>Parainfluenza virus (1/2/3/4)</td>
<td>HPIV</td>
<td>Nucleoprotein</td>
</tr>
<tr>
<td>Rhinovirus (A/B/C)*</td>
<td>RV</td>
<td>5'UTR</td>
</tr>
</tbody>
</table>

*a Rhinovirus typing used primer pairs reported by Lee et al.\textsuperscript{224}
2.3.5 Positive sequencing control generation

A mock community of 16 bacterial species (MOCK) was used as a positive sequencing control (see Table 2.2). This included a mixture of Gram positive and Gram negative organisms, including some with known prevalence in the upper respiratory tract and some not expected to be found in this environment. Genomic DNA was extracted from glycerol stocks of each species using the Wizard SV Genomic DNA Purification System, with the exception of *N. meningitidis* which was obtained as a heat-killed stock. Each species separately underwent PCR amplification and AMPure XP purification (described in the following section). The PCR products were quantified by Qubit fluorometry, diluted to equal concentrations and pooled at equal volumes to create the positive sequencing control that was included on each sequencing run. The theoretical expected relative abundance is 6.25% for each species. The positive sequencing control was aliquoted before freezing at -80°C to minimise freeze/thaw effects on the DNA. Each sequencing run utilised a separate aliquot, in duplicate.

2.3.6 Amplicon sequencing

Samples were prepared for amplicon sequencing following the Illumina protocol for 16S rRNA gene sequencing (Part #15044223, Rev. B) with modification to ensure sufficient amplification from samples that yielded low amounts of DNA. The recommended primers (forward: 5’CCTACGGGNGGCWGCAG, reverse: 5’GACTACHVGGGTATCTAATCC) target the V3/V4 region of the 16S rRNA gene with Illumina adapters attached to the 5’ end. The expected length of this targeted region is approximately 465 bp including the amplicon primers. The PCR reaction mix contained 9.5 µl of genomic DNA with a final concentration of 300 nM for each amplicon primer and 1x KAPA HiFi HotStart ReadyMix (KAPA Biosystems). A negative extraction control and a no-template negative PCR control were included on each PCR plate. PCR cycling conditions were 95°C for 3 minutes, 30 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds before a final 72°C for 5 minutes and holding at 4°C. All PCR products were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter) as described in the Illumina protocol. The samples and mock community aliquots were then barcoded by Illumina’s dual indexing strategy (Nextera XT Index Kit v2, Sets A and B, Illumina) as described in the Illumina protocol using the default barcode layout from the Illumina Experiment Manager software v1.11.0. Samples were sent to the Australian Genome Research
Table 2.2: Species included in the positive sequencing control.
All cultures were obtained from the University of Western Australia’s School of Biomedical Sciences culture collection, with the exception of *N. meningitidis* which was kindly provided by A/Prof Charlene Kahler (UWA) and originally described in Stephens *et al.* (1991).\(^{225}\)

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain/ATCC/NCTC no.</th>
<th>Gram</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>ATCC 29213</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Staphylococcus epidermidis</em></td>
<td>ATCC 14990</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Streptococcus pneumoniae</em></td>
<td>Strain D39</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Staphylococcus warneri</em></td>
<td>ATCC 27836</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Moraxella catharralis</em></td>
<td>ATCC 25238</td>
<td>Negative</td>
</tr>
<tr>
<td>Non-typeable <em>Haemophilus influenzae</em></td>
<td>Strain 86-028NP</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Haemophilus haemolyticus</em></td>
<td>ATCC 33390</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em></td>
<td>Strain M7</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Corynebacterium jeikeium</em></td>
<td>ATCC 43216</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Propionibacterium acnes</em></td>
<td>ATCC 6919</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Gemella haemolysans</em>(^a)</td>
<td>NCTC 10244</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em></td>
<td>NCTC 8172</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
<td>ATCC 15692</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Streptococcus salivarrius</em></td>
<td>not available</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Veillonella parvula</em></td>
<td>ATCC 10790</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Alloioococcus otitidis</em></td>
<td>ATCC 51267</td>
<td>Positive</td>
</tr>
</tbody>
</table>

\(^a\) This species may have been mislabelled as it was identified as *Globicatella* by SILVA (v123) taxonomy.\(^{226,227}\)
Facility where they underwent further purification and quality checking followed by dilution and equimolar pooling of all samples at 1 ng/µl. The final pool underwent a band excision with the QIAquick Gel Extraction kit (Qiagen) to select the V3/V4 band (approximately 600 bp with adapters and indexes) for sequencing. The pool was sequenced on the Illumina MiSeq with 2 x 300 bp V3 chemistry and a spike-in of 20% PhiX control. A total of 581 samples including all positive and negative controls were sequenced across four separate sequencing runs of 140-149 samples each. Samples of all types were included on each sequencing run.

2.3.7 Preprocessing of sequence data

A diagrammatic overview of the entire analysis pipeline is provided in Appendix A.3. Demultiplexed paired-end reads in FASTQ format were received from the sequencing centre. Overall run quality was observed with FastQC v0.11.3. The sequence data was processed via the UPARSE pipeline (using the USEARCH v8.1.1861 algorithm) and QIIME v1.9.1. Paired-end reads were merged with USEARCH with maximum expected errors set to 1 and length restricted to 440-470 bp. Amplicon primer sequences were removed from these high quality reads using a custom script before conversion to FASTA format with QIIME, concatenation of the data from all four sequencing runs and dereplication (collapsing into unique sequences) with USEARCH. Reads that aligned to the human genome (GRCh38_p2) were removed with Deconseq v0.4.3 using identity threshold 94% and coverage threshold 90%. Remaining reads were clustered into OTUs (operational taxonomic units) at a 97% identity threshold with USEARCH to create a high quality chimera-filtered representative set of OTU sequences. The original raw paired-end reads were then merged as before but without expected error filtering. This set of reads was then aligned to the high quality OTU representative set with USEARCH to assign an OTU to each sequence (97% identity threshold) and create an OTU table. From this point, the sequences were processed within QIIME. Taxonomy was assigned with UCLUST v1.2.22q at 90% identity with the SILVA database v123 and then aligned to the SILVA core alignment with PyNAST v1.2.2. We selected the SILVA database over the QIIME default database (GreenGenes release 13_8) as Dolosigranulum is not present in the 13_8 release, and is misclassified as Alloiococcus. The alignment was filtered and a phylogenetic tree generated with FastTree v2.1.3. OTUs that failed to align and those with an abundance below 0.005% were removed from the OTU table. Samples with low sequencing depth (77 out of 581 total samples; threshold 1499 reads) were also removed from the table. This OTU table was used
for all downstream analysis.

Common practice is to then rarefy this data, which involves subsampling to an even number of sequences per sample to render them comparable. However, this can discard a large volume of usable data and reduces statistical power.\textsuperscript{237} We have avoided the statistically inadmissible\textsuperscript{237} practice of rarefying throughout this chapter.

### 2.3.8 Covariate and diversity analyses

Data from completed demographic questionnaires included day care attendance, duration of breastfeeding, presence of siblings and smokers in the household, relevant clinical diagnoses (i.e. asthma, allergy and chest, heart or kidney problems), previous admission to hospital for infection and use of antibiotics in the previous month. Statistical analysis of covariates (including detection of respiratory viruses) was done in R v3.3.2\textsuperscript{238} using the Wilcoxon rank sum test for continuous data and Pearson’s Chi-squared test for categorical data (unless any values were below 5, then Fisher’s exact test was used). Odds ratios and associated confidence intervals and $p$-values for the viral detection data were calculated in R with the exact2x2 function in the exact2x2 package.\textsuperscript{239} All subjects who contributed at least one analysable sample were included in the analysis of covariates.

Summaries of taxonomic relative abundance were generated with QIIME. Correlation of taxa summaries in QIIME was done in either expected mode (all samples against one theoretical expected mock community) or paired mode (pairs of samples) using the Pearson correlation coefficient with a two-sided permutation test (999 permutations) to calculate a non-parametric $p$-value.

Alpha (within-sample) diversity was measured using Faith’s Phylogenetic Diversity (PD)\textsuperscript{240} and inverse Simpson (IS)\textsuperscript{241} indices, calculated per sample in QIIME on a cumulative-sum-scale (CSS) normalised and logged OTU table (see differential abundance method for CSS normalisation). PD is a phylogenetic measure that reflects how much of the phylogenetic tree is covered by the OTUs found in the samples. The IS index is a non-phylogenetic measure that takes into account the richness (number of OTUs) and evenness (relative abundance of OTUs) in the sample, with the reciprocal taken so that larger values represent greater diversity. Median values and statistical tests between groups of samples were calculated in R using the Wilcoxon rank sum test for unpaired samples, and Wilcoxon signed rank for paired samples. Alpha diversity boxplots were generated in R with ggplot2 v2.2.0\textsuperscript{242} and
Beta diversity (between-sample diversity) was calculated using the weighted UniFrac metric in QIIME on the raw OTU table. McMurdie and Holmes (2014) demonstrated that weighted UniFrac distances are accurate on raw (unrarefied) data. Two-dimensional principal coordinates analysis (PCoA) plots were also generated in QIIME. Procrustes analysis was used to determine the similarity of the PCoA between pairs of samples collected or treated in two different ways. Comparisons were made between raw and rarefied (subsampling) data, MEF and MER sample types from the same ear, and left and right ears from the same child. The analysis calculates an $m^2$ value which describes how similar the paired datasets are, with a smaller number indicating the datasets are more similar. The associated non-parametric $p$-value describes the chance of seeing an $m^2$ value at least this extreme in 999 permutations using Monte Carlo simulations. A $p$-value of 0 means that no value as extreme as the calculated $m^2$ was observed in these permutations. This was carried out within QIIME and plotted with Emperor.

### 2.3.9 Co-occurrence analyses

To identify positive and negative correlations between individual OTUs in each sample type, we used SparCC, developed specifically for quantifying correlations in compositional microbiome data. Correlations between OTUs were determined separately for each sample type (case NPS, control NPS, MEF, MER) with only one sample per child chosen at random if more than one was available. OTUs that were represented by less than 2 reads per sample on average were removed from the raw OTU table before calculating correlations. The default settings were used and one-sided pseudo $p$-values were calculated (taking into account the direction of correlation) using 100 simulated datasets. For the MEF and MER, where one ear per child was selected at random, the analysis was repeated on a second set of samples with the opposite ear. Correlations with a significant ($<0.05$) $p$-value in both sets of samples were considered true positives. Values from the first set are reported for correlations that were validated in this way.

### 2.3.10 Differential abundance analyses

Differential abundance of OTUs was assessed using the fitZIG function in the R package metagenomeSeq v1.18.0 in R v3.4.1. fitZIG fits a zero-inflated Gaussian
mixture model to test for differentially abundant OTUs between groups, and is also designed for use with microbiome data. Five models were fitted to compare the abundance of OTUs between sample types (case/control NPS, MEF/MER, MEF/NPS, MER/NPS and ECS/MEF). In all models, OTUs that were not present in at least 25% of the samples in that model were filtered from the analysis to reduce false positives. The data were normalised using metagenomeSeq’s CSS normalisation. The models were fitted as follows:

Model 1 (case NPS/control NPS)

The model included sex (Male/Female), recent antibiotic usage (within the past month; Yes/No) and length of breastfeeding (Never/Under 6 months/6 to 12 months/Over 12 months/Current) to control for these potential confounders. All nasopharyngeal swabs from both groups were included in the analysis except those with missing data for the above covariates (n = 4), leaving 98 control NPS and 86 case NPS for comparison.

Models 2-5 (MEF/MER, MEF/NPS, MER/NPS, ECS/MEF)

These models included only samples from the cases and were analysed as within-child pairs. Where multiple samples were available (both left and right ear), one was selected at random (sample Set 1). A second set (sample Set 2) containing the sample from the opposite ear was used for validation (i.e. differentially abundant OTUs with an adjusted $p$-value $\leq 0.05$ in only one of the two sets were considered false positives). For the MEF/MER and ECS/MEF comparisons, the pair included samples from the same ear of each child. Subject ID was a covariate in each of the four models. Included in the analysis were 50 pairs of MEF/MER samples, 75 pairs of MEF/NPS samples, 54 pairs of MER/NPS samples and 33 pairs of ECS/MEF samples.

OTUs whose abundance was significantly different between the two groups compared (FDR-adjusted $p$-value $\leq 0.05$) were retained if their median or mean relative abundance was at least 0.35% in one of the two groups compared. We chose this threshold as it discarded known environmental contaminants abundant in the negative controls (e.g. *Delftia, Lysinibacillus*) but retained low-abundance OTUs not found in the negative controls that are known upper respiratory tract colonisers (e.g. *Veillonella*). Heatmaps representing log CSS normalised OTU counts in the compared
groups were created with superheat v0.1.0. Model coefficients (log$_2$ fold-changes; logFC) and $p$-values were derived from the MRcoefs function in metagenomeSeq.

2.4 Results

A total of 93 cases and 103 controls were included in the study. Samples from all children underwent respiratory viral typing and 16S rRNA gene sequencing, though some samples were excluded from the sequencing analyses due to low sequencing depth. NPS samples from 87 cases and 101 controls were available for comparison between the groups.

2.4.1 Study population characteristics

In our cohort, we assessed several environmental and clinical variables that may influence the risk of OM (see Table 2.3). The most significant difference observed between cases and controls was recent antibiotic usage, which was significantly more common in the cases. The length of breastfeeding and the presence of any other chronic illness were also significantly different between the two groups; the cases were breastfed for a shorter median time (5 months compared to 10.3 months) and had a higher incidence of chronic illness than the controls. Controls were not significantly different to cases in terms of age, season of collection or sex confirming successful matching of controls to cases at recruitment. We did note a higher proportion of males than females with rAOM in the case group.

2.4.2 Bacterial taxa identified across the samples

A total of 31.7 million raw reads were generated across the four sequencing runs. After read pre-processing and OTU picking, the reads were clustered into a total of 123 OTUs. The full taxonomy of these OTUs is provided in Appendix A.4, and the results from the positive and negative sequencing controls can be found in Appendix A.5. Table 2.4 shows the aggregated relative abundance of genus-level taxa from all samples excluding those with a low read count (less than 1499 reads) and sequencing controls. *Moraxella*, *Haemophilus* and *Streptococcus* were abundant in the nasopharynx of the cases and healthy controls, but the control samples contained *Dolosigranulum* and *Corynebacterium* at higher abundance than the cases. The cases
Table 2.3: Demographic characteristics of children recruited to the study. This table includes all children who contributed at least one sample to any analysis. *P*-values were calculated by Wilcoxon rank-sum test for continuous data, and Pearson’s chi-squared test for categorical data (unless any values were less than 5, then Fisher’s exact test was used).

<table>
<thead>
<tr>
<th></th>
<th>Case (n=93)</th>
<th>Control (n=103)</th>
<th>Missing data</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.9 years (1.3 – 2.8 years)</td>
<td>1.6 years (1.5 – 3.2 years)</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>58 (62.4%)</td>
<td>53 (51.5%)</td>
<td>0</td>
<td>0.12</td>
</tr>
<tr>
<td>Male</td>
<td>35 (37.6%)</td>
<td>50 (48.5%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Aboriginal or Torres Strait Islander</td>
<td>2 (2.1%)</td>
<td>1 (0.97%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 (7.5%)</td>
<td>4 (3.9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Season</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>20 (21.5%)</td>
<td>16 (15.5%)</td>
<td>0</td>
<td>0.39</td>
</tr>
<tr>
<td>Autumn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>45 (48.4%)</td>
<td>53 (51.5%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>21 (22.6%)</td>
<td>30 (29.1%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Breastfeeding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever breastfed (yes/no)</td>
<td>83 (89.2%)</td>
<td>98 (95.1%)</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>Duration of breastfeeding Median (IQR)</td>
<td>5 months (1.6 – 10 months)</td>
<td>10 months (6 – 13 months)</td>
<td>1 case (1.1%)</td>
<td>6.4 x 10^{-5}</td>
</tr>
<tr>
<td>Currently breastfeeding</td>
<td>7 (7.5%)</td>
<td>14 (13.6%)</td>
<td>0</td>
<td>0.17</td>
</tr>
<tr>
<td>over 12 months</td>
<td>10 (10.8%)</td>
<td>23 (22.3%)</td>
<td>1 control (1.6%)</td>
<td></td>
</tr>
<tr>
<td>&lt; 6 months (excluding never)</td>
<td>31 (33.3%)</td>
<td>45 (43.7%)</td>
<td>0.00077</td>
<td></td>
</tr>
<tr>
<td>never</td>
<td>9 (9.7%)</td>
<td>4 (3.9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Day care or school attendance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currently attending</td>
<td>74 (79.6%)</td>
<td>80 (77.7%)</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>Days of day care or school per week Median (IQR)</td>
<td>2.5 days (2 - 3 days)</td>
<td>2 days (1.75 – 3 days)</td>
<td>1 control (1.0%)</td>
<td></td>
</tr>
<tr>
<td>no day care or school</td>
<td>9 (20.4%)</td>
<td>22 (21.4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 2 days/week</td>
<td>13 (14.0%)</td>
<td>20 (19.4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-3 days/week</td>
<td>52 (55.9%)</td>
<td>46 (44.7%)</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>&gt; 3 days/week</td>
<td>9 (9.7%)</td>
<td>14 (13.6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at day care/school start Median (IQR)</td>
<td>13.5 months (10 months – 1.9 years)</td>
<td>12 months (10 months – 1.5 years)</td>
<td>34 controls (33%)</td>
<td></td>
</tr>
<tr>
<td>Siblings of 5 years of age or younger</td>
<td>46 (49.5%)</td>
<td>70 (68.0%)</td>
<td>1 case (1.1%)</td>
<td></td>
</tr>
<tr>
<td>Smoker in the householdf</td>
<td>15 (16.1%)</td>
<td>8 (7.7%)</td>
<td>0</td>
<td>0.607</td>
</tr>
<tr>
<td>Antibiotic usage in the past month Median (IQR)</td>
<td>61 (65.6%)</td>
<td>11 (10.7%)</td>
<td>1 case (1.1%)</td>
<td>1.9 x 10^{-5}</td>
</tr>
<tr>
<td>Any chronic illnessd</td>
<td>18 (19.4%)</td>
<td>10 (9.7%)</td>
<td>0</td>
<td>0.026</td>
</tr>
<tr>
<td>Ever admitted to hospital for infection 17 (18.3%)</td>
<td>11 (10.7%)</td>
<td>0</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

* Season was categorised by months.

b Median values exclude children who are currently breastfeeding as breastfeeding has not ceased.

c The age at which the child started day care or school; whichever they are currently attending.

d Any chronic respiratory, cardiovascular or renal illness including asthma and allergies; or other chronic illness identified by the parents.
additionally contained *Neisseria* (10.9%), *Gemella* (3.2%), *Porphyromonas* (3.6%), *Alloprevotella* (2.4%) and *Fusobacterium* (2.7%) where these were almost absent in the controls (each below 0.3%). Of the three major otopathogen genera, *Haemophilus* was the most abundant in the middle ear, contributing to 18.5% of the total reads in middle ear fluids and 3.2% of the rinses. *Streptococcus* and *Moraxella* were less common, and were similarly lower in the MER than the MEF. While *Haemophilus* is prevalent in the fluids, overall the ear samples were dominated by *Alloiococcus* and *Staphylococcus*, with *Turicella* also abundant. These genera were not abundant in the nasopharynx of cases or controls. The ear canal did not contain any taxon that was not also seen at similar or higher aggregated relative abundance in the rinses or fluids.

Species-level identification could not be achieved with the V3/V4 region. Some taxa contain multiple OTUs (which may not be the same species). It is also possible that multiple species have been classified as the same OTU if the sequences are more than 97% identical.

### 2.4.3 Diversity within the nasopharynx and middle ear

We measured alpha (within-sample) diversity with the Faith’s phylogenetic diversity (PD) and inverse Simpson (IS, alternatively reciprocal Simpson) metrics. Alpha diversity by sample type is shown in Figure 2.1. The nasopharynx of children with rAOM was significantly more diverse than the nasopharynx of the healthy controls. Within the same ear of the same child, the middle ear rinse was also significantly more diverse than the fluid but this difference was not as pronounced. The nasopharynx was also more diverse than the middle ear fluid when comparing within the same child.

### 2.4.4 Comparing the microbiome of the nasopharynx between rAOM-prone and rAOM-resistant children

Beta diversity represents the differences between samples; how similar or dissimilar they are to each other. We calculated beta diversity with the weighted UniFrac metric on the raw read counts to determine whether there was a distinct microbiome related to case/control status, sample type or other covariates. Calculating beta diversity on raw counts is accurate and was very similar to using rarefied counts (Procrustes analysis, see Appendix A.6).
Table 2.4: Genus-level community composition by sample type. Relative abundance was calculated for aggregated counts across all samples of each sample type and is summarised at genus level (i.e. aggregates all OTUs with the same genus assignment). Samples below a total read count of 1499 are not included. Genera with an average relative abundance below 1% are grouped as “Other”.

<table>
<thead>
<tr>
<th>Genus level taxonomy</th>
<th>Control NPS (%)</th>
<th>Case NPS (%)</th>
<th>MEF (%)</th>
<th>MER (%)</th>
<th>ECS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Corynebacterium</em></td>
<td>13.3</td>
<td>1.63</td>
<td>1.35</td>
<td>1.34</td>
<td>2.73</td>
</tr>
<tr>
<td><em>Turicella</em></td>
<td>0.03</td>
<td>0.03</td>
<td>6.72</td>
<td>11.72</td>
<td>13.06</td>
</tr>
<tr>
<td><em>Staphylococcus</em></td>
<td>1.42</td>
<td>0.81</td>
<td>9.94</td>
<td>22.23</td>
<td>27.01</td>
</tr>
<tr>
<td><em>Alloiooccus</em></td>
<td>0.17</td>
<td>0.19</td>
<td>49.84</td>
<td>57.04</td>
<td>53.62</td>
</tr>
<tr>
<td><em>Dolosigranulum</em></td>
<td>16.34</td>
<td>1.86</td>
<td>0.05</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>7.05</td>
<td>15.29</td>
<td>3.52</td>
<td>1.2</td>
<td>1.44</td>
</tr>
<tr>
<td><em>Neisseria</em></td>
<td>0.27</td>
<td>10.95</td>
<td>0.19</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td><em>Haemophilus</em></td>
<td>12.43</td>
<td>18.96</td>
<td>18.52</td>
<td>3.18</td>
<td>1.3</td>
</tr>
<tr>
<td><em>Moraxella</em></td>
<td>47.55</td>
<td>30.8</td>
<td>2.17</td>
<td>0.21</td>
<td>0.1</td>
</tr>
<tr>
<td><em>Pseudomonas</em></td>
<td>0.02</td>
<td>0.01</td>
<td>6.34</td>
<td>1.34</td>
<td>0.12</td>
</tr>
<tr>
<td>Other (49 other taxa)</td>
<td>1.42</td>
<td>19.48</td>
<td>1.35</td>
<td>1.59</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Figure 2.1: Alpha diversity measured by a) Faith’s Phylogenetic Diversity and b) Inverse Simpson metrics grouped by sample type. Alpha diversity values were calculated on CSS normalised and logged read counts. The p-values represent the difference between groups determined by Wilcoxon rank sum (case/control NPS) or Wilcoxon signed rank (MEF/MER and MEF/NPS) test where paired samples were from the same child.
2.4. Results

We compared the microbiome of the nasopharynx of cases to that of controls to identify genera that may be associated with rAOM (potential novel pathogens), or with apparent resistance to rAOM (potential candidates for probiotic therapy). Figure 2.2 shows a principal coordinates analysis (PCoA) plot of the UniFrac distances on the nasopharyngeal samples from cases and controls, where each sample is an individual point. The closer two points are, the more similar the microbiome of those samples. The nasopharyngeal microbiome of the cases was distinct from that of the controls, separating along the PC1 axis; indicating that much of the variability between samples is explained by case/control status. No grouping of samples was observed with any other covariate (including age, presence of chronic illness, attendance at day care or school, sex, previous hospital admission for infection, season of collection and presence of siblings), with the exception of those previously shown to differ significantly between cases and controls.

The pattern of recent antibiotic usage overlapped with case/control status, which is expected, as most of the cases had recently taken antibiotics whilst usage in the controls was low. Current breastfeeding and breastfeeding for at least 12 months were also concentrated towards the controls along PC1, who had a higher median length of breastfeeding (see Table 2.3). There were no batch effects by sequencing run (see Appendix A.7).

We then determined which OTUs were causing this difference between the two groups by fitting a model using the fitZIG function in metagenomeSeq. The model controlled for recent antibiotic usage, sex and length of breastfeeding. Of the 33 significantly differentially abundant OTUs, 16 were above a threshold of ≥ 0.35% mean or median relative abundance in either group (see Section 2.3.10) and are shown in Figure 2.3. Appendix A.8 contains a full list of significant OTUs with adjusted p-values and log2 fold-changes between groups for all fitZIG models. Of these 16 OTUs, 14 were more abundant in the nasopharynx of the cases. *Haemophilus* (OTU6) was one of these with a logFC of 2.7 ($p = 0.004$), though it was commonly seen across both groups. The other otopathogen genera *Moraxella* (OTU2) and *Streptococcus* (OTU4) were not significantly different. The remaining 13 OTUs were found at low abundance in the cases, but were very low or almost absent in the controls (logFC = 0.96 to 4.4, $p$-values < 0.01). Those with the highest median log CSS counts were *Gemella* (OTU12) and *Neisseria* (OTU13). The representative sequences for these OTUs did not match any species definitively with BLASTN 2.6.1 against 16S rRNA sequences; OTU12 matched *G. taiwanensis* and *G. haemolysans* at 100%, with OTU13 hitting several *Neisseria* species at 98-99% identity. However, the SILVA
Figure 2.2: Principal coordinates analysis (PCoA) on nasopharyngeal samples from cases and controls. Distances between samples were calculated using the weighted UniFrac metric.244
database classified the sequence as \textit{N. lactamica}.

Three OTUs were more abundant in the nasopharynx of controls compared to the cases. \textit{Moraxella} (OTU10, distinct from OTU2 which was abundant across all nasopharyngeal samples and was not significantly different here) was found at low relative abundance in both groups, but was prevalent in several controls (logFC = 2.0, \( p = 0.05 \)). The representative OTU sequence for this \textit{Moraxella} OTU matches \textit{M. lincolnii} using BLASTN.\textsuperscript{250} \textit{Dolosigranulum} (logFC = 3.0, \( p = 0.002 \)) and \textit{Corynebacterium} (logFC = 4.1, \( p = 8.9 \times 10^{-6} \)) were both low in the cases but significantly more abundant in the controls. \textit{Dolosigranulum} only has one species (\textit{D. pigrum}), but this \textit{Corynebacterium} OTU (OTU8) could not be identified at species level with BLASTN.\textsuperscript{250}

\section*{2.4.5 Comparing the microbiome of different sample types within children with rAOM}

We compared the microbiome of the nasopharynx, middle ear and ear canal in children with rAOM to identify potential novel OM pathogens. The NPS are distinct from the ear samples in a weighted UniFrac PCoA plot (Figure 2.4). The MEF, MER and ECS samples do not form distinct groups and are more sparsely distributed than the NPS. The ear samples (MEF, MER and ECS) were not observed to group by any other covariates (including antibiotic use, age, duration of breastfeeding, presence of other chronic illness, current attendance at day care or school, sex, previous hospital admission for infection, season of collection and presence of siblings), and there were no batch effects by sequencing run (see Appendix A.9); suggesting that the sparse layout of these samples in Figure 2.4 is most likely due to large differences between individual children.

Pairs of left and right ear samples from the same child were strongly correlated. When comparing pairs of taxa summaries, the fluids had a Pearson correlation coefficient of 0.833 (95\% CI 0.821-0.844, non-parametric \( p = 0.001 \), 999 permutations) and the rinses, 0.858 (95\% CI 0.849-0.867, non-parametric \( p = 0.001 \), 999 permutations). They are also moderately similar following Procrustes analysis (see Appendix A.10; \( m^2 = 0.340 \), \( p < 0.001 \), 999 permutations) but many pairs are distant. A lower \( m^2 \) value indicates higher similarity between datasets. To ensure robust results, downstream analyses that required independent samples (i.e. only one ear per child where two were available) were run with the left or right ear randomly chosen (Set 1),
Chapter 2. Characterising the microbiota with 16S rRNA gene sequencing

Figure 2.3: Log CSS normalised counts of differentially abundant OTUs between rAOM-prone and rAOM-resistant children. OTUs shown are significantly differentially abundant between the nasopharyngeal samples of the cases and controls and are additionally found above the threshold of ≥ 0.35% mean or median relative abundance in at least one group. Differential abundance analysis controlled for recent antibiotic usage, length of breastfeeding and sex; children with missing data for any of these covariates were excluded (n = 4). log2FC refers to the log fold change of OTU abundance from cases to controls.
Figure 2.4: Principal coordinates analysis (PCoA) on samples from children with rAOM. Distances between samples were calculated using the weighted UniFrac metric.244
and then repeated with the sample from the opposite ear (Set 2) for agreement.

### Middle ear fluid and middle ear rinse

Sampling of the MEF in a subset (60%) of the cases was followed by a saline MER to attempt to capture bacteria that may not otherwise be detected in the fluid (for example, bacteria present in biofilm adhered to the mucosa). The taxonomic composition of the two sample types were strongly correlated when comparing samples from the same ear (Pearson correlation coefficient 0.835, 95% CI 0.826-0.843, non-parametric \( p = 0.001 \), 999 permutations) and they occupy a similar area on the PCoA plot (Figure 2.4). Procrustes analysis on the first three principal coordinates between these pairs indicated that the sample types are only moderately similar (\( m^2 = 0.335, p < 0.001 \), 999 permutations; see Appendix A.11). To determine which OTUs were differentially abundant between the MEF and MER, we fitted a fitZIG model on MEF/MER pairs from the same ear with the Set 1/Set 2 approach (see Section 2.3.10). The 7 significantly differentially abundant OTUs above threshold are shown in Figure 2.5a (two additional OTUs above threshold were dropped for disagreement between Set 1 and Set 2). *Staphylococcus* (OTUs 3, 212, 269 and 1003), *Alloiococcus* (OTU1) and *Turicella* (OTU7) were found at higher abundance in the MER than in the MEF (logFC = 0.9 to 2.1, \( p \)-values < 0.0007). *Haemophilus* (OTU6) was more abundant in the MEF than in the MER (logFC = 1.2, \( p = 2.5 \times 10^{-9} \)). While three *Staphylococcus* OTUs (OTU212, OTU269 and OTU1003) were found at very low abundance, all other OTUs were found at moderate abundance in both sites.

### Middle ear and nasopharynx

We found that the middle ear and nasopharynx of children with rAOM were not highly concordant. In Figure 2.4, the NPS samples are distinct from the ear samples (including MEF, MER and ECS). Differential abundance analysis revealed 23 significant OTUs above threshold between the MEF and NPS (an additional 2 OTUs above threshold were dropped for disagreement between Set 1 and Set 2), and 18 OTUs when comparing the MER to the NPS (4 OTUs above threshold dropped for disagreement). The abundance of significant OTUs are shown in Figure 2.5b and 2.5c. Overall, the genera *Staphylococcus* (OTU3), *Alloiococcus* (OTU1) and *Turicella* (OTU7) were highly abundant in the middle ear, with very low abundance in the nasopharynx (logFC = 2.6 to 5.2, \( p \)-values < 0.001); these genera appear to be charac-
Figure 2.5: Log CSS normalised counts of differentially abundant OTUs amongst sample types within the cases. OTUs plotted are significantly differentially abundant between paired (within-child) a) MEF and MER samples; b) MEF and NPS samples; c) MER and NPS samples; d) MEF and ECS samples. Only OTUs with an adjusted $p \leq 0.05$ and above the threshold of $\geq 0.35\%$ mean or median relative abundance in at least one of the groups in each comparison are shown. $\log_2 FC$ refers to the log fold change between the two groups, with the value representing the change from a) fluids to rinses; b) fluids to NPS; c) NPS to rinses; d) ear canals to fluids.
teristic of the middle ear. OTU3 matched several species of *Staphylococcus* above 97% with BLASTN, however *Alloiococcus* and *Turicella* both contain one known species each (*A. otitidis* and *T. otitidis* respectively). The otopathogen genera *Haemophilus* (OTU6), *Moraxella* (OTU2) and *Streptococcus* (OTU4) were moderately abundant in both sites but were higher in the nasopharynx (logFC = 2.3 to 6.4, p-values < 0.002). Several OTUs including *Gemella* (OTU12) and *Neisseria* (OTU13) were at very low abundance in the middle ear and higher in the nasopharynx, though still low overall (logFC = 0.8 to 3.6, p-values < 0.05). These low-abundance OTUs appear to be contributing to the increased diversity seen in the nasopharynx compared to the middle ear. Differences between the middle ear and nasopharynx were generally more pronounced when comparing to the MER samples.

### Middle ear fluid and ear canal

The ear canal samples were taken to assess which bacteria may potentially contami-
inate the MEF and MER samples during collection. No bacterial taxa were present in the ear canal that were not present in the middle ear, and the dominant bacteria in the canal were also dominant in the middle ear (see Table 2.4). Shown in Figure 2.5d, only three OTUs were significantly differentially abundant between pairs of ECS and MEF samples. *Alloiococcus* (OTU1), *Turicella* (OTU7) and *Staphylococcus* (OTU3) occurred at higher abundance in the ear canal (logFC = 2.9, 1.7 and 2.7 respectively; p-values < 0.004), though they were common in both sites. *Haemophilus* (OTU6) was higher in the middle ear fluid with a log fold change of 3.6 (p = 2.5 x 10^{-13}).

### 2.4.6 Patterns of bacterial co-occurrence

We searched for strong correlations between the dominant OTUs representing the known otopathogen genera (*Haemophilus, Streptococcus* and *Moraxella*) and also between the OTUs we found to be characteristic of the case or control nasopharynx, or the middle ear. Correlations were determined separately for each sample type with SparCC. Correlograms showing the overall pattern of correlations for each sample type are shown in Figure 2.6, and a full list of correlations can be found in Appendix A.12.

Within the case nasopharynx, there were several moderately strong (0.6-0.8) corre-
lations with p < 0.01 (100 simulations). The strongest correlations observed were between low-abundance OTUs *Haemophilus* (OTU1033) and *Moraxella* (OTU1073)
2.4. Results

Figure 2.6: Correlations between OTUs. Correlation coefficients between OTUs were calculated by SparCC\textsuperscript{246} within the a) Case NPS; b) Control NPS; c) MEF and d) MER. One sample per child was included in each set. Non-significant correlations (one-sided \( p > 0.05 \)) are coloured white. \( N \) refers to the number of samples included in each correlation analysis, which tested for correlations between all OTUs observed in those samples.
with coefficient 0.80; between the most abundant *Moraxella* OTU (OTU2) and *Haemophilus* (OTU1033) with coefficient 0.77 and between *Gemella* (OTU12) and *Porphyromonas* (OTU23) with coefficient 0.70. *Gemella* (OTU12) and *Neisseria* (OTU13), OTUs that were significantly more prevalent in the cases than in controls, were also positively correlated (0.65). Correlations between the dominant OTUs for *Haemophilus* (OTU6), *Moraxella* (OTU2) and *Streptococcus* (OTU4) were not significant.

Within the control NPS samples, there were fewer significant correlations observed. A moderately strong correlation (0.722, \( p < 0.01, 100 \) simulations) was detected between *Corynebacterium* (OTU5) and *Dolosigranulum* (OTU8), which were significantly more prevalent in the nasopharynx of controls than cases. *Moraxella* (OTU10) correlated only weakly with other OTUs.

In the middle ear, the dominant *Staphylococcus* (OTU3) correlated with other low-abundance *Staphylococcus* OTUs (1003 and 212), with coefficients 0.61 to 0.73 (\( p \)-values < 0.01, 100 simulations) in both the MEF and MER. These minor *Staphylococcus* OTUs were also moderately to strongly correlated with each other (coefficient 0.83 in MER, 0.71 in MEF). A moderate negative correlation was observed between *Staphylococcus* (OTU3) and *Alloiococcus* (OTU1) in the MER (coefficient = -0.52, \( p < 0.01, 100 \) simulations), although this association was weak in the MEF (coefficient = -0.3, \( p < 0.01, 100 \) simulations). The genera that appear to be characteristic of the middle ear, *Alloiococcus* (OTU1) and *Turicella* (OTU7), only weakly correlated with each other in the MEF (coefficient = 0.37, \( p < 0.01, 100 \) simulations) but not in the MER. No significant correlations were found between *Haemophilus* (OTU6), *Streptococcus* (OTU4) and *Moraxella* (OTU2) in either the MEF or MER.

### 2.4.7 Detection of respiratory viruses

We tested all MEF and NPS samples for common respiratory viruses. The identification rates (the percentage of positive samples out of the total number of samples tested) are shown in Table 2.5. Amongst the NPS samples, 61.3% of the cases were positive for at least one virus compared to 47.6% of the healthy controls, though the odds of detecting any virus were not significantly higher in the cases (OR 1.4; CI 0.8, 2.6). The odds of detecting respiratory syncytial virus (RSV) were 9.6 times higher (CI 2.2, 60.5) in the cases than the controls. HMPV was also substantially different between the groups; while the detection rate was low in the cases (8.6%) the virus was not detected in any of the controls. For all other viruses tested, the odds of
2.5 Discussion

The microbiome of the middle ear in children with rAOM has remained relatively unexplored. Modest proportions of children with AOM carry no detectable otopathogen in the MEF by culture or PCR, so there is an opportunity for the identification of novel otopathogens by studying the microbiome of the otitis-prone middle ear. Additionally, studies of the microbiome can help to identify protective bacteria that give rise to probiotic therapies, which present a novel treatment option for rAOM. Our study explored both of these opportunities and found evidence of bacterial genera that may represent novel otopathogens as well as genera that are worth further investigation as probiotic candidates. In doing so, we have, to our knowledge, characterised for the first time the microbiome of both the middle ear and nasopharynx in children with rAOM.

Our study determined that the nasopharyngeal microbiome of rAOM-resistant children is distinct from that of rAOM-prone children. Specifically, we identified three bacteria that had higher relative abundances in our healthy controls: Corynebacterium (OTU8), Dolosigranulum (OTU5) and, to a lesser extent, Moraxella (OTU10). The exact Corynebacterium species could not be determined, but the genus Dolosigranulum contains only one species, *D. pigrum*, and the Moraxella OTU was most likely *M. lincolnii* by BLASTN. This Moraxella OTU was distinct from the dominant Moraxella (OTU2) found abundantly in the nasopharynx of both cases and controls, but OTU10 was not prevalent in the nasopharynx of all controls. The results of our study are supported by previous reports that Corynebacterium and Dolosigranulum are found in the nasopharynx of healthy adults and children and may be
Chapter 2. Characterising the microbiota with 16S rRNA gene sequencing

Table 2.5: Viral identification rates in the nasopharynx of cases and controls. The identification rate is the percentage of positive samples out of the total number of samples collected for that group.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Case NPS (n = 93)</th>
<th>Control NPS (n = 103)</th>
<th>Odds Ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFV</td>
<td>1 (1.1%)</td>
<td>2 (1.9%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HAdV</td>
<td>18 (19.4%)</td>
<td>6 (5.8%)</td>
<td>3.9 (1.5, 10.3)</td>
<td>0.004</td>
</tr>
<tr>
<td>HBoV</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RSV</td>
<td>15 (16.1%)</td>
<td>2 (1.9%)</td>
<td>9.6 (2.2, 60.5)</td>
<td>0.0005</td>
</tr>
<tr>
<td>HCoV</td>
<td>14 (15.1%)</td>
<td>5 (4.9%)</td>
<td>3.5 (1.1, 10.4)</td>
<td>0.03</td>
</tr>
<tr>
<td>HPIV</td>
<td>4 (4.3%)</td>
<td>4 (3.9%)</td>
<td>1.1 (0.3, 4.8)</td>
<td>1</td>
</tr>
<tr>
<td>HMPV</td>
<td>8 (8.6%)</td>
<td>0 (0%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RV</td>
<td>40 (43.0%)</td>
<td>35 (34.0%)</td>
<td>1.5 (0.8, 2.6)</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>57 (61.3%)</strong></td>
<td><strong>54 (47.6%)</strong></td>
<td><strong>1.4 (0.8, 2.6)</strong></td>
<td><strong>0.2</strong></td>
</tr>
</tbody>
</table>

Table 2.6: Viral concordance rates between the middle ear and nasopharynx of cases. The concordance rate is the number of cases where the virus was identified in both the MEF and NPS out of the total number of cases where the virus was detected at all.

<table>
<thead>
<tr>
<th>Virus</th>
<th>IFV</th>
<th>HAdV</th>
<th>HBoV</th>
<th>RSV</th>
<th>HCoV</th>
<th>HPIV</th>
<th>HMPV</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases with virus in MEF</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>13</td>
<td>3</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>(50.0%)</td>
<td>(14.3%)</td>
<td>(0%)</td>
<td>(40.0%)</td>
<td>(48.1%)</td>
<td>(42.9%)</td>
<td>(33.3%)</td>
<td>(48.7%)</td>
<td></td>
</tr>
<tr>
<td>Cases with virus in NPS</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>15</td>
<td>14</td>
<td>4</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>(50.0%)</td>
<td>(85.7%)</td>
<td>(0%)</td>
<td>(60.0%)</td>
<td>(51.9%)</td>
<td>(57.1%)</td>
<td>(66.7%)</td>
<td>(51.3%)</td>
<td></td>
</tr>
<tr>
<td>Concordance rate</td>
<td>0/2</td>
<td>2/19</td>
<td>0/0</td>
<td>6/19</td>
<td>5/22</td>
<td>1/6</td>
<td>3/9</td>
<td>24/54</td>
</tr>
<tr>
<td>(0%)</td>
<td>(10.5%)</td>
<td>(0%)</td>
<td>(3.16%)</td>
<td>(22.7%)</td>
<td>(16.7%)</td>
<td>(33.3%)</td>
<td>(44.4%)</td>
<td></td>
</tr>
</tbody>
</table>

associated with a decreased risk of AOM.\textsuperscript{169} We also observed a positive correlation between \textit{Corynebacterium} and \textit{Dolosigranulum} which has been previously reported.\textsuperscript{168} It has been suggested that the production of lactic acid by \textit{Dolosigranulum} makes the surrounding environment more habitable for \textit{Corynebacterium} species,\textsuperscript{254} which would explain the co-occurrence we observed in our study. Some studies have observed a decrease in the abundance of these genera in the nasopharynx of children who had taken antibiotics in the preceding months,\textsuperscript{169,253} and they were also detected more frequently in the nasopharynx of breastfed compared to formula-fed 6 week old infants.\textsuperscript{255} Breastfeeding was extended in our healthy controls (median 10 months) compared to our cases (median 5 months), which is consistent with the strong evidence for breastfeeding in the first two years of life being protective against AOM.\textsuperscript{256} Together these results indicate that \textit{Corynebacterium} and \textit{Dolosigranulum} may be part of the normal flora of the nasopharynx, which might be strengthened by breastfeeding and disrupted by antibiotic use. In particular, \textit{D. pigrum} is susceptible to amoxicillin\textsuperscript{257}, the recommended antibiotic for treatment of AOM. Our cases had a shorter median breastfeeding duration and higher recent antibiotic use than the controls, suggesting that maintenance of this normal flora may be important in preventing rAOM.

Regarding the potential for \textit{Corynebacterium} and \textit{Dolosigranulum} to actively protect against the development of rAOM in the nasopharynx of controls, protective bacteria are often closely related to the pathogens they inhibit.\textsuperscript{197,219} \textit{Corynebacterium} is a member of the same family as \textit{Turicella} (one species; \textit{T. otitidis}) and \textit{D. pigrum}'s closest relative is \textit{Alloiococcus} (one species; \textit{A. otitidis}).\textsuperscript{258} \textit{Alloiococcus} and \textit{Turicella} have both been identified as potential otopathogens\textsuperscript{95,259} and both were abundant in the MEF, MER and ECS of cases in this study, although they were not commonly found in the nasopharynx of our cases. Currently, the pathogenic role of \textit{Alloiococcus} and \textit{Turicella}, if any, is yet to be determined. Other studies have identified bacteria with a protective role that are not closely related to the pathogens they inhibit.\textsuperscript{191,196}

In the anterior nares, \textit{Dolosigranulum} has been linked to decreased rates of colonisation with \textit{S. aureus}\textsuperscript{260} and \textit{Corynebacterium} has demonstrated activity against \textit{S. pneumoniae in vitro},\textsuperscript{261} indicating the potential for pathogen inhibition by these bacteria \textit{in vivo}. While we have shown that \textit{Corynebacterium} and \textit{Dolosigranulum} are characteristic of rAOM-resistant children, determining whether they play a role in protecting against rAOM is more challenging. We controlled for antibiotic usage within the month prior to sampling, however we did not have information relevant to long-term or repeat antibiotic usage, which is common in children with rAOM. We therefore cannot exclude the possibility that \textit{Corynebacterium} and \textit{Dolosigranulum}
are depleted in the cases due to long-term or repeated antibiotic use. Additionally, as this is a cross-sectional study we are not able to determine whether decreased Corynebacterium and Dolosigranulum precedes the development of rAOM, or if their depletion is a result of the disease. As 16S rRNA sequencing measures relative, not absolute abundance, it is also important to consider that the lower relative abundance of Corynebacterium and Dolosigranulum in children with rAOM may be due to pathogen overgrowth. The absolute abundance of these organisms may be unchanged, and an increase in the abundance of otopathogen species may have correspondingly influenced the relative abundance proportions. While our study is unable to answer these questions, the abundance of these genera in the controls warrants further investigation to determine whether they have a protective role in the nasopharynx.

