

ORIGINAL ARTICLE

Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age

This article has been corrected since Advance Online Publication and an erratum is also printed in this issue

Jennifer C Stearns¹, Carla J Davidson², Suzanne McKeon², Fiona J Whelan³, Michelle E Fontes¹, Anthony B Schryvers², Dawn ME Bowdish⁴, James D Kellner^{2,5} and Michael G Surette^{1,2,3}

¹Department of Medicine, McMaster University, Hamilton, Ontario, Canada; ²Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada; ³Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada; ⁴Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada and ⁵Department of Paediatrics, University of Calgary, Calgary, Alberta, Canada

The upper respiratory tract (URT) is a crucial site for host defense, as it is home to bacterial communities that both modulate host immune defense and serve as a reservoir of potential pathogens. Young children are at high risk of respiratory illness, yet the composition of their URT microbiota is not well understood. Microbial profiling of the respiratory tract has traditionally focused on culturing common respiratory pathogens, whereas recent culture-independent microbiome profiling can only report the relative abundance of bacterial populations. In the current study, we used both molecular profiling of the bacterial 16S rRNA gene and laboratory culture to examine the bacterial diversity from the oropharynx and nasopharynx of 51 healthy children with a median age of 1.1 years (range 1–4.5 years) along with 19 accompanying parents. The resulting profiles suggest that in young children the nasopharyngeal microbiota, much like the gastrointestinal tract microbiome, changes from an immature state, where it is colonized by a few dominant taxa, to a more diverse state as it matures to resemble the adult microbiota. Importantly, this difference in bacterial diversity between adults and children accompanies a change in bacterial load of three orders of magnitude. This indicates that the bacterial communities in the nasopharynx of young children have a fundamentally different structure from those in adults and suggests that maturation of this community occurs sometime during the first few years of life, a period that includes ages at which children are at the highest risk for respiratory disease.

The ISME Journal (2015) 9, 1246–1259; doi:10.1038/ismej.2014.250; published online 9 January 2015

Introduction

As the point of entry to both the lower respiratory tract and the gastrointestinal tract, the upper respiratory tract (URT) is continuously exposed to the outside world. The microbiota is involved in resistance to colonization by incoming pathogens (Kelley *et al.*, 2005; Margolis *et al.*, 2010), education of the immune system (Lathrop *et al.*, 2011) and regulation of host immunocompetence in the lung in response to infection (Ichinohe *et al.*, 2011).

In addition to its beneficial role, the commensal microbiota can also be a reservoir of respiratory pathogens as well as antibiotic resistance and virulence genes (Garcia-Rodriguez and Fresnadillo Martinez, 2002; Liu *et al.*, 2013). For instance, healthy people can have long periods of asymptomatic carriage of classic respiratory pathogens (Garcia-Rodriguez and Fresnadillo Martinez, 2002) and invasive disease can often arise from normally benign bacterial residents (Bakaletz, 2004). This duality blurs the definition of a commensal member of the respiratory tract. Instead, it's likely that this body site harbours an indigenous microbiota whose members behave differently, depending on factors such as their location in the body (Blaser and Falkow, 2009), bacterial community disturbance (Lynch, 2013), environmental pressures (Feldman and Anderson, 2013) and/or immune responses in

Correspondence: MG Surette, Department of Medicine, McMaster University, 1200 Main St West, Hamilton, Ontario L8S 4L8, Canada.

E-mail: surette@mcmaster.ca

Received 11 July 2014; revised 18 November 2014; accepted 24 November 2014; published online 9 January 2015

the host (Starkey *et al.*, 2013). Unlike many acute infectious diseases where a single microbe can be targeted and eradicated, lung infections are often polymicrobial (Bakaletz, 2004; Han *et al.*, 2012; Huang *et al.*, 2012, 2014; Dickson *et al.*, 2013) and the organisms recovered from respiratory and invasive infections are often a mixture of common URT microbes (Laupland *et al.*, 2000, Sibley *et al.*, 2008; 2012). Respiratory infections have a higher impact on health worldwide than all other infectious diseases combined (Mizgerd, 2006) and mortality rates associated with lung infections have not significantly improved in over 50 years (Mizgerd, 2008). Children, in particular, are very susceptible to respiratory illness (Liu *et al.*, 2012), therefore, an understanding of the makeup of the URT microbiota in this population is important to our understanding of the origin and progression of infection. Interactions between resident airway pathogens are known to occur, for instance, competitive exclusion between *Streptococcus pneumoniae* and *Staphylococcus aureus* (Lijek and Weiser, 2012); however, these interactions have not been studied in the context of the complete URT microbiota.

Until recently, the focus of most URT microbiology in healthy individuals has been on carriage of a few important pathogens such as *S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis* and *S. aureus* (Dunne *et al.*, 2013), however, our knowledge of colonization and succession of bacterial communities in the healthy human URT is lacking. A few microbiome studies have focused on describing the communities at these sites (Lemon *et al.*, 2010; Charlson *et al.*, 2011; Faust *et al.*, 2012), however, only Bogaert *et al.* (2011) looked in depth at children under 2 years of age, the population at greatest risk of respiratory illness and invasive pneumococcal disease. No reports, however, have looked at the healthy URT microbiota from both the perspective of molecular profiling and quantitative culture for all bacteria, an approach that provides a more complete picture of bacterial colonization. Molecular profiling can provide a snapshot of community structure, however, this picture is imperfect because of its inability to distinguish between living cells and dead cells or cell-free DNA. It also suffers from bias against some bacterial lineages, as no primer pair is perfect, and a lack of taxonomic resolution, as relatively short DNA sequences are used.

For the study herein, a subset of the swabs collected during a large point prevalence study of *S. pneumoniae* serotype carriage in the Calgary area (Ricketson *et al.*, 2014) were used. A total of 51 healthy children, with a median age of 1.1 years, from those attending community health centers for routine immunization, along with 19 accompanying parents had nasopharyngeal (NP) and oropharyngeal (OP) swabs collected. Molecular and quantitative culture profiles of the bacterial communities within the oropharynx and nasopharynx is presented in

order to describe the healthy child URT microbiome in the context of the adults with whom they have the most contact. We found that the nasopharynx of young children is dominated by a small number of bacterial groups that are present in high total numbers, in contrast to that in adults, which had much lower bacterial carriage and a more diverse bacteria community. The oropharynx communities were dominated by streptococci in all subjects and with higher bacterial biomass than the nasopharynx. This study provides the first comprehensive look at the healthy URT microbiota of young children along with those of their parents.

Materials and methods

Cohort

The Calgary Area *Streptococcus pneumoniae* Epidemiology Research (team has conducted 10-point prevalence surveys of pneumococcal NP colonization in healthy children <5 years of age attending Community Health Centres for routine immunization visits in Calgary, Canada since 2003 (Kellner *et al.*, 2008; Ricketson *et al.*, 2014). The study was approved by the University of Calgary Conjoint Health Research Ethics Board. Each survey was conducted over approximately a 6-week period in seven Community Health Centres. After obtaining written informed consent, study nurses obtained demographic information and conducted a health survey. An average of 615 children were enrolled in each survey. During the 2011 and 2012 surveys, additional written informed consent was obtained for a convenience subset of the study population to provide additional samples that were used in the present study to describe the healthy child URT microbiome. NP swabs were taken nasally and OP swabs orally according to the standard WHO method (O'Brien and Nohynek, 2003) by the public health nurse from 51 children and 19 adults (Supplementary Table S4). We used Copan eSwabs (Alere Canada, Ottawa, ON, Canada), a flocked swab with 1 ml of Amies transfer fluid. All swabs were stored at room temperature and processed within 6–8 h. For processing, the swabs in their transfer fluid were vortexed vigorously for 15–30 s, then a 100 µl subsample was taken for plating. The remainder was frozen at –20 °C for molecular analysis.

