A Study of Infant Nasal Microbiome Diversity Applying *cpn60* Universal Target (UT) as a Phylogenetic Marker

By

By Shelley W. Peterson

A Thesis Submitted to the

Faculty of Graduate Studies at the University of Manitoba

In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

Department of Medical Microbiology and Infectious Diseases

University of Manitoba

Winnipeg, Manitoba

Abstract

A protocol was developed to study the nasal microbiome applying a *cpn60* universal target (UT) amplicon as a bacterial phylotyping marker using next generation sequencing. The UT can be used to identify organisms in metagenomic samples to the species level. This study aimed to determine the typical composition of anterior nares microbiota for developing infants over time, and to explore diversity of the infant nasal microflora and the relation to microflora observed within their primary caregivers.

Nares swabs were collected from 40 two-week old infants and their primary caregivers over a 1-year period using a single swab in one nare for each subject. These samples were treated to diminish contaminating human DNA, and bacterial template DNA was augmented using a whole-genome linear augmentation procedure. The *cpn60* UT target was PCR amplified using degenerate primers and sequenced using high throughput 454/GS FLX pyrosequencing. Reads were quality-filtered and matched to the *cpn60* database using an 80% nucleotide identity cutoff. Diversity and metagenomics analyses were performed using *mothur* v1.27.0 and SIMCA v. 13.

Throughout the first year of life, infant nares microflora increased in diversity. Infants showed differential representation of organisms compared with their primary caregivers. The nares microbiota predominantly belong to the phyla Actinobacteria, Firmicutes, and Proteobacteria. Individuals appear to have only one or two predominating genera; however intra-subject variability was noted. High variability between study subjects and within study subjects over time was also observed.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AOM	Acute Otitis Media
bp	Base Pair(s)
BLASTn	Basic Local Alignment Search Tool nucleotides
CA-	Community-Associated
Cpn60 UT	Chaperonin protein-60 Universal Target (region)
dATP	Deoxyadenosine Triphosphate
dCTP	Deoxycytidine Triphosphate
dGTP	Deoxyguanosine Triphosphate
dTTP	Deoxythymidine Triphosphate
dH_2O	Distilled water
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
emPCR	Emulsion Polymerase Chain Reaction
G+C	G and C content of DNA
HIV	Human Immunodeficiency Virus
HMP	Human Microbiome Project
IgA	Immunoglobulin A
IgM	Immunoglobulin M
IPD	Invasive Pneumococcal Disease
Kb	Kilobase pairs

kDa	kiloDalton
M	Molar
$MgCl_2$	Magnesium Chloride
$MgSO_4$	Magnesium Sulfate
MID	Multiplex Identifier
µL	Microlitre
mL	Millilitre
µM	Micromolar
mM	Millimolar
MPC	Magnetic Particle Concentrator
MRSA	Methicillin-resistant *Staphylococcus aureus*
NP	Nasopharyngeal
NTC	Negative Template Control
OMP	Otitis Media Pathogen
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PCV	Pneumococcal Conjugate Vaccine
PLS-DA	Partial Least Squares Discriminate Analysis
PPB	Potentially Pathogenic Bacteria
PTP	PicoTiterPlate
RNA	Ribonucleic Acid
16S rRNA	16S Subunit of ribosomal Ribonucleic Acid
RT-PCR	Real Time Polymerase Chain Reaction

SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SPRI	Solid Phase Reverse Immobilization
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
UV	Ultraviolet

Glossary

Abundance – Percentage of an OTU in a sample.

Alpha diversity – Mean phylogenetic diversity in sites or habitats at a local scale

Beta diversity – Phylogenetic differentiation among habitats.

Chao richness index – An abundance-based index measuring alpha diversity, meant to calculate the estimated true species diversity of a sample.

Microbiome – The community of microorganisms that live within the human body.

Morisita horn diversity index – An abundance-based index measuring beta diversity that is represented by a value between 0 and 1, with 1 representing identical communities. This index is resistant to the effects of undersampling.

Shannon diversity – Diversity index that measures both the richness (number of different taxa/OTUs) and the evenness (relative abundance of taxa/OTUs).

Sobs – Observed species richness, the number of different taxa/OTUs found in a sample.

Ubiquity – The percentage of samples that possess an OTU.

Acknowledgements

First of all, I'd like to thank my committee members: Dr. Michael Mulvey, Dr. Morag Graham, Dr. Gary Van Domselaar and Dr. Joanne Embree for providing guidance throughout the extent of my degree. I was very fortunate in having a committee that was varied in expertise and each member was integral to my project, each lending help in their particular area. I'd also like to extend my thanks to Dr. Charles Bernstein who agreed at the last minute to be a member of my examining committee. Thanks to Dr. Fiona Fleming and the staff at Manitoba clinic for providing me with nares swabs and providing participants with questionnaires. A big thanks to Natalie Knox who helped guide me through the early stages of data analysis. I'd also like to thank the staff of Nosocomial Infections laboratory for all their support and advice. I'd particularly like to thank Megan Rea, my fellow Masters student for the shared understanding and jokes (it's ALWAYS awkward).

I'd also like to thank my wonderful family and friends for their support. My parents, - thanks for letting me live rent-free and supporting me through everything. Thanks to the Distinguished Ladies and Gentlemen of the Teddy Bob's Crew for all the ridiculous Friday nights and all the amazing times; I think we all somehow managed to keep each other sane. And to the rest of my friends who listened to my complaints through all the ups and downs of the last few years.

1. Introduction

1.1 Human Microflora

There are approximately 10^{14} bacterial cells living within the human body, while it is estimated that the body itself is composed of 10^{13} cells. The organisms that naturally live on or within the human body are referred to as indigenous flora or microbiota. These flora consist of bacteria, fungi and viruses that live in a commensal relationship with their human host. Currently, there are believed to be over 1000 different species of commensal bacteria in the human body [2]. Most commensal bacteria are located in the intestinal tract, but they can also be found in the upper airways, skin, and vaginal mucosa [3,4]. The composition of the microbiome varies with the age of the patient, and is also affected by underlying disease conditions and administration of antibiotics or other drugs [5].

The human microflora interact in a commensal relationship with the host to form a complex ecosystem that is usually beneficial to both [2]. These bacteria are considered to be beneficial to the host by protecting against the colonization of invading pathogens [5-7]. Protection is conferred by competing for resources, secretion of substances that are inhibitory to pathogenic bacteria, and activation of the host's immune system against potentially pathogenic bacteria [8,9]. The intestinal microbiota facilitate metabolism of some glycans, amino acids, and xenobiotics that the human body cannot metabolize [10]. The indigenous microbiota also aid in tissue development, drug metabolism and immune system development, along with providing a physical barrier against pathogenic bacteria [10].

1.2 Disruption of Human Microflora

Occasionally the composition of the microbiota is disrupted, often due to viral infections, infections by non-resident bacteria, or antibiotic use [9]. This can cause a normally commensal bacterial community to undergo a disruptive imbalance, and opportunistic infections can develop [2]. Removal of these species reduces bacterial competition and allows pathogenic species to proliferate and cause disease [3]. Examples of this include pseudomembranous colitis in the large intestine caused by toxigenic strains of *Clostridium difficile* and bacterial vaginosis caused by an increase of Gram-negative organisms [3,11]. Both of these infections occur due to significant reduction in one or more microbial species within these areas, usually due to antibiotic use. Often, disease status is not represented by the growth or introduction of a specific new organism, but a shift in the ratio of each species present [12].

1.3 Microbiome Studies

After the completion of the Human Genome Project, various researchers claimed that the study of the human genome was incomplete without the genetic material of the vast numbers of microorganisms that also colonize the human body [3]. The term "microbiome" was coined by Joshua Lederberg in 2001 to represent the community of microorganisms that live within the human body, and in a way, make up part of the human body [13]. Our bodies can be viewed as the sum of all of the *Homo sapiens* genes contained within our genome as well as those contained within the genomes of our microbial occupants; thus we can be considered a "supraorganism" [14,15]. When viewing the human body in this manner, our metabolic processes can include the metabolic ability of our microbiome as well [14].

Only about 20-60% of the bacteria in the human microbiome have been described through culture techniques, limiting the number of species that can be characterized and studied phenotypically [2,3]. In fact, an estimate of up to 80% of the human intestinal microflora are currently non-cultivable [3]. The cultivable flora have been highly characterized and well-studied, especially those that are considered pathogenic, but little is known about those that are not-yet cultivable.

With the development of massively parallel pyrosequencing technology, the focus on the microbial community has changed from the phenotypic study of single organisms to the entire genomes and transcriptomes of whole communities, termed *metagenomics* [15]. Through advancements in technology, the field of metagenomics is expanding, allowing for studies of complex microbial communities [10]. Also, advances in computer technology have increased the amount of data that can be gathered and analyzed, facilitating large-scale metagenomic studies [10].

1.3.1 New Sequencing Technologies

Prior to the advent of massively parallel pyrosequencing, metagenomic analyses were carried out using cloning and Sanger-based capillary sequencing. This was both tedious and expensive, which limited the number and scale of metagenomic projects that could be performed [16]. Pyrosequencing is a newer approach allowing parallel in-depth analysis of many samples [16]. Pyrosequencing is a sequencing-by-synthesis approach in which each of the four bases are independently flowed across single DNA molecules statically attached to hundreds of thousands of beads. Each bead contains a clonally amplified template sequence prepared from genomic DNA fragments, or amplicons. When a base complementary to the

bead-bound template molecule is washed across a bead, it is incorporated into the elongating nascent strand and light is emitted in proportion to the number of nucleotides added. This light is recorded in a flowgram and used to determine the DNA sequence [17]. Pyrosequencing is more cost effective and high throughput than Sanger sequencing; however, read lengths are shorter and pyrosequencing has difficulty in discerning long homopolymers (mononucleotide runs), often leading to sequencing errors (base miscalling generating insertion-deletion (indel) errors).

Most metagenomic studies identify organisms by sequencing the 16S rRNA genes and comparing them to online sequence database collections of known organisms. This gene in particular was chosen due to its manageable size (~1.5kb), presence in all organisms, and rare incidence of horizontal gene transfer; as well as being highly conserved, yet variable enough to distinguish between organisms [14]. However, 16S rRNA has limited ability to differentiate organisms beyond the genus or species level. For example, in a study by Aas *et al., Streptococcus mitis* and *Streptococcus pneumoniae* could not be differentiated using 16S rRNA [18]. Also, 16S rRNA cannot be used to distinguish between certain species of *Bordetella* and *Neisseria* [19]. In addition, organisms can carry multiple copies of 16S rRNA and copy number can vary within a species, making absolute quantitation difficult. For these reasons, alternative gene targets are often used to differentiate between or quantitate species [20].

Cpn60 (type I chaperonins, GroEL) is a possible alternate gene target for species identification [21]. *Cpn60* is a 60kDa chaperonin that aids in posttranslational protein folding and assembly of protein complexes [20]. It is universal in prokaryotic organisms, with the exception of some intracellular obligate organisms, which appear to lack other "essential"

prokaryotic genes as well [22]. *Cpn60* can also be found in eukaryotes and the Archaeal genus *Methanosarcina*, consisting of homo-14-mers in bacteria and mitochondria and heteromers in chloroplasts [23]. *Cpn60* is a useful target for phylogenetic studies and identification of organisms in part due to its universal target (UT) region [20]. This 549-567 bp region is amplified by universal degenerate PCR primers corresponding to nucleotides 274-288 of the *E. coli cpn60* sequence [24]. This region has sufficient resolution to discriminate closely related bacterial species and is more discriminating and phylogenetically informative than 16S rRNA [20]. For example, Jian *et al*. determined that *cpn60* could be used to distinguish between several species of *Bifidobacterium* that are indistinguishable using 16S rRNA [25]. The variability of *cpn60* is distributed uniformly along the target molecule, allowing for more sequence variation than 16S rRNA [20]. *Cpn60* is also useful for quantitative assays and studies of complex microbial populations, as there is usually only one copy of the gene per prokaryotic cell [26,27]. A reference sequence database (designated cpnDB; http://www.cpndb.ca) is available for the *cpn60* gene, albeit not as large as those databases available for 16S rRNA [20].

1.3.2 Human Microbiome Project

The Human Microbiome Project (HMP) was implemented to identify the members of the human microbiota and the factors that influence the distribution and evolution of human commensal organisms [15] . The results of the project may also provide information on modern human evolution – the effect of technology and change in lifestyle on the "micro-evolution" of humans and disease progression [15] . The human microbiome project aims to describe the

"core" microbiome shared by the majority of humans, as well as the extent of variability surrounding it. This can be determined at the genus or species level.

An undertaking such as the HMP comes with a variety of challenges. For instance, sequencing technology has improved much more quickly than the bioinformatics technology required to analyze the large amounts of data produced in metagenomics studies [28]. 454 pyrosequencing was initially designed for use in whole-genome shotgun sequencing in which there is a large amount of redundancy and overlap of reads, reducing the importance of sequencing errors [28]. Microbiome studies analyze reads individually, requiring development of stringent error checking methods. In addition to error checking, new methods of analyzing large amounts of multivariable data were needed for analysis of data generated from the HMP [28]. Because of this, many bioinformatics tools were developed in order to complete the HMP, and many required extensive computing power for their use [28].

Other microbiome project challenges include: 1) determining which body sites to sample; 2) determining how many participants should be recruited for the study; 3) creating a database in order to identify organisms; and 4) developing laboratory protocols for sampling and processing the samples [3,15]. Additional considerations are as follows: Different depth of information is acquired when deep sequencing is performed on a fewer number of individuals in many body sites, *vs*. a large number of individuals in a few body sites [15]. Generating a large number of reads per sample requires extensive computing power for analysis; meanwhile generating a lower number of reads per sample risks incomplete sampling of the entire diversity. In addition, a study is not effective in capturing the full diversity of a sample if organisms cannot be identified due to a diversity-limited database. Because of this, extensive databases were created to represent the vast diversity of organisms

within the human microbiome [3]. Finally, a method of sampling and processing samples based on body site needed to be developed to ensure the most accurate and reproducible data. Some body sites pose problems such as low microbial concentrations leading to limited amounts of DNA available for PCR or large amounts of human DNA contamination. Linear whole genome amplification and human DNA removal techniques have been developed to deal with these issues.

1.4 Normal Flora of the Upper Respiratory Tract

The anterior nares share characteristics with both the outer skin and the nasopharynx, and consequently, they share characteristics of both microbiomes [29]. The outer lining of the nostrils consists of keratinized squamous epithelium containing sebaceous glands, sweat glands and hairs more similar to that of outer skin than the ciliated columnar epithelium of the nasopharynx [7]. Mucus from the nasopharynx drains into the nasal cavity, carrying nasopharyngeal bacteria along with it [29]. Thus, human nasal microflora are very similar to those of the nasopharynx, with some exceptions. For example, there is an abundance of *Propionibacterium acnes*, which is a common member of the skin microflora, in the nares, which is not commonly present in the nasopharynx [30]. The nares favour more diverse bacterial species than the nasopharynx, as they are constantly being exposed to different environments through inhalation [29] and can come into contact with skin microflora on the fingers. The nares are mainly inhabited by Actinobacteria and Firmicutes, similar to that of skin, but they have a lower incidence of Proteobacteria and Bacteroidetes [31].

The normal flora of the human nares and nasopharynx have not been studied extensively, except in the case of potentially pathogenic organisms [7]. However, it is known

that the normal flora of the nasopharynx vary significantly between age groups [7]. Throughout a person's life, the composition of the microbiota changes, as bacteria colonize, disappear, and re-colonize the upper airways. For example, an immature immune system and increased close contacts may contribute to young children exhibiting different microflora predominating from those observed in adults [32]. In addition, as an individual progresses from infant to adulthood, physiological changes occur in skin, including increased presence of sebum and skin surface lipids, which can affect microbial growth within the anterior nares [30], as well as in the physical structure of the nose. Sections 1.5 and 1.6 outline roles of the upper respiratory tract microbiota and factors affecting the microbial composition.

1.5 Role of Microflora in the Upper Respiratory Tract

The commensal microflora can affect the human body both locally, by competition with potentially pathogenic organisms, as well as systemically, by activating the immune system [2]. The organisms within the nasal cavity are responsible for many immunological functions such as early priming of the immune system; providing protective immunity; mediating autoreactivity and immunological tolerance; and balancing allergic sensitization [7]. These functions are believed to be affected by endotoxins produced by Gram-negative bacteria, and IgA1 protease produced by several species of bacteria that commonly colonize the upper respiratory tract [7]. In particular, IgA1 proteinase cleaves human IgA1 in the hinge region, which can influence presentation of inhaled antigens, changing the reaction generated by host innate defenses [7].

The normal flora in the nasopharynx is an important defence against infection. These organisms both compete with, and inhibit, colonization and proliferation of invading

pathogens [9]. Alpha-haemolytic streptococci provide much of this protection, by inhibiting growth of other potentially pathogenic bacteria [9]. In a study by Johanson *et al.*, it was determined that about 90% of the inhibitors of potentially pathogenic bacteria were α-haemolytic streptococci [33]. These organisms produce inhibitory substances similar to bacteriocins, peptides that inhibit growth of, or kill, other bacteria—including potentially pathogenic bacteria [9]. Alternatively, other organisms in the nasopharynx can compete with potentially pathogenic bacteria for host binding sites and adhesion molecules, thereby preventing them from attachment and colonization of the anterior nares [9].

The nares microflora are very important for establishing and regulating the immune system at the epithelial barrier [34]. Ichyma *et al.* showed that germ-free mice contained fewer immunoglobulin IgM B cells and $CD4^+$ T cells in the upper respiratory tract than pathogen-free conventional mice [34]. Through this study, it was determined that the nasal microflora stimulate the nasal mucosa to recruit B and T cells to the area, increasing the immune response to invading pathogens [34].

1.6 Factors that Affect the Composition of the Upper Respiratory Tract

1.6.1 Age

1.6.1.1 Nasal/Nasopharyngeal Microflora in Infants

Before birth, the infant resides within the sterile environment of the uterus [2]. The infant first encounters microorganisms upon entry into the vaginal canal during birth. Colonization with microflora occurs within the first few hours of life through respiration, ingestion, and contact of skin to foreign surfaces. The upper respiratory tract comes into contact with large numbers

of aerobic bacteria through respiration [2]. Anaerobes can enter the nasopharynx via aspiration of saliva [6]. Most of these bacteria will only colonize the body briefly; however, some will remain on the body for long periods of time, or even for life [35]. Each different body site is colonized by a different community of bacteria owing to various factors; including nutrient availability, oxygen tension, host immune response, access to binding receptors, chemotherapeutic agents and pH. The infant nares microbiome may be affected by colonization status of the mother and close contacts, as well as a difference in anatomy between the adult and infant nasal cavity and skin. [12]. The nasopharyngeal flora is generally established in the first year of life [36].

One of the relatively few studies focusing on infant nasopharyngeal microbiota was performed by Ostfeld *et al.* in 1983 [35]. In this culture-based study of 132 newborn infants, 39% of infants carried Gram-positive flora, 23% carried potentially pathogenic Gram-negative cocci, 16% contained Gram-negative enteric rods and 22% of cultures did not show growth. The most frequent bacteria found were *Staphylococcus epidermidis* (39%), *Staphylococcus aureus* (11%), *Escherichia coli* (8%), viridans streptococci (8%) and group B streptococci (2%) [35]. Colonization of Gram-negative organisms was more common in infants that were hospitalized for more than three days. This could be due to nasotracheal intubation or use of nasogastric feeding tubes [35]. *S. aureus* and group B streptococci were found only in infants delivered by vaginal birth, suggesting vaginal transmission of these microbes [35]. In 2002, a culture based study by Kononen *et al.*, found anaerobes in 15 out of 220 infants sampled. The most common anaerobic organisms in the infant nasopharynx were *Fusobacterium nucleatum* and *Prevotella melaninogenica* [37].

In a recent 16S rRNA sequence-based metagenomic study documenting the microbial composition of the nares and nasopharynx of healthy 18 month-old infants, the 5 most predominant phyla in the nasopharynx were Proteobacteria (64%), Firmicutes (21%), Bacteroides (11%), Actinobacteria (3%) and Fusobacteria (1.4%) [38]. Other organisms found in the nasopharynx included *Cyanobacteria*, *Deinococcus-Thermus*, *Nitrospira, Planctomycetes, Chloroflexi* and the candidate phyla OD1, TM7, BRC1 (only identified through 16S rRNA sequencing). The most commonly found genera were *Moraxella* (40%), *Haemophilus* (20%), *Streptococcus* (12%), and *Flavobacterium* (10%). *Dolosigranulum, Corynebacterium, Neisseria* and *Fusobacterium* were also found, but less commonly [38]. Bogaert *et al.* also tested the nares of these infants, determining that roughly 80% of detected organisms were Gram-positive, consisting mostly of Actinobacteria and Firmicutes [38]. The nares samples were highly variable with each subject, and a limited number of common organisms could be said to make up a nares core microbiome. Bogaert *et al.* defined their core microbiome as the operational taxonomic units (OTUs) representing >0.1% of total sequences generated and present in all 96 infants [38]. Only two core phyla were observed (Firmicutes and Proteobacteria) and no single OTU was found in all sampled children [38]. However, *Moraxella, Haemophilus influenzae, Enhydrobacter, Streptococcus, Dolosigranulum* and *Corynebacterium* were found in over 50% of the infants [38].

The carriage rate of otitis media pathogens (OMP) is higher in infants and children than in adults [39]. OMP organisms include *Streptococcus pneumoniae, Moraxella catarrhalis* and *Haemophilus influenzae. S. pneumoniae* carriage increases during the first few months of life and then begins to decrease at about age 3-5 years [40]. In infants under one year of age, *S. pneumoniae* and *M. catarrhalis* can be identified up to 54% and 72% of the time,

respectively [36]. In one study, approximately 34% of infants became colonized with non-typeable *H. influenzae* within the first year of life, and up to 44% were colonized by 2 years of age [36]. On average, children first acquire at least one OMP by 6 months of age, with a range of 1-30 months [36]. By one year of age, 50-100% of infants are generally found to be colonized by at least one OMP [36]. The carriage rate of another potentially pathogenic bacterial species, *S. aureus*, is very high in the anterior nares within the first few weeks of life, but then colonization decreases significantly by the age of six months [41].

1.6.1.2 Nasal/Nasopharyngeal Microflora in Children

In a culture-based study by Tano *et al.* non-pathogenic bacteria represented about 50% of the total nasopharyngeal flora in healthy children aged 3-9. Most of these were α-haemolytic streptococci [9]. Alpha-haemolytic streptococci are predominantly responsible for inhibition of pathogenic bacteria. This is done by competitive adhesion to the upper respiratory tract cells and diffusion of inhibitory substances such as bacteriocins and bacteriocin-like inhibitory substances [9,42]. In this study, 90% of healthy children harboured potential OMPs in the nasopharynx and *H. influenzae* was the most common potential OMP found [9]. In a study of healthy Swedish school children, 27% were nasal carriers of *M. catarrhalis*, 29% carried *S. pneumoniae* and 13% carried *H. influenzae*, which are all potential OMPs [43]. The most common types of non-pathogenic bacteria found in the nasopharynx of children are *Streptococcus oralis*, *S. sanguis*, *S. mitis*, *S. salivarius*, and *S. intermedius* [9].

In a culture-based study of nasopharyngeal flora in healthy Japanese children ages six and under, 52.9% carried *Corynebacterium sp.*, 25.7% carried viridans group streptococci, 22.9% carried *S. aureus*, and 14.3% carried coagulase-negative staphylococci [39]. Less

abundant organisms included *Enterobacteriaceae*, *Haemophilus sp.*, glucose-nonfermenting bacteria, and other *Streptococcus* spp. *M. catarrhalis, S. pneumoniae* and *H. influenzae* were found in 35.7%, 47.1% and 55.7% of children, respectively [39].

S. pneumoniae represents a potentially pathogenic member of the normal flora, and is commonly found in children but not adults [7,44]. *S. aureus,* which is a common resident of infants, is found in only 10-35% of children [44]. This could possibly be due to an inverse (competitive) relationship between *S. aureus* and *S. pneumoniae* carriage in children [44]. However, individuals under 20 years of age tend to have higher rates of persistent *S. aureus* carriage than adults. In children from 0-9 years, the persistent *S. aureus* carriage rate is 10%, while it is 24% in adolescents from 10-19 years of age [45]. Children under 4 years are more likely to carry *M. catarrhalis* than older children [32].

1.6.1.3 Nasal/Nasopharyngeal Microflora in Adults

The adult nasopharynx is dominated by staphylococci, streptococci, *Corynebacterium* and Gram-negative cocci [2]. The central nasal cavity contains very limited microbiota dominated by diphtheroids, staphylococci, and other Gram-positive bacteria [7]. Viridans streptococci, *Haemophilus* species, as well as Gram-negative species that are characteristic of the oropharynx are quite rare [7]. Common aerobic genera include *Staphylococcus, Corynebacterium, Stomatococcus, Micrococcus, Streptococcus, Lactobacillus* and *Mycoplasma.* Anaerobic genera that are commonly found in the nasopharynx include *Veillonella, Peptostreptococcus, Fusobacterium, Porphyromonas, Bacteroides, Prevotella, Actinomyces, Lactobacillus, Bifidobacterium,* and *Propionibacterium* [7,44].

In a study of 10 healthy adult volunteers, *Corynebacterium* was the most common genus found in the nasopharynx, with *C. propinquum* and *C. accolens* being the most common species [7]. Staphylococci were second most prevalent with *S. epidermidis, S. capitis, S. hominis, S. haemolyticus, S. lugdunensis* and *S. warneri* being isolated most often. *S. aureus* was only found in a single individual likely due to low sampling numbers [7]. *Aureobacterium spp., Rhodococcus spp., Stomatococcus mucilaginosus, Micrococcus luteus* and *Gemella spp.* were also found in smaller proportions [7].

A high percentage of adults contain opportunistic pathogens, mainly *S. pneumoniae, S. aureus* and *H. influenzae*, within the nasopharynx [2]. Many of these organisms can cause pneumonia or meningitis, as well as less severe diseases such as otitis, sinusitis or pharyngitis [2]. In a study of 534 male clerical worker subjects 24 to 45 years old in Lucerne, Switzerland, 77% of subjects carried potentially pathogenic bacteria in the nasopharynx, of which *S. aureus* was found most frequently [43]. Other nasopharyngeal opportunistic pathogens include *Corynebacterium pseudodiphtherium, Corynebacterium accolens, Aureobacterium spp.* and *Rhodococcus spp.* [7].

In a metagenomic study of nares samples from healthy adults, the majority of sequences were obtained from only two different phyla: Actinobacteria (68%) and Firmicutes (27%) [46]. Other common phyla included Proteobacteria, Bacteroidetes, Fusobacteria, Cyanobacteria, Tenericutes, and Deinococcus. *P. acnes* was found in all subjects, and *S. epidermidis* was also prevalent [46].

1.6.2 Sex

The sex of the individual can predispose a person to carriage of certain organisms. For example, Caugant *et al*. determined that males are more than twice as likely to be carriers of *N. meningitidis* than females [47]. Also, among those working outside of the home, this male/female ratio was increased when compared with the population as a whole. In a separate study, researchers showed that amongst Dutch infants, boys have a significantly higher risk of testing positive for *S. aureus* carriage than girls [45].

In contrast, Harrison *et al*. discovered that the rate of microbial colonization (number of individuals carrying a specific organism) of the nasopharynx in infants was highly similar between boys and girls with the exception of a few genera. He determined that female infants were significantly more likely to carry *Corynebacterium pseudodiphtheriticum,* and *Enterobacteriaceae* than males. Male infants were more likely to carry β-hemolytic streptococci. They also found that although the number of infants carrying *S. aureus* did not differ between males and females, it was found that males had a much higher overall bacterial load than females [48].

1.6.3 Socioeconomic Status

Many infections, such as community-associated, methicillin-resistant *S. aureus* (CA-MRSA) are found more often in people of low socioeconomic status for a variety of reasons, such as limited access to health care, crowded living conditions, lack of indoor plumbing, inadequate nutrition, and inadequate shelter [49,50]. In a study of children in India, it was determined that living in a mud-thatched house increased the chance of *S. aureus* colonization [50]. This enhanced carriage could be due to the aforementioned socioeconomic factors, along with

poor ventilation. Intravenous drug use, prior antibiotic usage and contact with a health care facility also increase the chance of *S. aureus* nasal colonization [50,51].

N. meningitidis is another potential human pathogen of which the only known reservoir is the human upper respiratory tract [47]. Most patients who develop this disease have not come into contact with another infected patient [47]. During times of endemic disease, carriage can reach over 50% in semi-secluded populations, such as military recruits. During epidemics, the carriage rate can be up to 95% [47]. In suburban areas, people of low economic status are three times more likely to become carriers of *N. meningitidis* than upper middle class families. Caugant *et al*. hypothesized that carriage and transmission are associated with number of social contacts, as there is a marked decrease of carriage in the elderly, who exhibit reduced social activity [47]. People who work at home are less likely to be carriers of *N. meningitidis* compared with those who work outside the home [47]. Also, individuals with university level education, and those under 18 years whose parents have completed junior high school were less likely to be carriers than those with less education [47]. People who are single, widowed, divorced or separated are more likely to be carriers than those who are married [47].

A study by Jourdain *et al.* determined that children living in low economic status households were more likely to carry *M. catarrhalis* and *S. aureus* than children with higher economic status [32]. Conversely, *S. pneumoniae* and *H. influenzae* carriage did not appear to be affected by socioeconomic status [32]. Children living in homes containing multiple children were more likely to carry *S. pneumoniae* and less likely to carry *M. catarrhalis* [32]. *H. influenzae* and *S. pneumonia*e have been found in 65.4% and 55.7% of children living in a crowded household, respectively, compared with 5.8% and 10.8%, respectively in a more

spacious household with higher economic status. Coles *et al*. have discovered that nasopharyngeal colonization by upper respiratory pathogens is higher in children who have parents with low educational levels [36]. Disadvantaged populations, such as aboriginal children, have increased nasopharyngeal and nasal colonization with *S. pneumoniae* and *S. aureus* due to overcrowded living conditions, lack of running water, high carriage rates, and circulation of multiple strains [36,49].

1.6.4 Underlying Disease

Underlying disease conditions and overall health status often have impacts on bacterial colonization; for example, diseases such as diabetes mellitus, or immunodeficiency disorders such as HIV [50]. Presence of chronic obstructive pulmonary disease, obesity, immunosuppression, allergy, or acute sinusitis have been shown to affect carriage of pathogens in the nasopharynx [36]. Individuals with sickle cell anemia or allergies are more likely to carry *S. pneumoniae*; and HIV patients are at higher risk for invasive pneumococcal disease [36]. Cases of acute upper respiratory tract infection caused by normal flora such as *S. pneumoniae, H. influenzae,* and *M. catarrhalis* are more common in patients with allergies, especially in people between the ages of 16 and 39 [52]. Patients between the ages of 60 and 74 with malignant intranasal tumors or people with diabetes are also more likely to develop upper respiratory tract infections [52]. Carriage of *N. meningitidis* is more likely when a patient has undergone tonsillectomy [47].

Infections with non-resident microflora can also have an effect on the composition of the flora within the upper airways. Bacterial infections in the upper respiratory tract are often preceded by viral infections such as influenza, and carriage of acute otitis media (AOM)

pathogens is more common during acute respiratory illness [6]. During viral infections, the environment of the nasopharynx changes owing to increased nasal secretions and temperature (fever), which can cause increased growth of anaerobes [6]. Anaerobes are rarely found in the nasopharynx of a healthy individual; however, they can enter through aspiration of saliva from the oropharynx [6]. The nasopharyngeal environment of a healthy individual will not support extensive growth of these organisms; however, growth and secondary infection can occur in a virus-infected individual or during convalescence from viral infection [6]. For example, Harrison *et al*. discovered that infants with viral upper respiratory infections were more likely to carry *M. catarrhalis* and α- or non-hemolytic streptococci than healthy infants. In contrast, infants without upper respiratory tract infection were more likely to carry *S. epidermidis* [48].

Seriously ill, hospitalized, and/or elderly patients are more likely to carry Gram-negative organisms, especially facultative bacilli [5]. In a study by Frank *et al*. of nares microbiota from inpatients and healthy adults, inpatients carried significantly more Firmicutes (*S. aureus* and *S. epidermidis*), and fewer Actinobacteria (*P. acnes*) [46]. The healthy subject microbiota were significantly more diverse, with approximately twice as many species-level OTUs than those of inpatients [46].

1.6.5 Seasonality

Some studies have shown higher nasal carriage of potentially pathogenic organisms during colder months of the year, presumably owing to closer contact with other individuals during the school year and cold winter months, as well as a lack of adequate ventilation, increased incidence of viral infections, and increased antibiotic use [36,53]. This may be because the adult

nasal glands produce an increased amount of nasal secretions during cold months, which may increase the amount of bacterial colonization during these months, or possibly increase detection of bacteria due to increased amount of secretions [53]. Some of the evidence for higher carriage rates comes from increased incidence during acute rhinitis, which occurs more often during colder months. Syrjanen *et al*. noted that *S. pneumoniae* carriage was more common in children under two years old with upper respiratory tract illness than for healthy children [40]. Faden *et al*., also detected increased amounts of the three AOM pathogens in children during times of acute illness [53].

In a study of 16 children followed through at least three seasons of the year, carriage rates of potentially pathogenic bacteria (PPB) were highest in the winter season, with 85% of children carrying one or more PPBs during both illness and wellness. PPB carriage decreased to 56% during the summer. On average, carriage of *M. catarrhalis* was 40%, and *S. pneumoniae* and *H. influenzae* were about 33% in studied individuals [53]. During spring and autumn, children with acute respiratory illnesses were about 10% more likely to carry potential pathogens than healthy children. In winter, the carriage rates for *S. pneumoniae* and *H. influenzae* were 50% in schoolchildren and more than 70% in preschoolers [53]. Harrison *et al*. found that *S. aureus* and *S. epidermidis* are also more common in winter and autumn than during summer in healthy infants [48].

In a metagenomic study of 96 18-month old children by Bogaert *et al*., there were significant differences in the nasopharyngeal microflora during autumn/winter in comparison to spring. Proteobacteria were found in 75% of children sampled, compared with 51% of subjects in spring; Cyanobacteria was detected at an abundance of 64% vs. 30% in spring; and Fusobacteria were found at 14% versus 2% in spring. Bacteroidetes were found more

frequently in spring (91% compared to 54% in autumn/winter) [38]. More *Bacillus, Brevibacillus, Lactobacillus, Flavobacterium* and *Bacteroides fragilis* were found in spring than in autumn/winter. *Bacillus, Brevibacillus* and *Lactobacillus* were found almost exclusively in spring. These organisms have also been associated with increased protection against pathogenic bacteria due to production of bacteriocins [38]. In autumn/winter, α-proteobacteria, Oxalobacteriaceae, Icrobacteriaceae, *Ralstonia, Pseudomonas, Acidovorax,* Cyanobacteria, and *Porphyromonas catoniae* were more abundant than in spring [38].

1.6.6 Antibiotic Usage

Antibiotic usage can disturb the balance of the normal bacterial flora within the upper airways, the digestive tract, and other body sites[9]. This can cause secondary infections to occur, as well as increase incidence of antibiotic resistance [9]. However, antibiotics are still necessary for patients with AOM, as well as those carrying potential pathogens who undergo surgery or suffer from injuries of the head, nasal sinuses or lungs [9,43].

Treatment with antimicrobials reduces carriage rates of *S. pneumoniae* for a brief period of time, about 7 days [40]. This change, however, is only temporary and does not eradicate the organism from the nasopharynx [36]. In the nasopharynx and anterior nares, coagulase-negative staphylococci and diphtheroids prevent growth of *S. aureus,* whereas non-pathogenic *Neisseria* species inhibit colonization by *N. meningitidis* [5]. If the numbers of these organisms are reduced by antibiotic usage, pathogenic organisms will be able to grow and cause disease [5]. Following antibiotic treatment, resistant strains and other organisms that survived the antibiotics can flourish within the nasal cavities, changing the make-up of the microflora either temporarily or for a longer term [36].

1.6.7 Probiotics

Probiotics are defined as live cultures of microorganisms administered orally and acting beneficially on host health [2]. Elie Metchinkoff discovered that the ingestion of bacteria in fermented products are beneficial, as they can compete with potentially pathogenic bacteria [54]. This idea could be used in the maintenance of normal flora by introducing non-pathogenic bacteria to fill the ecological niche usually taken up by commensal organisms, ensuring that potentially pathogenic bacteria cannot colonize the host [55]. Probiotics influence the development and stability of normal microflora, while inhibiting growth of pathogens, as well as influencing the mucosal barrier and stimulating the immune system [2].