As well as identifying potentially protective bacterial species we also investigated the potential for novel otopathogen identification. It is generally accepted that the otopathogens originate from the nasopharynx and ascend the Eustachian tube to cause infections in the middle ear. We found that the middle ear and nasopharynx of children with rAOM were not highly concordant, supporting similar findings from previous reports.\textsuperscript{76,262} Haemophilus was the most abundant otopathogen genus in the MEF, which corresponds with surveillance studies reporting a predominance of non-typeable Haemophilus influenzae (NTHi) in children with rAOM since the introduction of the pneumococcal vaccine targeting S. pneumoniae.\textsuperscript{64} However, Haemophilus, Moraxella and Streptococcus were observed at low abundance in the MEF and MER compared to Alloiococcus (OTU1), Staphylococcus (OTU3) and Turicella (OTU7), with Alloiococcus the overall most abundant genus in the MEF, MER and ECS. There are many species of Staphylococcus, and we couldn’t identify OTU3 at species level by BLASTN. S. aureus has previously been isolated from the MEF of children with AOM\textsuperscript{263} but is usually associated with chronic suppurative OM where the tympanic membrane is perforated.\textsuperscript{264,265} The other genera each contain only one species (A. otitidis and T. otitidis respectively) which have previously been identified as possible otopathogens.\textsuperscript{95,259} At present, the role of these bacteria in OM is still under debate.\textsuperscript{105,107} A. otitidis has been frequently detected\textsuperscript{75,102,172,176} at high abundance\textsuperscript{172,176} in MEF, usually from children with OME; but T. otitidis has been less well studied. Recently, its abundance was associated with older children (>24 months) with COME,\textsuperscript{178} though we did not observe this pattern in our cohort. Unlike the known otopathogens, Alloiococcus, Staphylococcus and Turicella have been identified as members of the normal ear canal flora in both children and adults\textsuperscript{108} and in our study were more abundant in the ECS than in the MEF. Also unlike the
characterising the microbiota with 16S rRNA gene sequencing

otopathogens, *Turicella* was almost absent from the nasopharynx of our cases, whilst *Alloiococcus* and *Staphylococcus* were also uncommon in this niche. *Alloiococcus* has previously been isolated infrequently from the nasopharynx of children with upper respiratory tract infections or a history of AOM.\textsuperscript{75,107} The rarity of *Alloiococcus* and *Turicella* in the nasopharynx suggests that they may typically colonise the middle ear. Few studies have investigated the normal flora of the healthy middle ear; when healthy, this site contains no fluid and is inaccessible without a surgical procedure. Those that have sampled this site in healthy children did not report *Alloiococcus*\textsuperscript{266,267} and *Turicella* appears to have only been observed once in the healthy middle ear of an adult.\textsuperscript{267} While *Alloiococcus*, *Staphylococcus* and *Turicella* seem to have the potential to be novel otopathogens, we cannot yet exclude the possibility that they are normal aural flora.

We also observed increased abundance of *Alloiococcus*, *Staphylococcus* and *Turicella* in the MER compared to the MEF; whilst *Haemophilus* was more common in the MEF and *Streptococcus* and *Moraxella* were not significantly different between the two sample types. It is possible that *Alloiococcus*, *Staphylococcus* and *Turicella* adhere to the mucosa and were more efficiently sampled by the MER; differences between the middle ear and NPS were more pronounced when comparing the MER than MEF, suggesting the MER might better represent the microbiome of the middle ear. Alternatively, the MER may include contamination from the ear canal, as *Alloiococcus*, *Staphylococcus* and *Turicella* were the most dominant organisms in the canal and were significantly more abundant here than in the MEF. We observed increased diversity in the MER compared to the MEF, though our differential abundance analysis revealed significant differences only between genera that were common in both sites. The increased diversity may therefore include low abundance contaminant OTUs, suggesting that sampling methods to obtain the MER may be more prone to environmental or skin contamination than the MEF. However, as there were no genera unique to the ear canal and the genera at this location were found abundantly in both the ear canal and middle ear it is difficult to determine whether the canal flora contaminates the middle ear fluid with *Alloiococcus*, *Staphylococcus* and *Turicella* during sampling.

We observed that the nasopharyngeal microbiome of the cases was significantly more diverse than that of controls. This is in contrast to studies of the gut microbiome (a high-diversity environment\textsuperscript{268}), where greater diversity has been associated with a healthy state, and a decrease in diversity is characteristic of disease.\textsuperscript{269,270} However, studies of the vaginal microbiome (a low-diversity environment\textsuperscript{271}) have shown a
similar pattern to that observed in the nasopharynx in our study; that a lower diversity is observed in the healthy environment with an increased diversity being characteristic of disease.\textsuperscript{272,273} Previous studies suggest that the nasopharynx is dominated by only a few taxa\textsuperscript{252,253} so it may be that this pattern is characteristic of less complex microbiomes. Additionally, a study in the adult nasopharynx determined that a more diverse microbiome was more susceptible to colonisation with \textit{S. pneumoniae}.\textsuperscript{251} We cannot determine from our cross-sectional study whether the higher diversity in the cases occurs before or after the development of rAOM, or as a consequence of repeat antibiotic treatment where the normal nasopharyngeal microbiome is disrupted, perhaps allowing the community to diversify. A study in adults has indicated that the gut microbiome can begin to resemble its original state in as little as one week after antibiotic treatment,\textsuperscript{274} though this recovery is often incomplete. The microbiome of adults with cystic fibrosis is similarly resilient to short-term antibiotics.\textsuperscript{275} However, antibiotic-induced changes in the microbiome in children may be more long-term. In the gut microbiome of children, reduced richness can persist for up to 2 years after the use of macrolides.\textsuperscript{140} Similarly, after long-term oral and inhaled antibiotic use the taxonomic richness of the microbiome in children with cystic fibrosis is markedly reduced.\textsuperscript{141} It is possible that in our cohort, the nasopharyngeal microbiome of rAOM-prone children has been altered by antibiotic use that extends beyond the month prior to sampling that we controlled for. However, based on these studies, a reduction in microbial diversity would be expected with repeated antibiotic use; this is the opposite effect to what we observed in our cases.

The increased diversity in the case nasopharynx was contributed to in part by the presence of \textit{Gemella} (OTU12) and \textit{Neisseria} (OTU13), which positively correlated with each other. The OTU representative sequences matched multiple species above 97\% identity with BLASTN, however the SILVA database classified OTU13 as \textit{N. lactamica}, which is a commensal of the nasopharynx in children.\textsuperscript{276} While \textit{Gemella} and \textit{Neisseria} were characteristic of the nasopharynx in children with rAOM, their abundance in the middle ear was significantly lower, suggesting these genera are not involved in pathogenesis in the middle ear and are unlikely to be novel otopathogens. Both \textit{Neisseria} and \textit{Gemella} have been observed in the nasopharynx of children with upper respiratory illness and AOM,\textsuperscript{169} though \textit{Neisseria} has also previously been reported amongst the healthy flora of the nasopharynx.\textsuperscript{174,252} It is therefore unclear whether these genera play a role in the nasopharynx of children with rAOM or if they represent a shift in the composition of the microbiome due to repeated antibiotic usage or other factors.
Contrary to the results of our study, Hilty et al. (2012)\textsuperscript{174} and Pettigrew et al. (2012)\textsuperscript{169} both reported decreased diversity in the nasopharynx of children with AOM compared to children without AOM. The nasopharyngeal samples in these studies were taken during an episode of AOM, so possibly better represent the nasopharyngeal microbiome at this time. However, the majority of healthy controls used for comparison did not attend day care and most of the cases did not have a history of recurrent AOM (or this information was not provided), so these cohorts may represent less extreme ends of the phenotypic spectrum of disease.

Episodes of AOM often occur after a viral upper respiratory tract infection,\textsuperscript{14} with bocavirus (HBoV),\textsuperscript{277} rhinovirus (RV)\textsuperscript{81} and respiratory syncytial virus (RSV)\textsuperscript{278} commonly found in children with AOM. Rhinovirus was the most frequently detected virus in the nasopharynx of both our cases and controls. We detected other viruses more frequently in the cases than controls, though the odds were not always significant as relatively few viruses were detected overall. The odds of detecting respiratory syncytial virus were significantly higher in the cases than controls, which supports its association with an increased risk of AOM.\textsuperscript{279} We also observed human metapneumovirus (HMPV) in the nasopharynx of a small number of cases, but it was absent from all of the controls, suggesting an association with rAOM. HMPV has previously been observed in 5% of children with upper respiratory tract infections, with half of these also experiencing AOM.\textsuperscript{280} Our results were similar to the patterns of viral detection in the nasopharynx in an independent Western Australian cohort of children with and without rAOM,\textsuperscript{81} though the overall rate of viral detection was higher in that study (rAOM: 94% with at least one virus detected, controls: 71%, compared to our 61% and 48% respectively). This may be because they additionally tested for polyomavirus and enterovirus, which they found in 36% and 17% of children with rAOM respectively; for polyomavirus this was significantly higher than in their healthy controls. Bocavirus was not detected in any of the cases or controls in our study, although it has previously been seen in children with rAOM.\textsuperscript{81,277} We detected viruses less often in the middle ear than in the nasopharynx, which Wiertsema et al. (2011)\textsuperscript{81} also reported. Despite lower viral identification rates, results from our cohort therefore seem to follow the same pattern as previously reported in Western Australian children.

Rhinovirus\textsuperscript{85} and respiratory syncytial virus\textsuperscript{86} have been observed to enhance the adherence of \textit{S. pneumoniae} to epithelial cells \textit{in vitro}. It has previously been reported that rhinovirus presence correlates with presence by standard culture of each of the three major otopathogens,\textsuperscript{67} and adenovirus with \textit{M. catarrhalis} presence assessed
by culture and PCR.\textsuperscript{67,81} We could not observe this in our study as 16S rRNA gene sequencing could not identify the otopathogen species; sequences that are more than 97% identical are grouped into the one OTU, which we could only identify at genus level. It is possible that \textit{Haemophilus}, \textit{Streptococcus} and \textit{Moraxella} OTUs include both the otopathogens and commensal species from the same genus, which would obscure viral/otopathogen correlations. 

One limitation specific to our study is that grommet surgeries are generally not performed during active infection, so the microbiome of our rAOM-prone children may not be fully representative of the microbiome during an episode of AOM. 16S rRNA studies in general can provide a comprehensive overview of the taxonomic composition of the microbiome, but are limited in that they do not provide information on microbiome function or gene content. Additionally, there are important biases to consider when conducting these studies. DNA extraction methods\textsuperscript{281} and amplicon primers\textsuperscript{161} work with variable efficiency across bacterial taxa, which can result in the underrepresentation of some bacteria. Our DNA extraction protocol and amplicon primers were chosen based on recommendations by Yuan \textit{et al.} (2012)\textsuperscript{281} and Klindworth \textit{et al.} (2013)\textsuperscript{161} respectively to reduce this bias. The number of copies of the 16S rRNA gene can vary amongst bacteria,\textsuperscript{282} even between strains\textsuperscript{283} which can inflate OTU abundance. Additionally, copies within a single genome are not always identical,\textsuperscript{284} which can inflate the number of OTUs detected. Samples can be contaminated with DNA found in reagents and the laboratory environment,\textsuperscript{285} and there is the potential in our study for contamination during sample collection (i.e. from the ear canal or anterior nares). Contamination can heavily influence low biomass environments,\textsuperscript{285,286} however we sequenced positive and negative controls to address this (see Appendix A.5). Furthermore, it is difficult to achieve species-level identification with 16S rRNA sequencing as related species are often very similar in this region and the efficiency of classification also varies depending on the region of the 16S rRNA gene.\textsuperscript{287} There is the possibility that there are commensal bacteria that we could not detect which are within the same OTU as the otopathogens. For example, \textit{Haemophilus haemolyticus} and \textit{Streptococcus salivarius} are closely related to otopathogens and have shown promise in other studies as candidates for the prevention of rAOM.\textsuperscript{211,288} Metagenomic shotgun sequencing addresses the issue of species-level identification as it sequences across the entire genome, and can thereby also provide information on gene content and function. However, DNA sequencing itself can only detect bacterial presence and does not indicate bacterial viability or activity. Metatranscriptomics addresses this issue, but this is a relatively new field and has not yet been applied to OM. The precautions taken in our study aimed to
reduce the biases inherent to 16S rRNA sequencing, however we acknowledge that the relative abundance of taxa may not reflect the true proportions of bacteria and the genera we have detected may contain multiple species which are not necessarily viable or active.

2.6 Conclusions

Our study has provided the first comprehensive exploration of the microbiome of the middle ear and nasopharynx in children with rAOM. We have taken an important step in the identification of candidate therapeutic bacteria derived from the healthy microbiome by observing significantly higher proportions of *Corynebacterium* and *Dolosigranulum* in the nasopharynx of healthy controls. Further research should focus on investigating their potential to inhibit the known otopathogens, and it would be of interest for longitudinal studies to determine whether the abundance of these genera change prior to or as a result of rAOM. We have also identified *Alloiococcus*, *Staphylococcus* and *Turicella* as potential otopathogens, though their relative absence in the nasopharynx and abundance in the ear canal means we cannot rule out their role as normal aural flora. *Gemella* and *Neisseria* contribute to increased diversity in the nasopharynx of children with rAOM, but do not appear to colonise the middle ear and are therefore not likely to be novel otopathogens. Shotgun metagenomics and metatranscriptomics are the next steps towards achieving species-level or strain-level resolution of these bacteria of interest and confirming their viability and investigating their activity in the middle ear. Our study has contributed significantly towards greater understanding of the microbiome of the upper respiratory tract in children with rAOM, and has taken an important step towards the development of specific probiotic therapies for the disease.
Chapter 3

Exploration of the microbiome with shotgun metagenomics sequencing

In Chapter 2, 16S rRNA amplicon sequencing analysis identified important genera associated with resistance to rAOM (*Corynebacterium, Dolosigranulum* and *Moraxella*) and potential novel otopathogens (*Alloiococcus, Staphylococcus* and *Turicella*). In this chapter, shotgun metagenomics is applied to further identify these organisms at species level and to explore differences in the functional capability of the middle ear and nasopharyngeal microbiota. This will facilitate an understanding of the role of these organisms in the development of, or resistance to rAOM.
3.1 Introduction

The microbiome of OM has only recently begun to be explored, with relatively few studies having investigated the microbiome of the middle ear or nasopharynx in children with OM; most of which focus on the COME phenotype. In Chapter 2, the microbiota of the middle ear, EAC and nasopharynx of children with rAOM (cases) and the nasopharyngeal microbiome of healthy children (controls) were characterised by 16S rRNA amplicon sequencing. In that study, *Corynebacterium*, *Dolosigranulum* and *Moraxella* were enriched in the nasopharynx of healthy children, who had been exposed to some of the major environmental risk factors for rAOM but had not developed the disease. *Corynebacterium* and *Dolosigranulum* have previously been reported together in the nasopharynx of healthy children and adults and appear to be members of the healthy nasopharyngeal flora. If these organisms play a protective role in the nasopharynx of healthy children, they may be therapeutically useful in preventing rAOM through translation into a probiotic formulation. However, as the study in Chapter 2 was cross-sectional, it is not yet clear whether these organisms actively contribute to health or are simply out-competed by the otopathogens or reduced by antibiotic use in children with rAOM.

*Alloiococcus*, *Turicella* and *Staphylococcus* were the dominant genera observed in the middle ear of children with rAOM in Chapter 2, which points to a possible role for them in the pathogenesis of rAOM. The involvement of *Alloiococcus* and *Turicella* in OM has been under debate for several years, as unlike the known otopathogens (NTHi, *S. pneumoniae* and *M. catarrhalis*) they are also prevalent in the EAC of children with and without OM and are infrequently detected in the nasopharynx. This was also observed in Chapter 2 and due to their abundance in the EAC, it remains possible that contamination of the MEF from the canal during sampling contributes to their detection in the middle ear. For both the health-associated genera and these potential novel otopathogen genera, a species-level understanding of these organisms is required, as this cannot be achieved with 16S rRNA profiling.

16S rRNA gene sequencing may be referred to as ‘metataxonomics’, as its purpose is to identify the members of the microbiota based on the amplification of a universal bacterial marker gene. It is a relatively inexpensive method for high-throughput profiling of microbial communities, but only provides taxonomic information and thus has several limitations. First, it cannot provide information on the functional
3.1. Introduction

The capability of the microbial community as it does not involve the sequencing of genes other than the 16S rRNA marker gene. Function can be inferred from metataxonomic profiles by assuming the presence of known genome content from available reference genomes, though this method is indirect and does not perform well on poorly-characterised communities. Secondly, the achievable taxonomic resolution is limited to bacteria at genus level or above. Species-level distinction can be important in the upper respiratory tract microbiota as some genera including *Streptococcus*, *Haemophilus* and *Moraxella* contain both pathogenic and non-pathogenic species. The 16S rRNA gene is found only in bacteria and archaea, and so the viral, fungal and microeukaryotic components of the microbiota are missed by metataxonomics, unless they are targeted via another conserved marker gene such as the 18S rRNA gene or internal transcribed spacer (ITS) region (though no such universal marker exists for viruses). In metataxonomics, the amplicon is generally too short to provide sufficient sequence information to discriminate between species; some of which may not exhibit much variation in the 16S rRNA gene. Finally, metataxonomics is prone to multiple sources of bias affecting the reported taxonomic composition of the microbiota. The choice of primers can influence the amplicon PCR, with some bacterial taxa amplified more efficiently than others depending on which segments of the 16S rRNA gene are sequenced. Additionally, different bacterial genera harbour varying numbers of copies of the 16S rRNA gene in their genome, which can influence the measurement of relative abundance. Sequence variation between copies from the same genome as well as lateral gene transfer amongst some taxa also complicates the delineation of taxonomy by this marker gene. It is challenging to account for all of these factors when analysing metataxonomic data and so the method is generally not considered to reflect absolute abundances of organisms.

The application of whole genome shotgun (WGS) sequencing to the microbiome is known as metagenomics. Distinct from metataxonomics, metagenomics involves sequencing the entire genomes of the organisms present in a specimen and is not subject to biased amplification. Metagenomics can provide higher resolution taxonomic classification, as each organism is represented by more than a small part of a single gene; allowing species and sometimes strain-level descriptions of the microbiota. By sequencing all parts of the genomes, the method also provides information on the protein-coding genes within the organisms, enabling the characterisation of the metabolic functions that the microbial community is potentially capable of carrying out. The metagenomics approach has been successfully used on specimens from a wide range of body sites in the Human Microbiome Project. It is also a very powerful technique for producing high quality assembled genomes of unculturable
Chapter 3. Exploration of the microbiome with shotgun metagenomics...

or unknown organisms. The method is, however, more expensive than metataxonomics in terms of both cost and computational resources and without amplification of a specific group of organisms, it also requires consideration of the impact of host DNA, particularly in human clinical samples.

While other environments have been extensively studied with metagenomics, the application of this method to the respiratory tract is novel. Metagenomics in the upper respiratory tract can be challenging due to the low sample volume and low DNA yield from respiratory specimens, of which the majority is host, not microbial DNA. Most studies utilising metagenomics to explore the functional capability of the respiratory microbiota have primarily focused on the lower airways, reviewed by Pienkowska et al. (2017). Metagenomics in the nasopharynx is still in its infancy, with a limited number of studies exploring this field. These studies have mostly focused on taxonomy, particularly that of viruses; with only a few studies having explored functional potential in the nasopharyngeal microbiota, and none having taken this approach in otitis-prone children or the middle ear.

The work in this chapter follows on from the metataxonomic study described in Chapter 2. The NPS samples most representative of cases (children with rAOM) and controls (healthy children resistant to rAOM) along with matching MEF samples from the cases were selected to undergo shotgun metagenomics sequencing. The aims of this approach were to 1) further characterise the bacterial genera of interest identified in Chapter 2 – the potentially protective Corynebacterium, Dolosigranulum and Moraxella and the potential otopathogens Alloiococcus, Staphylococcus and Turicella – by identifying their members at species level; and 2) to identify the major differences in functional capability between the case and control nasopharyngeal microbiota, and between the middle ear and nasopharyngeal microbiota of cases. This will enable a better understanding of the role of these organisms in the pathogenesis of rAOM or resistance to the development of rAOM.

3.2 Methods

3.2.1 Patient recruitment and sample collection

The samples analysed in this study were obtained from children recruited to the Perth Otitis Media Microbiome (biOMe) Study. Details on the recruitment of children to the study, sample collection and DNA extraction can be found in Chapter 2.
3.2.2 Selection of samples for metagenomic shotgun sequencing

Samples from 93 case children and 103 control children were collected in the biOMe study. As the depth of sequencing required for metagenomics (and therefore the cost) is much greater than for metataxonomics, a subset of 28 samples was selected to undergo metagenomic shotgun sequencing. The aim was to select samples most representative of the case and control groups. Samples were selected via a random forest classification model in the randomForest package v4.6-12\textsuperscript{314} in R v3.4.1.\textsuperscript{238} This is a supervised machine-learning method that constructs a large number of decision trees from the data it is given. These trees represent a series of characteristics (abundances of different OTUs) that describe whether a sample is a case or a control. By constructing several thousand trees using different subsets of samples and OTUs, each sample can then be classified by the trees in which it was not used, providing internal validation. Samples that are ‘easily’ classified by the model (i.e. those identified correctly as a case or control in the vast majority of trees) were considered to be the most representative of each group.

To build the classification model, a rarefied (subsampled to 1499 reads per sample) OTU table generated from the 16S rRNA gene sequencing analysis in Chapter 2 containing only NPS samples from cases and controls was used as input. Additionally, samples from children that had taken antibiotics in the previous month were excluded from the model as this was determined to be a significant confounding factor in the composition of the microbiota (Chapter 2). The random forest classification model was built with 5000 trees and produced an overall out-of-bag error rate (the average percentage of times the classification was incorrect) of 10.9%, with less error in classifying controls (2.3% class error) than cases (35.5% class error). The votes (the proportion of times a particular sample was classified as either a case or control) were sorted in descending order to list the samples that were most often classified correctly. Samples were selected for sequencing from this list based on their genomic DNA yield, the presence of the taxa of interest identified by the 16S rRNA analysis, and a similar distribution of age and sex amongst the cases and controls. 10 control NPS samples and 9 case NPS samples were selected for metagenomic sequencing. These samples were classified correctly by the model at least 96% (controls) or 80% (cases) of the time. One MEF sample from each of the 9 selected cases was also sequenced, for a total of 28 samples undergoing metagenomic shotgun sequencing.

Two further samples also underwent metagenomic sequencing. To explore strain-
level variation in *A. otitidis* and *T. otitidis* DNA from the middle ear compared to that from the EAC, five pairs of MEF and ECS samples containing high relative abundances of these organisms in the metataxonomics data were combined into two pools; one MEF pool and one ECS pool. The pairs consisted of a MEF and an ECS sample obtained from the same ear of a child with no previous tympanic membrane perforations. They were sequenced in this way to provide insight on whether these organisms are present in the MEF due to contamination of the sample from the EAC, or if they inhabit both environments independently, given an intact tympanic membrane. As this is a within-child comparison, other variables did not need to be taken into account when selecting these samples. These samples were pooled due to the very low (sometimes unquantifiable by Qubit) DNA yield in the ECS samples. Thus, a total of 30 samples were prepared for metagenomic sequencing.

### 3.2.3 Removal of human DNA

Due to the high proportion of human DNA expected in low biomass clinical specimens, it was necessary to remove as much of this from the specimens as possible prior to sequencing. This was to avoid the generation of large volumes of redundant human sequence data that cannot be utilised in this study, and to ensure sufficient depth of sequencing for low-abundance bacterial species. The NEBNext Microbial DNA Enrichment Kit (New England Biolabs) was used to selectively remove human DNA from the total genomic DNA extracted from the samples. This kit contains magnetic beads bound to the protein MBD2-Fc, which binds specifically to CpG methylated DNA. On the premise that human DNA is heavily CpG methylated and prokaryotic DNA is not, the magnetic beads selectively remove human DNA (except unmethylated mitochondrial DNA) from solution. All samples except the ear canal pool underwent human DNA removal, as the total amount of DNA in this pool was minimal and may have been lost with further processing.

The NEBNext Microbiome DNA Enrichment Kit was used as per the manufacturer’s protocol, with a maximum of 1 µg total DNA as input. Samples above this amount were divided in half and recombined at the end of the protocol. Enriched microbial DNA was purified with Agencourt AMPure XP magnetic beads (Beckman Coulter) as described in the NEBNext kit protocol, with samples again divided in half if they exceeded the maximum input volume for this step of the protocol. Purified microbial DNA was resuspended in a final volume of 50 µl of low TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, FisherBiotec). The enriched microbial DNA was quantified before
and after purification with the Qubit 2.0 fluorometer (Invitrogen, dsDNA HS assay kit). Enriched samples were stored at -20°C in DNA lo-bind tubes (Eppendorf) prior to sequencing.

### 3.2.4 Metagenomic shotgun sequencing

The 30 samples were prepared for sequencing using the Nextera XT DNA Library Prep Kit (Illumina) according to manufacturer’s instructions. Samples were multiplexed in groups of 8, and these libraries were sequenced by the Australian Genome Research Facility (AGRF) on two lanes each of an Illumina HiSeq 2500 with 125 bp paired-end chemistry. The target sequencing yield per sample was 12.5Gb (minimum guaranteed output 50Gb per lane), based on a yield of 4-8Gb achieved in previous metagenomics studies\textsuperscript{316,317} and the estimation that no more than 50% of the remaining sequence would be human (determined from the quantification of total DNA pre- and post-human DNA removal).

### 3.2.5 Data pre-processing

Data were received from the sequencing centre in FASTQ format. The data from duplicate lanes was concatenated and read quality was assessed for each file with FastQC v0.11.3\textsuperscript{228} and summarised with MultiQC v1.0.\textsuperscript{318} Illumina adapter sequences were trimmed from the reads with Trimmomatic v0.36\textsuperscript{319} in paired-end mode using the NexteraPE adapter file. Reads shorter than 60 bp in length after trimming were discarded, and all remaining reads (pairs and orphans) that aligned with bowtie2 v2.3.3.1\textsuperscript{320} to a combined reference containing the human genome (GRCh38 patch 11) and PhiX genome (Enterobacteria phage phiX174 sensu lato) were removed. Trimmed reads that remained paired were aligned to these genomes via the \texttt{bowtie2_discordant_pairs.py} script from KneadData v0.6.1 (https://bitbucket.org/biobakery/kneaddata/wiki/Home) to remove read pairs that mapped concordantly, discordantly or as individual reads. After human and PhiX sequence removal, the remaining microbial reads were concatenated into one file to be treated as single-end reads as the downstream analyses do not use paired-end information. The quality of these microbial reads was again assessed by FastQC and MultiQC to ensure that the removal of the substantial proportion of human sequence reads did not enrich for poor quality sequences.
3.2.6 Taxonomy analysis

Case and control samples

To identify the taxa present in the samples, microbial sequence reads were analysed with MetaPhlAn2 v2.7.1 on default settings. MetaPhlAn2 uses a database that contains short clade-specific marker sequences to estimate the relative abundance of taxa. The MetaPhlAn2 profiles for each sample were merged into one file, and visualised with R package ggplot2 v3.0.0 in R v3.5.1.

MEF and ear canal pools

Sequences from the MEF and ECS pooled samples were compared to determine if the strains of *A. otitidis* and *T. otitidis* were identical in both sites (potentially indicative of contamination) or if they exhibited some variation (potentially indicative of separate habitats). Microbial reads from these samples were first aligned via bowtie2 to the genus-specific markers for *Alloioccocus* and *Turicella*, extracted from the MetaPhlAn2 marker database with StrainPhlAn. There were 200 marker sequences for each genus. The microbial reads from these samples had not been filtered for sequences matching the PhiX genome, however MetaPhlAn2 is robust to such sequences as the database contains only curated markers unique to these clades. The resulting alignment (SAM) file was sorted and converted to BAM format with SAMtools v1.3.1 and the alignments were viewed in IGV 2.3.72. Single nucleotide polymorphisms (SNPs) across the 200 marker sequences for each species were manually inspected in IGV to observe variation between the populations of reads in the MEF and ECS pools.

3.2.7 Functional analysis

Functional assignment of shotgun metagenomics data

Microbial reads were analysed with HUMAnN2 v0.11.1, (https://bitbucket.org/biobakery/humann2/wiki/Home) to identify the gene families and metabolic pathways present in the metagenome. Reads were aligned to the UniRef50 peptide database (--search-mode uniref50). The resulting UniRef50 gene families and MetaCyc metabolic pathway abundance tables for each sample were then renormalised from reads per kilobase (RPK) to copies per million (CPM) units,
excluding the reads that did not map to UniRef50. The gene families table was then regrouped into informative Gene Ontology (infoGO) terms. The infoGO and pathway abundance tables per sample were then merged into a single table each.

**Functional prediction from 16S rRNA amplicon data**

The functional capability of the nasopharyngeal and middle ear microbiome was also explored via metagenome prediction from the 16S rRNA amplicon data. The OTU table generated in Chapter 2 was analysed with Piphillin, which uses a nearest-neighbour approach to identify the OTU representative sequences and predict the metagenomic content from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (release 82, May 2017). The counts generated by Piphillin were normalised per sample to counts per million (i.e. a relative abundance out of 1 million), which is required for LEfSe analysis below. Predicted KEGG orthologs and pathways were produced by Piphillin.

**Functional differences between microbial communities**

To identify the features (functional assignments) most likely to explain differences between groups of samples, the linear discriminant analysis tool LEfSe v1.0.8 (via BioConda) was used to compare a) the case NPS to the control NPS samples, and b) the case NPS to paired case MEF samples; for both the shotgun metagenomics and 16S rRNA amplicon (predicted metagenome) datasets. These 8 LEfSe analysis models are described below. Prior to input into LEfSe, the number of features was reduced by removing those with low abundance (based on the median normalised value across all samples).

*Shotgun metagenomics dataset (functions assigned by HUMAnN2)*

- Case NPS (n = 9) vs. Control NPS (n = 10)
  - Model 1: InfoGO terms with median \( \geq \) 10 (524 terms)
  - Model 2: Metabolic pathways with pathway coverage (probability that a given pathway is present) > 50% and median \( \geq \) 0 (49 pathways)
- Paired samples: Case NPS (n = 9) vs. Case MEF (n = 9)
  - Model 3: InfoGO terms with median \( \geq \) 10 (509 terms)
  - Model 4: Metabolic pathways with pathway coverage > 50% and median \( \geq \) 0 (49 pathways)
Chapter 3. Exploration of the microbiome with shotgun metagenomics...

Predicted metagenome dataset (functions assigned by Piphillin from metataxonomics data)

- Case NPS (n = 87) vs. Control NPS (n = 101)
  - Model 5: KEGG Orthologs with median $\geq 100$ (1293 terms)
  - Model 6: KEGG Pathways with median $\geq 600$ (136 pathways)
- Paired samples: Case NPS (n = 75) vs. Case MEF (n = 75)
  - Model 7: KEGG Orthologs with median $\geq 100$ (1520 terms)
  - Model 8: KEGG Pathways with median $\geq 600$ (134 pathways)
  - Note that the median thresholds for models 7 and 8 were calculated across all case MEF and NPS samples (n = 214)

In the models comparing pairs of case NPS and MEF samples, when two MEF samples were available for a given patient, one MEF was chosen at random (sample Set 1). The analysis was repeated with the opposite ear (sample Set 2) and features that were significant ($p < 0.05$) in both analyses were retained. See Chapter 2, section 2.3.10 Differential abundance analyses for further detail on this approach. Heatmaps of the significantly discriminative features showing the normalised feature abundance and linear discriminant analysis (LDA) score (effect size) from these analyses were generated in R v3.5.1 with the package superheat v1.0.0.

3.3 Results

3.3.1 Characteristics of participants

Children under the age of 5 years were recruited into the biOMe study as either cases (diagnosed with rAOM) or as healthy controls (no history of rAOM despite being exposed to at least one of two major risk factors for OM; attendance at day care or living with siblings under 5 years of age). The characteristics of the subset of participants whose samples underwent shotgun metagenomics are shown in Table 3.1 below, for those whose samples were sequenced individually (n = 19, Table 3.1a) and those whose samples were sequenced as part of the MEF and ECS sample pools (n = 5, Table 3.1b). None of the participants identified as Aboriginal or Torres Strait Islander, or had a known current or previous spontaneous tympanic membrane perforation. One case whose samples were sequenced individually had previously had grommets inserted.
### Table 3.1: Demographic characteristics of participants.

This table includes all children who contributed at least one sample to a) shotgun metagenomics on individual samples or b) the MEF and ECS sample pools.

|                      | Case  
|----------------------|--------|
|                      | (n = 9) | Control  
|                      | (n = 10) | Pooled cases  
|                      | (n = 5) |
| **Age**              | Mean 2.2 years | Mean 2.2 years | Mean 1.6 years |
| **Sex**              | Female: 5 | Female: 6 | Female: 2 |
|                      | Male: 4 | Male: 4 | Male: 3 |
| **Season**           | Summer: 2 | Summer: 0 | Summer: 0 |
|                      | Autumn: 1 | Autumn: 3 | Autumn: 1 |
|                      | Winter: 3 | Winter: 4 | Winter: 4 |
|                      | Spring: 3 | Spring: 3 | Spring: 0 |
| **Ever breastfed**   | 7/9 | 10/10 | 5/5 |
| **Duration of breastfeeding** | Mean 2.9 months | Mean 11.4 months | Mean 6.8 months |
| **Currently attending day care or school** | 7/9 | 6/10 | 5/5 |
|                      | (1 unknown) |
| **Siblings of 5 years of age or younger** | 5/9 | 8/10 | 3/5 |
| **Smoker in the household** | 1/9 | 1/10 | 1/5 |
| **Antibiotic usage in the past month** | 0/9 | 0/10 | 4/5 |

*Excludes children that were breastfeeding at time of collection as breastfeeding has not ceased.*
3.3.2 Removing contaminating host DNA and sequences from nasopharyngeal swabs and middle ear fluids

Across the 28 individual samples (9 case NPS, 9 case MEF and 10 control NPS), an average of 19.6% (3.9% - 57.9%) of the original input DNA (by total nanograms calculated from DNA concentration quantified by Qubit 2.0) remained following processing by the NEBNext Microbiome DNA Enrichment Kit and purification by AMPure XP Beads. Two of the control NPS were below the quantifiable range after this processing. Shotgun metagenomic sequencing of these enriched and purified samples generated an average of 62.5 million raw paired-end reads per sample (15.6Gb). Following trimming of Illumina adapter sequences and removal of short reads, an average of 95.1% of read pairs remained, with an additional 3.9% surviving as orphans. The overwhelming majority of remaining reads then aligned to the human/PhiX reference, with 1.2 million – 13.6 million individual reads per sample (average 3.8 million, or 3.2% of the original raw reads) remaining for taxonomic and functional analysis. The human DNA sequences detected \textit{in silico} were not restricted to a particular region of the genome, containing both mitochondrial and chromosomal sequences. A summary of the proportions of sequence assignment by sample type is provided in Appendix B.

3.3.3 Species level taxonomy of bacteria and viruses

Sufficient microbial reads remained to explore the species-level taxonomic profiles of the 28 individual samples. The genus and species-level bacterial taxonomy for these samples generated by MetaPhlAn2 is shown in Figure 3.1, in comparison to the taxonomic profiles for these samples as determined by 16S rRNA amplicon sequencing in Chapter 2. Within these samples, 16S rRNA amplicon sequencing identified a total of 51 genera, of which 9 genera were present above an average relative abundance of 5% in at least one group (control NPS, case NPS or case MEF). In the shotgun metagenomics data, 41 bacterial genera and 10 viruses were identified by MetaPhlAn2. 8 bacterial genera were present above the 5% threshold. Overall, no new major genera were detected with shotgun metagenomics sequencing compared to 16S rRNA amplicon sequencing, and diversity remained low with each major genus containing 1-3 species.

In Chapter 2, the genera \textit{Corynebacterium}, \textit{Dolosigranulum} and \textit{Moraxella} (most likely \textit{M. lincolnii}) were enriched in the nasopharynx of healthy controls. These
3.2 Results

Figure 3.1: Comparison of taxonomic summaries by 16S rRNA amplicon sequencing and shotgun metagenomics sequencing. Genera with at least 5% average relative abundance in at least one sample type (control NPS, case NPS or case MEF) and their dominant species are shown.
remained the dominant genera in the healthy control samples in the metagenomics data, enabling species-level identification. The most dominant *Corynebacterium* species was *C. pseudodiphtheriticum*. Two of the ten control NPS samples contained *C. propinquum* or *C. accolens* as the dominant *Corynebacterium* species. The overall most abundant species in the control samples was *Dolosigranulum pigrum*, the only species in the genus *Dolosigranulum*. In the metataxonomic data, *Moraxella* was abundant (> 25%) in 9 of the 10 control samples and included both the *M. lincolnii* OTU and another more dominant *Moraxella* OTU (most likely catarrhalis). However, *Moraxella* abundance was greatly reduced in the control samples when sequenced by shotgun metagenomics and only *M. catarrhalis* was observed. *Moraxella* was absent or at < 1% estimated abundance in 7 of the 10 control NPS samples.

*Gemella* and *Neisseria* were the most abundant genera that were significantly enriched in the nasopharynx of cases, determined in Chapter 2. Despite all 9 selected case NPS samples containing *Gemella* above 1% relative abundance by 16S amplicon sequencing, the genus was observed above 1% estimated relative abundance in only two of these samples by shotgun metagenomics. Within these samples, two species were identified; *G. haemolysans* and an unclassified *Gemella* species. *Neisseria* was observed above 1% abundance by both sequencing methods in 7 of 9 case NPS, and predominantly contained *N. lactamica* and an unclassified *Neisseria* species which often occurred together in the same sample.

16S rRNA amplicon analysis in Chapter 2 also identified *Alloiococcus*, *Turicella* and *Staphylococcus* as the dominant taxa in the MEF of children with rAOM. *Alloiococcus otitidis* and *Turicella otitidis* (the only species in their respective genera) were confirmed to be abundant in the MEF microbiota by shotgun metagenomics, with *A. otitidis* most dominant. However, *Staphylococcus* was completely absent from all but one of the MEF samples despite a high (mean 20.7%) relative abundance across the 9 MEF samples by 16S rRNA profiling and presence in 4 of these 9 samples at 1% relative abundance or above. This one MEF sample contained an estimated 4.3% relative abundance of *S. haemolyticus*. *Haemophilus* was also a prevalent genus in the metataxonomic analysis of the MEF (≥ 1% relative abundance in 5 of 9 MEF samples). It was confirmed by shotgun metagenomics to be the otopathogen species *H. influenzae*, with one sample also containing a lower abundance of *H. aegyptius*. *H. influenzae* was also the most common species observed in the case NPS by shotgun sequencing.

Shotgun metagenomics sequencing is also capable of detecting DNA viruses, though many respiratory viruses are RNA viruses and cannot be detected by this method.
3.3. Results

In 13 of the 28 samples, at least 1% of the sample’s estimated relative abundance was attributed to viruses and these are shown in Table 3.2. The most common virus in the case NPS samples was a *Streptococcus* phage, EJ-1. While the host *Streptococcus* was also present in these samples, it was not always at high abundance. Viruses contributed to a higher amount of estimated relative abundance in the MEF than in the nasopharyngeal samples, and were more commonly detected in the case NPS than in control NPS.

3.3.4 Comparisons between the *A. otitidis* and *T. otitidis* sequences in the MEF and ECS sample pools

There was a substantial difference between the MEF and ECS sample pools in the number of reads aligning to the *Alloiococcus* and *Turicella*-specific marker sequences. There were 3,408,897 reads from the ECS pool which aligned across the 200 *Alloiococcus* markers, and 4,455,583 to the 200 *Turicella* markers. From the MEF pool, only 2016 and 4482 reads mapped to the *Alloiococcus* and *Turicella* markers respectively, resulting in only a few alignments per marker sequence. The level of human sequence contamination in the ECS pool was minimal, with over 90% of the read pairs remaining after alignment against the human reference genome. For the MEF pool, only 20% of read pairs remained. With such low coverage of markers compared to the ECS pool, it was therefore not possible to assess whether there were variable SNPs between the pools. For SNPs present in both pools, they were commonly the same SNP; though coverage was insufficient to provide any evidence for different *Alloiococcus* and *Turicella* strains in the MEF and ECS pools.

3.3.5 The functions of the nasopharyngeal and middle ear microbiome

The microbial reads in each of the 28 samples were given a functional annotation either by alignment to the pangenomes of the species detected by MetaPhlAn2 (via the ChocoPhlAn pangenome database packaged with HUMAnN2) or by translated search against the UniRef50 database, chosen due to the large percentage (>90%) of reads that remained unmapped when tested with the more stringent UniRef90 database. Across all 28 samples, only 37.9% of remaining microbial reads received a functional assignment (average 1,711,400 reads per sample).
Table 3.2: DNA viruses detected by shotgun metagenomics sequencing. Samples where a total of at least 1% of the estimated relative abundance was from the viral kingdom are included. Individual viruses with an abundance of at least 0.5% within the sample are shown.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Virus</th>
<th>Estimated relative abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control NPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8014</td>
<td>KI polyomavirus</td>
<td>6.60%</td>
</tr>
<tr>
<td>MGO012S1</td>
<td>WU polyomavirus</td>
<td>4.60%</td>
</tr>
<tr>
<td>MGO020S1</td>
<td>KI polyomavirus</td>
<td>12.00%</td>
</tr>
<tr>
<td>MGO038S1</td>
<td><em>Streptococcus</em> phage EJ1</td>
<td>2.10%</td>
</tr>
<tr>
<td></td>
<td>Human adenovirus B</td>
<td>1.10%</td>
</tr>
<tr>
<td>MGO065S1</td>
<td><em>Streptococcus</em> phage EJ1</td>
<td>11.10%</td>
</tr>
<tr>
<td></td>
<td>Torque teno virus</td>
<td>4.00%</td>
</tr>
<tr>
<td></td>
<td>Torque teno virus 7</td>
<td>3.90%</td>
</tr>
<tr>
<td>MGO071S1</td>
<td><em>Streptococcus</em> phage EJ1</td>
<td>6.00%</td>
</tr>
<tr>
<td>Case NPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGO091S1</td>
<td><em>Streptococcus</em> phage EJ1</td>
<td>6.00%</td>
</tr>
<tr>
<td>MGO019FL</td>
<td><em>Propionibacterium</em> phage PAS50</td>
<td>84.70%</td>
</tr>
<tr>
<td>MGO020FL</td>
<td>Human adenovirus B</td>
<td>0.80%</td>
</tr>
<tr>
<td></td>
<td>Torque teno virus</td>
<td>24.50%</td>
</tr>
<tr>
<td></td>
<td>Probionibacterium phage PAS50</td>
<td>23.00%</td>
</tr>
<tr>
<td>MGO38FL</td>
<td>Human adenovirus B</td>
<td>2.60%</td>
</tr>
<tr>
<td></td>
<td>Human adenovirus D</td>
<td>1.10%</td>
</tr>
<tr>
<td></td>
<td>Human herpesvirus 5</td>
<td>10.40%</td>
</tr>
<tr>
<td>Case MEF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGO063FR</td>
<td><em>Propionibacterium</em> phage PAS50</td>
<td>6.80%</td>
</tr>
<tr>
<td></td>
<td>Torque teno virus</td>
<td>2.40%</td>
</tr>
<tr>
<td>MGO065FR</td>
<td>TTV-like mini virus</td>
<td>84.80%</td>
</tr>
<tr>
<td></td>
<td>Torque teno virus</td>
<td>0.90%</td>
</tr>
<tr>
<td>MGO071FL</td>
<td>Torque teno virus 7</td>
<td>11.70%</td>
</tr>
<tr>
<td></td>
<td>Torque teno virus</td>
<td>1.50%</td>
</tr>
</tbody>
</table>
3.3. Results

Differences between the case and control nasopharyngeal microbiome

The genomic features most likely to explain the differences between the case NPS and control NPS samples are shown in Figure 3.2 with their LDA score (effect size) and normalised abundance in each sample. Figures 3.2a and 3.2b show the infoGO terms and MetaCyc pathways significantly enriched in either group, respectively. Some of the MetaCyc pathways should not be present in the microbiome; the mammalian coenzyme A pathway is found in humans, and the pyruvate fermentation pathway is engineered and does not occur naturally. This indicates that some human sequence may not have been removed, suggesting that these functional assignments may be spurious and represent noise rather than biological differences. A total of 29 infoGO terms with an LDA score over 2.5 were statistically significant (p < 0.05), with many of these enriched in the control NPS. Terms enriched in the control group corresponded to essential biological processes such as ATP binding, phosphorelay signal transduction system, tricarboxylic acid cycle and proton-transporting ATP synthase activity. In contrast, the terms enriched in the case NPS were at very low abundance in both groups, with the exception of ‘virion attachment to host cell’, which is substantially higher in one case NPS sample. The UniRef50 gene families that contribute reads to this infoGO category are the penton and fiber proteins from human adenovirus. Adenovirus was detected in this sample, though the total estimated viral abundance was small (3.3%). There were 8 MetaCyc pathways significantly enriched with an LDA score over 2.0. The pathways that were not spurious also corresponded to essential biological functions. These results suggest that there are no distinct differences in the functional capability of the case and control nasopharyngeal microbiomes.