Bacterial DNA isolation and Illumina sequencing of bacterial tags

DNA was extracted from NP and OP swab samples with a custom DNA extraction protocol involving mechanical and enzymatic lysis followed by a phenol:chloroform extraction and a clean-up step. First, 300 µl of sample was added to a tube containing 0.1-mm glass beads (MoBio Laboratories Inc., Carlsbad, CA, USA) along with 800 µl of 200 mM sodium phosphate monobasic (pH 8) and 100 µl guanidinium thiocyanate EDTA *N*-lauroylsarkosine buffer (50.8 mM guanidine thiocyanate, 100 mM ethylenediaminetetraacetic acid

and 34 mM *N*-lauroylsarcosine). These were then homogenized in the PowerLyzer 24 Bench Top Homogenizer (MoBio Laboratories Inc.) for 3 min at 3000 revolutions per minute. Next, two enzymatic lysis steps were performed. In the first, the sample was incubated with 50 μ l of 100 mg ml⁻¹ lysozyme, 500 U mutanolysin and 10 μ l of 10 mg ml⁻¹ RNase for 1 hour at 37 °C. In the second, the sample was incubated with 25 μ l 25% sodium dodecyl sulphate, 25 μ l of 20 mg ml⁻¹ Proteinase K and 62.5 μ l of 5 M NaCl at 65 °C for 1 h. Next, debris was pelleted in a tabletop centrifuge at maximum speed for 5 min and the supernatant added to 900 μ l of phenol:chloroform:isoamyl alcohol (25:24:1). The sample was then vortexed and centrifuged at maximum speed in a tabletop centrifuge for 10 min. The aqueous phase was removed and the sample run through the Clean and Concentrator-25 column (Zymo Research, Irvine, CA, USA) according to kit directions except for elution, which was done with 50 μ l of ultrapure water and allowed to sit for 5 min before elution. The DNA was quantified using a Nanodrop 2000c Spectrophotometer.

Amplification of bacterial 16S rRNA gene v3 region tags was done as in (Bartram *et al.*, 2011) with the following changes: 5 pmol of primer, 200 μ M of each dNTP, 1.5 mM MgCl₂, 2 μ l of 10 mg ml⁻¹ bovine serum albumin and 1.25 U Taq polymerase (Life Technologies, Carlsbad, CA, USA) were used in a 50 μ l reaction volume. The PCR program used was as follows: 94 °C for 2 min followed by 30 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s, then a final extension step at 72 °C for 10 min. Illumina libraries were sequenced in the McMaster DNA Sequencing Facility with the following steps. Pooled libraries were first tested on an Agilent BioAnalyzer High Sensitivity DNA chip and then quantified with qPCR using Illumina's PhiX control library as a standard, SYBR fast 2 \times qPCR mastermix (KAPABiosystems, Wilmington, WA, USA) and primers that bind to the distal ends of the adaptors (flowcell-binding regions): P5 5'-AATGATACGGC GACCACCGA-3', P7 5'-CAAGCAGAAGACGGCATA CGA-3'. 16S rRNA gene v3 region pools were then combined with PhiX control DNA in a 9:1 ratio and 250bp were sequenced in the forward and reverse direction on the Illumina MiSeq instrument. The completed run was demultiplexed with Illumina's Casava software (version 1.8.2).

Sequence processing and data analysis

Custom, in-house Perl scripts were developed to process Illumina sequences and are available from the authors. First, Cutadapt (Martin, 2011) was used to trim any reads that exceeded the v3 region of the bacterial 16S rRNA gene. The resulting paired-end sequences were aligned with PANDAseq (Masella *et al.*, 2012) and the sequences with any mismatches or ambiguous bases were culled. Input sequences from all samples were clustered into operational taxonomic units (OTUs) using Abundant OTU +

(Ye, 2011), with a clustering threshold of 97%. Output from this tool was then formatted for input into QIIME, where taxonomy was assigned using the Ribosomal Database Project classifier (Wang *et al.*, 2007), with a minimum confidence cutoff of 0.8 (QIIME default) against the Greengenes (4 February 2011 release) reference database to the genus level (DeSantis *et al.*, 2006). All OTUs classified as 'Root:Other' were excluded (Supplementary Table S1) along with three samples with poor sequence coverage and aberrant microbial profiles.

The α -diversity measures used were observed OTUs, which is an estimate of community richness, and Shannon diversity, which is an estimate of community diversity based on richness and evenness of the community (Shannon and Weaver, 1949). β -diversity, which is a measure of the differences between communities, was calculated with the weighted UniFrac distance and clustering of the samples based on distance was illustrated with principal coordinate analysis. OTUs containing a single sequence (singletons) were removed prior to β -diversity calculations. Weighted UniFrac distance was calculated on evenly sampled profiles (sample depth of 2400 sequences, without replacement). The number of clusters were evaluated with the gap statistic, which provides an estimation of how well each *k*, or number of clusters, fits the data. Gap statistic estimation was done with 500 Monte Carlo samples in R with the use of the phyloseq (McMurdie and Holmes, 2013) clusGap wrapper. In order to resolve sample-clustering jackknife analysis using weighted UniFrac distance was performed on evenly sampled OTU tables (10 tables, 2400 sequences per sample) and a composite unweighted pair group method with arithmetic mean (UPGMA) tree of the samples was created. α - and β -diversity calculations and plots were done in R with the phyloseq package, except for jackknifed UPGMA trees, which were done in QIIME version 1.8.0 (Caporaso *et al.*, 2010). Taxonomic summaries and multiple response permutation procedure with 1000 permutations, were also both done with QIIME. A χ^2 -test of each age group among NP clades (Ib, IIa and IIb) was done in R with 500 *P*-value simulations. OTUs that differed significantly between adult and child samples were selected from NP and OP data separately using LefSe, which combines a non-parametric Kruskal–Wallis sum-rank test to determine which OTUs have a different abundance between ages, a unpaired Wilcoxon rank-sum test to determine which age is statistically relevant and effect size is estimated using linear discriminant analysis (Segata *et al.*, 2011), a default cutoff of $P < 0.05$ and linear discriminant analysis effect of > 2.0 was used. OTUs with a total relative abundance across all samples of $< 0.11\%$ were removed from the data prior to LefSe analysis, as were samples from children over 19 months age (all of which were between 4 and 4.5 years of age) owing to the small relative number of samples from this group (6/69 for NP and 4/49 for OP). All OTU

comparisons between groups were corrected for multiple testing using the false discovery rate (Benjamini and Hochberg, 1995) with the *p.adjust* function in R.

Streptococcus species phylogeny

In order to resolve the species distribution of OTUs and cultured isolates assigned to the *Streptococcus* genus, a phylogenetic placement method was used, whereby these short sequences were placed onto a reference tree made from reference strains of *Streptococcus* species from the Human Oral Microbiome Database (Chen *et al.*, 2010). All full-length 16S rRNA gene reference sequences for *Streptococcus* species (29 in total) from Human Oral Microbiome Database were aligned using MUSCLE (multiple sequence comparison by log-expectation; Edgar, 2004) then used to create a maximum likelihood phylogeny with RaXML (randomized accelerated maximum likelihood; Stamatakis, 2006; Silvestro and Michalak, 2012) using the Generalized Time Reversible + Gamma evolutionary model + a proportion of (I) invariable sites and the rapid bootstrap method over 10 000 reps (Supplementary Figure S1A). A separate phylogeny created with MrBayes (Ronquist and Huelsenbeck, 2003) using the same evolutionary model (GTR + G + I) over 10 million generations had an identical topology to the maximum likelihood tree with better confidence for branching (Supplementary Figure S1B). Variation in likelihood values visualized with Tracer v1.6 (Rambaut and Drummond, 2009) were used to confirm convergence of the bayesian tree. Short sequences were placed onto the maximum likelihood phylogeny with pplacer (Matsen *et al.*, 2010). Confidence in each placement was measured with expected distance between placement locations (Supplementary Table S8) and was good for most OTU sequences.