In a study by Gluck *et al*., people were given a probiotic drink containing *Lactobacillus* GG, *Streptococcus thermophilus, Lactobacillus acidophilus* and *Bifidobacterium* for three weeks [8]. Within this time period, the occurrence of potentially pathogenic bacteria within the nasal cavity decreased by 19%. This implies that it is possible to reduce the carriage of potentially pathogenic bacteria in the nasal flora with regular intake of probiotics [8]. Ingesting *Lactobacillus* GG improves colonization resistance against harmful bacteria, lowers oxidative enzyme activity, reinforces the mucosal barrier and stimulates immunologic memory [8].

1.6.8 Primary and Second-hand Smoke Exposure

Smoking has been associated with increased carriage of respiratory pathogens by causing inflammation of the respiratory mucosa, which increases susceptibility to bacterial infection [36]. This may be associated with increased carriage of AOM pathogens in adults with chronic respiratory diseases [36]. Exposure to cigarette smoke also is associated with increased carriage

of potentially pathogenic bacteria. For example, smokers are nearly three times more likely to be carriers of *N. meningitidis* than nonsmokers [47]. Exposure to second hand smoke doubles the risk of *N. meningitidis* carriage [47]. Smoke exposure also increases carriage of *M. catarrhalis* in children [32,56]. This is possibly due to enhanced adhesion of bacteria to epithelial cells of the airways of smokers, disruption of nasal mucociliary clearance, or reduction of the host immune response [56,57]. Thus, smokers and those who are chronically exposed to cigarette smoke have increased susceptibility to respiratory tract infections and invasive diseases caused by respiratory pathogens [56].

Smoking is believed to reduce the diversity and number of commensal flora, while allowing potential pathogens to proliferate [57]. In a study of anaerobic bacterial carriage, it was determined that anaerobic organisms are more prevalent in non-smokers than in those exposed to cigarette smoke. It was also shown that these anaerobic bacteria play a role in interfering with colonization within the nasopharynx of pathogenic bacterial species such as *S. pneumoniae* and *H. influenzae* [58]. Smokers carry a less diverse nasopharyngeal flora with increased carriage of pathogens such as *S. pneumoniae, H. influenzae,* and *M. catarrhalis* [57]. Carriage of *Prevotella* and *Peptostreptococcus spp.* is also drastically reduced in smokers, with subsequent reinstatement of these organisms following cessation of smoking [57].

1.7 Potentially Pathogenic Bacteria (PPB)

Pathogens enter the host through respiration and colonize the nasopharyngeal mucosal epithelium [36]. Proteins on their cell walls generally adhere to host cell carbohydrate receptors [36]. There are three general patterns of carriage: removal of the initial colonizing strain, continuous carriage of the initial colonizing strain, and colonization of subsequent different

strains [36]. Carriage of multiple strains of the same species or multiple pathogenic species can occur [36].

There are four main potentially pathogenic members of the nares/nasopharynx normal resident flora. These include *S. aureus* and the three AOM pathogens: *S. pneumoniae, M. catarrhalis* and *H. influenzae*. Carriage of these organisms varies throughout life, with *S. aureus* being found most often in adults and young infants, and the AOM pathogens being more common in children. Studies have shown that early colonization with AOM pathogens can lead to greater incidences of AOM in childhood, and that frequent colonization has been associated with increased episodes of AOM [59]. About 70% of healthy children can be found to carry one or more of these potentially pathogenic bacteria [53]. In a study done by Jourdain *et al.* over a two year period, 83% of children were colonized at least once with *H. influenzae*, 69% with *S. pneumoniae*, 67% with *M. catarrhalis*, and 50% with *S. aureus* [32]. In a Finnish study of nasopharyngeal carriage, most children carried *S. pneumoniae* at least once during the first two years of life. Several studies have also shown this same trending in children sampled in Sweden and Israel [40]. Children with AOM had a colonization rate of 49%, while 9% of healthy children were PPB carriers at 2 months, and 43% at two years [40].

The carriage rate of potentially pathogenic bacteria in the nasopharynx of healthy individuals generally decreases with age [36]. They are found very commonly in children under 7 years of age, and are more rare in people older than 16 [36]. Gunnarsson *et al*. studied a population of pre-school children (<7 years of age), schoolchildren (7-16) and adults (>16) in Sweden and found that *M. catarrhalis* rates are 27%, 4%, and 2%, respectively; S. *pneumoniae* is found in 19%, 6%, and 0.8% of each population; and *H. influenzae*

frequencies were 13%, 6% and 3% [36]. Frequent colonization and early colonization are both associated with development of otitis media [36].

1.7.1 *Staphylococcus aureus*

While *S. aureus* is a commensal organism, it is also an important cause of both hospital and community acquired infections [60]. It is most commonly found in the human nares and on parts of the skin [45]. There are three different types of *S. aureus* nasal carriage: namely *persistent* carriers, defined as individuals who carry the same strain over time; *intermittent* carriers, who are temporarily colonized by different strains; and *non* carriers, representing those who do not test positive for *S. aureus* [60]. Infants aged three months and under have the highest rate of *S. aureus* persistent carriage [45]. Lebon *et al*. determined that *S. aureus* carriage rates decreased significantly from 52.1% at 1.5 months of age to 21.7% at 6 months [45]. In a study by Gries *et al*., 35% of healthy infants became nasal carriers of *S. aureus* by two weeks of age, and the majority also carried *S. aureus* on other parts of the body as well [41]. Thirteen percent of *S. aureus* carriers had a parent that was a carrier of the same strain and 8.6% of the infants in the study were carriers of methicillin-resistant *S. aureus*.

Approximately 30-35% of the adult population carries *S. aureus,* with the nares being the main carriage site [44]. *S. aureus* colonization in the nares is an important risk factor for infection, as nasal carriage triples the chance of becoming bacteremic sometime during one's lifetime [60]. Approximately 20% of adults are persistent carriers, 30% are intermittent and 50% are non carriers [45]. *S. aureus* carriage is inhibited by the presence of *Corynebacterium* or other staphylococci [45].

1.7.2 Acute Otitis Media (AOM) Pathogens

1.7.2.1 *Moraxella catarrhalis*

M. catarrhalis is the first potentially pathogenic organism found to colonize infants, and is found throughout infancy more often than the other AOM pathogens [59]. Some studies have shown that *M. catarrhalis* can be found in up to 50% of children [9,48,59,61]. Carriage patterns are similar to that of *H. influenzae*, with a child acquiring up to three or four different molecular subtypes within the first year of life [36]. Carriage of *M. catarrhalis* begins to decline after a few years of life, until it reaches roughly 5% of sampled adults [9]. During episodes of respiratory illness, carriage of *M. catarrhalis* is increased, and is very high during the first two years of life, with up to 80% of children under 2 being colonized while ill [9].

1.7.2.2 *Streptococcus pneumoniae*

Despite being a member of the normal nasopharyngeal flora, *S. pneumoniae* causes a wide variety of diseases from otitis media to severe invasive pneumococcal disease, which causes meningitis and bacteremia. Colonization generally occurs in the first few years of life [40]. *S. pneumoniae* causes significant morbidity and mortality worldwide, especially in the developing world [62,63]. *S. pneumoniae* causes 40% of all pneumonia deaths in children less than 5 years of age, as well as being the major cause of pneumonia in the elderly. In developing countries, 70,000 deaths occur annually from both pneumococcal meningitis and sepsis. *S. pneumoniae* is also responsible for significant morbidity, including 43-59% of AOM cases each year [62].

There are more than 90 distinct serotypes of *S. pneumoniae*, based on their capsular polysaccharides [36]. Some children never carry this organism, while others can be colonized by up to four different types before they reach 12 months of age [36]. Individuals generally carry *S. pneumoniae* for one to four months at a time, with serotypes that induce weaker immune responses being carried for longer periods of time [36]. Duration of *S. pneumoniae* carriage is also inversely proportional to age, with adults carrying the organism for only about 2 to 4 weeks [36].

In developing countries, *S. pneumoniae* is acquired at a much younger age than in industrialized countries [64]. For example, in Papua New Guinea, infants become colonized with *S. pneumoniae* during the first few weeks of life, with all infants studied found to be colonized by the age of 3 months. In Sweden, infants normally acquire *S. pneumoniae* during the first 6 months of life, with a large number of children remaining uncolonized by their first birthday and peak colonization occuring around their second birthday [64]. In developing countries, approximately 95% of children below the age of three are *S. pneumoniae* carriers, as well as 40% of adults [62].

Individuals who carry viridans streptococci may be less likely to carry *S. pneumoniae*, as viridans streptococci are antagonistic to other streptococci [36]. Alpha-haemolytic streptococci have also been shown to inhibit nasopharyngeal colonization of *S. pneumoniae*, *M. catarrhalis* and *H. influenzae*. It has been shown that colonization with vaccine-type *S. pneumoniae* is inversely associated with *S. aureus* nasal carriage in unvaccinated populations [65].

1.7.2.2.1 Pneumococcal Conjugate Vaccine

Introduction of a 7-valent pneumococcal conjugate vaccine (PCV7), and more recently, a thirteen-valent vaccine (PCV13), has affected *S. pneumoniae* carriage in vaccinated children. This vaccine targets the 7 serotypes (4, 6B, 9V, 14, 18C, 19F and 23F) most commonly associated with invasive pneumococcal disease (IPD) in children and adults in North America [66]. PCV7 is given in three doses in early infancy (2, 4, 6 months of age), with a booster shot given at age two [64].

Since the introduction of the vaccine, IPD caused by PCV7-serotype *S. pneumoniae* in children under 2 years has decreased to almost zero [66,67]. Owing to reduction of *S. pneumoniae* carriage in infants, the incidence of IPD among unvaccinated older children and adults also has decreased [66]. In the USA, efficacy trials showed a 93% reduction in consolidative pneumonia cases, a 7% reduction in AOM and a 20% reduction in ventilatory tube placements [62]. According to one study, the rates of IPD decreased by 32% in adults 20-39 years, 8% in adults 40-64, and 18% in those over 65 years [67]. In another study, there was a 68-70% reduction in IPD rates in children, 42% in adults 18-49 years and 30% in adults over 64 years [67].

The reduction of carriage of common *S. pneumoniae* serotypes may leave a niche for other serotypes or other organisms to colonize [32]. However, in a study by Garcia-Rodriguez *et al.*, PCV7 immunization did not appear to have an effect on the four main nasopharyngeal pathogens [36]. The PCV7 vaccine causes little or no effect on the amount of pneumococcal colonization within the nasopharynx because the non-vaccine-type organisms become more prevalent as the vaccine-type organisms are no longer carried [68]. Several studies in the USA

have found increases in AOM and IPD caused by *S. pneumoniae* serotype 19A. However, it is unclear whether or not this is due to widespread PCV7 use, as it is not universal throughout areas with widespread PCV7 use [66]. Also, serotype 19A *S. pneumoniae* was becoming more prevalent in other countries before introduction of PCV7. Other possible explanations for this increase are antibiotic pressure and capsular switching [66].

One implication of widespread PCV7 vaccination is that reduction of vaccine-type *S. pneumoniae* (the seven most predominant *S. pneumoniae* serotypes found in the vaccine) could lead to an increase in *S. aureus* carriage in the anterior nares of the general population and also an increase in serious disease caused by MRSA [44,65]. Factors that are positively associated with *S. pneumoniae* carriage tend to be negatively associated with *S. aureus* carriage, leading to an inverse relationship [44]. In children, simultaneous carriage of *S. pneumoniae* and *S. aureus* is found significantly less frequently than expected. This relationship could be due to bacterial interference or confounding effects, such as age [44]. *S. pneumoniae* is most commonly found in children under three, while *S. aureus* carriage increases with age. The inverse association between the two pathogens was not found in adults; however, when controlling for age, the association among children persisted. In a study by Regev-Yochay *et al.*, daycare attendance was found to be directly associated with *S. pneumoniae* carriage and inversely related to *S. aureus* carriage [44]. Day care attendance may prolong the carriage of *S. pneumoniae* or another organism could be present that has an inverse relationship with *S. aureus* [44]. There is no conclusive evidence on whether widespread use of PCV7 is causing an increase in *S. aureus* carriage: Lee *et al.*, have found no significant difference in colonization of children three months of age to seven years since

the implementation of the vaccine [65]; conversely Benninger *et al.*, have found an increase in AOM cases caused by *S. aureus* and *M. catarrhalis* post vaccination [67].

1.7.2.3 *Haemophilus influenzae*

Nontypeable (non-encapsulated) *H. influenzae* is one of the major causes of otitis media. It is commonly found in the nasopharynx of children, with up to 60% of children harbouring these bacteria [69]. On average, 41% of children in a study by Dabernat *et al*. in France carried this organism. Adults typically carry only one strain of *H. influenzae*, while children generally carry many different strains simultaneously [36]. Within the first year of life, 50% of colonized children will carry one strain, while the other half will carry up to seven strains sequentially [36]. *H. influenzae* is often carried in the nasopharynx for between two weeks to 5 months at a time.

1.8 Study Objective

Following the completion of the Human Genome Project, the realization occurred that humans are not composed of the human genome alone. As our microbial inhabitants outnumber us 10 to 1, they must have a significant and important impact on our general well-being. Knowledge of the nature of our microbial inhabitants is important, as it can help us gain insight into differences between the microbiome composition in times of health and of disease. If we can determine the composition of the "core microbiome" shared between all humans, it can help us determine disease risk and evaluate general health among at-risk patients, such as those with cystic fibrosis, asthma, and immunocompromised individuals.

An important aspect of understanding the nature of our microbiota is to understand how it develops within the first year of life. Within the womb an infant is normally sterile; however, it becomes colonized with bacteria within the first few hours of life. The infant is exposed to these bacteria through contact with the outside environment, including breathing, eating, and skin-to-skin contact with the mother and other individuals. As the infant grows and their immune system matures, their microbial composition changes as well.

This study aims to determine the composition of the microbiota typically found in the anterior nares of developing infants as well as their primary caregivers. This information will be used to determine the inter-subject and intra-subject diversity of the nares microflora. The intra-subject diversity is compared over five sampling periods ranging from 2 weeks to one year of age. The inter-subject diversity is determined during each sampling period between the infants themselves, between primary caregivers and between the two groups.

1.9 Study Hypotheses

This study has three major but interrelated hypotheses: 1) The composition of the infant nasal microbiome will become more diverse over time, but also more stable as the infant progresses through its first year of life; 2) Intra-subject diversity over different time points will be greater than inter-subject diversity at the same time point, and 3) Infants will also have microbiomes that are more similar to those of other infants than to those of their primary caregivers.

2. Materials and Methods

2.1 Protocol Development

2.1.1 Swabbing and Storage Procedures

Swabbing techniques were tested on two healthy female adult volunteers using either dry swabs or swabs pre-moistened with sterile saline. Volunteers were selected by convenience and provided informed consent. Pre-moistened Copan flocked nylon swabs (Copan Diagnostics Inc., Murrieta USA) caused less nasal discomfort than dry swabs (informal comments). Swabs were immediately suspended in 500 µl Dulbecco's phosphate-buffered saline (Invitrogen, Carlsbad, USA) or chaotropic Oragene®-DNA solution (DNA Genotek, Kanata, Canada) in pre-labelled 1.5 ml LoBind DNA tube (Eppendorf Canada, Missisauga, Canada). Oragene®-DNA solution is a chaotropic buffer that lyses human cells, while bacterial cells remain intact and was chosen due to ease of storage. As it is a nuclease-inhibiting chaotropic solution, it can be stored at room temperature, while saline solution requires freezing to prevent growth of microorganisms.

2.1.2 Template Extraction

Template extraction procedures were developed using swabs sampled from the anterior nares of healthy adult volunteers (described above), as well as serially diluted MRSA sham swabs. Two different template extraction methods were tested: 1) Agencourt GenFind v2 (Beckman Coulter Genomics, Danvers, USA), a solid phase reversible immobilization (SPRI) magnetic bead-based assay (described in Section 2.2.4), and 2) Oragene®-DNA Purifier solution, which is a precipitation-based method (DNA Genotek Inc., Ottawa, Canada) (described in

Appendix A). The GenFind v2 protocol was chosen, as it is plate-based and enables improved batching of templates for enhanced processing Extraction efficiency (in terms of both DNA yield and reproducibility across replicates) was evaluated by quantification of the extracted DNA using Quant-it^{TM} PicoGreen dsDNA assays (Invitrogen, Carlsbad, USA) (described in Section 2.2.10) with a DTX 880 spectrophotometer (Beckman Coulter Genomics, Danvers USA).

Two different pre-treatment methods were tested (relative to no pre-extraction treatment): 1) incubation with 10 µl of lytic enzyme cocktail (comprised of 2 mg/ml Lysostaphin, 12.5 IU/ml Lysozyme, and 200 IU/ml Mutanolysin); and 2) physical disruption (bead beating). To ensure breakdown of bacterial cell membranes, a 4 m incubation at 60 ^{o}C with 1% SDS (v/v) and subsequent vortexing was introduced in between pre-treatment and GenFind v2 purification. 35 mg/ml RNase A was added to the lytic enzyme cocktail to reduce RNA carryover via binding to the high-affinity SPRI beads and thereby increase DNA recovery during the GenFind v2 procedure.

2.1.3 Amplification of *cpn60* Universal Target (*cpn60* UT)

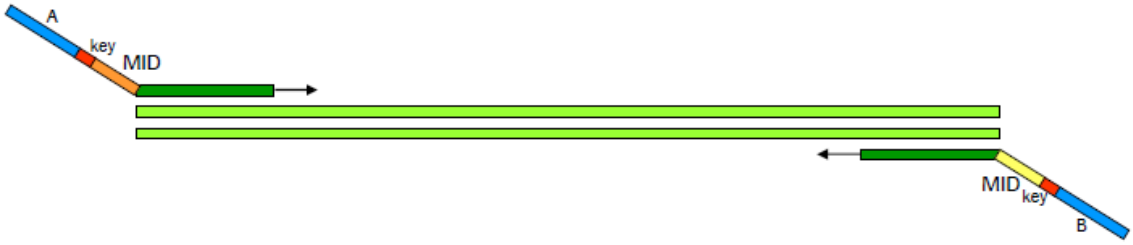
The ~670 bp Universal Target (UT) region of group I chaperonin 60 (*cpn60*) can be used for bacterial identification [20]. Two degenerate primer sets were designed to amplify this region, designated H279/H280 and H1612/H1613 [24]. Primer set H279/H280 was originally designed to amplify *cpn60* UT, and the H1612/H1613 primer set was designed to prevent under-representation of organisms with high G+C content. These primers were adapted to be compatible with 454 amplicon pyrosequencing (LibA emulsion PCR (emPCR) reagents); generating UT fusion primer pairs designated P279/P280 and P1612/P1613. These fusion

primers contain 454 Titanium LibA-specific adaptor A sequences (21nt) linked 5′ to the UT-specific sequences of P279 and P1612, and adaptor B sequences linked to primers P280 and P1613 (Tables 1 and 2, Section 2.2.8). The primers also contain one of the available multiplex identifier (MID) tags to distinguish between samples during pooled 454 pyrosequencing. An example of a fusion primer sequence is depicted in Figure 1. Fusion primers for G+C rich targets and for G+C sparse targets were then mixed in a 1:3 ratio, as per Hill *et al.*, to allow for amplification from both G+C rich and G+C sparse organisms [24]. Following the initial development of the PCR protocol, new fusion primers were created containing 454 Titanium LibL-specific adaptor A sequences, and more MID tags were incorporated to facilitate multiplexing of additional subjects during pyrosequencing. This was necessary to accommodate the experimental design of the sequencing runs as well as to take advantage of the simplified LibL-based workflows.

2.1.3.1 Touchdown PCR

The PCR amplification reactions were originally tested using a touchdown PCR method on each MID-tagged fusion primer set. The reaction was performed using cold bench top aluminum working racks with the following components per reaction (total 25 µl): 2.5 µl of 10X PCR buffer (Applied Biosystems, Inc. Foster City, USA), 1.5 µl of 25 mM $MgCl_2$, 1.25µl of 4 mM deoxynucleoside triphosphates (dATP, dCTP, dTTP, dGTP), 2.5 µl of 10X primer mix (containing 1 µM each of primers P279/P280 and 3 µM each of primers P1612/P1613), 0.25 µl Ampli *Taq* Gold™ DNA polymerase (5 U/µl) (Applied Biosystems, Inc. Foster City, CA) (1.25 U final), 15-15.5 µl of sterile dH_2O and 1.5-2 µl template DNA (~150-200 ng/reaction depending on DNA concentration). PCR amplification was performed using a Veriti® 96-well thermal cycler (Applied Biosystems, Inc. Foster City, USA).

Figure 1: Diagram of a set of fusion primers. The template specific sequence (green) anneals to the template. The MID tag (orange) is located only on the forward primer, and is used as a barcode to distinguish between samples during pooled 454 pyrosequencing. The key (red) is a four base identifier allowing the sequencing software to distinguish between the GS FLX Titanium Primers A or B (blue) sections of the read, and the informative sections of the read (used with permission from Roche) [1].



Forward primer (Primer A-Key):

5’-CGTATCGCCTCCCTCGCGCCATCAG-*{MID}*-*{template-specific-sequence}*-3’

Reverse primer (Primer B-Key):

5’-CTATGCGCCTTGCCAGCCCGCTCAG-*{MID}*-*{template-specific-sequence}*-3’

Cycling conditions were as follows: initial 5 m denaturation at 95 ^{o}C; followed by 15 cycles of 30 s denaturation at 94 ^{o}C, 90 s annealing at 65 ^{o}C, decreasing by 0.8 ^{o}C every subsequent cycle until the annealing temperature reached 44 ^{o}C, and 90 s extension at 68 ^{o}C. An additional 12 cycles were run with a 44 ^{o}C annealing temperature followed by a final extension step at 68 ^{o}C for 10 m. This PCR reaction appeared to be effective for *S. aureus* and *Pseudomonas aeruginosa*, as well as a complex mixture of various fecal organisms isolated using the Qiagen Stool DNA extraction kit (source: NML Genomics Core Facility, *courtesy of Dr. Michelle Alfa's Laboratory*); however, the amplified products from concentrated nares DNA (50 µl) appeared to be slightly smaller (~600 bp), and the negative template control (NTC) water also yielded an amplicon of expected size for bacterial *cpn60* UT (~670bp including fusion primers) from the prepared templates. Consequently, PCR reactions were repeated using "certified" DNA-free PCR water (Ambion, Austin, USA) within a PCR-clean hood previously decontaminated with DNA AWAY™ (Molecular BioProducts, San Diego, USA) and 20 m exposure to UV-irradiation from the in-line UV source lamp. Blunt-end cloning, Sanger sequencing and BLASTn analysis [70] (Appendix A) of the template-positive nares sample bands determined them to be mostly of human derivation, suggesting human DNA contamination within the nares templates (as might be expected) resulting from non-specificity of the fusion primers under the low temperature annealing conditions of the PCR; hence, steps would be needed in the study protocol to limit the contaminating human amplicon carryover within bacterial *cpn60* UT amplicons. Using the same method of identification, the PCR products from the negative template control NTC reactions were found to derive from *Pseudomonas fluorescens cpn60* UT. A literature search revealed that commercial *Taq* polymerase enzymes often contain variable amounts of

contaminating *P. fluorescens* DNA [71]. Combined, these findings implied that the PCR products being generated were due to both human contamination and bacterial contamination from within the *Taq* polymerase (Invitrogen, Carlsbad, USA) and that the developed protocol would have to be adjusted to limit their production.

Alternative commercial sources of *Taq* polymerase enzyme were evaluated in an attempt to increase the specificity of the PCR amplification: namely, HiFi Taq polymerase (Qiagen Inc, Mississauga, Canada) and MTP™ *Taq* polymerase (Sigma-Aldrich Canada Ltd, Oakville, Canada). MTP *Taq* is a highly purified *Taq* polymerase that is certified to be DNA-free. Also evaluated was use of diluted *Taq* as per Spangler *et al.* [71]. In an attempt to increase the yield of the *cpn60* UT product and reduce the amount of human DNA coamplification, addition of Q solution (Qiagen Inc., Missisauga, Canada) and $MgSO_4$ were tested in amounts varying from 1.5 mM to 2.5 mM. After addition of a linear amplification step (described in Section 2.1.4) and removal of human contaminating DNA steps (Section 2.1.5), the resulting *cpn60* UT amplicons from the nares test samples were sequenced using 454 pyrosequencing and evaluated for their specificity.

2.1.3.2 Pooled PCR

Owing to limitations of the touchdown PCR, the PCR cycling program was modified to instead pool four independent reactions carried out at annealing temperatures of 44 ^{o}C, 47 ^{o}C, 49 ^{o}C, and 54 ^{o}C. The reactions proceeded to 35 cycles, as follows: initial 5 m denaturation at 95 ^{o}C; followed by 35 cycles of 30 s denaturation at 94 ^{o}C, 90 s annealing at one of the four designated temperatures, and 60 s extension at 72 ^{o}C, followed by a final 10 m extension at 72 ^{o}C. Amplicons of ~660 bases in length were only detected at an annealing temperature of

54 ^{o}C. To allow *cpn60* UT to be amplified more evenly across all organisms, the primer ratio was changed from 1:3 to 1:1 (P279/280:P1612/1613), and the PCR was tested at each previously mentioned temperature using a high G+C content organism (*Bacillus cereus*) and a low G+C content organism (*Campylobacter upsaliensis*) as test templates. In an attempt to reduce amplification of human DNA and increase yield of the *cpn60* UT amplicon, several different commercial and homemade PCR enhancers were tested, including: 1) twelve different FailSafe™ premix buffers (Epicentre Technologies Inc., Madison, USA), 2) Q-solution (QIAGEN Inc., Missisauga, Canada), 3) SequalPrep™ Enhancer A (Invitrogen, Carlsbad, USA), 4) SequalPrep™ Enhancer B (Invitrogen, Carlsbad, USA), 5) ethylene glycol, 6) dimethyl sulfoxide (DMSO), 7) $MgCl_2$, 8) Pre-CES I as per Ralser *et al.* [72], 9) Pre-Ces II [72], and 10) Pre-CES III [72] following the manufacturer's recommended usage (Appendix A). Amplifications were also tested with Mol*Taq* 16S *Taq* polymerase (Molzym GmbH & Co. KG, Bremen, Germany), another certified DNA-free enzyme, as well as MTP™ *Taq* polymerase with the Mol*Taq* 16S buffer system (Molzym) (Appendix A). The primer concentration was also doubled to 2 µM for each of the four primers per reaction, increasing the yield of the *cpn60* UT amplicon as well as the human amplicon.

Cpn60 UT fusion primers were created using 454 Titanium LibL-specific adaptor A sequences, and 24 MID tags. These new primers were tested using twelve different FailSafe™ buffers along with MTP *Taq* polymerase. Buffer G was chosen as the most effective buffer to use with these new LibL-based primers. To reduce PCR bias, the number of pooled annealing temperatures was reduced by one to three: now 44^{o}C, 49^{o}C and 54^{o}C. To ensure that enough amplicons would be produced, PCR reactions were performed in

triplicate, with a total of nine PCR reactions per nares swab sample (three different annealing temperatures, each amplified in triplicate) (Final PCR conditions described in Section 2.2.8).

2.1.4 Linear Augmentation of Nares Template DNA

To augment the low yield of bacterial DNA templates, a linear whole-genome amplification reaction was performed using Phi29 polymerase coupled with random nonamers labelled WGA3 and WGA4 (Appendix A). This protocol was unsuccessful in augmenting the nares template. The GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp., Piscataway, USA), also using Phi29 polymerase, was also tested and was shown to effectively augment the nares template (described in Section 2.2.5). An electrophoresis step was included to assess the PCR reactions (Section 2.2.6). A Qiagen MinElute purification step (QIAGEN Inc., Mississauga, Canada) was added to remove residual reaction components (Section 2.2.7) before amplification of *cpn60* UT (described in Section 2.2.8). With the addition of template DNA augmentation and these other modifications, the *cpn60* UT PCR reactions were found to generate bacterial amplicons of expected size (~670 bp) as the majority amplicon product.

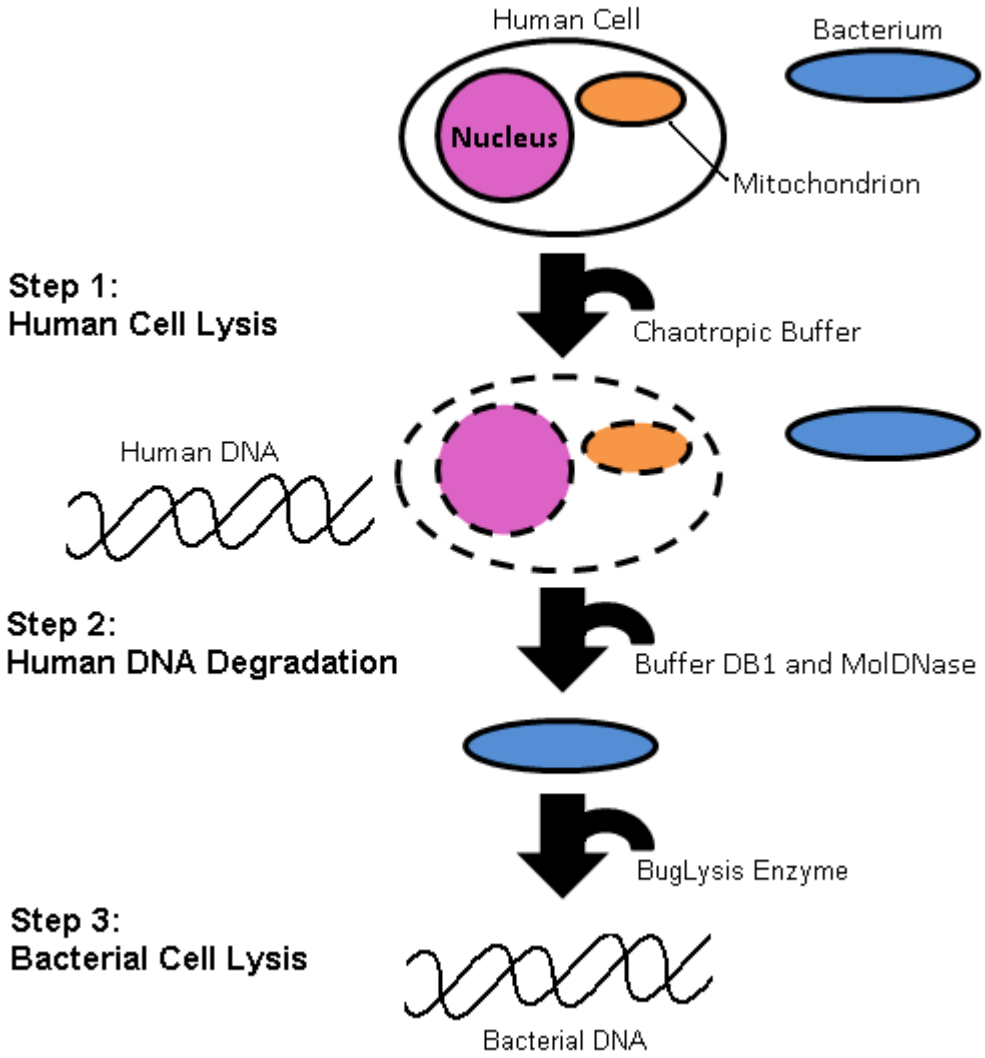
2.1.5 Removal of Human DNA Contamination

The LOOXSTER® (SIRS-Lab GmbH, Jena, Germany) protocol was employed post-template isolation as a procedure to selectively enrich for bacterial DNA prior to GenomiPhi linear whole-genome augmentation, using manufacturer supplied reagents unless otherwise specified. This kit uses an affinity chromatography column to selectively bind to hypo-methylated CpG-containing DNA (representing bacterial and fungal genomic DNA), while methylated human DNA passes through the column. The matrix was prepared by application

of 400 µl of matrix solution to the LOOXSTER® cartridge matrix with centrifugation at 1000 x g for 30 s, followed by addition of 300 µl of Buffer D to the matrix and centrifugation for 30 s at 1000 x g. One hundred fifty microlitres of Buffer D was added to 150 µl of nares template DNA (suspended in ddH_2O), the mixture was applied to the matrix and mixed carefully by pipet-tipmixing. This was allowed to incubate for 30 m at room temperature. The matrix was then centrifuged for 30 s at room temperature and the flow-through was discarded. The matrix was washed twice by addition of 300 µl of Buffer D and centrifugation for 30 s at 1000 x g. The enriched bacterial DNA was eluted by adding 300 µl of Buffer E, heating to 65 ^{o}C and pipet-tipmixing, followed by a 5 m incubation at room temperature. The matrix was then centrifuged for 30 s at 1000 x g. The elution step was repeated once more for a total of 600 µl eluted bacterial DNA. The DNA was precipitated using a mixture of 5 µl Solution G, 60 µl 3M NaAc pH 5.2 and 480 µl isopropanol and vortexing for 10 s. The sample was then centrifuged for 20 m at 16,000 x g at 4^{o}C and the supernatant was discarded. The pellet was washed twice with 1 ml cold 70% ethanol, centrifuged for 5 m at 16,000 x g, and the supernatant was discarded. The pellet was allowed to dry at room temperature for 5 m and then was resuspended in 10 mM Tris and allowed to incubate for 15 m at 50 ^{o}C, followed by a short vortexing to ensure the pellet was completely resuspended.

The MolYsis® procedure (Molzym GmbH & Co. KG, Bremen, Germany), a method of removing human DNA contamination, was tested on the prepared DNA templates prior to DNA template isolation (Figure 2) [73]. Human DNA was removed by selectively lysing human cells in chaotropic buffer, and then enzymatically digesting the human DNA using MolDNase, a proprietary chaotrope-resistant endonucleolytic enzyme. The bacterial cells

Figure 2: Diagram of MoLysis procedure. Step 1 involves addition of a chaotropic buffer to lyse human cells. Step 2 is the enzymatic degradation of human DNA. Step 3 is the enzymatic degradation of bacterial cells to release bacterial DNA.


Human Cell
Bacterium
Nucleus
Mitochondrion
Step 1:
Human Cell Lysis
Chaotropic Buffer
Human DNA
Step 2:
Human DNA Degradation
Buffer DB1 and MolDNase
BugLysis Enzyme
Step 3:
Bacterial Cell Lysis
Bacterial DNA

were pelleted, washed, and then lysed (described in Section 2.2.3). The bacterial DNA was then purified using the GenFind v2 DNA purification system (described in Section 2.2.4).

2.1.6 Purification of *cpn60* Amplicons

The *cpn60* PCR reactions produced ~670 bp *cpn60* UT amplicons, as well as a background of ~150 bp minor amplicons. As these lower molecular weight amplicons could not be removed using a Qiagen MinElute PCR purification column, other purification methods were investigated. Four main amplicon purification methods were tested: namely 1) Gel purification, including E-Gel® SizeSelect gels (Invitrogen, Carlsbad, USA); 2) Gel purification using Bio-Rad certified™ low-melt agarose (Bio-Rad, Mississauga, Canada), and Freeze ‘n Squeeze DNA Gel Extraction Spin Columns (Bio-Rad, Mississauga, Canada); 3) SPRI bead-based purification using Agencourt AmPure XP (Beckman Coulter Genomics, Mississauga, Canada); 4) SPRI bead-based purification using Aline DNA SizeSelector-454 (Aline Bioscience, Woburn, USA) (Appendix A). The Freeze ‘n Squeeze gel extraction method was the most efficient of the gel extraction methods and therefore was used for initial tests. For high throughput tests (thereafter), after the PCR was refined and validated to confirm that it yielded fewer extraneous bands, the AmPure purification method was chosen, as it is automatable, quick and can remove any bands below 500 bp in length with minimal DNA loss (source: NML Genomics Core Facility).

2.1.7 Rarefaction Analyses

A pyrosequencing run was performed on the in-house GS FLX to assess the number of reads required for each sample to ensure maximum alpha (within sample species) diversity representation. This run was prepared using the methods outlined in Participant Sample

Processing (Section 2.2) on amplicons prepared from DNA templates isolated from a single swab, from a healthy adult volunteer. The *mothur* program version 1.27.0 was employed to generate a rarefaction curve for the test data [28]. This program uses a bootstrapping method to generate a random subset of sequences and calculate their alpha diversity [28] . The number of sequences in this subset is continually increased until it contains the set of all sequences (quality-filtered reads). This is then plotted on a graph of alpha diversity vs. subset size to determine the minimum number of sequences required for optimal species representation for each data set. Rarefaction curves were generated using 97% identity OTU clustering with 1000 randomizations.