Due to the very low coverage of microbial genomes with the shotgun metagenomics dataset, the 16S rRNA amplicon data from Chapter 2 was used to predict the functional content of the nasopharyngeal microbiome based on the KEGG functional database. Figures 3.2c and 3.2d show the KEGG orthologs and KEGG pathways identified by this method which are significantly enriched in either the cases or controls. For many of these functions, the pattern of abundance is very similar in both the case and control nasopharynx; few features are consistently lower or higher in one group compared to the other. One exception is the enrichment of transposase orthologs in some of the case NPS samples; a pattern not present in the control NPS. These case NPS samples tended to contain a high abundance of an OTU corresponding to *Neisseria*. 
a) InfoGO terms

Based on shotgun metagenomics data, a) Informative gene ontology terms with \( p < 0.05 \) and LDA score > 2.5 and b) MetaCyc pathways with \( p < 0.05 \) and LDA score > 2.0 are shown. Predicted from the 16S rRNA amplicon data, c) KEGG functional orthologs with \( p < 0.05 \) and LDA score > 2.3 and d) KEGG pathways with \( p < 0.05 \) and LDA score > 2.5 are shown. The normalised (copies or counts per million) abundances of each of these features in the samples are displayed in the heatmaps.
b) MetaCyc pathways

- PWY-6123: inosine-5’-phosphate biosynthesis I
- PWY-6124: inosine-5’-phosphate biosynthesis II
- PWY-6700: queuosine biosynthesis
- PWY-5103: L-isoleucine biosynthesis III
- PWY-7111: pyruvate fermentation to isobutanol (engineered)
- PWY-6387: UDP-N-acetylmuramoyl-pentapeptide biosynthesis I
- COA-PWY-1: coenzyme A biosynthesis II (mammalian)
- PWY-5686: UMP biosynthesis

Case NPS (n = 9) | Control NPS (n = 10)

LDA Score
c) KEGG ontology terms

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>K07482</td>
<td>transposase, IS30 family</td>
</tr>
<tr>
<td>ABC.CD.A</td>
<td>putative ABC transport system ATP-binding protein</td>
</tr>
<tr>
<td>K07488</td>
<td>transposase</td>
</tr>
<tr>
<td>ABC.CD.P</td>
<td>putative ABC transport system permease protein</td>
</tr>
<tr>
<td>ABC.FEVP</td>
<td>iron complex transport system permease protein</td>
</tr>
<tr>
<td>lacI, galR</td>
<td>LacI family transcriptional regulator</td>
</tr>
<tr>
<td>K07052</td>
<td>uncharacterized protein</td>
</tr>
<tr>
<td>opuC</td>
<td>osmoprotectant transport system substrate-binding protein</td>
</tr>
<tr>
<td>pqqJ</td>
<td>zinc protease [EC:3.4.24.-]</td>
</tr>
<tr>
<td>opuBD</td>
<td>osmoprotectant transport system permease protein</td>
</tr>
<tr>
<td>pcaC</td>
<td>4-carboxymuconolactone decarboxylase [EC:4.1.1.44]</td>
</tr>
<tr>
<td>ybgC</td>
<td>acyl-CoA thioester hydrolase [EC:3.1.2.-]</td>
</tr>
<tr>
<td>NRT, narK, narP, nasA; MFS transporter, NNP family, nitrate/nitrite transporter</td>
<td></td>
</tr>
<tr>
<td>E2.8.3.5B, scoB; 3-oxoacid CoA-transferase subunit B [EC:2.8.3.5]</td>
<td></td>
</tr>
<tr>
<td>E2.8.3.5A, scoA; 3-oxoacid CoA-transferase subunit A [EC:2.8.3.5]</td>
<td></td>
</tr>
<tr>
<td>K07182</td>
<td>CBS domain-containing protein</td>
</tr>
<tr>
<td>actP</td>
<td>cation/acetate symporter</td>
</tr>
</tbody>
</table>

Case NPS (n = 87) Control NPS (n = 101)

![LDA Score Plot](chart.png)
3.3. Results

d) KEGG pathways

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Case NPS (n = 87)</th>
<th>Control NPS (n = 101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphotransferase system (PTS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC transporters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fructose and mannose metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminoacyl-tRNA biosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quorum sensing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mismatch repair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycolysis / Gluconeogenesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptomycin biosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyoxylate and dicarboxylate metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of antibiotics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine, serine and threonine metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valine, leucine and isoleucine degradation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-component system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folate biosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butanoate metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfur relay system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of secondary metabolites</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LDA Score
Differences between paired middle ear and nasopharyngeal microbiomes

The genomic features most likely to explain the differences between pairs of case NPS and case MEF from the same child are shown in Figure 3.3a (infoGO terms) and 3.3b (MetaCyc pathways). Similar to the comparison between case and control NPS, the MetaCyc pathways are either common to all bacteria or originating from organisms that should not be present, again indicating potentially spurious results or the absence of detectable functional differences between these sites. A total of 15 infoGO terms with an LDA score over 3.0 were significantly enriched (p < 0.05). Each one of these was enriched in the case NPS; the only feature enriched in the case MEF was the ‘ungrouped’ category (UniRef50 gene families that did not correspond to an infoGO term). This suggests that the MEF contains a large proportion of unknown functions. As observed in the comparison between case and control NPS samples, many of the features enriched in the case NPS represent functions essential for life including ATP binding, ribosomal components and cell membrane components. Some ion binding functions (magnesium, zinc and 4 iron, 4 sulfur cluster) were enriched in the case NPS compared to the MEF. Ten MetaCyc pathways were significantly enriched in the 9 case NPS samples, with none enriched in the case MEF.

Figures 3.3c and 3.3d show the KEGG functional orthologs and KEGG pathways most likely to explain the differences between the paired case MEF and case NPS. In both sets of features, very similar patterns of abundance are seen; with no functional orthologs or pathways displaying a consistent, substantial difference in abundance between the two groups despite moderate effect sizes. The significantly enriched features again represent functions shared by all prokaryotic life, indicating no distinct functional differences between the groups.
a) InfoGO terms

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0015935</td>
<td>[CC] small ribosomal subunit</td>
</tr>
<tr>
<td>GO:0015934</td>
<td>[CC] large ribosomal subunit</td>
</tr>
<tr>
<td>GO:0030170</td>
<td>[MF] pyridoxal phosphate binding</td>
</tr>
<tr>
<td>GO:0009279</td>
<td>[CC] cell outer membrane</td>
</tr>
<tr>
<td>GO:0009252</td>
<td>[BP] peptidoglycan biosynthetic process</td>
</tr>
<tr>
<td>GO:0003924</td>
<td>[MF] GTPase activity</td>
</tr>
<tr>
<td>GO:0008360</td>
<td>[BP] regulation of cell shape</td>
</tr>
<tr>
<td>GO:0000049</td>
<td>[MF] RNA binding</td>
</tr>
<tr>
<td>GO:0051539</td>
<td>[MF] 4 iron, 4 sulfur cluster binding</td>
</tr>
<tr>
<td>GO:0008270</td>
<td>[MF] tRNA binding</td>
</tr>
<tr>
<td>GO:0008360</td>
<td>[BP] regulation of cell shape</td>
</tr>
<tr>
<td>GO:0009270</td>
<td>[MF] zinc ion binding</td>
</tr>
<tr>
<td>GO:0002877</td>
<td>[MF] magnesium ion binding</td>
</tr>
<tr>
<td>GO:0003735</td>
<td>[MF] GTP binding</td>
</tr>
<tr>
<td>GO:0003735</td>
<td>[MF] structural constituent of ribosome</td>
</tr>
<tr>
<td>GO:0015934</td>
<td>[CC] large ribosomal subunit</td>
</tr>
<tr>
<td>GO:0015935</td>
<td>[CC] small ribosomal subunit</td>
</tr>
</tbody>
</table>

Figure 3.3: Enriched functional categories in the middle ear and nasopharynx of children with rAOM. Based on shotgun metagenomics data, a) Informative gene ontology terms with p < 0.05 and LDA score > 3.0 and b) MetaCyc pathways with p < 0.05 and LDA score > 2.0 are shown. Predicted from the 16S rRNA amplicon data, c) KEGG functional orthologs and d) KEGG pathways each with p < 0.05 and LDA score > 2.5 are shown. The normalised (copies or counts per million) abundances of each of these features in the samples are displayed in the heatmaps. Note that the ‘UNGROUPED’ term (in a) and ‘Metabolic pathways’ (in d) are not shown as their abundance was very high in both sample groups.
b) MetaCyc pathways

<table>
<thead>
<tr>
<th>Pathway Code</th>
<th>Pathway Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWY-6123</td>
<td>inosine-5'-phosphate biosynthesis I</td>
</tr>
<tr>
<td>PWY-6124</td>
<td>inosine-5'-phosphate biosynthesis II</td>
</tr>
<tr>
<td>CDA-PWY</td>
<td>coenzyme A biosynthesis (mammalian)</td>
</tr>
<tr>
<td>PWY-5103</td>
<td>L-isoleucine biosynthesis III</td>
</tr>
<tr>
<td>BRANCHED-CHAIN-AA-SYN-PWY</td>
<td>superpathway of branched amino acid biosynthesis</td>
</tr>
<tr>
<td>PWY-6147</td>
<td>6-hydroxymethyl-dihydropterin diphosphate biosynthesis I</td>
</tr>
<tr>
<td>PWY-2942</td>
<td>L-lysine biosynthesis III</td>
</tr>
<tr>
<td>PWY-7539</td>
<td>6-hydroxymethyl-dihydropterin diphosphate biosynthesis III (Chlamydia)</td>
</tr>
<tr>
<td>PWY-3001</td>
<td>superpathway of L-isoleucine biosynthesis I</td>
</tr>
<tr>
<td>THRESYN-PWY</td>
<td>superpathway of L-threonine biosynthesis</td>
</tr>
</tbody>
</table>

Case MEF (n = 9) | Case NPS (n = 9)
c) KEGG ontology terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC−2.A</td>
<td>ABC−2 type transport system ATP−binding protein</td>
</tr>
<tr>
<td>ABC.FE.V.P</td>
<td>Iron complex transport system permease protein</td>
</tr>
<tr>
<td>ABC.FE.V.S</td>
<td>Iron complex transport system substrate−binding protein</td>
</tr>
<tr>
<td>rpoE</td>
<td>RNA polymerase sigma−70 factor, ECF subfamily</td>
</tr>
<tr>
<td>K07090</td>
<td>Uncharacterized protein</td>
</tr>
<tr>
<td>mqp</td>
<td>Malate dehydrogenase (quinone) [EC:1.1.5.4]</td>
</tr>
<tr>
<td>gluQ, uguQ</td>
<td>Glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46]</td>
</tr>
<tr>
<td>fhaB</td>
<td>Filamentous hemagglutinin</td>
</tr>
<tr>
<td>tRNA−Ile; tRNA Ile</td>
<td></td>
</tr>
<tr>
<td>tRNA−Leu; tRNA Leu</td>
<td></td>
</tr>
<tr>
<td>RNA−Ala; tRNA Ala</td>
<td></td>
</tr>
<tr>
<td>hemoglobin/transferrin/lactoferrin receptor protein</td>
<td></td>
</tr>
</tbody>
</table>
d) KEGG pathways

- Two-component system
- ABC transporters
- Microbial metabolism in diverse environments
- Quorum sensing
- Carbon metabolism
- Pyruvate metabolism
- Glycerolipid metabolism
- Valine, leucine and isoleucine degradation
- Propanoate metabolism
- Glycolysis / Gluconeogenesis
- Glyoxylate and dicarboxylate metabolism
- Citrate cycle (TCA cycle)
- Tryptophan metabolism
- Pentose phosphate pathway
- Arginine and proline metabolism
- Arginine biosynthesis
- Phenylalanine, tyrosine and tryptophan biosynthesis
- Amino sugar and nucleotide sugar metabolism
- Biosynthesis of amino acids
- Homologous recombination
- Nicotinate and nicotinamide metabolism
- Pyrimidine metabolism
- Lipopolysaccharide biosynthesis
- Oxidative phosphorylation
- Ribosome
- Aminoacyl-tRNA biosynthesis

| Case MEF (n = 75) | Case NPS (n = 75) |

LDA Score

10000 30000 40000

-2 0 2
3.4 Discussion

The microbiome of the upper respiratory tract in children with OM has not been well characterised. Several studies, including the work described in Chapter 2, have explored the microbiome of the middle ear or nasopharynx using 16S rRNA profiling to describe taxonomic composition.\textsuperscript{109,168,169,171,172,174,176,178,267,289,290,313} This method is limited to genus-level profiling and is prone to PCR and gene copy number bias; issues which can be overcome by the use of metagenomic shotgun sequencing. This technique has not previously been applied to the upper respiratory tract of children with OM, and this chapter describes the first use of the technique to characterise human middle ear and nasopharyngeal microbiomes. The aims of this work were to obtain species-level resolution of taxonomic profiles, providing a more detailed picture of the bacterial and viral communities in these sites; and to identify the major differences in the functional capabilities of these microbiota to explore the possible roles of major bacterial genera identified in Chapter 2.

3.4.1 Removal of human DNA from specimens with high host contamination: a persistent challenge

The most striking finding was the overwhelming proportion of human DNA in MEF and NPS samples. Despite the removal of approximately 80% of the total DNA in each sample with the NEBNext Microbiome DNA Enrichment Kit prior to sequencing, an average of 97% of the adapter-trimmed sequence reads were unusable due to their alignment to the human or PhiX reference genome. With substantially lower coverage of the metagenome than expected, the analysis of this data was challenging and the results must be interpreted with caution. On average, only 3.8 million total reads per sample remained. Covering about 0.475 Gb of sequence (125 bp read length x 3.8 million reads), this is far below the 4-8 Gb of data recommended by other metagenomics studies on which the depth of sequencing for this study was based.\textsuperscript{316,317} Thus, \textit{de novo} metagenome assembly and annotation was not attempted, and the analysis was instead restricted to short-read alignment to reference databases containing previously characterised genomes. This was ideal for taxonomic identification, but was sub-optimal for the characterisation of genomic functions.

The proportion of host DNA was similarly high in other studies that have undertaken shotgun sequencing in the upper respiratory tract. The Human Microbiome Project
observed values in excess of 95% in the anterior nares of healthy individuals, and virome studies in the nasopharynx of children have observed similar values over 90%. The inefficiency of microbial DNA enrichment is surprising, though at least one other study conducted since this work has also observed a substantial proportion of host reads even after the use of this particular commercial kit. The necessity of attempting to deplete host DNA prior to sequencing has also precluded the ability to identify microbial eukaryotes in this environment, leaving this aspect of the microbiome poorly characterised.

3.4.2 Bacterial taxonomy – shotgun metagenomics compared to 16S rRNA profiling

Comparing the samples that have undergone both 16S rRNA amplicon sequencing and shotgun metagenomics, the mean relative abundance of bacterial genera in each sample type was similar. There were no genera above 1% relative abundance detected by shotgun metagenomics that were not also detected by 16S rRNA profiling. This may indicate that while the latter method has limitations, it may be sufficient to capture the vast majority of bacteria in the upper respiratory tract at genus level. However, it may also be a product of the small volume of shotgun metagenomics data remaining for analysis; less common genera may not have been detectable. Importantly, shotgun metagenomics allowed the species-level identification of key members of the genera of interest identified in Chapter 2. These were Corynebacterium, Dolosigranulum and Moraxella, which were significantly enriched in the healthy nasopharynx; and Alloiococcus, Turicella and Staphylococcus, potentially novel otopathogens which were highly abundant in the MEF and ECS.

Corynebacterium, Dolosigranulum and Moraxella: characteristic of a healthy nasopharynx

These three genera were significantly enriched in the nasopharynx of healthy controls in the 16S rRNA work described in Chapter 2. Here, C. pseudodiphtheriticum was identified as the dominant Corynebacterium species, with C. propinquum and C. accolens also dominant in one control NPS sample each. These species have previously been demonstrated to exhibit beneficial effects, with C. accolens capable of producing antipneumococcal compounds and viable C. pseudodiphtheriticum demonstrating improved resistance to respiratory syncytial virus and S. pneumoniae.
infection in mice by boosting innate immunity.\textsuperscript{331} \textit{D. pigrum} was confirmed as the only \textit{Dolosigranulum} species present (the only known species in its genus). The relative abundance of \textit{Moraxella} was reduced in the shotgun metagenomics data compared to the metataxonomics data. In Chapter 2, the \textit{Moraxella} OTU enriched in healthy controls matched \textit{M. lincolnii} most closely by BLASTN. However, only \textit{M. catarrhalis} was present in the shotgun metagenomics data despite moderate abundances of the presumed \textit{M. lincolnii} OTU in the selected control NPS samples. This species designation could therefore not be confirmed. The genus itself was also absent from some control NPS where it was detected at high abundance by 16S rRNA profiling, though other samples did contain \textit{Moraxella} at similar abundance in both datasets. As MetaPhlAn2 only estimates organism abundance based on the coverage of specific genomic markers, a variation in the small number of reads remaining for microbial analysis may produce large variations in abundance estimates. Each profile generated by the tool must add up to 100%; when the coverage across all organisms is low, there may not be sufficient data available to accurately estimate the taxonomic composition. It is therefore possible that this influenced the abundance of \textit{M. catarrhalis}, and additional species present at low abundance did not produce sufficient coverage of the species-specific marker sequences to be detected by MetaPhlAn2.

\textbf{Alloiococcus, Turicella and Staphylococcus: dominant in the OM-prone middle ear}

The genera \textit{Alloiococcus} and \textit{Turicella} each contain one known species, \textit{A. otitidis} and \textit{T. otitidis}, which were confirmed to be the only species from these genera present in the MEF samples. However, the abundance of \textit{Staphylococcus} in the MEF samples was substantially reduced in the shotgun data compared to the 16S rRNA amplicon data (average ~21\% to < 1\%). \textit{S. haemolyticus} was the only \textit{Staphylococcus} species present in the MEF, observed in one sample. This species colonises the skin and is amongst the coagulase-negative staphylococci that cause opportunistic skin and soft tissue infections,\textsuperscript{332} but it has not previously been associated with OM. More commonly, \textit{S. aureus} is reported in MEF samples.\textsuperscript{20}

The substantially reduced abundance of \textit{Staphylococcus} in the shotgun data compared to the metataxonomics data may be the product of overestimated abundance by metataxonomics and the shallow depth of the microbial shotgun sequencing. \textit{Staphylococcus} species tend to have 5-6 copies of the 16S rRNA gene.\textsuperscript{296} This could be a greater number of copies than \textit{Alloiococcus} and \textit{Turicella} have, as copy numbers for these genera are currently unknown. The abundance of \textit{Staphylococcus} in 16S
rRNA amplicon data may also be inflated by primer bias, where the 16S rRNA gene primers could amplify *Staphylococcus* sequences more efficiently than those of the other genera. Similar to *Moraxella*, it is also possible that the small volume of shotgun sequencing data remaining after removal of human sequences has resulted in very low coverage of MetaPhlAn2’s species-specific marker genes. *Staphylococcus* may have been missed by chance, where the low volume of data results in a smaller chance that reads covering species-specific marker genes will be present.

### 3.4.3 Identifying strain variation in *Alloiococcus* and *Turi-cellae*

While several studies have described the presence of *A. otitidis* and *T. otitidis* in MEF samples, it remains unclear whether the organisms actually reside within the middle ear. Unlike the known otopathogens, they are rarely observed in the nasopharynx and are present at high abundance in the EAC. The microbiome of the healthy middle ear is difficult to study and appears to contain minimal biomass where *A. otitidis* and *T. otitidis* have not been consistently found. Determining their role in OM relies on the knowledge of whether these species are indeed inhabiting the MEF, or if their presence in these samples is a result of contact with the EAC during sampling. Unfortunately, the issue of substantial host sequence contamination in the MEF samples made this impossible to determine as very few reads mapped to the marker genes.

A more effective approach would be to sequence several cultured isolates of *A. otitidis* and *T. otitidis* from each site; genomic data from a pure culture is more easily assembled and this would also improve the characterisation of these organisms in reference databases such as NCBI’s RefSeq, where these genomes remain in an incomplete stage. Studies focusing on *A. otitidis* and *T. otitidis* may additionally take care to sterilise the EAC prior to sampling of the MEF and carefully avoiding contact with the EAC. Understanding whether these organisms are even inhabiting the MEF – and how they get there – is integral to understanding their role in the development of OM.

### 3.4.4 Viral taxonomy

Another advantage of shotgun metagenomics over metataxonomics is that it is not bacteria-specific and will also detect DNA viruses. Other eukaryotes may also be
detected, however in this study as the samples were pre-treated to remove CpG-methylated DNA, any microbial eukaryotes present would also have been removed. Similar to the results from the respiratory virus detection panel in Chapter 2, viruses were detected less often in the nasopharynx of healthy control children than of those with rAOM. This suggests that nasopharyngeal colonisation with DNA or RNA viruses is either directly or indirectly associated with rAOM. By metagenomics, the most commonly detected DNA virus in the case NPS was a Streptococcus bacteriophage, EJ-1. Described as a temperate phage isolated from an atypical strain of S. pneumoniae, it is capable of integrating into the bacterium’s genome and has previously been detected in the nasopharynx of children with pneumonia. Its effect on S. pneumoniae in relation to disease appears to be unknown. Polyomaviruses were also observed in the NPS, and was the only viral group detected in the nasopharynx of healthy controls (one sample). The KI and WU polyomaviruses are closely related and both found in respiratory secretions, and have been frequently observed coinfecting with other viruses. Polyomavirus was also reported in a previous cohort of Australian children with and without rAOM, where they were found commonly in both groups and appeared to have limited association with rAOM.

A substantial portion of the average relative abundance in the MEF samples was attributed to viruses. The majority of estimated abundance came from Propionibacterium phage PAS50 and TTV-like mini virus, which were detected at high abundance in one MEF sample each and lower abundance in others. Propionibacteriaceae were detected in the MEF sample in which this phage was most abundant, but this family was not common amongst the MEF. This was very dissimilar to the metataxonomic profile of this MEF sample, which was instead predominantly Staphylococcus. The phage has also previously been reported in the nasopharynx of children with pneumonia. TTV-like mini virus was named due to its similarity to the Torque Teno virus (TTV), which was also detected in the case MEF and NPS. TTV has previously been associated with a wide range of diseases including observations of increased replication in inflamed nasal epithelium. However, as TTV is a relatively ubiquitous virus present in over 90% of healthy adults, it has been difficult to establish any causative role in disease. It has been observed in metagenomic virome studies of the nasopharynx, where it tended to be ubiquitous and was not associated with disease.

It is important to remember that shotgun metagenomics is restricted to DNA viruses; this is a selective and potentially underestimated view of the virome, as many
respiratory viruses (including most of those detected in Chapter 2) have an RNA genome and cannot be detected by metagenomic DNA sequencing. Other studies that have focused on viral detection by metagenomics have isolated RNA from the specimen separately for sequencing. The samples included in this study were not adequately prepared nor of sufficient volume for RNA isolation, which is a non-trivial task for MEF samples in particular (see Chapter 5).

3.4.5 Characterising microbiome function with shotgun metagenomics

Functional characterisation of the microbiome in nasopharyngeal and middle ear samples is new territory, and is difficult to interpret; primarily due to the very low coverage of microbial genomes and the lack of prior knowledge about these microbial communities. There appears to be a large portion of unknown function in these communities, as only ~38% of reads per sample received a functional assignment after a generic translated search to UniRef50 (an improvement upon the ~10% of reads mapping to the more stringent UniRef90 database). Furthermore, similarity to UniRef50 families can only give a broad overview of function rather than indicate specific genes that are present in the metagenome; the database is generic with sequences matched at a 50% identity threshold to proteins clustered at 50% similarity. The pangenomes of the bacterial species most substantially different between sample groups (C. pseudodiphtheriticum, D. pigrum, A. otitidis and T. otitidis) are not complete, with a whole genome sequence available for only two isolates of each species, each of which are in scaffold or contig (not complete) stage in RefSeq. Thus, the lack of prior knowledge about these genomes likely contributes to the functions considered ‘unknown’ and also to the apparent difference in functional content. Furthermore, the sequences without assigned function may originate from uncharacterised organisms that cannot be detected by software reliant on alignment of reads to a reference, like MetaPhlAn2 and HUMAnN2.

For both infoGO terms and MetaCyc metabolic pathways, most of the significantly enriched features were conserved functions common to all prokaryotic life. As all bacterial genomes should carry these functions, this enrichment may be a result of lower coverage in one group rather than of true differences in genome content. For example, the case MEF samples were significantly enriched in gene families which did not fall into an informative Gene Ontology category (‘ungrouped’). Many features were apparently enriched in the case NPS compared to the MEF, and are likely to be
due to the absence of reads corresponding to these functions in the MEF or missing functional annotations in A. otitidis and T. otitidis rather than the absence of the functions themselves.

Few previous studies have employed metagenomics in the nasopharynx, with most focusing on viral taxonomy. A small number of studies have begun to explore microbial functions within the nasopharynx with metagenomics. These studies do not report the proportion of metagenomic sequencing reads that were human, but have associated some microbial metabolic pathways with asthma, severe bronchiolitis and RSV or RV infection in children; the latter two studies also relied on accompanying metabolomics data. Early metatranscriptomics work in the nasal epithelium has indicated a potential role for Moraxella gene expression to produce differences in microbial metabolic function in children with asthma, though this has not yet been explored in detail. Ultimately, the functional capability of the nasopharyngeal microbiome remains largely uncharacterised and difficult to study via direct sequencing, due to issues with human sequence contamination, sequences lacking functional annotation and the current reliance on reference databases for this incompletely characterised community.

3.4.6 Predicting microbiome function from 16S rRNA amplicon data

An alternative approach to identifying differences in microbiome function was to infer KEGG orthologs and pathways from the metataxonomic data with the Pip Hilin tool. This circumvented the issue of low sequencing depth and added power to the analysis by the use of all NPS and MEF samples in the biOMe study. However, this approach becomes less direct as the organisms actually present are not directly sequenced and the inference again relies on what is already known and present in reference databases. These results were also unclear, with many highly conserved functions deemed most likely to explain the differences between groups. Additionally, the patterns of abundance were often similar in both groups, suggesting that there is not a consistent difference in abundance of these functions between groups. This may indicate that there are no substantial differences in function when comparing only the bacterial portion of the microbiota, using what is currently known about the organisms present. The only distinct functional difference was in the abundance of transposase KEGG orthologs. Transposases were predicted at higher abundance in the genomic content of the case NPS microbiome, perhaps related to horizontal transfer.
of virulence or antibiotic resistance genes amongst the otopathogens. However, these samples each had a high abundance of the *Neisseria* OTU. *Neisseria* genomes contain many transposon-like Correia elements; potentially ancestral sequences\(^{344}\) which are thought to regulate gene expression, including that of virulence genes.\(^{345}\)

Pipihillin performs best on well-characterised microbiomes, such as human clinical specimens.\(^{293}\) However, as the major differentially abundant genera in this study are not well-characterised they may have been missed from this analysis. Only 40% of the 123 OTUs input into Pipihillin were given a match in the 16S rRNA database used to identify the most closely-related reference sequence, supporting the notion that the nasopharyngeal and middle ear microbiomes have not been well-characterised and indicating that the most important genera may have been excluded from functional prediction.

Regarding the identification of functional differences between microbial communities, methods for statistical analysis of microbiome data are still in their infancy. LEfSe was designed for the discovery of features most likely to explain the differences between groups, however it does not appear to implement multiple testing correction within its statistical tests, and cannot control for confounding factors such as age or sex. DESeq\(^{346}\) and metagenomeSeq\(^{248}\) designed to identify differentially abundant features between groups, were developed to achieve robust statistical analysis with improved normalisation approaches and the ability to account for confounding factors. However, Hawinkel et al. (2017) notes that many methods for differential abundance analysis of microbiome data, including these, are prone to producing alarming rates of false positives; leaving such data challenging to interpret at present.\(^{347}\)

### 3.4.7 Towards further characterisation of the upper respiratory tract microbiome

Metagenomic assembly is the approach often taken in environments where the members of the microbial community are largely unknown or uncharacterised, such as in the cow rumen\(^{298}\) and extreme environments\(^{348,349}\) as it does not require the sequences present in samples to align to already-known sequences in reference databases. Annotation of function in assembled metagenomes does however rely on homology to known, publically available sequences. Based on the challenges observed in this chapter, two major areas for improvement may facilitate further metagenomics work in the upper respiratory tract.
First, the sequencing depth of these microbial communities must be increased if *de novo* assembly of metagenomes is to be made possible. The ‘over-sequencing’ of heavily human contaminated libraries is not economical; while a higher number of microbial sequences should be produced, there will be an overwhelming proportion of unusable human sequences to be discarded and the proportion of poor quality sequences may increase. Ideally, the removal of human DNA prior to sequencing should be prioritised. The NEBNext Microbiome DNA Enrichment Kit appears to perform poorly for microbial enrichment in the nasopharynx, as observed in this study and recently by Romero-Espinoza *et al.* (2018). Repeating enrichment with the kit may improve bacterial yield, but this is not viable for low biomass samples due to the loss of DNA during handling.

Other methods for human DNA removal have been developed, and involve implementation at the point of DNA extraction. The MolYsis Basic 5 kit selectively lyses human cells, degrading human DNA with a DNase enzyme prior to lysis and DNA extraction from bacterial cells. This kit performed well in synovial joint fluids and has demonstrated substantially improved microbial enrichment in this specimen type when compared to the NEBNext Microbiome DNA Enrichment Kit. Other selective lysis methods utilising benzonase, saponin or propidium monoazide for the degradation of human DNA have also been successful in enriching microbial DNA from respiratory samples. In the current study, the NEBNext kit was the only available option as it is currently the only method that enriches microbial DNA from a pool of mixed genomic DNA. With a selection of methods available for microbial enrichment at the point of DNA extraction, optimisation of human DNA removal in nasopharyngeal and middle ear fluid samples is essential for further deep sequencing attempts in the upper respiratory tract. This should include the detection of microbial DNA that may be removed in the process, and particularly whether this results in the biased removal of certain organisms.

The second area for improvement is in the quality and quantity of reference genomes for nasopharyngeal and middle ear organisms. The two health-associated commensals (*C. pseudodiphtheriticum* and *D. pigrum*) and two potential novel otopathogens (*A. otitidis* and *T. otitidis*) identified in this work have not had many isolates sequenced and their genome assemblies are incomplete and therefore poorly annotated. Whole genome sequencing efforts focused on characterising these individual species will provide a better understanding of their genomic content, towards an understanding of their role in the upper respiratory tract. This may be a more effective way to study these organisms, as well as contributing to the coverage of their genomes in
public reference databases. Furthermore, microbiome research is largely exploratory
and hypothesis-generating; any biomarkers or features of interest discovered via
exploration of the microbiome should be validated with more direct measurements
to test hypotheses and validate findings.

3.5 Conclusion

The use of shotgun metagenomics sequencing revealed the key commensal species in
the nasopharynx of healthy controls; *C. pseudodiphtheriticum* and *D. pigrum*. These
organisms are of interest for further research into their possible role in protection
against the development of rAOM. However, shotgun metagenomics sequencing of
the nasopharynx and middle ear remains highly challenging due to the overwhelming
proportion of host DNA. When insufficient data remains to conduct *de novo* metage-
nomic assembly, methods that rely on alignment to existing genome databases must
be used. This is problematic as the low coverage of genomes can produce spurious or
incomplete results and existing reference databases do not contain sufficient informa-
tion on organisms that are poorly characterised, such as *C. pseudodiphtheriticum*,
*D. pigrum*, *A. otitidis* and *T. otitidis* – the major genera of interest in this study.
The removal of host sequences from NPS and MEF samples must be carefully op-
timised prior to further work in this area as this is currently the major barrier
to deep sequencing in the upper respiratory tract. Further characterisation of the
genomes of these individual species would also assist in understanding their role in
the development of rAOM.
Chapter 4

*In vitro* assessment of inhibitory activity by the commensal species *C. pseudodiphtheriticum* and *D. pigrum*

In this chapter, the species *C. pseudodiphtheriticum* and *D. pigrum* (shown in Chapters 2 and 3 to be characteristic of the rAOM-resistant nasopharyngeal microbiome) are tested *in vitro* for inhibitory activity. Agar plate assays are undertaken to observe bacterial interference between these species and clinical otopathogen strains to investigate the mechanisms by which these organisms may contribute to resistance to the development of rAOM.
4.1 Introduction

Many studies of the human microbiome have focused on the difference in microbiome composition between a healthy and a disease state. This research has naturally progressed into ‘mining the microbiota for therapeutics’; the process of identifying bacterial strains or molecular products within a healthy microbiota for development as therapeutic agents, either in the form of probiotics (live, beneficial microorganisms) or isolated antimicrobial compounds. The concept of deriving therapy from the microbiome may be most widely known in the form of faecal microbiota transplants, where a healthy donor microbiota is transplanted into an individual with relapsing *C. difficile* disease. This treatment has been extremely effective, with resolution of infection in 81% of patients after the first infusion compared to 23-31% in control groups in the initial clinical trial.

An approach to identifying the individual members of the microbiota that contribute to health has recently been outlined in a proposed variant of Koch’s postulates. Neville et al. (2018) suggest that to establish an individual commensal strain as a causative promoter of health, it must be 1) associated with host health, found regularly in healthy hosts but less often in disease; 2) possible to isolate as a pure culture; 3) capable of reducing disease when introduced to a new host; and 4) detected following introduction to the new host. Several examples mirroring this process have already been demonstrated. In mice, Lawley et al. (2012) developed a defined mixture of six commensal bacteria and used it to successfully resolve experimentally-induced *C. difficile* disease. Brugiroux et al. (2016) developed a small community, informed by the dominant phyla and functions of a healthy murine gut, that provided resistance to colonisation by *Salmonella enterica* serovar Typhimurium. In human volunteers, *Staphylococcus epidermidis* isolated from healthy people was shown to produce the Esp protease which inhibits nasal colonisation and the formation of biofilm by *S. aureus*.

To provide a better understanding of the effect of commensal bacteria in preventing or relieving disease, an additional stage of research may be carried out between postulate 2 (isolating the health-associated commensals in pure culture) and postulate 3 (introducing them to a diseased host to resolve or mitigate disease): an *in vitro* assessment of bacterial interference. This approach would be a useful addition to the process, to determine whether the commensal strains compete with pathogenic species. It offers a relatively simple and inexpensive way to begin to understand the mechanism by which commensal strains may contribute to health, prior to moving
4.1. Introduction

into animal models or human clinical trials.

The concept of bacterial interference is not new, with several early studies occurring in the late 19th century, reviewed by Brook in 1999. Since that time, methods have been developed to detect the various mechanisms of competition which probiotic bacteria are now understood to use, including the production of antimicrobial substances, competitive adhesion, induction of mucin production and modulation or stimulation of the immune system. Besides the use of an in vivo model, the use of in vitro co-culture of a candidate probiotic strain, pathogenic strain and host cells can be used to experimentally detect these mechanisms of bacterial interference. However, the production of antimicrobial substances can be detected via simple bacteriological experiments on agar if a host factor is not required for the production or action of that substance; it may be considered the ‘simplest’ probiotic mechanism to detect. Genes encoding bacteriocins (antimicrobial proteins or peptides) may also be identified from genomes or metagenomes in silico using tools such as BAGEL3, (now BAGEL4) or antiSMASH. This approach can be useful for members of the microbiota which are difficult to culture, and has successfully been used to identify lactocillin genes in the vaginal microbiota and bacteriocin gene clusters in the gut microbiota from species not previously associated with bacteriocin production.

The antimicrobial substances that bacteria produce can be broadly classified into low molecular weight compounds (e.g. lactic and acetic acid) and bacteriocins (proteins or peptides; e.g. colicin, nisin). Lactic acid, produced by lactic acid bacteria (LAB) such as Lactobacillus can act on Gram negative bacteria, disrupting their outer membrane. Some compounds produced by LAB have previously been used as food preservatives. Bacteriocins can have broad-spectrum activity; for example nisin, produced by Lactobacillus lactis is effective against a range of Gram positive, and some Gram negative organisms. Others have a very narrow range, targeting only closely related organisms; colicin (produced primarily by Escherichia coli) and many bacteriocins that fall into the same class (class IIa, determined by their structure) act in this way.

Bacterial interference does occur in the upper respiratory tract (URT), including interference by antimicrobial substances. Alpha-haemolytic streptococci demonstrate good inhibitory activity via the production of bacteriocins and have shown promise in clinical trials (particularly Streptococcus salivarius) as a probiotic to reduce the risk of AOM. Recently, advances have been made on bacterial interference in the context of otitis media. H. haemolyticus was recently found to
produce what is thought to be a novel bacteriocin that inhibits NTHi, currently the major otopathogen in Australia. The substance is specific to *Haemophilus* species and seems to be distinct from haemocin, another bacteriocin produced by *Haemophilus*. Bomar et al. (2016) identified *Corynebacterium* and *Dolosigranulum* as overrepresented genera in children that were negative for pneumococcal colonisation. They chose to study *C. accolens* in vitro due to its prevalence in the nose of adults and children and because there were fewer reports implicating it in human infection than there were for *C. pseudodiphtheriticum* or *C. propinquum*. A lipophilic *Corynebacterium* species, *C. accolens* requires an external source of fatty acids. It was discovered that *C. accolens* is capable of utilising triacylglycerols found on human skin (also present in the nostrils) to produce the free fatty acids they require for growth, and that these fatty acids inhibit the growth of *S. pneumoniae*, another major otopathogen. The lipase produced by *C. accolens* for this process to occur appears to be unique to this species.

In Chapter 2 and Chapter 3, *Corynebacterium pseudodiphtheriticum* and *Dolosigranulum pigrum* were identified as candidate probiotic species due to their abundance in the nasopharynx of healthy children with no history of rAOM compared to those with rAOM. As a cross-sectional study, it is not known whether the high abundance of these species in the healthy children is circumstantial or indicative of a protective role. In this chapter, these organisms are characterised further with *in vitro* experiments to bridge the gap between the identification of commensals that are associated with host health and further work in animal models or clinical trials, aiming to determine whether they exhibit bacterial interference in the URT and are playing a role in resistance to the development of rAOM.

*C. pseudodiphtheriticum* and *D. pigrum* are not commonly studied organisms; they are considered to be commensals of the URT, though both species have occasionally been associated with opportunistic human infection. *C. pseudodiphtheriticum* is a Gram positive, non-lipophilic coryneform highly similar to *C. propinquum* by 16S rRNA homology. *D. pigrum* is a Gram positive lactic acid bacterium, the only species of its genus, and is most closely related to *Alloiococcus otitidis*, which is commonly isolated from the middle ear effusions of children with otitis media. In several studies that have focused on the microbiome of the URT in children, the genera *Corynebacterium* and *Dolosigranulum* frequently co-occur and are associated with a range of health-related outcomes (Table 4.1). They also exhibit reduced colonization in infants born via caesarean section compared to those born vaginally. Their co-occurrence suggests that they support each other’s growth
Table 4.1: Studies observing the presence of *Corynebacterium* and *Dolosigranulum* in the upper respiratory tract microbiota and their association with a healthy state. No studies have reported the presence of these genera in the nasopharyngeal microbiota to be associated with disease.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Major <em>Corynebacterium</em> and <em>Dolosigranulum</em> findings</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bogaert et al. (2011)²⁵²</td>
<td>Both identified as common genera in the nasopharynx of healthy children.</td>
<td>96 healthy children</td>
</tr>
<tr>
<td>Laufer et al. (2011)¹⁶⁸</td>
<td>Both genera associated with decreased risk of pneumococcal colonisation and OM diagnosis.</td>
<td>108 children with upper respiratory tract infection symptoms (including 25 with OM and 47 culture-positive <em>S. pneumoniae</em>)</td>
</tr>
<tr>
<td>Pettigrew et al. (2012)¹⁶⁹</td>
<td>Both genera associated with decreased risk of AOM when <em>Streptococcus</em> and <em>Haemophilus</em> abundance was low and no antibiotics used in the previous 6 months.</td>
<td>73 healthy children, 95 children with upper respiratory infection (URI), and 72 children with both URI and AOM</td>
</tr>
<tr>
<td></td>
<td>Levels of <em>Dolosigranulum</em> were lower in children receiving antibiotics in the previous 6 months.</td>
<td></td>
</tr>
<tr>
<td>Biesbroek et al. (2014a)²⁵⁵</td>
<td>Increased presence and abundance of both genera in breastfed children at 6 weeks of age.</td>
<td>101 breastfed infants vs 101 formula-fed infants</td>
</tr>
<tr>
<td></td>
<td><em>Dolosigranulum</em> negatively associated with wheezing (adjusted for feeding type) and number of respiratory tract infections.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Corynebacterium</em> OTU highly homologous to <em>C. pseudodiphtheriticum</em> and <em>C. propinquum</em></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.1: (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Major <em>Corynebacterium</em> and <em>Dolosigranulum</em> findings</th>
<th>Subjects</th>
</tr>
</thead>
</table>
| Biesbroek *et al.* (2014b)\(^{291}\) | Both genera were associated with breastfeeding and lower rates of respiratory infections.  
Infants at 1.5 months of age dominated by *Corynebacterium* and *Dolosigranulum* shifted to *Moraxella* and *Dolosigranulum* at 6 months of age. These profiles were more stable than those dominated by *Haemophilus* or *Streptococcus*. | 60 healthy children                    |
| Bomar *et al.* (2016)\(^{261}\)  | Both genera associated with the absence of *S. pneumoniae* colonisation.  
Demonstrated *in vitro* that *C. accolens* inhibits *S. pneumoniae* with antipneumococcal free fatty acids. | 9 children colonised with *S. pneumoniae* vs 18 children without |
| Prevaes *et al.* (2016)\(^{376}\) | Both genera associated with healthy infants.  
Antibiotic treatment related to a reduction in *Corynebacterium* and *Dolosigranulum* in both groups. | 20 infants with cystic fibrosis vs 45 healthy infants |
| Bosch *et al.* (2017)\(^{292}\)   | Prolonged reduction of *Corynebacterium* and *Dolosigranulum* associated with a higher number of respiratory tract infections.  
Both genera significantly reduced in abundance after antibiotic usage. | Prospective cohort of 112 healthy infants |
### Table 4.1: (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Major <em>Corynebacterium</em> and <em>Dolosigranulum</em> findings</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chomnittree et al. (2017)(^{171})</td>
<td><em>Corynebacterium</em> found to be associated with infants without AOM.</td>
<td>65 infants with AOM vs 74 infants without AOM.</td>
</tr>
<tr>
<td></td>
<td>Antibiotics did not affect otopathogen genera but significantly decreased <em>Corynebacterium</em> and <em>Dolosigranulum</em>.</td>
<td></td>
</tr>
<tr>
<td>Hasegawa et al. (2017)(^{377})</td>
<td>Both genera associated with low likelihood of severe bronchiolitis.</td>
<td>40 hospitalised infants with bronchiolitis vs 110 healthy infants.</td>
</tr>
<tr>
<td>Kelly et al. (2017)(^{378})</td>
<td>Both genera associated with healthy children.</td>
<td>204 children with pneumonia (14 with HIV infection) vs 60 healthy children</td>
</tr>
<tr>
<td></td>
<td>HIV-infected children with pneumonia had a near complete absence of <em>Dolosigranulum</em>.</td>
<td></td>
</tr>
<tr>
<td>Copeland et al. (2018)(^{379})</td>
<td>Both genera associated with healthy adults.</td>
<td>21 adults with chronic rhinosinusitis vs 12 healthy adults.</td>
</tr>
<tr>
<td>Lappan et al. (2018)(^{313}) - Chapter 2</td>
<td>Both genera associated with rAOM-resistant children.</td>
<td>86 children with rAOM vs 98 rAOM-resistant children.</td>
</tr>
</tbody>
</table>
in some way; it has been speculated that *D. pigrum*, as a lactic acid bacterium, lowers the pH of the local environment via the production of lactic acid, selecting for *Corynebacterium* species.\textsuperscript{254} This is a similar mechanism utilised by commensal *Lactobacillus* species in the vaginal microbiome, a low-diversity environment like the nasopharynx, where lactic acid has antimicrobial and anti-inflammatory properties.\textsuperscript{385} *Dolosigranulum* alone has also been associated with decreased *S. aureus* colonisation in adults.\textsuperscript{260}

While it is clear that *Corynebacterium* and *Dolosigranulum* are strongly associated with health, many of the studies in Table 4.1 are recent, and little work has been done to assess them \textit{in vitro} to understand how they contribute to health. No studies appear to have directly tested *D. pigrum* \textit{in vitro} or \textit{in vivo} for therapeutic potential. Of those that have looked at *Corynebacterium*, one early study assessing *C. pseudodiphtheriticum* demonstrated competitive binding with *M. catarrhalis* to pharyngeal cells, though no zone of inhibition was observed on blood agar.\textsuperscript{386} An unknown *Corynebacterium* species was reported to eradicate *S. aureus* in the nose of adult volunteers.\textsuperscript{387} A similar result was subsequently observed with *C. pseudodiphtheriticum*; however this study was based on a small number of volunteers and results were not well-replicated across subjects.\textsuperscript{388} A recent study inoculating *C. pseudodiphtheriticum* strain 090104 into mice found that it improved resistance to infection by respiratory syncytial virus and secondary pneumococcal infection, where non-viable 090104 had a weaker effect.\textsuperscript{331} The whole genome sequence of the commensal strain 090104 was reported to not contain any bacteriocin-related genes,\textsuperscript{389} so it has been suggested that the mechanism may rely on modulation of the host immune system.\textsuperscript{331} Bomar \textit{et al}. (2016) recently described the inhibition of *S. pneumoniae* by *C. accolens* via antimicrobial free fatty acids produced as a requirement for its growth.\textsuperscript{261}

The work presented in this chapter aimed to detect direct bacterial antagonism by *C. pseudodiphtheriticum* and *D. pigrum* against one well-characterised strain of each otopathogen species and 6 clinical isolates of each from the middle ear and nasopharynx of children with otitis media. This work is intended as a starting point to determine whether these candidate commensal species are capable of directly inhibiting otopathogen strains independently or in synergy, and if inhibition is effected via a bacteriocin-like inhibitory substance; as a precursor to further work in cellular or animal models to detect more complex or subtle mechanisms by which they may contribute to health. This is the first time these species have been tested together for protective characteristics against a range of clinically relevant otopathogen strains.
This work can be used to underpin further studies addressing Neville et al.’s postulates 3 and 4; to determine, via animal studies or clinical trials, whether these species can be utilised in a probiotic therapy to prevent or resolve rAOM.