Quantitative culture

Quantitative culture for NP swabs from 20 children and five of their parents and OP swabs for four of these adult-child pairs was done using five media types (BHI CO, brain heart infusion supplemented with colistin sulfate ($10 \mu\text{g ml}^{-1}$) and oxolinic acid ($5 \mu\text{g ml}^{-1}$); CBA, Columbia blood agar; Chocolate agar; McKay agar; and MSA, mannitol salt agar) and incubated aerobically at 37°C with 5% CO_2 . Three dilutions (10^{-1} , 10^{-3} and 10^{-5}) were plated on each media type and incubated for 48 h. Morphologically similar colonies from each plate were counted and used to calculate colony-forming unit (CFU) ml^{-1} for each isolate. A representative isolate from each type and media was picked and stored in 10% skim milk at -80°C . The 16S rRNA gene from each of >300 cultured isolates was amplified using the 8F–926R primers (Liu *et al.*, 1997), and the resulting 700–900 bp product was

sequenced in the forward direction by Beckman Coulter Genomics (Danvers, MA, USA). Low-quality regions of each sequences were trimmed for 1% error probability in Geneious (v5.6.6) then taxonomy was assigned by BLAST alignment (Altschul *et al.*, 1990) to the Human Oral Microbiome Database and counts were compiled into Supplementary Table S7. All *e*-values were below $7.26\text{e}-115$ over the entire length of the partial 16S rRNA gene sequences used.

Results

Bacterial communities within the nasopharynx and oropharynx are distinct from one another

The development of the URT microbiota is poorly understood, although some studies have described pathogen carriage in healthy people, few have looked at the commensal microbiota in healthy young children. To investigate the microbiome of young children, we obtained NP and OP swabs from healthy children during routine immunization in Calgary, AB, Canada. Most of the children were between 12 and 19 months, with a small subset aged between 4 and 5 years. The parents of these children were asked to volunteer samples, and in total, we analyzed NP swabs from 69 participants (51 from children and 18 from adults and OP swabs for 34 children and 15 adults, for a total of 118 samples, all except two of which were matching (Supplementary Table S4).

Ecological diversity measures were used to distinguish differences in bacterial 16S rRNA gene molecular profiles from the nasopharynx and the oropharynx in healthy children and adults. Figure 1a shows α -diversity estimates, which provide an indication of how the bacterial communities in each sample were structured. The number of observed OTUs is a straightforward count of the number of unique OTUs in each sample and Shannon diversity is an estimate of community diversity based on the richness and evenness of all OTUs within a sample. Both the observed number of OTUs and Shannon diversity estimates for NP samples had a broader range than did those for OP samples. Also, adults had a significantly higher number of observed OTUs and higher Shannon diversity in the nasopharynx than did the children <19 months of age at either site swabbed ($P < 0.05$). Together, this suggests that bacterial communities in the adult nasopharynx were more diverse than those in the nasopharynx or the oropharynx of young children.

β -diversity estimates complement α -diversity by illustrating the difference in community membership between samples and the weighted UniFrac distance shown in Figure 1b uses both the abundance and phylogenetic relationship of the OTUs in each sample to determine how similar samples are to one another. We suspected that the relative

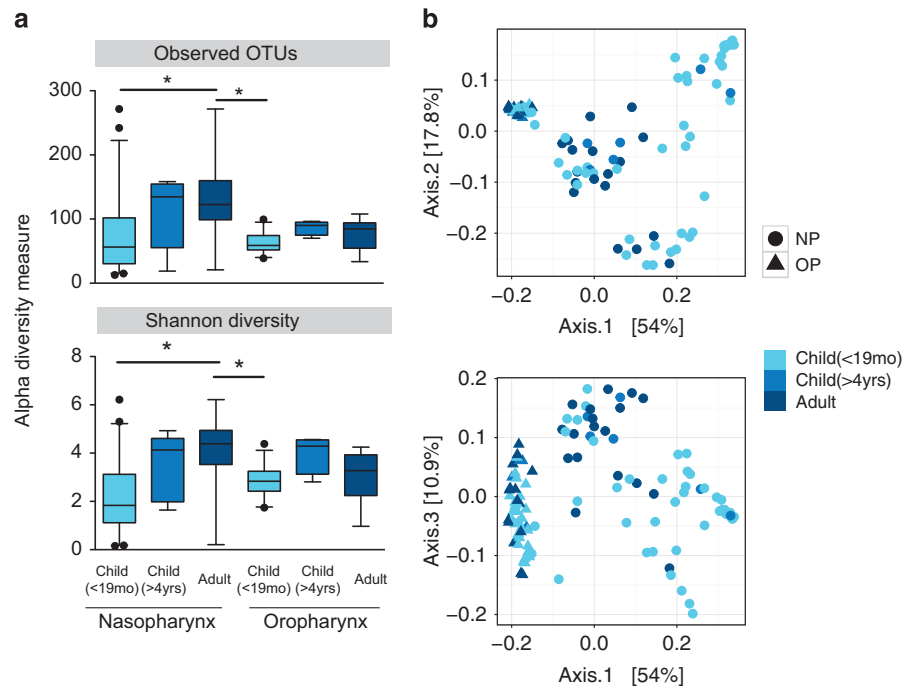


Figure 1 Oropharyngeal and nasopharyngeal communities are distinct. (a) Alpha diversity of oropharyngeal and nasopharyngeal microbial communities in adults and children have a range of values, however, adult NP samples appear to have higher diversity than OP or child NP samples. Whiskers are from the 10–90 percentile and statistical significance are for $P < 0.05$ (Kruskal–Wallis and Dunn’s multiple comparison test). (b) Beta diversity of the microbial profiles within the oropharyngeal and nasopharyngeal samples show separation by sample site. The first three principal coordinates of a principal coordinate analysis, based on weighted UniFrac distance, shows distinct clustering of samples by swab geography as opposed to age of the subject. Tight clustering of adult and child OP samples suggests similarity in these profiles, whereas scattering of NP samples suggests dissimilarity in the profiles.

abundance of bacterial groups would be essential for describing differences between samples and indeed jackknife β -diversity showed good reproducibility for weighted measures, weighted UniFrac and Bray–Curtis (not shown), and less confidence for unweighted UniFrac distances (Supplementary Figure S2). Of the metadata available, only swab site and age corresponded to separation of samples in the principal coordinate analysis (Figure 1b). Bacterial communities from the nasopharynx or oropharynx were distinctly separated from each other in the ordination in Figure 1b and were significantly different from each other based on a multiple response permutation procedure ($A = 0.25$, significance of delta = 0.001), indicating that distinct bacterial communities likely colonize each site. It is also apparent that OP bacterial communities in different people were all very similar to one another, compared with the heterogeneity in the NP samples, as distances were smaller between any two OP samples (0.127 ± 0.054) than between any two NP samples (0.276 ± 0.095) ($P < 0.0001$; Supplementary Figure S3).

Adults and children have different bacterial communities in the nasopharynx

The relative taxonomic profiles for adult or child samples from each site were averaged in order to clearly illustrate the differences between them

(Figure 2a). These plots show the diversity differences calculated using α -diversity estimates (Figure 1a). Note that bacterial groups within each profile were identified down to the best classification possible on the basis of taxonomic information contained within the v3 region of the 16S rRNA genes, hence many bacterial groups are labeled with the genus and others by only the family or order name. The oropharynx of both children and adults not only had a high proportion of *Streptococcus* but also contained bacterial groups often cultured from this site such as *Rothia*, *Prevotella*, *Gemella*, *Veillonella*, *Fusobacteria*, *Haemophilus* and *Neisseria* (Figure 2 and for a complete list see Supplementary Table S2). NP samples from adults had a large proportion of Firmicutes such as Lachnospiraceae, *Staphylococcus* and *Streptococcus*, Bacteroidetes such as *Sphingobacterium* and *Prevotella* and Actinobacteria other than *Corynebacterium* such as *Bifidobacteria*, *Rothia* and *Propionibacterium*. Child NP bacterial communities contained similar groups but had larger proportions of Proteobacteria such as *Moraxella*, Enterobacteriaceae and *Haemophilus*, as well as the Firmicute *Enterococcus* (Figure 2 and for a complete list see Supplementary Table S3).