2.2 Participant Sample Processing

2.2.1 Study Participants

This study was approved by the University of Manitoba Ethics Board as well as the Health Canada Research Ethics Board (REB). Participants were recruited from Manitoba Clinic (Winnipeg, Manitoba) during their first post-natal appointment, when infants were two weeks of age. Brochures and posters were distributed throughout Manitoba Clinic and the Women's Hospital at Health Sciences Centre (Winnipeg, Manitoba) to inform patients about the study for recruitment purposes. Fifty infants and their primary caregivers were enrolled in the study. A consent form was reviewed and signed by each patient caregiver during the first visit (Figure I in Appendix B). Patients were asked to fill out a study questionnaire to acquire information about their living conditions, their general health, smoke exposure, infant delivery method, antibiotic use, and infant feeding method (Figures II and III in Appendix B).

2.2.2 Nares Sampling

A sample collection protocol was based on recommendations contained within the Core Microbiome Sampling Protocol A (HMP Protocol # 07-001, Version Number: 11.0, 29 Mar 2010; Manual of Procedures – Human Microbiome Project, accessed at <http://hmpdacc.org/doc/sops_2/manual_of_procedures_v11.pdf>). One sterile Copan flocked nylon swab (either an adult swab or an infant swab, Copan Diagnostics Inc., Murrieta USA) per subject, per sampling period was moistened with UV-irradiated sterile saline. These swabs were inserted into one of the anterior nares, and used to sample the nares by swabbing the area twice using a twisting motion. Care was taken to avoid contact with the nostril exterior. Each swab was then placed into a pre-labelled 1.5 ml LoBind DNA tube (Eppendorf Canada, Missisauga, Canada) containing 500 µl of Oragene®-DNA solution (DNA Genotek, Kanata, Canada). A negative control pair of swabs (one adult nares swab and one infant nares swab) was processed in parallel, without swabbing the nares. Nares swab samples were stored at ambient temperature in the Manitoba Clinic until they could be collected and subsequently processed in batches. Nares swab samples were stored at room temperature for a period ranging from a few hours to 2 weeks.

2.2.3 Removal of Human Contaminating DNA

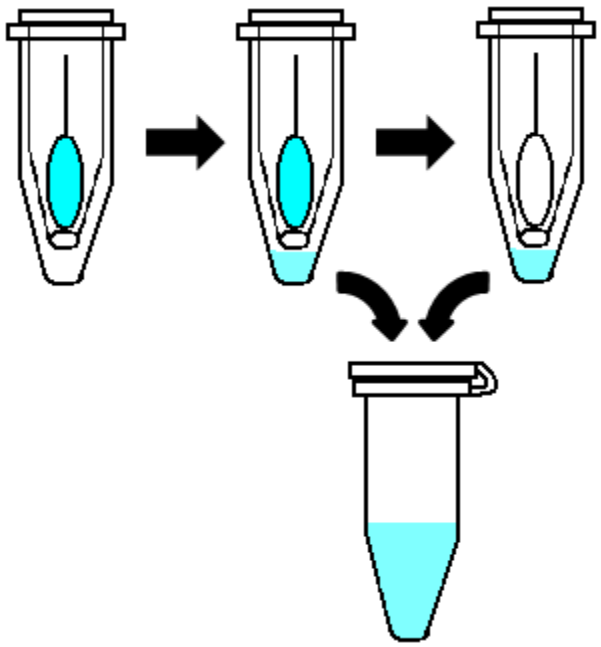
The MolYsis Basic kit for pathogen DNA isolation (Molzym Life Science, Bremen, Germany) was used to reduce contaminating human DNA and isolate bacterial DNA, as per the manufacturer's recommendations. Namely, both 50 µl buffer DB1 and 10 µl MolDNase A were added to the 500 µl Oragene solution containing the nasal swab. The resulting mixture was vortexed for 10 s and allowed to incubate at room temperature for 15 m to digest

human DNA contained within the samples. The tips of 0.6 ml Eppendorf tubes were cut off using scissors (pre-treated with 70% EtOH), the tubes were placed into 1.5 ml Eppendorf tubes using forceps (pre-treated with 70% EtOH), and the swabs were placed into the 0.6 ml Eppendorf tubes as per Figure 3. These were centrifuged twice at 800 x g with transfer of liquid from the 1.5 ml tube to the tube containing the remaining liquid following each centrifugation. This was done to remove bacteria and solution from the swab. Swabs were then discarded. Bacterial cells were harvested by centrifuging the samples at 12,100 x g for 5 m. The supernatant was removed and the pelleted bacterial cells were washed with 1 ml buffer RS, vortexed for 10 s and centrifuged again for 10 m. The supernatant was removed and the pellet was resuspended in 80 µl buffer RL and vortexed for 10 s to homogenize. Finally, 20 µl of BugLysis enzyme was added to the solution, and incubated for 45 m at 37 ^{o}C. This BugLysis incubation was followed immediately with the procedure for purification of bacterial DNA (see Section 2.2.4).

2.2.4 Purification of Bacterial DNA

Bacterial DNA was purified using the Agencourt GenFind v2 purification system (Beckman Coulter Genomics, Danvers, USA). Briefly, 200 µl of Lysis Buffer and 4.5 µl of 96 µg/µl Proteinase K were added to the samples and pipette-mixed 10 times. These were then incubated at 37 ^{o}C for 10 m. The binding buffer was inverted 20 times to ensure that the magnetic particles were fully resuspended; then 150 µl was added to each sample mixture and incubated for 5 m to allow binding of DNA to the magnetic beads. The samples were then placed on a magnet for 10 m to allow the beads to separate from the supernatant. The supernatant was then removed and the magnetic beads were resuspended by pipette tip-mixing in 400 µl of Wash Buffer 1. The samples were then placed back on the magnet rack

Figure 3: Removal of liquid from nares swabs. Swabs are placed into perforated 0.6ml Eppendorf tubes that are contained within 1.5ml Eppendorf tubes and centrifuged multiple times. Liquid is removed from 1.5ml tube and placed in original sample tube following each centrifugation until all of the liquid is removed from the swabs.



until the solution became clear, which took approximately 5 m. Bead washing was repeated once more, followed by two washes with Wash Buffer 2, performed in the same manner as the previous wash, but substituting 250 µl Wash Buffer 2 for the 400 µl Wash Buffer 1. Following the final wash, the supernatant was removed with a manual pipette, ensuring that very little or no buffer remained. The beads were then resuspended in 40 µl of RT-PCR grade water (Ambion, Austin, USA), and incubated for 2 m at room temperature for optimum DNA elution. The samples were pipette tip-mixed again and then placed on a magnet to facilitate bead separation from the aqueous phase. The supernatant was then transferred to clean Eppendorf tubes and stored at 4 ^{o}C until further use (weeks to several months).

2.2.5 Linear Whole-Genome Augmentation of Nares Template DNA

As there was only a trace amount of DNA yield from each study nares sample, bacterial genomic DNA was augmented linearly using the Phi29 polymerase GenomiPhi v2 kit (GE Healthcare Life Sciences, Baie d'Urfe, Canada). Each study subject sample was augmented in duplicate, alongside a negative template control (NTC) as per the manufacturer's recommendations. Nine microliters of manufacturer supplied Sample Buffer was added to 1 µl of template DNA and heated to 95 ^{o}C for 3 m to heat-denature the template. The reactions were then immediately placed on ice, and 9 µl of Reaction buffer and 1 µl of Phi29 Enzyme Mix were added. The mixtures were then placed in a thermal cycler at 30 ^{o}C for 1.5 h, followed by a 10 m incubation at 65 ^{o}C to deactivate the Phi29 enzyme.

2.2.6 Gel Electrophoresis

Gel electrophoresis was performed using 1% agarose gels in 0.5X TBE buffer (0.045 M Tris-borate, 0.001 M EDTA, Fisher Scientific, Ottawa, Canada) made by heating the agarose and buffer solution until the agarose was completely dissolved. Approximately 4 µl of whole-genome augmented DNA template was loaded into each well, along with bromophenol loading dye (0.25% w/v bromophenol blue, 40% w/v sucrose). In addition, 4 µl of 1 kb DNA extension ladder (Invitrogen, Burlington, Canada) was added to one well of each gel in order to inspect and approximate the general size of the augmented genomic DNA fragments. Electrophoresis was at 120 V/cm conducted for 45 m to 1 h. Gels were stained using ethidium bromide (0.5 mg/L) in ddH_2O for 20 m. The gels were then viewed using a Multi-Image Light Cabinet transilluminator (Alpha Innotech, San Leandro, USA).

2.2.7 Removal of Residual Reaction Components

Residual reaction components left over from the whole-genome augmentation reactions were removed using the MinElute PCR purification kit (Qiagen, Toronto, Canada). This kit uses columns containing a silica membrane that binds DNA in high-salt buffer, and elutes the DNA when low salt buffer is applied. Briefly, 100 µl of buffer PB was added to each 20 µl whole-genome augmentation reaction, and placed into a single MinElute column. Both duplicate reactions were applied to a single MinElute column to recover more concentrated DNA in the eluant. Columns were then centrifuged at 17,900 x g for 1 m to allow DNA to bind. Flow-through was discarded and 750 µl of buffer PE was added to each tube and centrifuged at 17,900 x g, for 1 m as a wash step. Flow-through was again discarded, and the samples were centrifuged for another minute to ensure removal of residual ethanol. The

columns were then placed into clean 1.5 ml Eppendorf tubes and 25 µl of Buffer EB (10mM Tris-Cl pH 8.5) was applied to each membrane to elute the bound DNA. After a 1 m incubation at room temperature, the tubes were centrifuged at 17,900 x g for 1 m to collect the eluted DNA. Samples were then stored at 4 ^{o}C until further use.

2.2.8 PCR Amplification of *cpn60* Universal Target

Amplification of *cpn60* UT was achieved using a standard PCR method with a 1:1 ratio of each set of LibL *cpn60* UT fusion primers (P279/P280 and P1612/P1613) (Table 1). Primers P279 and P1612 contained MID tag sequences to allow for identification during a 454 GS FLX sequencing run. There were 24 different MID-tagged primer sets used in total (Table 2). PCR amplification was performed in either a Tetrad PTC-225 thermocycler or a PTC-200 thermocycler (Bio-Rad, Mississauga, Canada). Negative controls containing no DNA template (NTC controls) were included for all sets of PCR reactions. Positive controls were not included in order to prevent possible contamination.

PCR reactions were performed using 12.5 µl FailSafe PCR Premix G (Epicentre, Madison, USA), 2.5 µl 10x primer mix (containing 2 µM of each primer), 0.25 µl MTP *Taq* DNA Polymerase (Sigma-Aldrich, St. Louis, USA), 8.75 µl RT-PCR Grade water (Ambion, Austin, USA) and 1 µl of linearly augmented whole-genome template DNA. The PCR cycling program consisted of a 95 ^{o}C initial denaturation step, followed by 27 cycles of 94 ^{o}C for 30 s, annealing temperature for 1 m, and 72 ^{o}C for 1 m. This was followed by a final elongation step at 72 ^{o}C for 7 m and a final incubation at 4 ^{o}C. Three annealing temperatures (44 ^{o}C, 49 ^{o}C, 54 ^{o}C) were used to ensure greater amplification coverage over the broad range of G+C contents for the various organism templates expected to be present. PCR

Table 1: Oligonucleotides used for validating *cpn60* universal target (UT) molecular procedures

Name	Oligonucleotide Sequence (5'-3')*
P279	CGTATCGCCTCCCTCGCGCCA**TCAG**nnnnnnnnnn GAIIIIGCIGGIGAYGGIACIACIAC
P280	CTATGCGCCTTGCCAGCCCGC**TCAG**YKIYKITC ICCRAAICCIGGIGCYTT
P1612	CGTATCGCCTCCCTCGCGCCA**TCAG**nnnnnnnnnn GAIIIIGCIGGYGACGGYACSACSAC
P1613	CTATGCGCCTTGCCAGCCCGC**TCAG**CGRCGRT CRCCGAAGCCSGGIGCCTT

*Degenerate nucleotides indicated by standard IUPAC ambiguity codes;
I = deoxyinosine ('universal pairing base'); n=MID barcode sequence; TCAG 454 GS FLEX identification key is in bold

Table 2: Multiplex Identifier (MID) Barcode Sequences

Name	Sequence
MID1	ACGAGTGCGT
MID2	ACGCTCGACA
MID3	AGACGCACTC
MID4	AGCACTGTAG
MID5	ATCAGACACG
MID6	ATATCGCGAG
MID7	CGTGTCTCTA
MID8	CTCGCGTGTC
MID9	TAGTATCAGC
MID10	TCTCTATGCG
MID11	TGATACGTCT
MID12	TACTGAGCTA
MID13	CATAGTAGTG
MID14	CGAGAGATAC
MID15	ATACGACGTA
MID16	TCACGTACTA
MID17	CGTCTAGTAC
MID18	TCTACGTAGC
MID19	TGTACTACTC
MID20	ACGACTACAG
MID21	CGTAGACTAG
MID22	TACGAGTATG
MID23	TACTCTCGTG
MID24	TAGAGACGAG

reactions were performed in triplicate for each of three different annealing temperatures, with a total of nine reactions combined together for optimal yield and G+C coverage.

2.2.9 Purification of PCR Products

Excess salts, primer dimers and PCR products below ~300 bp were removed from PCR reactions using the AMPure XP purification system (Agencourt, Danvers, USA). Triplicate reactions were pooled using the Biomek FX robot (Beckman Coulter Genomics, Danvers, USA), and 0.7X Agencourt AMPure XP was added to each reaction. The mixture was mixed thoroughly by pipette tip-mixing and allowed to incubate for 5 m at room temperature. The plates were then placed onto Agencourt SPRIPlate 96 Super Magnet Plates (Agencourt, Danvers, USA) for 2 m to allow the magnetic beads to separate from the solution. The supernatant was aspirated from the plates and discarded. Two wash steps were then performed by adding 200 µl of 70% ethanol to each well, incubating for 30 s at room temperature, and then removing the ethanol from each well. A total of 80 µl of nuclease-free sterile ddH_2O was mixed into each well to elute the DNA. The plates were then placed back on the magnet plate to allow the beads to separate from the solution, and the eluted DNA was transferred to a clean plate and stored at 4 ^{o}C. Given beads are paramagnetic and can leach metal ions, great care was taken to ensure no bead carryover.

2.2.10 Pooling and Quantification of PCR Reactions

Following purification, each set of reactions was quantified using Quant-it^{TM} PicoGreen dsDNA assays (Invitrogen, Carlsbad, USA) as per the manufacturer's protocol. A standard curve was generated using the DNA standard provided in the Quant-it PicoGreen dsDNA assay kit. A total of 594 µl of 1X TE buffer was transferred to a 1.5 ml microcentrifuge tube,

labelled Tube 1. Then, 300 µl of 1X TE buffer was added to each of 7 other 1.5 ml microcentrifuge tubes. Six microlitres of DNA standard was transferred to Tube 1 and vortexed for 10 s. Serial dilutions were performed by transferring 300 µl from each tube to the next and vortexing for 10 s, until a set of standards was created ranging from 100 ng/well to 1.56 ng/well and a negative template control. One microlitre of each purified amplicon DNA sample was added to 99 µl 1X TE buffer in a black MicroFluor plate (VWR, West Chester, USA) and mixed by pipetting up and down. 100 µl of a 1:200 dilution of PicoGreen reagent was added to each well and pipet-tip mixed. Samples were then analyzed using an in-line DTX 880 spectrophotometer (Beckman-Coulter Genomics). Samples were analyzed with reference to the standard curve, ensuring that the standard curve had a minimum R^2 value of 0.98.

Each set of reactions (differing in the annealing temperature, but from common host source) was then pooled at equi-volume amounts with the other reactions from a specific patient and time point, combining a total of nine PCR reactions, using three different annealing temperatures, per patient. Reactions were further pooled into groups of 22 reactions, each with different MID tags. Pooling was done on the Biomek FX liquid handling robot using scripted normalization programming. To create these final pools, concentrations were standardized to 2 ng/µl (where possible) and 6 ng of total DNA from each sample was added to the 22-sample pool, to ensure even representation during the GS FLX sequencing run.

2.2.11 emPCR Amplification

emPCR amplification was performed by the NML Genomic Core sequencing facility as per the Roche GS FLX Titanium Series emPCR Method Manual – Lib-L SV [1] using reagents supplied with the corresponding kit (unless otherwise noted). To prepare the Capture Beads, 1 ml Capture Bead Wash Buffer TW was combined with 9 ml Molecular Biology Grade water. DNA Capture Beads were vortexed and 80 µl per reaction was transferred to separate 1.7 ml tubes. The beads were pelleted by centrifugation for 10 s at 2000 x g, rotating by 180^{o} and spinning for another 10 s (spin-rotate-spin). Supernatants were discarded and each tube of beads was washed twice with 1 ml of 1X Capture Bead Wash Buffer TW (Roche). Beads were resuspended by vortexing, spun down, and the supernatant was discarded following each wash. Each pellet was then resuspended in 640 µl of 1X Capture Bead Wash Buffer TW and distributed evenly among the 1.7 ml tubes. These were then pelleted again, followed by removal of the supernatant.

The amplicon DNA library was then heat denatured for 2 m at 95 ^{o}C, cooled to 4 ^{o}C, and added to the washed DNA Capture beads. The tubes were then vortexed for 5 s and 215 ul of Live Amplification Mix was added to each. The captured library mix was transferred into a prepared emulsion tube, inverted 3 times and placed in the TissueLyser at 15 Hz for 5 m.

Following emulsification, 100 µl of emPCR amplification mix was dispensed into each well of a 96-well thermocycler plate. The emPCR amplification was comprised of 4 m at 94 ^{o}C and 50 cycles of 30 s at 94 ^{o}C, 4.5 m at 58 ^{o}C, 30 s at 68 ^{o}C, followed by a 10 ^{o}C hold. The emulsion was then drawn from each well into 10 ml syringes with 16 gauge blunt

flat-tip needles. A total of 100 µl of isopropanol was added to each well and mixed by pipetting up and down, before being drawn into the same syringe as the samples. Approximately 3 ml of air was added to this mixture and a GS FLX Titanium emPCR Filter SV was attached to the syringe (after removal of the needle). The syringe was then vortexed for 5 s at maximum speed with the filter pointing down onto a Kimwipe. The contents were then expelled out of the syringe through the filter and discarded. Three washes were performed by drawing 3 ml of air into the syringe, followed by 3-5 ml of isopropanol, vortexing for 5 s with the filter pointing down onto a Kimwipe, and expulsion of the contents of the syringe. Three more washes were performed by drawing 3 ml of air into the syringe, followed by 3 ml of Enhancing Fluid XT, vortexing for 5 s with the filter pointing down onto a Kimwipe and expulsion of the syringe contents. Then 3 ml of air was then drawn into the syringe followed by Enhancing Fluid XT until it reached the 0.5 ml mark and vortexing for 5 s. Another 3 ml of air was drawn into the syringe and the bead suspension was dispensed into a new 1.7 ml collection tube, and spun-rotated-spun. The supernatant was then discarded. Finally, 3 ml of air was drawn into the syringe, and the filter was reattached. Enhancing Fluid XT was drawn to the 0.5ml mark, vortexed for 5 s with the filter pointing down onto the Kimwipe, inverted, and 3 ml air was drawn in. The filter was then removed from the syringe and the bead suspension was dispensed into the same 1.7 ml collection tube.

To enrich the DNA bead library, the tubes were spun-rotated-spun and the supernatant was discarded. Twice, 1 ml of the manufacturer-supplied Melt Solution (Roche) was added per tube and incubated for 2 m at room temperature. A 1 ml volume of Annealing Buffer XT (Roche) was added per tube and spun-rotated-spun before the supernatant was discarded, and then repeated once more. To this, 30 µl of Annealing Buffer XT and 12 µl of

Enrichment Primer (Roche) was added to each tube and vortexed. The tubes were then heated to 65^{o}C for 5 m then cooled on ice for 2 m. Then, 500 µl of Enhancing Fluid XT (Roche) was added per tube and vortexed. The tube was spun-rotated-spun and the supernatant was discarded. This was repeated, followed by addition of 800 µl of Enhancing Fluid XT per tube and vortexing.

A total of 40 µl of Enrichment Beads per sample was vortexed and then pelleted using a DynaMag-2 Magnetic Particle Concentrator (MPC) (Invitrogen, Carlsbad, USA). The supernatant was then discarded and 500 µl of Enhancing Fluid XT was added to the mixture. This was vortexed and then pelleted using an MPC. The supernatant was again discarded and this was repeated once more. The tube was then removed from the MPC and 20 µl of Enhancing Fluid XT was added per emulsion and vortexed. Then 40 µl of washed and collected Enrichment Beads were added to each tube of amplified DNA beads and vortexed. The tubes were rotated on a Labquake rotator (Thermo Fisher Scientific, Ottawa, Canada) at room temperature for 5 m and then placed in an MPC for 3-5 m and inverted several times. The supernatant was removed from each tube and the beads were washed with Enhancing Fluid XT until no visible beads remained in the supernatants.

The enriched DNA Beads were collected by removal from the MPC and resuspended in 700 µl of Melt Solution, vortexed for 5 s, then placed back into the MPC. The supernatants containing enriched DNA beads were transferred to a 1.7 ml collection tube, spun-rotated-spun, and the supernatant was discarded. Seven hundred microliters of Melt solution was again added to the enrichment tubes, vortexed for 5 s and placed into the MPC until pelleted. Supernatants were transferred again into the same collection tube, spun-rotated-spun followed by discarding of the supernatant. Three washes were performed by adding 500 µl of

Annealing Buffer XT per collection tube, vortexing for 5 s, spin-rotate-spin and discarding the supernatant. Each bead pellet was resuspended in 60 µl Annealing Buffer XT.

Twelve microliters of Sequencing Primer was then added to each sample, heated for 5 m and then placed on ice for 2 m. Three washes were performed with 500 µl of Annealing Buffer XT added per collection tube and vortexed for 5 s, spun-rotated-spun, and removal of the supernatant. A total of 100 µl of Annealing Buffer XT was added to each bead pellet and vortexed. The beads were then stored at 4 ^{o}C.

2.2.12 GS FLX Pyrosequencing

The nares cpn60 UT amplicons were sequenced in 8 region PicoTiterPlates (PTP) using the GS FLX+ XLR70 sequencer (Roche Applied Science, Laval, Canada) as per the manufacturer's protocol. The GS FLX sequencer was pre-washed and beads were prepared as described in the Roche GS FLX+/XLR70 Sequencing Manual Version 9. Beads were deposited into the PTPs by injecting the suspensions into the Bead Deposition Device, and then centrifuging as follows: A total of 325 µl of Bead Layer 1 (the Enzyme Beads Pre-layer) was loaded into each region of the PTP device, followed by a 5 m centrifugation at 1620 x g. The supernatants were removed, and then 325 µl Bead Layer 2 (DNA and Packing Beads) was added to each PTP region and centrifuged for 10 m at 1620 x g. The supernatant was then removed from each region, and 325 µl of Bead Layer 3 (Enzyme Beads Post Layer) was added to each region and centrifuged for 5 m at 1620 x g. The supernatant was aspirated from each region, and 325 µl of Bead Layer 4 (PPiase Beads) was added to each region, followed by a 5 m centrifugation at 1620 x g. The PTP was then inserted into the sequencer, the seal was removed and the surface was cleaned with 50% ethanol. Sequencing was performed

unidirectionally from the 5' end of each Adaptor A amplicon using 200 flows of each nucleotide (ordered TCAG), as per the instrument standard operating procedure. Data was processed using gsRunProcessor v2.6 using the default amplicon settings.

2.2.13 Raw Data Analysis

Sequence reads were processed and analyzed using *mothur* v1.27.0 as per Schloss *et al* [28]. Pyrosequencing reads were excluded if they contained any of the following: ambiguous bases; more than two nucleotide differences to the primer or to the barcode; more than 8 homonucleotide runs (also called homopolymers); or read lengths less than 150 bases. Screening and exclusion of these sequences, presumed to be of low quality, was performed using the trim.flows and trim.seqs commands within *mothur*. The PyroNoise algorithm was employed using the shhh.flows command in *mothur*, to differentiate between noise generated during the pyrosequencing process and actual diversity found within a sample. PyroNoise [74] was included in the study pipeline to remove base calling errors (largely generating insertion-deletion (indel) errors) generated by pyrosequencing.

Unique sequences were aligned to a *cpn60* UT reference alignment generated using reference sequences from the cpnDB sequence database[20] (accessed at <www.cpndb.ca/cpnDB/home.php>, October 2012). All available bacterial *cpn60* UT sequences were aligned using ClustalW (default settings) to create the reference alignment. 454 sequences were aligned to the reference alignment using *align.seqs* in *mothur* (default settings), which employs a NAST-based aligner[75]. Chimera removal was performed using the *chimera.uchime* command (default settings), which employs the UCHIME program in *mothur*.

Sequences were run through an in-house phylogeny-based taxonomic profiling pipeline (manuscript in preparation) to identify reads that were found to be human contaminant reads. Reads were compared against the hg19 human reference database available from NCBI using a 40bp cutoff for HSP length and an 80% ID cutoff. These sequences were then removed using the *remove.seqs* command in *mothur*.

2.2.14 Sequence Classification and Diversity Analysis

Grouping and classification of resulting sequence reads was performed using *mothur*. The *dist.seqs* command was used to generate a distance matrix (cutoff=0.15) and then sequences were grouped using the average neighbour algorithm into OTUs with 97% identity (%ID) using the *cluster* command. These OTUs were then compared with the cpnDB reference database [20] (accessed at <www.cpndb.ca/cpnDB/home.php>, October 2012) to identify the organisms represented by each group. This was done using the *classify.otu* command in *mothur* with an 80% identity cutoff value and identification to the species level whenever possible. The following parameters were used: basis=sequence, cutoff=80, label=0.03. *Mothur* also was used to calculate α-diversity and measurements for each individual sample including Shannon diversity, Chao richness index, S_{obs} (observed species richness) and Good's estimation of coverage, as well as to generate rarefaction curves. β diversity (between sample diversity) was measured between groups (divided by time point and infant or caregiver) using the Morisita-Horn diversity index (visualized using *heatmap.sim* and *tree.shared* commands in *mothur*). Number of shared OTUs between each group was determined using the *venn* command in *mothur*. Significance testing was conducted by applying Wilcoxon Rank Sum Tests in R version 3.0.1 for Windows and using the *metastats* command in *mothur*. Multivariate analysis was done using SIMCA v. 13 (MKS Umetrics

AB, Sweden). Partial Least Squares Discriminant Analysis (PLS-DA) plots were generated for infants and caregivers at each time point to visualize any relationships between groups, as well as for all time points together to determine relationships between time points.

3. Results

3.1 Protocol Development

3.1.1 Template Extraction Procedures

Nares swabs stored in either saline or Oragene showed similar DNA recovery when measured using Quant-it™ PicoGreen dsDNA assays (data not shown). Swabs were tested with one of three pre-treatments: (1) no treatment, (2) bead beating, and (3) enzyme pre-treatment, along with either the Genfind v2 DNA purification system, or the Oragene DNA purification system. Only trace amounts of DNA were recovered from all swabs using both pre-treatment procedures as well as no pre-treatment. DNA yields were improved following addition of a 4 m incubation at 60 ^{o}C with 1% SDS (v/v); however they were still very low. Prior to purification, half of the samples were treated with 7 µl of a 4 mg/ml solution of RNase A and incubated for 30 m at 37 ^{o}C. The enzyme pre-treatment showed greater DNA recovery than the bead beating treatment. The GenFind v2 DNA purification system also yielded more DNA than the Oragene purification system (Table 3).[38]

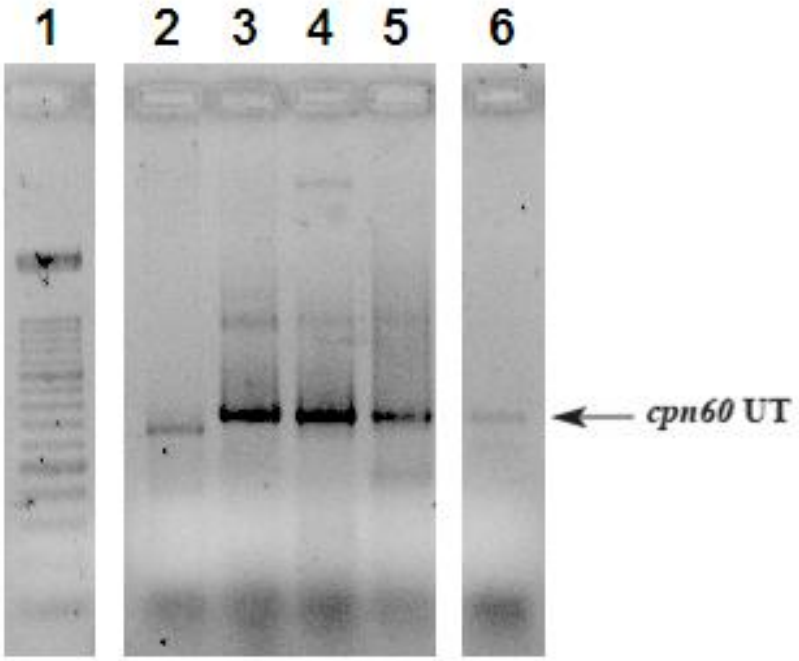
3.1.2 Touchdown PCR

Initial touchdown PCR reactions performed on DNA taken from nares swabs of healthy adult volunteers demonstrated that the primers were able to generate an amplicon of expected size (~670 bp). The major amplicon of concentrated nares templates, however, was ~600 bp, along with a minor ~670 bp amplicon (Figure 4). In initial tests with Invitrogen *Taq* polymerase, the negative template control (NTC) produced a ~670 bp amplicon. To determine its origin, after blunt-end cloning and Sanger sequencing of 16 amplicons from the

Table 3: Yield of extracted DNA from mock-up nares swabs after a variety of pre-treatments (shown in ng)

	Oragene		GenFind	
	RNase	No RNase	RNase	No RNase
No Treatment	34	96	88	62
Bead Beating	196	132	136	178
Enzyme	1274	282	3400	2260

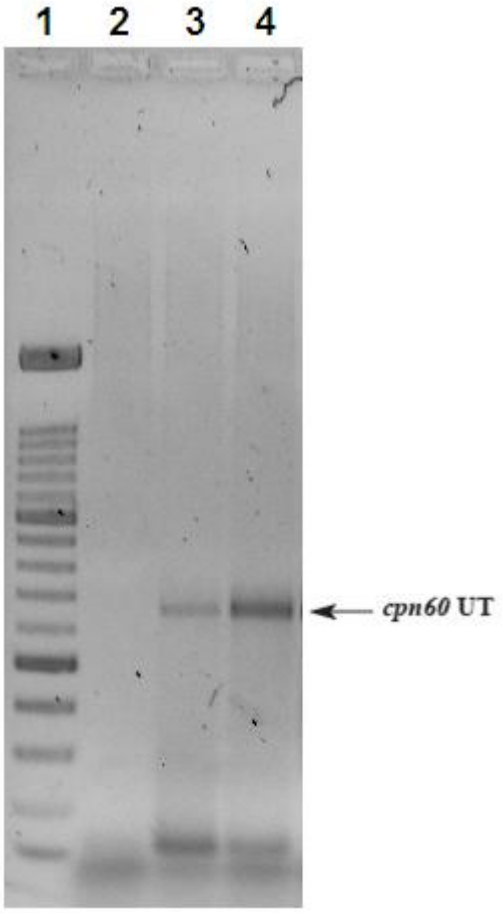
Figure 4: Touchdown PCR reactions using Invitrogen *Taq* polymerase. Lane 1 is the size marker (1 kb DNA extension ladder, Invitrogen). Lane 2 consists of concentrated nares template, showing a ~600 bp contaminating band along with a minor ~670 bp *cpn60* UT amplicon. Lanes 3-5 contain *cpn60* UT amplicons from *S. aureus*, *P. aeruginosa* and a mixture of fecal DNA. Lane 6 is the NTC, clearly showing a contaminating *cpn60* UT amplicon band. Gel image was inverted for visual clarity.


1
2
3
4
5
6
cpn60 UT

positive nares sample bands, 10 randomly selected clones were determined to be human DNA contamination, whereas 6 amplicons were identified as bacterial *cpn60* derived from *Pseudomonas fluorescens* – and thus, represent contaminating DNA in the NTC reactions with Invitrogen *Taq* polymerase. Use of MTP *Taq* (vendor-certified as being DNA-free) was confirmed to lack contaminants and thus the template free control was negative. However, when amplified, each nares sample contained a lower molecular weight amplicon (~600 bp) as the major PCR product alongside the ~670 bp *cpn60* UT product. Addition of 1X Q solution in an attempt to increase the yield of the *cpn60* UT product and reduce the amount of human DNA co-amplification succeeded in increasing the bacterial amplicon yield; however, it was still the minor reaction product.

Gel analysis following *cpn60* UT PCR reactions after LOOXSTER® treatment showed only trace amounts of human amplicons, although decreased yields of *cpn60* UT amplicons as well. The addition of the LOOXSTER® protocol (human DNA removal), the GenomiPhi procedure (linear whole-genome augmentation) and the use of certified DNA-free MTP *Taq* polymerase yielded amplicons of expected size, with no visible human DNA contamination and no visible contamination in the negative template controls (Figure 5). Following 454 sequencing of these *cpn60* nares amplicons as a preliminary investigation, a high number of high annealing temperature organisms (with high G+C templates) were found, as exemplified by higher read abundances observed for *Corynebacterium* (67% G+C), *Pseudomonas* (32-67% G+C), and *Bradyrhizobium sp.* (64% G+C). This was most likely due to biases introduced through the touchdown PCR as well as a high number of environmental organisms.

Figure 5: Touchdown PCR reactions using MTP *Taq* polymerase. Lane 1 is the size marker (1 kb DNA extension ladder, Invitrogen). Lane 2 is the NTC. Lanes 3 and 4 contain nares template DNA following the MoLysis (human DNA removal) and GenomiPhi (linear augmentation of genomic DNA) procedures, respectively. Clear ~660 bp *cpn60* UT bands can be seen with enhanced amplicon yield with GenomiPhi protocol (lane 4). Gel image was inverted for visual clarity.

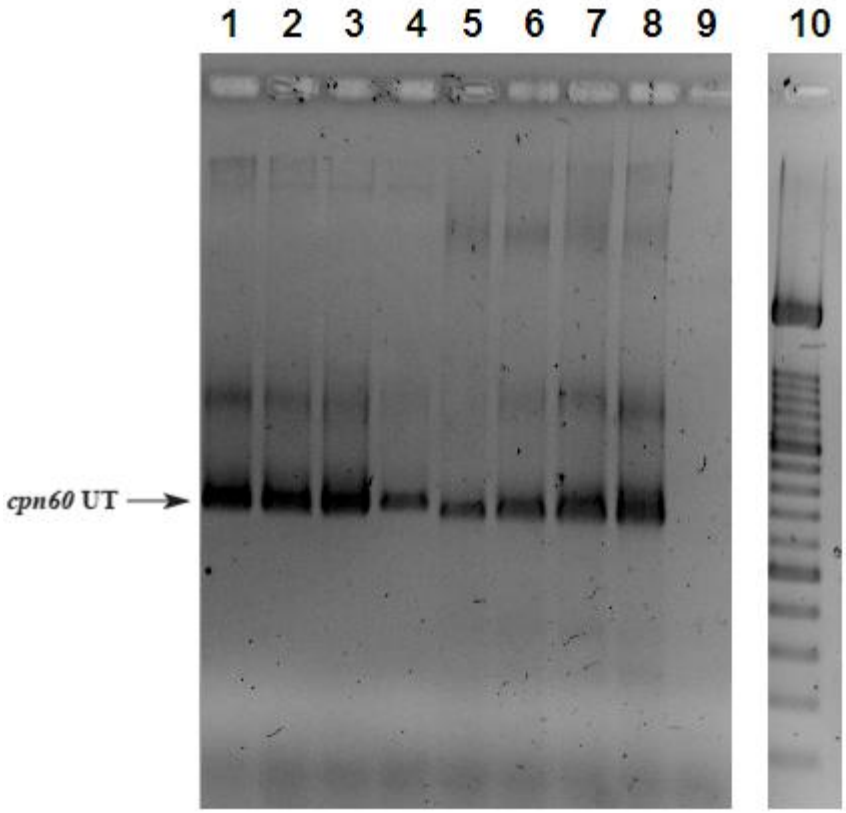

1
2
3
4
cpn60 UT

3.1.3 Pooled PCR

Initial PCR tests utilizing four separate temperatures (44 ^{o}C, 49 ^{o}C, 51 ^{o}C, 54 ^{o}C) produced amplicons only when the annealing temperature was 54 ^{o}C. Tests using *Bacillus cereus* as a high G+C content control and *Campylobacter upsaliensis* as a low G+C content control revealed that only *B. cereus* was amplified at both higher and lower annealing temperatures. After adjusting the ratio of the PCR primers to 1:1, both organisms produced ~670 bp PCR products of roughly equivalent yield at all temperatures (Figure 6). However, this primer adjustment, along with various PCR enhancers failed to adequately amplify typical nares swab templates with robustness and consistency, with the majority of tests producing little to no visible amplicons.