### 4.2 Methods

#### 4.2.1 Bacterial strains and growth conditions

Two clinical isolates each of *C. pseudodiphtheriticum* and *D. pigrum* were obtained from PathWest Laboratory Medicine (Perth, Western Australia) as candidate probiotic species. *C. pseudodiphtheriticum* strain 52091 (referred to here as *C. pseudodiphtheriticum* 1, Cp1) was isolated from a blood culture, and strain 65442 (*C. pseudodiphtheriticum* 2, Cp2) from a continuous ambulatory peritoneal dialysis (CAPD) fluid. *D. pigrum* strains 61470 (*D. pigrum* 1, Dp1) and 71319 (*D. pigrum* 2, Dp2) were isolated from blood cultures. It was unconfirmed whether these isolates were the cause of an opportunistic infection or were contaminants from the skin puncture site. All four strains were initially propagated onto blood agar plates (PathWest Media) and incubated at 36°C with 5% CO$_2$ until sizeable colonies appeared (1-2 days for *C. pseudodiphtheriticum*, 3-4 days for *D. pigrum*) before storage at -80°C in brain-heart infusion broth (BHIB) with 15% glycerol (PathWest Media).

For all experimental work, *C. pseudodiphtheriticum* isolates were grown from glycerol stocks on blood agar (PathWest Media) at 37°C with 5% CO$_2$ for 48 hours. Colonies were white-grey, smooth and round. *D. pigrum* isolates were cultured on blood agar (PathWest) at 37°C with 5% CO$_2$ for 72 hours. Colonies were very small but varied in size from < 1 mm to approximately 1 mm and were white-grey, smooth, round and alpha haemolytic. *D. pigrum* 2 colonies were slightly larger and flatter than those of *D. pigrum* 1.

A total of 21 strains of the major otopathogen species were tested for inhibition of growth by *C. pseudodiphtheriticum* and *D. pigrum*. Table 4.2 contains the isolates used. *M. catarrhalis* and *S. pneumoniae* strains were cultured on blood agar plates (PathWest Media) and NTHi strains on CHOC agar plates (PathWest Media), incubated overnight at 37°C with 5% CO$_2$, and stored at -80°C in BHIB with 15% glycerol.
Table 4.2: Otopathogen strains used in the interference assays.
Strains kindly made available by Lea-Ann Kirkham, Caitlyn Granland, Deborah Lehmann and Jacinta Bowman.

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. catarrhalis</em> 93 (ID 1091)</td>
<td>Middle ear effusion</td>
<td></td>
</tr>
</tbody>
</table>
| *M. catarrhalis* 59 (ID 1010) | Nasopharynx | GROMIT study, Wiertsema et al. (2011) 
| *M. catarrhalis* 65 (ID 1085) | Middle ear effusion | |
| *M. catarrhalis* 38 (ID 1091) | Nasopharynx | GROMIT study, Wiertsema et al. (2011) 
| *M. catarrhalis* 55 (ID 1103) | Middle ear effusion | |
| *M. catarrhalis* 93 (ID 2063) | Nasopharynx | GROMIT study, Wiertsema et al. (2011) 
| *M. catarrhalis* (ATCC® 25238™) Type strain | | |
| *S. pneumoniae* LN004 (ID 003) | Nasopharynx | |
| *S. pneumoniae* LN92 (ID 24) | Middle ear effusion | Kalgoorlie Otitis Media Project (KOMP), Watson et al. (2006) 
| *S. pneumoniae* LN123 (ID 64) | Middle ear effusion | |
| *S. pneumoniae* LN1091 (ID 1252) | | |
| *S. pneumoniae* LN3179 (ID 3355) | Nasopharynx | GROMIT study, Wiertsema et al. (2011) 
| *S. pneumoniae* LN852 (ID 981) | | |
| *S. pneumoniae* strain D39 Historical strain | | Avery et al. (1944) 
| NTHi H122 (ID 1040) | Middle ear effusion | |
| NTHi H76 (ID 1014) | Nasopharynx | GROMIT study, Wiertsema et al. (2011) 
| NTHi H146 (ID 1019) | | |
| NTHi H134 (ID 1036) | | |
| NTHi H94 (ID 1028) | | |
| NTHi H108 (ID 1041) | | |
| NTHi 86-028NP | | Bakaletz et al. (1988) 


4.2. Methods

4.2.2 Bacterial growth curves

To understand the behaviour of *C. pseudodiphtheriticum* and *D. pigrum* in broth culture, strain 1 of each species was monitored across 24 hours for growth in BHIB. Four to five pure colonies from each species were added to 3 ml of BHIB (PathWest Media) and incubated with agitation at 37°C for approximately 13 hours. This overnight broth was adjusted to an optical density (OD) at 600 nm of 0.08-0.12 with BHIB and 500 µl of this standardised broth was transferred to 8 ml of BHIB. The OD600 of these broths was measured immediately (timepoint 0) and then at regular intervals across a 24 hour period. At each measurement, an aliquot of broth was taken and serially diluted from $10^{-1}$ to $10^{-6}$ for a colony-forming unit (CFU) viable count assay. Drops of 15 µl were plated in triplicate on blood agar plates and incubated as described above for each species. Colony counts were used to determine the viable number of bacteria at each time point. Three replicates for each species were grown and monitored concurrently.

4.2.3 Agar drop assay

To observe interference between the candidate probiotic species (*C. pseudodiphtheriticum* and *D. pigrum*) and the otopathogens (*M. catarrhalis, S. pneumoniae* and NTHi) a side-by-side drop assay was undertaken on agar plates. All strains were revived from glycerol stocks and subcultured once. The two *C. pseudodiphtheriticum* strains (Cp1 and Cp2) were then cultured in 3 ml of BHIB overnight then diluted to an OD600 of between 0.3 and 0.4, as described in Bomar et al. (2016). D. pigrum 1 (Dp1), which produced limited turbidity in broth, was grown to an OD600 of 0.25-0.3. D. pigrum 2 (Dp2) grew poorly and inconsistently in broth and so colonies were resuspended to an OD600 of 0.3-0.4. Mixtures containing 1 part standardised *C. pseudodiphtheriticum* and 1 part standardised *D. pigrum* were also prepared for all four combinations of the strains. Drops of 5 µl were placed onto blood agar and CHOC agar plates (PathWest Media). These plates were prepared in duplicate and incubated for 2 days (*C. pseudodiphtheriticum*-only and combination plates) or 3 days (*D. pigrum*-only plates) at 37°C with 5% CO$_2$. Negative controls with 5 µl drops of BHIB were also prepared for both 2-day and 3-day incubation in duplicate. Subsequently, 5 µl of a colony resuspension (also in BHIB to an OD600 of 0.3-0.4) of each otopathogen strain was placed adjacent to the previous growth or negative control drop. Plates were photographed after overnight incubation at 37°C with 5% CO$_2$. 
4.2.4 Agar overlay assay

To determine whether any inhibition of otopathogen growth by the candidate probiotic species is due to a secreted molecule, an agar overlay assay similar to that described in Tano et al. (1999) was carried out. The candidate probiotic strains were first revived from glycerol stocks and subcultured once. For plates containing only *C. pseudodiphtheriticum*, the strains were grown to an OD600 of 0.6-0.7 (approximately equivalent to $1 \times 10^8$ viable cells, as determined by the Cp1 growth curve assay) in 4 ml of BHIB. For the *D. pigrum*-only and *C. pseudodiphtheriticum/D. pigrum* combination plates, colonies were resuspended in BHIB to the same turbidity due to the limited growth of *D. pigrum* in broth. Blood agar base was made with BBL Mueller Hinton Agar II base (BD) to which citrinated horse blood was added at 5% volume. 1 ml of candidate probiotic culture (*C. pseudodiphtheriticum*, *D. pigrum* or a combination of both) was added to 20 ml of molten blood agar and poured into a petri dish to set. These plates were incubated at 37°C with 5% CO$_2$ for 2 days (*C. pseudodiphtheriticum* and combination plates) or 3 days (*D. pigrum* plates). Negative controls contained 1 ml of BHIB per 20 ml of agar, and negative controls were included for 2-day and 3-day incubation periods. After incubation, 20 ml of molten blood agar was poured over the bottom layer and 5 µl drops of each otopathogen strain (standardised via colony resuspension to an OD600 of 0.2) were inoculated on the plates once set. 5 µl of BHIB was included as negative control, and all drops were placed in duplicate on the same plate. The plates were photographed after overnight incubation.

4.2.5 Disc diffusion assay

Extracellular and intracellular supernatants of *C. pseudodiphtheriticum* were tested in a disc diffusion assay against the *M. catarrhalis* strains. Cp1 and Cp2 were cultured on blood agar as previously described. For an overnight broth culture, 3-4 colonies of each strain were inoculated into 5 ml of BHIB and grown at 37°C with agitation for 17 hours. The cultures were centrifuged at 3220 x g for 10 minutes to pellet the bacterial cells, and the broth supernatant was collected and filtered through a 0.22 µm filter (extracellular supernatant). A fresh aliquot of BHIB was also centrifuged and filtered as a negative extracellular control. These supernatants were kept at 4°C prior to inoculation onto the discs. The cell pellets were then resuspended in 2 ml of 1 x PBS, and 1 ml of this resuspension was lysed via the OmniLyse HL kit (Claremont Bio) for 7.5 minutes on ice to produce intracellular supernatant. A
negative intracellular control (1 x PBS) was also lysed. The lysates were centrifuged at 3220 x g for 10 minutes, filtered through a 0.22 µm filter and kept at 4°C.

Blank 6 mm Oxoid filter paper discs (ThermoFisher) were laid out in sterile petri dishes inside a biosafety cabinet. Seven types of discs were prepared by inoculating the disc with 20 µl of the following solutions: extracellular and intracellular supernatant from each of the two *C. pseudodiphtheriticum* strains (solutions 1-4), extracellular and intracellular negative controls (solutions 5 and 6) and 1.5 mg/ml kanamycin as a positive control (solution 7). The inoculated discs were dried for approximately 1 hour inside the biosafety cabinet. The seven strains of *M. catarrhalis* in Table 4.2 were tested for susceptibility to these solutions. Strains cultured on blood agar plates as previously described were resuspended in 2 ml of BHIB to an OD600 of approximately 0.1, which is equivalent to the McFarland standard of 0.5 commonly used in antimicrobial disc diffusion assays. Two blood agar plates per strain were lawn-inoculated with a cotton swab. One disc of each type was pressed onto each plate with sterile forceps. The plates were incubated at 37°C with 5% CO₂ overnight and photographed.

### 4.3 Results

#### 4.3.1 Growth of *C. pseudodiphtheriticum* and *D. pigrum* in broth

Figure 4.1 contains growth curves displaying the relationship of turbidity (measured by the OD at 600nm, OD600) over time (Fig. 4.1a) and the relationship between turbidity and CFU per ml (Fig. 4.1b) for strain 1 of each species. The three Cp1 replicate broths were consistent and reached a viable count of $1 \times 10^8$ CFU/ml at an optical density of 0.6-0.7. After 24 hours the growth was very turbid. Dp1 exhibited very limited growth in broth, with the turbidity remaining below OD600 0.2 even after 24 hours of incubation. Subsequent tests on both *D. pigrum* strains with a higher starting inoculum (colonies directly from the plate) improved this growth, though neither strain exceeded an OD600 of approximately 0.4 after 24 hours of incubation, with Dp2 sometimes failing to produce turbidity. Colony resuspension in broth was determined to be a more reliable method for preparing a standardised suspension of *D. pigrum*. 
Figure 4.1: Growth curve results for *C. pseudodiphtheriticum* and *D. pigrum*. The relationship between a) turbidity and time, and b) turbidity and viable count for each species is shown. Three replicates for each strain (Cp1 and Dp1) were grown concurrently.
4.3.2 Detecting inhibition via an agar drop assay

In the agar drop assay, the otopathogen strains were grown on agar adjacent to the candidate probiotic strains. There were no observable effects of either *D. pigrum* strain on any of the otopathogen strains tested. However, *D. pigrum* grew poorly and inconsistently on CHOC agar so its interaction with NTHi in this assay could not be assessed. *C. pseudodiphtheriticum* did not inhibit any of the *S. pneumoniae* or NTHi isolates, however both strains of *C. pseudodiphtheriticum* reduced the growth of all *M. catarrhalis* strains tested. Figure 4.2 contains photographs of representative plates, where *M. catarrhalis* exhibits reduced growth from the direction of the *C. pseudodiphtheriticum* drop; forming a ‘crescent-moon’ pattern of growth (Fig. 4.2a). This was distinct from the uniform *M. catarrhalis* growth observed next to drops of *D. pigrum* (Fig. 4.2b), which did not appear to inhibit *M. catarrhalis*, and BHIB as the negative control (Fig. 4.2c). This effect occurred on both duplicate plates for all seven *M. catarrhalis* strains. The effect was not enhanced or reduced by the presence of *D. pigrum* (Fig. 4.2d).

4.3.3 Testing for inhibitory compounds via diffusion through agar

The agar overlay assay was used to determine whether the candidate commensal species would be capable of inhibiting the otopathogen strains when separated by agar. Neither *D. pigrum* strain inhibited any of the otopathogen strains in this assay, including NTHi (Figure 4.3a). Both *C. pseudodiphtheriticum* strains were unable to produce clear results in this assay. This species migrated to the surface of the bottom-layer agar during incubation, then mixed with the top-layer agar upon pouring; resulting in a lawn of growth of *C. pseudodiphtheriticum* on the agar surface where the otopathogen strains were plated. This effect was more pronounced with Cp2 than with Cp1 and obscured the visibility of the otopathogens grown on the top layer (Figure 4.3b).
Figure 4.2: Agar drop assay results for *M. catarrhalis*. The effect of a) *C. pseudodiphtheriticum*, b) *D. pigrum*, c) BHIB (negative control) and d) A mixture of *C. pseudodiphtheriticum* and *D. pigrum* on the growth of *M. catarrhalis* is shown. Inner drops are *C. pseudodiphtheriticum*/*D. pigrum*; outer drops are *M. catarrhalis*. Representative images are shown.
Figure 4.3: Overlay assay results. 

a) *D. pigrum* in the bottom layer of agar did not inhibit any of the strains of NTHi spotted on the top layer. b) *C. pseudodiphtheriticum* migrated to the surface of the agar and obscured results of the *M. catarrhalis* strains spotted on the top layer. Representative images are shown.
4.3.4 Testing for inhibitory compounds in the extracellular and intracellular supernatants of *C. pseudodiphtheriticum*

To begin to understand the mechanism of inhibition of *M. catarrhalis* by *C. pseudodiphtheriticum*, the activity of extracellular and intracellular *C. pseudodiphtheriticum* supernatants against *M. catarrhalis* growth was assessed by a disc diffusion assay. None of the seven *M. catarrhalis* strains tested showed any zones of inhibition or reduced growth around either the extracellular or intracellular *C. pseudodiphtheriticum* supernatants (Figure 4.4). All strains were susceptible to kanamycin and had a zone of inhibition 19-22 mm in diameter.
4.3. Results

Figure 4.4: Disc diffusion assay results for *M. catarrhalis* against *C. pseudodiphtheriticum* extracts. No inhibition was observed surrounding the discs impregnated with *C. pseudodiphtheriticum* supernatant or the negative controls. The central disc containing 30 µg kanamycin (positive control) shows a zone of inhibition. A representative image is shown; the results for each *M. catarrhalis* strain and replicate plate were identical.
4.4 Discussion

The presence of *Corynebacterium* and *Dolosigranulum* in the nasopharynx of children has consistently been associated with a healthy state. The work in Chapters 2 and 3 on the microbiome of the nasopharynx in children with and without rAOM showed that *C. pseudodiphtheriticum* and *D. pigrum* were overrepresented in rAOM-resistant. Despite their consistent association with nasopharyngeal health, little work has been done to understand how they might contribute to a healthy state. In this chapter, two strains of each species were tested *in vitro* for inhibitory activity against seven strains each of the major otopathogens; *S. pneumoniae*, non-typeable *H. influenzae* and *M. catarrhalis*. While no inhibitory effects were exhibited by *D. pigrum*, both strains of *C. pseudodiphtheriticum* reduced the growth of all *M. catarrhalis* strains on blood agar. However, the extracellular and intracellular supernatant of *C. pseudodiphtheriticum* had no effect on the growth of *M. catarrhalis* in a disc diffusion assay.

4.4.1 The inhibition of *M. catarrhalis*

The growth of all *M. catarrhalis* strains was reduced when adjacent to *C. pseudodiphtheriticum*. This effect was not replicated with cell-free supernatants from *C. pseudodiphtheriticum* or on the other otopathogen species. This is the first time that direct inhibition has been observed between *C. pseudodiphtheriticum* and *M. catarrhalis*. A study in 1989 suggested that *C. pseudodiphtheriticum* was capable of competitively adhering to cells when co-cultured with *M. catarrhalis*, but direct inhibition on agar was not observed.\(^{386}\) *S. pneumoniae* and *S. mitis* have previously been reported to inhibit the growth of *M. catarrhalis*,\(^{395,396}\) but direct antagonism by *C. pseudodiphtheriticum* against any species has not been previously shown, as its *in vitro* activity has not been explored in detail.

*D. pigrum* did not exhibit inhibitory activity in this chapter. Several species of *Lactobacillus* have recently been shown to inhibit *M. catarrhalis* via the antimicrobial effects of lactic acid,\(^{397}\) which has also been demonstrated in the context of the vaginal microbiota.\(^{385}\) The inhibition of *M. catarrhalis* by *L. rhamnosus* GG was shown to be pH-dependent (more effective at lower pH), and the strain also demonstrated competitive exclusion of *M. catarrhalis* on epithelial cells and prevention of *M. catarrhalis* biofilm formation.\(^{397}\) As the pH or the production of lactic acid by *D. pigrum* was not measured in the work in this chapter, it remains a possibility that *D. pigrum* plays a similar role in the healthy nasopharynx.
4.4.2 Understanding the mechanism of inhibition

While this work demonstrated a reduction in *M. catarrhalis* growth by *C. pseudodiphtheriticum*, no extracellular or intracellular component from *C. pseudodiphtheriticum* that was capable of producing the same effect could be detected. This suggests that the inhibition may not be the result of a diffusible antimicrobial substance. However, there are several other possibilities as to why evidence for a bacteriocin-like molecule was not observed. Such a molecule may be produced by *C. pseudodiphtheriticum*, but may not have been present in sufficient concentrations to produce an observable zone of complete inhibition. It is also possible that the substance degraded upon extraction; general care was taken to keep the sample cold and undertake gentle lysis, but this may not have been sufficient to prevent the degradation or loss of activity of particular compounds, and may have also caused the release of intracellular proteases.

The BAGEL4 webserver for detection of bacteriocin genes\(^{362}\) indicates that neither of the *C. pseudodiphtheriticum* genomes available in the RefSeq\(^{343}\) database (090104 and DSM44287) contain any such genomic regions of interest. Interestingly, *D. pigrum* strain ATCC 51524 contains three areas of interest, while strain 15S00348 contains only two of these. Both of these strains are potentially opportunistic pathogens\(^{398,399}\), suggesting that the production of such molecules is not restricted to commensal strains. However, as no inhibition of otopathogen strains by *D. pigrum* was observed in this chapter, these regions may not be present in all strains. Alternatively, they may only be expressed under certain environmental conditions or require the presence of another, perhaps host-derived molecule not provided in blood agar to trigger production or activation.

Alternatively, it may be that the inhibitory effect of *C. pseudodiphtheriticum* is not the result of an antimicrobial substance but rather an effect that requires the activity of live bacteria; for example via the competitive acquisition of nutrients that *M. catarrhalis* requires, or some other change in the local environment that is not the result of a secreted molecule. The pattern of reduced growth of *M. catarrhalis* may support this; rather than a clear zone with no growth as is commonly seen with antimicrobial substances, a zone of reduced growth was observed. This perhaps indicates suboptimal conditions near *C. pseudodiphtheriticum* rather than direct bactericidal effects. Additionally, the presence of *C. pseudodiphtheriticum* did not interfere with the growth of *S. pneumoniae* or NTHi; there may be some essential component for growth of *M. catarrhalis* that is utilised or made inaccessible by *C. pseudodiphtheriticum* but is not required by the other otopathogen species.
Importantly, *C. pseudodiphtheriticum* is likely to inhibit *M. catarrhalis* via a different mechanism than the one characterised by Bomar *et al.* (2016). They described the inhibition of *S. pneumoniae* by *C. accolens*, which was mediated by an essential metabolic process. *C. accolens* is a lipophilic *Corynebacterium* species, and converts triacylglycerols into fatty acids that are required for its growth. These fatty acids inhibit the growth of *S. pneumoniae*. As *C. pseudodiphtheriticum* is non-lipophilic, this may be why it did not inhibit the growth of *S. pneumoniae* - it does not carry out the same fatty acid mechanism as *C. accolens*. It is therefore possible that multiple *Corynebacterium* species are beneficial to the URT, acting via different mechanisms against different pathogens in a complementary fashion.

### 4.4.3 Antimicrobial effects may be strain-specific

The assays in this chapter utilised *C. pseudodiphtheriticum* and *D. pigrum* clinical isolates obtained from opportunistic infections unrelated to the URT. Ideally, a wide range of commensal nasopharyngeal strains should be tested for their activity against the otopathogens because these strains are likely to be more similar to those present in the nasopharynx of healthy children, and opportunistic pathogens are not good candidates for probiotic therapy. Commensal nasopharyngeal carriage strains could not be obtained for this work, as these are not routinely cultured and stored by pathology labs, and the swabs collected in the biOMe study (Chapters 2 and 3) had insufficient material remaining for the isolation of individual species. While the inhibition of *M. catarrhalis* by *C. pseudodiphtheriticum* was broad (all strains of both species exhibited the same interaction) and no inhibition by *D. pigrum* was noted, it is possible that commensal strains of *C. pseudodiphtheriticum* or *D. pigrum* will behave differently.

The interaction between *C. pseudodiphtheriticum* and *M. catarrhalis* was not influenced by the presence of *D. pigrum*; these two species may not interact, however there may be interactions that are not observable in this context. This may be due to their different rates of growth on agar, strain-specific effects, or because their true *in situ* interactions are more complex than what can be observed on agar plates. For example, it has been speculated that the production of lactic acid by *D. pigrum* creates a more acidic local environment, selecting for the growth of *Corynebacterium* species and as discussed above, lactic acid is an important antimicrobial compound. There are also other mechanisms of bacterial interference and probiotic activity such as competitive adherence and stimulation of the immune system that can only be
observed experimentally in the presence of host cells or in an *in vivo* model, which may complement the effect observed here on agar.

### 4.4.4 Further research to determine how *Corynebacterium* and *Dolosigranulum* contribute to health in the URT

Recent nasopharyngeal microbiome studies have been clear in showing that the presence of both *Corynebacterium* and *Dolosigranulum* is associated with a healthy state (Table 4.1), and may be an important characteristic of nasopharyngeal microbiomes that are resistant to disease. Significant reductions in *Corynebacterium* and *Dolosigranulum* abundance after antibiotic use have been reported;\textsuperscript{171,292,376} this may contribute to further episodes of OM or other disease if they are unable to recolonise the nasopharynx before the otopathogens do, indicating that they are important components of the healthy nasopharyngeal microbiome. The work in this chapter represents an important preliminary step towards investigating the role of these genera in the context of children with otitis media, reporting an inhibitory interaction between *C. pseudodiphtheriticum* and *M. catarrhalis* that may be indicative of bacterial interference in the nasopharynx.

As recent studies have shown a strong and consistent association between these commensal organisms and nasopharyngeal health, further research should focus on understanding whether they play an active role in contributing to a healthy state. A range of commensal nasopharyngeal *C. pseudodiphtheriticum* and *D. pigrum* strains should be tested against clinical otopathogen strains for multiple bacterial interference mechanisms, as effects may be strain-specific. The presence of antimicrobial substances should not be ruled out; genomic analyses of commensal strains may reveal the presence of bacteriocin genes, particularly in *D. pigrum*. Additionally, Latham *et al.*’s recent study revealing a bacteriocin-like substance produced by *H. haemolyticus* against NTHi utilised an enrichment method prior to screening for strains with activity against NTHi.\textsuperscript{375} This may prove useful for *Corynebacterium* and *Dolosigranulum* if they produce novel bacteriocins or small antimicrobial molecules not recognisable by genomic analysis and are not present at a sufficient concentration in supernatant to produce a visible effect on agar. Quantification of any inhibitory effects in comparison with other antimicrobial molecules is therefore also important. However, in light of this preliminary work and the connections between *Corynebacterium, Dolosigranulum*, lactic acid and pH it is perhaps more plausible that their
potential role in maintaining a healthy nasopharyngeal microenvironment involves other mechanisms of bacterial interference and perhaps the host immune response. van den Broek et al. (2018) thoroughly explored the role of lactic acid and pH in the inhibition of \textit{M. catarrhalis} by \textit{Lactobacillus} species,\textsuperscript{397} and a similar approach could be used to study commensal \textit{D. pigrum} strains as the species is also a lactic acid bacterium.

Epithelial cell models have been successful in understanding the competition between \textit{H. haemolyticus} and NTHi.\textsuperscript{288} Biofilm and adhesion assays in NTHi and \textit{M. catarrhalis} have also shown that commensal bacteria can inhibit the formation of biofilm and competitively exclude the otopathogens from adhesion to host cells.\textsuperscript{288,397} There are some previous studies that have used a microfluidics apparatus to explore the interactions between pathogenic and commensal bacteria in a host cell environment in the gut.\textsuperscript{400,401} Microfluidics provides the benefit of an environment that more closely mimics the \textit{in vivo} environment than regular cell culture. However, the use of microfluidics in bacterial/epithelial co-culture has not been widely explored, and the expertise required to build such a system is not commonly found within a microbiological lab. Current commercial systems may need improvement as the CellASIC® ONIX Microfluidics System (Merck) failed to sustain even robust cell lines in work preliminary to this thesis.\textsuperscript{402} Experiments assessing bacterial interference in a host cell environment can be coupled with proteomics and/or transcriptomics to obtain a deeper insight into the mechanisms and host responses occurring in the culture.

If further investigation into \textit{Corynebacterium} and \textit{D. pigrum} suggest that they are actively involved in nasopharyngeal health rather than simply being reduced in abundance upon infection, then Neville \textit{et al}’s commensal postulates can be revisited; 3) to assess their capability of reducing disease when introduced to a new host; and 4) to see if they are detectable following their introduction to the new host. Such work would involve experiments in an animal model of AOM\textsuperscript{403} understand the \textit{in vivo} effects of their colonisation and importantly, to assess the feasibility of their use as a probiotic therapy for otitis media. By investigating the mechanisms by which they contribute to health, the mechanism may be reproduced without the live organisms if the organisms themselves are not suitable for probiotic therapy in children.
4.5 Conclusions

The association between co-occurring *Corynebacterium* and *Dolosigranulum* and a healthy nasopharyngeal environment is well-established. The work in this chapter demonstrated direct antagonism between *C. pseudodiphtheriticum* and 7 clinically relevant strains of *M. catarrhalis*, but the growth of *S. pneumoniae* and NTHi were unaffected. The presence of *D. pigrum* does not influence the effect. The mechanism for this inhibition remains unknown, and this interaction is likely one small part of a complex host and microbial environment in the nasopharynx. Further work involving relevant commensal strains, exploration of the role of pH and lactic acid and interactions in a host cell environment are the immediate next step towards characterising the mechanism by which *Corynebacterium* and *Dolosigranulum* contribute to health in the nasopharynx. This work will provide direction for the use of animal models or clinical trials to evaluate the potential for these organisms, alone or in synergy, to be developed into a probiotic therapy to treat or prevent rAOM in children.
Chapter 5

Profiling the microbial metatranscriptome and host transcriptome of middle ear fluid using dual RNA-seq

This chapter describes a pilot study designed to evaluate the feasibility of a new technique, dual RNA-seq, in characterising both the host and microbial expression profiles in the context of otitis media. The goal of the dual RNA-seq approach is to characterise the host response – particularly the role of the hypoxia pathway – in children with rAOM and COME, and to simultaneously profile the activity of the microorganisms in the middle ear.
5.1 Introduction

Studying the microbiome using DNA-based sequencing technologies can offer important insight into the presence, abundance and potential function of the microbes in the sample environment. One of the major limitations of 16S rRNA amplicon and metagenomic sequencing is that DNA-based methods cannot distinguish between organisms that are alive, dead, dormant or metabolically active; and while the potential protein content of their genomes can be determined with metagenomics, it cannot inform whether an organism is using those genes. To understand which organisms are actively carrying out their genomic functions – and which functions are in use at a particular time – the transcriptome can be sequenced via RNA-seq.

RNA-seq is used to profile the gene expression of a single organism to understand the dynamics of disease, responses to certain conditions or stimuli, or how metabolic activity changes over time. Dual RNA-seq is a relatively new approach that involves concurrent sequencing of a host and pathogen transcriptome. It is used primarily to measure the expression of the host and a single pathogenic organism within in vitro culture models, though the term is occasionally used to refer to the sequencing of the host transcriptome and the microbial metatranscriptome. Metatranscriptomics is the study of the transcriptome from the entire population of organisms in an environment, and is used to understand the activity and dynamics of microbial communities. To determine the applicability of dual RNA-seq to otitis media, a pilot study was performed to concurrently sequence the microbial metatranscriptome and the host transcriptome in the middle ear fluid of children with OM. This study describes the first application of this technique to investigate OM, and is also the first study aiming to characterise the host transcriptome or the microbial metatranscriptome of the human middle ear.

The pilot study had three major aims. First, to determine the host expression profile of children with rAOM compared to children with COME; and to identify the genes that are differentially expressed between these two conditions. As OM has been described as a continuum of disease phenotypes, transcriptome data from human middle ear effusions (MEE) will provide information as to how these two phenotypes differ at the molecular level, further defining this spectrum of disease. This aim includes a comparison of the genes expressed in the middle ear, to identify differences in the host response in the local environment; as well as the genes expressed in the blood, to identify differences in the systemic response in these two phenotypes. Secondly, the study aimed to assess whether RNA-seq data obtained from MEE
samples can be utilised to additionally profile the microbial metatranscriptome in a dual RNA-seq approach. Information on the activity of microbes in the middle ear of children with rAOM compared to those with COME will complement studies using DNA-based metataxonomics or metagenomics to characterise the microbiome of these environments, such as the work described in Chapters 2 and 3, and will enable some understanding of the roles of *A. otitidis* and *T. otitidis* in MEE. The final aim of the pilot project was to look at the specific role of the host-transcribed hypoxia pathway in the human middle ear, assessing whether genes in this pathway are upregulated in MEE. As MEE is only produced in OM, there is no equivalent sample to be obtained from the middle ears of healthy children; so this is done by comparing the transcriptome of the MEE to that of a matching blood sample from the same patient. To ensure that the analysis is relevant to this research question and to limit spurious results, the analysis was targeted to genes involved in the hypoxia pathway.

During low-oxygen (hypoxic) conditions such as those found in inflamed environments, the activation of the HIF-VEGF pathway results in the upregulation of hypoxia inducible factor (HIF) transcription factors, which induce the expression of genes such as vascular endothelial growth factor (VEGF) that allow tissues to respond to hypoxic conditions and maintain homeostasis. Previous work in animal models of chronic OM has suggested a role for cellular hypoxia in OM pathogenesis, demonstrating both a hypoxic middle ear environment and the upregulation of these pathways. It has been proposed that hypoxia-inducible pathways may be targetable for therapies in inflammatory diseases like chronic OM, potentially achievable via drugs that are already approved for use in other contexts. However, the potential role for the hypoxia pathway in chronic OM has only been demonstrated in animal models. The pilot study described in this chapter therefore sought to determine whether these pathways are upregulated in the middle ear of children with COME and rAOM. In an animal model, an unaffected ear can be used as a control sample (by removal post-mortem) to determine upregulation or downregulation of hypoxia-related genes. With this being inapplicable to human studies, another approach is the use of venous blood as a ‘normoxic’ control, demonstrated in a mouse model by Cheeseman *et al.* (2011). The healthy middle ear lumen is thought to have an oxygen tension similar to that of mixed venous blood, and middle ear effusions have been shown to contain a high number of leukocytes. In the absence of an unaffected middle ear specimen, venous blood is the best available ‘normoxic’ control for human middle ear effusions.
In Chapters 2 and 3, DNA-based methods were used to investigate the microbiome of the middle ear in children with rAOM, where two potentially novel otopathogens (A. otitidis and T. otitidis) were identified. These organisms, particularly A. otitidis, have also been commonly detected in the MEE of children with COME, however unlike the known otopathogens A. otitidis is rarely detected in the nasopharynx and both organisms also inhabit the EAC. Their role in the development of OM remains unknown, as it is not yet clear how these organisms migrate to the middle ear when the tympanic membrane is intact. It is possible that the presence of A. otitidis and T. otitidis in MEE samples is a result of contamination from the EAC, and not reflective of their growth or activity in the middle ear. Metatranscriptome data from the otitis-prone middle ear would complement what is known about the metagenome, providing insight into which of the organisms detected in this environment are metabolically active, and describing what functions they are carrying out. This information would contribute to a more accurate picture of the overall microbial dynamics of the middle ear in OM, and would particularly assist in the understanding of A. otitidis and T. otitidis and their role in the disease. Furthermore, the comparison between MEEs from children with rAOM and those with COME will help to elucidate whether these organisms exhibit different behaviours, functions or roles in these two phenotypes.

The purpose of this pilot study was to evaluate the feasibility of a dual RNA-seq approach to studying OM in human clinical samples. If feasible, this approach can be used in larger studies to gain a more accurate picture of the microbial environment in the OM middle ear, to define the molecular differences in host response between rAOM and COME and to determine whether the hypoxia pathway is a valid target for therapy in human OM. Dual RNA-seq OM datasets may be analysed in conjunction with host and microbial genomic datasets to provide a systems biology view of the interaction between host and pathogens in otitis media.

## 5.2 Methods

### 5.2.1 Patient recruitment, sample collection and storage

Children under the age of 5 years with physician-diagnosed severe otitis media (rAOM or COME) scheduled for grommet surgery were recruited to the study. Children were recruited from a private day surgery in Perth, Western Australia between
July and December 2016. None of the children had previously undergone grommet surgery or had their adenoids removed. Written parental consent was obtained prior to sample collection and the study recruitment was approved by the University of Western Australia Human Research Ethics Committee (RA/4/1/7870) and the relevant hospital governance committee. On the day of surgery, a questionnaire on relevant environmental risk factors for OM was completed by the parent.

At the time of surgery, MEE from each ear was aspirated into separate sterile Argyle™ specimen traps (Covidien) followed by 2 ml of sterile saline to flush the tubing. If the tympanic membrane was perforated, the MEE was collected and processed but did not undergo RNA-seq due to the potential influence of a perforation on both the oxygen tension and the microbiome of the middle ear. MEEs were immediately placed on ice. MEEs were transferred into a 1.5 ml microfuge tube and centrifuged at maximum speed (16,100 x g) at 4°C to pellet the mucus. The saline supernatant was removed and the centrifugation repeated until the entire sample volume was pelleted. Ten volumes of TRIzol Reagent (Ambion) was added to the sample within 1 hour of collection. The sample was vortexed vigorously to break up the mucus, making the sample as homogeneous as possible. Homogenised samples were incubated at room temperature for 5 minutes before storage at -80°C (within 1.5 hours of collection) for later RNA extraction.

Also at the time of surgery, approximately 2.5 ml of venous blood was collected via a cannula in the patient’s hand and transferred into a PAXgene Blood RNA Tube (PreAnalytiX, QIAGEN), inverted several times and stored at room temperature for at least 2 hours. Blood samples were then frozen at -20°C for 24 hours, then moved to -80°C for later RNA extraction.

5.2.2 RNA extraction

All pipettes, surfaces and gloves were treated with RNase-Zap (Ambion) prior to and during RNA extraction to prevent degradation of RNA.

Homogenised MEE samples in TRIzol Reagent were brought to room temperature for extraction of RNA. Chloroform was added to the sample at a ratio of 0.2 ml per 1 ml of TRIzol Reagent. Samples were shaken vigorously by hand for 15 seconds and incubated for 2-3 minutes at room temperature. The samples were centrifuged at 12,000 x g for 15 minutes at 4°C and the aqueous phase containing RNA was collected. The aqueous phase was then processed with the RNeasy MinElute Cleanup Kit (QIAGEN) as per the manufacturer’s protocol and eluted in 14 µl of RNase-free
water. Eluted total RNA was quantified with the Nanodrop 2000 (Thermo Scientific) and frozen at -80°C prior to sequencing.

Blood samples in PAXgene Blood RNA tubes were thawed to room temperature and processed with the PAXgene Blood RNA kit (PreAnalytiX, QIAGEN) as per the manufacturer’s protocol for manual purification of total RNA. Eluted total RNA was quantified with the Nanodrop 2000 and frozen at -80°C prior to sequencing.

5.2.3 Dual RNA sequencing

All samples were sent to the Australian Genome Research Facility (AGRF) where they were assessed for RNA quality on an Agilent 2100 Bioanalyzer. Based on this quality assessment, ten blood samples (five rAOM and five COME) and five MEE samples (paired to five of the selected blood samples) were selected for RNA sequencing from a total of 16 patients recruited. All 15 samples underwent ribosomal RNA depletion using the Ribo-Zero Gold rRNA Removal Kit (Illumina), using a 1:1 mixture of human and bacterial probes for the MEE samples; and the 10 blood samples additionally underwent globin mRNA depletion with the Globin-Zero Gold rRNA Removal Kit (Illumina). Sequencing libraries were prepared with the TruSeq Stranded Total RNA (Illumina) sample preparation protocol which involves fragmentation of RNA, cDNA synthesis with random priming using SuperScript II Reverse Transcriptase (Invitrogen), 3’ adenylation of DNA fragments, ligation of sequencing adapters and then enrichment of DNA fragments via 13-cycle PCR. Samples were sequenced across two lanes of a HiSeq 2500 (Illumina) with single-end 100 bp reads. With the HiSeq 2500 producing about 200M clusters per lane, the ten blood samples were sequenced on one lane (expected yield 20M reads per sample) and the five MEEs on the second lane (expected yield 40M reads per sample). These expected yields were selected based on Pérez-Losada et al. (2015), who achieved dual host and microbial metatranscriptome sequencing with approximately 40M raw reads per sample in a similar environment (nasal epithelium). This increased depth of sequencing was not required for the blood samples as they are not relevant to the microbial metatranscriptome; due to budget constraints these ten samples were sequenced on a single lane.
5.2.4 Data processing and analysis

Raw sequencing reads were analysed for quality via FastQC v0.11.7. Illumina TruSeq adapters and low quality bases were trimmed, and reads shorter than 30 bp were removed with Trimmomatic v0.38. Reads mapping to the PhiX genome with Bowtie2 v2.3.4.1 were discarded, and reads matching ribosomal RNA were removed with SortMeRNA v2.1 (all bacterial, archaeal and eukaryotic ribosomal RNAs). The remaining sequencing reads were mapped to the human reference genome GRCh38 with HISAT2 v2.1.0 using the provided pre-built index (Ensembl release 84), and both mapped (human) and unmapped (microbial) reads were retained for further analysis.

5.2.5 Metatranscriptomics pipeline

Microbial reads were analysed with the SAMSA2 metatranscriptomics pipeline v2.0.0. Only the five MEE samples underwent metatranscriptome analysis as the aim of this study was to explore the microbial expression profile of the middle ear only. Briefly, microbial reads were aligned to the non-redundant bacterial proteins and all viral proteins present in release 89 of the NCBI RefSeq database, as well as the SEED Subsystems hierarchical database (a hierarchical collection of genomic functions) using DIAMOND v0.8.38.100 in sensitive mode. SAMSA2’s ‘specific organism’ function was used to summarise the reads mapping to the bacterial genera of interest Alloiococcus and Turicella. Additionally, all microbial reads were mapped to a small database containing only the RefSeq proteins from Alloiococcus and Turicella genomes, to detect proteins from these organisms which may have erroneously mapped to more well-characterised organisms in the full database. Modified SAMSA2 R scripts were used to calculate Shannon and Simpson diversity and create summarised figures and tables using R v3.5.1 and package ggplot2 v3.0.0.

5.2.6 Human transcriptomics pipeline

Reads that mapped to the human genome with HISAT2 were further processed as recommended by Pertea et al. (2016). The abundance of known human transcripts was estimated using StringTie v1.3.4, and the resulting transcript and gene abundance tables were tested for differential expression in Ballgown v2.12.0. Two separate model designs were used to fit a differential expression model at both the gene and transcript level. For both model designs, transcripts with a variance in
FPKM (fragments per kilobase of transcript per million) values of less than 1 across the included samples were filtered out prior to fitting the models.

Model 1 addressed the first aim of the study; to identify differentially expressed host genes in the blood from children with rAOM (n = 5) compared to those with COME (n = 5). Diagnosis (rAOM or COME) was the covariate of interest, and the model adjusted for the potential confounding effect of age. An unadjusted model was also fitted.

Model 2 addressed the third aim of the study; to compare the expression levels of genes involved in the hypoxia pathway in MEE samples (a potentially hypoxic environment, n = 5) to blood samples (an environment with an oxygen tension equivalent to the healthy middle ear, n = 5) from the same child. Sample type (MEE or blood) was the covariate of interest and this model adjusted for patient ID to account for the paired sample design. After fitting the model, the results were restricted to a list of hypoxia-related genes. This list included the genes assessed on the Human Hypoxia Signaling Pathway and the Human VEGF Signaling Pathway RT² Profiler PCR Arrays (QIAGEN), the latter of which (the equivalent mouse array, previously sourced from SA Biosciences) was utilised in Cheeseman et al. (2011); in addition to all human genes annotated to the Gene Ontology (GO) term ‘cellular response to hypoxia’. The GO annotation list was generated from QuickGO (www.ebi.ac.uk/QuickGO) with GO version 2018-09-23. A complete list of these genes (n = 296) and their Ensembl ID numbers can be found in Appendix C. Q-values were calculated by adjusting the p-values for these genes only with the false discovery rate (FDR) method.

For both model designs, significantly differentially expressed genes were considered to be those with a q-value of less than 0.05.

5.3 Results

5.3.1 Isolation and sequencing of total RNA from middle ear effusion and blood specimens

A total of 8 patients with rAOM and 8 with COME were recruited to the study. One blood sample was collected from each child and bilateral effusion was present in all but 1 patient, so a total of 16 blood samples and 31 MEE samples were collected and processed for RNA isolation.
The RNA yield from the 31 MEE samples ranged from 32.5 ng/µl to 1657 ng/µl. However, the RNA integrity numbers (RINs) for all MEE samples were generally poor, with 6/31 (19%) samples having no RIN, 18/31 (58%) having a RIN below 4.5, and the remaining 7 samples (23%) with RINs between 6.1 and 7.3. The RIN was not influenced by the MEF type (mucoid, purulent, mucopurulent or serous). Figure 5.1 summarises the selection of five MEE samples and ten blood samples for dual RNA-seq. Five within-child pairs of MEE and blood samples were chosen based on RNA quality of the MEEs. As the blood samples did not have these quality issues, an additional 5 blood samples were selected to provide an equal number of rAOM and COME samples for the comparison of blood expression profiles between these phenotypes (Aim 1). These blood samples were selected based on RNA yield and a balance of age and sex.

The MEE samples chosen for sequencing had RINs of 6.4 or above, however the Bioanalyzer fluorescence peaks were minimal and noisy compared to the blood samples, with low molecular weight RNA fragments and weaker rRNA bands. Bioanalyzer gel images and fluorescence peaks are shown in Figure 5.2 for the 5 MEE, 5 paired blood, and 5 additional blood samples that were sequenced.