To determine whether there was distinct clustering of samples, as suggested by Figure 1b, weighted UniFrac distances calculated above were used to determine how well estimates of each of 1–10

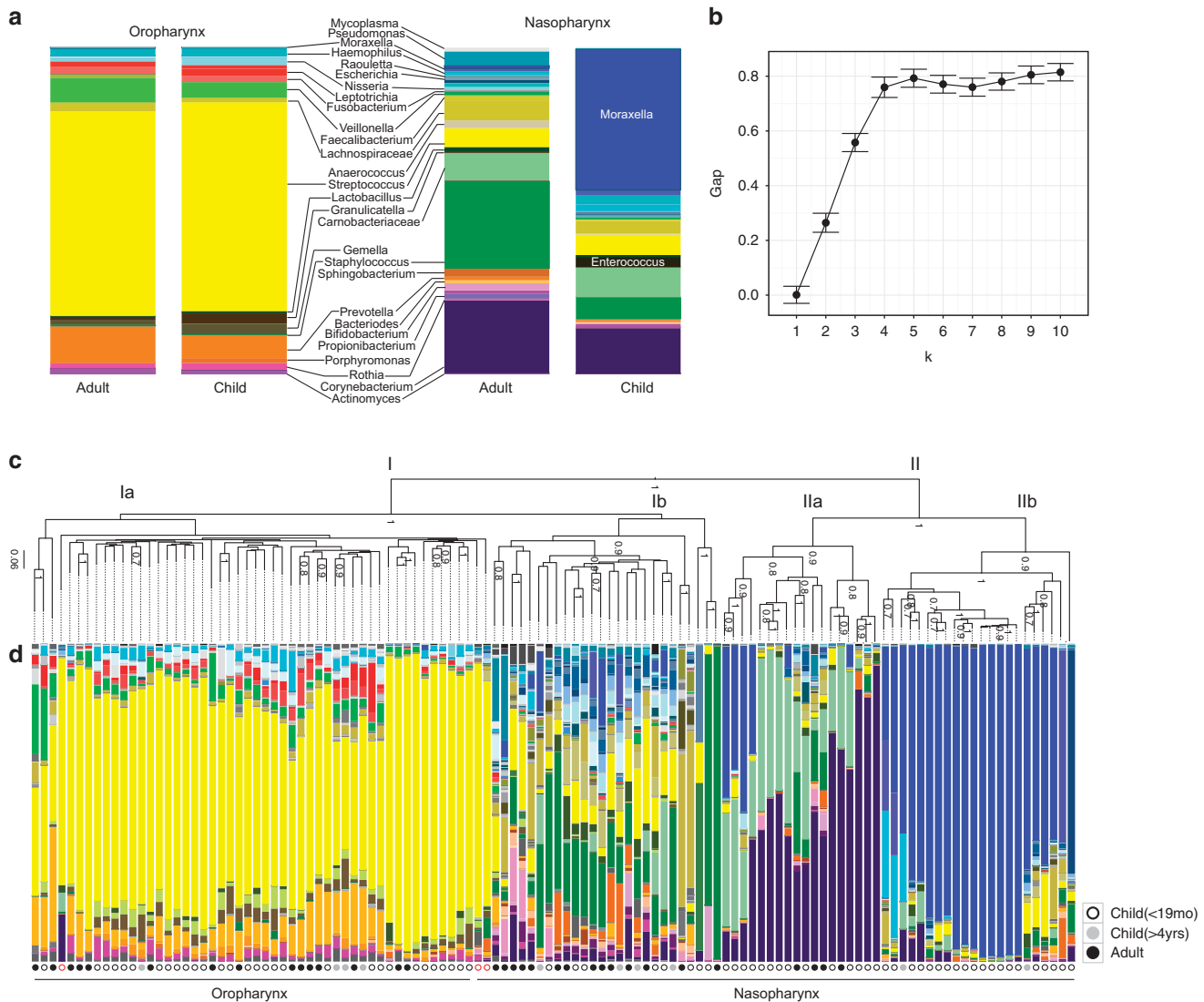


Figure 2 Taxonomic profiles illustrate the differences between groups. **(a)** Microbial community profiles, summarized down to the genus level when possible, for adults and children at both sample sites. **(b)** Estimation of the gap statistic for each of 1–10 clusters of samples, based on weighted UniFrac distances. **(c)** UPGMA tree of weighted UniFrac distances with jackknife support (only values above 0.6 are shown in the tree). **(d)** Taxonomic profiles were ordered based on their position in the UPGMA phenogram labeled by age and sample site (colours are the same as in **a**). Similar colour groups were used for each phyla: Actinobacteria (purple/pink), Bacteroidetes (orange), Cyanobacteria (grey), Firmicutes (green/yellow/brown), Fusobacteria (red), Proteobacteria (blue), and Tenericutes (black). A complete list of taxa is available in Supplementary Tables S2 (OP) and S3 (NP). Red circles in **d** indicate NP samples that were grouped with OP samples.

clusters fit the data with a gap statistic calculation. Gap statistic values increased steadily then level off for $k=4$, suggesting that there were four clusters of samples (Figure 2b). Differences between profiles were illustrated using UPGMA hierarchical clustering with jackknife support, also based on weighted UniFrac distances. The resulting tree, shown in Figure 2c, branches into two main clades (I and II) and these each branch again into two subclades (Ia, Ib, IIa and IIb), separation of which had good jackknife support. Taxonomic profiles for each sample were ordered according to the UPGMA phenogram (Figure 2d) and the pattern that emerged illustrates the differences between samples that were suggested in Figure 1.

Separation of samples into clades and subclades was due to taxonomic differences within them, as UPGMA was based on the weighted UniFrac distance between samples. For instance, clade I contains profiles with a high proportion of Firmicutes, whereas clade II contains samples with a high proportion of Gram-negative bacterial taxa as well as Actinobacteria. Subclade Ia includes all bacterial profiles from the oropharynx, as well as three NP samples that likely grouped with them owing to the high proportion of *Streptococcus*. Subclades Ib, IIa and IIb contain all of the NP swab samples, with the exception of the previous three. Subclade Ib was the most taxonomically diverse, whereas subclades IIa was dominated by Actinobacteria (specifically

Corynebacterium) and/or Firmicutes (specifically Carnobacteriaceae) and subclade IIb was dominated by Proteobacteria (*Moraxella*). A small number of samples (4/41) had all three taxonomic groups. A subset of subclade IIb (3/22) had a substantial proportion of *Haemophilus* without *Streptococcus* and another subset (3/22) with the opposite composition. From this ordering of samples, we can see that subclades containing NP samples clustered by age of the subjects (χ^2 , $P=0.002$, over that expected by chance). Subclade Ib was enriched with samples from adult nasopharynx and subclades IIa and IIb were enriched with samples from the nasopharynx of young children (<19 months). Of note is the proportion of the children over the age of 4 years that were in subclade Ib (4/6).