The MolYsis® procedure was chosen over the LOOXSTER® protocol, as it appeared to remove the contaminating human-derived DNA amplicons without significantly decreasing the yield of the *cpn60* UT amplicons. Doubling the primer concentration to 2 µM of each primer, alongside use of the MoLysis protocol (in lieu of the LOOXSTER® protocol) yielded visible *cpn60* UT amplicon bands when viewed on an agarose gel at all four temperatures. Following the switch to LibL fusion primers from LibA fusion primers, tests with FailSafe buffers A-L revealed that buffer G produced the most uniform amplification at both high and low annealing temperatures, as well as the most distinct bands in comparison with the other buffers (data not shown).

Figure 6: PCR tests utilizing four separate temperatures (44 ^{o}C, 49 ^{o}C, 51 ^{o}C, 54 ^{o}C) using *Bacillus cereus* (lanes 1-4) as a high G+C content control and *Campylobacter upsaliensis* (lanes 5-8) as a low G+C content control. *Cpn60* UT bands (~660bp) can be seen at all four temperatures for each organism. Lane 9 was the NTC and lane 10 contained the size marker (1 kb DNA extension ladder, Invitrogen). Gel image was inverted for visual clarity.


1 2 3 4 5 6 7 8 9 10
cpn60 UT

3.1.4 Results of Original Pilot Data

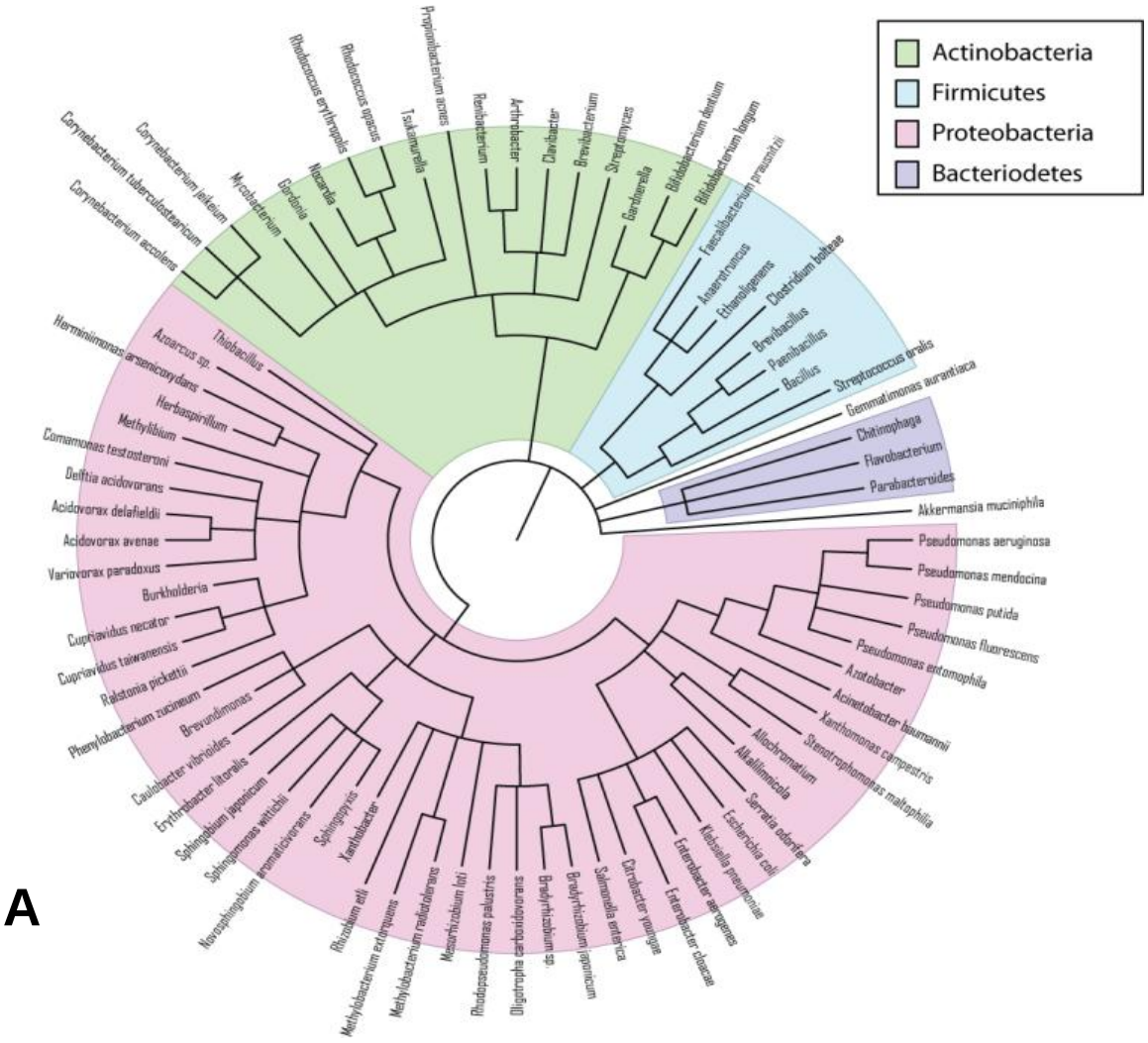
A pilot nares template pyrosequencing run generated with MID-tagged LibL amplicons yielded 416,453 reads ranging in length from 150 to 516 nt (mean = 438 nt). Quality-filtering and trimming resulted in the loss of 293,741 reads (70.5%) owing to short sequence length (after removal of the barcode and 5' terminal 26 nucleotides corresponding to the degenerate forward primer) or because there was no identifiable barcode. The remaining 122,712 quality-filtered, trimmed pilot data reads were submitted for taxonomic assignment. Of these, 93,592 (76.3% of reads input to *BLASTn*) yielded hits within the CpnDB database with over 150 nt match length and >80% identity. These were assigned to 85 unique phylotypes/taxa. The 20 most abundant phylotypes accounted for >98% of community membership. All but two phylotypes were classified within one of four phyla (Proteobacteria, Actinobacteria, Firmicutes orBacteroidetes) (Figure 7).

3.2 Results from Participant Samples

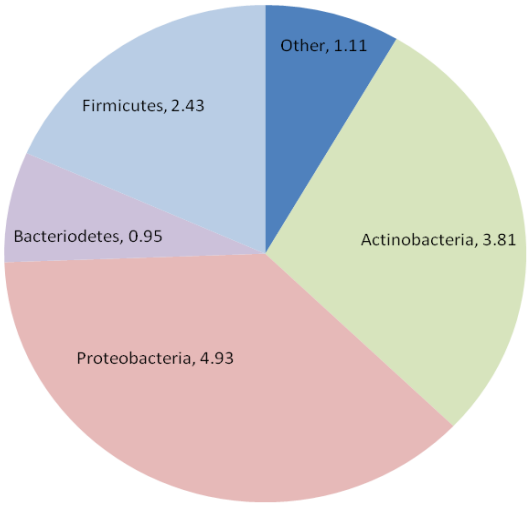
3.2.1 Study Participants

Fifty infants and their primary caregivers were enrolled in the study during the post-natal check-up appointment two weeks after birth. Eight participant pairs did not complete the study – 1 withdrew, 3 moved or changed clinics, and 4 failed to book follow-up appointments. Of the 50 original participants, all primary caregivers were female, 30 infants were male and 20 were female. The ethnicities of the infants were as follows: approximately half (24) of the infants were Caucasian, 8 were Filipino, 7 were Aboriginal/Metis, 2 were East Indian, 1 Latin American, 1 West Asian, 1 Chinese, 3 of mixed descent and 2 failed to specify. Thirty-seven (74%) infants were born via vaginal delivery, and 12 (24%) were

Figure 7: Taxonomic distribution (A) and log abundance (B) for *cpn60* UT reads acquired by 454 unidirectional pyrosequencing in a pilot run using LibL amplicons. Anterior nare templates acquired from a single individual were sampled, sequenced, and data processed using procedures described in the Methods text. Eighty five unique phylotypes grouped predominantly into four phyla.


Actinobacteria
Firmicutes
Proteobacteria
Bacteriodetes
Propionibacterium acnes
Rhodococcus erythropolis
Rhodococcus opacus
Tsukamurella
Rothia
Nocardia
Mycobacterium
Corynebacterium jeikeium
Corynebacterium tuberculostearicum
Corynebacterium accolens
Renibacterium
Arthrobacter
Clavibacter
Brevibacterium
Streptomyces
Gordonella
Bifidobacterium adolescentis
Bifidobacterium longum
Bifidobacterium pseudolongum
Faecalibacterium prausnitzii
Anaerotruncus
Ethanoligenens
Clostridium botulinum
Brevibacillus
Paenibacillus
Bacillus
Streptococcus gralis
Gemmatimonas aurantiaca
Chitinophaga
Flavobacterium
Parabacteroides
Akkermansia muciniphila
Pseudomonas aeruginosa
Pseudomonas mendocina
Pseudomonas putida
Pseudomonas fluorescens
Pseudomonas entomophila
Azotobacter
Acinetobacter baumannii
Xanthomonas campestris
Stenotrophomonas maltophilia
Alkalilimnicola
Serratia odorifera
Escherichia coli
Klebsiella pneumoniae
Enterobacter aerogenes
Enterobacter cloacae
Citrobacter youngae
Salmonella enterica
Bradyrhizobium japonicum
Bradyrhizobium sp.
Oligotropha carboxidovorans
Rhodopseudomonas palustris
Mesorhizobium loti
Methylobacterium radiotolerans
Methylobacterium extorquens
Rhizobium etli
Xanthobacter
Sphingopyxis
Novosphingobium aromaticivorans
Sphingomonas wittichii
Sphingobium japonicum
Erythrobacter litoralis
Caulobacter vibrioides
Brevundimonas
Phenylobacterium zucineum
Ralstonia pickettii
Cupriavidus taiwanensis
Cupriavidus necator
Burkholderia
Variovorax paradoxus
Acidovorax avenae
Acidovorax delafieldii
Delftia acidovorans
Comamonas testosteroni
Methylibium
Herbaspirillum
Azoarcus sp.
Herminiimonas arsenicoxydans
Thiobacillus

A


Other, 1.11
Actinobacteria, 3.81
Proteobacteria, 4.93
Bacteriodetes, 0.95
Firmicutes, 2.43

B

delivered by caesarean section, with one subject failing to specify. All participants received recommended vaccinations at each appointment, with approximately one quarter receiving the rotavirus vaccination (9/42) at two months of age, (8/41) at 4 months and (6/37) at 6 months. No infants were enrolled in daycare before they had reached one year of age, at which point 14 infants (41%) attended daycare. Nine of the 50 primary caregivers signed up for the study were smokers, but all infants were reported to be unexposed to secondary smoke.

3.2.2 Observations and Processing of Participant Samples

Upon visual observation, the consistency of the fluid attached to the nares swabs of two week old infants was quite different than that of the primary caregivers. Infant nares swabs contained thicker, more viscous and more copious mucus than those of primary caregivers. By two months of age, the consistency appeared to be more similar to the adult nares swabs, and by 4 months of age, they were indistinguishable.

Four participant samples arrived missing the Oragene solution, and one arrived in saline solution instead of chaotropic Oragene solution, leaving these samples unfit for processing owing to potential uninhibited growth of microorganisms prior to sample processing. Ten pairs were excluded from the study owing to participant withdrawal or insufficient number of swabs received. In total, samples from all 5 time periods (if available) of 40 participant pairs were *cpn60* amplified and processed for sequencing.

Concentrations of each set of PCR reactions (after pooling of replicates) were generally between, 0.5-20 ng/µl, with the lowest at 0.08 ng/µl. These were standardized to 2

ng/µl whenever possible, and 6 ng total DNA was added into each 22-sample pool, each containing a different MID tag. In total, 354 samples were sent for pyrosequencing.

3.2.3 454 Pyrosequencing Results

A total of 2,075,480 sequence reads were obtained from two GS FLX sequencing runs. After sequence read quality filtering and filtering for human contamination, only 1,505,017 reads remained for further analysis. The majority of excluded reads were removed due to short read length (<150 bases). *Cpn60* UT reads fewer than 150 nt cannot be used to distinguish between species, and therefore were excluded. In addition, the majority of the reads generated in the first two lanes of the first sequencing run were of too poor quality for analysis. Owing to the low read coverage, another supplemental 8 lane sequencing run was performed to acquire supplemental reads for the participant samples that lacked sufficient coverage for a robust analysis.

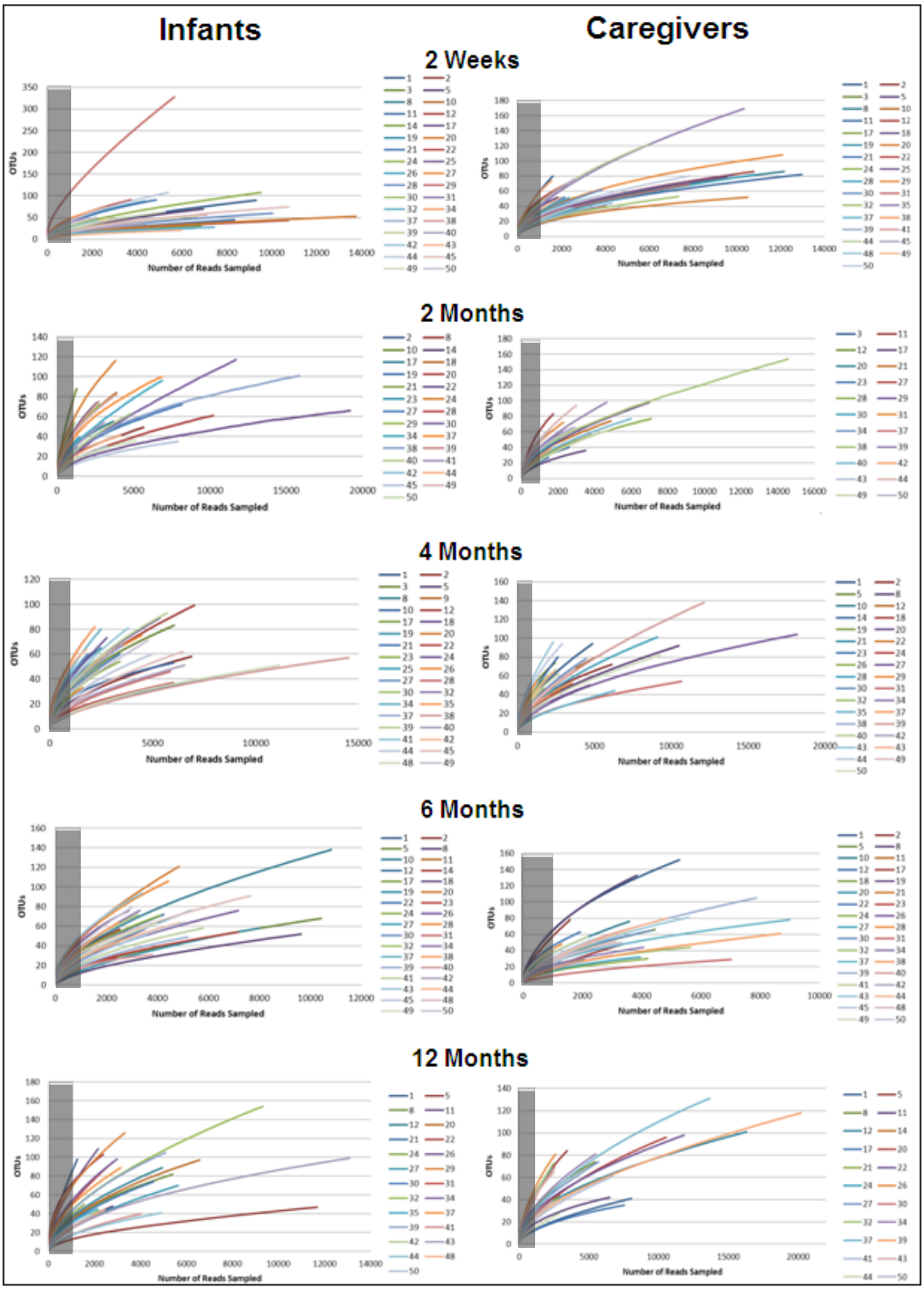
A total of 2,969,857 sequence reads were obtained from the three combined sequencing runs. After filtering and removal of 826 chimeric reads, the final dataset going forward to analysis contained 1,512,776 reads with 95,197 unique sequences. Of these, 1,470,543 (97.2%) could be identified via sequence identity as bacterial *cpn60.* On average, 4,154 reads per subject sample were obtained, with the fewest being 1 read, and the most being 20,215. Participant samples with fewer than 1000 reads were removed from analysis owing to insufficient coverage (see 3.2.4 below). This left 297 remaining individual samples originating from 38 study participant pairs.

Clustering the unique reads into OTUs at a 3% genetic distance resulted in 7758 different OTUs, of which 6700 (86.3%) could be matched to the *cpn60* database. These OTUs were matched to sequences from the *cpn60* database (cpnDB) with an 80% threshold value using *mothur*. The largest OTUs (containing the most reads) were matched to *Corynebacterium* spp. (209,598 reads), *Dolosigranulum pigrum* (156,550 reads), and Actinobacteria (Class) (138,541 reads). A total of 4259 OTUs contained singleton reads, while 273 OTUs contained more than 100 reads. A high percentage (97.8%) of reads could be matched to cpnDB, with the remaining being labelled as unclassified bacteria. Almost all nares microbiota were members of three main phyla: Actinobacteria (53.9%), Firmicutes (40.7%), and Proteobacteria (3.1%), with trace amounts of Bacteroidetes and Deinococcus-Thermus.

3.2.4 Alpha Diversity

Rarefaction curves were generated for each participant at each time point using mothur in order to estimate species richness (Figure 8). Infant 29 from Time point 1 was observed to be an extreme outlier containing nearly double the OTUs relative to any other sample, and more than triple the number from any other infant sample during that time point. Upon investigation of this subject's clinical data, the only possible outlying factor noted was the presence of a pet gecko in the participant's home. This sample was included in Figure 8 (Infants, 2 weeks), but then removed from further analyses due to possible contamination. Good's coverage values ranged from 94.4%-99.8% per subject, with an average of 99%. Such a level of coverage implies that a sampling depth of approximately 1000 reads per

Figure 8: Rarefaction curves generated from nares swab samples taken over the course of one year from 40 infants and their primary caregivers. Greyed out areas indicate cutoff point of 1000 reads.


Infants
Caregivers
2 Weeks
2 Months
4 Months
6 Months
12 Months
OTUs
Number of Reads Sampled

sample should be sufficient to sample the true biodiversity of each specimen with relative confidence.

Overall, the mean Chao score for species richness in infants across all time points was 155.27 (range 39 – 391.25) and in caregivers was 143.59 (range 49 – 413.5). The mean Shannon diversity index for infants was 1.11, ranging from 0.10 to 3.47, and in caregivers was 1.18 (range 0.10 to 2.48). Average diversity values for all five time points are shown in Table 4A. The Chao score is a measure of species richness, which was similar between caregivers and infants across all time periods, with the exception of the 6 month time period, where caregivers had a lower richness ($p = 0.039$, Wilcoxon rank sum test) (Table 5). The Shannon diversity, a measure that takes into account both richness and species distribution, was significantly lower in two week old infants than in their primary caregivers ($p = 0.0047$, Wilcoxon rank sum test), but there was no significant difference during any other time point (Table 5). This reveals that there is a similar richness of lineages (number of different species taking into account how many species are rare) between two-week old infants and their primary caregivers, but the caregivers have a more diverse microbiota (relative abundances of species, along with number of species).

When comparing average relative diversity over time of infant nares microflora, as hypothesized, it was observed that infants at two weeks of age contain significantly less diverse nares microbiota (calculated using the average Shannon diversity index) relative to infants at 4 months ($p = 0.038$, Wilcoxon rank sum test), 6 months ($p = 0.037$) and 12 months ($p = 0.018$) (Data not shown). Otherwise, there was no difference in average diversity (comparing across infants) between 2 weeks and 2 months of age, or between any

Table 4A: Average alpha diversity measurements of infants and their primary caregivers over the course of one year

	Chao Richness		Shannon Diversity		Species Richness (S_{obs})		*n*
Infants	Mean	SD	Mean	SD	Mean	SD	
2 Weeks	130.4	68.7	0.89	0.61	56.7	23.2	33
2 Months	146.6	74.6	1.01	0.58	66.6	24.6	27
4 Months	142.8	64.6	1.16	0.57	62.9	17.2	35
6 Months	155.5	68.9	1.16	0.57	66.2	23.3	34
12 Months	174.0	83.6	1.33	0.71	77.6	30.7	25
All	155.3	109.4	1.12	0.65	67.0	32.3	40
Caregivers							
2 Weeks	139.7	68.0	1.25	0.53	68.6	27.5	31
2 Months	151.8	84.2	1.26	0.48	67.0	26.7	23
4 Months	151.6	91.2	1.21	0.46	64.4	24.6	32
6 Months	122.3	67.2	1.01	0.62	57.7	28.6	34
12 Months	161.6	71.3	1.15	0.66	70.2	27.1	22
All	143.6	78.0	1.18	0.56	65.0	27.0	40

SD = Standard Deviation

Table 4B: Median alpha diversity measurements of infants and their primary caregivers over the course of one year

	Chao Richness	Shannon Diversity	Species Richness (S_{obs})	*n*
Infants	Median	Median	Median	
2 Weeks	129.25	0.87	51	33
2 Months	131	1.06	63	27
4 Months	127	1.16	60	35
6 Months	147.1	0.99	63.5	34
12 Months	165.67	1.39	82	25
All	137.34	1	61	40
Caregivers				
2 Weeks	129.5	1.24	68	31
2 Months	131	1.3	65	23
4 Months	116.8	1.29	58	32
6 Months	104.5	0.99	50	34
12 Months	151.4	1.14	71	22
All	122.67	1.24	62.5	40

Table 5: Species Richness and Diversity Measures of Nares Microbiota in Infants and their Primary Caregivers

Time point	No. of Participants	Richness Score,* mean ± SD	*p* value	Diversity index,* mean ± SD	*p* value
Two Weeks					
Caregivers	31	139.72 ± 68.00	0.52	1.25 ± 0.53	**0.0047**
Infants	33	130.42 ± 68.70		0.88 ± 0.61	
Two Months					
Caregivers	23	151.79 ± 84.19	0.93	1.26 ± 0.48	0.12
Infants	27	146.63 ± 74.58		1.01 ± 0.58	
Four Months					
Caregivers	32	151.60 ± 91.23	0.75	1.21 ± 0.46	0.63
Infants	35	142.75 ± 64.66		1.16 ± 0.57	
Six Months					
Caregivers	34	122.35 ± 67.17	**0.039**	1.07 ± 0.62	0.5
Infants	34	155.50 ± 68.89		1.16 ± 0.57	
Twelve Months					
Caregivers	22	161.64 ± 71.28	0.73	1.15 ± 0.66	0.56
Infants	25	174.02 ± 83.62		1.33 ± 0.71	

SD = standard deviation

[1] Richness was measured with the Chao1 score. Diversity was measured with the Shannon diversity index

[2] Wilcoxon rank sum test

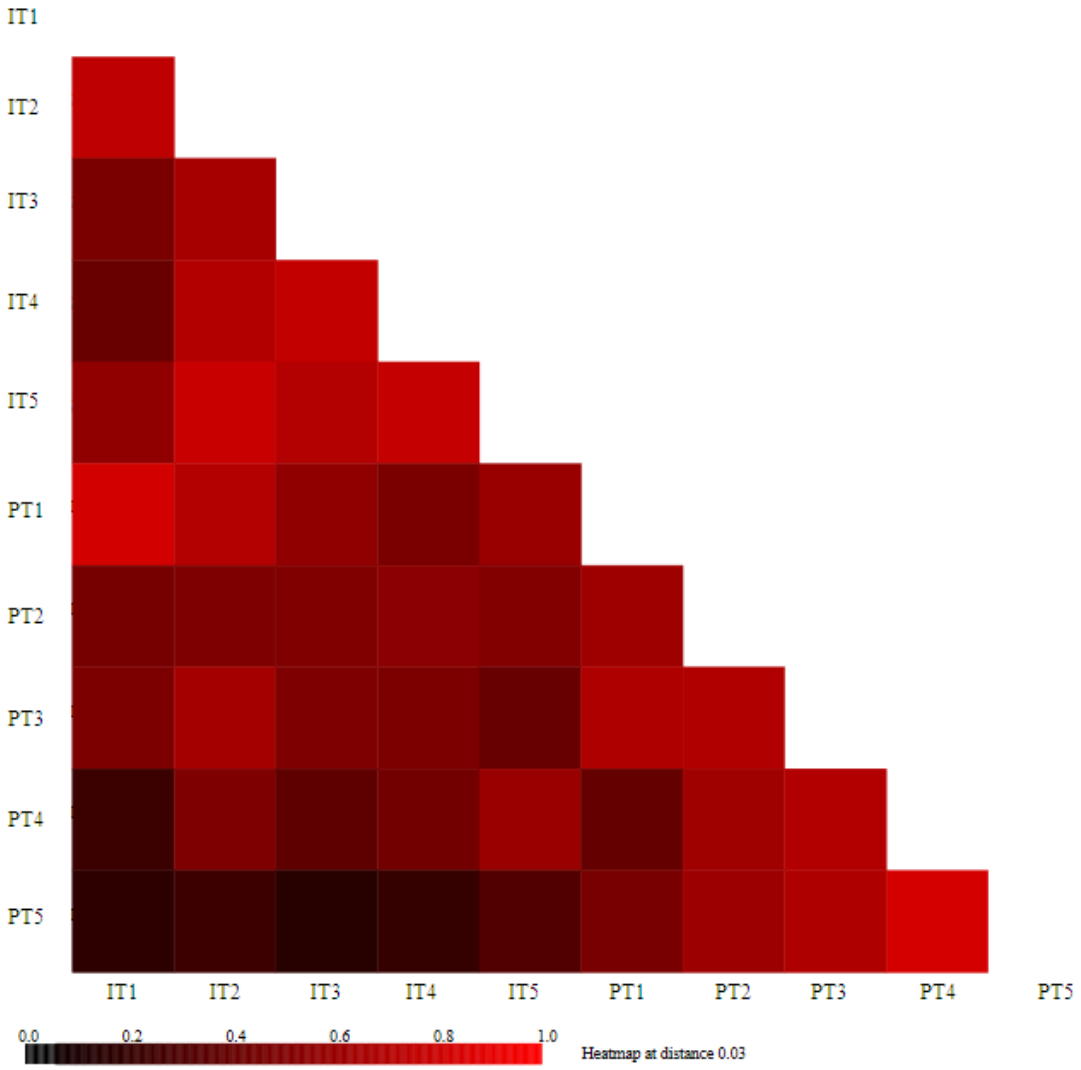
other time points. The average species richness also was significantly different for 2 week old and 12 month old infants (p = 0.032, Wilcoxon rank sum test). Similar trends were observed for the median values of richness and diversity (Table 4B). As expected, the diversity of the primary caregiver nares microbiota did not change with time overall, with the exception of a unexpected significant difference in Chao diversity between time points 4 (at 6 months) and 5 (at one year) (p = 0.032) (Data not shown).

3.2.5 Beta Diversity

Beta diversity was measured using the Morisita-Horn index (chosen due to its ability to evaluate differences of relative abundances of organisms in populations of different sizes), calculated in *mothur*, and shown as the Morisita-Horn index between infants and their primary caregivers at each time point (Figure 9). Infants aged two weeks were most similar to their primary caregivers during the same time period, but quite dissimilar to both infants and caregivers at all remaining time periods. Infants were most similar to each other during time points 2 (2 months) and 5 (12 months); and least similar during time points 1 (2 weeks) and 4 (6 months). The caregivers at time point 5 were most dissimilar to children of all time points (Figure 9), although this time point appears to be somewhat of an outlier in caregivers as seen in the previous section. When comparing caregivers across time points, 4 and 5 appeared to be the most similar, and time points 1 and 4 appeared to be the least similar.

The number of shared OTUs between infants and caregivers was derived using the *venn* command in mothur. This command calculates the number of OTUs shared between each user-defined group, as well as the number exclusive to each group. The number of shared OTUs between infants and caregivers (range: 225 to 347) was relatively stable across

Figure 9: Heatmap of Morisita-Horn diversity indices between infants (I) and their primary caregivers (P) at each time point (T1-T5). Gradations in shading intensity reflect the degree of similarity between time points, with the darkest intensities denoting most dissimilar (approaching a Morisita-Horn diversity index value of 0.0) and the brightest denoting the most similar (index value approaching 1.0).


IT1
IT2
IT3
IT4
IT5
PT1
PT2
PT3
PT4
PT5
IT1
IT2
IT3
IT4
IT5
PT1
PT2
PT3
PT4
PT5
0.0
0.2
0.4
0.6
0.8
1.0
Heatmap at distance 0.03

all time points. Time points 3 and 5 were less similar, with 11.8% and 11.2% shared OTUs, respectively; time points 1, 2, and 4 had 14.5%, 13.5% and 13.5% shared OTUs, respectively. When looking at infants alone over the course of one year, the number of shared OTUs ranged from 287 to 444. The time points that had the fewest shared OTUs were 2 weeks and 12 months (both 11.3%) and the greatest were 4 months and 6 months (18.4%). In adults, the number of shared OTUs ranged from 297 to 380. Time points that had the least OTU overlap were time points 1 and 4 (16.0%). Time points 1 and 5 had the most shared OTUs with 19.3% similarity. These findings are evident in Figure 10, a phylogenetic tree generated by calculating the Morisita-Horn index on a subsampling of 25,000 reads per group to determine distance between each group. This tree demonstrates a seasonal trending of observed nares diversity in caregivers, with groupings of the 2 week, 2 month and 1 year time points (winter) and the 4 month and 6 month time points (summer). In infants, the figure depicts the observed dissimilarity of the 2 week time point with all other time points.

3.2.6 Profiles of Nares Microbiota

The nares microbiota of study subjects varied at both the phyla and the genera levels. Figure 11 shows the average relative abundance of phyla and genera for infants and primary caregivers at each time point. The nares microbiota of both infant and adult groups were generally dominated by the phyla Actinobacteria (mean 46.8%, standard deviation (SD) 37.7), and Firmicutes (mean 45.5%, SD 38.3). The majority of participants also harboured Proteobacteria (mean 5.4%, SD 14.9) in smaller amounts, and some participants harboured Bacteroidetes (7-13% of participants per time point, in low abundance), and a few participants carried one member of the phylum Deinococcus-Thermus in very low abundance. In both groups, the relative abundance of Actinobacteria and Firmicutes

Figure 10: Phylogenetic tree showing the relationship between nares microbiota derived from infants (I) and their primary caregivers (P) across 5 different time points (T1-T5) during the first year of life. Phylogenetic distance was calculated using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) for each time point based on a random subsampling of 25,000 reads from all subjects per group. The tree was generated and viewed using the *tree.shared* command in *mothur* (parameters: calc=morisitahorn, subsample=25000). The scale bar depicts an inferred distance (Morisita-Horn diversity index) of 0.04 (no units).