The 5 MEE samples produced an average yield of 45.2M raw sequencing reads per sample (expected yield 40M reads). The 10 blood samples yielded an average of 22.9M reads per sample (expected yield 20M reads). An average of 91.5% of reads from MEE samples and 95% from the blood samples remained after pre-processing steps. Of these ‘clean’ reads, 97.8% mapped to the human genome in the MEE samples, with a 99.0% mapping rate in the blood samples. This resulted in an average of 2.2% of the clean reads in the MEE samples (0.94M reads) remaining as microbial reads for metatranscriptome analysis.

5.3.2 Differences in the host expression profiles: rAOM compared to COME (Aim 1)

The pilot study aimed to explore the differentially expressed genes between the two OM phenotypes in both the MEE (local host response to disease) and blood (systemic host response to disease). However, as the poor quality of MEE samples precluded the sequencing of a sufficient number from each phenotype, only a comparison of blood expression profiles could be performed. Analysis with StringTie and Ballgown did not reveal any significantly differentially expressed genes between these two phenotypes, both in an unadjusted model and in one adjusting for the patients’ age.
Figure 5.1: Description of samples selected for RNA-seq. Issues with the quality of RNA from MEE samples led to a selection process for samples undergoing RNA-seq. Grey boxes indicate the exclusion of MEE samples. Green boxes indicate samples that underwent RNA-seq. *Note that this patient had a parental report of two previous tympanic membrane perforations, but the surgeon did not note any perforation at the time of surgery. No other patients whose MEEs underwent RNA-seq had current or past tympanic membrane perforations.
Figure 5.2: Bioanalyzer gel images and peaks for the sequenced samples. Shown are the Bioanalyzer results for the a) 5 MEE samples, b) 5 corresponding blood samples and c) 5 additional blood samples selected for RNA-seq. Note that the positions of the 28S and 18S rRNA bands vary as not all samples were run on the same Bioanalyzer chip but do correspond to the correct nucleotide length.
At the transcript level, one transcript was significantly differentially expressed ($q = 0.016$), but only in the unadjusted model. This transcript was from the gene \textit{DGAT2} (diacylglycerol O-acyltransferase 2), and was upregulated in the COME blood samples compared to the rAOM blood samples (fold change $= 3.6$). The scaled FPKM values are shown in Figure 5.3.

### 5.3.3 Characterisation of the microbial metatranscriptome of MEE (Aim 2)

In the five MEE samples, an average of 26.8\% of the remaining microbial reads (those that did not map to the human genome with HISAT2) mapped to the RefSeq bacterial and viral protein database. Only 3.1\% of the reads (about 24,300 reads) per sample mapped to the SEED subsystems hierarchical database. Figure 5.4 shows the relative and absolute abundance of organisms from the RefSeq database contributing reads to the metatranscriptome. The five MEE samples are highly similar in composition and diversity (mean Shannon index $= 3.40$, SD 0.13; mean Simpson index $= 0.93$, SD 0.0091). The ten most abundant organisms are bacterial species that are not known to be found in the human middle ear and most of these species are well-characterised human pathogens, indicating that this is not an accurate representation of the organisms contributing to the activity in the middle ear.

The functional assignments of the reads mapping to the RefSeq database are shown in Table 5.1. The ten most abundant functions are largely unknown. The annotations consist primarily of hypothetical proteins, proteins with domains of unknown function (DUF) and include a function that should not be found in the microbial metatranscriptome (immunoglobulin heavy chain). Very few reads mapped to the SEED subsystems database; the most abundant functions grouped at the broadest hierarchy level are listed in Table 5.2.

\textit{Alloiococcus otitidis} and \textit{Turicella otitidis} are potential novel otopathogens commonly detected in the middle ear fluid of otitis-prone children. To determine their transcriptional activity in the middle ear, the RefSeq functional results were subsetted to the functions annotated to these organisms. This subset included a minimal number of reads, with only 1255 reads attributed to \textit{A. otitidis} and 73 to \textit{T. otitidis}. The dominant functions that these reads represented are shown in Table 5.3a and 5.3b. The alternative approach, aligning all reads to a small database containing only \textit{A. otitidis} and \textit{T. otitidis} RefSeq proteins, resulted in a total of 28,619 mapped reads.
Figure 5.3: Abundance of the differentially expressed $DGAT2$ transcript. The log$_2$ transformed FPKM values for this transcript are shown for the COME blood samples and rAOM blood samples.
Figure 5.4: Organism composition of the middle ear metatranscriptome. The proportion and number of reads mapped to proteins for the ten most abundant organism annotations (by total reads across all samples) are shown.
### 5.3. Results

**Table 5.1: RefSeq functions detected in the middle ear metatranscriptome.** The ten most abundant (by total reads across all samples) RefSeq proteins are shown.

<table>
<thead>
<tr>
<th>Function</th>
<th>TSO002FR</th>
<th>TSO009FR</th>
<th>TSO011FL</th>
<th>TSO012FR</th>
<th>TSO016FL</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypothetical protein</td>
<td>81967</td>
<td>367534</td>
<td>58749</td>
<td>387524</td>
<td>29188</td>
</tr>
<tr>
<td>hypothetical protein, partial</td>
<td>10852</td>
<td>60182</td>
<td>5772</td>
<td>63103</td>
<td>3784</td>
</tr>
<tr>
<td>DUF1725 domain-containing protein</td>
<td>22612</td>
<td>37652</td>
<td>18535</td>
<td>37292</td>
<td>14284</td>
</tr>
<tr>
<td>DUF1725 domain-containing protein, partial</td>
<td>7023</td>
<td>16889</td>
<td>6309</td>
<td>15431</td>
<td>4445</td>
</tr>
<tr>
<td>MULTISPECIES: hypothetical protein</td>
<td>1548</td>
<td>7464</td>
<td>529</td>
<td>8233</td>
<td>1051</td>
</tr>
<tr>
<td>genome polyprotein</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>14534</td>
<td>0</td>
</tr>
<tr>
<td>molecular chaperone DnaK</td>
<td>2489</td>
<td>4194</td>
<td>1188</td>
<td>620</td>
<td>1654</td>
</tr>
<tr>
<td>immunoglobulin heavy chain variable domain-containing protein</td>
<td>48</td>
<td>36</td>
<td>6297</td>
<td>245</td>
<td>64</td>
</tr>
<tr>
<td>endonuclease, partial</td>
<td>944</td>
<td>1258</td>
<td>754</td>
<td>1171</td>
<td>580</td>
</tr>
<tr>
<td>tetracycline efflux MFS transporter Tet(C)</td>
<td>329</td>
<td>453</td>
<td>583</td>
<td>314</td>
<td>496</td>
</tr>
<tr>
<td>Other</td>
<td>15582</td>
<td>6181</td>
<td>61867</td>
<td>7586</td>
<td>22994</td>
</tr>
</tbody>
</table>
### Table 5.2: Functional content of the middle ear metatranscriptome categorised by the SEED subsystems database.

Reads mapping to the SEED subsystems database are grouped at level 1 hierarchy and the ten most abundant (by average count) are shown.

<table>
<thead>
<tr>
<th>Subsystems Level 1 hierarchy</th>
<th>Average reads mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein biosynthesis</td>
<td>5200</td>
</tr>
<tr>
<td>Proteolytic pathway</td>
<td>2128.2</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>1824.8</td>
</tr>
<tr>
<td>No hierarchy</td>
<td>1563</td>
</tr>
<tr>
<td>Transcription</td>
<td>1488.2</td>
</tr>
<tr>
<td>Resistance to antibiotics and toxic compounds</td>
<td>997.2</td>
</tr>
<tr>
<td>Gram-Negative cell wall components</td>
<td>613.8</td>
</tr>
<tr>
<td>Central carbohydrate metabolism</td>
<td>592.2</td>
</tr>
<tr>
<td>Membrane Transport</td>
<td>545.2</td>
</tr>
<tr>
<td>Lysine, threonine, methionine, and cysteine</td>
<td>481.8</td>
</tr>
</tbody>
</table>
The most abundant functions are shown in Table 5.3c. Many of these functions are involved in basic metabolism and occur widely in the bacterial kingdom.

5.3.4 The role of hypoxia in the otitis-prone middle ear (Aim 3)

A total of 199,184 known human transcripts were detected across the 5 pairs of MEE and blood samples, which was reduced to 36,928 transcripts after removing those with a variance in FPKM values of less than 1. The transcript-level and gene-level differential expression results were then subsetted to the list of 296 genes involved in the response to hypoxia; 205 of these genes were present in the gene-level results, however none of the 36,928 transcripts above the variance threshold were from any of these genes. After FDR correction of the \( p \)-values, none of the \( q \)-values for these genes were below 0.05 (the lowest was 0.061). Furthermore, \( HIF1A \), the major transcription factor involved in the cellular response to hypoxia, was not amongst the genes with the lowest \( q \)-values (\( q = 0.20 \), fold change from blood to MEE = 3.03).
Table 5.3: Activity of *Alloiococcus* and *Turicella* in the middle ear. The number of reads per sample mapped to a) *Alloiococcus* and b) *Turicella* in the RefSeq functional database and c) either of these organisms in a database containing only these species are shown. For each database, the 10 most abundant (by total reads mapped across samples) functions annotated are listed followed by followed by the reads that mapped to other functions (‘Other’).

<table>
<thead>
<tr>
<th>Function</th>
<th>TSO002FR</th>
<th>TSO009FR</th>
<th>TSO011FL</th>
<th>TSO012FR</th>
<th>TSO016FL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) <em>Alloiococcus</em> subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hypothetical protein</td>
<td>117</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>YSIRK-type signal peptide-containing protein</td>
<td>95</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>ABC transporter ATP-binding protein</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dihydrodipolpyl dehydrogenase</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>formate C-acetyltransferase</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DNA starvation/stationary phase protection protein</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3-hydroxyacyl-CoA dehydrogenase</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>metal ABC transporter permease</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>SDR family NAD(P)-dependent oxidoreductase</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>elongation factor Tu</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>805</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>90</td>
</tr>
<tr>
<td>b) <em>Turicella</em> subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hypothetical protein</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>amidohydrolase</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>glycine dehydrogenase (aminomethyl-transferring)</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>formate dehydrogenase</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ribonuclease III</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>universal stress protein</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3-isopropylmalate dehydrogenase</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50S ribosomal protein L9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ABC transporter permease</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aldehyde dehydrogenase family protein</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>40</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>c) <em>Alloiococcus</em> and <em>Turicella</em>-specific database</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elongation factor Tu</td>
<td>120</td>
<td>35</td>
<td>842</td>
<td>16</td>
<td>251</td>
</tr>
<tr>
<td>molecular chaperone DnaK</td>
<td>348</td>
<td>426</td>
<td>186</td>
<td>57</td>
<td>229</td>
</tr>
<tr>
<td>MULTISPECIES: molecular chaperone DnaK</td>
<td>51</td>
<td>544</td>
<td>166</td>
<td>76</td>
<td>212</td>
</tr>
<tr>
<td>MULTISPECIES: elongation factor Tu</td>
<td>77</td>
<td>17</td>
<td>561</td>
<td>9</td>
<td>142</td>
</tr>
<tr>
<td>elongation factor G</td>
<td>92</td>
<td>13</td>
<td>458</td>
<td>8</td>
<td>181</td>
</tr>
<tr>
<td>metal ABC transporter substrate-binding protein</td>
<td>16</td>
<td>2</td>
<td>470</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>50S ribosomal protein L2</td>
<td>29</td>
<td>4</td>
<td>387</td>
<td>1</td>
<td>61</td>
</tr>
<tr>
<td>chaperonin GroEL</td>
<td>79</td>
<td>19</td>
<td>171</td>
<td>10</td>
<td>95</td>
</tr>
<tr>
<td>DNA-directed RNA polymerase subunit beta</td>
<td>40</td>
<td>10</td>
<td>238</td>
<td>1</td>
<td>77</td>
</tr>
<tr>
<td>MULTISPECIES: elongation factor G</td>
<td>24</td>
<td>3</td>
<td>280</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>Other</td>
<td>3531</td>
<td>827</td>
<td>12732</td>
<td>450</td>
<td>3891</td>
</tr>
</tbody>
</table>
5.4 Discussion

This pilot study was carried out to assess the feasibility of a dual RNA-seq approach to study both the host transcriptome and the microbial metatranscriptome in the context of otitis media. Using this approach, the study aimed to 1) compare the host expression profiles in the middle ear and the blood from children with rAOM to those with COME; 2) to gain insight into the functions actively carried out by microbes in the middle ear and 3) to determine whether the hypoxia pathway is upregulated in the otitis-prone human middle ear. The study has demonstrated that the isolation and sequencing of total RNA from MEE samples is challenging, and requires optimisation prior to undertaking larger-scale dual RNA-seq projects on otitis media.

5.4.1 Insights from the pilot data

The host response to rAOM compared to the response to COME (Aim 1)

The first aim of the study could not be entirely fulfilled, as the poor quality of RNA from MEE samples precluded the sequencing of more than 5 samples. However, a preliminary comparison of the systemic response to rAOM and COME via the blood expression profiles could be carried out. Only one transcript, from the DGAT2 gene, was differentially expressed between these phenotypes. This gene is involved in lipid metabolism and has not previously been associated with OM. Due to the small number of replicates for each phenotype in this pilot study, this may be a spurious result. Natural variation in the blood expression profiles between individuals can be influenced by gender, age and time of day and there were insufficient replicates to adequately control for each of these variables in this pilot study. There were no other significantly differentially expressed transcripts or genes, suggesting that this pilot study does not have sufficient power to detect differences between these two phenotypes; the differences may not be substantial and could be in genes with a naturally low expression level, which would require additional replicates and greater depth of sequencing to detect. Furthermore, the HISAT2/StringTie/Ballgown pipeline is designed for transcript-level analysis. As the expression profiles of children with OM has not previously been explored, an initial exploratory gene-level analysis may be more appropriate. It has been suggested by the author of the Ballgown tool that count-based models such as DESeq2 are
more suitable for differential gene expression than an FPKM analysis in Ballgown, which is more appropriate for transcript level analysis. It would be worthwhile analysing the data with multiple count-based tools as all RNA-seq analysis tools perform differently and no one method performs well on all data. This approach would improve confidence that any differentially expressed genes are not false positives, and could be further validated by qPCR or Western blotting.

Activity of microbial organisms in the middle ear (Aim 2)

Previous studies using 16S rRNA gene profiling of MEE samples have indicated that Alloiococcus, Haemophilus, Streptococcus, Moraxella and Staphylococcus are typically the dominant genera in this environment. While the method is not quantitative or necessarily reflective of the true proportions of these bacteria in situ, it would be expected that some of the transcriptional activity in MEE samples be attributed to these organisms. Amongst the organisms apparently contributing to the microbial activity in the middle ear in this study were S. pneumoniae and S. aureus, however these organisms are also well-characterised pathogens of other diseases – as are many of the other organisms contributing to the majority of microbial activity in the 5 MEE samples. Activity from organisms such as Mycobacterium tuberculosis and Acinetobacter baumannii, which can be highly pathogenic, is unexpected and these organisms are not known to infect the middle ear. The five MEE samples were similar in composition and were diverse, contrasting what has been demonstrated in the same population of children via 16S rRNA profiling and shotgun metagenomics (see Chapters 2 and 3). The transcriptional activity of the microbes in the middle ear should represent a subset of the organisms observed in this population using DNA-based methods, as not all organisms detected in this way are necessarily viable or active; instead, these results apparently detect several new organisms. Due to the poor RNA quality of MEE samples, it is most likely that the profile of organisms observed in the current study via RNA-seq are not biologically meaningful and are observed as noise, contaminants or a product of the degraded input RNA.

Future studies on the metatranscriptome of low biomass clinical samples may therefore benefit from the use of negative sequencing controls. While the presence of bacterial DNA in sequencing reagents is known to affect the DNA sequencing profiles of low biomass specimens, such controls are not commonly sequenced in metatranscriptomics studies. RNA sequencing studies are most often comparative, using samples with baseline gene expression profiles as negative controls as opposed to “blank” samples. Future studies that aim to explore which microbial genes are
expressed rather than to compare gene expression between conditions could therefore use blank samples (kit reagents only) or deliberately degraded RNA as a point of comparison to determine whether the detected organisms and functions are likely to be spurious.

Understanding the role of hypoxia in the otitis-prone middle ear (Aim 3)

None of the genes involved in hypoxia were significantly differentially expressed between the MEE and blood of children with severe OM. As described above, it is likely that the minimal number of replicates may inhibit the detection of these differences. However, it may also be necessary to revisit the use of the venous blood control.

In this pilot study venous blood was used as a baseline ‘normoxic’ control for comparison with MEE samples to determine whether genes involved in the cellular response to hypoxia are upregulated in the middle ear. This was the approach taken by a similar study, which determined that in mouse models of OM, the middle ear is hypoxic, and the HIF-VEGF pathway is upregulated and may be a useful target for therapy.\textsuperscript{408} Mixed venous blood is potentially the best available normoxic control for MEE samples, as it has been reported that the normal middle ear lumen and mixed venous blood contain similar partial pressures of oxygen (PO\textsubscript{2}),\textsuperscript{410} which is approximately 40 mmHg.\textsuperscript{410,426} However, some studies measuring the PO\textsubscript{2} in MEE have reported similar\textsuperscript{427} or lower\textsuperscript{428} values to mixed venous blood, so the literature is not consistent on this point and no recent studies have measured PO\textsubscript{2} in MEE. Furthermore, true mixed venous blood is defined as blood entering the pulmonary artery (after traversing the body), which can only be invasively sampled.\textsuperscript{429} Venous blood collected from the hand vein in this study may contain a different level of PO\textsubscript{2}; it is therefore possible that the blood sampled from the hand does not contain a PO\textsubscript{2} similar to the patient’s healthy middle ear. If it is instead similar to the PO\textsubscript{2} in MEE, hypoxia-related genes may not be differentially expressed despite hypoxic (relative to normal) conditions in the middle ear. If venous blood has a higher PO\textsubscript{2} than the healthy middle ear, false positives may be observed where the hypoxia pathways are upregulated in comparison to an environment with a higher PO\textsubscript{2} rather than to the normal state of the middle ear. Confirming these values in children would require further research on the PO\textsubscript{2} in these conditions.

Ideally, the most appropriate control is the healthy middle ear. This is straightforward
in studies utilising animal models where the entire middle ear can be harvested, but is not applicable to studies in humans. A potential alternative approach to the use of a blood control may be to sample epithelial cells from the middle ear mucosa in both cases and controls. Control samples may be obtained from children undergoing cochlear implant surgery, or any other procedure where the middle ear is healthy but becomes accessible during surgery. However, this would remain a challenging task due to the infrequency of cochlear implant procedures and the potentially low number of epithelial cells obtained via mucosal swabs. Due to the nature of MEE, finding an appropriate sample type for use as a ‘healthy’ or ‘normoxic’ control for the middle ear will always remain problematic.

5.4.2 Preparation of RNA from MEE specimens

This pilot study demonstrated that freshly-collected MEE samples kept on ice for no longer than 1 hour prior to homogenisation in TRIzol Reagent produced a moderate RNA yield, but did not produce RNA of good quality; this was unrelated to the appearance of the MEF specimen. The majority of MEF samples in this pilot study had a RIN below 4.5; the only other study to have attempted dual RNA-seq on clinical respiratory specimens sequenced samples with a RIN above 5.0. The otitis-prone middle ear environment likely promotes the degradation of RNA due to the presence of neutrophils and apoptosing cells in an inflamed environment where RNAse enzymes are released. Bacterial mRNAs in particular are very rapidly degraded and may remain difficult to capture even with RNA of good quality. This presents a challenge for extracting intact RNA from MEE samples, and few studies have previously done so. These studies have primarily been performed in animal models in a laboratory where the RNA can be processed immediately or the middle ear mucosal tissue can be sampled. Those studies that analysed RNA from human MEE samples have processed them fresh; Kerschner et al. (2007) extracted RNA “immediately” after collection and Kaur et al. (2016) did not specify how soon the samples were processed, but did not freeze them or introduce an RNA protection agent. This study obtained good quality MEF samples with a 260/280 absorbance ratio of at least 2.0; suggesting that RNA isolation in the current study can be improved. Similar to these studies, the current pilot study aimed to avoid immediate freezing of the samples, as degradation of RNA would begin to occur upon thawing. The addition of a protective agent such as RNAlater or RNAProtect was also avoided, as these reagents are designed for solid tissues or cultured cells and have not been optimised for mucoid samples. There has also been evidence that
the expression profile can be altered if the reagent takes time to diffuse through the sample; protective reagents may not rapidly penetrate the thick mucus that can be found in MEE samples. Treatment of human specimens with dithiothreitol (DTT) to break down mucus has also been associated with substantial changes in RNA-seq expression profiles and is not an ideal option.

Results from the current study suggest that if the samples can be processed fresh it must be done immediately after collection, which was not possible due to the need to travel between operating theatre and laboratory. Additionally, the time spent on ice is intended to slow the degradation of RNA, but may also alter the expression profile as the cells respond to cold temperatures. Immediate freezing of the sample on dry ice for RNA extraction at a later date (avoiding prolonged storage at -80°C and thawing on ice) may be the best option to prevent degradation of RNA in MEE samples, where RNA extraction cannot be performed immediately. Ideally, all samples would then be thawed on ice and undergo RNA extraction in as few batches as possible after a similar time in storage to limit the confounding of expression data with batch effects and storage time. Regardless, optimisation of an RNA protection reagent for mucoid samples would be beneficial, as the current study has shown that any time spent on ice has the potential to contribute to RNA degradation.

5.4.3 Achieving appropriate depth of sequencing for the dual analysis of host and microbiome transcriptomes from MEE specimens

Only about 0.5% of the original raw data per sample (~0.25M reads) aligned to bacterial and viral sequences in the most recent RefSeq release (89). This is a very small volume of data, which makes it highly challenging to estimate the contribution of different microbial organisms to the activity occurring in the samples. It was apparent in this pilot study that these remaining reads did not represent true biological signal; the unexpected organisms may have been detected via the mapping of reads corresponding to highly conserved, commonly expressed genes that may represent the proportion of intact RNA in the samples, or via erroneous mapping to common protein sequences. The functional results also demonstrate the most abundant proteins detected are of unknown function, human origin or are constitutively expressed in a wide range of organisms. While constitutively expressed or widely conserved proteins are also likely to be the most abundant functions observed in
datasets generated from good-quality RNA, a much greater depth of sequencing than that achieved in the current study is required to explore the species-specific transcripts occurring at a lower abundance which may be of interest.

Many dual RNA-seq studies that focus on the transcriptomes of the host and a single pathogen produce at least 20M reads per sample, sometimes up to a few hundred million; with a range of < 1% to 67% of these reads being microbial (reviewed in Westermann et al. (2017)). Typically, reads from the host constitute the majority of the sample and enrichment of microbial transcripts is often required. There is no universal recommended depth of sequencing for analysis of a metatranscriptome, as this depends on the read length, diversity of the sample, size of the microbial genomes and to what extent the detection of rare transcripts from low-abundance species is required. A depth of 5-10 million reads has been recommended to extensively cover the transcriptome of an individual bacterial species, with a minimum of 10 million annotated reads (40 million raw reads) recommended to achieve accuracy in transcript abundance in the human gut metatranscriptome. For dual RNA-seq, a minimum of 200 million rRNA-depleted reads has been suggested for studying the transcriptome of the host and a single pathogen.

Few studies have taken the approach of sequencing the transcriptome of both the host and the microbiome. Pérez-Losada et al. (2015) generated 41.4M single-end 100 bp reads per sample on average, with 33.7M aligning to the human genome and 1.8M remaining as microbial reads (~4% of original data); which was sufficient to detect differences in the functions of the nasal microbiota between asthmatic and non-asthmatic children. Wesolowska-Andersen et al. (2017) generated a lower average of 12.6M reads per sample in nasal brushings, though this was sufficient for their study which focused on the detection of respiratory viruses, rather than developing a profile of all microbial functions. Ilott et al. (2016) conducted a study on the colon contents of mice, and generated an average of 45.4M raw reads per sample, with 17.0% of reads mapping to the host genome, 8M aligning to the NCBI NR protein database and 6.8M receiving a genus assignment. Based on the read counts from our current pilot study, < 1 million microbial reads is insufficient for metatranscriptomics analysis as this does not match the depth achieved in these other studies, and a further large proportion of these reads did not map to known sequences. Dual RNA-seq is challenging in sample environments where the overwhelming majority of RNA comes from the host; utilising only 0.5% of the input reads for metatranscriptomic analysis is highly cost ineffective, as deeper sequencing – at greater cost – is then required to adequately capture this small proportion.
Thus, improvements in RNA yield and quality from MEE samples and the dual RNA-seq approach to capture sufficient microbial information are required prior to the application of dual RNA-seq in larger studies of otitis media.

5.4.4 Improved methods for studying microbial activity in the middle ear

This pilot study has demonstrated some of the major issues faced when applying transcriptomics to MEE samples. Given the rapid degradation of RNA – particularly bacterial mRNA – in MEE samples and the high proportions of host RNA, it will remain challenging to obtain sufficient data of quality from both the host and microbial transcriptomes in this context. The issue of sequencing depth may be improved by microbial RNA enrichment; several in vitro dual RNA-seq studies have utilised such a step; even within in vitro culture models the proportion of uninfected host cells is far in excess of the infected cells. Some studies focusing on the bacterial transcriptome in a host cell environment have used the MICROBEnrich kit to deplete host RNA, though unlike the NEBNext Microbial DNA Enrichment Kit (see Chapter 3) the host RNA cannot be eluted separately.

One possibility is a ‘separated’ dual RNA-seq, where the host RNA is physically separated from the microbial RNA and sequenced separately. This could be achieved via poly-A tail selection to capture eukaryotic mRNAs, which are then sequenced as separate samples to the remaining microbial RNAs, similar to the approach taken by Afonso-Grunz et al. (2015). Poly-A tail selection has also been successful as an enrichment method for dual RNA-seq by applying it to half of the sample and recombining the samples before sequencing. These techniques could facilitate deeper sequencing of the microbial component, ideally to a sufficient depth required to detect the less common microbial transcripts from low-abundance but biologically important species. However, additional processing steps must be undertaken with caution as they present further opportunity for the degradation of or loss RNA.

While enrichment techniques hold promise for dual RNA-seq studies, the question of the roles of A. otitidis and T. otitidis in the middle ear may be better investigated by other methods. The genomes of these organisms, while annotated, currently remain in draft stage and contain many unknown functions and hypothetical proteins; limiting the information on their activity in the middle ear that can be inferred from metatranscriptomic data. This is indeed the case for any organisms of interest that have not been previously well-characterised; the functions cannot be detected if they...
are not known. *In vitro* cell models utilising human middle ear epithelial cells may generate more valuable information on how these organisms interact with the epithelium in the middle ear and whether they exhibit pathogenic characteristics. These organisms have not been extensively studied *in vitro*; the majority of the literature in which they appear simply describe their prevalence in the middle ear and EAC. However, determining the role of *A. otitidis* and *T. otitidis* in the middle ear may always remain a low priority research topic due to their current status as commensals of the EAC and thus far limited evidence for pathogenicity.

### 5.5 Conclusions

This pilot study has demonstrated that the isolation of RNA from human MEE samples is challenging due to the rapid degradation of RNA in an inflammatory environment. The dual RNA-seq approach is further complicated by the overwhelming proportion of host-derived RNA. These issues led to an insufficient volume of microbial data to accurately assess the microbial metatranscriptome in the human middle ear. No substantial differences between the blood expression profiles of children with rAOM compared to COME were detected, and there was no upregulation of hypoxia-related genes in the MEE. Differential expression of host genes may not be detectable in this preliminary analysis with few replicates. Optimisation of RNA isolation from MEE samples is required to produce better quality RNA, and enrichment or separation of microbial RNA from host RNA is necessary to achieve the depth of sequencing required for dual host and microbial transcriptome profiling. Metatranscriptomic data, in conjunction with metagenomics data, still holds promise for the development of a more accurate view of the microbial dynamics of the otitis-prone middle ear.
Chapter 6

General discussion

Recurrent acute otitis media remains a significant problem in young children. High rates of antibiotic prescriptions and the need for surgical intervention contribute to the substantial economic burden of this disease on the healthcare system, and impact the quality of life of children with rAOM and their families. rAOM is a polymicrobial disease that can be resistant to treatment, and the composition and role of the microbial communities in the URT of children with rAOM have remained poorly understood. A comprehensive understanding of the organisms in the URT that may be contributing to the pathogenesis of rAOM, or increasing a child’s resistance to developing rAOM, is an important step towards the development of new, more effective therapies.

This thesis aimed to explore two avenues of alternative therapy for rAOM by studying the URT microbiota: 1) the possibility of novel otopathogens contributing to rAOM, which may be new targets for therapy; and 2) the potential for rAOM-resistant children to harbour beneficial microbes, with a view to developing new agents for therapy. The use of high throughput next-generation sequencing technologies for the characterisation of microbial communities has been pioneered in the human gut and in environmental contexts, and the state of the URT microbiome field is somewhat behind in comparison. This thesis therefore also presents a novel analysis of the URT microbiota using metataxonomics, metagenomics and metatranscriptomics techniques; applying the latter two to the URT of children with OM for the first time.
Understanding the role of *Alloiococcus otitidis* and *Turicella otitidis* in the otitis-prone middle ear

One of the major findings of this work was the prevalence of the organisms *A. otitidis* and *T. otitidis* in children with rAOM. These species have a reputation as ‘controversial’ otopathogens; they have frequently been associated with OM in the literature due to their prevalence in the middle ear of children with OM, however their role in the disease, if any, remains unknown. It is challenging to understand the role of *A. otitidis* and *T. otitidis* in OM, as unlike the known otopathogens they are not well-characterised species and can be more difficult to detect. *A. otitidis* requires extended incubation to isolate in culture\textsuperscript{102} and there are currently no species-specific primers for *T. otitidis* described in the literature; *T. otitidis* isolates are also not often distinguished from commensal *Corynebacteriaceae*.\textsuperscript{95} However, as further studies investigating the URT microbiome in children with OM emerge, the prevalence of these organisms in the MEF and EAC of children with OM is becoming increasingly apparent. This thesis investigated the microbiome of the MEF, EAC and nasopharynx of children with rAOM for the first time. In conjunction with what is known about the prevalence of *A. otitidis* and *T. otitidis* in these sites in children with different phenotypes of OM, this work has begun to reveal some patterns about these organisms and their potential role in OM.

The first observation is that *A. otitidis*, and perhaps *T. otitidis*, appears to be involved in chronic inflammation more so than acute infection. *A. otitidis* has been reported at high prevalence (often greater than that of the major otopathogen species) in children with OME or COME,\textsuperscript{77,78,96,97,99,102,103} and it has been reported at a somewhat lower prevalence in purulent fluid and children with AOM.\textsuperscript{77,78,104} In this thesis, both organisms were observed at high prevalence and relative abundance in the MEF of children with rAOM via both metataxonomics and metagenomics. This may seem contradictory to what these previous studies have observed, however in Western Australia, grommet surgeries are not performed during active episodes of AOM; the cohort therefore represents children who are between episodes of AOM, during which it is common for children to experience OME. This suggests that if *A. otitidis* and *T. otitidis* are involved in the pathogenicity of OM, their role may be in the perpetuation of inflammation rather than directly causing acute infection.

This thesis has also advanced the understanding of the way in which *A. otitidis* and *T. otitidis* are likely to colonise the middle ear, which appears to be different than that of the known otopathogens. *A. otitidis* and *T. otitidis* have been reported as uncommon or absent from the nasopharynx,\textsuperscript{75,96,106,107,110} but also inhabit the
EAC. The work in Chapter 2 supports these findings and demonstrates that the relative abundance of these organisms is high in both the MEF and ECS of children with rAOM, but is comparatively very low in the nasopharynx of the same children. This suggests, and has been speculated by some, that the organisms enter the middle ear from the ear canal rather than ascending from the nasopharynx as the known otopathogen species do. However, some studies including the work described in this thesis detected *A. otitidis* or *T. otitidis* in specimens from children in which tympanic membrane perforations (including previous grommets) had not occurred. This suggests three possibilities: 1) *A. otitidis* and *T. otitidis* inhabit the normal middle ear, prior to the onset of OM; 2) they enter the middle ear through undetectable, minor tympanic membrane perforations; 3) they are detectable in MEF specimens as a result of contamination from the EAC during sampling.

The first possibility is challenging to address, as the healthy middle ear is a very low biomass environment, where samples are highly prone to environmental contamination and the area is inaccessible without surgery. Some attempts have been made to survey the microbiota of the healthy middle ear, and these studies have not consistently reported the presence of *A. otitidis* and *T. otitidis* and some have not found evidence for the presence of bacteria in the normal middle ear at all. This will remain a difficult possibility to address, as it is also possible that the organisms are only present in the normal middle ears of children who go on to develop OM.

Chapters 2 and 3 attempted to address the other two possibilities; that *A. otitidis* and *T. otitidis* are present in MEF because they entered the middle ear from the EAC, or because MEF specimens are inadvertently contaminated with EAC flora. In Chapter 2, there were no additional bacterial genera detected in the EAC that were not present in the MEF. With no EAC-specific organisms detected, the possibility of contamination cannot be ruled out. In Chapter 3, the overwhelming proportion of human DNA sequence in MEF specimens, resulting in a very low volume of remaining microbial data, meant that a strain-level comparison of *A. otitidis* and *T. otitidis* sequences from each site could not be carried out. The findings of this thesis demonstrate that there are three key areas of future research that need to be established in order to understand the role of *A. otitidis* and *T. otitidis* in the pathogenesis of OM.

It must first be established whether these organisms are inhabiting the MEF or whether they are merely contamination from the EAC during sampling. To determine this, the bacterial load and viability of these two organisms from both sites within the same child should be measured to determine the absolute abundance of these
organisms in comparison with the known otopathogens, with strict criteria; no prior tympanic membrane perforations including grommets, sterilisation of the ear canal prior to the incision in the tympanic membrane, and specific care taken to ensure the MEF or the surgical instrument do not contact the ear canal.

If this provides evidence that *A. otitidis* and *T. otitidis* do inhabit the middle ears of children with OM and no previous tympanic membrane perforations, the second thing to be established is how the organisms gain entry into the middle ear. Quantitative measurements (i.e. qPCR) of these species in the middle ear, ear canal and nasopharynx of children with OM as well as a genome-level comparison of isolates from these sites may indicate the origin of middle ear strains of *A. otitidis* and *T. otitidis*. If the ear canal appears to be a reservoir for these organisms in the middle ear, even in children with previously intact tympanic membranes, the ability of the organisms to pass through microscopic tympanic membrane perforations or to cross the tympanic membrane by some other means should be considered as a possible mechanism.

Finally, the activity of these organisms in the middle ear of children with OM needs to be explored. This was the focus of the metatranscriptomics work in Chapter 5, however the application of this method to MEF specimens requires further optimisation before it can provide reliable results. Further experiments exploring their behaviour including gene and protein expression *in vitro* with middle ear epithelial cells or *in vivo* in animal models, perhaps in a coinfection model with the otopathogen species, will contribute to an understanding of whether these organisms are pathogenic, commensal, or something in between. It is important to further characterise the role of these organisms in OM as they may be a useful target for therapy; further understanding about the microbial pathogenesis of OM may lead towards new approaches to resolving recurrent or chronic OM.

The role of *Corynebacterium spp.* and *Dolosigranulum pigrum* in nasopharyngeal health and the feasibility of probiotic treatment for children with rAOM

Another major finding of this thesis was the association of *Corynebacterium* and *Dolosigranulum* with the rAOM-resistant nasopharyngeal microbiota. As reviewed in Chapter 4, several studies have now reported a strong and consistent association between nasopharyngeal health and the co-occurrence of *Corynebacterium* species and *D. pigrum* in a range of URT diseases. However, little work has been done to take
this information one step further and determine whether this association indicates a causative relationship between these organisms and resistance to disease. The *in vitro* assays in Chapter 4 demonstrated that the species *C. pseudodiphtheriticum* exhibits inhibitory activity against OM-derived strains of *M. catarrhalis*, providing a foundation for future work aimed at understanding the contribution of these organisms to the healthy nasopharyngeal microbiota, particularly in the context of OM.

*M. catarrhalis* is the least common of the known otopathogens. However, as reviewed by Perez and Murphy (2017) the organism may be underestimated as an otopathogen due to its ability to synergistically support NTHi and *S. pneumoniae* rather than being directly responsible for severe OM alone.66 The finding in this thesis was that *M. catarrhalis*, but not NTHi or *S. pneumoniae*, is inhibited by *C. pseudodiphtheriticum*. A possible mechanism by which *C. pseudodiphtheriticum* contributes to resistance to rAOM in the nasopharynx may be that the reduction of *M. catarrhalis* interferes with the pathogenic potential of NTHi and *S. pneumoniae*, which benefit from the activity of *M. catarrhalis*; as Perez and Murphy (2017) suggest, the reduction of *M. catarrhalis* may improve the success of antimicrobial treatment for coinfections in OM by removal of its contribution to the resistance of the other otopathogens.66 However, the role of *D. pigrum* to nasopharyngeal health remains unclear; in Chapter 4, it did not appear to contribute to inhibition.

It has been speculated that *Corynebacterium* and *Dolosigranulum* co-occur because *D. pigrum* produces lactic acid, which acidifies the local environment and selects for the growth of *Corynebacterium* species.254 Lactic acid itself has been shown to be beneficially antimicrobial and anti-inflammatory in the vaginal microbiota, where it is produced by lactobacilli abundant in the healthy vagina.385 Like the nasopharynx, the vaginal environment normally contains a low diversity of microbes. As the work in this thesis suggested that the inhibition of *M. catarrhalis* by *C. pseudodiphtheriticum* was not mediated by a secreted molecule, it is possible that the effect of pH is involved in this process, the effects of which may be enhanced in the presence of commensal *D. pigrum* strains that could not be assessed in this thesis.

Future studies aiming to define the role of these organisms in resistance to rAOM should focus on the dynamics of *Corynebacterium* and *Dolosigranulum* abundance in the nasopharynx over time. Longitudinal nasopharyngeal microbiome data, such as the cohort analysed in Chonmaitree *et al.* (2017),171 to specifically track the abundance of these two genera would be useful. It should be determined whether
their low abundance in children with rAOM contributes to disease susceptibility (i.e. a low abundance of *Corynebacterium* and *Dolosigranulum* leads to rAOM) or whether antibiotic use and otopathogen colonisation cause a reduction in these organisms (i.e. their low abundance is a result of rAOM). Measuring the absolute abundance in terms of bacterial load is essential, as fluctuations in the relative abundance of organisms are dependent on the abundance of other organisms in the sample, and may not reflect true changes in species populations.

As well as further defining the role of *Corynebacterium* and *Dolosigranulum* in resistance to rAOM, the immediate next step towards assessing whether these organisms can be utilised as probiotics is to identify appropriate strains for further testing. Isolates of *D. pigrum* and *C. pseudodiphtheriticum* as well as *C. accolens* and *C. propinquum* from the nasopharynx of children without a history of OM should be screened for activity against clinical otopathogen strains. The work in Chapter 4 was preliminary and utilised non-commensal strains to identify species-wide inhibitory effects, and these strains are not suitable for probiotic development. The commensal strains most effective at otopathogen inhibition should then be further investigated in cellular co-culture models with clinical otopathogen strains and in an animal model of OM. Ascension models of AOM would be ideal for testing candidate probiotic species. In these models, the otopathogen is delivered to the nasopharynx rather than injected directly into the middle ear. The animal is also co-infected with an upper respiratory virus to promote dysfunction of the Eustachian tube and ascension of the otopathogen to the middle ear, mimicking the process of human AOM. This system would be ideal for testing promising candidate strains for their ability to prevent otopathogen colonisation of the nasopharynx and ascent to the middle ear. Previous clinical trials assessing the efficacy of nasal spray probiotics in children with rAOM have indicated that pre-treatment with antibiotics to allow re-establishment of the healthy microbiota with probiotic species may be the most effective method for preventing the re-emergence of otopathogens and the recurrence of AOM. Antibiotic treatment should be incorporated into the study design to determine whether this improves the colonisation of probiotic strains. Finally, selected commensal strains will need to undergo rigorous testing to confirm that they are incapable of causing disease before they can be developed into a formula for use in a clinical trial with children.

There is some debate as to whether probiotic treatment is effective at restoring the normal microbiota. A recent study by Suez *et al.* (2018) examining the effects of probiotic treatment after the administration of antibiotics in the gut of humans and
mice reported that an 11-strain probiotic treatment actually impaired the recovery of the normal microbiota, but recovery was improved by a faecal microbiota transplant (FMT) from the subjects’ gut pre-antibiotics (autologous FMT). The probiotic demonstrated the ability to rapidly colonise the gut after antibiotics, however this delayed the recovery of the initial, indigenous microbiota which returned rapidly with administration of the autologous FMT. This study suggests that the use of antibiotics does create an opportunity for the colonisation of administered bacteria, but that an autologous microbiota transplant or person-specific mixture of probiotic strains may be more effective in returning the microbiota to its original state.

However, the human gut harbours a substantially more diverse microbiota than the nasopharynx of children and the probiotic strains selected were not targeted to a specific disease. In the low diversity nasopharynx, the administration of probiotics containing several commensal clinical isolates selected specifically from the same niche in children resistant to the disease which the probiotics are aiming to treat may remain an effective option and is worth exploring.

The importance of Corynebacterium spp. and D. pigrum in nasopharyngeal health is clear, and this thesis has demonstrated that they are also specifically associated with rAOM-resistant children. In Chapter 3, it was also shown for the first time that multiple Corynebacterium species (pseudodiphtheriticum, propinquum and accolens) may contribute to the resistance to rAOM. The use of multiple species as a niche-specific, disease-specific targeted probiotic therapy administered locally and after antibiotic administration still holds great promise for the prevention of further episodes of AOM in children with rAOM. Niche-specific probiotics applied to a naturally low-diversity environment have been extensively trialled with success in the treatment of bacterial vaginosis, and are likely to be more effective than a generic probiotic therapy used to repopulate a very diverse environment. As demonstrated by the efficacy of faecal transplants, it may be necessary to administer an entire healthy microbiota to observe substantial beneficial effects; this would require a large number of perhaps person-specific probiotic strains in the high-diversity human gut, but may involve only a few species in the low-diversity nasopharynx of children. As Corynebacterium and Dolosigranulum have also been associated with health in the context of cystic fibrosis and bronchiolitis, it is likely that a nasally-delivered probiotic formulation of these organisms could be useful in treating other respiratory diseases in children. Attempting to reverse the effects of antibiotics in the treatment of childhood infections instead of solely relying on antibiotic treatment may also be important for the child’s future health, as early antibiotic use in children and subsequent alteration of the microbiota has been shown to have substantial health
implications later in life.\textsuperscript{140}

The challenges associated with undertaking ’omics work in the human upper respiratory tract

The most common next-generation sequencing method utilised in studies of the URT microbiota is metataxonomics. This thesis presents the novel application of metagenomics and metatranscriptomics to the URT of children with OM. These methods provide a great amount of additional information over metataxonomics, enabling the characterisation of the genomic functions of the microbiota, assessment of microbial activity and finer resolution in the taxonomic identification of the members of the microbial community. These advantages are essential for understanding the roles of \textit{A. otitidis} and \textit{T. otitidis} in the pathogenesis of OM, and of \textit{Corynebacterium spp.} and \textit{D. pigrum} in resistance to the development of rAOM. However, in applying these methods to MEF and NPS specimens from children with OM for the first time, this thesis has identified several challenges that must be overcome before further studies of this kind are able to provide reliable information.

The first major challenge is the nature of URT specimens. Nasopharyngeal swabs and MEF specimens are both low in biomass and high in host nucleic acids. The issue of low biomass is not something that can be altered, but must be taken into account with consideration given to the sampling and DNA extraction methods. Laboratory reagents are often contaminated with bacterial DNA, and a method has recently been published recommending the pre-treatment of PCR reagents with DNase for metataxonomic studies in low biomass specimens.\textsuperscript{425} Data from low biomass specimens must be interpreted in conjunction with additional information from negative controls to identify background contamination; several approaches for dealing with this have been described.\textsuperscript{451} The high proportion of contaminating host DNA may be reduced, however the work in this thesis demonstrated that the NEBNext Microbiome DNA Enrichment kit is not efficient at removing human DNA from nasopharyngeal and MEF specimens. There are several other approaches now available for selective human DNA removal at the point of DNA extraction, which should be evaluated in these sample types alongside qPCR to quantify the human and bacterial DNA load before and after enrichment. However, the low biomass of these samples means that two qPCR reactions in addition to metagenomic sequencing may not be possible with the limited material available. Additional samples will need to be collected for the optimisation of host removal methods, or several replicates can be analysed through shotgun sequencing to observe the impact of remaining host
DNA on metagenomics analysis. One persistent issue with attempting to remove host nucleic acids is that any additional steps with extra reagents have the potential to introduce contamination to the sample, and there is also the possibility that rarer species or particular taxa may be removed.