To test whether adults and children from the same family had URT microbial communities that were more similar to each other than to other people, weighted UniFrac distances were calculated between each sample and shown for each related pair (adult vs child from the same family), each unrelated pair (adult vs child not from the same family) and between all adult or all child samples (Figure 3). In the nasopharynx, the microbial profiles of adults had the smallest distance from one another. The profiles of child NP communities were most different from that of other children, the mean of which was significantly different from the mean distances between adults and adults with unrelated children ($P<0.05$). In the oropharynx, however, the microbial profiles of children had the smallest distance from one another. Distances between adults and adults and unrelated children were higher and significantly different from the child–child distances ($P<0.05$).

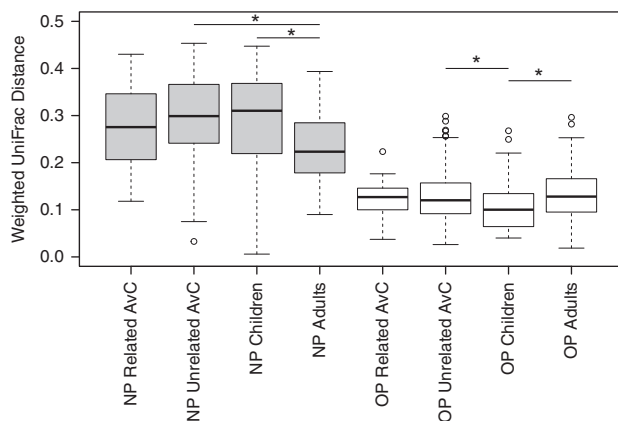


Figure 3 Distances between pairs of samples show community differences between sites and ages of subjects. Weighted UniFrac distances between each pair of samples were averaged for: related adults and children, adult and child samples from all other subjects (i.e., called ‘unrelated’), between all adults and between all children. Asterisks indicate significantly different means tested with a Kruskal–Wallis test and a Dunn’s multiple comparison test, $P<0.05$).

Bacterial OTUs differ significantly between adult and child microbial communities

Bacterial OTUs, rather than taxa summaries, were analyzed with LefSe to identify statistically significant differences between adult and child microbial profiles, as OTUs with the same taxonomic assignment sometimes varied between adults and children. Owing to the small number of samples from older children (4/49 for OP and 6/69 for NP), these were excluded from the analysis. In the oropharynx, microbial communities in young children had a significantly higher proportion of 91 different OTUs, including *Streptococcus* (OTU3), *Gemella* (OTU 14), *Haemophilus* (OTU 15), *Neisseria* (OTU 6), *Porphyromonas* (OTU 30), *Granulicatella* (OTU 24 and 25), an unclassified Fusobacteriaceae (OTU 36), among others (Figure 4a). Interestingly, several different *Prevotella* OTUs were significantly different between adults and children, with some more prevalent in children and others in adults. Microbial communities in the oropharynx of adults had a higher proportion of *Veillonella* (OTU 2) and Lachnospiraceae (OTU 13) among others. Within the nasopharynx, there were 82 statistically significant OTUs, many more that were associated with adult samples than with child samples (Figure 4b). Of the three OTUs significantly associated with communities in young children, only *Moraxella* (OTU 1) occurred in high abundance. OTUs significantly associated with adult samples included *Staphylococcus* (OTU 7), Lachnospiraceae (OTU 13), *Streptococcus* (OTU 2), *Anaerococcus* (OTU 41) and *Pseudomonas* (OTU 27), among others (a complete list is shown in Supplementary Table S5). Individual abundance histogram plots for all OTU shown in Figure 4a (NP) and 4b (OP) are presented in Supplementary Figure S4 and S5, respectively, and complete lists of linear discriminant analysis effect sizes and P -values for OP and NP are presented in Supplementary Table S5 and S6, respectively.

Streptococcus diversity in the oropharynx and the nasopharynx

Some of the most prevalent bacteria found in this study were the *Streptococcus*, especially in the oropharynx. Although OTU classification may have underrepresented species as defined by classical taxonomic methods, all of the dominant populations except the *Streptococcus* were represented by a single OTU. It is difficult to resolve the taxonomy of the streptococci based on differences within the 16S rRNA gene and particularly based on only the short V3 region. We nonetheless endeavored to identify, which species of *Streptococcus* were represented by the most abundant OTUs using the pplacer phylogenetic placement method. The *Streptococcus* phylogeny created from full-length 16S rRNA gene sequences for species within the Human Oral Microbiome Database (Supplementary Figures S6)

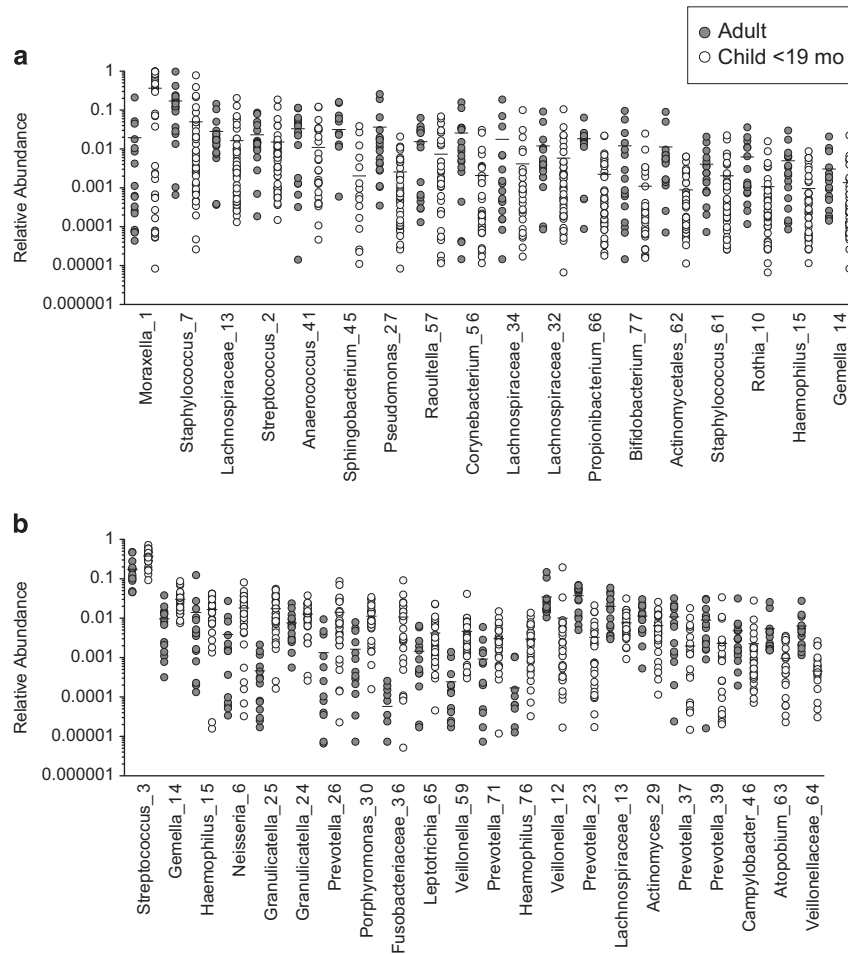


Figure 4 Bacterial OTUs are significantly different between profiles from adults or young children at each site. **(a)** Bacterial OTUs from the nasopharynx and **(b)** the oropharynx. Only those OTUs with a total relative abundance across all samples of $>0.11\%$ are shown here along with an abundance histogram for each OTU (Supplementary Figures S4 and S5). A complete list of OTUs tested along with effect sizes and *fdr* adjusted *P*-values for each site is given in Supplementary Tables S5 and S6.

was able to separate most species with the exception of a few closely related species such as *Streptococcus oralis*, *Streptococcus mitis* and *S. pneumoniae*; and between *Streptococcus intermedius*, *Streptococcus constellatus* and *Streptococcus anginosus* (known as the *Streptococcus milleri/anginosus* group). Despite having good separation between *S. salivarius* and *S. vestibularis* in the reference phylogeny pplacer was unable to place many of sequences from isolated strains squarely with either one, therefore although OTU 2 was placed closer to *S. vestibularis*, we don't have confidence about its exact identity and believe that it likely represents a mix of members of this clade (called salivarius group here; Supplementary Figure S6). The second most highly abundant *Streptococcus* OTU (OTU 3) was placed within the clade containing *S. mitis*, *S. pneumoniae* and *S. oralis*, which we designated as the mitis group. Lastly, OTU 55 was placed with a smaller more distinct clade made up of members of the *Streptococcus milleri/anginosus* group (*S. anginosus*, *S. intermedius* and *S. constellatus*) and likely represents one, or more, of these species.