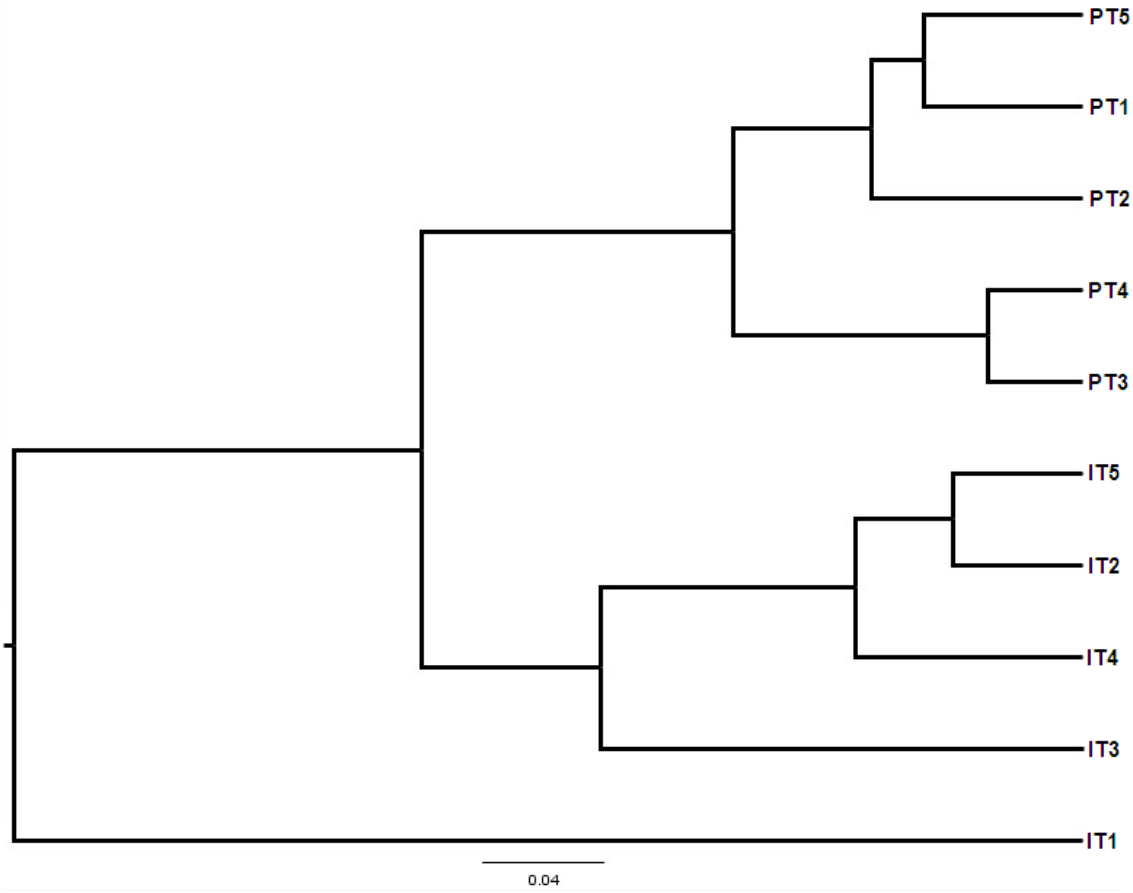
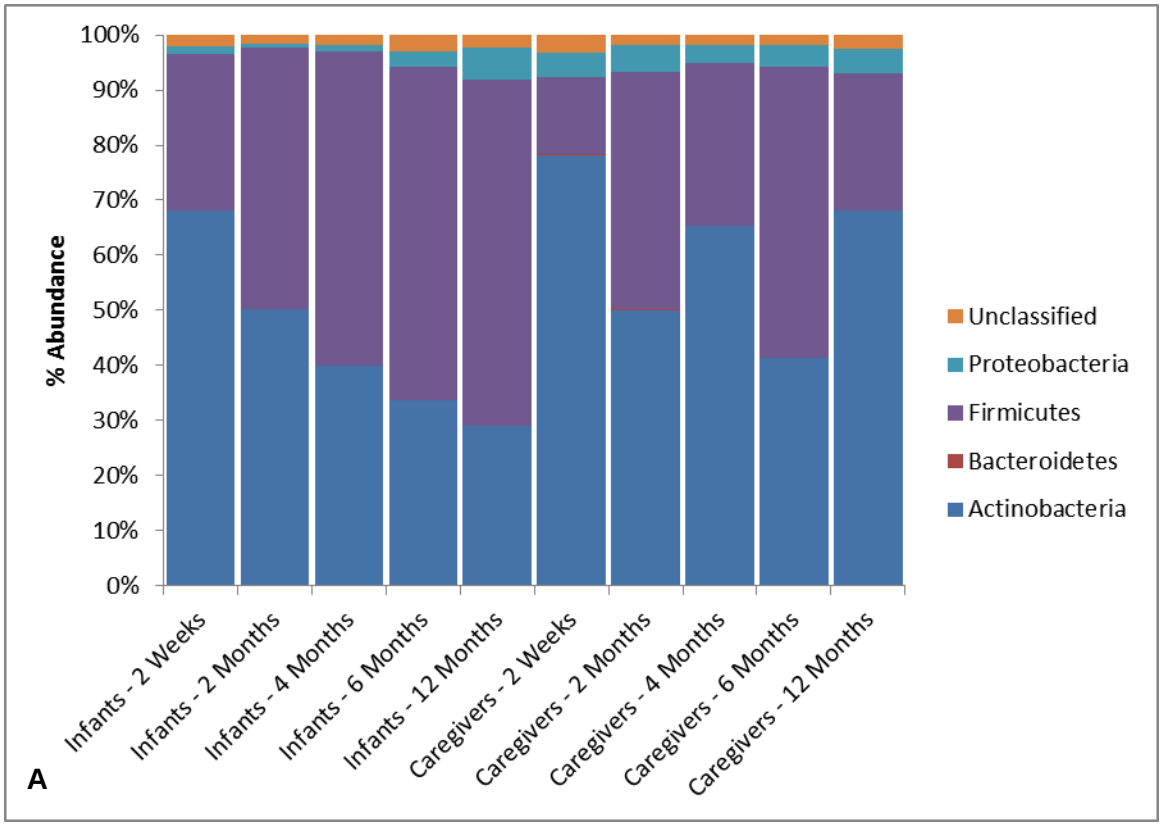
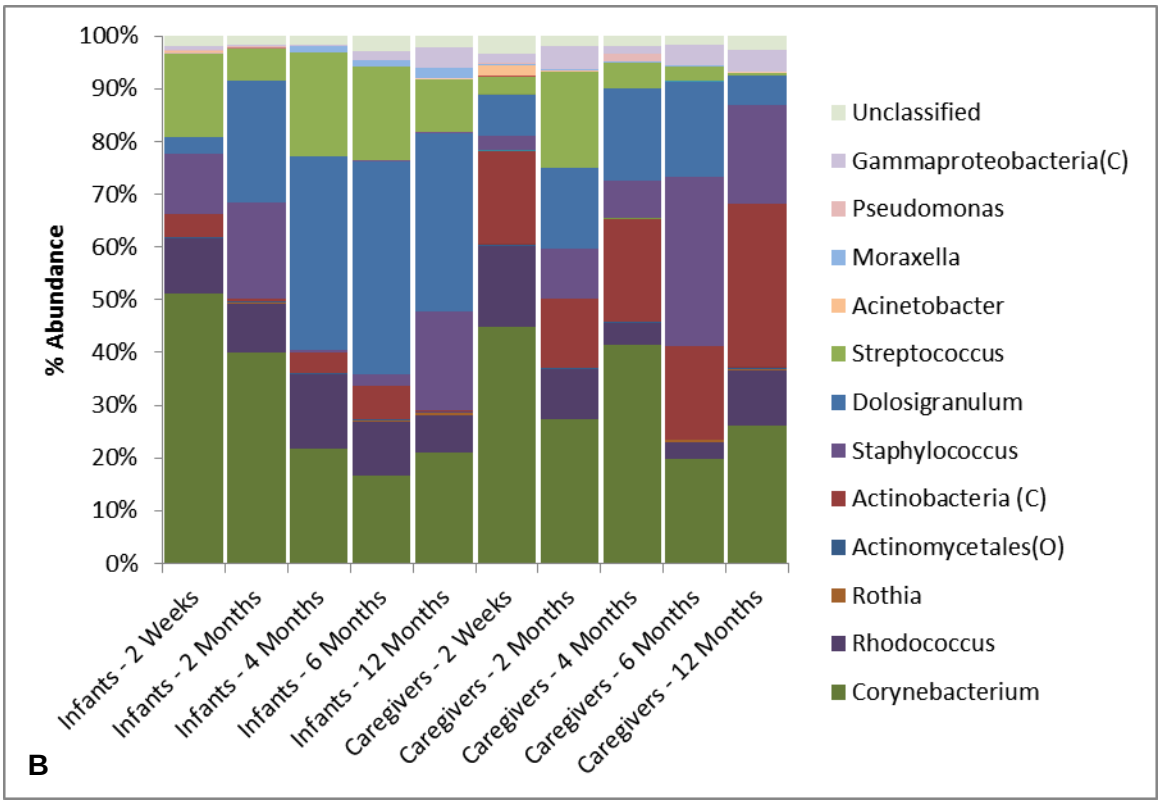

PT5
PT1
PT2
PT4
PT3
IT5
IT2
IT4
IT3
IT1
0.04

Figure 11: Average relative abundance of phyla (A) and genera (B) in the nares microbiota of infants and their associated primary caregivers across five time points during the first year of life. The abundance represents the average values derived from 22-35 subjects in each group (of a possible 40 subjects maximum). Legend entries with (C) or (O) notations could only be classified to the class or order taxonomic levels, respectively.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% Abundance
Infants - 2 Weeks
Infants - 2 Months
Infants - 4 Months
Infants - 6 Months
Infants - 12 Months
Caregivers - 2 Weeks
Caregivers - 2 Months
Caregivers - 4 Months
Caregivers - 6 Months
Caregivers - 12 Months
Unclassified
Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
A


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% Abundance
Infants - 2 Weeks
Infants - 2 Months
Infants - 4 Months
Infants - 6 Months
Infants - 12 Months
Caregivers - 2 Weeks
Caregivers - 2 Months
Caregivers - 4 Months
Caregivers - 6 Months
Caregivers - 12 Months
Unclassified
Gammaproteobacteria(C)
Pseudomonas
Moraxella
Acinetobacter
Streptococcus
Dolosigranulum
Staphylococcus
Actinobacteria (C)
Actinomycetales(O)
Rothia
Rhodococcus
Corynebacterium
B

appeared to be inversely related, however in infants Actinobacteria decreased over the course of the year and Firmicutes increased in abundance with each subsequent time point of sampling. The abundance of Proteobacteria showed a marked increase between the ages of 6 months and 1 year. In adults, the relative abundance of all three major phyla fluctuated throughout the year, but generally maintained a steady state, without showing an increase of any phylum. In infants, Actinobacteria appeared to dominate during the first two time points, with Firmicutes being the most dominant phylum at 4 months, 6 months and 12 months of age. In adults, Actinobacteria was the most prevalent phylum during all time points except 6 months, in which 50% of detected organisms were Firmicutes, and 40% were Actinobacteria. In general, infants had fewer Actinobacteria and Proteobacteria, and more Firmicutes than primary caregivers.

Upon observation of average relative abundance at the genera taxonomic level (Figure 11B), it is apparent that *Dolosigranulum* and *Streptococcus* are more prevalent in infants than in primary caregivers. Infants carry very little *Dolosigranulum* at 2 weeks of age, with the relative abundance increasing at 2 months and again at 4 months, then levels remaining relatively steady for the rest of the year. Relative abundance of *Corynebacterium* decreased in infants over the first year of life. Members of the Actinobacteria class are more commonly found in caregivers than infants at all time points. There was no apparent trending in general of specific genera in adults over the course of one year.

Differences in relative abundances of genera between groups were detected using Metastats, available within the *mothur* program. Metastats generates a *p* value, representing the false positive rate, and a *q* value, representing the false discovery rate. Each group was compared to all others in pairwise comparisons applying 1000 permutations and a *p* value

threshold of 0.05. When comparing infants *vs*. primary caregivers at each time point, caregivers were observed to have significantly more Actinobacteria during time points 2 (p = 0.007, q = 1), 3 (p = 0.01, q = 1), 4 (p = 0.03, q = 1), and 5 ($p < 0.00001$, $q < 0.00001$). Caregivers carry more *Dolosigranulum* at time point 1 (p = 0.002, q = 1), but infants have a higher relative abundance at time points 3 (p = 0.01 q = 1), 4 (p = 0.04, q = 1) and 5 (p = 0.001, q = 0.2). Infants also carry significantly more *Streptococcus* than caregivers at time points 3 (p = 0.005, q = 0.8), 4 (p = 0.002, q = 0.6) and 5 (p = 0.001, q = 0.2), while caregivers have a higher relative abundance of *Staphylococcus* at time points 3 ($p < 0.00001$, $q < 0.0001$) and 4 (p = 0.002, q = 0.6).

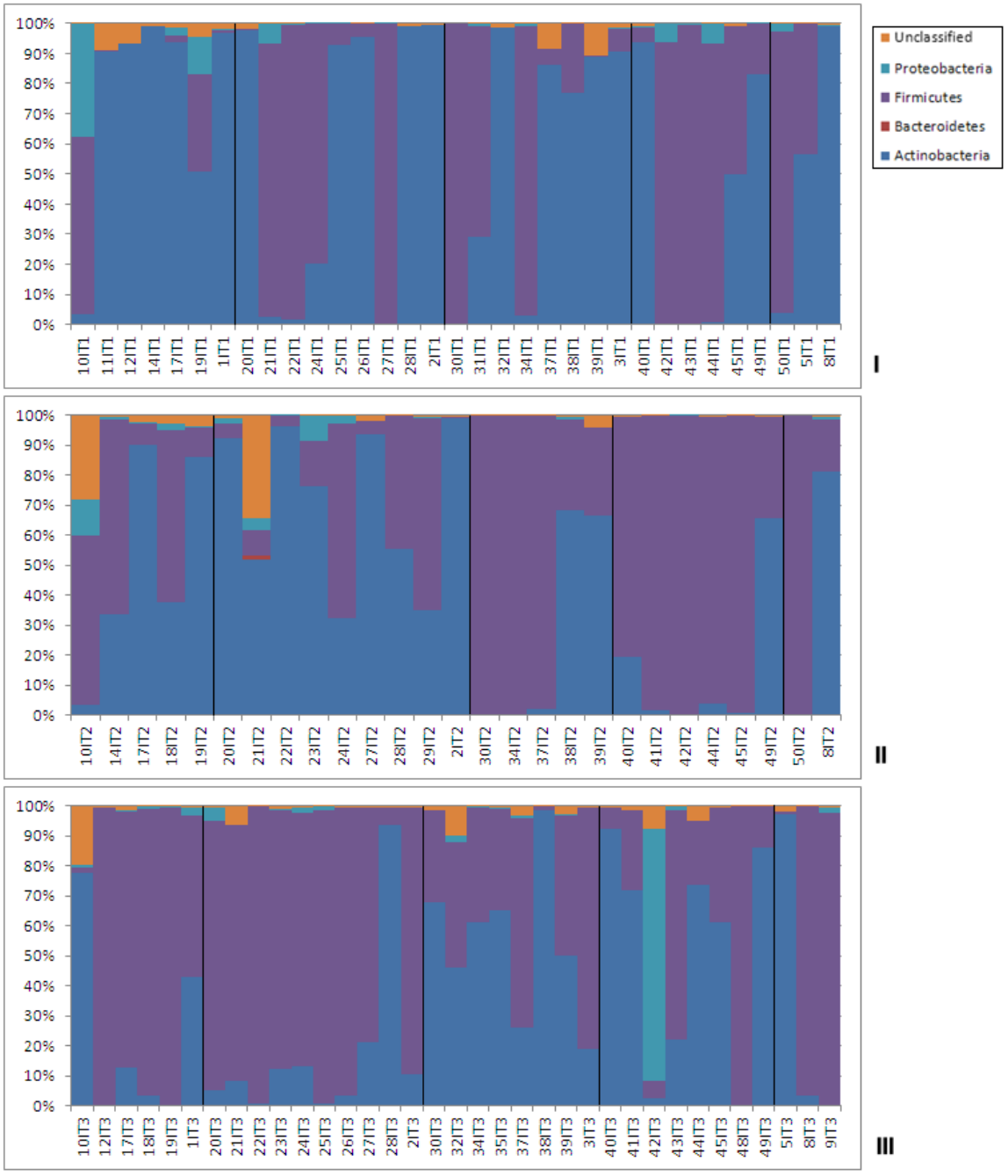
Infants at two weeks showed significantly higher relative abundance of *Corynebacterium* compared to infants at 6 months (p = 0.007, q = 0.6) and 12 months (p = 0.009, q = 1). Four month old infants had significantly fewer *Staphylococcus* than infants at all other time points (T1 (p = 0.001, q = 0.1), T2 (p = 0.02, q = 10), T4 ($p < 0.00001$, $q < 0.00001$), T5 ($p < 0.00001$, $q < 0.00001$)). Two week old and 12 month old infants also carried more *Staphylococcus* than infants at 6 months of age (p = 0.02, q = 1 and p = 0.009, q = 1, respectively). *Dolosigranulum* carriage was significantly lower at the 2 week time point in infants than all other sampling time points (p = 0.003, q = 1 at T2; p = 0.001, q = 0.1 at T3, T4, and T5). Infants carried significantly less *Streptococcus* at 2 months than at 4 months (p = 0.03, q = 1). At the first 2 time points, infants carried significantly less *Moraxella* on average than infants at all other time points ($p < 0.00001$, $q < 0.0001$). Adults were less variable across time points with only *Corynebacterium* being significantly more abundant at 4 months than 6 months (p = 0.02, q = 1), *Dolosigranulum* abundance decreasing significantly from 6 months to 12 months (p = 0.02, q = 1), and *Acinetobacter* having a

higher relative abundance at 2 weeks than at 2 months, 4 months and 6 months ($p < 0.0001$, $q < 0.0001$ for all).

Intra-subject variability was noted, with subjects containing a high proportion of either Actinobacteria or Firmicutes, and only a few subjects with a high proportion of Proteobacteria (Figure 12). The most common phyla of infants do not appear to correlate with those of their primary caregivers, nor do they appear to remain steady over time. Participants in general show no patterns in which Actinobacteria or Firmicutes persist as the main phyla over time. Generally, individuals appear to have only one or two major genera (Figure 13). At two weeks of age, the majority of infants predominantly carry *Corynebacterium, Streptococcus,* and *Staphylococcus*. By age 2 months, *Dolosigranulum* replaces *Streptococcus* as a predominant genus. At 4-6 months of age, *Staphylococcus* virtually disappeared in the infants, with *Corynebacterium, Dolosigranulum* and *Streptococcus* prevalence increasing to dominate the nares flora in most infants; although a few participants revealed predominantly *Rhodococcus* or the class Actinobacteria. At the one year time point, *Staphylococcus* was again a major component of the infant nares microbiota, along with *Corynebacterium, Dolosigranulum,* and *Streptococcus.* An increase in Gammaproteobacteria and *Moraxella* were also noted at the one year sampling.

In adults, *Corynebacterium* was the most prevalent genus, followed by a mix of *Rhodococcus, Dolosigranulum, Staphyloccocus, Streptococcus,* and the classes of Actinobacteria and Gammaproteobacteria. *Corynebacterium* was prevalent in all time points except for the 6 month time point, in which there was a higher abundance of *Staphylococcus.*

Figure 12A: Relative abundance of phyla in the nares microbiota of infants across five time points (I-V: 2 weeks, 2 months, 4 months, 6 months, 12 months, respectively) during the first year of life.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Unclassified
Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
10IT1 11IT1 12IT1 14IT1 17IT1 19IT1 1IT1 20IT1 21IT1 22IT1 24IT1 25IT1 26IT1 27IT1 28IT1 2IT1 30IT1 31IT1 32IT1 34IT1 37IT1 38IT1 39IT1 3IT1 40IT1 42IT1 43IT1 44IT1 45IT1 49IT1 50IT1 5IT1 8IT1
I
10IT2 14IT2 17IT2 18IT2 19IT2 20IT2 21IT2 22IT2 23IT2 24IT2 27IT2 28IT2 29IT2 2IT2 30IT2 34IT2 37IT2 38IT2 39IT2 40IT2 41IT2 42IT2 44IT2 45IT2 49IT2 50IT2 8IT2
II
10IT3 12IT3 17IT3 18IT3 19IT3 1IT3 20IT3 21IT3 22IT3 23IT3 24IT3 25IT3 26IT3 27IT3 28IT3 2IT3 30IT3 32IT3 34IT3 35IT3 37IT3 38IT3 39IT3 3IT3 40IT3 41IT3 42IT3 43IT3 44IT3 45IT3 48IT3 49IT3 5IT3 8IT3 9IT3
III

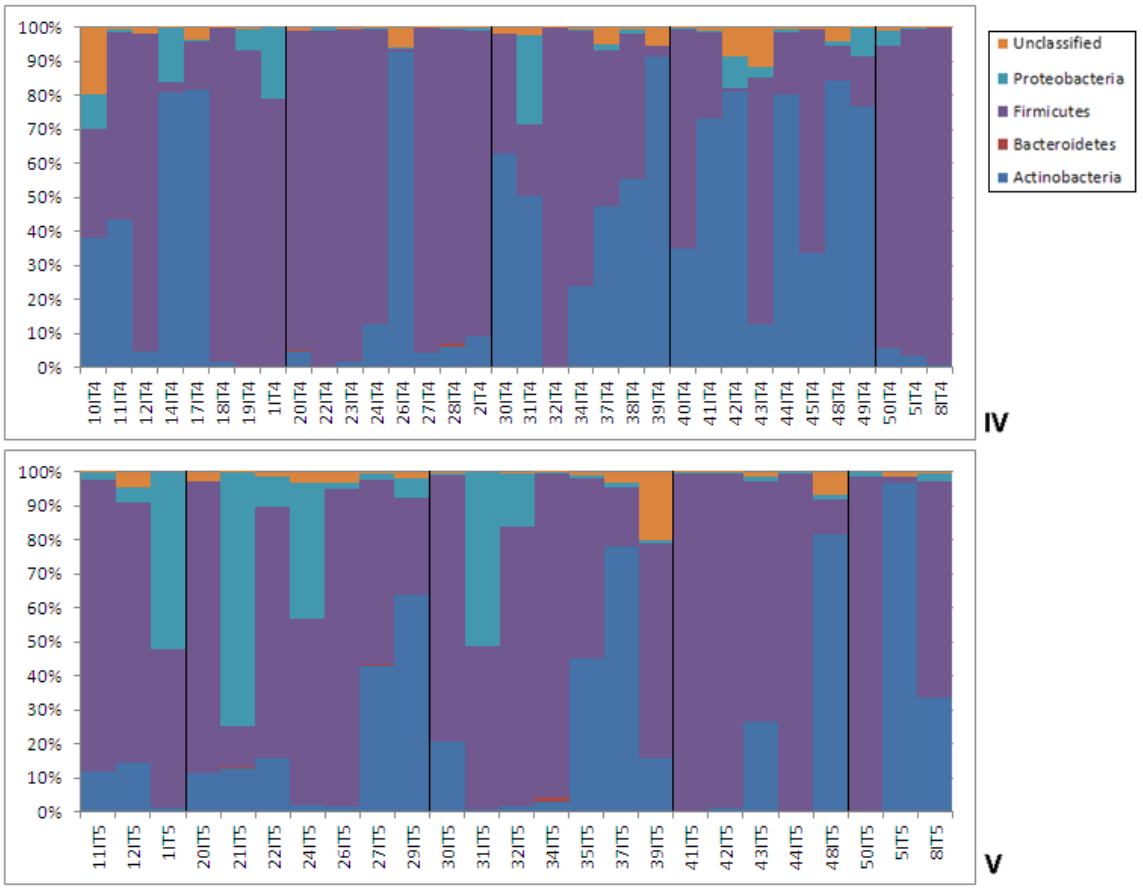
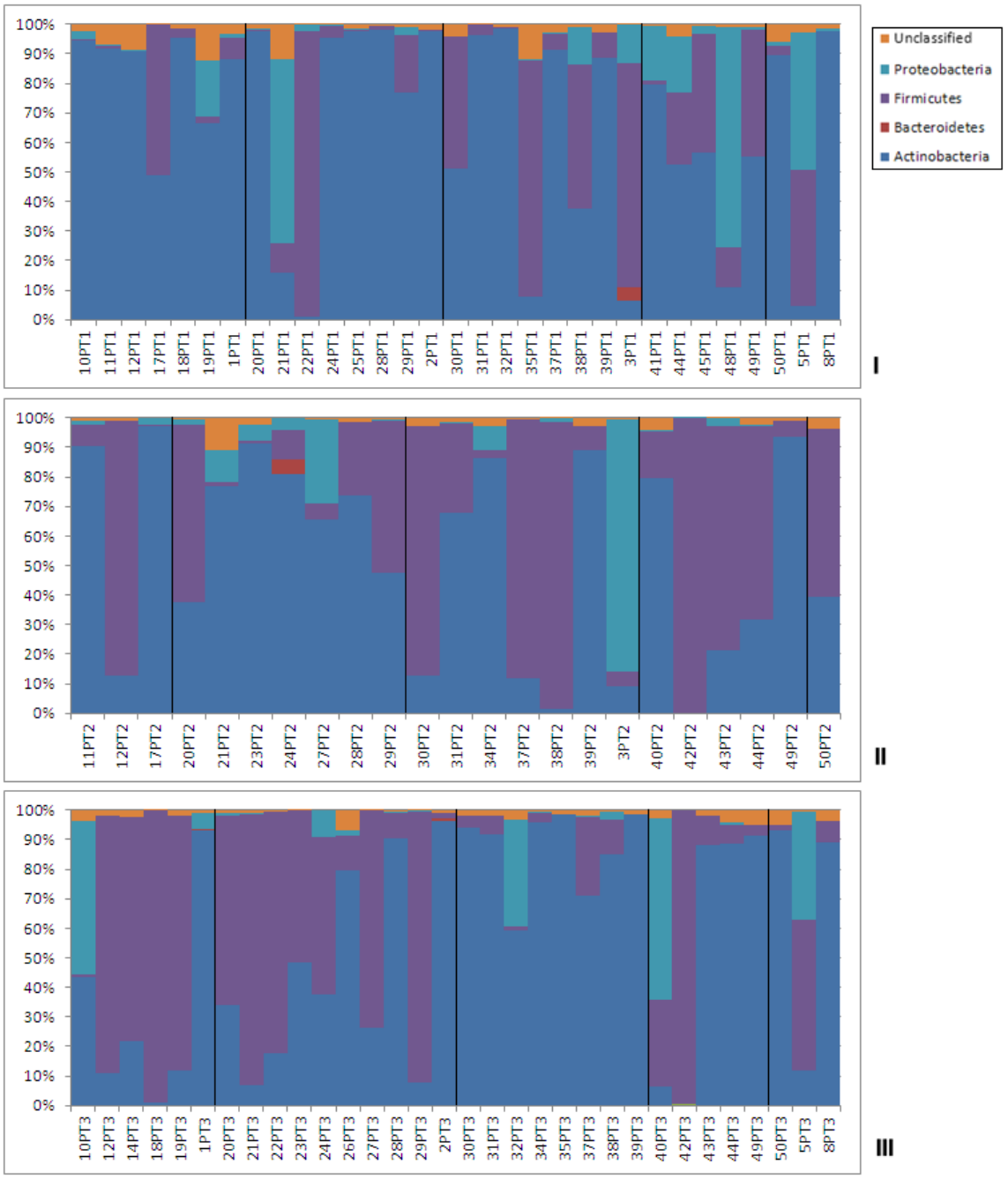

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
10IT4
11IT4
12IT4
14IT4
17IT4
18IT4
19IT4
1IT4
20IT4
22IT4
23IT4
24IT4
26IT4
27IT4
28IT4
2IT4
30IT4
31IT4
32IT4
34IT4
37IT4
38IT4
39IT4
40IT4
41IT4
42IT4
43IT4
44IT4
45IT4
48IT4
49IT4
50IT4
5IT4
8IT4
Unclassified
Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
IV
11IT5
12IT5
1IT5
20IT5
21IT5
22IT5
24IT5
26IT5
27IT5
29IT5
30IT5
31IT5
32IT5
34IT5
35IT5
37IT5
39IT5
41IT5
42IT5
43IT5
44IT5
48IT5
50IT5
5IT5
8IT5
V

Figure 12B: Relative abundance of phyla in the nares microbiota of primary caregivers across five time points (I-V: 2 weeks, 2 months, 4 months, 6 months, 12 months, respectively) during the first year.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Unclassified
Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
10PT1 11PT1 12PT1 17PT1 18PT1 19PT1 1PT1 20PT1 21PT1 22PT1 24PT1 25PT1 28PT1 29PT1 2PT1 30PT1 31PT1 32PT1 35PT1 37PT1 38PT1 39PT1 3PT1 41PT1 44PT1 45PT1 48PT1 49PT1 50PT1 5PT1 8PT1
I
11PT2 12PT2 17PT2 20PT2 21PT2 23PT2 24PT2 27PT2 28PT2 29PT2 30PT2 31PT2 34PT2 37PT2 38PT2 39PT2 3PT2 40PT2 42PT2 43PT2 44PT2 49PT2 50PT2
II
10PT3 12PT3 14PT3 18PT3 19PT3 1PT3 20PT3 21PT3 22PT3 23PT3 24PT3 26PT3 27PT3 28PT3 29PT3 2PT3 30PT3 31PT3 32PT3 34PT3 35PT3 37PT3 38PT3 39PT3 40PT3 42PT3 43PT3 44PT3 49PT3 50PT3 5PT3 8PT3
III

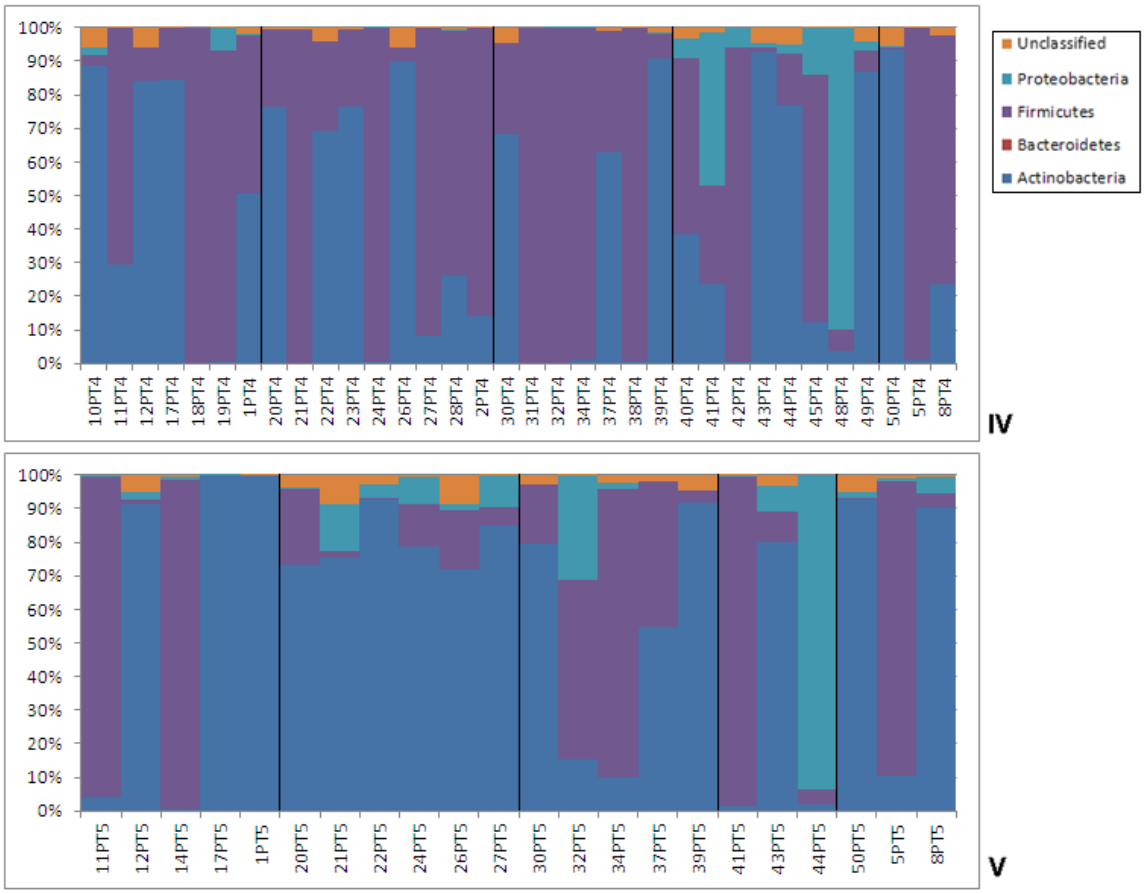
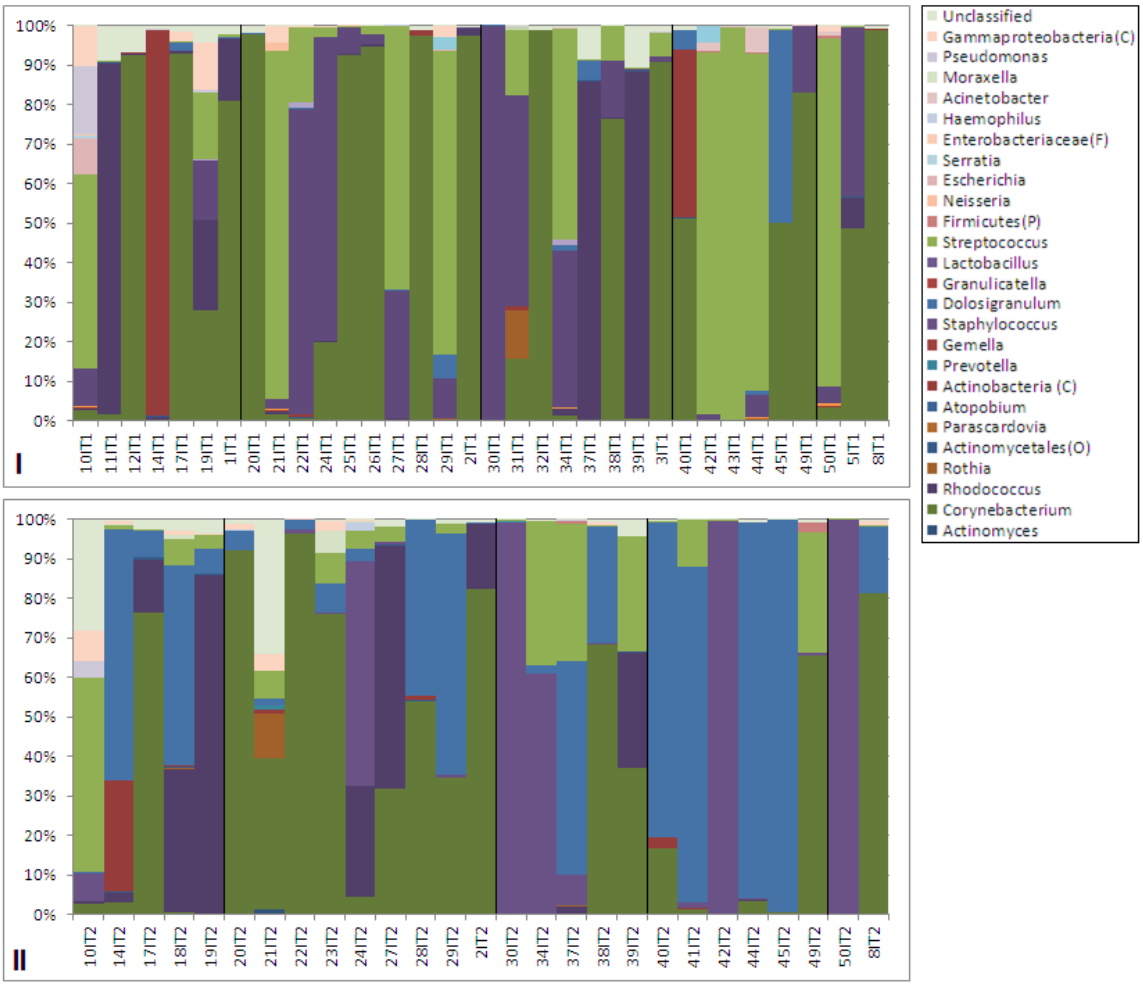

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
10PT4
11PT4
12PT4
17PT4
18PT4
19PT4
1PT4
20PT4
21PT4
22PT4
23PT4
24PT4
26PT4
27PT4
28PT4
2PT4
30PT4
31PT4
32PT4
34PT4
37PT4
38PT4
39PT4
40PT4
41PT4
42PT4
43PT4
44PT4
45PT4
48PT4
49PT4
50PT4
5PT4
8PT4
Unclassified
Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
IV
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
11PT5
12PT5
14PT5
17PT5
1PT5
20PT5
21PT5
22PT5
24PT5
26PT5
27PT5
30PT5
32PT5
34PT5
37PT5
39PT5
41PT5
43PT5
44PT5
50PT5
5PT5
8PT5
V

Figure 13A: Relative abundance of genera existing in the nares microbiota of infants (I) across five time points (I-V: 2 weeks (T1), 2 months (T2), 4 months (T3), 6 months (T4), 12 months (T5)) during the first year of life. Taxa listed with (P), (C), or (F) notations could only be classified to the phylum, class, or family level respectively.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
I
10IT1
11IT1
12IT1
14IT1
17IT1
19IT1
1IT1
20IT1
21IT1
22IT1
24IT1
25IT1
26IT1
27IT1
28IT1
29IT1
2IT1
30IT1
31IT1
32IT1
34IT1
37IT1
38IT1
39IT1
3IT1
40IT1
42IT1
43IT1
44IT1
45IT1
49IT1
50IT1
5IT1
8IT1
II
10IT2
14IT2
17IT2
18IT2
19IT2
20IT2
21IT2
22IT2
23IT2
24IT2
27IT2
28IT2
29IT2
2IT2
30IT2
34IT2
37IT2
38IT2
39IT2
40IT2
41IT2
42IT2
44IT2
45IT2
49IT2
50IT2
8IT2
Unclassified
Gammaproteobacteria(C)
Pseudomonas
Moraxella
Acinetobacter
Haemophilus
Enterobacteriaceae(F)
Serratia
Escherichia
Neisseria
Firmicutes(P)
Streptococcus
Lactobacillus
Granulicatella
Dolosigranulum
Staphylococcus
Gemella
Prevotella
Actinobacteria (C)
Atopobium
Parascardovia
Actinomycetales(O)
Rothia
Rhodococcus
Corynebacterium
Actinomyces