Another major challenge in undertaking ‘omics studies in the URT is the lack of reference genomes and annotation for key species. In this thesis, *A. otitidis*, *T. otitidis*, *C. pseudodiphtheriticum* and *D. pigrum* were important species warranting further investigation. Tied to the issue of contaminating human DNA, the small amount of microbial sequence data generated makes metagenome assembly highly challenging; forcing reliance on reference databases in which information on these organisms is largely incomplete. The detailed genomic information required to understand the functions of these organisms is simply not available. High quality assemblies of several strains of this species will advance the knowledge of these organisms substantially, whether they are derived from whole genome sequencing of pure cultures or from assembled metagenomes enabled by deeper sequencing. Without this information, utilising reference databases for metagenomics and metatranscriptomics where these organisms are poorly characterised and the volume of data is limited due to the issues described above makes the interpretation of URT functional data challenging; a task of separating signal from noise when both are uncharacterised. A combination of improvements in the use of metagenomics and metatranscriptomics in the URT and further characterisation of these particular species may be required for future studies of this nature.

The URT microbiome field is beginning to catch up with the study of microbial communities in gut, soil and marine environments. However, this thesis has characterised several remaining challenges that need to be overcome before high quality metagenomics and metatranscriptomics data can be produced. Single-cell metagenomics is an emerging development that may have the capacity to circumvent some of these issues; the technique is still very much in its infancy but appears very promising for the fine-scale characterisation of microbial communities.\(^{452,453}\) Despite the rapidly increasing number of studies utilising next-generation sequencing techniques to study the microbiome, not only ‘omics technologies should be used to understand the roles of organisms in the URT. Discussed by Hill (2018), exploratory methods such as metagenomics and metataxonomics need to be complemented by basic microbiology or molecular biology experiments to target organisms of interest, validate findings observed in microbiome studies and directly test hypotheses generated from microbiome research.\(^{454}\) This approach is particularly beneficial for the URT microbiota,
where ‘omics methods will always remain challenging to undertake due to the nature of available specimens.

Caution must also be taken in interpreting microbiome data in general due to its compositional nature. Compositional data has been notoriously challenging to interpret, where an increase in relative abundance of one organism will always result in the apparent decrease in others, which may not reflect a true reduction in the population of these other organisms and ignores the effect of total bacterial load.\textsuperscript{455,456} It has been well-documented that the methods of sample collection and storage, DNA extraction and choice of amplicon primers can influence the measured composition of the microbiome,\textsuperscript{161–163,281} obscuring the true relative abundances of organisms. It can be challenging and potentially misleading to view a picture of the DNA extracted from a specimen as a true representation of the living, functioning organisms in that habitat and their contribution to the disease or health status of the subject in question. Complementary and follow-up studies are therefore essential to accompany microbiome work, particularly in the understanding of pathogenic or beneficial microbes in rAOM. These studies should utilise information from a range of ‘omics techniques including metabolomics, proteomics and host genomics/transcriptomics as well as basic microbiology to build a more complete picture of the interactions between the microbiome and the human host.

\section*{Conclusion}

This thesis has presented a combination of metataxonomics, metagenomics and metatranscriptomics analyses aimed at understanding the role of organisms that may be involved in the pathogenesis of and resistance to rAOM. In doing so, the thesis provides a comprehensive view of the challenges associated with applying these techniques to URT specimens. This work has contributed to an increased understanding of the composition of the microbiota in the middle ear, nasopharynx and external auditory canal in children with rAOM; and also that of the nasopharynx of children resistant to the development of rAOM. This has provided further evidence to suggest that the potentially novel otopathogens \textit{Alloiococcus otitidis} and \textit{Turicella otitidis} are worth investigating for their contribution to the pathogenesis of OM, and that the health-associated commensal organisms \textit{Corynebacterium spp.} and \textit{Dolosigranulum pigrum} are involved in resistance to the development of rAOM. This work will help to shape future ‘omics-based studies in children with OM, as well as targeted research into these organisms of interest, which may lead to the development of effective alternative therapies for this recalcitrant disease.
References


167


Perez AC, Murphy TF. A Moraxella catarrhalis vaccine to protect


75. Kaur R, Adlowitz DG, Casey JR, Zeng M, Pichichero ME. Simultaneous Assay for Four Bacterial Species Including Alloiococcus otitidis Using Multiplex-PCR in


137. Leach AJ, Morris PS. Antibiotics for the prevention of acute and chronic


139. Shin JJ. Antibiotics administered for acute otitis media have modest benefits and adverse effects. BMJ Evid Based Med. 2016 Oct 1;21(5):181–1, DOI:10.1136/ebmed-2016-110398.


184

References

1;126(12):2844–51, DOI:10.1002/lary.26128.


185


205. Rautava S, Salminen S, Isolauri E. Specific probiotics in reducing the risk of acute infections in infancy a randomised, double-blind, placebo-controlled study. Br


266. Westerberg BD, Kozak FK, Thomas EE, Blondel-Hill E, Brunstein JD, Patrick DM. Is the Healthy Middle Ear a Normally Sterile Site? Otol Neurotol. 2009


018-1154-3.


344. Buisine N, Tang CM, Chalmers R. Transposon-like Correia elements: Structure,


373. Tano K, Olofsson C, Grahn-Håkansson E, Holm SE. In vitro inhibition of


390. American Type Culture Collection. ATCC Product Sheet: Corynebacterium pseudodiphtheriticum (ATCC 10701) [Internet]. Available from: https://www.atcc.org/-/ps/10701.ashx

References


441. Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Wal-}


References


Appendix A

Published manuscript and supplementary material (Chapter 2)

This appendix contains the typeset published manuscript which constitutes Chapter 2 and the supplementary material published with this article:

A.1 Published manuscript

The publication as it appears in BMC Microbiology.
A microbiome case-control study of recurrent acute otitis media identified potentially protective bacterial genera

Rachael Lappan 1,2*, Kara Imbrogno 1,3, Chisha Sikazwe 4, Denise Anderson 3, Danny Mok 3, Harvey Coates 5, Shyan Vijayasekaran 5,6, Paul Bumbak 5,6, Christopher C. Blyth 2,4,5,6, Sarra E. Jamieson 3 and Christopher S. Peacock 1,3*

Abstract

Background: Recurrent acute otitis media (rAOM, recurrent ear infection) is a common childhood disease caused by bacteria termed otopathogens, for which current treatments have limited effectiveness. Generic probiotic therapies have shown promise, but seem to lack specificity. We hypothesised that healthy children with no history of AOM carry protective commensal bacteria that could be translated into a specific probiotic therapy to break the cycle of re-infection. We characterised the nasopharyngeal microbiome of these children (controls) in comparison to children with rAOM (cases) to identify potentially protective bacteria. As some children with rAOM do not appear to carry any of the known otopathogens, we also hypothesised that characterisation of the middle ear microbiome could identify novel otopathogens, which may also guide the development of more effective therapies.

Results: Middle ear fluids, middle ear rinses and ear canal swabs from the cases and nasopharyngeal swabs from both groups underwent 16S rRNA gene sequencing. The nasopharyngeal microbiomes of cases and controls were distinct. We observed a significantly higher abundance of Corynebacterium and Dolosigranulum in the nasopharynx of controls. Alloiococcus, Staphylococcus and Turicella were abundant in the middle ear and ear canal of cases, but were uncommon in the nasopharynx of both groups. Gemella and Neisseria were characteristic of the case nasopharynx, but were not prevalent in the middle ear.

Conclusions: Corynebacterium and Dolosigranulum are characteristic of a healthy nasopharyngeal microbiome. Alloiococcus, Staphylococcus and Turicella are possible novel otopathogens, though their rarity in the nasopharynx and prevalence in the ear canal means that their role as normal aural flora cannot be ruled out. Gemella and Neisseria are unlikely to be novel otopathogens as they do not appear to colonise the middle ear in children with rAOM.

Keywords: Otitis media, Microbiome, 16S rRNA, Corynebacterium, Dolosigranulum, Alloiococcus, Turicella, Nasopharynx, Middle ear

Background

Otitis media (OM) refers to a group of inflammatory conditions of the middle ear and is commonly seen in young children. The disease can be divided into two broad categories; acute OM (AOM) and OM with effusion (OME). AOM involves signs of active infection including fever and irritability, and the middle ear contains purulent fluid with a bulging tympanic membrane. OME is characterised by non-purulent effusion and no signs of acute infection. Recurrent AOM (rAOM), defined as 3 or more episodes of AOM within 6 months; or 4 or more in 12 months [1] is also common, with 17% of children experiencing at least 3 episodes before the age of 1 year [2]. Children with rAOM are commonly prescribed repeated courses of antibiotics, and are often referred to an ear, nose and throat (ENT) surgeon for insertion of ventilation tubes (grommets) to prevent rupturing of the tympanic membrane. In Australia, 73% of children under 12 months of age will have experienced OM at least once, with costs to the health care system estimated at $100 to $400 million AUD [3].

* Correspondence: rachael.lappan@research.uwa.edu.au; christopher.peacock@uwa.edu.au

1The Marshall Centre for Infectious Diseases Research and Training, School of Biomedical Sciences, The University of Western Australia, Perth, WA, Australia

Full list of author information is available at the end of the article.

© The Author(s). 2018 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Bacterial pathogens that cause AOM are referred to as otopathogens. The three bacterial species widely recognised as otopathogens are Streptococcus pneumoniae, non-typeable Haemophilus influenzae and Moraxella catarrhalis, which are thought to originate from the nasopharynx and are capable of migrating to the middle ear and persisting within biofilms [4]. Children with AOM are commonly colonised with multiple otopathogens [5], and coinfection with respiratory viruses is common [6]. However, a proportion of children with AOM do not appear to be colonised with any of the known otopathogens [5, 7, 8], implying that there may be other bacteria involved in AOM. Alloiococcus otitidis and Turicella otitidis have been associated with OM, but their role in the pathogenesis of AOM remains unknown [9, 10].

Antibiotic treatment has been shown to be of limited benefit for AOM [11], and all three otopathogens have exhibited resistance to the antibiotics commonly used to treat it [12]. Additionally, one in five children fitted with grommets will require reinsertion in the future [13]. Considering the limited effectiveness of current treatments, there is a need for alternative therapies for rAOM. Probiotic treatment for rAOM is one alternative that has been investigated in several clinical trials. Probiotics are defined as "live microorganisms, which when administered in adequate amounts, confer a health benefit on the host" [14] and act via mechanisms such as competition for nutrients, stimulation of the immune system and direct inhibition with antibacterial molecules [15]. The effect of probiotics on the recurrence of AOM has been variable, with some studies showing a significant improvement [16–21] and others reporting no change [22–27]. The contradictory results of these studies highlight the importance for the development of probiotics containing bacterial strains that are relevant to the upper respiratory tract environment, active against otopathogens and able to colonise the nasopharynx.

A relatively new area of microbiome research involves the identification of commensal bacteria or bacterial products from a ‘healthy’ microbiome for therapeutic use. Therapies identified by this approach have been successful in resolving relapsing Clostridium difficile disease in mice [28], resolving Salmonella enterica serovar Typhimurium disease in pigs [29] and inhibiting colonisation and biofilm formation of Staphylococcus aureus in the nasopharynx of adult humans [30]. These studies demonstrate that the microbiome can be a source for effective probiotic treatments, which may be a single species that acts specifically against the pathogen of interest, or a combination of multiple commensal species.

We had two hypotheses in this study. Firstly, we hypothesised that there are bacterial pathogens involved in rAOM other than the known otopathogens. We sought to identify potential novel otopathogens by characterising for the first time the microbiome of the middle ear in children with rAOM and comparing it to their nasopharyngeal microbiome. Secondly, under the indication that exposure to other children via attendance at day care or the presence of multiple children in the home is a major risk factor for rAOM [31, 32], we hypothesised that children exposed to this risk factor but who have not developed rAOM are carrying nasopharyngeal bacteria that provide protection against the disease. We aimed to identify potentially protective commensal bacteria, which may be of use as a specific probiotic therapy for rAOM, by comparing the nasopharyngeal microbiomes of these rAOM-resistant children with those of children with rAOM.

**Methods**

**Patient recruitment**

Children under the age of 5 years were recruited into either the case (rAOM-prone) or control (rAOM-resistant) group of the Perth Otitis Media Microbiome (biOMe) study in the Perth metropolitan area of Western Australia from December 2013 to December 2015. Cases were undergoing grommet insertion for a physician-diagnosed rAOM and were identified as eligible for inclusion by their ENT surgeon. Children undergoing grommet insertion for OME were excluded from recruitment. Middle ear fluid (MEF) from each ear and a nasopharyngeal swab (NPS) were collected at the time of surgery. A saline middle ear rinse (MER) from each ear and one ear canal swab (ECS) were also collected from a subset of patients (56/93).

Healthy controls with no history of rAOM were recruited from a community immunisation clinic. Controls were attending day care or had a sibling up to 5 years of age at the time of collection (i.e. were exposed to previously described major risk factors for rAOM [31, 32]). A NPS sample was collected from controls. A questionnaire on demographics, risk factors and recent antibiotic use was completed for all case and control subjects (see Additional file 1). Controls were matched to cases by age (within 3 months if case less than 1 year of age; within 6 months if case between 1 and 2 years of age; within 12 months if case between 2 and 5 years of age) and season (within 2 weeks of collection time; to the previous year if no match found in recruitment year). Subjects were matched by sex where possible. Exclusion criteria for both groups included diagnosis of cleft lip or palate, immune deficiency or genetic syndrome. All specimens and questionnaire data were obtained with informed written consent from a parent or guardian. Recruitment to the study was approved by the Human Research Ethics Committees (HREC) at Princess Margaret Hospital for Children (2013119/EP), St John of God Health Care (#708) and the University of Western Australia (RA/4/1/6839) as well as by all relevant hospital governance committees.
Sample collection and storage
All specimens obtained from the cases were collected at the time of grommet surgery by the performing surgeon. ECS specimens were obtained from one ear prior to myringotomy (incision of the tympanic membrane) with a sterile FLOQswab (Copan) and were placed in 1 ml skim milk tryptone glucose glycerol broth (STGGB, PathWest). Following myringotomy, MEF specimens from each ear were aspirated into a sterile Argyle™ specimen trap (Covidien) with 2 ml of sterile saline used to flush out the tubing. MER specimens were collected from each ear after aspiration of MEF whereby 2 ml of sterile saline was injected into the middle ear, then aspirated into an Argyle™ trap. A NPS was taken with a sterile FLOQswab, rotating for at least 3 s at the nasopharynx before transferring into 1 ml of STGGB. Specimens from the cases were immediately frozen on dry ice or kept on wet ice and transported to the laboratory on the same day. NPS specimens were collected from controls in the same manner and kept on wet ice until transport to the lab. All specimens were frozen at ~80 °C until DNA was extracted.

DNA extraction and sequencing preparation
Swab samples (NPS and ECS) were first prepared by vortexing followed by blending of the swab, inverted, to a new sterile tube with sterile forceps. This was centrifuged to collect mucus attached to the swab which was then transferred back into the milk broth. All samples were then aliquoted for DNA extraction (500 μl for ECS and NPS, 750 μl for MEF and MER) and each MEF and NPS specimen was also aliquoted for viral typing (200 μl). The remainder of all specimens were archived at −80 °C. DNA was extracted with the Wizard SV Genomic DNA Purification System (Promega) and FastPrep Lysing Matrix B tubes (MP Biomedicals) as described in Teo et al. [33] with some modifications. In brief, extractions were carried out inside a class II biohazard hood with UV-sterilised plastics and pipettes wiped with DNA Away (Molecular BioProducts) to minimise contamination. A negative extraction control (reagents with no specimen) was included in each extraction batch and each batch included samples of each type. DNA extraction aliquots were then processed as previously described [33], with the final purified genomic DNA stored in DNA Lo-Bind tubes (Eppendorf). Samples were quantified by the Qubit 2.0 fluorometer (dsDNA HS assay, Invitrogen) and diluted to 5 ng/μl with low TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, Fisher Biotech). All genomic DNA was frozen at ~80 °C until sequencing preparation.

Viral detection
DNA from the viral typing aliquots for all MEF and NPS specimens (cases and controls) was extracted using an automatic extraction platform (MagMAX express 96) according to manufacturer’s instructions. These samples were screened for 19 common respiratory viruses (see Table 1) using a routine multiplex PCR at PathWest (Perth, Western Australia). This method was developed and described by Chidlow et al. [34].

Positive sequencing control generation
A mock community of 16 bacterial species (MOCK) was used as a positive sequencing control (see Table 2). This included a mixture of Gram positive and Gram negative organisms, including some with known prevalence in the upper respiratory tract and some not expected to be found in this environment. Genomic DNA was extracted from glycerol stocks of each species using the Wizard SV Genomic DNA Purification System, with the exception of N. meningitidis which was obtained as a heat-killed stock. Each species separately underwent PCR amplification and AMPure XP purification (described in the following section). The PCR products were quantified by Qubit fluorometry, diluted to equal concentrations and pooled at equal volumes to create the positive sequencing control that was included on each sequencing run. The theoretical expected relative abundance is 6.25% for each species. The positive sequencing control was aliquoted before freezing at ~80 °C to minimise freeze/thaw effects on the DNA. Each sequencing run utilised a separate aliquot, in duplicate.

Amplicon sequencing
Samples were prepared for amplicon sequencing following the Illumina protocol for 16S rRNA gene sequencing (Part # 15044223, Rev. B) with modification to ensure sufficient amplification from samples that yielded low amounts of DNA. The recommended primers (forward: 5’ CCTACGGGNGGCWGCAG, reverse: 5’ GACTACHVGGGTATCTAATCC) target the V3/V4 region of the 16S rRNA gene [39] with Illumina adapters attached.

Table 1: Gene targets for multiplex respiratory virus PCR. Where multiple strains of a virus were detected, results were combined.

<table>
<thead>
<tr>
<th>Virus name (strains targeted)</th>
<th>Abbreviation</th>
<th>PCR target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human adenovirus</td>
<td>HaDNA</td>
<td>Hexon gene</td>
</tr>
<tr>
<td>Human bocavirus</td>
<td>HBoV</td>
<td>VP1 gene</td>
</tr>
<tr>
<td>Influenza virus (A/B/C)</td>
<td>IFV</td>
<td>Haemagglutinin &amp; Matrix gene</td>
</tr>
<tr>
<td>Respiratory syncytial virus (Type A/Type B)</td>
<td>RSV</td>
<td>Nucleoprotein</td>
</tr>
<tr>
<td>Human metapneumovirus</td>
<td>HMPV</td>
<td>Nucleoprotein</td>
</tr>
<tr>
<td>Human coronavirus (OC43/229E/HKU1/NL63)</td>
<td>HCoV</td>
<td>Nucleocapsid (OC43, 229E, NL63)</td>
</tr>
<tr>
<td>Human coronavirus (DC43/229E/HKU1/NL63)</td>
<td>HCoV</td>
<td>ORF1a/b (HKU1)</td>
</tr>
<tr>
<td>Parainfluenza virus (1/2/3/4)</td>
<td>HPV</td>
<td>Nucleoprotein</td>
</tr>
<tr>
<td>Rhinovirus (A/B/C)*</td>
<td>RV</td>
<td>3'UTR</td>
</tr>
</tbody>
</table>

* Rhinovirus typing used primer pairs reported by Lee et al. [33]
to the 5' end. The expected length of this targeted region is approximately 465 bp including the amplicon primers. The PCR reaction mix contained 9.5 μl of genomic DNA with a final concentration of 300 nM for each amplicon primer and 1× KAPA HiFi HotStart ReadyMix (KAPA Biosystems). A negative extraction control and a no-template negative PCR control were included on each PCR plate. PCR cycling conditions were 95 °C for 3 min, 30 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s before a final 72 °C for 5 min and holding at 4 °C. All PCR products were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter) as described in the Illumina protocol. The samples and mock community aliquots were then barcoded by Illumina's dual indexing strategy (Nextera XT Index Kit v2, Sets A and B, Illumina) as described in the Illumina protocol using the default barcode layout from the Illumina Experiment Manager software v1.11.0. Samples were sent to the Australian Genome Research Facility where they underwent further purification and quality checking followed by dilution and equimolar pooling of all samples at 1 ng/μl. The final pool underwent a band excision with the QIAquick Gel Extraction kit (Qiagen) to select the V3/V4 band (approximately 600 bp with adapters and indexes) for sequencing. The pool was sequenced on the Illumina MiSeq with 2×300 bp V3 chemistry and a spike-in of 20% PhiX control.

A total of 581 samples including all positive and negative controls were sequenced across four separate sequencing runs of 140–149 samples each. Samples of all types were included on each sequencing run.

Preprocessing of sequence data

A diagrammatic overview of the entire analysis pipeline is provided in Additional file 2. Demultiplexed paired-end reads in FASTQ format were received from the sequencing centre. Overall run quality was observed with FastQC v0.11.3 [40]. The sequence data was processed via the UПARSE pipeline [41] (using the USEARCH v8.1.1861 algorithm [42]) and QIIME v1.9.1 [43]. Paired-end reads were merged with USEARCH with maximum expected errors set to 1 and length restricted to 440–470 bp. Amplicon primer sequences were removed from these high quality reads using a custom script [44] before conversion to FASTA format with QIIME, concatenation of the data from all four sequencing runs and dereplication (collapsing into unique sequences) with USEARCH. Reads that aligned to the human genome (GRCm38_p0) were removed with Deconseq v0.4.3 [45] using identity threshold 94% and coverage threshold 90%. Remaining reads were clustered into OTUs (operational taxonomic units) at a 97% identity threshold with USEARCH to create a high quality chimera-filtered representative set of OTU sequences. The original raw paired-end reads were then merged as before but without expected error filtering. This set of reads was then aligned to the high quality OTU representative set with USEARCH to assign an OTU to each sequence (97% identity threshold) and create an OTU table. From this point, the sequences were processed within QIIME. Taxonomy was assigned with UCLUST v1.2.22q [42] at 90% identity with the SILVA database v123 [37, 38] and then aligned to the SILVA core alignment with PyNAST v1.2.2 [46]. We selected the SILVA database over the QIIME default database (GreenGenes release 13.8 [47]) as Dolosigranulum is not present in the 13.8 release, and is misclassified as Alloiococcus. The alignment was filtered and a phylogenetic tree generated with FastTree v2.1.3 [48]. OTUs that failed to align and those with an abundance below 0.005% were removed from the OTU table. Samples with low sequencing depth (77 out of 581 total samples; threshold 1499 reads) were also removed from the table. This OTU table was used for all down-stream analysis.

Common practice is to then rarefy this data, which involves subsampling to an even number of sequences per sample to render them comparable. However, this can discard a large volume of usable data and reduces statistical power [49]. We have avoided the statistically inadmissible [49] practice of rarefying throughout this study.

### Table 2

Species included in the positive sequencing control. All cultures were obtained from the University of Western Australia’s School of Biomedical Sciences culture collection, with the exception of N. meningitidis which was kindly provided by A/Prof Charlene Kahler (UWA) and originally described in Stephens et al. [36]

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain/ATCC/NCTC no.</th>
<th>Gram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staphylococcus aureus</td>
<td>ATCC 29213</td>
<td>Positive</td>
</tr>
<tr>
<td>Staphylococcus epidermidis</td>
<td>ATCC 14990</td>
<td>Positive</td>
</tr>
<tr>
<td>Streptococcus pneumoniae</td>
<td>Strain D39</td>
<td>Positive</td>
</tr>
<tr>
<td>Staphylococcus warneri</td>
<td>ATCC 27836</td>
<td>Positive</td>
</tr>
<tr>
<td>Moraxella catarrhalis</td>
<td>ATCC 25238</td>
<td>Negative</td>
</tr>
<tr>
<td>Non-typeable Haemophilus influenzae</td>
<td>Strain 86-028NP</td>
<td>Negative</td>
</tr>
<tr>
<td>Haemophilus haemolyticus</td>
<td>ATCC 33390</td>
<td>Negative</td>
</tr>
<tr>
<td>Neisseria meningitidis</td>
<td>Strain M7</td>
<td>Negative</td>
</tr>
<tr>
<td>Corynebacterium jeikeium</td>
<td>ATCC 43216</td>
<td>Positive</td>
</tr>
<tr>
<td>Propionibacterium acnes</td>
<td>ATCC 6919</td>
<td>Positive</td>
</tr>
<tr>
<td>Gemella haemolysans</td>
<td>NCTC 10244</td>
<td>Positive</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>NCTC 8172</td>
<td>Negative</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>ATCC 15692</td>
<td>Negative</td>
</tr>
<tr>
<td>Streptococcus salivarius</td>
<td>not available</td>
<td>Positive</td>
</tr>
<tr>
<td>Veillonella parvula</td>
<td>ATCC 10790</td>
<td>Negative</td>
</tr>
<tr>
<td>Alloiococcus esteticus</td>
<td>ATCC 51267</td>
<td>Positive</td>
</tr>
</tbody>
</table>

*This species may have been mislabelled as it was identified as Globicatella by SILVA v123 taxonomy [37, 38]*
Covariate and diversity analyses

Data from completed demographic questionnaires included day care attendance, duration of breastfeeding, presence of siblings and smokers in the household, relevant clinical diagnoses (i.e. asthma, allergy and chest, heart or kidney problems), previous admission to hospital for infection and use of antibiotics in the previous month. Statistical analysis of covariates (including detection of respiratory viruses) was done in R v3.3.2 [50] using the Wilcoxon rank sum test for continuous data and Pearson’s Chi-squared test for categorical data (unless any values were below 5, then Fisher’s exact test was used). Odds ratios and associated confidence intervals and p-values for the viral detection data were calculated in R with the exact2x2 function in the exact2x2 package [51]. All subjects who contributed at least one analysable sample were included in the analysis of covariates.

Summaries of taxonomic relative abundance were generated with QIIME. Correlation of taxa summaries in QIIME was done in either expected mode (all samples against one theoretical expected mock community) or paired mode (pairs of samples) using the Pearson correlation coefficient with a two-sided permutation test (999 permutations) to calculate a non-parametric p-value.

Alpha (within-sample) diversity was measured using Faith’s Phylogenetic Diversity (PD) [52] and inverse Simpson (IS) [53] indices, calculated per sample in QIIME on a cumulative-sum-scale (CSS) normalised and logged OTU table (see differential abundance method for CSS normalisation). PD is a phylogenetic measure that reflects how much of the phylogenetic tree is covered by the OTUs found in the samples. The IS index is a non-phylogenetic measure that takes into account the richness (number of OTUs) and evenness (relative abundance of OTUs) in the sample, with the reciprocal taken so that larger values represent greater diversity. Median values and statistical tests between groups of samples were calculated in R using the Wilcoxon rank sum test for unpaired samples, and Wilcoxon signed rank for paired samples. Alpha diversity boxplots were generated in R with ggplot2 v2.2.0 [54] and gridExtra v2.2.1 [55].

Beta diversity (between-sample diversity) was calculated using the weighted UniFrac metric [56] in QIIME on the raw OTU table. McMurdie and Holmes [49] demonstrated that weighted UniFrac distances are accurate on raw (unrarefied) data. Two-dimensional principal coordinates analysis (PCoA) plots were also generated in QIIME. Procrustes analysis was used to determine the similarity of the PCoA between pairs of samples collected or treated in two different ways. Comparisons were made between raw and rarefied (subsampled) data, MEF and MER sample types from the same ear, and left and right ears from the same child. The analysis calculates an m² value which describes how similar the paired datasets are, with a smaller number indicating the datasets are more similar.

The associated non-parametric p-value describes the chance of seeing an m² value at least this extreme in 999 permutations using Monte Carlo simulations. A p-value of 0 means that no value as extreme as the calculated m² was observed in these permutations. This was carried out within QIIME and plotted with Emperor [57].

Co-occurrence analyses

To identify positive and negative correlations between individual OTUs in each sample type, we used SparCC [58], developed specifically for quantifying correlations in compositional microbiome data. Correlations between OTUs were determined separately for each sample type (case NPS, control NPS, MEF, MER) with only one sample per child chosen at random if more than one was available. OTUs that were represented by less than 2 reads per sample on average were removed from the raw OTU table before calculating correlations [58]. The default settings were used and one-sided pseudo p-values were calculated (taking into account the direction of correlation) using 100 simulated datasets. For the MEF and MER, where one ear per child was selected at random, the analysis was repeated on a second set of samples with the opposite ear. Correlations with a significant (≤ 0.05) p-value in both sets of samples were considered true positives. Values from the first set are reported for correlations that were validated in this way.

Differential abundance analyses

Differential abundance of OTUs was assessed using the fitZIG function in the R package metagenomeSeq v1.18.0 [59] in R v3.4.1. fitZIG fits a zero-inflated Gaussian mixture model to test for differentially abundant OTUs between groups, and is also designed for use with microbiome data. Five models were fitted to compare the abundance of OTUs between sample types (case/control NPS, MEF/MER, MEF/NPS, MER/NPS and ECS/MEF). In all models, OTUs that were not present in at least 25% of the samples in that model were filtered from the analysis to reduce false positives. The data were normalised using metagenomeSeq’s CSS normalisation [60]. The models were fitted as follows:

**Model 1 (case NPS/control NPS)**

The model included sex (Male/Female), recent antibiotic usage (within the past month; Yes/No) and length of breastfeeding (Never/Under 6 months/6 to 12 months/Over 12 months/Current) to control for these potential confounders. All nasopharyngeal swabs from both groups were included in the analysis except those with missing data for the above covariates (n = 4), leaving 98 control NPS and 86 case NPS for comparison.


Models 2–5 (MEF/MER, MEF/NPS, MER/NPS, ECS/MEF)
These models included only samples from the cases and were analysed as within-child pairs. Where multiple samples were available (both left and right ear), one was selected at random (sample Set 1). A second set (sample Set 2) containing the sample from the opposite ear was used for validation (i.e. differentially abundant OTUs with an adjusted \( p \)-value ≤0.05 in only one of the two sets were considered false positives). For the MEF/MER and ECS/MEF comparisons, the pair included samples from the same ear of each child. Subject ID was a covariate in each of the four models. Included in the analysis were 50 pairs of MEF/MER samples, 75 pairs of MEF/NPS samples, 54 pairs of MER/NPS samples and 33 pairs of ECS/MEF samples.

OTUs whose abundance was significantly different between the two groups compared (FDR-adjusted \( p \)-value ≤0.05) were retained if their median or mean relative abundance was at least 0.35% in one of the two groups compared. We chose this threshold as it discarded known environmental contaminants abundant in the negative controls (e.g. *Delftia, Lysinibacillus*) but retained low-abundance OTUs not found in the negative controls that are known upper respiratory tract colonisers (e.g. *Veillonella*). Heatmaps representing log CSS normalised OTU counts in the compared groups were created with superheat v0.1.0 [61]. Model coefficients (log; fold-changes; logFC) and \( p \)-values were derived from the MRcoefs function in metagenomeSeq.

Results
Study population characteristics
In our cohort, we assessed several environmental and clinical variables that may influence the risk of OM (see Table 3). The most significant difference observed between cases and controls was recent antibiotic usage, which was significantly more common in the cases. The length of breastfeeding and the presence of any other chronic illness were also significantly different between the two groups; the cases were breastfed for a shorter median time (5 months compared to 10.3 months) and had a higher incidence of chronic illness than the controls. Controls were not significantly different to cases in terms of age, season of collection or sex confirming successful matching of controls to cases at recruitment. We did note a higher proportion of males than females with rAOM in the case group.

Bacterial taxa identified across the samples
A total of 31.7 million raw reads were generated across the four sequencing runs. After read pre-processing and OTU picking, the reads were clustered into a total of 123 OTUs. The full taxonomy of these OTUs is provided in Additional file 3, and the results from the positive and negative sequencing controls can be found in Additional file 4. Table 4 shows the aggregated relative abundance of genus-level taxa from all samples excluding those with a low read count (less than 1499 reads) and sequencing controls. *Moraxella, Haemophilus* and *Streptococcus* were abundant in the nasopharynx of the cases and healthy controls, but the control samples contained *Dolosigranulum* and *Corynebacterium* at higher abundance than the cases. The cases additionally contained *Neisseria* (10.9%), *Gemella* (3.2%), *Porphyromonas* (3.6%), *Alloprevotella* (2.4%) and *Fusobacterium* (2.7%) where these were almost absent in the controls (each below 0.3%). Of the three major otopathogen genera, *Haemophilus* was the most abundant in the middle ear, contributing to 18.5% of the total reads in middle ear fluids and 3.2% of the rinses. *Streptococcus* and *Moraxella* were less common, and were similarly lower in the MER than the MEF. While *Haemophilus* is prevalent in the fluids, overall the ear samples were dominated by *Alloprevotella* and *Staphylococcus*, with *Turicella* also abundant. These genera were not abundant in the nasopharynx of cases or controls. The ear canal did not contain any taxon that was not also seen at similar or higher aggregated relative abundance in the rinses or fluids.

Species-level identification could not be achieved with the V3/V4 region. Some taxa contain multiple OTUs (which may not be the same species). It is also possible that multiple species have been classified as the same OTU if the sequences are more than 97% identical.

Diversity within the nasopharynx and middle ear
We measured alpha (within-sample) diversity with the Faith’s phylogenetic diversity (PD) and inverse Simpson (IS, alternatively reciprocal Simpson) metrics. Alpha diversity by sample type is shown in Fig. 1. The nasopharynx of children with rAOM was significantly more diverse than the nasopharynx of the healthy controls. Within the same ear of the same child, the middle ear rinse was also significantly more diverse than the fluid but this difference was not as pronounced. The nasopharynx was also more diverse than the middle ear fluid when comparing within the same child.

Comparing the microbiome of the nasopharynx between rAOM-prone and rAOM-resistant children
Beta diversity represents the differences between samples; how similar or dissimilar they are to each other. We calculated beta diversity with the weighted UniFrac metric [56] on the raw read counts to determine whether there was a distinct microbiome related to case/control status, sample type or other covariates. Calculating beta diversity on raw counts is accurate [49] and was very similar to using rarefied counts (Procrustes analysis, see Additional file 5).
We compared the microbiome of the nasopharynx of cases to that of controls to identify genera that may be associated with rAOM (potential novel pathogens), or with apparent resistance to rAOM (potential candidates for probiotic therapy). Figure 2 shows a principal coordinates analysis (PCoA) plot of the UniFrac distances on the nasopharyngeal samples from cases and controls, where each sample is an individual point. The closer two points are, the more similar the microbiome of those samples. The nasopharyngeal microbiome of the cases was distinct from that of the controls, separating along the PC1 axis; indicating that much of the variability between samples is explained by case/control status. No grouping of samples was observed with any other covariate (including age, sex, etc.).

Table 3 Demographic characteristics of children recruited to the study. This table includes all children who contributed at least one sample to any analysis. P-values were calculated by Wilcoxon rank-sum test for continuous data, and Pearson’s $\chi^2$ test for categorical data (unless any values were less than 5, then Fisher’s exact test was used).

<table>
<thead>
<tr>
<th></th>
<th>Case ($n = 93$)</th>
<th>Control ($n = 103$)</th>
<th>Missing data</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (IQR)</td>
<td>1.9 years (1.3–2.8 years)</td>
<td>1.6 years (1.5–3.2 years)</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>Sex</td>
<td>58 male (62.4%)</td>
<td>53 male (51.5%)</td>
<td>0</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>35 female (37.6%)</td>
<td>50 female (48.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aboriginal or Torres strait islander</td>
<td>2 (2.1%)</td>
<td>1 (0.97%)</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Season</td>
<td>Summer: 7 (7.5%)</td>
<td>Summer: 4 (3.9%)</td>
<td>0</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Autumn: 20 (21.5%)</td>
<td>Autumn: 16 (15.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Winter: 45 (48.4%)</td>
<td>Winter: 53 (51.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spring: 21 (22.6%)</td>
<td>Spring: 30 (29.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breastfeeding (yes/no)</td>
<td>83 (89.2%)</td>
<td>98 (95.1%)</td>
<td>1 case (1.1%)</td>
<td>0.15</td>
</tr>
<tr>
<td>Duration of breastfeeding (IQR)</td>
<td>5 months (1.6–10 months)</td>
<td>10 months (6–13 months)</td>
<td>1 control (1.0%)</td>
<td>6.4 × 10^-5</td>
</tr>
<tr>
<td>Currently breastfeeding</td>
<td>7 (7.5%)</td>
<td>14 (13.6%)</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>over 12 months</td>
<td>10 (10.8%)</td>
<td>23 (22.3%)</td>
<td>0.00077</td>
<td></td>
</tr>
<tr>
<td>6–12 months</td>
<td>31 (33.3%)</td>
<td>45 (43.7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 6 months (excluding never)</td>
<td>35 (37.6%)</td>
<td>16 (15.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>never</td>
<td>9 (9.7%)</td>
<td>4 (3.9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day care or school attendance</td>
<td>74 (79.6%)</td>
<td>80 (77.7%)</td>
<td>1 control (1.0%)</td>
<td>–</td>
</tr>
<tr>
<td>Days of day care or school per week</td>
<td>Median 2.5 days (IQR 2–3 days)</td>
<td>Median 2 days (IQR 1.75–3 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no day care or school</td>
<td>19 (20.4%)</td>
<td>22 (21.4%)</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>2 days/week</td>
<td>13 (14.0%)</td>
<td>20 (19.4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 3 days/week</td>
<td>52 (55.9%)</td>
<td>46 (44.7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at day care/school start</td>
<td>13.5 months (IQR 10 months – 1.9 years)</td>
<td>12 months (IQR 10 months – 1.5 years)</td>
<td>34 controls (33%)</td>
<td>0.94</td>
</tr>
<tr>
<td>Siblings of 5 years of age or younger</td>
<td>46 (49.5%)</td>
<td>70 (68.0%)</td>
<td>1 case (1.1%)</td>
<td>–</td>
</tr>
<tr>
<td>Smoker in the household</td>
<td>16 (16.1%)</td>
<td>8 (7.7%)</td>
<td>0.069</td>
<td></td>
</tr>
<tr>
<td>Antibiotic usage in the past month</td>
<td>61 (65.6%)</td>
<td>11 (10.7%)</td>
<td>1 case (1.1%)</td>
<td>1.9 × 10^-15</td>
</tr>
<tr>
<td>Any chronic illness</td>
<td>18 (19.4%)</td>
<td>10 (9.7%)</td>
<td>14 cases (15.1%)</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 controls (6.8%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever admitted to hospital for infection</td>
<td>17 (18.3%)</td>
<td>11 (10.7%)</td>
<td>0</td>
<td>0.13</td>
</tr>
</tbody>
</table>

* Season was categorised by months
b Median values exclude children who are currently breastfeeding as breastfeeding has not ceased
c The age at which the child started day care or school; whichever they are currently attending
d Any chronic respiratory, cardiovascular or renal illness including asthma and allergies; or other chronic illness identified by the parents
presence of chronic illness, attendance at day care or school, sex, previous hospital admission for infection, season of collection and presence of siblings), with the exception of those previously shown to differ significantly between cases and controls. The pattern of recent antibiotic usage overlapped with case/control status, which is expected, as most of the cases had recently taken antibiotics whilst usage in the controls was low. Current breastfeeding and breastfeeding for at least 12 months were also concentrated towards the controls along PC1, who had a higher median length of breastfeeding (see Table 3). There were no batch effects by sequencing run (see Additional file 6).

We then determined which OTUs were causing this difference between the two groups by fitting a model using the fitZIG function in metagenomeSeq [59]. The model controlled for recent antibiotic usage, sex and length of breastfeeding. Of the 33 significantly differentially abundant OTUs, 16 were above a threshold of $\geq 0.35\%$ mean or median relative abundance in either group (see Methods) and are shown in Fig. 3. Additional file 7 contains a full list of significant OTUs with adjusted $p$-values and log2-fold-changes between groups for all fitZIG models. Of these 16 OTUs, 14 were more abundant in the nasopharynx of the cases. Haemophilus (OTU6) was one of these with a logFC of 2.7 ($p = 0.004$), though it was commonly seen across both groups. The other otopathogen genera Moraxella (OTU2) and Streptococcus (OTU4) were not significantly different. The remaining 13 OTUs were found at low abundance in the cases, but were very low or almost absent in the controls (logFC = 0.96 to 4.4, $p$-values < 0.01). Those with the highest median log CSS counts

<table>
<thead>
<tr>
<th>Genus level taxonomy</th>
<th>Control NPS (%)</th>
<th>Case NPS (%)</th>
<th>MEF (%)</th>
<th>MER (%)</th>
<th>ECS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corynebacterium</td>
<td>13.30</td>
<td>1.63</td>
<td>1.35</td>
<td>1.34</td>
<td>2.73</td>
</tr>
<tr>
<td>Turicella</td>
<td>0.03</td>
<td>0.03</td>
<td>6.72</td>
<td>11.72</td>
<td>13.06</td>
</tr>
<tr>
<td>Staphylococcus</td>
<td>1.42</td>
<td>0.81</td>
<td>9.04</td>
<td>22.23</td>
<td>27.01</td>
</tr>
<tr>
<td>Alloiococcus</td>
<td>0.17</td>
<td>0.19</td>
<td>49.84</td>
<td>57.04</td>
<td>53.62</td>
</tr>
<tr>
<td>Dolosigranulum</td>
<td>16.34</td>
<td>1.86</td>
<td>0.05</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>7.05</td>
<td>15.29</td>
<td>3.52</td>
<td>1.20</td>
<td>1.44</td>
</tr>
<tr>
<td>Neisseria</td>
<td>0.27</td>
<td>10.95</td>
<td>0.19</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>Haemophilus</td>
<td>12.43</td>
<td>18.96</td>
<td>18.52</td>
<td>3.18</td>
<td>1.44</td>
</tr>
<tr>
<td>Moraxella</td>
<td>47.55</td>
<td>30.80</td>
<td>0.21</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Pseudomonas</td>
<td>0.02</td>
<td>0.01</td>
<td>6.34</td>
<td>1.34</td>
<td>0.12</td>
</tr>
<tr>
<td>Other (49 other taxa)</td>
<td>1.42</td>
<td>19.48</td>
<td>1.35</td>
<td>1.59</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table 4 Genus-level community composition by sample type. Relative abundance was calculated for aggregated counts across all samples of each sample type and is summarised at genus level (i.e. aggregates all OTUs with the same genus assignment). Samples below a total read count of 1499 are not included. Genera with an average relative abundance below 1% are grouped as "Other".

**Fig. 1** Alpha diversity measured by a) Faith’s Phylogenetic Diversity and b) Inverse Simpson metrics grouped by sample type. Alpha diversity values were calculated on CSS normalized and logged read counts. The $p$-values represent the difference between groups determined by Wilcoxon rank sum (case/control NPS) or Wilcoxon signed rank (MEF/MER and MEF/NPS) test where paired samples were from the same child.
were *Gemella* (OTU12) and *Neisseria* (OTU13). The representative sequences for these OTUs did not match any species definitively with BLASTN 2.6.1 [62] against 16S rRNA sequences; OTU12 matched *G. taiwanensis* and *G. haemolysans* at 100%, with OTU13 hitting several *Neisseria* species at 98–99% identity. However, the SILVA database classified the sequence as *N. lactamica*.

Three OTUs were more abundant in the nasopharynx of controls compared to the cases. *Moraxella* (OTU10, distinct from OTU2 which was abundant across all nasopharyngeal samples and was not significantly different here) was found at low relative abundance in both groups, but was prevalent in several controls (logFC = 2.0, *p* = 0.05). The representative OTU sequence for this *Moraxella* OTU matches *M. lincolnii* using BLASTN [62]. *Dolosigranulum* (logFC = 3.0, *p* = 0.002) and *Corynebacterium* (logFC = 4.1, *p* = 8.9 × 10⁻⁵) were both low in the cases but significantly more abundant in the controls. *Dolosigranulum* only has one species (*D. pigrum*), but this *Corynebacterium* OTU (OTU8) could not be identified at species level with BLASTN [62].

**Comparing the microbiome of different sample types within children with rAOM**

We compared the microbiome of the nasopharynx, middle ear and ear canal in children with rAOM to identify novel bacteria that may be involved in the disease. The NPS are distinct from the ear samples in a weighted UniFrac PCoA plot (Fig. 4). The MEF, MER and ECS samples do not form distinct groups and are more sparsely distributed than the NPS. The ear samples (MEF, MER and ECS) were not observed to group by any other covariates (including antibiotic use, age, duration of breastfeeding, presence of other chronic illness, current attendance at day care or school, sex, previous hospital admission for infection, season of collection and presence of siblings), and there were no batch effects by sequencing run (see Additional file 8); suggesting that the sparse layout of these samples in Figure 4 is most likely due to large differences between individual children.

Pairs of left and right ear samples from the same child were strongly correlated. When comparing pairs of taxa summaries, the fluids had a Pearson correlation coefficient of 0.833 (95% CI 0.821–0.844, non-parametric *p* = 0.001, 999 permutations) and the rinses, 0.858 (95% CI 0.849–0.867, non-parametric *p* = 0.001, 999 permutations). They are also moderately similar following Procrustes analysis (see Additional file 9). A lower *m*² value indicates higher similarity between datasets. To ensure robust results, downstream analyses that required independent samples (i.e. only one ear per child where two were available) were run with the left or right ear randomly chosen (Set 1), and then repeated with the sample from the opposite ear (Set 2) for agreement.