The taxa summaries from Figure 2a were recoloured to highlight the distribution of the three most abundant *Streptococcus* OTUs, which were different in each group of samples (Figure 5). The oropharynx profiles of children had more mitis group than salivarius group *Streptococcus*, however, adults harboured more salivarius group than mitis group *Streptococcus*, as well as a small proportion of the milleri/anginosus group *Streptococcus* that was lacking in children. Likewise in the nasopharynx, child profiles carried more mitis group than salivarius group *Streptococcus* and lacked the milleri/anginosus group *Streptococcus* found in the adult nasopharynx. This group fluctuated greatly in the adult subset sometimes representing up to 23% of the community. This illustrates the diversity of *streptococci* within the URT, which was masked by the taxa summary profiles at the genus level and suggests a role for this group in the maturation of the NP bacterial community, where either a lack of *streptococci* or dominance of one population is replaced by a different distribution of the three groups of species.

Quantitative culture adds context to molecular profiles
Direct sequencing of bacterial 16S rRNA genes provided very detailed profiles for bacterial communities in the URT of healthy people. Quantitative culture, however, can provide absolute numbers of each cultivable bacterial population, providing context to relative abundance profiles by adding total abundances. Although only a small proportion of environmental bacteria have been successfully cultured (Stewart, 2012), ~84% of human airway-

associated bacteria have been cultured with these methods from Sibley *et al.* (2011). Quantitative culturing of samples from 20 children and five adults illustrated that total bacterial load varied greatly with location and age (Figure 6a). Bacterial load in the oropharynx was higher than that in the nasopharynx for both adults and children, with children carrying a higher total CFU ml⁻¹ than adults. In the nasopharynx, cultures from children showed a higher overall median number of

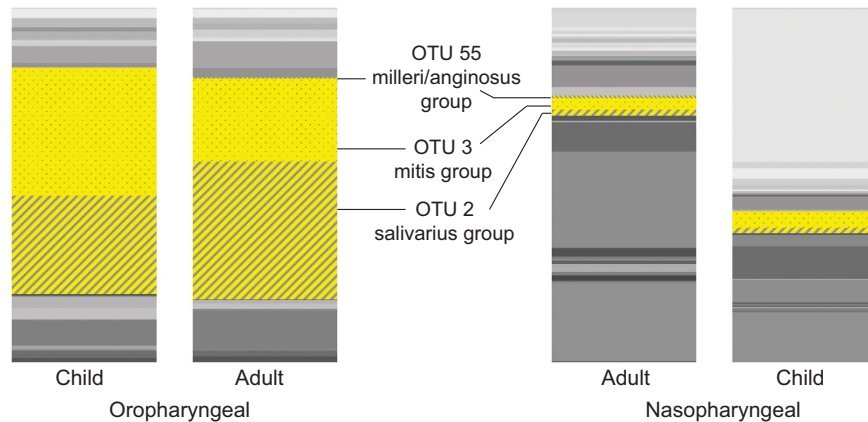


Figure 5 Diversity of *Streptococcus* OTUs differs between groups of samples. Taxa summaries from Figure 2a were recoloured for the three most abundant *Streptococcus* OTUs. The species group assigned to each OTU is derived from placement of each OTU within a phylogeny of *Streptococcal* species (Supplementary Figure S6) and from the taxonomy of cultured isolates, based on similarity with sequences within HOMD as shown by BLAST (Supplementary Table S8).

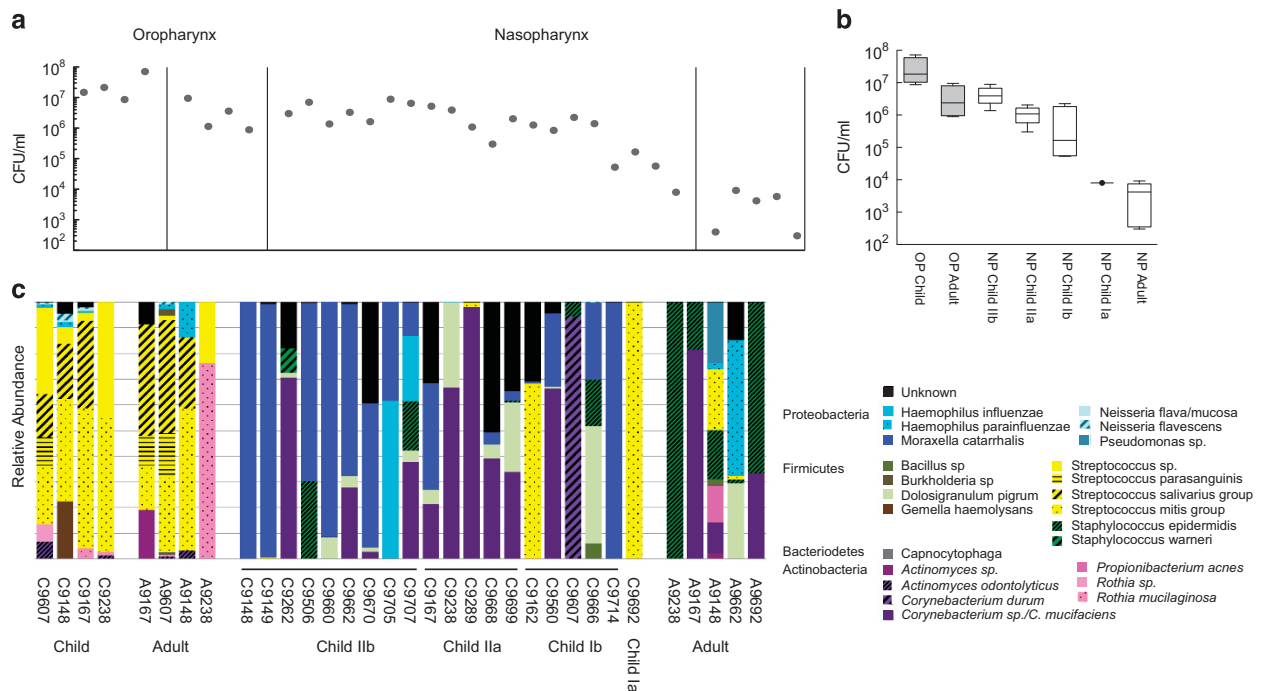


Figure 6 Total bacterial biomass differs between samples. (a) Total bacterial counts for all bacterial species cultured. Samples from the oropharynx had the highest bacterial biomass, whereas samples from the nasopharynx of adults had the lowest bacterial biomass. None of the differences here were statistically significant based on a Kruskal–Wallis test and a Dunn's multiple comparison's test. (b) Range of total bacterial counts obtained for both sample site and age represented by geometric mean and 10–90 percent indicated by the whiskers. (c) Stacked bar chart of the relative abundance of each bacterial species/group cultured. Owing to the length of the sequence obtained from cultured isolates, species designations could be made. Members of the same Phylum are coloured with the same hue (Actinobacteria: pink/purple; Bacteroidetes: grey; Firmicutes: yellow/green/brown; Proteobacteria: blues; unknown: black).

CFU ml⁻¹ (1.5×10^6 vs 4.2×10^3) than for adults, a difference of almost three orders of magnitude.