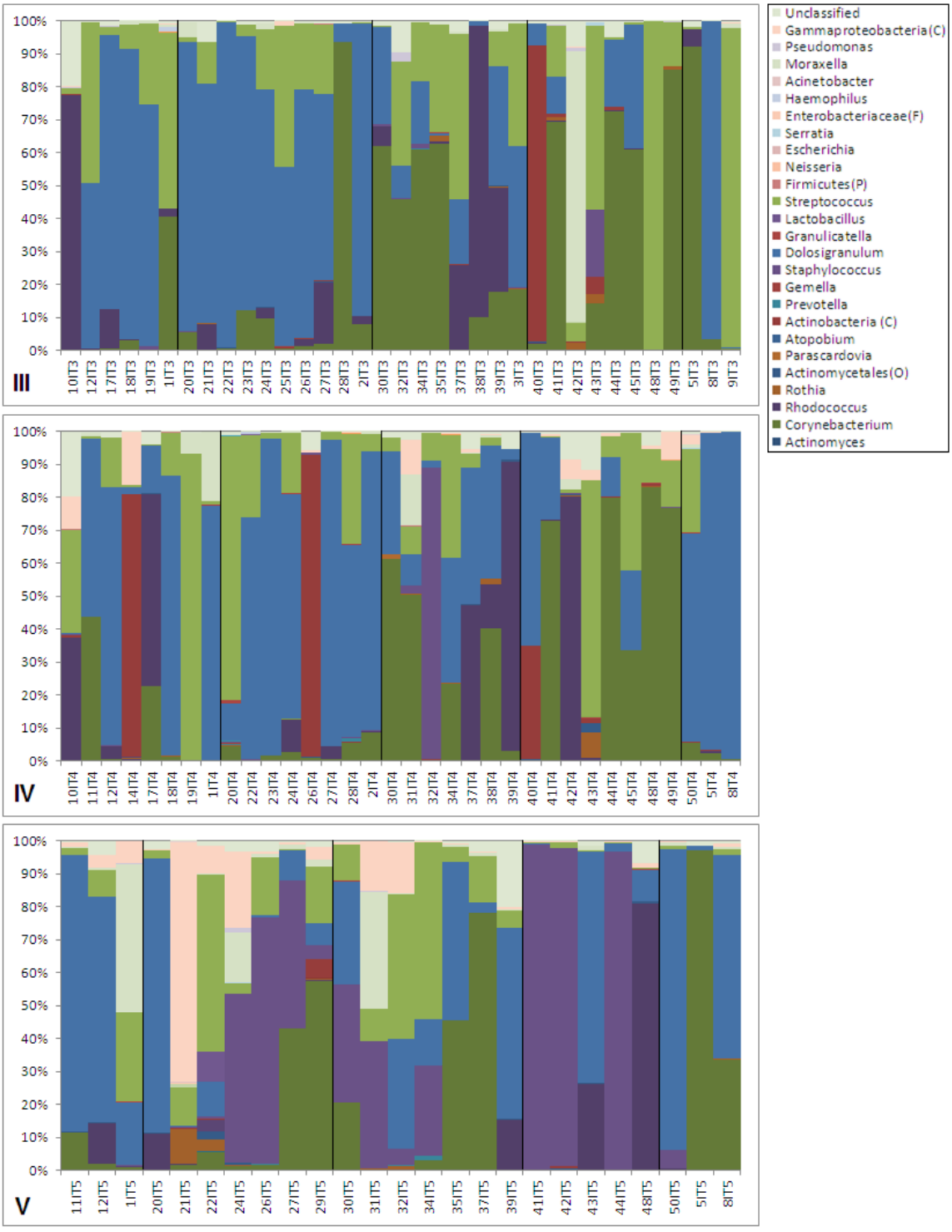
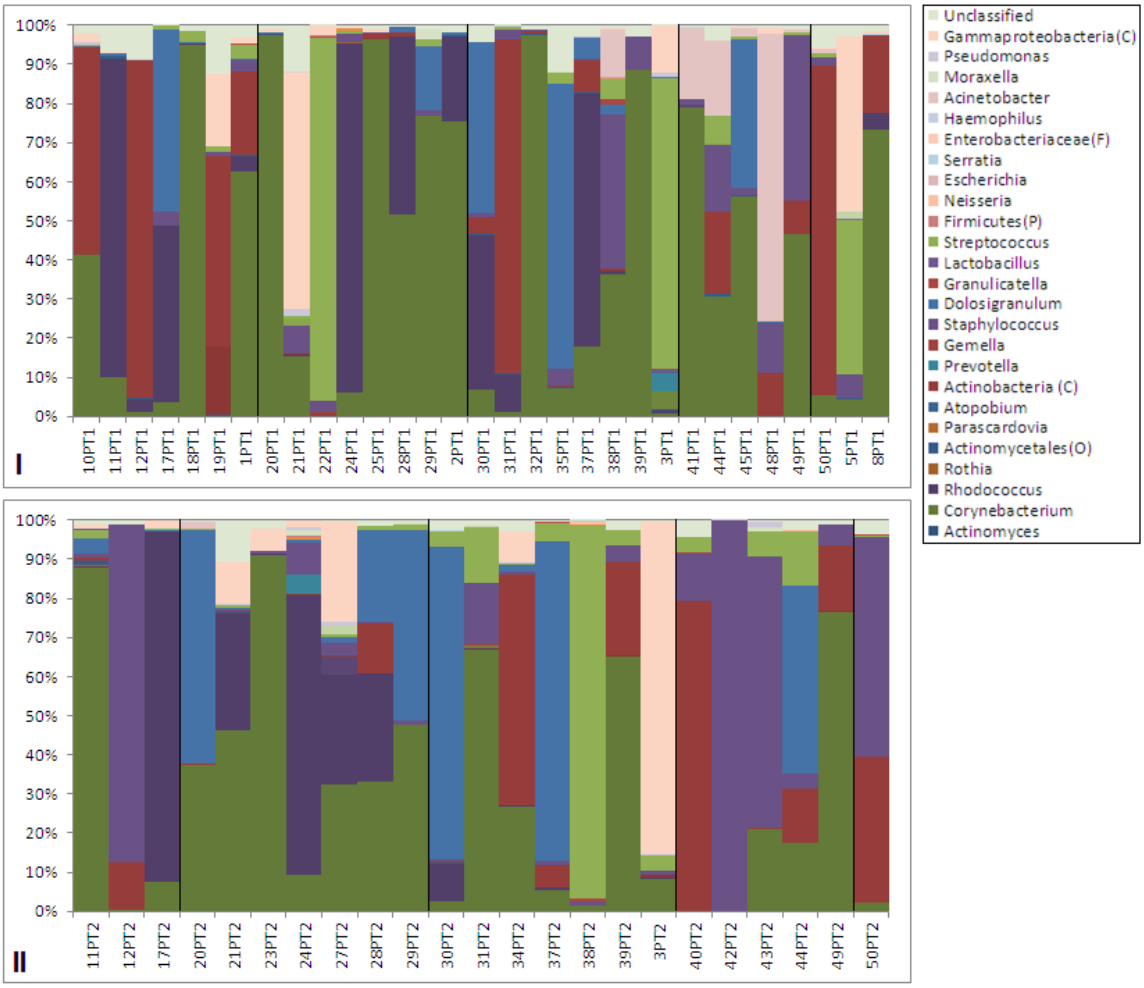
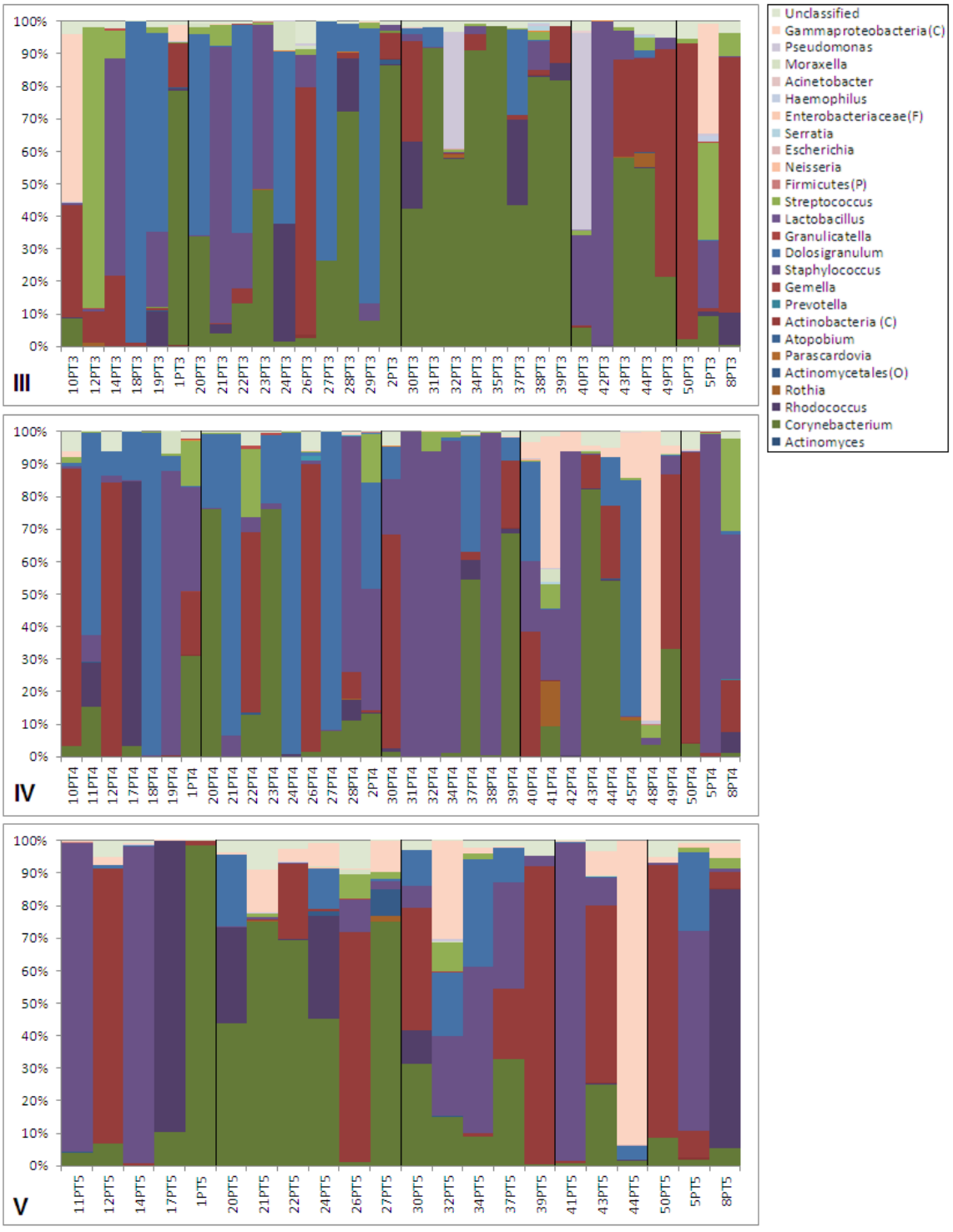

III
IV
V
Unclassified
Gammaproteobacteria(C)
Pseudomonas
Moraxella
Acinetobacter
Haemophilus
Enterobacteriaceae(F)
Serratia
Escherichia
Neisseria
Firmicutes(P)
Streptococcus
Lactobacillus
Granulicatella
Dolosigranulum
Staphylococcus
Gemella
Prevotella
Actinobacteria (C)
Atopobium
Parascardovia
Actinomycetales(O)
Rothia
Rhodococcus
Corynebacterium
Actinomyces

Figure 13B: Relative abundance of genera in the nares microbiota of primary caregivers (C) across five time points (I-V: 2 weeks (T1), 2 months (T2), 4 months (T3), 6 months (T4), 12 months (T5)) during the first year of life. Taxa listed with (P), (C), or (F) notations could only be classified to the phylum, class, or family level, respectively.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
10PT1
11PT1
12PT1
17PT1
18PT1
19PT1
1PT1
20PT1
21PT1
22PT1
24PT1
25PT1
28PT1
29PT1
2PT1
30PT1
31PT1
32PT1
35PT1
37PT1
38PT1
39PT1
3PT1
41PT1
44PT1
45PT1
48PT1
49PT1
50PT1
5PT1
8PT1
I
11PT2
12PT2
17PT2
20PT2
21PT2
23PT2
24PT2
27PT2
28PT2
29PT2
30PT2
31PT2
34PT2
37PT2
38PT2
39PT2
3PT2
40PT2
42PT2
43PT2
44PT2
49PT2
50PT2
II
Unclassified
Gammaproteobacteria(C)
Pseudomonas
Moraxella
Acinetobacter
Haemophilus
Enterobacteriaceae(F)
Serratia
Escherichia
Neisseria
Firmicutes(P)
Streptococcus
Lactobacillus
Granulicatella
Dolosigranulum
Staphylococcus
Gemella
Prevotella
Actinobacteria (C)
Atopobium
Parascardovia
Actinomycetales(O)
Rothia
Rhodococcus
Corynebacterium
Actinomyces


III
10PT3
12PT3
14PT3
18PT3
19PT3
1PT3
20PT3
21PT3
22PT3
23PT3
24PT3
26PT3
27PT3
28PT3
29PT3
2PT3
30PT3
31PT3
32PT3
34PT3
35PT3
37PT3
38PT3
39PT3
40PT3
42PT3
43PT3
44PT3
49PT3
50PT3
5PT3
8PT3
IV
10PT4
11PT4
12PT4
17PT4
18PT4
19PT4
1PT4
20PT4
21PT4
22PT4
23PT4
24PT4
26PT4
27PT4
28PT4
2PT4
30PT4
31PT4
32PT4
34PT4
37PT4
38PT4
39PT4
40PT4
41PT4
42PT4
43PT4
44PT4
45PT4
48PT4
49PT4
50PT4
5PT4
8PT4
V
11PT5
12PT5
14PT5
17PT5
1PT5
20PT5
21PT5
22PT5
24PT5
26PT5
27PT5
30PT5
32PT5
34PT5
37PT5
39PT5
41PT5
43PT5
44PT5
50PT5
5PT5
8PT5
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Unclassified
Gammaproteobacteria(C)
Pseudomonas
Moraxella
Acinetobacter
Haemophilus
Enterobacteriaceae(F)
Serratia
Escherichia
Neisseria
Firmicutes(P)
Streptococcus
Lactobacillus
Granulicatella
Dolosigranulum
Staphylococcus
Gemella
Prevotella
Actinobacteria (C)
Atopobium
Parascardovia
Actinomycetales(O)
Rothia
Rhodococcus
Corynebacterium
Actinomyces

Dolosigranulum levels were elevated at all time points except 12 months. At each time period, *Rhodococcus, Streptococcus,* and Gammaproteobacteria are predominant in at least one or two participants.

3.2.7 Core Microbiome of the Human Nares

Tables 6 and 7 summarize the taxa that best make up the detected nares core microbiome in this study. Table 6 highlights the median relative abundance of genera found in the nares microbiome with an interquartile range (IQR) greater than 0.0-0.0. These organisms were present in the anterior nares of at least 25% of study participants during at least one time point. Firmicutes tended to predominate in infants, whereas Actinobacteria tended to predominate in adult caregivers. Study participants appeared to host only a few truly abundant genera in the nares; otherwise, the majority of other taxa were rare or subject specific. Infants also tended to exhibit higher median relative abundances for *Dolosogranulum* and *Streptococcus* than did their primary caregivers.

Table 7 shows the frequency of dominant genera carriage in the infants and their primary caregivers over the first year of life. *Corynebacterium* was present in almost all study participants across all time points. *Streptococcus* was found in over 75% of participants at each time point, while *Rhodococcus, Staphylococcus,* and *Dolosigranulum* were found in over 50% of participants at each time point. *Propionibacterium* was more prevalent in adults than in infants, present in 30-45% of caregivers at each time point and fewer than 10% of infants. Members of the class Actinobacteria were more likely to be found in caregivers than in infants (68-91% and 49-63%, respectively). *Staphylococcus* was generally more prevalent in caregivers (84-100%) than in infants (65-96%), while *Dolosigranulum* was more likely

Table 6A: Median relative abundance of dominant taxa (Interquartile range in parentheses) found in nares swabs from healthy infants from 5 time points during the first year of life

Taxon*	2 Weeks	2 Months	4 Months	6 Months	12 Months
Actinobacteria	83.2 (3.2-93.8)	37.7 (2.8-78.8)	21.1 (3.6-66.6)	28.7 (4.5-70.6)	13.0 (1.5-33.8)
Corynebacterium	28.1 (0.6-92.5)	16.6 (0.6-66.9)	8.1 (0.5-53.3)	2.4 (0.4-31.1)	1.6 (0.1-20.7)
Rhodococcus	0.2 (0.0-0.7)	0.0 (0.0-7.8)	0.2 (0.0-4.1)	0.0 (0.0-3.1)	0.0 (0.0-0.1)
Rothia	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)
Actinomycetales (Order)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)
Propionibacterium	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Actinobacteria (Class)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.2)	0.0 (0.0-0.1)
Firmicutes	16.6 (0.7-90.8)	56.6 (12.5-96.5)	76.3 (28.5-92)	69.1 (21.8-93.4)	70.7 (48.3-86.0)
Staphylococcus	1.5 (0.0-15.4)	0.3 (0.0-4.1)	0.0 (0.0-0.2)	0.0 (0.0-0.2)	0.8 (0.1-38.6)
Dolosigranulum	0.1 (0.0-0.3)	6.0 (0.4-52.2)	29.6 (0.8-72.9)	34.5 (2.4-72.0)	10.5 (1.4-57.7)
Streptococcus	0.3 (0.1-19.1)	1.3 (0.1-7.3)	12.6 (0.7-32.1)	5.8 (0.9-23.3)	4.6 (1.0-14.0)
Proteobacteria	0.1 (0.0-0.6)	0.1 (0.0-1.0)	0.3 (0.0-0.8)	0.5 (0.2-2.8)	1.7 (0.6-5.9)
Moraxella	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.4)	0.0 (0.0-0.3)	0.3 (0.0-1.1)
Pseudomonas	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Gammaproteobacteria (Class)	0.0 (0.0-0.2)	0.0 (0.0-0.6)	0.0 (0.0-0.0)	0.2 (0.0-1.2)	0.6 (0.3-3.9)
Unclassified	0.3 (0.0-1.3)	0.6 (0.1-1.4)	0.6 (0.3-1.4)	0.5 (0.2-2.4)	0.6 (0.2-2.6)

*Taxa with an interquartile range of 0.0-0.0 for both groups during all time points are omitted

Table 6B: Median relative abundance of dominant taxa (Interquartile range in parentheses) found in nares swabs from primary caregivers from 5 time points over the course of one year

Taxon*	2 Weeks	2 Months	4 Months	6 Months	12 Months
Actinobacteria	88.3 (49.9-95.5)	65.6 (17.1-83.7)	65.1 (16.3-91.5)	27.7 (1.1-76.9)	77.2 (11.8-91.2)
Corynebacterium	24.2 (4.1-73.9)	21.0 (4.1-56.8)	23.9 (2.4-61.6)	3.3 (0.1-14.8)	9.6 (2.5-41.1)
Rhodococcus	0.1 (0.0-5.4)	0.2 (0.0-5.2)	0.1 (0.0-1.6)	0.0 (0.0-0.3)	0.1 (0.0-0.3)
Rothia	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Actinomycetales (Order)	0.0 (0.0-0.3)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Propionibacterium	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.1)
Actinobacteria (Class)	0.6 (0.0-13.2)	0.4 (0.1-13.3)	1.5 (0.1-23.4)	0.5 (0.0-21.9)	1.1 (0.0-34.0)
Firmicutes	5.7 (1.1-41.6)	25.1 (5.2-70.7)	12.0 (3.5-74.1)	49.9 (15.0-93.4)	10.8 (2.2-51.0)
Staphylococcus	1.3 (0.1-3.8)	1.2 (0.6-10.2)	1.2 (0.2-11.8)	5.0 (0.3-42.7)	2.8 (0.4-30.4)
Dolosigranulum	0.0 (0.0-1.0)	0.1 (0.0-13.6)	0.2 (0.0-12.8)	2.9 (0.1-28.8)	0.2 (0.0-11.1)
Streptococcus	0.5 (0.0-1.9)	1.0 (0.1-3.9)	0.6 (0.0-1.9)	0.3 (0.1-1.3)	0.1 (0.0-1.2)
Proteobacteria	0.8 (0.2-7.9)	0.6 (0.1-3.4)	0.1 (0.0-1.3)	0.1 (0.0-1.8)	1.9 (0.3-7.0)
Moraxella	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Pseudomonas	0.0 (0.0-0.1)	0.0 (0.0-0.2)	0.0 (0.0-0.1)	0.0 (0.0-0.0)	0.0 (0.0-0.1)
Gammaproteobacteria (Class)	0.2 (0.0-0.8)	0.0 (0.0-1.9)	0.0 (0.0-0.0)	0.1 (0.0-1.0)	1.4 (0.3-6.6)
Unclassified	1.3 (0.5-3.5)	1.1 (0.4-2.5)	1.4 (0.7-2.4)	0.4 (0.1-4.0)	1.4 (0.2-3.5)

*Taxa with an interquartile range of 0.0-0.0 for both groups during all time points are omitted

Table 7A : Frequency of dominant taxa (%) in nares swabs from healthy infants from 5 time points over the first year of life

Time point	2 Weeks		2 Months		4 Months		6 Months		12 Months	
N	33		27		35		34		25	
Actinobacteria										
Actinomyces	0	(0)	2	(7)	2	(6)	1	(3)	3	(12)
Corynebacterium	31	(94)	26	(96)	35	(100)	34	(100)	24	(96)
Rhodococcus	26	(79)	16	(59)	25	(71)	22	(65)	15	(60)
Rothia	6	(18)	5	(19)	15	(43)	17	(50)	14	(56)
Actinomycetales(Order)	18	(55)	18	(67)	19	(54)	18	(53)	11	(44)
Propionibacterium	1	(3)	2	(7)	1	(3)	1	(3)	2	(8)
Bifidobacterium	2	(6)	2	(7)	0	(0)	1	(3)	3	(12)
Actinobacteria (Class)	21	(64)	17	(63)	17	(49)	17	(50)	13	(52)
Bacteroidetes										
Prevotella	1	(3)	2	(7)	3	(9)	4	(12)	6	(24)
Firmicutes										
Gemella	7	(21)	1	(4)	2	(6)	5	(15)	3	(12)
Staphylococcus	28	(85)	22	(81)	30	(86)	22	(65)	24	(96)
Dolosigranulum	23	(70)	26	(96)	30	(86)	31	(91)	24	(96)
Granulicatella	0	(0)	0	(0)	6	(17)	9	(26)	8	(32)
Lactobacillus	3	(9)	1	(4)	0	(0)	9	(26)	3	(12)
Streptococcus	28	(85)	25	(93)	35	(100)	32	(94)	25	(100)
Anaerococcus	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)
Finegoldia	0	(0)	0	(0)	1	(3)	0	(0)	0	(0)
Peptoniphilus	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)
Clostridiales(Order)	1	(3)	0	(0)	1	(3)	1	(3)	0	(0)
Veillonella	0	(0)	0	(0)	4	(11)	4	(12)	0	(0)
Unclassified Firmicutes	6	(18)	4	(15)	2	(6)	9	(26)	4	(16)
Proteobacteria										
Ralstonia	2	(6)	0	(0)	1	(3)	2	(6)	2	(8)
Neisseria	2	(6)	1	(4)	0	(0)	3	(9)	4	(16)
Escherichia	2	(6)	2	(7)	0	(0)	1	(3)	2	(8)
Serratia	3	(9)	2	(7)	2	(6)	1	(3)	2	(8)
Haemophilus	1	(3)	2	(7)	2	(6)	1	(3)	0	(0)
Acinetobacter	9	(26)	3	(11)	0	(0)	0	(0)	3	(12)
Moraxella	5	(15)	13	(48)	18	(51)	19	(56)	18	(72)
Pseudomonas	15	(45)	6	(22)	15	(43)	4	(12)	8	(32)
Xanthomonadaceae(Family)	2	(6)	1	(4)	0	(0)	1	(3)	2	(8)
Gammaproteobacteria(Class)	18	(55)	14	(52)	10	(29)	22	(65)	25	(100)
Unclassified Proteobacteria	3	(9)	0	(0)	3	(9)	1	(3)	1	(4)
Unclassified	26	(79)	24	(89)	35	(100)	33	(97)	21	(84)

*Taxa observed less consistently (in fewer than 10 individuals (total)) are not reported

Table 7B : Frequency of dominant taxa (%) in nares swabs from healthy caregivers from 5 time points over the course of one year

Time point	2 Weeks		2 Months		4 Months		6 Months		12 Months	
N	31		23		32		34		22	
Actinobacteria										
Actinomyces	2	(6)	0	(0)	0	(0)	0	(0)	0	(0)
Corynebacterium	31	(100)	21	(91)	31	(97)	30	(88)	22	(100)
Rhodococcus	24	(77)	16	(70)	25	(78)	18	(53)	17	(77)
Rothia	12	(39)	8	(35)	9	(28)	6	(18)	6	(27)
Actinomycetales (Order)	23	(74)	7	(30)	15	(47)	10	(29)	9	(41)
Propionibacterium	11	(35)	7	(30)	13	(41)	14	(41)	10	(45)
Bifidobacterium	1	(3)	1	(4)	2	(6)	1	(3)	0	(0)
Actinobacteria (Class)	26	(84)	20	(87)	27	(84)	23	(68)	20	(91)
Bacteroidetes										
Prevotella	2	(6)	1	(4)	1	(3)	2	(6)	0	(0)
Firmicutes										
Gemella	1	(3)	1	(4)	3	(9)	3	(9)	0	(0)
Staphylococcus	26	(84)	23	(100)	32	(100)	33	(97)	21	(95)
Dolosigranulum	19	(61)	16	(70)	22	(69)	29	(85)	19	(86)
Granulicatella	4	(13)	4	(17)	2	(6)	3	(9)	2	(9)
Lactobacillus	2	(6)	0	(0)	1	(3)	1	(3)	0	(0)
Streptococcus	26	(84)	23	(100)	30	(94)	32	(94)	17	(77)
Anaerococcus	3	(10)	4	(17)	4	(13)	5	(15)	0	(0)
Finegoldia	3	(10)	2	(9)	3	(9)	3	(9)	1	(5)
Peptoniphilus	5	(16)	3	(13)	7	(22)	5	(15)	3	(14)
Clostridiales (Order)	1	(3)	2	(9)	1	(3)	4	(12)	1	(5)
Veillonella	0	(0)	1	(4)	1	(3)	2	(6)	0	(0)
Unclassified Firmicutes	4	(13)	4	(17)	3	(9)	8	(24)	2	(9)
Proteobacteria										
Ralstonia	5	(16)	5	(22)	2	(6)	1	(3)	2	(9)
Neisseria	0	(0)	1	(4)	1	(3)	0	(0)	1	(5)
Escherichia	1	(3)	2	(9)	1	(3)	3	(9)	3	(14)
Serratia	8	(26)	7	(30)	6	(19)	2	(6)	4	(18)
Haemophilus	0	(0)	1	(4)	2	(6)	1	(3)	1	(5)
Acinetobacter	9	(29)	3	(13)	2	(6)	0	(0)	0	(0)
Moraxella	10	(32)	8	(35)	9	(28)	8	(24)	5	(23)
Pseudomonas	15	(48)	14	(61)	16	(50)	8	(24)	9	(41)
Xanthomonadaceae (Family)	3	(10)	3	(13)	5	(16)	0	(0)	2	(9)
Gammaproteobacteria(Class)	22	(71)	14	(61)	7	(22)	22	(65)	21	(95)
Unclassified Proteobacteria	5	(16)	2	(9)	1	(3)	5	(15)	1	(5)
Unclassified	31	(100)	20	(87)	30	(94)	28	(82)	19	(86)

*Taxa observed less consistently (in fewer than 10 individuals (total)) are not reported

to be found in infants (70-96% *vs*. 61-86% in caregivers). The lower abundance organisms, *Ralstonia* and *Serratia* were more commonly found in caregivers than in infants. When comparing time points, *Rothia* carriage increased from 18% to 56% over the 12 month period in infants. *Granulicatella* was absent in all infants during the 2 weeks and 2 month time periods, but reached 17% carriage at 4 months and increased to 32% at 12 months. *Acinetobacter* was absent during the summer months (4 and 6 month time points). *Moraxella* showed a steady increase across the 5 time points, with 15% carriers at 2 weeks, increasing to 48% after 2 months, and finally reaching 72% of infants carrying the organism at 12 months. In caregivers, *Dolosigranulum* appeared to increase over the one year period from 61% to 86% of participants carrying the organism. The rate of *Streptococcus* carriage appeared to decrease during the winter months (time points 1 and 5).

3.2.8 Potentially Pathogenic Taxa Detected in the Nares

Potentially pathogenic taxa were found in the anterior nares, including *Staphylococcus aureus, Streptococcus pneumoniae, Moraxella catarrhalis*¸ and *Pseudomonas aeruginosa*. Table 8 summarizes the average relative abundances and number of carriers for each potential pathogen at each time point. Carriage of *S. aureus* and *S. pneumoniae* appeared to be inversely related to each other. It is notable that *S. pneumoniae* carriage is much more common in infants than in caregivers. During the summer months (4 months and 6 months), relative abundance of *S. aureus* decreased in infants, whereas *S. pneumoniae* increased. In contrast, adults had extremely low average relative abundance detected for *S. pneumoniae*; and *S. aureus* doesn't exhibit any decreasing trend during the summer months. In both adults and infants, carriage of *S. aureus* increased at the 12

Table 8: Relative abundance and carriage frequency of potentially pathogenic organisms in the nares microbiota of infants and their primary caregivers over the course of one year

Time point	No. of Participants	*Staphylococcus aureus*		*Streptococcus pneumoniae*		*Moraxella catarrhalis*		*Pseudomona aeruginosa*	
Two Weeks		Mean	N (%)	Mean	N (%)	Mean	N (%)	Mean	N (%
Caregivers	31	0.2	11 (35)	0.1	6 (19)	<0.1	1 (3)	0.1	8 (2
Infants	33	6.4	16 (49)	<0.1	9 (27)	0	0 (0)	0.5	12 (3
Two Months									
Caregivers	23	8.3	13 (57)	<0.1	4 (17)	0	0 (0)	0.2	14 (6
Infants	27	13.2	13 (48)	0.8	5 (19)	<0.1	5 (19)	0.2	6 (2
Four Months									
Caregivers	32	3.3	13 (41)	<0.1	2 (6)	0	0 (0)	3.1	13 (4
Infants	35	0.1	12 (34)	3.3	14 (40)	0.1	7 (20)	0.1	13 (3
Six Months									
Caregivers	34	18.8	17 (50)	<0.1	4 (12)	0	0 (0)	0.1	5 (1
Infants	34	2.6	10 (29)	3	15 (44)	0.1	7 (21)	<0.1	1 (3
Twelve Months									
Caregivers	22	18.3	16 (73)	0.2	2 (9)	<0.1	1 (5)	0.1	5 (2
Infants	25	23.1	20 (80)	0.2	10 (40)	0.1	7 (28)	<0.1	2 (8

month time point coinciding with the end of winter-spring season (January-April). At the 6 month time point, caregivers had a significantly higher relative abundance of *S. aureus* than infants ($p = 0.02$, two-tailed student's t-test). *S. aureus* carriage in infants at 12 months was significantly higher than at 2 weeks ($p = 0.04$), 4 months ($p = 0.003$), and 6 months ($p = 0.009$). In caregivers, relative *S. aureus* abundance during the 2 week time point was significantly lower than at 6 months ($p = 0.007$) and 12 months ($p = 0.02$). *S. pneumoniae* carriage was significantly higher in infants than caregivers during the 4 month time point ($p = 0.03$). Infants carried significantly more *S. pneumoniae* during time point 3 than during either time points 1 ($p = 0.04$) and 5 ($p = 0.04$).

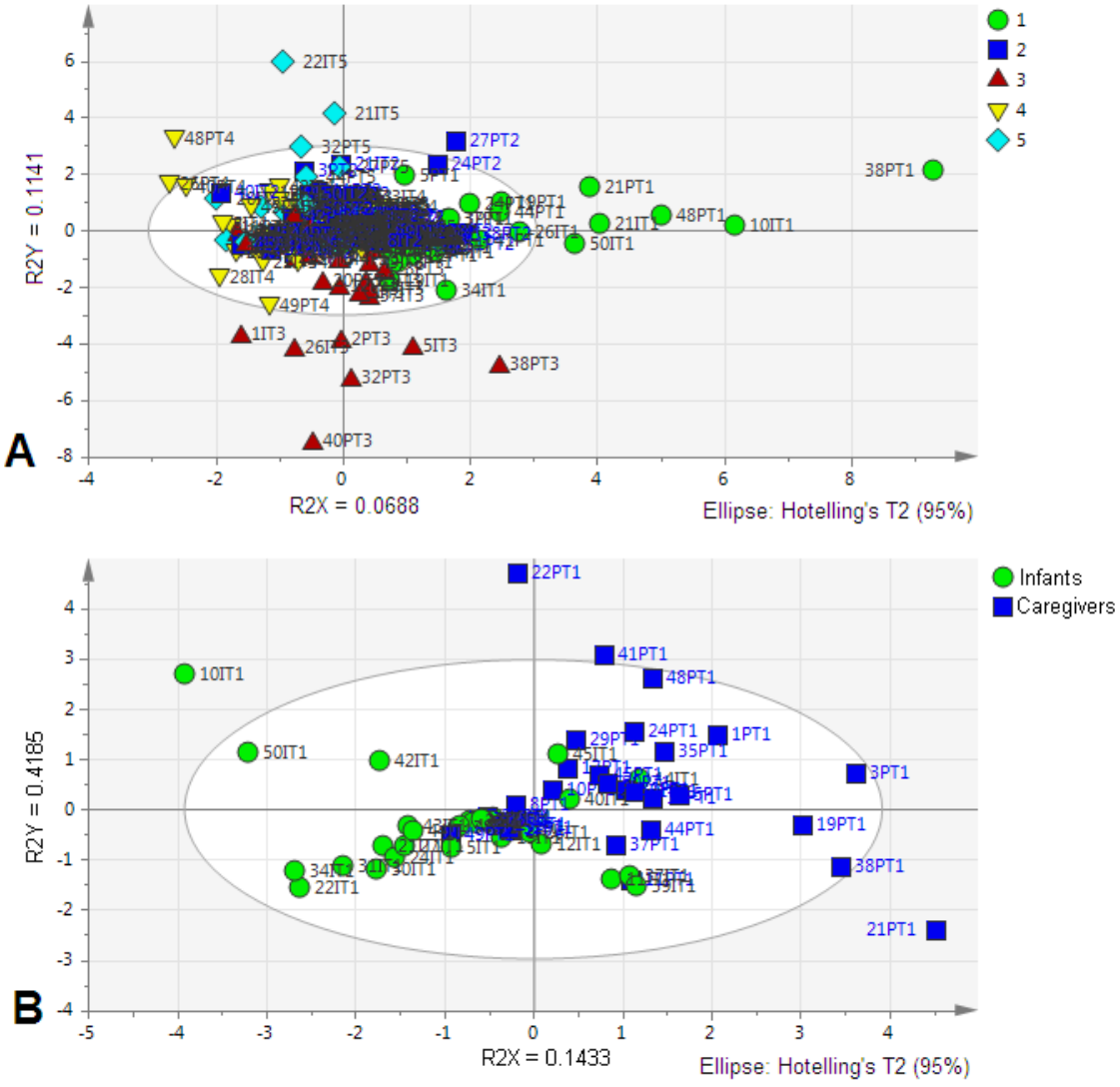
M. catarrhalis carriage was virtually nonexistent in caregivers, but beginning at the 2 month time point, approximately 20% of infants carried the organism, albeit in low abundance. *P. aeruginosa* carriage in caregivers ranged from 15-61%, and appeared to be variable over the course of the year. In infants, during the first 4 months of life, carriage ranged from 22-37% and decreased drastically during the latter two time points. Lastly, during the 4 month time point, 2 participants exhibited high relative abundance of *P. aeruginosa*, and both participants reported a flu-like illness at the time.