**Middle ear fluid and middle ear rinse**

Sampling of the MEF in a subset (60%) of the cases was followed by a saline MER to attempt to capture bacteria that may not otherwise be detected in the fluid (for example, bacteria present in biofilm adhered to the mucosa). The taxonomic composition of the two sample types were strongly correlated when comparing samples from the same ear (Pearson correlation coefficient 0.835, 95% CI 0.826–0.843, non-parametric *p* = 0.001, 999 permutations) and they occupy a similar area on the PCoA plot (Figure 4). Procrustes analysis on the first three principal coordinates between these pairs indicated that the sample types are only moderately similar (*m*² = 0.335, *p* < 0.001, 999 permutations; see Additional file 10).

To determine which OTUs were differentially abundant between the MEF and MER, we fitted a fitZIG model on MEF/MER pairs from the same ear with the Set 1/Set 2 approach (see Methods). The 7 significantly differentially
abundant OTUs above threshold are shown in Fig. 5a (two additional OTUs above threshold were dropped for disagreement between Set 1 and Set 2). Staphylococcus (OTUs 3, 212, 269 and 1003), Alloiococcus (OTU1) and Turicella (OTU7) were found at higher abundance in the MER than in the MEF (logFC = 0.9 to 2.1, p-values < 0.0007). Haemophilus (OTU6) was more abundant in the MEF than in the MER (logFC = 1.2, p = 2.5 × 10^{-9}). While three Staphylococcus OTUs (OTU212, OTU269 and OTU1003) were found at very low abundance, all other OTUs were found at moderate abundance in both sites.

**Middle ear and nasopharynx**

We found that the middle ear and nasopharynx of children with rAOM were not highly concordant. In Fig. 4, the NPS samples are distinct from the ear samples (including MEF, MER and ECS). Differential abundance analysis revealed 23 significant OTUs above threshold between the MEF and NPS (an additional 2 OTUs above threshold were dropped for disagreement between Set 1 and Set 2), and 18 OTUs when comparing the MER to the NPS (4 OTUs above threshold dropped for disagreement). The abundance of significant OTUs are shown in Fig. 5b and c. Overall, the genera Staphylococcus (OTU3), Alloiococcus (OTU1) and Turicella (OTU7) were highly abundant in the middle ear, with very low abundance in the nasopharynx (logFC = 2.6 to 5.2, p-values < 0.001); these genera appear to be characteristic of the middle ear. OTU3 matched several species of Staphylococcus above 97% with BLASTN, however Alloiococcus and Turicella both contain one known species each (A. otitidis and T. otitidis respectively). The otopathogen genera Haemophilus (OTU6), Moraxella (OTU2) and Streptococcus (OTU4) were moderately abundant in both sites but were higher in the nasopharynx (logFC = 2.3 to 6.4, p-values < 0.002). Several OTUs including Gemella (OTU12) and Neisseria (OTU13) were
at very low abundance in the middle ear and higher in the nasopharynx, though still low overall (logFC = 0.8 to 3.6, p-values < 0.05). These low-abundance OTUs appear to be contributing to the increased diversity seen in the nasopharynx compared to the middle ear. Differences between the middle ear and nasopharynx were generally more pronounced when comparing to the MER samples.

**Middle ear fluid and ear canal**

The ear canal samples were taken to assess which bacteria may potentially contaminate the MEF and MER samples during collection. No bacterial taxa were present in the ear canal that were not present in the middle ear, and the dominant bacteria in the canal were also dominant in the middle ear (see Table 4). Shown in Fig. 5 d, only three OTUs were significantly differentially abundant between pairs of ECS and MEF samples. *Alloiooccus* (OTU1), *Turicella* (OTU7) and *Staphylococcus* (OTU3) occurred at higher abundance in the ear canal (logFC = 2.9, 1.7 and 2.7 respectively; p-values < 0.004), though they were common in both sites. *Haemophilus* (OTU6) was higher in the middle ear fluid with a log fold change of 3.6 ($p = 2.5 \times 10^{-13}$).

**Patterns of bacterial co-occurrence**

We searched for strong correlations between the dominant OTUs representing the known otopathogen genera (*Haemophilus*, *Streptococcus* and *Moraxella*) and also between the OTUs we found to be characteristic of the case or control nasopharynx, or the middle ear. Correlations were determined separately for each sample type with SparCC [58]. Correlograms showing the overall pattern of correlations for each sample type are shown in Fig. 6, and a full list of correlations can be found in Additional file 11.

Within the case nasopharynx, there were several moderately strong (0.6–0.8) correlations with $p < 0.01$ (100 simulations). The strongest correlations observed were between low-abundance OTUs *Haemophilus* (OTU1033) and *Moraxella* (OTU1073) with coefficient 0.80; between the most abundant *Moraxella* OTU (OTU2) and *Haemophilus* (OTU1033) with coefficient 0.77 and between *Gemella* (OTU12) and *Porphyromonas* (OTU23) with coefficient 0.70. *Gemella* (OTU12) and *Neisseria* (OTU13), OTUs that were significantly more prevalent in the cases than in controls, were also positively correlated (0.65). Correlations between the dominant OTUs for *Haemophilus* (OTU6), *Moraxella* (OTU2) and *Streptococcus* (OTU4) were not significant.
Within the control NPS samples, there were fewer significant correlations observed. A moderately strong correlation (0.722, \( p < 0.01, 100 \) simulations) was detected between \( \text{Corynebacterium} \) (OTU5) and \( \text{Dolosigranulum} \) (OTU8), which were significantly more prevalent in the nasopharynx of controls than cases. \( \text{Moraxella} \) (OTU10) correlated only weakly with other OTUs.

In the middle ear, the dominant \( \text{Staphylococcus} \) (OTU3) correlated with other low-abundance \( \text{Staphylococcus} \) OTUs (1003 and 212), with coefficients 0.61 to 0.73 (\( p \)-values < 0.01, 100 simulations) in both the MEF and MER. These minor \( \text{Staphylococcus} \) OTUs were also moderately to strongly correlated with each other (coefficient 0.83 in MER, 0.71 in MEF). A moderately negative correlation was observed between \( \text{Staphylococcus} \) (OTU3) and \( \text{Alloiococcus} \) (OTU1) in the MER (coefficient \( = -0.52, p < 0.01, 100 \) simulations), although this association was weak in the MEF (coefficient \( = -0.3, p < 0.01, 100 \) simulations). The genera that appear to be characteristic of the middle ear, \( \text{Alloiococcus} \) (OTU1) and \( \text{Turicella} \) (OTU7), only weakly correlated with each other in the MEF (coefficient \( = 0.37, p < 0.01, 100 \) simulations) but not in the MER. No significant correlations were found between \( \text{Haemophilus} \) (OTU6), \( \text{Streptococcus} \) (OTU4) and \( \text{Moraxella} \) (OTU2) in either the MEF or MER.

### Detection of respiratory viruses

We tested all MEF and NPS samples for common respiratory viruses. The identification rates (the percentage of positive samples out of the total number of samples tested) are shown in Table 5. Amongst the NPS samples, 61.3% of the cases were positive for at least one virus compared to 47.6% of the healthy controls, though the odds of detecting any virus were not significantly higher in the cases (OR 1.4, (0.8, 2.6)). The odds of detecting respiratory syncytial virus (RSV) were 9.6 times higher (2.2, 60.5) in the cases than the controls. HMPV was also substantially different between the groups; while the detection rate was low in the cases (8.6%) the virus was not detected in any of the controls. For all other viruses tested, the odds of detection were also significantly higher in the cases than in the controls. RV was the most frequently detected virus overall, seen in 43% of case NPS and 34% of control NPS. No samples in either the cases or controls tested positive for HBoV.

We calculated the rate of concordance in viral detection between the MEF and NPS for each virus to determine whether any of these viruses are found at both sites in the same child (Table 6). Overall, the concordance rates were low; RV showed the highest concordance rate with 44.4% of cases with RV having detectable virus in both sites. Each virus was detected more frequently in the nasopharynx than in the middle ear, with the exception of IFV which was only detected in one NPS and one MEF from different children.

### Discussion

The microbiome of the middle ear in children with rAOM has remained relatively unexplored. Modest proportions of children with AOM carry no detectable otopathogen in the MEF by culture [5, 7] or PCR [5, 8], so there is an opportunity for the identification of novel otopathogens by studying the microbiome of the otitis-prone middle ear. Additionally, studies of the microbiome can help to identify protective bacteria that give rise to probiotic therapies [28–30], which present a novel treatment option for...
Table 5 Viral identification rates in the nasopharynx of cases and controls. The identification rate is the percentage of positive samples out of the total number of samples collected for that group.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Case NPS (n = 93)</th>
<th>Control NPS (n = 103)</th>
<th>Odds ratio (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFV</td>
<td>1 (1.1%)</td>
<td>2 (1.9%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HAdV</td>
<td>18 (19.4%)</td>
<td>6 (5.8%)</td>
<td>3.9 (1.5, 10.3)</td>
<td>0.004</td>
</tr>
<tr>
<td>HBoV</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RSV</td>
<td>15 (16.1%)</td>
<td>2 (1.9%)</td>
<td>9.6 (2.2, 60.5)</td>
<td>0.0005</td>
</tr>
<tr>
<td>HCoV</td>
<td>14 (15.1%)</td>
<td>5 (4.9%)</td>
<td>3.5 (1.1, 10.4)</td>
<td>0.03</td>
</tr>
<tr>
<td>HPIV</td>
<td>4 (4.3%)</td>
<td>4 (3.9%)</td>
<td>1.1 (0.3, 4.8)</td>
<td>1</td>
</tr>
<tr>
<td>HMPV</td>
<td>8 (8.6%)</td>
<td>0 (0%)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RV</td>
<td>40 (43.0%)</td>
<td>35 (34.0%)</td>
<td>1.5 (0.8, 2.6)</td>
<td>0.2</td>
</tr>
<tr>
<td>Total</td>
<td>57 (61.3%)</td>
<td>54 (47.6%)</td>
<td>1.4 (0.8, 2.6)</td>
<td>0.2</td>
</tr>
</tbody>
</table>

* Viral abbreviations: IFV Influenza virus, HAdV Adenovirus, HBoV Human bocavirus, RSV Respiratory syncytial virus, HCoV Human coronavirus, HPIV Parainfluenza virus, HMPV Human metapneumovirus, RV Rhinovirus.

rAOM. Our study explored both of these opportunities and found evidence of bacterial genera that may represent novel otopathogens as well as genera that are worth further investigation as probiotic candidates. In doing so, we have, to our knowledge, characterised for the first time the microbiome of both the middle ear and nasopharynx in children with rAOM.

Our study determined that the nasopharyngeal microbiome of rAOM-resistant children is distinct from that of rAOM-prone children. Specifically, we identified three bacteria that were more abundant in our healthy controls: Corynebacterium (OTU8), Dolosigranulum (OTU5) and, to a lesser extent, Moraxella (OTU10). The exact Corynebacterium species could not be determined, but the genus Dolosigranulum contains only one species, D. pigrum, and the Moraxella OTU was most likely M. lincolnii by BLASTN. This Moraxella OTU was distinct from the dominant Moraxella (OTU2) found abundantly in the nasopharynx of both cases and controls, but OTU10 was not prevalent in the nasopharynx of all controls. The results of our study are supported by previous reports that Corynebacterium and Dolosigranulum are found in the nasopharynx of healthy adults [63] and children [64, 65] and may be associated with a decreased risk of AOM [66]. We also observed a positive correlation between Corynebacterium and Dolosigranulum which has been previously reported [67]. It has been suggested that the production of lactic acid by Dolosigranulum makes the surrounding environment more habitable for Corynebacterium species [68], which would explain the co-occurrence we observed in our study. Some studies have observed a decrease in the abundance of these genera in the nasopharynx of children who had taken antibiotics in the preceding months [65, 66], and they were also detected more frequently in the nasopharynx of breastfed compared to formula-fed 6 week old infants [69]. Together these results indicate that Corynebacterium and Dolosigranulum may be part of the normal flora of the nasopharynx, which might be strengthened by breastfeeding and disrupted by antibiotic use. In particular, D. pigrum's closest relative is Alloiooccus (one species; A. otitidis) [71]. Alloiooccus and Turicella have both been identified as potential otopathogens [9, 72] and both were abundant in the MEF, MER and ECS of cases in this study, although they were not commonly found in the nasopharynx of our cases. Currently, the pathogenic role of Alloiooccus and Turicella, if any, is yet to be determined. Other studies have identified bacteria with a protective role that are not closely related to the pathogens they inhibit [18, 30]. Corynebacterium is a member of the same family as Turicella (one species; T. otitidis) and D. pigrum's closest relative is Alloiooccus (one species; A. otitidis) [71]. Alloiooccus and Turicella have both been identified as potential otopathogens [9, 72] and both were abundant in the MEF, MER and ECS of cases in this study, although they were not commonly found in the nasopharynx of our cases. Currently, the pathogenic role of Alloiooccus and Turicella, if any, is yet to be determined. Other studies have identified bacteria with a protective role that are not closely related to the pathogens they inhibit [28, 29]. In the anterior nares, Dolosigranulum has been linked to decreased rates of colonisation with S. aureus [73] and Corynebacterium has demonstrated activity against S. pneumoniae in vitro [74], indicating the potential for pathogen inhibition by these bacteria in vivo. While we have shown that Corynebacterium and Dolosigranulum are characteristic of rAOM-resistant children, determining whether they play a role in protecting against rAOM is more

Table 6 Viral concordance rates between the middle ear and nasopharynx of cases. The concordance rate is the number of cases where the virus was identified in both the MEF and NPS out of the total number of cases where the virus was detected at all.

<table>
<thead>
<tr>
<th>Virus</th>
<th>IFV</th>
<th>HAdV</th>
<th>HBoV</th>
<th>RSV</th>
<th>HCoV</th>
<th>HPIV</th>
<th>HMPV</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases with virus in MEF</td>
<td>1 (50.0%)</td>
<td>3 (14.3%)</td>
<td>0 (0%)</td>
<td>10 (40.0%)</td>
<td>13 (48.1%)</td>
<td>3 (42.9%)</td>
<td>4 (33.3%)</td>
<td>38 (48.7%)</td>
</tr>
<tr>
<td>Cases with virus in NPS</td>
<td>1 (50.0%)</td>
<td>18 (85.7%)</td>
<td>0 (0%)</td>
<td>15 (60.0%)</td>
<td>14 (51.9%)</td>
<td>4 (57.1%)</td>
<td>8 (66.7%)</td>
<td>40 (51.3%)</td>
</tr>
</tbody>
</table>

* Viral abbreviations: IFV Influenza virus, HAdV Adenovirus, HBoV Human bocavirus, RSV Respiratory syncytial virus, HCoV Human coronavirus, HPIV Parainfluenza virus, HMPV Human metapneumovirus, RV Rhinovirus.
challenging. We controlled for antibiotic usage within the month prior to sampling, however we did not have information relevant to long-term or repeat antibiotic use. Additionally, as this is a cross-sectional study we cannot determine whether decreased *Corynebacterium* and *Dolosigranulum* precedes the development of *rAOM*, or if their depletion is a result of the disease. While our study is unable to answer these questions, the abundance of these genera in the controls warrants further investigation to determine whether they have a protective role in the nasopharynx.

As well as identifying potentially protective bacterial species we also investigated the potential for novel otopathogen identification. It is generally accepted that the otopathogens originate from the nasopharynx and ascend the Eustachian tube to cause infections in the middle ear. We found that the middle ear and nasopharynx of children with *rAOM* were not highly concordant, supporting similar findings from previous reports [75, 76]. *Haemophilus* was the most abundant otopathogen genus in the MEF, which corresponds with surveillance studies reporting a predominance of non-typeable *Haemophilus influenzae* (NTHi) in children with *rAOM* since the introduction of the pneumococcal vaccine targeting *S. pneumoniae* [5]. However, *Haemophilus*, *Moraxella* and *Streptococcus* were observed at low abundance in the MEF and MER compared to *Alloiococcus* (OTU1), *Staphylococcus* (OTU3) and *Turicella* (OTU7), with *Alloiococcus* the overall most abundant genus in the MEF, MER and ECS. There are many species of *Staphylococcus*, and we couldn’t identify OTU3 at species level by BLASTN. *S. aureus* has previously been isolated from the MEF of children with AOM [77] but is usually associated with chronic suppurative OM where the tympanic membrane is perforated [78, 79]. The other genera each contain only one species (*A. otitidis* and *T. otitidis* respectively) which have previously been identified as possible otopathogens [9, 72]. At present, the role of these bacteria in OM is still under debate [10, 80]. *A. otitidis* has been frequently detected [8, 81–83] at high abundance [81, 82] in MEF, usually from children with OME; but *T. otitidis* has been less well studied. Recently, its abundance was associated with older children (> 24 months) with COME [84], though we did not observe this pattern in our cohort. Unlike the known otopathogens, *Alloiococcus*, *Staphylococcus* and *Turicella* have been identified as members of the normal ear canal flora in both children and adults [85] and in our study were more abundant in the ECS than in the MEF. Also unlike the otopathogens, *Turicella* was almost absent from the nasopharynx of our cases, whilst *Alloiococcus* and *Staphylococcus* were also uncommon in this niche. *Alloiococcus* has previously been isolated infrequently from the nasopharynx of children with upper respiratory tract infections or a history of AOM [8, 10]. The rarity of *Alloiococcus* and *Turicella* in the nasopharynx suggests that they may typically colonise the middle ear. Few studies have investigated the normal flora of the healthy middle ear; when healthy, this site contains no fluid and is inaccessible without a surgical procedure. Those that have sampled this site in healthy children did not report *Alloiococcus* [86, 87] and *Turicella* appears to have only been observed once in the healthy middle ear of an adult [87]. While *Alloiococcus*, *Staphylococcus* and *Turicella* seem to have the potential to be novel otopathogens, we cannot yet exclude the possibility that they are normal aural flora.

We also observed increased abundance of *Alloiococcus*, *Staphylococcus* and *Turicella* in the MER compared to the MEF; whilst *Haemophilus* was more common in the MEF and *Streptococcus* and *Moraxella* were not significantly different between the two sample types. It is possible that *Alloiococcus*, *Staphylococcus* and *Turicella* adhere to the mucosa and were more efficiently sampled by the MER; differences between the middle ear and NPS were more pronounced when comparing the MER than MEF, suggesting the MER might better represent the microbiome of the middle ear. Alternatively, the MER may include contamination from the ear canal, as *Alloiococcus*, *Staphylococcus* and *Turicella* were the most dominant organisms in the canal and were significantly more abundant here than in the MEF. We observed increased diversity in the MER compared to the MEF, though our differential abundance analysis revealed significant differences only between genera that were common in both sites. The increased diversity may therefore include low abundance contaminant OTUs, suggesting that sampling methods to obtain the MER may be more prone to environmental or skin contamination than the MEF. However, as there were no genera unique to the ear canal and the genera at this location were found abundantly in both the ear canal and middle ear it is difficult to determine whether the canal flora contaminates the middle ear fluid with *Alloiococcus*, *Staphylococcus* and *Turicella* during sampling.

We observed that the nasopharyngeal microbiome of the cases was significantly more diverse than that of controls. This is in contrast to studies of the gut microbiome (a high-diversity environment [88]), where greater diversity has been associated with a healthy state, and a decrease in diversity is characteristic of disease [89, 90]. However, studies of the vaginal microbiome (a low-diversity environment [91]) have shown a similar pattern to that observed in the nasopharynx in our study; that a lower diversity is observed in the healthy environment with an increased diversity being characteristic of disease [92, 93]. Previous studies suggest that the nasopharynx is dominated by only a few taxa [64, 65] so it may be that this pattern is characteristic of less complex microbiomes. Additionally, a study in the
adult nasopharynx determined that a more diverse microbiome was more susceptible to colonisation with *S. pneumoniae* [63]. We could not determine from our cross-sectional study whether the higher diversity in the cases occurs before or after the development of rAOM, or as a consequence of repeat antibiotic treatment where the normal nasopharyngeal microbiome is disrupted, perhaps allowing the community to diversify. A study in adults has indicated that the gut microbiome can begin to resemble its original state in as little as one week after antibiotic treatment [94], though this recovery is often incomplete. The microbiome of adults with cystic fibrosis is similarly resilient to short-term antibiotics [95]. However, antibiotic-induced changes in the microbiome in children may be more long-term. In the gut microbiome of children, reduced richness can persist for up to 2 years after the use of macrolides [96]. Similarly, after long-term oral and inhaled antibiotic use the taxonomic richness of the microbiome in children with cystic fibrosis is markedly reduced [97]. It is possible that in our cohort, the nasopharyngeal microbiome of rAOM-prone children has been altered by antibiotic use that extends beyond the month prior to sampling that we controlled for. However, based on these studies, a reduction in microbial diversity would be expected with repeated antibiotic use; this is the opposite effect to what we observed in our cases.

The increased diversity in the case nasopharynx was contributed to in part by the presence of *Gemella* (OTU12) and *Neisseria* (OTU13), which positively correlated with each other. The OTU representative sequences matched multiple species above 97% identity with BLASTN, however the SILVA database classified OTU13 as *N. lactamica*, which is a commensal of the nasopharynx in children [98]. While *Gemella* and *Neisseria* were characteristic of the nasopharynx in children with rAOM, their abundance in the middle ear was significantly lower, suggesting these genera are not involved in pathogenesis in the middle ear and are unlikely to be novel otopathogens. Both *Neisseria* and *Gemella* have been observed in the nasopharynx of children with upper respiratory illness and AOM [66], though *Neisseria* has also previously been reported amongst the healthy flora of the nasopharynx [64, 99]. It is therefore unclear whether these genera play a role in the nasopharynx of children with rAOM or if they represent a shift in the composition of the microbiome due to repeated antibiopic usage or other factors.

Contrary to the results of our study, Hilty et al. [99] and Pettigrew et al. [66] both reported decreased diversity in the nasopharynx of children with AOM compared to children without AOM. The nasopharyngeal samples in these studies were taken during an episode of AOM, so possibly better represent the nasopharyngeal microbiome at this time. However, the majority of healthy controls used for comparison did not attend day care and most of the cases did not have a history of recurrent AOM (or this information was not provided), so these cohorts may represent less extreme ends of the phenotypic spectrum of disease.

Episodes of AOM often occur after a viral upper respiratory tract infection, [100] with bocavirus (HBoV) [101], rhinovirus (RV) [6] and respiratory syncytial virus (RSV) [102] commonly found in children with AOM. Rhinovirus was the most frequently detected virus in the nasopharynx of both our cases and controls. We detected other viruses more frequently in the cases than controls, though the odds were not always significant as relatively few viruses were detected overall. The odds of detecting respiratory syncytial virus were significantly higher in the cases than controls, which supports its association with an increased risk of AOM [103]. We also observed human metapneumovirus (HMPV) in the nasopharynx of a small number of cases, but it was absent from all of the controls, suggesting an association with rAOM. HMPV has previously been observed in 5% of children with upper respiratory tract infections, with half of these also experiencing AOM [104]. Our results were similar to the patterns of viral detection in the nasopharynx in an independent Western Australian cohort of children with and without rAOM [6], though the overall rate of viral detection was higher in that study (rAOM: 94% with at least one virus detected, controls: 71%, compared to our 61% and 48% respectively). This may be because they additionally tested for polyomavirus and enterovirus, which they found in 36% and 17% of children with rAOM respectively; for polyomavirus this was significantly higher than in their healthy controls. Bocavirus was not detected in any of the cases or controls in our study, although it has previously been seen in children with rAOM [6, 101]. We detected viruses less often in the middle ear than in the nasopharynx, which Wiertsema et al. [6] also reported. Despite lower viral identification rates, results from our cohort therefore seem to follow the same pattern as previously reported in Western Australian children.

Rhinovirus [105] and respiratory syncytial virus [106] have been observed to enhance the adherence of *S. pneumoniae* to epithelial cells in vitro. It has previously been reported that rhinovirus presence correlates with presence by standard culture of each of the three major otopathogens [107], and adenovirus with *M. catarrhalis* presence assessed by culture and PCR [6, 107]. We could not observe this in our study as 16S rRNA gene sequencing could not identify the otopathogen species; sequences that are more than 97% identical are grouped into the one OTU, which we could only identify at genus level. It is possible that *Haemophilus*, Streptococcus and *Moraxella* OTUs include both the otopathogens and
commensal species from the same genus, which would obscure viral/otopathogen correlations.

One limitation specific to our study is that grommet surgeries are generally not performed during active infection, so the microbiome of our rAOM-prone children may not be fully representative of the microbiome during an episode of AOM. 16S rRNA studies in general can provide a comprehensive overview of the taxonomic composition of the microbiome, but are limited in that they do not provide information on microbiome function or gene content. Additionally, there are important biases to consider when conducting these studies. DNA extraction methods [108] and amplicon primers [39] work with variable efficiency across bacterial taxa, which can result in the underrepresentation of some bacteria. Our DNA extraction protocol and amplicon primers were chosen based on recommendations by Yuan et al. [108] and Klindworth et al. [39] respectively to reduce this bias. The number of copies of the 16S rRNA gene can vary amongst bacteria [109], even between strains [110] which can inflate OTU abundance. Additionally, copies within a single genome are not always identical [111], which can inflate the number of OTUs detected. Samples can be contaminated with DNA found in reagents and the laboratory environment [112], and there is the potential in our study for contamination during sample collection (i.e. from the ear canal or anterior nares). Contamination can heavily influence low biomass environments [112, 113], however we sequenced positive and negative controls to address this (see Additional file 4). Furthermore, it is difficult to achieve species-level identification with 16S rRNA sequencing as related species are often very similar in this region and the efficiency of classification also varies depending on the region of the 16S rRNA gene [114]. There is the possibility that there are commensal bacteria that we could not detect which are within the same OTU as the otopathogens. For example, Haemophilus haemolyticus and Streptococcus salivarius are closely related to otopathogens and have shown promise in other studies as candidates for the prevention of rAOM [21, 115]. Metagenomic shotgun sequencing addresses the issue of species-level identification as it sequences across the entire genome, and can thereby also provide information on gene content and function. However, DNA sequencing itself can only detect bacterial presence and does not indicate bacterial viability or activity. Metatranscriptomics addresses this issue, but this is a relatively new field and has not yet been applied to OM. The precautions taken in our study aimed to reduce the biases inherent to 16S rRNA sequencing, however we acknowledge that the relative abundance of taxa may not reflect the true proportions of bacteria and the genera we have detected may contain multiple species which are not necessarily viable or active.

Conclusions

Our study has provided the first comprehensive exploration of the microbiome of the middle ear and nasopharynx in children with rAOM. We have taken an important step in the identification of candidate therapeutic bacteria derived from the healthy microbiome by observing significantly higher proportions of Corynebacterium and Dolosigranulum in the nasopharynx of healthy controls. Further research should focus on investigating their potential to inhibit the known otopathogens, and it would be of interest for longitudinal studies to determine whether the abundance of these genera change prior to or as a result of rAOM. We have also identified Alloiooccus, Staphylococcus and Turicella as potential otopathogens, though their relative absence in the nasopharynx and abundance in the ear canal means we cannot rule out their role as normal aural flora. Gemella and Neisseria contribute to increased diversity in the nasopharynx of children with rAOM, but do not appear to colonise the middle ear and are therefore not likely to be novel otopathogens. Shotgun metagenomics and metatranscriptomics are the next steps towards achieving species-level or strain-level resolution of these bacteria of interest and confirming their viability and investigating their activity in the middle ear. Our study has contributed significantly towards greater understanding of the microbiome of the upper respiratory tract in children with rAOM, and has taken an important step towards the development of specific probiotic therapies for the disease.

Additional files

Additional file 1: Questionnaire completed by families recruited to the study. (PDF 28 kb)
Additional file 2: Figure S1. Diagrammatic overview of the 16S rRNA gene data analysis pipeline. Names of the software or tools used are in red. The SILVA database replaced the default taxonomy database in QIIME (GreenGenes) as GreenGenes 13_8 version does not discriminate between Alloiooccus and Dolosigranulum (PDF 366 kb)
Additional file 3: Table S1. Full taxonomy of all OTUs identified in this study. Taxonomy is from the SILVA database, v123 for QIIME. Taxonomy assigned to OTUs by UCLUST v1.22.2q within QIIME v1.9.1 (design_taxonomy.py). (XLSX 13 kb)
Additional file 4: Results from the positive and negative sequencing controls, including Table S2. (DOCX 27 kb)
Additional file 5: Figure S2. Procrustes analysis of raw and rarefied datasets. The rarefied dataset was subsampled at a threshold of 1499 reads per sample. The raw dataset excluded samples below this depth. P-values are non-parametric and are based on 999 Monte Carlo simulations. (PNG 174 kb)
Additional file 6: Figure S3. Beta diversity PCoA in the nasopharynx of cases and controls, sorted by other covariates. Case and control nasopharyngeal samples shown in Fig. 3 are coloured by other covariates. NA refer to samples where the covariate was not applicable or was missing (not given or recorded “unknown”) and the number represents individual samples. (PDF 564 kb)
Additional file 7: Table S3. Complete list of significantly differentially abundant OTUs determined by metagenomeSeq. All differentially abundant OTUs between a) MEF and MEF; b) MEF and NPS; c) MEF and NPS; d) ECS and MEF are shown with their log fold change, p-values and mean and median abundance. OTUs in bold/grey are those above the selected threshold of at least 0.95% mean or median abundance in at least one group, and were present in both sets of samples where applicable. Set 1 refers to the set of samples were the left or right ear was chosen at random; these numbers are reported in results. Set 2 refers to the set of samples where the opposite ear was chosen; these results were only used for validation of the differentially abundant OTUs from Set 1. (DLSX 29 kb)

Additional file 8: Figure S4. Beta diversity PCoA in the samples from the cases, sorted by other covariates. Samples from the cases shown in Fig. 5 are coloured by other covariates. NA refers to samples where the covariate was not applicable or was missing. The p-value is non-parametric and is based on 999 Monte Carlo simulations. (PNG 91 kb)

Additional file 9: Figure S5. Procrustes analysis of left and right ear samples. The dataset includes both MEF and MEF samples in left/right ear pairs from the same child. Samples with less than 1499 reads are excluded. The p-value is non-parametric and is based on 999 Monte Carlo simulations. (PNG 80 kb)

Additional file 10: Figure S6. Procrustes analysis of MEF and MER samples. The dataset includes pairs of MEF and MER samples from the same ear of the same child. Samples with less than 1499 reads are excluded. The p-value is non-parametric and is based on 999 Monte Carlo simulations. (DLSX 443 kb)

Abbreviations
AOA: Acute otitis media; CSS: Cumulative-sum-scale; DNA: Deoxyribonucleic acid; EGS: Ear canal swab; ENT: Ear, nose and throat; HAdV: Human adenovirus; HBCV: Human bocavirus; HCoV: Human coronavirus; HMPV: Human metapneumovirus; HPV: Human papillomavirus; HRBC: Human research ethics committee; IFV: Influenza virus; IS: Inverse Simpson; MEF: Middle ear fluid; MER: Middle ear rinse; NPS: Nasopharyngeal swab; OM: Otitis media; OME: Otitis media with effusion; OTU: Operational taxonomic unit; PCoA: Principal coordinates analysis; PCR: Polymerase chain reaction; PD: Phylogenetic diversity; rAOM: Recurrent acute otitis media; rRNA: Ribosomal ribonucleic acid; RSV: Respiratory syncytial virus; RV: Rhinovirus; STGB: Skim milk tryptone glucose glycerol broth; TE: Tris-ethylenediaminetetraacetic acid

Acknowledgements
We thank the children and their families who participated in the Perth Otitis Media Microbiome (biOMe) study, the Australian Genome Research Facility for preparing and sequencing the 16S rRNA gene libraries and A/Prof K. Carter, manager of the McCusker Charitable Foundation Bioinformatics Centre, Telethon Kids Institute for supplying and maintaining the computational resources used. We would also like to thank A. Hoskins for the recruitment and sample collection for several of the controls, and G. Platell for helpful discussion regarding the sequence pre-processing pipeline.

Funding
The Perth Otitis Media Microbiome (biOMe) study was funded by a Telethon New Children’s Hospital research grant. Control samples were provided by the Central Immunisation Clinic Pneumococcal Colonisation Study, funded by WA Health. RL is supported by an Australian Government Research Training Program Scholarship and a top-up scholarship from the Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, The University of Western Australia. CCB is supported by a NHMRC Career Development Fellowship (APP1111596). The funding bodies had no role in the design of the study, the collection, analysis or interpretation of data, or in writing the manuscript.

Availability of data and materials
The dataset generated and analysed during the current study is available in the NCBI Sequence Read Archive (SRR108387; available at https://www.ncbi.nlm.nih.gov/traces/study/?taxid=9606) and additional data is available from the corresponding author on reasonable request. Complete documentation of all 16S rRNA sequence analysis code can be found at https://raelahalledtannap.github.io/16S-analysis/.

Authors’ contributions
SL and CSP conceived the study, RL, KL, CS, HC, SV, PB, CCB, CSP and SEJ designed and planned the study, RL, KL, HC, SV, PB and CCB identified and recruited participants; RL, KL, HC, SV and PB collected specimens and RL and KL conducted questionnaire interviews. DM, SEJ, CS, DA, SHA and CG undertook the DNA extraction and bioinformatics analysis. RL and DA undertook the statistical analysis. RL, DA, DM, CSP and SEJ interpreted the data. RL, KL, DA, DM, CS, HC, CCB, CSP and SEJ prepared the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Recruitment to the study and the study protocol were approved by the Human Research Ethics Committees (HREC) at Princess Margaret Hospital for Children (2013119/EP), St John of God Health Care (K708) and the University of Western Australia (RA/1/1639) as well as by all relevant hospital governance committees. Parents or guardians of children recruited to the study provided written informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1The Marshall Centre for Infectious Diseases Research and Training, School of Biomedical Sciences, The University of Western Australia, Perth, WA, Australia. 2School of Biomedical Sciences, The University of Western Australia, Perth, WA, Australia. 3Department of Microbiology, PathWest, Perth, WA, Australia. 4Department of Microbiology, PathWest, Perth, WA, Australia. 5School of Medicine, The University of Western Australia, Perth, WA, Australia. 6Princess Margaret Hospital for Children, Perth, WA, Australia.

Received: 26 July 2017 Accepted: 8 February 2018

Published online: 20 February 2018

References
of respiratory viruses in the nasopharynx and the middle ear of children
8. Kaur R, Ackowitz DG, Casey JR, Zeng M, Pichichero ME. Simultaneous assay for four
bacterial species including Allobacillus otitidis using multiplex-PCR in children
11. Venekamp RP, Sanders SL, Glaziov PP, Del Mar CB, Rovers MM. Antibiotics
No.: CD002519.
13. Bostom M, McCook J, Burke B, Denker C. Incidence of and risk factors for additional
16. Rautava S, Salminen S, Isolauri E. Specific probiotics in reducing the risk of
acute infections in infancy – a randomised, double-blind, placebo-controlled
17. Steckel-Blicks C, Sjöström I, Twetman S. Effect of long-term consumption
of milk supplemented with Probiotic lactobacill and fluoride on dental
caries and general health in preschool children: a cluster-randomized study.
taurolosis K12 in preventing recurrent pharyngitis and/or tonsillitis caused by
streptococcus pyogenes and recurrent acute otitis media. Int J Gen Med.
19. Roos K, Håkansson EG, Holm S. Effect of recoinfection with “interfering”
(alpha) streptococci on recurrences of acute and secretory
Prophylactic administration of Probiotic lactobacilli in children attending day care
23. Taipale TJ, Pienihäkkinen K, Isolauri E, Jokela JT, Söderling EM. Rapid
identification of Staphylococcus aureus using multiplex real-time PCR
24. Teo SM, Mok D, Pham K, Kusel M, Serhal MA, Troy N, et al. The infant airway
microbiome in health and disease impacts later asthma development. Cell
Host Microbe. 2015;17:70–94.
multiplex real-time PCR technique for the detection of a comprehensive
previously unrecognized human rhinoviruses are common causes of
27. Stephens DS, Swarey JS, Kuthar P, Morse SA. Insertion of Tn916 in
Staphylococcus aureus using multiplex-PCR in children
bacterial species including Allobacillus otitidis using multiplex-PCR in children
five-strain probiotic combination reduces pathways shedding and alleviates
Epidermidis Esp inhibits Staphylococcus Aureus biofilm formation and nasal
RH, et al. Prevalence and risk factors for parent-reported recurrent otitis
media during early childhood in the western Australian pregnancy-cohort
33. Teo SM, Mok D, Pham K, Kusel M, Serhal MA, Troy N, et al. The infant airway
microbiome in health and disease impacts later asthma development. Cell
Host Microbe. 2015;17:70–94.
34. Chiodo GR, Hammett GL, Shelgam SM, Smith DW. An economical tandem
multiplex real-time PCR technique for the detection of a comprehensive
previously unrecognized human rhinoviruses are common causes of
36. Stephens DS, Swarey JS, Kuthar P, Morse SA. Insertion of Tn916 in
Staphylococcus aureus using multiplex-PCR in children
ribosomal RNA gene database project: improved data processing and web-
and “ali-species living tree project (LTP)” taxonomic frameworks. Nucleic Acids
of general 16S ribosomal RNA gene PCR primers for classical and
gen-next-generation sequencing-based diversity studies. Nucleic Acids
40. Andrews S. FastQC: a quality control tool for high throughput sequence
41. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon
42. Edgar RC. Search and clustering orders of magnitude faster than BLAST.
46. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R.
PyNAST, a flexible tool for aligning sequences to a template alignment.
gen-next-generation sequencing-based diversity studies. Nucleic Acids


A.2 Supplementary File 1

The questionnaire completed by families recruited to the study.
Defining the microbes in the middle ear and upper respiratory tract that lead to recurrent ear infections

Data Collection Form

<table>
<thead>
<tr>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s DOB __ __ / __ __ / __ __          I.D. Number: M G O __ __ __</td>
</tr>
<tr>
<td>Gender: □ Male    □ Female</td>
</tr>
<tr>
<td>Identify with being Aboriginal or Torres Straight Islander? □ Yes    □ No</td>
</tr>
<tr>
<td>Suburb (where participant lives): ___________________________ Postcode:_______</td>
</tr>
<tr>
<td>Who is providing the information? □ Mother □ Father □ Grandparent □ Guardian □ Foster Carer □ Other_________________</td>
</tr>
</tbody>
</table>

| Is the child currently breastfed? □ Yes □ No □ N/A □ Unknown |
| If Yes: □ Exclusive □ Partial □ Unknown |
| How old was child when breastfeeding ceased? ________________ □ Unknown |

| Is the child attending day care or school? □ Day care □ School □ Neither □ N/A □ Unknown |
| If at day care/school, how many hours/wk _________ |
| What age (in months) did they start day care_________ |

| Number of people normally living in your house |
| Total number __ __ Number of children (at or under 5 years of age) __ __ |
| Does anyone at home smoke? □ Yes □ No □ Unknown |
| If yes: How many people? __ __ □ Unknown |
| If yes: How often do they smoke? □ Occasionally □ Regularly □ Unknown |
| If yes: Does anyone smoke inside your house? □ Inside □ Outside only □ Unknown |
Does your child have any chronic illnesses such as: If Yes, describe

- Ear infection
  - Yes
  - No
  - Unknown

  *If yes,* How many ear infections have they had in total? __________________  Unknown

  Has your child previously had grommets inserted?  Yes  No  Unknown

  Has your child previously had their adenoids removed?  Yes  No  Unknown

- Asthma
  - Yes
  - No
  - Unknown

- Allergy
  - Yes
  - No
  - Unknown

- Chest problem
  - Yes
  - No
  - Unknown

- Heart problem
  - Yes
  - No
  - Unknown

- Kidney problem
  - Yes
  - No
  - Unknown

- Any other illnesses
  - Yes
  - No
  - Unknown

Has your child ever been admitted to hospital for infection?  Yes  No  Unknown

Please specify (Hospital, date and site - LRTI, meningitis, gastrointestinal, UTI etc)

___________________________________________________________________________________

___________________________________________________________________________________

___________________________________________________________________________________

Antibiotic usage during the last month:  Yes  No  Unknown

*If yes:* Date started _ ___/___/____ Date finished _ ___/___/____

Name of Antibiotic given:

- Penicillin
- Amoxycillin
- Augmentin
- Flucloxacillin
- Cepahlexin
- Cefaclor
- Ceftriaxone
- Cotrimoxazole
- Erythromycin
- Roxithromycin
- Azithromycin
- Other ____________

GP Name: ________________________________________________________________

Practice Name: ___________________________________________________________

Address: __________________________________________________________________

___________________________________________________________________________
### General health (on day of surgery)

#### Nose

**Any Nasal discharge?**
- □ Yes
- □ No
- □ Unknown

*If Yes, Discharge severity:*
- □ Mild
- □ Moderate
- □ Profuse
- □ Unknown

*Discharge Colour:*
- □ Clear
- □ Opaque
- □ Yellow/green
- □ Unknown

#### Ear

**Any ear discharge?**
- □ Yes
- □ No
- □ Unknown

*If yes, which ear?*
- □ Left
- □ Right
- □ Both
- □ Unknown

#### General

**Has the child been diagnosed with...**

* Cleft lip/palate?
  - □ Yes
  - □ No
  - □ Unknown

* Immune deficiency?*
  - □ Yes
  - □ No
  - □ Unknown

* Any genetic syndrome?*
  - □ Yes
  - □ No
  - □ Unknown

### Immunisation Status

Your child’s current immunisation status will be checked on the Australian Childhood Immunisation Register

ACIR checked by (researcher name): ________________________________  
Date ACIR checked  __ __ / __ __/____  
Name of Research Staff: ___________________________  
Signature: ___________________________  
Date of interview __ __ / __ __/____

---

Defining the microbes in the middle ear and upper respiratory tract that lead to recurrent ear infections  
Telethon Kids Institute, 100 Roberts Road, Subiaco WA 6008,  
Version 1: 30/10/2013  
A26  
Published manuscript and supplementary material (Chapter 2)
A.3 Supplementary File 2

A diagrammatic overview of the 16S rRNA gene data analysis pipeline.
Figure A.1: Diagrammatic overview of the 16S rRNA gene data analysis pipeline. Names of the software or tools used are in red. The SILVA database replaced the default taxonomy database in QIIME (GreenGenes) as GreenGenes 13.8 version does not discriminate between *Alloiococcus* and *Dolosigranulum*. 
A.4  Supplementary File 3

Full taxonomy of all OTUs identified in this study.

Taxonomy is from the SILVA database, v123 for QIIME. Taxonomy assigned to OTUs by UCLUST v1.2.22q within QIIME v1.9.1 (assign_taxonomy.py).

This table is too large to print. It can be accessed here:
A.5 Supplementary File 4

Results from the positive and negative sequencing controls.
Sequencing controls

Positive and negative sequencing controls are essential in a microbiome study. We included a positive mock community control to assess bias across independent sequencing runs. The negative controls were used to identify reagent or environmental contaminants during the DNA extraction process (negative extraction controls) or PCR amplification (negative PCR controls).

Table A.1 shows the summary of taxa for the positive sequencing control containing 16 bacterial species, which was included in duplicate on each of the four sequencing runs. These samples were highly correlated at genus level with the expected theoretical composition of 6.25% each species (Pearson correlation coefficient = 0.986, 95% CI 0.983 – 0.988, non-parametric p = 0.001, 999 permutations), indicating that separation of the samples onto four independent sequencing runs did not introduce bias.