This is especially interesting in the context of data presented in Figure 2, as profiles that were grouped together into subclades had a different amount of total bacteria than those in other subclades. Figure 6b illustrates this point by ordering profiles according to their bacterial load, with those from child NP samples split into their respective subclades. Next to oropharyngeal samples, bacterial load was highest in samples from subclade IIb (those dominated by *Moraxella*), with between 10^6 and 10^7 CFU ml⁻¹. Adult NP communities had the lowest, with between 5×10^2 and 10^4 CFU ml⁻¹ and samples in subclade IIa (dominated by *Corynebacterium* and *Carnobacteriaceae*) had an intermediate amount. Interestingly, child samples within subclade Ib, which had profiles that resembled to those in adults, had lower biomass than those of the other children in the study but a higher amount than adults. As there was a large disparity in the number of samples included in each group, a Kruskal–Wallis test found no statistically significant difference in the CFU ml⁻¹ between groups.

Partial 16S rRNA gene sequences (~500–700 bp) from cultured isolates were also useful in suggesting species level taxonomy for prevalent groups from direct sequenced molecular profiles in Figure 2. For instance, the only *Moraxella* species cultured was *M. catarrhalis* and it was present in high abundance. Likewise, the only species of *Carnobacteriaceae* found in NP samples was *Dolosigranulum pigrum*, which again was found in high numbers, providing a species designation for a group, which was only classified to the family level in the molecular profiles. The *Streptococcus* species, however, were numerous and varied, and it seems likely that, for this group especially, OTU designations were not adequate to describe a species. For example, OTU 2, which was identified above as belonging to the salivarius group, likely contained more than one species, as these were isolated in abundance from the oropharynx and their partial 16S rRNA gene sequences were placed between *S. vestibularis* and *S. salivarius* along with the representative sequence for OTU 2 on the reference phylogeny (Supplementary Figure S6). Likewise, OTU 3 likely contained members from several species, including *S. mitis*, *S. pneumoniae* and *S. oralis* for the same reason (for a complete list of species, see Supplementary Table S7).

Discussion

Characterization of the microbial communities in the URT is important as, even in healthy people, it can harbour both pathogens and potentially pathogenic commensal organisms (Garcia-Rodriguez and Fresnadillo Martinez, 2002; Bakaletz, 2004; Liu *et al.*, 2013). These are responsible for acute respiratory infections such as pneumonia and

invasive disease, such as empyema and sepsis. In addition to the prevalence of pathogens, a comprehensive investigation of microbes within the URT can provide insight into community structure, aspects of which are important to disease susceptibility. Children, who are at an increased risk of respiratory infection, are known to have high carriage rates of pathogenic bacteria (Pelton, 2012), however, their URT microbiome has not been fully characterized. Here, molecular and culture-based methods were used to compare microbial communities from the nasopharynx and oropharynx of young children (aged 1–4.5 years, median 1.1) and their parents in hopes of describing the immature URT microbiota.

As expected from molecular studies of the nasopharynx and oropharynx (Keijser *et al.*, 2008; Lazarevic *et al.*, 2009; Nasidze *et al.*, 2009; Lemon *et al.*, 2010; Charlson *et al.*, 2011), within this study, these two niches harboured distinct bacterial communities. OP bacterial communities of children and adults were dominated by *streptococci* (Figure 2) and the fact that bacterial profiles were so similar, regardless of age, suggests that the colonization of the oropharynx occurs early and that communities are quickly dominated by *Streptococcus*. The oropharynx is known to contain mainly *Prevotella*, *Veillonella* and *Streptococcus* (Keijser *et al.*, 2008; Lazarevic *et al.*, 2009; Nasidze *et al.*, 2009), with the latter being more prevalent here than previously shown. Earlier studies may have underestimated the prevalence of this group, and Gram-positive bacteria in general, as our DNA isolation protocol has been optimized to capture nearly all of the bacterial diversity within the respiratory tract (Sibley *et al.*, 2011). A test of the differences in UniFrac distances between samples showed that the bacterial communities in the oropharynx of children were all most similar to one another and that adult profiles were most different from one another or those of unrelated children (Figure 3). Although few obvious taxonomic differences could be seen between adult and child profiles from the oropharynx, several OTUs were significantly different (Figure 4a). Of interest is the fact that OTU 3, which was characterized as a mitis group *Streptococcus* and was highly abundant in all OP samples, was significantly associated with young children. Also, several OTUs belonging to the genera *Prevotella* had significantly different proportions in the adult and the young child oropharynx. This is a complex genus with many species that may colonize people differently as they age.

NP bacterial communities were heterogeneous but overall adult samples had higher diversity (Figure 1a) and lower bacterial abundance (Figure 6b). Samples from children contained a range of bacteria, with some dominated by Proteobacteria (*Moraxella*), others containing a large proportion of Actinobacteria and Firmicutes (*Corynebacterium* and *D. pigrum*), and others resembling

adult NP swabs. Interestingly, within the *Moraxella*-dominated group, the evidence of competitive exclusion between *Haemophilus* and *Streptococcus* was seen in a small number of samples, an observation previously made during studies on the interactions of individual URT resident pathogens (Lijek and Weiser, 2012). A small number of samples were dominated by streptococci and hence resembled swabs from the oropharynx. When all samples were clustered based on phylogenetic differences, four groups emerged two of which were enriched with samples from the nasopharynx of children (Figure 2c). Similarly, an ordination of NP microbial profiles from children in the Netherlands of similar age revealed four clusters, three dominated by a single OTU (*Moraxella*, *Haemophilus* or *Streptococcus*) and one, which was mixed (Bogaert *et al.*, 2011). It is unknown if the fourth mixed group of child samples resembled adult profiles, as none were included, however, this aligns well with our data, suggesting that NP bacterial communities of young children often have low diversity owing to the dominance by a small number of bacterial species. The prevalence of *Corynebacterium* and *D. pigrum* instead of *Haemophilus* could be due to seasonal or geographical influence on bacterial carriage, yet may also be due to better representation of Gram-positive bacteria in our work. The NP profiles of adults were most similar to one another and distances were greatest between communities in children and between those in children and unrelated adults (Figure 3). The distances between children does not take into account the fact that there were two different groups of children.

An important finding from this study was that not only did child NP samples differ in bacterial membership, but that they also had different total bacterial numbers (Figure 6). Overall, bacterial numbers differed by several orders of magnitude between adult and child samples, with children always carrying a higher number than adults at each site. Within each group of child NP samples, differences in the bacterial load suggests further differences between the microbial communities within these individuals that were not apparent with molecular profiles alone. This observation does echo alpha diversity, calculated from molecular profiles (Figure 1a), where NP samples from young children had lower species richness and community diversity, compared with adult profiles. These observations make sense in the context of community maturation, as lower biomass and higher diversity are hallmarks of a climax community (Loucks, 1970), which is stable (McCann, 2000). Caution needs to be taken with molecular profiles based on DNA and high throughput sequencing as, when there are differences in bacterial biomass, an equal depth of sampling will always uncover both real and spurious low abundance OTUs (Schloss *et al.*, 2011) and microbial community richness becomes difficult to estimate without inflating

richness and diversity estimates (Haegeman *et al.*, 2013). Why child NP communities should have such high bacterial loads is unknown but could be related to increased immune tolerance (Jaspan *et al.*, 2006) in the immature nasopharynx. Longitudinal quantitative culture studies designed to capture commensal bacterial diversity are needed to answer this more completely.