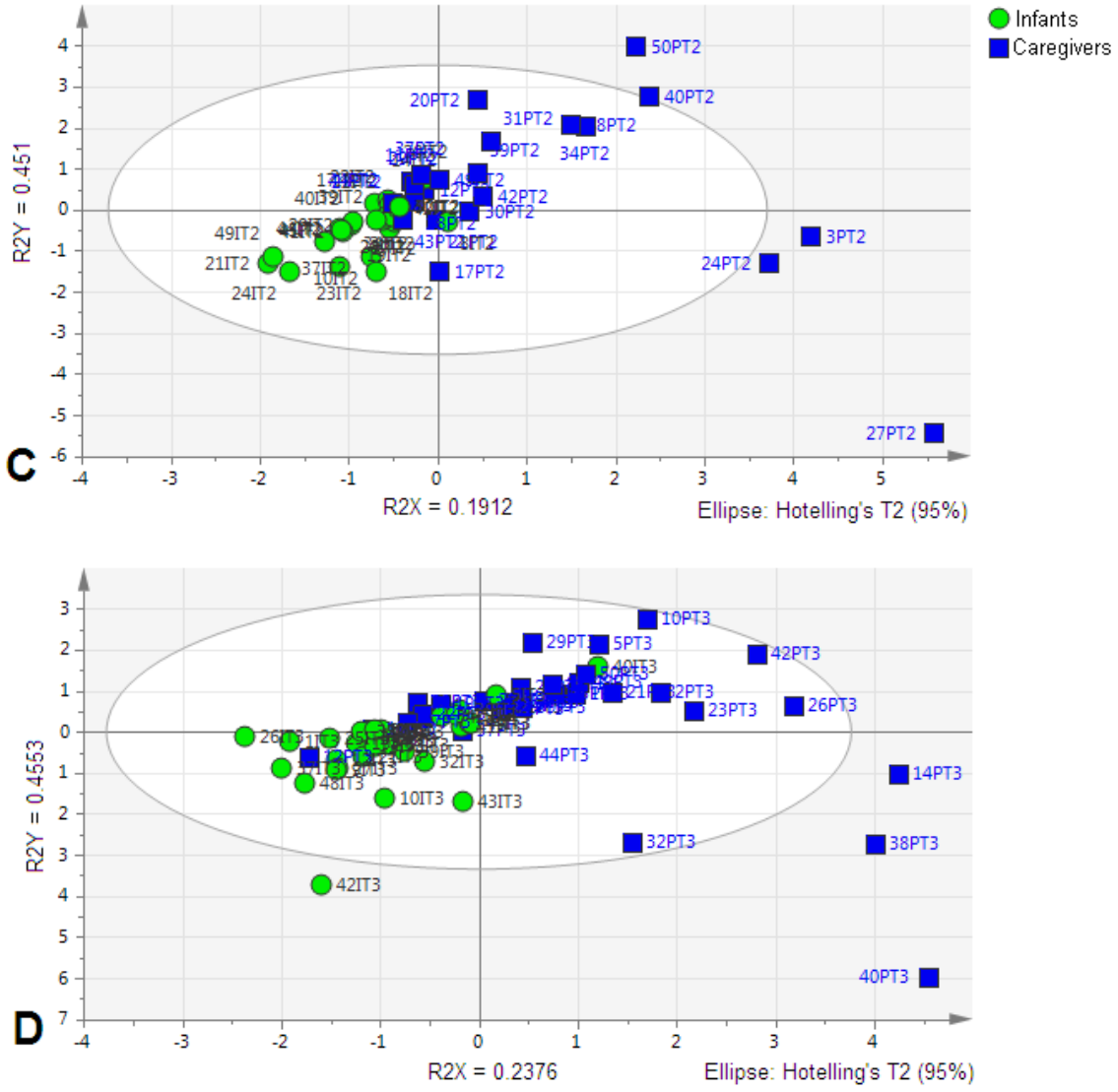
3.2.9 Multivariate Analysis

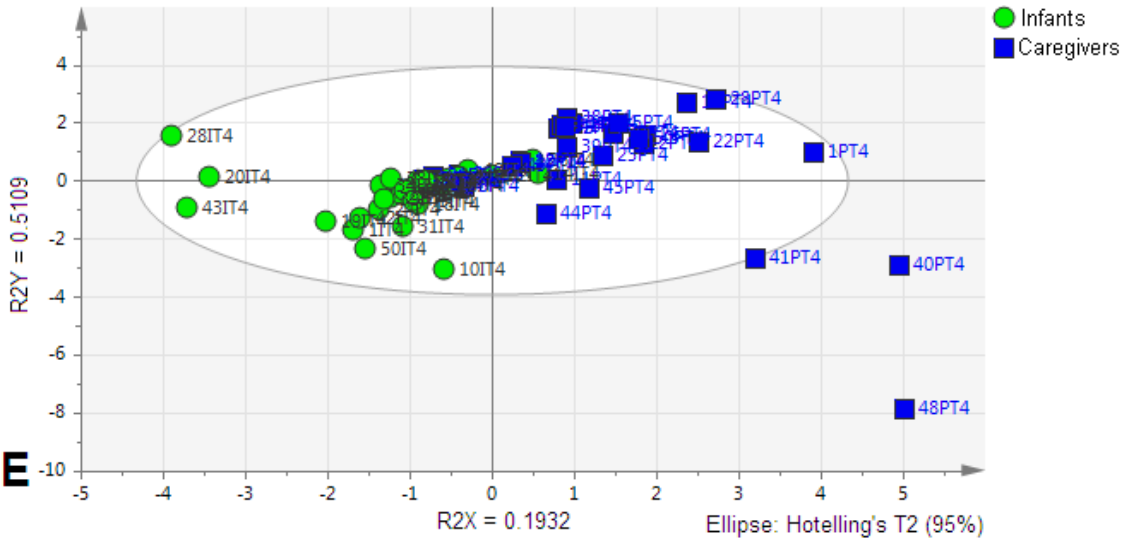
A Partial Least Squares Discriminant Analysis (PLS-DA) was performed on each group to determine the relationship between the microbial compositions of each participant nares sample (Figure 14). When considering all 5 time points on the same plot, there is much observed overlap between participants during each time point; however, there is also visible grouping of each time point. Participants from the 2 week time point (1) showed a more

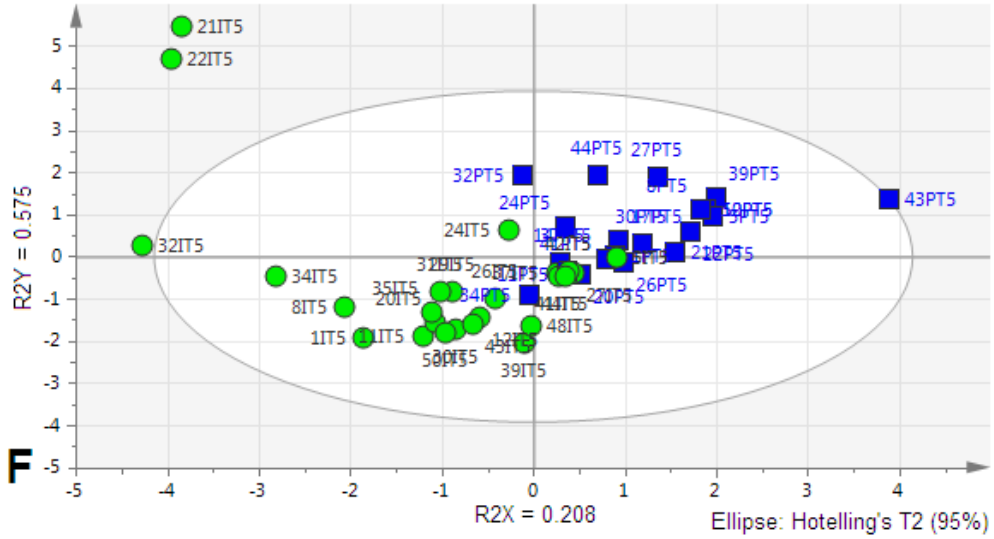
distinct separation from the other 4 time points. Participants also showed grouping by season, with the winter time points (1, 2, 5) located in the top right hand side of the plot, and the summer time points (3, 4) located mainly in the bottom left hand side of the plot. When observing plots of each individual time point, there was distinct grouping observed for infants and caregivers, with caregivers appearing to be more diverse than infants. At the 12 month time point (5), however, infants appear to be more diverse than at the other time points.

Figure 14: Partial Least Squares Discriminate Analysis (PLS-DA) plots of nares microbiota of infants and their primary caregivers. Principal Components 1 and 2 are represented by the X and Y axes, respectively. Panel A shows all subjects at all time points together (i.e., the full data set). Colour coding with variable symbols denotes the sampling time points. Panels B-E show the individual time points (2 weeks (B), 2 months (C), 4 months (D), 6 months (E), and 12 months of age (F)). Infants are depicted with green circles and caregivers are denoted using blue squares.


A
R2Y = 0.1141
R2X = 0.0688
Ellipse: Hotelling's T2 (95%)
1
2
3
4
5
22IT5
21IT5
48PT4
32PT5
27PT2
24PT2
38PT1
21PT1
48PT1
10IT1
50IT1
34IT1
28IT4
49PT4
1IT3
26IT3
2PT3
5IT3
38PT3
32PT3
40PT3
B
R2Y = 0.4185
R2X = 0.1433
Ellipse: Hotelling's T2 (95%)
Infants
Caregivers
22PT1
41PT1
48PT1
10IT1
50IT1
42IT1
24PT1
1PT1
35PT1
3PT1
19PT1
44PT1
37PT1
38PT1
21PT1
22IT1


Infants
Caregivers
C
R2Y = 0.451
R2X = 0.1912
Ellipse: Hotelling's T2 (95%)
50PT2
40PT2
20PT2
31PT2
34PT2
42PT2
3PT2
24PT2
27PT2
17PT2
49IT2
21IT2
24IT2
23IT2
18IT2
D
R2Y = 0.4553
R2X = 0.2376
Ellipse: Hotelling's T2 (95%)
10PT3
29PT3
5PT3
42PT3
23PT3
26PT3
44PT3
14PT3
32PT3
38PT3
40PT3
42IT3
10IT3
43IT3
48IT3
32IT3


Infants
Caregivers
E
R2Y = 0.5109
R2X = 0.1932
Ellipse: Hotelling's T2 (95%)
28IT4
20IT4
43IT4
31IT4
50IT4
10IT4
22PT4
1PT4
45PT4
44PT4
41PT4
40PT4
48PT4


F
R2Y = 0.575
R2X = 0.208
Ellipse: Hotelling's T2 (95%)
21IT5
22IT5
32IT5
34IT5
8IT5
1IT5
48IT5
39IT5
44PT5
27PT5
32PT5
39PT5
24PT5
43PT5
26PT5

4 Discussion

4.1 Anterior Nares Diversity

As hypothesized, these findings provide evidence that diversity of the nasal microbiota in infants increases over the first year of life. The median species richness in 25-35 infants per time point, (representing the total number of different OTUs found per individual as measured by the S_{obs} richness index) was found to increase over the course of the year, from 56.7 at 2 weeks (time point 1) to 77.6 at 12 months of age (time point 5), corresponding overall to a 26.9% increase. The infant nares diversity measured lower than caregivers averaging 56.7 (median 51) versus 68.6 (median 68) in caregivers at time point 1, but evolved over the year to higher than observed in caregivers by 12 months of age (average/median diversity: 77.6 /82 in infants versus 70.2/71 in caregivers). The caregivers (n = 40) were found to carry a relatively constant average number of OTUs throughout the year, with the exception of time point 4 (July-August), in which the Chao species richness dropped significantly for an unexplained reason, possibly due to less concentrated indoor air exposure in the summer months (average/median richness 122.3/104.5 in caregivers vs 155.5/147.1 in infants; p value = 0.039). The Shannon diversity index for infants also was significantly lower than for caregivers at time point 1. Lower nares diversity can be attributed to the lack of exposure of two week old infants to varied external environments. Presumably infants at 2 weeks of age are virtually unexposed to environments beyond the home (with only limited hospital exposure), and have minimal contacts other than their (close contact) caregivers.

Similar to other human microbiome research studies, this study showed high variability between study subjects. The Chao richness levels ranged from 39 - 391.25 in infants and 49 - 413.5 in caregivers and the Shannon diversity indices ranged from 0.1 – 3.47 in infants and 0.1 – 2.48 in caregivers. Hence, median values at each time point are most informative for inferring study trends. Lemon *et al*. measured anterior nares Chao richness levels in a healthy adult population (n = 7) from sequenced 16S rRNA gene clone libraries and found an average of 50 taxa (±7.2) at a 97% sequence similarity level [29]. Their Chao richness observation was much lower than the average Chao richness measurement of 143.6 (median: 122.67) in adult caregivers (n = 40) from this study based on sequencing; however they sequenced only 200-500 cloned reads per subject sample, whereas the minimum from this study was 1000 reads per subject sample. In contrast, using a phylogenetic microarray (based on 16S rRNA sequences) for detection, Lemon and colleagues identified 125-778 different taxa per nares sample, with an average of 342.7. In comparison, this sequence-based study observed an average of 67 taxa in infants and 65 in caregivers (medians: 61 and 62.5 respectively). As microarrays are subject to inherent subtleties and biases, the diversity is likely inflated; 454 pyrosequencing-based results are filtered to a higher level of stringency and are more likely to represent the true diversity of the sample. Oh *et al.* measured nares microbiome diversity in children at various stages of puberty by cloning and sequencing 16s rRNA targets. In prepubescent children, they found a Shannon diversity index of 2.3, whereas in teenagers the Shannon diversity was lower at 1.9 [30]. In contrast, the average Shannon diversity values from this study were much lower, at 1.12 (median: 1) in infants and 1.18 (median: 1.24) in caregivers. These lower values could possibly be attributed to the sample processing methods used in this study, including whole-genome

DNA augmentation. This could occur from preferential amplification of some organisms and a lack of amplification of others. In contrast, studies using 16S rRNA may have inflated diversity estimates owing to multiple copies of 16S rRNA genes in the chromosome [19]. A study by Li *et al*. from the Human Microbiome Project (HMP) found an average Shannon diversity index of 1.696 in their population, but also with a large variability between individual subjects [76]. Due to the large variability between subjects, the median values may be more informative than the means, as they are less likely to be skewed by outliers. As the nares represent a body site that is highly exposed to external environments through respiration as well as contact with fingers, it is reasonable to expect high variability within the nares microbiota resulting from variable exposure to external microbes over time and seasons. The wide range in diversity indices observed between these multiple research studies can be due to differences in sample collection, processing, and data analysis; differences between study populations (socioeconomic, climate, location etc.); as well as the high degree of natural variability between individual participant nares samples. Different genetic markers (16S rRNA vs *cpn60* UT) can yield different results, along with varying standards for quality filtering and chimera removal leading to differing measures of diversity. Microbiome studies themselves are in an infancy state of constant development, with a variety of methods of sample processing and data analysis currently in use.

4.2 Core Microbiome of the Anterior Nares

Consistent with previous reports, the nares microbiomes in participants from this study were made up primarily of Actinobacteria and Firmicutes, with a lesser number of Proteobacteria [29,38,46,77]**.** The HMP also detected a large proportion of Actinobacteria and Firmicutes in the nares, with smaller numbers of Proteobacteria, Cyanobacteria and Bacteroides. In

comparison to data acquired in the HMP, this study found a smaller proportion of Bacteroides and an absence of Cyanobacteria [31]**.** Differences in the numbers of detected Bacteroides and Cyanobacteria can be attributed to using *cpn60* UT instead of 16S rRNA. *Cpn60* has a much smaller database than 16S rRNA, and it is possible that some organisms belonging to these phyla were not detected owing to use of this phylogenetic marker. An inverse correlation in the relative proportions of Actinobacteria and Firmicutes as detected, which was also noted by Lemon *et al*. [29]. The core microbiome of the human anterior nares consists of a very small number of main genera and a variety of organisms found in low abundance. Individuals appear to have only one or two predominating genera; however intra-subject variability was noted. The predominant taxa in the nares varied between individuals at each time point, as well as in the same individual across time points. This study was congruent with a previous finding noting a large variation between predominant taxa found in each subject [76]. Together these findings imply that there may be a degree of functional redundancy within the nares microflora, allowing for a variety of organisms to perform essential functions within the nares ecosystem.

To define a core microbiome there are two main highly recommended criteria: abundance (the relative number of reads of a specific taxon in an individual or community) and ubiquity (the number of members of the community carrying a specific taxon) [78]. Based on the existing trend in the field, a cut-off point of >0.1% abundance in at least 50% of subjects during at least one time point was chosen. Using these criteria, *Corynebacterium, Rhodococcus, Staphylococcus, Dolosigranulum, Streptococcus, Moraxella*, and the classes Actinobacteria and Gammaproteobacteria were identified as members of the core nares microbiome for both infants and caregivers. According to these criteria, the class

Actinobacteria classified as a core nares organism only in caregivers, and *Moraxella* was exclusively a member of the infant core microbiome. A previous study by Bogaert *et al*. also defined the core microbiome of nasopharyngeal (NP) samples from 18 month old children as organisms that were present in more than 50% of all samples, and representing >0.1% of all the sequences. Their core nasopharyngeal microbiome consisted of *Moraxella, Haemophilus influenzae, Enhydrobacter, Streptococcus, Dolosigranulum,* and *Corynebacterium* [38]. Despite being measured from a distinct anatomical site, these core NP organisms were common in nares findings from this study as well, albeit with more acute otitis media (AOM) pathogens and fewer skin organisms detected in the latter. The HMP studying healthy adults described *Propionibacterium, Corynebacterium, Staphylococcus* and *Moraxella* as some of the most predominant nares flora, with smaller amounts of *Streptococcus, Bacteroides, Haemophilus, Prevotella, Veillonella,* and *Lactobacillus* [77]. Frank *et al*. found that *Propionibacterium* and *Staphylococcus* were prevalent in the nares of all their adult study subjects [46]; whereas *Propionibacterium* was found in 30-45% of caregivers and rarely in infants, and *Staphylococcus* in 65-100% of subjects. This discrepancy is likely attributed to Frank *et al*. having only five healthy adult subjects, whereas this study observed 40 study subjects over 5 time points.

In this study, a positive association was found between OTU abundance and ubiquity. Ubiquity does not necessarily correlate with abundance; however, organisms present in a majority of study subjects also tended to be predominant genera in at least a few subjects. This finding agrees with previous data from the HMP that found a positive association between the abundance and ubiquity of OTUs in the nares, with the most abundant OTUs being present in the greatest number of subjects [78,79]. *Moraxella* (which had

a high abundance but low ubiquity) and *Streptococcus* (with a low relative abundance of about 2% but present in >80% of participants) were exceptions to this HMP finding [79]. The HMP also reported that the majority of genera in the nares were in <1% abundance, but many of them were quite ubiquitous [79].

4.3 Nares Microbiota of Infants and their Primary Caregivers

In agreement with my hypothesis, this study showed differential representation of organisms in infants compared with their primary caregivers. The first time point had the greatest similarity in microbial composition, but the greatest difference in diversity between the two groups; which can be attributed to close contact between the infant and their primary caregiver, along with minimal environmental exposure [80]. During subsequent time points, however, an increase in diversity and a divergence in composition of nares microflora was noted, with time point 5 showing the greatest difference between the two study groups. This observation agrees with findings of a previous study in which cohabitation did not result in convergence of nares microbiotas [46]. Studies performed on other human body sites have also observed a differential composition of microflora between infants and adults, including a study by Capone *et al.* on the skin microflora showing an abundance of Bacilli, Clostridia, and Actinobacteria in infants, and Proteobacteria, Actinobacteria, and Firmicutes in adults [81].

Infants in this study generally carried more Firmicutes and fewer Actinobacteria and Proteobacteria than caregivers, which is in agreement with Oh *et al.* which found Firmicutes and Proteobacteria to be associated with pre-pubescent children and Actinobacteria to be associated with older teenagers [30]. In this study, infants carried a higher abundance of *Dolosigranulum* and *Streptococcus* than caregivers, and also carried *Dolosigranulum* more

often. Caregivers were more likely to be carriers of *Propionibacterium, Staphylococcus* and a variety of lower abundance organisms. Oh *et al*. also found that young children carried a higher abundance of *Streptococcus, Dolosigranulum, Gemella, Granulicatella, Moraxella, Haemophilus* and *Neisseria*, while older teenagers carried more *Corynebacterium, Propionibacterium* and *Turicella* [30]. Inter-subject diversity in adults at the genus level was greater than was observed for infants, contrary to findings by Oh *et al.*, in which they found a greater diversity in their youngest age group. However, the age range of their study was 2-17 years of age, which could account for this discrepancy [30]. Differences between infants and caregivers can be attributed to a wide variety of differences between the two groups, including different contacts and environmental exposures as well as immune system function and possibly even anatomical differences within the anterior nares [80,81].

4.4 Temporal Variation/Seasonality

A seasonal trend was observed in the composition of the nares microbiota, especially in the caregivers group. The 5 time points were taken over the course of one year, with time points 1 and 5 in January-February (winter), 2 in March-April (spring), 3 in May-June (spring/summer) and 4 in July/August (summer). It should be noted that 9 samples from time point 5 were collected in March and April. Testing was performed in Winnipeg, Canada, a city with a humid continental climate and great differences between summer and winter temperatures ranging from an average of 25^{o}C/11^{o}C (high/low) in August to -12 ^{o}C /-22^{o}C in January. Relative humidity values range from 77% in May to 87.7% in August [82]. A study by Rintala *et al.* (2008) measured the microbial composition of dust inside office buildings during all four seasons in Finland and found that all seasons were significantly different from each other [83]. People are more likely to be outside in spring and summer, coming into

contact through respiration with a greater variety of microbes and other airborne particles. In addition, winter air is very cold and could possibly have an effect on the microbiota of the nares [53]. Airborne microbes are expected to differ in Winnipeg's winter and summer. In winter, the Winnipeg ground is covered in snow, the windows are more likely to remain closed to retain circulating heated air, and people are more likely to remain inside for long periods of time. In summer, people spend more time outdoors when the temperature is warmer and there are more sources of aerosolized microbes carried from soil, wildlife, and plants.

It was also determined that the microbial composition of the anterior nares does not remain stable over time. There was trend between abundance of phyla or genera in individuals, with an individual predominantly carrying Actinobacteria during one time point carrying a high abundance of Firmicutes at the next. This is in accordance with studies by the HMP in which there was a weak correlation between nares swabs in adults taken at different times (30-359 days) [77].

It was observed that the caregiver nares microflora of both winter time points were similar in composition, as well as those of both summer time points. In contrast, infants did not follow this same trend, as they had the fewest number of shared OTUs between the two winter time points; however, in infants seasonality is difficult to examine as it is confounded with infant development causing physiological changes and immune system development, and mobility changes over the first year of life. An interesting observation, however, is the grouping in the phylogenetic tree of infant microfloras at time points 2 and 5, showing that seasonality perhaps plays a role, as time point 2 is in early spring, and time point 5 takes

place between winter and early spring. If season doesn't influence the nares microbiome, one might expect sampling period 5 to be most similar to 4 since these are temporally closest.

When looking at individual taxa, a distinctive decrease of *Staphylococcus* abundance was observed in infants in the summer months. It should be noted, however, that seasonal variation in infants may be confounded with general changes in the nares microbiota as the infant develops throughout the year. This seasonal trending was not seen in the adults, with *Staphylococcus* actually increasing during time point 4/summer. Members of the class Actinobacteria also appeared to increase in abundance in the summer in infants but remained relatively stable all year in adults. A few lower abundance organisms appeared to show seasonal tendencies with *Acinetobacter* being absent in infants during summer, and *Veillonella* being present in infants only in summer. Fewer caregivers carried *Streptococcus* in winter. These trending observations resemble a study by Bogaert *et al*. which demonstrated a seasonal trend in the nasopharyngeal flora of children [38].

4.5 Potentially Pathogenic Bacteria

This study showed variations in carriage of potentially pathogenic bacteria (PPB) between infants and their primary caregivers as well as temporal trending. PPB of the anterior nares include: *Moraxella catarrhalis, Pseudomonas aeruginosa, Staphylococcus aureus,* and *Streptococcus pneumoniae,* [59]. Although these organisms are generally commensal within the human epithelia, they have the potential to cause disease [84]. For example, *S. aureus* carriers are more likely to become infected and are three times more likely to develop *S. aureus* bacteraemia [84], while carriage of acute otitis media pathogens is associated with disease [59].

In infants, *M. catarrhalis* carriage rates increased over the course of the year, however the relative abundance remained generally level. At two weeks of age, *M. catarrhalis* was absent in infants, and was present in 28% of infants by 12 months, although in low abundance. This observation is congruent with previous studies that have found that carriage of acute otitis media pathogens in the nasopharynx increases throughout the first year of life [42,59]. Abundance of *P. aeruginosa* was low in the majority of participants; however two adults carried it as a predominant organism during time point 3. *P. aeruginosa* is not generally considered to be a commensal member of the normal nares flora, and it is interesting to note that the people who carried *P. aeruginosa* as a predominant member of the nares microbiota also reported a flu-like illness during the same time point.

The HMP found that 29% of individuals carried *S. aureus* in nares samples [77]; In this study carriage rates were between 35-73% in caregivers and 29-80% in infants. The abundance and ubiquity of *S. aureus* in caregivers fluctuated throughout the year, but showed a general increase in carriage with time both in abundance and ubiquity. *S. aureus* carriage levels were highest during time point 5, which is in agreement with previous studies in which *S. aureus* was most prevalent in winter [48], as well as anecdotal evidence of higher *S. aureus* infection rates (at least in seniors) during winter. This study found *S. aureus* carriage was lowest during time point 1, which is another winter time point. One possible explanation for this observation is that the majority of caregivers in this study were mothers of the infant subjects and would have given birth 2 weeks prior to time point 1. Physiological changes associated with pregnancy have been shown to influence the gut microflora [85] and it is reasonable to believe that pregnancy may also affect the microflora of mucus membranes,

such as the nares. During delivery approximately half of mothers were given antibiotics, which can also be expected to potentially influence the mother's microflora for a few weeks.

Several studies have described a negative association between *S. aureus* and *S. pneumoniae* [43,44]. Bogaert *et al*. noticed a negative correlation between *S. aureus* and *S. pneumoniae*, but only for PCV7 vaccine type strains [86]. This study also showed an inverse relationship between *S. aureus* and *S. pneumoniae* in infants but not in caregivers. Caregivers already had a very low abundance of *S. pneumoniae* throughout the year as it was present in less than 20% of the adult population. In infants, *S. aureus* carriage (both abundance and ubiquity) declined in summer, whereas carriage of *S. pneumoniae* increased. During the final time point (5/winter), an increase in *S. aureus* carriage was seen (both relative abundance and number of carriers) as well as a decrease in *S. pneumoniae* carriage (relative abundance). This is in concordance with past studies that have found an inverse correlation between *S. pneumoniae* and *S. aureus* carriage in children as well as studies demonstrating seasonality for *S. aureus* carriage [48]. Interference between *S. aureus* and *S. pneumoniae* may be caused by hydrogen peroxide production by *S. pneumoniae* [43,44]. Other organisms have been identified as possible competitors with *S. aureus* including Actinobacteria [29] and *Corynebacterium spp.* [60].

4.6 Limitations of the Study

Microbiome studies present a variety of challenges including recruitment and compliance of human subjects, complex sample processing pipelines, and challenges surrounding the complexity of measuring microbial populations inhabiting variable human populations. Despite such challenges, this study has provided important information regarding microbial communities of the human anterior nares that has not been reported previously. Known or

assumed limitations of this study arise from sampling bias, sample processing or template amplification bias, and sequencing and/or data processing errors. Attempts were made to address each in turn to the best of our ability.

Owing to failure of infant-caregiver pairs to show up to every scheduled appointment, as well as some samples lacking a minimum of 1000 quality reads, data was ultimately acquired from only 6 infant/caregiver pairs from each time point. This reduces capacity for performing longitudinal analysis on the subject data. To avoid undue discomfort to the study subjects, each participant was sampled from only one nostril; however a study performed by Frank *et al.* collected specimens at the same time from both the right and the left nares, and found no significant differences between the samples [46].

With respect to sample processing, the MoLysis procedure that was employed degrades DNA from previously lysed cells; hence, DNA from non-intact bacterial cells should be degraded prior to target amplification, possibly changing the proportion of OTUs amplified in the final purified DNA mixture relative to other studies without dead cell removal. Given the rapid turnover of bacteria in the nares, these results represent an important advance in the method of analyzing a body site that can contain many dead cells due to the presence of mucus. Cell wall structure and composition can alter DNA extraction efficiency [87]; however, the MoLysis protocol has been tested with a variety of Gram-positive and Gram-negative bacteria including: *Bacillus cereus, Corynebacterium diphteriae, Enterococcus sp., Lactobacillus sp., Staphylococcus sp., Streptococcus sp., Escherichia coli, Serratia marsescens, Neisseria sp., Proteus sp., Pseudomonas sp.,* and several others [88]. The DNA template purification procedure that was used employed SPRI magnetic beads (GenFind, Ampure) that could potentially leach metal ions into the DNA mixture if left for

too long without ensuring removal of all beads, potentially contributing to DNA degradation of some subject template mixtures. Although care was taken to avoid beads during the manual transfer of eluted DNA, this remains a potential source of sample-to-sample variability as there is no way of measuring this potential degradation. Lastly owing to the prospective sampling schedule, many DNA samples were stored at 4^{o}C for prolonged periods of time, leading to possible degradation of bacterial DNA.

As with all microbiome studies currently unable to achieve single molecule sequencing, this study may include template amplification bias, including potential bias introduced by whole-genome augmentation, the *cpn60* UT PCR amplification reaction itself, or the emPCR required to augment beads prior to 454 sequencing. Despite a high level of primer degeneracy, some UT sequences may be amplified preferentially over others, or some UT sequences might be more effectively amplified at each of the three annealing temperatures. Such amplification bias also applies to the linearly amplifying GenomiPhi whole-genome template augmentation process we employed. With random hexamer-primed linear augmentation, some oligonucleotides may bind more efficiently to some templates over others. Another source of potential bias leading to skewed relative abundance values may result from to the so-called "C_oT effect" [89,90]. The C_oT effect was demonstrated in a study by Mathieu-Daude *et al.,* and describes the phenomenon that occurs in a mixed-template PCR reactions wherein amplification of DNA templates found in highest abundance tend to decline more rapidly than those found in lower abundance owing to preferential annealing of the amplicons to each other rather than to the primers [90]. It should be noted, that most microbiome studies share these same technical limitations, perhaps with the exception of linear whole-genome amplification.

A major source of known error when using pyrosequencing stems from inaccurate measurement or production of light intensities in homonucleotide tracts (also called homopolymers), resulting in potential miscalling of the number of bases and possible insertion or deletion (indel) errors [17]. A study by Kunin *et al.* (2010) found that base calling errors inflated the actual diversity of 16S rRNA reads by two orders of magnitude when considering unique reads [59,89]. A large proportion of these sequencing errors are due to miscounted homopolymer runs (with their associated indel errors) that are not picked up during quality trimming [59,89]. These errors are also highly reproducible, producing phantom OTUs with multiple reads, leading to possible detection of false phylotypes that in some cases could be relatively abundant [59,89]. To minimize such errors, the *PyroNoise* algorithm, a flowgram preclustering software tool, was employed to reduce pyrosequencing errors, increase the accuracy of relative abundance measurements, and create more accurate phlogeny-based diversity measures. However, reads containing indels may remain within our pass-through dataset (following PyroNoise), as large numbers of singleton OTUs were present [17]. Clustering OTUs at 97% rather than 100% is another approach employed to minimize technology-induced indel errors by grouping reads containing errors with similar reads containing no errors. There were no technical replicates included in this study, so it was impossible to determine how much of the variation between samples was due to technical variation and how much was due to true biological variance; however, every effort was taken to ensure the soundness of the data.

4.7 Current Implications and Future Studies

Using high throughput targeted gene sequencing, this study has characterized the nares microbiota of developing infants and their primary caregivers. This study addresses an

important knowledge gap, as infant nares microflora have not been previously catalogued and studies of nares microbiota have been rare. This study has explored the infant nares microbiota throughout the first year of life, and has identified both temporal differences by season and differences by age. A major strength of this study is the size of this subject populations (40 infants and their associated primary caregivers), which represents a relatively large number compared to other microbiome studies [4,18,29]. The current study adds to the general knowledge base of the nares microbiota of a healthy individual. This study has also collected evidence on temporal trending in the carriage of potentially pathogenic organisms, better informing us when an individual may be predisposed to infections such as acute otitis media (AOM).

Calibrator studies are first needed to better understand the differences between the organisms that colonize specific body sites in times of health and in disease states, as well as those that are associated with specific symptoms and levels of disease severity. For example, chronic sinusitis patients show less diversity in the maxillary sinuses and nasopharynx than healthy controls [91], whereas increased diversity of vaginal flora is indicative of bacterial vaginosis [92]. Recent discoveries have been made as to the influencing or correlative roles of host microbiota in human conditions such as psoriasis, obesity, asthma, and inflammatory bowel disease [31]. For example, psoriatic lesions appear to have increased levels of Firmicutes and decreased levels of Actinobacteria [31]. Presence of potentially pathogenic organisms in the anterior nares may be informative, indicating possible presence of PPBs in the nasopharynx and thus, increased risk for acute otitis media. Such insight gathered regarding microflora of healthy human beings may ultimately lead to enhanced treatments for upper

respiratory tract infections, identification of protective probiotics, and possible earlier diagnoses of previously difficult to diagnose illnesses.

Although our study size (n = 40) is somewhat limited, additional insight may be gained from the collected data by investigating additional clinical information gathered from the questionnaires. For example, comparing nares microbiota for breastfed infants versus formula-fed infants, or comparing different delivery methods. Having demonstrated the feasibility of such work, future studies should include larger subject cohorts with more detailed questionnaires, to determine biological factors and/or socio-economic factors that influence the anterior nares flora. To explore the seasonality of nares microflora, a multi-centre study should be performed comparing seasonal trends between groups of individuals living in different climatic regions. Lastly, comparing the nares microbiota of healthy subjects versus those with upper respiratory tract infections (URTI), allergies, asthma, cystic fibrosis or other illnesses should allow us to better understand these human disease states, leading to potentially better treatment strategies.

5. Conclusion

This thesis demonstrated an effective protocol with which to study the human nares microbiome, as compared with data from prior studies. This study addresses an important knowledge gap, as infant nares microflora have not been previously catalogued and studies of nares microbiota have been rare. This study explored the infant nares microbiota throughout the first year of life and has identified both temporal differences by season and differences by age. In agreement with my hypothesis, this study found that the diversity of the nares microbiota in infants increases over the first year of life. A high intra-subject variability was noted between study subjects across all five time points; however the nares microbiome was made up primarily of three phyla (Actinobacteria, Firmicutes and Proteobacteria). Individuals contained only one or two predominating genera, with the predominant taxa varying between individuals at each time point, as well as in the same individual across multiple time points. The core nares microbiome in both infants and caregivers in this study was comprised of *Corynebacterium, Rhodococcus, Staphylococcus, Dolosigranulum, Streptococcus, Moraxella*, and the classes Actinobacteria and Gammaproteobacteria.

A seasonal trending in the composition of the nares microbiota was observed; with caregiver nares microflora of both winter time points being similar in composition, as well as those of both summer time points. Infants did not follow this same trend, as they had the fewest number of shared OTUs between the two winter time points. When looking at potentially pathogenic organisms, it was found that *M. catarrhalis* rates increased over the course of the year in infants; as well we observed an inverse relationship between *S. aureus* and *S. pneumoniae*, with an increase in *S. pneumoniae* in summer and a higher prevalence of

S. aureus in winter. The abundance and ubiquity of *S. aureus* in caregivers fluctuated throughout the year, but showed a general increase in *S. aureus* carriage over the T1 – T5 sampling period time frame (early spring to late winter).

6. References

1. Roche. emPCR method manual – lib-L SV GS FLX titanium series. . 2010.

2. Tlaskalova-Hogenova H, Stepankova R, Hudcovic T, et al. Commensal bacteria (normal microflora), mucosal immunity and chronic inflammatory and autoimmune diseases. *Immunol Lett*. 2004;93(2-3):97-108.

3. NIH HMP Working Group, Peterson J, Garges S, et al. The NIH human microbiome project. *Genome Res*. 2009;19(12):2317-2323.

4. Zaura E, Keijser BJ, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol*. 2009;9:259.

5. Hull MW, Chow AW. Indigenous microflora and innate immunity of the head and neck. *Infect Dis Clin North Am*. 2007;21(2):265-82, v.

6. Kononen E. Anaerobes in the upper respiratory tract in infancy. *Anaerobe*. 2005:131-132-136.

7. Rasmussen TT, Kirkeby LP, Poulsen K, Reinholdt J, Kilian M. Resident aerobic microbiota of the adult human nasal cavity. *APMIS*. 2000;108(10):663-675.

8. Gluck U, Gebbers JO. Ingested probiotics reduce nasal colonization with pathogenic bacteria (staphylococcus aureus, streptococcus pneumoniae, and beta-hemolytic streptococci). *Am J Clin Nutr*. 2003;77(2):517-520.

9. Tano K, Olofsson C, Grahn-Hakansson E, Holm SE. In vitro inhibition of S. pneumoniae, nontypable H. influenzae and M. catharralis by alpha-hemolytic streptococci from healthy children. *Int J Pediatr Otorhinolaryngol*. 1999;47(1):49-56.

10. Hsiao WW, Fraser-Liggett CM. Human microbiome project--paving the way to a better understanding of ourselves and our microbes. *Drug Discov Today*. 2009;14(7-8):331-333.

11. Schellenberg J, Links MG, Hill JE, et al. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl Environ Microbiol*. 2009;75(9):2889-2898.

12. Parahitiyawa NB, Scully C, Leung WK, Yam WC, Jin LJ, Samaranayake LP. Exploring the oral bacterial flora: Current status and future directions. *Oral Dis*. 2009.

13. Lederberg J MA. 'Ome sweet 'Omics—a genealogical treasury of words. *Scientist*. 2001;15:8.

14. Peterson DA, Frank DN, Pace NR, Gordon JI. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe*. 2008;3(6):417-427.

15. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007;449(7164):804-810.

16. Nasidze I, Quinque D, Li J, Li M, Tang K, Stoneking M. Comparative analysis of human saliva microbiome diversity by barcoded pyrosequencing and cloning approaches. *Anal Biochem*. 2009;391(1):64-68.

17. Quince C, Lanzen A, Curtis TP, et al. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods*. 2009;6(9):639-641.

18. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*. 2005;43(11):5721-5732.

19. Petti CA. Detection and identification of microorganisms by gene amplification and sequencing. *Clin Infect Dis*. 2007;44(8):1108-1114.

20. Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM. cpnDB: A chaperonin sequence database. *Genome Res*. 2004;14(8):1669-1675.

21. Viale AM, Arakaki AK, Soncini FC, Ferreyra RG. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int J Syst Bacteriol*. 1994;44(3):527-533.

22. Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. The complete sequence of the mucosal pathogen ureaplasma urealyticum. *Nature*. 2000;407(6805):757-762.

23. Klunker D, Haas B, Hirtreiter A, et al. Coexistence of group I and group II chaperonins in the archaeon methanosarcina mazei. *J Biol Chem*. 2003;278(35):33256-33267.