The negative controls were assessed for taxa that may have an impact on the composition of the clinical samples, shown in Table A.2. These controls typically had a low sequence read count with 13/16 negative controls falling below the selected threshold of 1499 reads for removing low-depth samples from the analysis. The most abundant taxa across all negative controls included *Delftia* (20.0%), *Pseudomonas* (14.4%), *Alloiococcus* (13.4%) and *Lysinibacillus* (10.7%). The dominant *Pseudomonas* OTU in the negative controls (OTU58) was distinct from the dominant *Pseudomonas* OTU found in abundance in some of the ear samples (OTU9, presumptively *Pseudomonas aeruginosa* by BLASTN at 100% identity). The presence of *Alloiococcus* in a few negative control samples suggests some cross-contamination from the clinical samples, as it is known to be found in the middle ear and ear canal and was greatly abundant here. This is not likely to be an indication that the MEF samples are heavily contaminated by negative control taxa, as other organisms abundant in the negative controls occurred at very low relative abundance in the MEF samples. *Delftia* and *Lysinibacillus* are known environmental colonisers and these are found at a very low relative abundance in the clinical samples (median <0.003% and mean <0.1% relative abundance over all samples included in analysis). Furthermore, the negative controls were moderately correlated with each other at genus level when comparing taxonomy summaries. Pairs of extraction and PCR negative controls from the same sequencing plate (one of each per plate) had a Pearson correlation coefficient of 0.508 (95% CI 0.537-0.653, non-parametric p = 0.001, 999 permutations). Pairs of negative extraction controls on the same sequencing run had a Pearson correlation coefficient of 0.507 (95% CI 0.405-0.596, non-parametric p = 0.001, 999 permutations).
and pairs of negative PCR controls had a coefficient of 0.519 (95% CI 0.420-0.607, non-parametric p = 0.004, 999 permutations). This indicates that while there is a pattern amongst the negative controls, they do not contain a strong, consistent contaminant signal and are unlikely to influence the conclusions drawn from the samples due to the extremely low abundance of major contaminant taxa.
Table A.1: Taxonomic composition of the positive control (mock community) replicates.
Numbers represent column percentages. Taxa below an average relative abundance of 1% across all replicates were collapsed into “Other”.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Expected</th>
<th>MOCK1</th>
<th>MOCK2</th>
<th>MOCK3</th>
<th>MOCK4</th>
<th>MOCK5</th>
<th>MOCK6</th>
<th>MOCK7</th>
<th>MOCK8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corynebacterium</td>
<td>6.25</td>
<td>4.83</td>
<td>5.03</td>
<td>4.08</td>
<td>3.94</td>
<td>4.8</td>
<td>4.89</td>
<td>4.16</td>
<td>3.61</td>
</tr>
<tr>
<td>Propionibacterium</td>
<td>6.25</td>
<td>5.15</td>
<td>4.88</td>
<td>4.09</td>
<td>3.73</td>
<td>5.19</td>
<td>5.37</td>
<td>4.97</td>
<td>4.81</td>
</tr>
<tr>
<td>Staphylococcus</td>
<td>18.75</td>
<td>20.67</td>
<td>19.98</td>
<td>21.94</td>
<td>22.12</td>
<td>21.38</td>
<td>21.06</td>
<td>22</td>
<td>22.8</td>
</tr>
<tr>
<td>Globicatella</td>
<td>6.25</td>
<td>5.92</td>
<td>4.81</td>
<td>5.6</td>
<td>5.52</td>
<td>6.33</td>
<td>5.45</td>
<td>5.22</td>
<td>4.8</td>
</tr>
<tr>
<td>Alloiococcus</td>
<td>6.25</td>
<td>7.71</td>
<td>8</td>
<td>8.76</td>
<td>8.73</td>
<td>8.05</td>
<td>7.97</td>
<td>7.83</td>
<td>7.99</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>12.5</td>
<td>12.71</td>
<td>13.64</td>
<td>11.16</td>
<td>10.83</td>
<td>12.39</td>
<td>12.57</td>
<td>12.35</td>
<td>13.02</td>
</tr>
<tr>
<td>Veillonella</td>
<td>6.25</td>
<td>6.13</td>
<td>5.95</td>
<td>5.7</td>
<td>5.56</td>
<td>5.95</td>
<td>5.5</td>
<td>6.09</td>
<td>5.76</td>
</tr>
<tr>
<td>Neisseria</td>
<td>6.25</td>
<td>6.66</td>
<td>6.6</td>
<td>7.3</td>
<td>6.83</td>
<td>6.22</td>
<td>5.7</td>
<td>7.06</td>
<td>6.59</td>
</tr>
<tr>
<td>Klebsiella</td>
<td>6.25</td>
<td>7.14</td>
<td>7.3</td>
<td>8.17</td>
<td>8.22</td>
<td>7.02</td>
<td>7.11</td>
<td>7.03</td>
<td>6.47</td>
</tr>
<tr>
<td>Haemophilus</td>
<td>12.5</td>
<td>10.39</td>
<td>10.8</td>
<td>9.71</td>
<td>11.6</td>
<td>10.27</td>
<td>11.97</td>
<td>10.35</td>
<td>11.83</td>
</tr>
<tr>
<td>Pseudomonas</td>
<td>6.25</td>
<td>6.01</td>
<td>6</td>
<td>6.52</td>
<td>6.31</td>
<td>6.11</td>
<td>5.83</td>
<td>6.09</td>
<td>5.34</td>
</tr>
<tr>
<td>Other (47 other taxa)</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Table A.2: Taxonomic composition of the negative extraction and PCR controls. Numbers represent column percentages. Taxa below an average relative abundance of 1% across all replicates were collapsed into “Other”.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Overall</th>
<th>Ex1</th>
<th>Ex2</th>
<th>Ex3</th>
<th>Ex4</th>
<th>Ex5</th>
<th>Ex6</th>
<th>Ex7</th>
<th>Ex8</th>
<th>PCR1</th>
<th>PCR2</th>
<th>PCR3</th>
<th>PCR4</th>
<th>PCR5</th>
<th>PCR6</th>
<th>PCR7</th>
<th>PCR8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corynebacterium</td>
<td>1.12</td>
<td>0.39</td>
<td>0.05</td>
<td>7.29</td>
<td>0.24</td>
<td>3.98</td>
<td>1.36</td>
<td>0</td>
<td></td>
<td>0.04</td>
<td>0</td>
<td>0.68</td>
<td>3.29</td>
<td>0.12</td>
<td>0.26</td>
<td>0.17</td>
<td>0</td>
</tr>
<tr>
<td>Propionibacterium</td>
<td>1.18</td>
<td>0.19</td>
<td>1.57</td>
<td>2.13</td>
<td>0</td>
<td>13.93</td>
<td>0.78</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0.13</td>
<td>0.17</td>
<td>0</td>
</tr>
<tr>
<td>Bacillus</td>
<td>8.27</td>
<td>28.9</td>
<td>0</td>
<td>29.46</td>
<td>16.82</td>
<td>6.93</td>
<td>0</td>
<td>4.08</td>
<td>12.9</td>
<td>0.04</td>
<td>12.67</td>
<td>0</td>
<td>0</td>
<td>0.84</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Geobacillus</td>
<td>1.49</td>
<td>13.87</td>
<td>0</td>
<td>4.07</td>
<td>2.94</td>
<td>0</td>
<td>0</td>
<td>1.94</td>
<td>1.08</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lysinibacillus</td>
<td>10.69</td>
<td>0</td>
<td>26.5</td>
<td>0</td>
<td>9.02</td>
<td>0</td>
<td>44.78</td>
<td>4.08</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12.62</td>
<td>12.76</td>
<td>0</td>
<td>14.14</td>
</tr>
<tr>
<td>Staphylococcus</td>
<td>3.64</td>
<td>4.43</td>
<td>3.18</td>
<td>0.16</td>
<td>8.61</td>
<td>0</td>
<td>8.96</td>
<td>0.58</td>
<td>15.05</td>
<td>0.38</td>
<td>1.73</td>
<td>3.3</td>
<td>1.18</td>
<td>1.2</td>
<td>0.65</td>
<td>5.79</td>
<td>3.01</td>
</tr>
<tr>
<td>Alloccoccus</td>
<td>13.51</td>
<td>0</td>
<td>18.02</td>
<td>4.61</td>
<td>14.79</td>
<td>53.65</td>
<td>7.46</td>
<td>17.28</td>
<td>20.43</td>
<td>0.5</td>
<td>0.58</td>
<td>2.34</td>
<td>60.82</td>
<td>1.93</td>
<td>6.58</td>
<td>6.3</td>
<td>0.84</td>
</tr>
<tr>
<td>Rhizobium</td>
<td>6.51</td>
<td>5.01</td>
<td>1.77</td>
<td>5.33</td>
<td>2.84</td>
<td>0</td>
<td>0.5</td>
<td>16.12</td>
<td>0</td>
<td>20.4</td>
<td>16.12</td>
<td>2.51</td>
<td>0</td>
<td>10.23</td>
<td>21.03</td>
<td>5.79</td>
<td>1.84</td>
</tr>
<tr>
<td>Delftia</td>
<td>19.97</td>
<td>27.55</td>
<td>27.56</td>
<td>44.87</td>
<td>2.03</td>
<td>1.22</td>
<td>0</td>
<td>12.23</td>
<td>0</td>
<td>73.39</td>
<td>72.18</td>
<td>0</td>
<td>0.16</td>
<td>29.36</td>
<td>18.91</td>
<td>10.03</td>
<td></td>
</tr>
<tr>
<td>Escherichia-Shigella</td>
<td>5.47</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87.57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Haemophilus</td>
<td>4.21</td>
<td>9.44</td>
<td>0</td>
<td>2.01</td>
<td>8</td>
<td>3.77</td>
<td>3.98</td>
<td>0.97</td>
<td>15.05</td>
<td>1.57</td>
<td>0.91</td>
<td>0.08</td>
<td>0.31</td>
<td>19.01</td>
<td>1.16</td>
<td>1.02</td>
<td>0</td>
</tr>
<tr>
<td>Acinetobacter</td>
<td>1.26</td>
<td>0</td>
<td>0</td>
<td>2.17</td>
<td>11.04</td>
<td>0.85</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.14</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Morazella</td>
<td>3.66</td>
<td>2.89</td>
<td>0.35</td>
<td>3.74</td>
<td>3.75</td>
<td>17.4</td>
<td>2.49</td>
<td>3.88</td>
<td>12.9</td>
<td>0.81</td>
<td>0.66</td>
<td>1.44</td>
<td>2.12</td>
<td>0.96</td>
<td>3.87</td>
<td>1.36</td>
<td>0</td>
</tr>
<tr>
<td>Pseudomonas</td>
<td>14.52</td>
<td>0.19</td>
<td>18.02</td>
<td>0</td>
<td>6.89</td>
<td>2.43</td>
<td>11.44</td>
<td>32.62</td>
<td>7.53</td>
<td>0.15</td>
<td>3.6</td>
<td>14.03</td>
<td>16.61</td>
<td>41.81</td>
<td>44.12</td>
<td>32.94</td>
<td></td>
</tr>
<tr>
<td>Other (45 other taxa)</td>
<td>4.5</td>
<td>7.13</td>
<td>4.59</td>
<td>1.74</td>
<td>3.85</td>
<td>13.5</td>
<td>2.49</td>
<td>4.08</td>
<td>15.05</td>
<td>2.72</td>
<td>0.99</td>
<td>0.68</td>
<td>2.98</td>
<td>0.84</td>
<td>4.9</td>
<td>2.21</td>
<td>4.18</td>
</tr>
</tbody>
</table>
A.6 Supplementary File 5

Procrustes analysis of raw and rarefied datasets.
Figure A.2: Procrustes analysis of raw and rarefied datasets. The rarefied dataset was subsampled at a threshold of 1499 reads per sample. The raw dataset excluded samples below this depth. P-values are non-parametric and are based on 999 Monte Carlo simulations.
A.7 Supplementary File 6

Beta diversity PCoA in the nasopharynx of cases and controls, sorted by other covariates.
Figure A.3: Beta diversity PCoA in the nasopharynx of cases and controls, sorted by other covariates. Case and control nasopharyngeal samples shown in Figure 2.2 are coloured by other covariates. NA refers to samples where the covariate was not applicable or was missing (not given or recorded “unknown”) and the number represents individual samples. (Continued on following page).
A.8 Supplementary File 7

Complete list of significantly differentially abundant OTUs determined by metagenome-Seq.

All differentially abundant OTUs between a) MEF and MEF; b) MEF and NPS; c) MER and NPS; d) ECS and MEF are shown with their log fold change, $p$-values and mean and median abundance. OTUs in bold/grey are those above the selected threshold of at least 0.35% mean or median abundance in at least one group, and were present in both sets of samples where applicable. Set 1 refers to the set of samples where the left or right ear was chosen at random; these numbers are reported in results. Set 2 refers to the set of samples where the opposite ear was chosen; these results were only used for validation of the differentially abundant OTUs from Set 1.

This table is too large to print. It can be accessed here:
A.9 Supplementary File 8

Beta diversity PCoA in the samples from the cases, sorted by other covariates.
**Figure A.4**: Beta diversity PCoA in the samples from the cases, sorted by other covariates. Samples from the cases shown in Figure 2.4 are coloured by other covariates. NA refers to samples where the covariate was not applicable or was missing (not given or recorded “unknown”) and the number represents individual samples (multiple samples per child). (Continued on the following pages.)
A.10 Supplementary File 9

Procrustes analysis of left and right ear samples.
Figure A.5: Procrustes analysis of left and right ear samples. The dataset includes both MEF and MER samples in left/right ear pairs from the same child. Samples with less than 1499 reads are excluded. The p-value is non-parametric and is based on 999 Monte Carlo simulations.
A.11 Supplementary File 10

Procrustes analysis of MEF and MER samples.
Figure A.6: Procrustes analysis of MEF and MER samples. The dataset includes pairs of MEF and MER samples from the same ear of the same child. Samples with less than 1499 reads are excluded. The p-value is non-parametric and is based on 999 Monte Carlo simulations.
A.12 Supplementary File 11

All correlations between OTUs determined by SparCC.

Correlations within each sample type are listed in separate sheets. This includes correlations between all possible pairs of OTUs found in the samples. \( P \)-values are non-parametric and were calculated as the proportion of times a correlation coefficient more extreme than the observed correlation coefficient occurred in 100 simulated datasets.

This table is a large Excel spreadsheet which can be accessed here: https://static-content.springer.com/esm/art%3A10.1186%2Fs12866-018-1154-3/MediaObjects/12866_2018_1154_MOESM11_ESM.xlsx
Appendix B

Sequence read counts in the shotgun metagenomics data (Chapter 3)
**Table B.1: Summary of usable metagenomics reads per sample.**

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Raw read pairs</th>
<th>Percentage of trimmed pairs remaining</th>
<th>Percentage (of raw pairs) of trimmed orphans remaining</th>
<th>Total single end reads remaining after removal of human and PhiX sequences (%)</th>
<th>Percentage of usable reads receiving functional annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>7773</td>
<td>62790228</td>
<td>94.8</td>
<td>4</td>
<td>5321099 (4.2%)</td>
<td>46.6</td>
</tr>
<tr>
<td>7836</td>
<td>62549083</td>
<td>93.2</td>
<td>5.3</td>
<td>4601768 (3.7%)</td>
<td>49.7</td>
</tr>
<tr>
<td>7844</td>
<td>64454500</td>
<td>97.1</td>
<td>2.3</td>
<td>3015919 (2.3%)</td>
<td>27.8</td>
</tr>
<tr>
<td>8014</td>
<td>52414876</td>
<td>97.7</td>
<td>1.8</td>
<td>4997086 (4.8%)</td>
<td>43</td>
</tr>
<tr>
<td>8070</td>
<td>65194133</td>
<td>92.2</td>
<td>6.1</td>
<td>8586668 (6.6%)</td>
<td>58.3</td>
</tr>
<tr>
<td>8081</td>
<td>60567229</td>
<td>92</td>
<td>6.3</td>
<td>3928930 (3.2%)</td>
<td>41.2</td>
</tr>
<tr>
<td>8086</td>
<td>59821232</td>
<td>94.7</td>
<td>4.2</td>
<td>2183377 (1.8%)</td>
<td>25.8</td>
</tr>
<tr>
<td>8101</td>
<td>67650164</td>
<td>96</td>
<td>3.2</td>
<td>4513863 (3.3%)</td>
<td>48.1</td>
</tr>
<tr>
<td>8117</td>
<td>59709939</td>
<td>96.3</td>
<td>3</td>
<td>4091599 (3.4%)</td>
<td>42.9</td>
</tr>
<tr>
<td>8118</td>
<td>64938692</td>
<td>93</td>
<td>5.5</td>
<td>2731477 (2.1%)</td>
<td>36.7</td>
</tr>
<tr>
<td>MGO012FR</td>
<td>61908641</td>
<td>93.5</td>
<td>5.1</td>
<td>1653863 (1.3%)</td>
<td>29.9</td>
</tr>
<tr>
<td>MGO012S1</td>
<td>62425876</td>
<td>94.2</td>
<td>4.5</td>
<td>2677828 (2.1%)</td>
<td>42.9</td>
</tr>
<tr>
<td>MGO019FL</td>
<td>64436305</td>
<td>95.6</td>
<td>3.5</td>
<td>1908472 (1.5%)</td>
<td>30.7</td>
</tr>
<tr>
<td>MGO019S1</td>
<td>59509870</td>
<td>95.7</td>
<td>3.4</td>
<td>2439353 (2%)</td>
<td>38.6</td>
</tr>
<tr>
<td>MGO020FL</td>
<td>62720714</td>
<td>93.3</td>
<td>5.2</td>
<td>1391594 (1.1%)</td>
<td>24</td>
</tr>
<tr>
<td>MGO020S1</td>
<td>59413661</td>
<td>96.7</td>
<td>2.6</td>
<td>3675287 (3.1%)</td>
<td>42.5</td>
</tr>
<tr>
<td>MGO038FL</td>
<td>65494274</td>
<td>95.2</td>
<td>3.7</td>
<td>1704328 (1.3%)</td>
<td>25.6</td>
</tr>
<tr>
<td>MGO038S1</td>
<td>62585924</td>
<td>97.3</td>
<td>2.2</td>
<td>7474267 (6%)</td>
<td>53.1</td>
</tr>
<tr>
<td>MGO063FR</td>
<td>66545018</td>
<td>95.2</td>
<td>3.7</td>
<td>2082478 (1.6%)</td>
<td>27.9</td>
</tr>
<tr>
<td>MGO063S1</td>
<td>66434127</td>
<td>95.9</td>
<td>3.3</td>
<td>6007496 (4.5%)</td>
<td>53.8</td>
</tr>
<tr>
<td>MGO065FR</td>
<td>65005991</td>
<td>89.2</td>
<td>8.5</td>
<td>1238226 (1%)</td>
<td>10</td>
</tr>
<tr>
<td>MGO065S1</td>
<td>57395736</td>
<td>96.3</td>
<td>3</td>
<td>13642913 (11.9%)</td>
<td>62.7</td>
</tr>
<tr>
<td>MGO071FL</td>
<td>67711561</td>
<td>96.1</td>
<td>3.2</td>
<td>2248183 (1.7%)</td>
<td>33.7</td>
</tr>
<tr>
<td>MGO071S1</td>
<td>54346038</td>
<td>96.9</td>
<td>2.4</td>
<td>2117507 (1.9%)</td>
<td>27.4</td>
</tr>
<tr>
<td>MGO091FR</td>
<td>62471031</td>
<td>94.7</td>
<td>3.9</td>
<td>2299532 (1.8%)</td>
<td>30</td>
</tr>
<tr>
<td>MGO091S1</td>
<td>56367537</td>
<td>96.8</td>
<td>2.5</td>
<td>5558033 (4.9%)</td>
<td>48.7</td>
</tr>
<tr>
<td>MGO094FL</td>
<td>67103058</td>
<td>96.2</td>
<td>3.1</td>
<td>2181470 (1.6%)</td>
<td>34.8</td>
</tr>
<tr>
<td>MGO094S1</td>
<td>66831624</td>
<td>96.1</td>
<td>3.1</td>
<td>2306797 (1.7%)</td>
<td>25.4</td>
</tr>
<tr>
<td>Mean</td>
<td>62457038</td>
<td>95.1</td>
<td>3.9</td>
<td>3806404 (3.1%)</td>
<td>37.9</td>
</tr>
</tbody>
</table>
Appendix C

List of hypoxia-related genes
(Chapter 5)
Table C.1: Human genes analysed in Chapter 5 as part of the hypoxia pathway (n = 296).

<table>
<thead>
<tr>
<th>Ensembl gene ID</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG00000166501</td>
<td>PRKCB</td>
</tr>
<tr>
<td>ENSG0000002330</td>
<td>BAD</td>
</tr>
<tr>
<td>ENSG00000115758</td>
<td>ODC1</td>
</tr>
<tr>
<td>ENSG00000124181</td>
<td>PLCG1</td>
</tr>
<tr>
<td>ENSG00000121022</td>
<td>COPS5</td>
</tr>
<tr>
<td>ENSG00000165197</td>
<td>FIGF</td>
</tr>
<tr>
<td>ENSG00000120910</td>
<td>PPP3CC</td>
</tr>
<tr>
<td>ENSG00000175054</td>
<td>ATR</td>
</tr>
<tr>
<td>ENSG00000135018</td>
<td>UBQLN1</td>
</tr>
<tr>
<td>ENSG00000104812</td>
<td>GYS1</td>
</tr>
<tr>
<td>ENSG00000114023</td>
<td>FAM162A</td>
</tr>
<tr>
<td>ENSG00000095015</td>
<td>MAP3K1</td>
</tr>
<tr>
<td>ENSG00000100968</td>
<td>NFATC4</td>
</tr>
<tr>
<td>ENSG00000136492</td>
<td>BRIP1</td>
</tr>
<tr>
<td>ENSG00000198836</td>
<td>OPA1</td>
</tr>
<tr>
<td>ENSG00000171303</td>
<td>KCNK3</td>
</tr>
<tr>
<td>ENSG00000197299</td>
<td>BLM</td>
</tr>
<tr>
<td>ENSG00000078237</td>
<td>TIGAR</td>
</tr>
<tr>
<td>ENSG00000101152</td>
<td>DNAJC5</td>
</tr>
<tr>
<td>ENSG00000133639</td>
<td>BTG1</td>
</tr>
<tr>
<td>ENSG00000059804</td>
<td>SLC2A3</td>
</tr>
<tr>
<td>ENSG00000170425</td>
<td>ADORA2B</td>
</tr>
<tr>
<td>ENSG00000265972</td>
<td>TXNIP</td>
</tr>
<tr>
<td>ENSG00000197930</td>
<td>ERO1A</td>
</tr>
<tr>
<td>ENSG00000118689</td>
<td>FOXO3</td>
</tr>
<tr>
<td>ENSG00000104765</td>
<td>BNIP3L</td>
</tr>
<tr>
<td>ENSG00000156711</td>
<td>MAPK13</td>
</tr>
<tr>
<td>ENSG00000173511</td>
<td>VEGFB</td>
</tr>
<tr>
<td>ENSG00000116016</td>
<td>EPAS1</td>
</tr>
<tr>
<td>ENSG00000240972</td>
<td>MIF</td>
</tr>
<tr>
<td>ENSG00000240583</td>
<td>AQP1</td>
</tr>
<tr>
<td>ENSG00000101280</td>
<td>ANGPT4</td>
</tr>
<tr>
<td>Ensembl gene ID</td>
<td>Gene name</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>ENSG00000138305</td>
<td>PLA2G12B</td>
</tr>
<tr>
<td>ENSG00000084676</td>
<td>NCOA1</td>
</tr>
<tr>
<td>ENSG00000119630</td>
<td>PGF</td>
</tr>
<tr>
<td>ENSG00000178473</td>
<td>UCN3</td>
</tr>
<tr>
<td>ENSG00000181061</td>
<td>HIGD1A</td>
</tr>
<tr>
<td>ENSG00000227231</td>
<td>IER3</td>
</tr>
<tr>
<td>ENSG00000127472</td>
<td>PLA2G5</td>
</tr>
<tr>
<td>ENSG00000167772</td>
<td>ANGPTL4</td>
</tr>
<tr>
<td>ENSG00000206478</td>
<td>IER3</td>
</tr>
<tr>
<td>ENSG00000121879</td>
<td>PIK3CA</td>
</tr>
<tr>
<td>ENSG00000169032</td>
<td>MAP2K1</td>
</tr>
<tr>
<td>ENSG00000159399</td>
<td>HK2</td>
</tr>
<tr>
<td>ENSG00000051382</td>
<td>PIK3CB</td>
</tr>
<tr>
<td>ENSG00000102755</td>
<td>FLT1</td>
</tr>
<tr>
<td>ENSG00000149925</td>
<td>ALDOA</td>
</tr>
<tr>
<td>ENSG00000130427</td>
<td>EPO</td>
</tr>
<tr>
<td>ENSG00000069667</td>
<td>RORA</td>
</tr>
<tr>
<td>ENSG00000114268</td>
<td>PFKFB4</td>
</tr>
<tr>
<td>ENSG00000181856</td>
<td>SLC2A4</td>
</tr>
<tr>
<td>ENSG00000276870</td>
<td>PLA2G10</td>
</tr>
<tr>
<td>ENSG00000072736</td>
<td>NFATC3</td>
</tr>
<tr>
<td>ENSG00000116044</td>
<td>NFE2L2</td>
</tr>
<tr>
<td>ENSG00000106638</td>
<td>TBL2</td>
</tr>
<tr>
<td>ENSG00000184381</td>
<td>PLA2G6</td>
</tr>
<tr>
<td>ENSG00000137672</td>
<td>TRPC6</td>
</tr>
<tr>
<td>ENSG00000170315</td>
<td>UBB</td>
</tr>
<tr>
<td>ENSG00000103363</td>
<td>TCEB2</td>
</tr>
<tr>
<td>ENSG00000128340</td>
<td>RAC2</td>
</tr>
<tr>
<td>ENSG00000131196</td>
<td>NFATC1</td>
</tr>
<tr>
<td>ENSG00000105698</td>
<td>USF2</td>
</tr>
<tr>
<td>ENSG00000185650</td>
<td>ZFP36L1</td>
</tr>
<tr>
<td>ENSG00000108094</td>
<td>CUL2</td>
</tr>
<tr>
<td>ENSG00000276536</td>
<td>HRAS</td>
</tr>
</tbody>
</table>
Table C.1: (continued)

<table>
<thead>
<tr>
<th>Ensembl gene ID</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG00000100335</td>
<td>MIEF1</td>
</tr>
<tr>
<td>ENSG00000064601</td>
<td>CTSA</td>
</tr>
<tr>
<td>ENSG00000188130</td>
<td>MAPK12</td>
</tr>
<tr>
<td>ENSG00000127928</td>
<td>GNGT1</td>
</tr>
<tr>
<td>ENSG00000137331</td>
<td>IER3</td>
</tr>
<tr>
<td>ENSG00000078369</td>
<td>GNB1</td>
</tr>
<tr>
<td>ENSG00000072274</td>
<td>TFRC</td>
</tr>
<tr>
<td>ENSG00000102908</td>
<td>NFAT5</td>
</tr>
<tr>
<td>ENSG00000116329</td>
<td>OPRD1</td>
</tr>
<tr>
<td>ENSG00000100644</td>
<td>HIF1A</td>
</tr>
<tr>
<td>ENSG00000171791</td>
<td>BCL2</td>
</tr>
<tr>
<td>ENSG00000136238</td>
<td>RAC1</td>
</tr>
<tr>
<td>ENSG00000116711</td>
<td>PLA2G4A</td>
</tr>
<tr>
<td>ENSG00000152455</td>
<td>SUV39H2</td>
</tr>
<tr>
<td>ENSG00000105894</td>
<td>PTN</td>
</tr>
<tr>
<td>ENSG00000100292</td>
<td>HMOX1</td>
</tr>
<tr>
<td>ENSG00000276701</td>
<td>MIF</td>
</tr>
<tr>
<td>ENSG00000100387</td>
<td>RBX1</td>
</tr>
<tr>
<td>ENSG00000137193</td>
<td>PIM1</td>
</tr>
<tr>
<td>ENSG00000120738</td>
<td>EGR1</td>
</tr>
<tr>
<td>ENSG00000087250</td>
<td>MT3</td>
</tr>
<tr>
<td>ENSG00000176171</td>
<td>BNIP3</td>
</tr>
<tr>
<td>ENSG00000167034</td>
<td>NKX3-1</td>
</tr>
<tr>
<td>ENSG00000111669</td>
<td>TPI1</td>
</tr>
<tr>
<td>ENSG00000198625</td>
<td>MDM4</td>
</tr>
<tr>
<td>ENSG00000027869</td>
<td>SH2D2A</td>
</tr>
<tr>
<td>ENSG00000105373</td>
<td>GLTSCR2</td>
</tr>
<tr>
<td>ENSG00000138764</td>
<td>CCNG2</td>
</tr>
<tr>
<td>ENSG00000142208</td>
<td>AKT1</td>
</tr>
<tr>
<td>ENSG00000174775</td>
<td>HRAS</td>
</tr>
<tr>
<td>ENSG00000063176</td>
<td>SPHK2</td>
</tr>
<tr>
<td>ENSG00000165240</td>
<td>ATP7A</td>
</tr>
<tr>
<td>ENSG00000113739</td>
<td>STC2</td>
</tr>
<tr>
<td>Ensembl gene ID</td>
<td>Gene name</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>ENSG00000122884</td>
<td>P4HA1</td>
</tr>
<tr>
<td>ENSG00000176170</td>
<td>SPHK1</td>
</tr>
<tr>
<td>ENSG0000013585</td>
<td>VDAC1</td>
</tr>
<tr>
<td>ENSG00000114738</td>
<td>MAPKAPK3</td>
</tr>
<tr>
<td>ENSG00000105835</td>
<td>NAMPT</td>
</tr>
<tr>
<td>ENSG00000167136</td>
<td>ENDOG</td>
</tr>
<tr>
<td>ENSG00000141506</td>
<td>PIK3R5</td>
</tr>
<tr>
<td>ENSG00000105647</td>
<td>PIK3R2</td>
</tr>
<tr>
<td>ENSG00000107758</td>
<td>PPP3CB</td>
</tr>
<tr>
<td>ENSG00000078401</td>
<td>EDN1</td>
</tr>
<tr>
<td>ENSG00000132356</td>
<td>PRKAA1</td>
</tr>
<tr>
<td>ENSG00000101076</td>
<td>HNF4A</td>
</tr>
<tr>
<td>ENSG00000067225</td>
<td>PKM</td>
</tr>
<tr>
<td>ENSG00000184216</td>
<td>IRAK1</td>
</tr>
<tr>
<td>ENSG00000154582</td>
<td>TCEB1</td>
</tr>
<tr>
<td>ENSG00000164442</td>
<td>CITED2</td>
</tr>
<tr>
<td>ENSG00000126218</td>
<td>F10</td>
</tr>
<tr>
<td>ENSG00000171862</td>
<td>PTEN</td>
</tr>
<tr>
<td>ENSG00000127483</td>
<td>HP1BP3</td>
</tr>
<tr>
<td>ENSG00000141510</td>
<td>TP53</td>
</tr>
<tr>
<td>ENSG00000214575</td>
<td>CPEB1</td>
</tr>
<tr>
<td>ENSG00000100985</td>
<td>MMP9</td>
</tr>
<tr>
<td>ENSG00000132155</td>
<td>RAF1</td>
</tr>
<tr>
<td>ENSG00000168209</td>
<td>DDIT4</td>
</tr>
<tr>
<td>ENSG00000145936</td>
<td>KCNMB1</td>
</tr>
<tr>
<td>ENSG00000073756</td>
<td>PTGS2</td>
</tr>
<tr>
<td>ENSG00000141448</td>
<td>GATA6</td>
</tr>
<tr>
<td>ENSG00000134057</td>
<td>CCNB1</td>
</tr>
<tr>
<td>ENSG00000213281</td>
<td>NRAS</td>
</tr>
<tr>
<td>ENSG00000138814</td>
<td>PPP3CA</td>
</tr>
<tr>
<td>ENSG00000144791</td>
<td>LIMD1</td>
</tr>
<tr>
<td>ENSG00000099250</td>
<td>NRP1</td>
</tr>
<tr>
<td>ENSG00000112759</td>
<td>SLC29A1</td>
</tr>
<tr>
<td>Ensembl gene ID</td>
<td>Gene name</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ENSG00000102144</td>
<td>PGK1</td>
</tr>
<tr>
<td>ENSG00000119950</td>
<td>MXI1</td>
</tr>
<tr>
<td>ENSG00000160307</td>
<td>S100B</td>
</tr>
<tr>
<td>ENSG00000106211</td>
<td>HSPB1</td>
</tr>
<tr>
<td>ENSG00000100393</td>
<td>EP300</td>
</tr>
<tr>
<td>ENSG00000070495</td>
<td>JMJD6</td>
</tr>
<tr>
<td>ENSG00000156273</td>
<td>BACH1</td>
</tr>
<tr>
<td>ENSG00000169398</td>
<td>PTK2</td>
</tr>
<tr>
<td>ENSG00000164867</td>
<td>NOS3</td>
</tr>
<tr>
<td>ENSG00000168214</td>
<td>RBPJ</td>
</tr>
<tr>
<td>ENSG00000132906</td>
<td>CASP9</td>
</tr>
<tr>
<td>ENSG00000277445</td>
<td>CPEB1</td>
</tr>
<tr>
<td>ENSG00000170890</td>
<td>PLA2G1B</td>
</tr>
<tr>
<td>ENSG00000182718</td>
<td>ANXA2</td>
</tr>
<tr>
<td>ENSG00000117020</td>
<td>AKT3</td>
</tr>
<tr>
<td>ENSG00000105220</td>
<td>GPI</td>
</tr>
<tr>
<td>ENSG00000107159</td>
<td>CA9</td>
</tr>
<tr>
<td>ENSG00000164362</td>
<td>TERT</td>
</tr>
<tr>
<td>ENSG00000131981</td>
<td>LGALS3</td>
</tr>
<tr>
<td>ENSG00000183023</td>
<td>SLC8A1</td>
</tr>
<tr>
<td>ENSG00000105851</td>
<td>PIK3CG</td>
</tr>
<tr>
<td>ENSG00000128052</td>
<td>KDR</td>
</tr>
<tr>
<td>ENSG00000187840</td>
<td>EIF4EBP1</td>
</tr>
<tr>
<td>ENSG00000090020</td>
<td>SLC9A1</td>
</tr>
<tr>
<td>ENSG00000106366</td>
<td>SERPINE1</td>
</tr>
<tr>
<td>ENSG00000150630</td>
<td>VEGFC</td>
</tr>
<tr>
<td>ENSG0000003402</td>
<td>CFLAR</td>
</tr>
<tr>
<td>ENSG00000101412</td>
<td>E2F1</td>
</tr>
<tr>
<td>ENSG00000126583</td>
<td>PRKCG</td>
</tr>
<tr>
<td>ENSG00000152256</td>
<td>PDK1</td>
</tr>
<tr>
<td>ENSG00000089159</td>
<td>PXN</td>
</tr>
<tr>
<td>ENSG00000165272</td>
<td>AQP3</td>
</tr>
<tr>
<td>ENSG00000183207</td>
<td>RUVBL2</td>
</tr>
<tr>
<td>Ensembl gene ID</td>
<td>Gene name</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ENSG000000090339</td>
<td>ICAM1</td>
</tr>
<tr>
<td>ENSG000000173376</td>
<td>NDNF</td>
</tr>
<tr>
<td>ENSG00000067057</td>
<td>PFKP</td>
</tr>
<tr>
<td>ENSG00000105221</td>
<td>AKT2</td>
</tr>
<tr>
<td>ENSG00000105327</td>
<td>BBC3</td>
</tr>
<tr>
<td>ENSG00000101144</td>
<td>BMP7</td>
</tr>
<tr>
<td>ENSG00000185386</td>
<td>MAPK11</td>
</tr>
<tr>
<td>ENSG00000124253</td>
<td>PCK1</td>
</tr>
<tr>
<td>ENSG00000173531</td>
<td>MST1</td>
</tr>
<tr>
<td>ENSG00000122861</td>
<td>PLAU</td>
</tr>
<tr>
<td>ENSG00000103266</td>
<td>STUB1</td>
</tr>
<tr>
<td>ENSG00000089163</td>
<td>SIRT4</td>
</tr>
<tr>
<td>ENSG00000137449</td>
<td>CPEB2</td>
</tr>
<tr>
<td>ENSG00000170345</td>
<td>FOS</td>
</tr>
<tr>
<td>ENSG00000179094</td>
<td>PER1</td>
</tr>
<tr>
<td>ENSG00000134333</td>
<td>LDHA</td>
</tr>
<tr>
<td>ENSG00000154229</td>
<td>PRKCA</td>
</tr>
<tr>
<td>ENSG00000129521</td>
<td>EGLN3</td>
</tr>
<tr>
<td>ENSG00000221823</td>
<td>PPP3R1</td>
</tr>
<tr>
<td>ENSG00000112062</td>
<td>MAPK14</td>
</tr>
<tr>
<td>ENSG00000129946</td>
<td>SHC2</td>
</tr>
<tr>
<td>ENSG00000151651</td>
<td>ADAM8</td>
</tr>
<tr>
<td>ENSG00000123739</td>
<td>PLA2G12A</td>
</tr>
<tr>
<td>ENSG00000101096</td>
<td>NFATC2</td>
</tr>
<tr>
<td>ENSG00000117394</td>
<td>SLC2A1</td>
</tr>
<tr>
<td>ENSG00000067992</td>
<td>PDK3</td>
</tr>
<tr>
<td>ENSG00000068903</td>
<td>SIRT2</td>
</tr>
<tr>
<td>ENSG00000185624</td>
<td>P4HB</td>
</tr>
<tr>
<td>ENSG00000119772</td>
<td>DNMT3A</td>
</tr>
<tr>
<td>ENSG00000037280</td>
<td>FLT4</td>
</tr>
<tr>
<td>ENSG00000131055</td>
<td>COX4I2</td>
</tr>
<tr>
<td>ENSG00000135766</td>
<td>EGLN1</td>
</tr>
<tr>
<td>ENSG00000148926</td>
<td>ADM</td>
</tr>
</tbody>
</table>
### Table C.1: (continued)

<table>
<thead>
<tr>
<th>Ensembl gene ID</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG00000122691</td>
<td>TWIST1</td>
</tr>
<tr>
<td>ENSG00000102882</td>
<td>MAPK3</td>
</tr>
<tr>
<td>ENSG00000177885</td>
<td>GRB2</td>
</tr>
<tr>
<td>ENSG00000109819</td>
<td>PPARGC1A</td>
</tr>
<tr>
<td>ENSG00000102760</td>
<td>RGCC</td>
</tr>
<tr>
<td>ENSG00000186352</td>
<td>ANKRD37</td>
</tr>
<tr>
<td>ENSG00000124440</td>
<td>HIF3A</td>
</tr>
<tr>
<td>ENSG00000163586</td>
<td>FABP1</td>
</tr>
<tr>
<td>ENSG00000278139</td>
<td>PIK3R3</td>
</tr>
<tr>
<td>ENSG00000275199</td>
<td>AKT3</td>
</tr>
<tr>
<td>ENSG00000198793</td>
<td>MTOR</td>
</tr>
<tr>
<td>ENSG00000126934</td>
<td>MAP2K2</td>
</tr>
<tr>
<td>ENSG00000146674</td>
<td>IGFBP3</td>
</tr>
<tr>
<td>ENSG00000167315</td>
<td>ACAA2</td>
</tr>
<tr>
<td>ENSG00000197122</td>
<td>SRC</td>
</tr>
<tr>
<td>ENSG00000118257</td>
<td>NRP2</td>
</tr>
<tr>
<td>ENSG00000117461</td>
<td>PIK3R3</td>
</tr>
<tr>
<td>ENSG0000016490</td>
<td>CLCA1</td>
</tr>
<tr>
<td>ENSG00000165730</td>
<td>STOX1</td>
</tr>
<tr>
<td>ENSG00000124212</td>
<td>PTGIS</td>
</tr>
<tr>
<td>ENSG00000082482</td>
<td>KCNK2</td>
</tr>
<tr>
<td>ENSG00000184408</td>
<td>KCND2</td>
</tr>
<tr>
<td>ENSG00000105974</td>
<td>CAV1</td>
</tr>
<tr>
<td>ENSG00000069764</td>
<td>PLA2G10</td>
</tr>
<tr>
<td>ENSG00000113083</td>
<td>LOX</td>
</tr>
<tr>
<td>ENSG00000158828</td>
<td>PINK1</td>
</tr>
<tr>
<td>ENSG00000165168</td>
<td>CYBB</td>
</tr>
<tr>
<td>ENSG00000135245</td>
<td>HILPDA</td>
</tr>
<tr>
<td>ENSG00000117215</td>
<td>PLA2G2D</td>
</tr>
<tr>
<td>ENSG00000188784</td>
<td>PLA2G2E</td>
</tr>
<tr>
<td>ENSG00000074800</td>
<td>ENO1</td>
</tr>
<tr>
<td>ENSG00000230128</td>
<td>IER3</td>
</tr>
<tr>
<td>ENSG00000136381</td>
<td>IREB2</td>
</tr>
<tr>
<td>Ensembl gene ID</td>
<td>Gene name</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ENSG00000158786</td>
<td>PLA2G2F</td>
</tr>
<tr>
<td>ENSG00000112715</td>
<td>VEGFA</td>
</tr>
<tr>
<td>ENSG00000221983</td>
<td>UBA52</td>
</tr>
<tr>
<td>ENSG00000235030</td>
<td>IER3</td>
</tr>
<tr>
<td>ENSG00000282019</td>
<td>GPI</td>
</tr>
<tr>
<td>ENSG00000141526</td>
<td>SLC16A3</td>
</tr>
<tr>
<td>ENSG00000243708</td>
<td>PLA2G4B</td>
</tr>
<tr>
<td>ENSG00000150991</td>
<td>UBC</td>
</tr>
<tr>
<td>ENSG00000096717</td>
<td>SIRT1</td>
</tr>
<tr>
<td>ENSG00000080824</td>
<td>HSP90AA1</td>
</tr>
<tr>
<td>ENSG00000145431</td>
<td>PDGFC</td>
</tr>
<tr>
<td>ENSG00000100823</td>
<td>APEX1</td>
</tr>
<tr>
<td>ENSG00000036828</td>
<td>CASR</td>
</tr>
<tr>
<td>ENSG00000166135</td>
<td>HIF1AN</td>
</tr>
<tr>
<td>ENSG00000269858</td>
<td>EGLN2</td>
</tr>
<tr>
<td>ENSG00000104419</td>
<td>NDRG1</td>
</tr>
<tr>
<td>ENSG00000104332</td>
<td>SFRP1</td>
</tr>
<tr>
<td>ENSG00000197943</td>
<td>PLCG2</td>
</tr>
<tr>
<td>ENSG00000143947</td>
<td>RPS27A</td>
</tr>
<tr>
<td>ENSG00000064393</td>
<td>HIPK2</td>
</tr>
<tr>
<td>ENSG00000170525</td>
<td>PFKFB3</td>
</tr>
<tr>
<td>ENSG00000159167</td>
<td>STC1</td>
</tr>
<tr>
<td>ENSG00000112033</td>
<td>PPARD</td>
</tr>
<tr>
<td>ENSG00000148677</td>
<td>ANKRD1</td>
</tr>
<tr>
<td>ENSG00000143437</td>
<td>ARNT</td>
</tr>
<tr>
<td>ENSG00000100678</td>
<td>SLC8A3</td>
</tr>
<tr>
<td>ENSG00000117525</td>
<td>F3</td>
</tr>
<tr>
<td>ENSG00000155816</td>
<td>FMN2</td>
</tr>
<tr>
<td>ENSG00000171608</td>
<td>PIK3CD</td>
</tr>
<tr>
<td>ENSG00000188386</td>
<td>PPP3R2</td>
</tr>
<tr>
<td>ENSG00000117400</td>
<td>MPL</td>
</tr>
<tr>
<td>ENSG00000274334</td>
<td>VASN</td>
</tr>
<tr>
<td>ENSG00000145386</td>
<td>CCNA2</td>
</tr>
</tbody>
</table>
### Table C.1: (continued)

<table>
<thead>
<tr>
<th>Ensembl gene ID</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG00000134107</td>
<td>BHLHE40</td>
</tr>
<tr>
<td>ENSG00000070831</td>
<td>CDC42</td>
</tr>
<tr>
<td>ENSG00000168140</td>
<td>VASN</td>
</tr>
<tr>
<td>ENSG00000100030</td>
<td>MAPK1</td>
</tr>
<tr>
<td>ENSG00000100078</td>
<td>PLA2G3</td>
</tr>
<tr>
<td>ENSG00000171314</td>
<td>PGAM1</td>
</tr>
<tr>
<td>ENSG00000109320</td>
<td>NFKB1</td>
</tr>
<tr>
<td>ENSG00000135679</td>
<td>MDM2</td>
</tr>
<tr>
<td>ENSG00000204397</td>
<td>CARD16</td>
</tr>
<tr>
<td>ENSG00000162889</td>
<td>MAPKAPK2</td>
</tr>
<tr>
<td>ENSG00000145675</td>
<td>PIK3R1</td>
</tr>
<tr>
<td>ENSG00000141279</td>
<td>NPEPPS</td>
</tr>
<tr>
<td>ENSG00000133703</td>
<td>KRAS</td>
</tr>
<tr>
<td>ENSG00000148400</td>
<td>NOTCH1</td>
</tr>
<tr>
<td>ENSG00000141959</td>
<td>PFKL</td>
</tr>
<tr>
<td>ENSG00000237155</td>
<td>IER3</td>
</tr>
<tr>
<td>ENSG00000188257</td>
<td>PLA2G2A</td>
</tr>
<tr>
<td>ENSG00000149428</td>
<td>HYOU1</td>
</tr>
<tr>
<td>ENSG00000105976</td>
<td>MET</td>
</tr>
<tr>
<td>ENSG00000134086</td>
<td>VHL</td>
</tr>
<tr>
<td>ENSG00000136997</td>
<td>MYC</td>
</tr>
<tr>
<td>ENSG00000171132</td>
<td>PRKCE</td>
</tr>
<tr>
<td>ENSG00000173846</td>
<td>PLK3</td>
</tr>
<tr>
<td>ENSG00000114480</td>
<td>GBE1</td>
</tr>
<tr>
<td>ENSG00000137463</td>
<td>MGARP</td>
</tr>
<tr>
<td>ENSG00000160789</td>
<td>LMNA</td>
</tr>
<tr>
<td>ENSG0000005339</td>
<td>CREBBP</td>
</tr>
<tr>
<td>ENSG00000101945</td>
<td>SUV39H1</td>
</tr>
<tr>
<td>ENSG00000141682</td>
<td>PMAIP1</td>
</tr>
<tr>
<td>ENSG00000280682</td>
<td>HYOU1</td>
</tr>
<tr>
<td>ENSG00000283094</td>
<td>PTEN</td>
</tr>
<tr>
<td>ENSG00000283100</td>
<td>SIRT2</td>
</tr>
<tr>
<td>ENSG00000283087</td>
<td>CPEB2</td>
</tr>
</tbody>
</table>