24. Hill JE, Town JR, Hemmingsen SM. Improved template representation in *cpn60* polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. *Environ Microbiol*. 2006;8(4):741-746.

25. Jian W, Zhu L, Dong X. New approach to phylogenetic analysis of the genus bifidobacterium based on partial HSP60 gene sequences. *Int J Syst Evol Microbiol*. 2001;51(Pt 5):1633-1638.

26. Dumonceaux TJ, Hill JE, Briggs SA, Amoako KK, Hemmingsen SM, Van Kessel AG. Enumeration of specific bacterial populations in complex intestinal communities using quantitative PCR based on the chaperonin-60 target. *J Microbiol Methods*. 2006;64(1):46-62.

27. Hill JE, Hemmingsen SM, Goldade BG, et al. Comparison of ileum microflora of pigs fed corn-, wheat-, or barley-based diets by chaperonin-60 sequencing and quantitative PCR. *Appl Environ Microbiol*. 2005;71(2):867-875.

28. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*. 2011;6(12):e27310.

29. Lemon KP, Klepac-Ceraj V, Schiffer HK, Brodie EL, Lynch SV, Kolter R. Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. *MBio*. 2010;1(3):e00129-10.

30. Oh J, Conlan S, Polley EC, Segre JA, Kong HH. Shifts in human skin and nares microbiota of healthy children and adults. *Genome Med*. 2012;4(10):77.

31. Cho I, Blaser MJ. The human microbiome: At the interface of health and disease. *Nat Rev Genet*. 2012;13(4):260-270.

32. Jourdain S, Smeesters PR, Denis O, et al. Differences in nasopharyngeal bacterial carriage in preschool children from different socio-economic origins. *Clin Microbiol Infect*. 2011;17(6):907-914.

33. Johanson WG,Jr, Blackstock R, Pierce AK, Sanford JP. The role of bacterial antagonism in pneumococcal colonization of the human pharynx. *J Lab Clin Med*. 1970;75(6):946-952.

34. Henriksson G, Helgeland L, Midtvedt T, Stierna P, Brandtzaeg P. Immune response to mycoplasma pulmonis in nasal mucosa is modulated by the normal microbiota. *Am J Respir Cell Mol Biol*. 2004;31(6):657-662.

35. Ostfeld E, Segal J, Segal A, Bogokovski B. Bacterial colonization of the nose and external ear canal in newborn infants. *Isr J Med Sci*. 1983;19(12):1046-1049.

36. Garcia-Rodriguez JA, Fresnadillo Martinez MJ. Dynamics of nasopharyngeal colonization by potential respiratory pathogens. *J Antimicrob Chemother*. 2002;50 Suppl S2:59-73.

37. Kononen E, Syrjanen R, Takala A, Jousimies-Somer H. Nasopharyngeal carriage of anaerobes during health and acute otitis media by two years of age. *Diagn Microbiol Infect Dis*. 2003;46(3):167-172.

38. Bogaert D, Keijser B, Huse S, et al. Variability and diversity of nasopharyngeal microbiota in children: A metagenomic analysis. *PLoS One*. 2011;6(2):e17035.

39. Konno M, Baba S, Mikawa H, et al. Study of upper respiratory tract bacterial flora: First report. variations in upper respiratory tract bacterial flora in patients with acute upper respiratory tract infection and healthy subjects and variations by subject age. *J Infect Chemother*. 2006;12(2):83-96.

40. Syrjanen RK, Kilpi TM, Kaijalainen TH, Herva EE, Takala AK. Nasopharyngeal carriage of streptococcus pneumoniae in finnish children younger than 2 years old. *J Infect Dis*. 2001;184(4):451-459.

41. Gries DM, Zemzars TF, Gibson KA, et al. A pilot study to assess frequency of carriage and routes of acquisition of staphylococcus aureus by healthy infants. *Am J Infect Control*. 2009;37(7):598-600.

42. Walls T, Power D, Tagg J. Bacteriocin-like inhibitory substance (BLIS) production by the normal flora of the nasopharynx: Potential to protect against otitis media? *J Med Microbiol*. 2003;52(Pt 9):829-833.

43. Gluck U, Gebbers JO. The nose as bacterial reservoir: Important differences between the vestibule and cavity. *Laryngoscope*. 2000;110(3 Pt 1):426-428.

44. Regev-Yochay G, Dagan R, Raz M, et al. Association between carriage of streptococcus pneumoniae and staphylococcus aureus in children. *JAMA*. 2004;292(6):716-720.

45. Lebon A, Labout JA, Verbrugh HA, et al. Dynamics and determinants of staphylococcus aureus carriage in infancy: The generation R study. *J Clin Microbiol*. 2008;46(10):3517-3521.

46. Frank DN, Feazel LM, Bessesen MT, Price CS, Janoff EN, Pace NR. The human nasal microbiota and staphylococcus aureus carriage. *PLoS One*. 2010;5(5):e10598.

47. Caugant DA, Hoiby EA, Magnus P, et al. Asymptomatic carriage of neisseria meningitidis in a randomly sampled population. *J Clin Microbiol*. 1994;32(2):323-330.

48. Harrison LM, Morris JA, Telford DR, Brown SM, Jones K. The nasopharyngeal bacterial flora in infancy: Effects of age, gender, season, viral upper respiratory tract infection and sleeping position. *FEMS Immunol Med Microbiol*. 1999;25(1-2):19-28.

49. Golding GR, Levett PN, McDonald RR, et al. A comparison of risk factors associated with community-associated methicillin-resistant and -susceptible staphylococcus aureus infections in remote communities. *Epidemiol Infect*. 2010;138(5):730-737.

50. Chatterjee SS, Ray P, Aggarwal A, Das A, Sharma M. A community-based study on nasal carriage of staphylococcus aureus. *Indian J Med Res*. 2009;130(6):742-748.

51. Halablab MA, Hijazi SM, Fawzi MA, Araj GF. Staphylococcus aureus nasal carriage rate and associated risk factors in individuals in the community. *Epidemiol Infect*. 2010;138(5):702-706.

52. Konno M, Baba S, Mikawa H, et al. Study of nasopharyngeal bacterial flora. variations in nasopharyngeal bacterial flora in schoolchildren and adults when administered antimicrobial agents. *J Infect Chemother*. 2007;13(4):235-254.

53. Hendley JO, Hayden FG, Winther B. Weekly point prevalence of streptococcus pneumoniae, hemophilus influenzae and moraxella catarrhalis in the upper airways of normal young children: Effect of respiratory illness and season. *APMIS*. 2005;113(3):213-220.

54. Metchnikoff E. The prolongation of life. In: Heinemann W, ed. *Optimistic studies.* London: GP Putnam & Sons; 1907:1-100.

55. He X, Lux R, Kuramitsu HK, Anderson MH, Shi W. Achieving probiotic effects via modulating oral microbial ecology. *Adv Dent Res*. 2009;21(1):53-56.

56. Bakhshaee M, Naderi HR, Ghazvini K, Sotoudeh K, Amali A, Ashtiani SJ. Passive smoking and nasopharyngeal colonization by streptococcus pneumoniae, haemophilus influenzae, and moraxella catarrhalis in daycare children. *Eur Arch Otorhinolaryngol*. 2011.

57. Charlson ES, Chen J, Custers-Allen R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One*. 2010;5(12):e15216.

58. Brook I, Gober AE. Recovery of potential pathogens and interfering bacteria in the nasopharynx of smokers and nonsmokers. *Chest*. 2005;127(6):2072-2075.

59. Faden H, Duffy L, Wasielewski R, Wolf J, Krystofik D, Tung Y. Relationship between nasopharyngeal colonization and the development of otitis media in children. tonawanda/williamsville pediatrics. *J Infect Dis*. 1997;175(6):1440-1445.

60. Lina G, Boutite F, Tristan A, Bes M, Etienne J, Vandenesch F. Bacterial competition for human nasal cavity colonization: Role of staphylococcal agr alleles. *Appl Environ Microbiol*. 2003;69(1):18-23.

61. Konno M, Baba S, Mikawa H, et al. Study of nasopharyngeal bacterial flora. second report. variations in nasopharyngeal bacterial flora in children aged 6 years or younger when administered antimicrobial agents. part 1. *J Infect Chemother*. 2006;12(5):287-304.

62. Obaro S, Adegbola R. The pneumococcus: Carriage, disease and conjugate vaccines. *J Med Microbiol*. 2002;51(2):98-104.

63. Regev-Yochay G, Lipsitch M, Basset A, et al. The pneumococcal pilus predicts the absence of staphylococcus aureus co-colonization in pneumococcal carriers. *Clin Infect Dis*. 2009;48(6):760-763.

64. Kayhty H, Auranen K, Nohynek H, Dagan R, Makela H. Nasopharyngeal colonization: A target for pneumococcal vaccination. *Expert Rev Vaccines*. 2006;5(5):651-667.

65. Lee GM, Huang SS, Rifas-Shiman SL, et al. Epidemiology and risk factors for staphylococcus aureus colonization in children in the post-PCV7 era. *BMC Infect Dis*. 2009;9:110.

66. Van Effelterre T, Moore MR, Fierens F, et al. A dynamic model of pneumococcal infection in the united states: Implications for prevention through vaccination. *Vaccine*. 2010.

67. Benninger MS. Acute bacterial rhinosinusitis and otitis media: Changes in pathogenicity following widespread use of pneumococcal conjugate vaccine. *Otolaryngol Head Neck Surg*. 2008;138(3):274-278.

68. Cohen R, Levy C, de La Rocque F, et al. Impact of pneumococcal conjugate vaccine and of reduction of antibiotic use on nasopharyngeal carriage of nonsusceptible pneumococci in children with acute otitis media. *Pediatr Infect Dis J*. 2006;25(11):1001-1007.

69. Dabernat H, Plisson-Saune MA, Delmas C, et al. Haemophilus influenzae carriage in children attending french day care centers: A molecular epidemiological study. *J Clin Microbiol*. 2003;41(4):1664-1672.

70. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol*. 1990(215):403-410.

71. Spangler R, Goddard NL, Thaler DS. Optimizing taq polymerase concentration for improved signal-to-noise in the broad range detection of low abundance bacteria. *PLoS One*. 2009;4(9):e7010.

72. Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S. An efficient and economic enhancer mix for PCR. *Biochem Biophys Res Commun*. 2006;347(3):747-751.

73. Handschur M, Karlic H, Hertel C, Pfeilstocker M, Haslberger AG. Preanalytic removal of human DNA eliminates false signals in general 16S rDNA PCR monitoring of bacterial pathogens in blood. *Comp Immunol Microbiol Infect Dis*. 2009;32(3):207-219.

74. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*. 2011;12:38.

75. DeSantis TZ,Jr, Hugenholtz P, Keller K, et al. NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res*. 2006;34(Web Server issue):W394-9.

76. Li K, Bihan M, Yooseph S, Methe BA. Analyses of the microbial diversity across the human microbiome. *PLoS One*. 2012;7(6):e32118.

77. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207-214.

78. Huse SM, Ye Y, Zhou Y, Fodor AA. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One*. 2012;7(6):e34242.

79. Zhou Y, Gao H, Mihindukulasuriya KA, et al. Biogeography of the ecosystems of the healthy human body. *Genome Biol*. 2013;14(1):R1.

80. Fouhy F, Ross RP, Fitzgerald GF, Stanton C, Cotter PD. Composition of the early intestinal microbiota: Knowledge, knowledge gaps and the use of high-throughput sequencing to address these gaps. *Gut Microbes*. 2012;3(3):203-220.

81. Capone KA, Dowd SE, Stamatas GN, Nikolovski J. Diversity of the human skin microbiome early in life. *J Invest Dermatol*. 2011;131(10):2026-2032.

82. Government of Canada. **Canadian climate normals 1971-2000 station data**. http://climate.weather.gc.ca/climate_normals/results_e.html?stnID=3698&prov=&lang=e&dCode=1&dispBack=1&StationName=winnipeg&SearchType=Contains&province=ALL&provBut=&month1=0&month2=12. Updated 2013. Accessed Aug 29, 2013.

83. Rintala H, Pitkaranta M, Toivola M, Paulin L, Nevalainen A. Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiol*. 2008;8:56-2180-8-56.

84. van Belkum A. Staphylococcal colonization and infection: Homeostasis versus disbalance of human (innate) immunity and bacterial virulence. *Curr Opin Infect Dis*. 2006;19(4):339-344.

85. Konstantinov SR, van der Woude CJ, Peppelenbosch MP. Do pregnancy-related changes in the microbiome stimulate innate immunity? *Trends Mol Med*. 2013;19(8):454-459.

86. Bogaert D, van Belkum A, Sluijter M, et al. Colonisation by streptococcus pneumoniae and staphylococcus aureus in healthy children. *Lancet*. 2004;363(9424):1871-1872.

87. Carrigg C, Rice O, Kavanagh S, Collins G, O'Flaherty V. DNA extraction method affects microbial community profiles from soils and sediment. *Appl Microbiol Biotechnol*. 2007;77(4):955-964.

88. Molzym. Directed bacterial DNA enrichment from bacteremic blood. *MolYsis Basic 0 2ml blood protocol*. 2007.

89. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol*. 2010;12(1):118-123.

90. Mathieu-Daude F, Welsh J, Vogt T, McClelland M. DNA rehybridization during PCR: The 'cot effect' and its consequences. *Nucleic Acids Res*. 1996;24(11):2080-2086.

91. Abreu NA, Nagalingam NA, Song Y, et al. Sinus microbiome diversity depletion and corynebacterium tuberculostearicum enrichment mediates rhinosinusitis. *Sci Transl Med*. 2012;4(151):151ra124.

92. Schellenberg JJ, Links MG, Hill JE, et al. Molecular definition of vaginal microbiota in east african commercial sex workers. *Appl Environ Microbiol*. 2011;77(12):4066-4074.

Appendix A: Additional Protocols

I. Oragene Protocol

Purification of nares DNA was performed using the Laboratory Protocol for Manual Purification of DNA from 0.5ml of Oragene®-DNA/saliva protocol following pre-treatment as described in Section 2.1.2. Twenty microliters of Oragene-DNA Purifier were added to a 500µl nares sample, vortexed for a few seconds to mix and incubated on ice for 10 minutes. This mixture was then centrifuged at room temperature for 5 minutes at 15,000xg and the supernatant was transferred to a fresh LoBind DNA tube (Eppendorf Canada, Mississauga Canada). 500µl of room-temperature 95% ethanol was added to 500µl supernatant and mixed gently by inversion 10 times. The sample was then incubated at room temperature for 10 minutes to allow the DNA to fully precipitate and centrifuged at room temperature for 2 minutes at 15,000xg. The supernatant was removed and discarded and the pellet was washed by adding 250µl of 70% ethanol. After a 1 m incubation, the ethanol was removed and after a brief air dry to evaporate off residual ethanol, 40µl of TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) was added to resuspend the DNA pellet and the sample was incubated at 50^{o}C for approximately 1 h and at room temperature for approximately 3 h before storage at 4^{o}C.

II. Blunt-end Cloning and Sanger Sequencing

PCR products were purified using the MinElute PCR purification kit (Qiagen, Toronto, Canada), using the same protocol as described in Section 2.2.7 for the Qiaquick PCR Purification kit. Poly-A tails were added to the end of each PCR product to facilitate ligation with the vector by adding 1µl ThermoPol reaction buffer (New England Biolabs, Ipswitch,

USA), 1µl 2nM dNTPs, 0.2µl *Taq* polymerase and 3µl ddH_2O to 5µl PCR product (48.2ng/µl). This mixture was spun down in a microcentrifuge and incubated on a thermal cycler at 72^oC for 10 minutes. Cloning was performed as per the Topo XL PCR cloning kit (Invitrogen, Carlsbad, USA). One microliter of PCR-XL-TOPO® vector was added to 4µl of the poly A-PCR product, mixed gently and allowed to incubate for 5 minutes at room temperature (~25^oC). One microliter of 6X TOPO® Cloning Stop Solution was added and mixed for several seconds at room temperature, spun down and placed on ice. One µl of the TOPO® Cloning reaction was added to one vial of One Shot® electrocompetent cells (Invitrogen, Carlsbad, USA), mixed gently, and transferred to a chilled 0.2cm electroporation cuvette (Bio-Rad Mississauga, Canada), and an electric pulse of 1.8kV for 2.5 s was delivered using a Micropulser electroporator (Bio-Rad, Mississauga, ON). Then 200µ of room temperature S.O.C medium was added, mixed with the transformed cells, and the solution was transferred to a 1.5ml Eppendorf tube and incubated for 1 h at 37^oC. 50µl and 150µ of the cells were plated on LB + Km + IPTG + XGal plates (made by spread plating 10µl IPTG (isopropyl-beta-D-thiogalactopyranoside) and 50µl X-gal on an LB plate containing 50mg/ml kanamycin).

A plasmid prep was performed on 96 colonies as per Promega Wizard SV 96 DNA Purification protocol (Promega Corporation, Madison, USA). An enzyme digest was performed on plasmid DNA using a mixture of 5µl plasmid, 2µl EcoR1 buffer (Roche Applied Science, Laval, Canada), 1µl EcoR1 (Roche Applied Science, Laval, Canada) and 12µl ddH_2O and incubating at 37^oC for 30-60 minutes and visualizing on an agarose gel. The concentration of DNA was then determined by using a Nanodrop® spectrophotometer (ND-1000 V3.7.1, National Instruments Corporation, Austin, USA). The concentrations were

adjusted to 50ng/µl with ddH_2O and were sent to an in-house facility for sequencing with 1µm M13 primers. Sequences were visualized using Sequence Scanner (Applied Biosystems, Inc. Foster City, USA) and compared to the BLASTn database (NCBI) for sequence identification.

III. List of PCR Reaction Components

1. **FailSafe™ Premix buffers (A-L)**
    - Buffer = 25µl
    - 10X primer mix (A-L) = 5µl
    - mTp Taq = 0.5µl
    - Template = 2µl
    - ddH_2O = 17.5µl
2. **Enhancer evaluation experiments**
    - 10X MTP buffer = 5µl
    - 10mM dNTPs = 1µl
    - 10X Primer mix = 5µl
    - MTP *Taq* = 0.5µl
    - **Enhancer under test* (each enhancer was tested independently)**
    - Template
    - ddH_2O to make reaction volume to 25µl

***List of Enhancers:**

- *Q* solution – 2.5µl, 5µl, 6.25µl, 8.75µl, 10µl, 15µl, 20µl
- SequalPrep™ Enhancer A – 2.5µl

- SequalPrep™ Enhancer B – 2.5µl

- Ethylene Glycol – 3µl

- DMSO = 2.5µl (5%)

- MgCl2 = 1.5mM, 2.5mM

- PreCES (I, II, III) = 10µl

3. Molzym buffer + Molzym *Taq* polymerase evaluation experiments

- 2.5x Molzym buffer = 10µl
- Primers = 2.5µ
- Molzym *Taq* polymerase = 0.8µ
- Template = 5µl
- H_2O = 6.7µl

4. Molzym buffer and MTP *Taq* polymerase evaluation experiments

- Molzym buffer = 10µl
- Primers = 2.5ul
- MTP *Taq* polymerase = 0.25ul
- Template = 2ul
- H_2O = 10.25µl

5. MTP buffer + MTP *Taq* polymerase evaluation experiments

- 10x mtp buffer = 2.5ul
- Primers = 2.5ul
- dNTP = 0.5 ul
- MTP *Taq* polymerase = 0.25ul
- Template = 5ul

- H_2O = 14.25ul

IV. Linear Whole-genome Augmentation Procedure

One microliter of fresh Buffer A (400mM KOH, 10mM EDTA) and 1-5ng DNA were added to a cold PCR tube and allowed to denature for 3 minutes. One microliter of Buffer N (200mM HCl, 300mM Tris-HCl, pH 7.5) was added to the solution and mixed by gentle pipette tip-mixing, followed with an immediate addition of 5µl 1.2M trehalose and left on ice during master mix preparation. Master mix was prepared in a clean hood as follows: 19.8µl distilled water, 10µl 10X RXN buffer (500mM Tris-HCl pH 7.5, 100mM NaCl, 100mM $MgCl_2$, 100mM $(NH_4)_2SO_4$, 50mM DTT), 1.5µl 25mM dNTPs, 1.2µl 1nm/µl WGA primer (degenerate nonamer), 1µl 100ng/µl BSA, 57.5µl 1.2M trehalose. The master mix was briefly centrifuged and then 1µl of 1µg/µl Phi29 DNA polymerase was added to the solution. Ninety two microliters of master mix was added to each sample and pipet mixed, briefly centrifuged, and incubated at 30^oC overnight. The reaction was stopped by heating to 70^oC for 20 minutes. The reaction outcome was visualized by agarose gel electrophoresis using 4µl reaction product on a 1.5% agarose gel.

V. Purification Methods

i. E-Gel SizeSelect

E-Gel SizeSelect Agarose gels were used as per the protocol version A by Invitrogen (October 2008). The gel was loaded into the E-Gel iBase Power System and approximately 20µl of *cpn60* UT PCR product DNA was loaded into each top well of a manufacturer made E-Gel. 5µl of DNA ladder was loaded into lane M, and 25µl of ddH_2O was loaded into each bottom well, with 5µl being loaded into the corresponding lane M of the bottom row. The

system was run as per manufacturer protocol using the Run SizeSelect 2% program, for approximately 18 minutes, or until the bands reached the second set of wells. When the bands reached the reference line, the wells were refilled with 25µl of ddH_2O and DNA was collected from the wells using a pipette when the DNA bands approached the collection wells and transferred to clean Eppendorf tubes.

ii. BioRad certified low-melt agarose, Freeze 'n Squeeze DNA Gel Extraction Spin Column

Gel electrophoresis was performed as per Section 2.2.6, using Bio-Rad certified low-melt agarose (Bio-Rad, Mississauga, Canada). Gels were visualized on a DarkReader transilluminator (Clare Chemical Research, Dolores, USA) and the *cpn60* UT bands were excised and chopped into small fragments. Gel pieces were placed into a Quantum Prep Freeze 'N Squeeze DNA Gel Extraction Spin Column (Bio-Rad, Mississauga, Canada). The column was then placed in a -20^{o}C freezer for 5 minutes and then centrifuged for 3 minutes at 13,000xg at room temperature. The eluate was collected and then further purified and concentrated using the GenFind v2 SPRI magnetic bead system (Section 2.2.4).

iii. Agencourt Ampure XP

AmPure XP was run as per a modified version of the Agencourt Ampure XP protocol (Beckman-Coulter, Mississauga, Canada). This method is expected to remove DNA fragments >1000bp or <300 bp from the solution. Ten microliters of AmPure SPRI bead solution was added to 20µl of *cpn60* UT amplicon DNA, pipet tipmixed, and allowed to incubate for 5 minutes. The tubes were then placed on a magnet for 10-15 minutes and supernatant was aspirated into a tube containing 14µl AmPure magnetic beads, pipet tip-mixed, and allowed to incubate for 5 minutes. The tubes were then placed on a magnet for

10-15 minutes to allow the beads to separate from the solution and the supernatant was discarded. The beads were then washed twice with 200 µl 70% ethanol for 30 s each and allowed to air dry for 15-20 minutes. DNA was eluted from the beads by mixing with 20µl nuclease-free, DNA-free H_2O, placing the solution back on the magnet to separate the beads and transferring the eluted DNA solution to a fresh tube.

iv. Aline DNA SizeSelector-454

Seventy five microliters of DNA SizeSelector-454(Aline Biosciences, Woburn, USA) was added to thirty microliters of *cpn60* UT amplicon DNA, pipette mixed 5 times and allowed to incubate for 5 minutes at room temperature. The mixture was then placed on a magnet for 2 minutes to allow the beads to separate from the solution and the supernatant was moved to a fresh tube. Fifty microliters of DNA SizeSelector-454 was pelleted and resuspended in 20µl DNA SizeSelector-454 solution and pipette mixed with the DNA solution 5 times, then allowed to incubate for 5 minutes at room temperature. The mixture was then placed on a magnet for 2 minutes and the supernatant was discarded. The beads were washed thrice by mixing with 200µl 70% ethanol while the mixture was on the magnet. Following the final wash, the ethanol was removed and the plate was allowed to air dry for 5 minutes. The DNA was eluted by mixing the beads with 30µl of nuclease-free, DNA-free H_2O, placing on the magnet for 2 minutes and by transferring the supernatant to a clean tube.

Appendix B

I. Participant Consent Form

II. Initial Participant Questionnaire

III. Follow-up Questionnaire

RESEARCH PARTICIPANT INFORMATION AND CONSENT FORM
FOR
NASAL MICROBIAL COMMUNITY STUDY

Title of Study: "Preliminary study on the nasal microbiome of infants".

Principal Investigators: Michael Mulvey, PhD, *National Microbiology Laboratory, Assistant professor of University of Manitoba, Department of Medical Microbiology*

Co-Investigators: Morag Graham, PhD, *National Microbiology Laboratory, Adjunct professor of University of Manitoba, Department of Medical Microbiology;* George Golding, PhD, *National Microbiology Laboratory; Fiona Fleming, MD, University of Manitoba, Adjunct professor of U of M, Department of Medical Microbiology;* Gary Van Domselaar PhD, *National Microbiology Laboratory; Sergio Fanella, MD, University of Manitoba;* Joanne Embree, MD, *University of Manitoba.*

Sponsor: *Federal Genomics Research and Development Initiative (GRDI)*

You and your child are being asked to participate in a Research Study. Please take your time to review this consent form and discuss any questions you may have with the study staff. You may take your time to make your decision whether to participate in this study and you may discuss it with your friends, family or (if applicable) your doctor before you make your decision. This consent form may contain words that you do not understand. Please ask the study staff to explain any words or information that you do not clearly understand. The study investigators do not receive any professional fees but do receive financial support to conduct this study.

Purpose of Study: Bacteria are normally present in all of us and provide numerous important and protective functions. This research study is being done to look at bacteria present inside the nose of infants throughout their early development (2 weeks to ~ 12 months of age). The purpose of this study is to obtain data on the normal, developing bacterial communities inside the nose of children and compare it to the bacteria found inside the nose of the primary caregivers.

Questions to be addressed include: (a) when and what bacteria appear inside the nose (from 2 weeks - ~12 months of age)? ; (b) do bacterial communities change with standard infant vaccinations or antimicrobial usage?; (c) how different are bacterial communities inside the nose of unrelated infants? (d) Are the bacteria similar in infants and primary care givers?

Study procedures: If you agree to participate in this study, you will have the following procedures: Sterile swabs will be used to collect nasal specimens from the primary care giver (yourself) and your child. These samples will be collected at your first visit to the clinic following the birth of your child, as well as your child's next scheduled appointments of ~2, 4, 6 and 12 months. From these samples, DNA from the germs will be extracted and will be tested to identify the types of bacteria residing inside the nose of the primary care giver (yourself) and your child. Once tested, any remaining nasal specimen will be destroyed. In the unlikely event that bacterial DNA remains after testing, it will be stored at the National Microbiology Laboratory in case a new genetic test is identified in the future. There will be no health consequences of sudden withdrawal from the study. You can stop participating at any time.

Risks and Discomforts: There are no known risks during the collection of the samples from inside the nose, but there may be slight irritation at the sampling sites.

Benefits: By participating in this study you will be providing information to the investigators that will improve our understanding of germs in the nose. There may or may not be direct benefit to you or your child from participating in this study. We hope the information learned from this study will provide future benefit to other children who are prone to childhood diseases caused by bacteria. A report summarizing findings from the study will be made available to you at the conclusion of the study, which will not contain any individual test results.

Costs: All the procedures, which will be performed as part of this study, are provided at no cost to you. The study investigators are not receiving any professional fees to conduct this study.

Payment for participation: Subjects who consent to participate in the study will not be compensated for their time and effort.

Confidentiality: Information gathered in this research study may be published or presented in public forums, however your name and other identifying information will not be used or revealed. Medical records that contain your identity will be treated as confidential in accordance with the Personal Health Information Act of Manitoba. Despite efforts to keep your personal information confidential, absolute confidentiality cannot be guaranteed. Your personal information may be disclosed if required by law. All study documents related to you will bear only your assigned patient number or code. The study team at the National Microbiology Laboratory (NML) will have access to that study code, but will not be able to link it to your personal information.
The data will be entered into a study computer, located at the NML, and only the study team from NML will have access to the data. Identifying information linking your identity will be kept on a separate study computer located at the Manitoba Clinic and only the treating physicians will have access to it. No agencies will receive this electronic data. Everyone involved at the clinic in the handling of medical records and personal health information has pledged the oath of confidentiality. The rest of the investigating team will never have access to any personal information and will not be able to link the assigned study code number with your personal health records.

The Health Canada Research Ethics Board may review records related to the study for quality assurance purposes.

All records will be kept in a locked secure area and only authorized Manitoba Clinic personal will have access to these records. If any of your medical/research records need to be copied to any of the above, your name and all identifying information will be removed. No information revealing any personal information such as your name, address or telephone number will leave the clinic.

Voluntary Participation/Withdrawal from the Study: Your decision to take part in this study is voluntary. You may refuse to participate or you may withdraw from the study at any time. Your decision not to participate or to withdraw from the study will not affect your care at this centre.

Questions
You are free to ask any questions that you may have about your treatment and your rights as a research participant. If any questions come up during or after the study or if you have a research-related injury, contact the principal investigator Dr. Michael Mulvey at (204) 789-2133.

For questions about your rights as a research participant, you may contact The University of Manitoba, Bannatyne Campus Research Ethics Board Office at (204) 789-3389.

Do not sign this consent form unless you have had a chance to ask questions and have received satisfactory answers to all of your questions.

Statement of Consent

I have read this consent form. I have had the opportunity to discuss this research study with ……..and or his/her study staff. I have had my questions answered by them in language I understand. The risks and benefits have been explained to me. I believe that I have not been unduly influenced by any study team member to participate in the research study by any statements or implied statements. Any relationship (such as employer, supervisor or family member) I may have with the study team has not affected my decision to participate. I understand that I will be given a copy of this consent form after signing it. In the event that another individual has to bring in the child for a future scheduled appointment I am providing permission for samples to still be taken from the child. I understand that both my child and my participation in this study is voluntary and that I may choose to withdraw at any time. I freely agree to participate in this research study.

I understand that information regarding both my child and my personal identity will be kept confidential, but that confidentiality is not guaranteed. I authorize the inspection of any of my records that relate to this study by The University of Manitoba Research Ethics Board, for quality assurance purposes.

By signing this consent form, I have not waived any of the legal rights that I have as a participant in a research study.

Participant signature_____ Date _____

(day/month/year)

Participant printed name: _____

I, the undersigned, have fully explained the relevant details of this research study to the participant named above and believe that the participant has understood and has knowingly given their consent

Printed Name: _____Date _____

(day/month/year)

Signature: _____

Initial Questionnaire

Please complete this questionnaire to the best of your knowledge. Feel free to ask Dr. Fleming (study doctor) or your pediatrician if you have any questions.

Study ID_____

Today's Date (yyyy/mm/dd) _____

Date of Birth of Caregiver_____/____/____ (yyyy/mm/dd)
Gender of Caregiver † Male † Female

Date of Birth of Infant_____/____/____ (yyyy/mm/dd)
Gender of Infant † Male † Female

Self Claimed Ethnicity of Caregiver and Infant:

- ☐ Aboriginal
- ☐ Arab/West Asian (e.g. Armenian, Egyptian, Iranian, Lebanese, Moroccan)
- ☐ Black (e.g. African, Haitian, Jamaican, Somali)
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Latin American
- ☐ South Asian
- ☐ South East Asian
- ☐ White (Caucasian)
- ☐ Other (*specify)* _____
- ☐ Don't Know / Decline to Answer

Birth Delivery:

- ☐ Vaginal
- ☐ Caesarean
- ☐ Don't Know / Decline to Answer

Were antibiotics administered to the mother at delivery?

- ☐ Yes
- ☐ No
- ☐ Don't Know / Decline to Answer

Infant Feeding:

- ☐ Breast Milk (Approximate proportion %)_____
- ☐ Formula (Approximate proportion %)_____
- ☐ Solids (Approximate proportion %)_____

□ Don't Know / Decline to Answer

Are you a smoker?

□ Yes
□ No
□ Don't Know / Decline to Answer

Does anyone smoke inside your residence/car with you and/or your infant present?

	You	Infant
Daily	□	□
Weekly	□	□
Monthly	□	□
Less than Monthly	□	□
Never	□	□
Don't know/Decline to answer	□	□

Have you and/or your infant experienced any illness or infections in the last 2 months?

	You	Infant
Diarrhea	□	□
Ear Infection	□	□
Flu-Like Illness (ie. fever, aches, chills, cough, nausea)	□	□
Sinus (Nose) Infection	□	□
Throat Infection	□	□
Skin Infection	□	□
Urinary Tract Infection	□	□

Other (specify)_____

Have you and/or your infant taken antibiotics in the past two months?

□ No
□ Yes----------> □ Caregiver (specify)_____
_____
□ Infant (specify)_____
_____
□ Don't Know / Decline to Answer

Do you or your infant have chronic skin or medical conditions?

	You	Infant
Allergies		
Environmental (ie. Pollen, cats)	☐	☐
Food	☐	☐
Medication (ie. Penicillin)	☐	☐
Other, please specify	☐	☐
Asthma	☐	☐
Cancer	☐	☐
Diabetes	☐	☐
Eczema/Psoriasis	☐	☐
Emphysema	☐	☐
Heart Disease	☐	☐
HIV/AIDS	☐	☐
Immunosuppressive therapy	☐	☐
Kidney Disease	☐	☐
Liver Disease	☐	☐

Other chronic conditions (specify)_____

How many people are in your household?

- ☐ < 3
- ☐ 3 to 4
- ☐ 5 to 6
- ☐ > 6
- ☐ Don't Know
- ☐ Decline to Answer

Do you have any household pets?

- ☐ No
- ☐ Yes (specify)_____
- ☐ Don't Know
- ☐ Decline to Answer

Follow-up Questionnaire

Please complete this questionnaire to the best of your knowledge. Feel free to ask Dr. Fleming (study doctor) or your pediatrician if you have any questions.

*** If you are not the primary caregiver enrolled in this study - please forward this questionnaire to them. Please see the receptionist for a prepared envelope to return the completed form.**

Study ID_____

Today's Date (yyyy/mm/dd)_____

Infant Vaccinations:

- ☐ DaPTP (Diphtheria, acellular Pertussis, Tenanus, Polio)
- ☐ Hib (Haemophilus Influenzae B)
- ☐ MMR (Measles, Mumps, Rubella)
- ☐ PCV (Pneumococcal conjugate)
- ☐ MC (Meningococcal conjugate)
- ☐ Flu (Influenza)

Infant Feeding:

- ☐ Breast Milk (Approximate proportion %)_____
- ☐ Formula (Approximate proportion %)_____
- ☐ Solids (Approximate proportion %)_____
- ☐ Don't Know / Decline to Answer

Are you a smoker?

- ☐ Yes
- ☐ No
- ☐ Don't Know / Decline to Answer

Does anyone smoke inside your residence/car with you and/or your infant present?

	You	Infant
Daily	☐	☐
Weekly	☐	☐
Monthly	☐	☐
Less than Monthly	☐	☐
Never	☐	☐
Don't know/Decline to answer	☐	☐

Do you have any household pets?

- ☐ † No
- ☐ † Yes (specify)_____
- ☐ † Don't Know/Decline to Answer

Have you and/or your infant experienced any recent illness or infections within the past 2 months?

	You	Infant
Diarrhea	□	□
Ear Infection	□	□
Flu-like Illness (fever, aches, chills, cough, nausea)	□	□
Sinus (Nose) Infection	□	□
Throat Infection	□	□
Skin Infection	□	□
Urinary Tract Infection	□	□

Other (specify)_____

Antibiotic usage for you and/or your infant within the past 2 months?

□ No

□ Yes----------> □ Caregiver (specify)_____

_____

□ Infant (specify)_____

_____

□ Don't Know / Decline to Answer

□ Decline to Answer

Do you or your infant have chronic skin or medical conditions?

	You	Infant
Allergies		
Environmental (ie. Pollen, cats)	□	□
Food	□	□
Medication (ie. Penicillin)	□	□
Other, please specify	□	□
Asthma	□	□
Cancer	□	□
Diabetes	□	□
Eczema/Psoriasis	□	□
Emphysema	□	□
Heart Disease	□	□
HIV/AIDS	□	□
Immunosuppressive therapy	□	□
Kidney Disease	□	□
Liver Disease	□	□

Other chronic conditions (specify)_____

Does your infant attend daycare?

□ No

□ Yes

□ Don't Know/Decline to Answer