Maturation of the host-associated microbial community is known to occur in gut communities where progression from colonization to climax and stability (Adlerberth and Wold, 2009), as well as refraction of this community to disturbance (Dethlefsen *et al.*, 2008), has been studied. It is also, then, possible that a similar effect occurs in the airways, where a newborn child is colonized by bacteria from their surroundings and that these bacterial populations succeed one another until an equilibrium is reached and a stable community persists. In the gut, a healthy stable microbiota protects a person from disease (Round and Mazmanian, 2009), likewise a healthy, stable URT microbiota may have a similar role. In fact, the microbial composition has been shown to regulate immune modulation in response to viral pathogens in the URT (Ichinohe *et al.*, 2011). What impact a person's immune system has on the composition of the NP microbial community is not yet known. The profiles presented here of children whose nasopharynx had high bacterial load of one or two dominant organisms (that is, *M. catarrhalis*, *D. pigrum* and *Corynebacterium sp.*) could be an illustration of the immature NP bacterial community that is still dynamic and will eventually settle to one with lower biomass that is more stable and resembles what is found in mid-aged adults. Whether the children harbouring high *Moraxella* numbers will have different health outcomes from the others is impossible to say without longitudinal studies, however, this has been true for rates of acute otitis media in young children (Faden *et al.*, 1994). Children, along with the elderly, are at the highest risk for respiratory infection. In a similar study from our laboratory, the nasal microbiota within elderly nursing home patients were seen to have very little similarity between people, suggesting a loss of community order when compared with mid-aged adults (Whelan *et al.*, 2014). Differences in bacterial community structure within the URT of children and the elderly, when compared with mid-aged adults, may be contributing to the susceptibility of people within these two age groups to infection.

The *Streptococcus* have a range of phenotypes and behaviours in association with human mucosal sites. Whereas some, like *S. pneumoniae* can behave pathogenically in some people (Walker *et al.*, 2013), others such as *S. salivarius* and *S. mitis* are found commonly in healthy human mouths. Not surprisingly, this group was prevalent here, found abundantly in OP samples of both children and adults, seen to dominate a small number of immature NP samples and seen

often in mature NP samples. At the OTU level, the nasopharynx and oropharynx contained two abundant and one less abundant OTU with several other low abundant streptococci. However, due to the similarity of the 16S rRNA gene within these short sequences, it is difficult to resolve species identification. In fact when longer sequences, obtained from cultured isolates, were classified, hidden diversity within *Streptococcal* OTUs was revealed: OTUs 2 and 3 and 55 likely contained a mix of salivarius group, mitis group and milleri/anginosus group streptococci, respectively (Supplementary Figure S6). Therefore, from culture, we can identify more than one dominant species within each *Streptococcus* OTU, a fact that was not seen with the other abundant genera found. This was not surprising given both the diversity of streptococci traditionally found in the URT and the difficulty in distinguishing between streptococcal species with DNA profiling. Even culture methods based on colony morphology have been shown to underestimate the diversity of closely related species (Sibley *et al.*, 2011), and it is likely that the *Streptococcus* diversity is greater than estimated for these samples.

Several studies of URT colonization by bacteria, while measuring carriage rates of specific pathogens in the juvenile population, have described populations using mainly clinical culture methods (Vanechoutte *et al.*, 1990; Bogaert *et al.*, 2004; Regev-Yochay *et al.*, 2004). Although these studies describe the presence of specific pathogens, they neglect to describe the remainder of the bacterial community, which may contribute to polymicrobial lower airway infection (Bakaletz, 2004; Han *et al.*, 2012; Huang *et al.*, 2012; Dickson *et al.*, 2013; Huang *et al.*, 2014). Many of the bacterial strains living benignly in the healthy URT can act pathogenically in the lower airways and elsewhere. Here, the microbial landscape of the URT in healthy children and their parents presents aspects such as community membership and bacterial load that are likely important for child health and disease, as well as highlighting further questions about when and how the microbiota matures in the nasopharynx and what role the host immune system has in this process.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We would like to thank Laura Rossi and Christine King for their contribution to the Illumina library preps and sequencing and their contributions to this manuscript. This work was funded by the Canadian Institute for Health Research. MGS was supported as a Canada Research Chair in Interdisciplinary Microbiome Research. DMEB was supported in part by the McMaster Immunology Research Centre. DMEB and MGS were supported in part by the MG DeGroote Institute for Infectious Disease Research.

References

- Adlerberth I, Wold AE. (2009). Establishment of the gut microbiota in western infants. *Acta paediatr* **98**: 229–238.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Bakaletz LO. (2004). Developing animal models for polymicrobial diseases. *Nat Rev Microbiol* **2**: 552–568.
- Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl Environ Microbiol* **77**: 3846–3852.
- Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* **57**: 289–300.
- Blaser MJ, Falkow S. (2009). What are the consequences of the disappearing human microbiota? *Nat Rev Microbiol* **7**: 887–894.
- Bogaert D, Van Belkum A, Sluijter M, Lijndijk A, De Groot R, Rümke HC *et al.* (2004). Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *Lancet* **363**: 1871–1872.
- Bogaert D, Keijsers B, Huse S, Rossen J, Veenhoven R, van Gils E *et al.* (2011). Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* **6**: e17035.
- Caporaso GJ, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A *et al.* (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J of Respir Crit Care Med* **184**: 957–963.
- Chen T, Yu WH, Izard J, Baranova O, Lakshmanan A, Dewhirst F. (2010). The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* **2010**: baq013.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.
- Dethlefsen L, Huse S, Sogin ML, Relman DA. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.
- Dickson RP, Erb-Downward JR, Huffnagle GB. (2013). The role of the bacterial microbiome in lung disease. *Expert Rev Respir Med* **7**: 245–257.
- Dunne EM, Smith-Vaughan HC, Robins-Browne RM, Mulholland EK, Satzke C. (2013). Nasopharyngeal microbial interactions in the era of pneumococcal conjugate vaccination. *Vaccine* **31**: 2333–2342.
- Edgar RC. (2004). MUSCLE: a multiple sequence alignment with high accuracy and high throughput. *BMC Bioinformatics* **5**: 113.
- Faden H, Harabuchi Y, Hong J, Tonawanda/Williamsville Pediatrics (1994). Epidemiology of *Moraxella catarrhalis* in children during the first 2 years of life: relationship to otitis media. *J Infect Dis* **169**: 1312–1317.

- Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J *et al.* (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comp Biol* **8**: e1002606.
- Feldman C, Anderson R. (2013). Cigarette smoking and mechanisms of susceptibility to infections of the respiratory tract and other organ systems. *J Infect* **67**: 169–184.
- Garcia-Rodriguez JA, Fresnadillo Martinez MJ. (2002). Dynamics of nasopharyngeal colonization by potential respiratory pathogens. *J Antimicrob Chemother* **50**: 59–74.
- Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS. (2013). Robust estimation of microbial diversity in theory and in practice. *ISME J* **7**: 1092–1101.
- Han MK, Huang YJ, LiPuma JJ, Boushey HA, Boucher RC, Cookson WO *et al.* (2012). Significance of the microbiome in obstructive lung disease. *Thorax* **67**: 456–463.
- Huang YJ, Lynch SV, Wiener-Kronish JP. (2012). From microbe to microbiota: considering microbial community composition in infections and airway diseases. *Am J Resp Crit Care Med* **185**: 691–692.
- Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. (2014). Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *J Clin Microbiol* **52**: 2813–2823.
- Ichinohe T, Pang IK, Kumamoto Y, Peaper DR, Ho JH, Murray TS *et al.* (2011). Microbiota regulates immune defense against respiratory tract influenza A virus infection. *PNAS* **108**: 5354–5359.
- Jaspan HB, Lawn SD, Safrit JT, Bekker LG. (2006). The maturing immune system: Implications for development and testing HIV-1 vaccines for children and adolescents. *Aids* **20**: 483–494.
- Keijser B, Zaura E, Huse S, van der Vossen J, Schuren F, Montijn R *et al.* (2008). Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* **87**: 1016–1020.
- Kellner JD, Scheifele D, Vanderkooi OG, MacDonald J, Church DL, Tyrrell GJ. (2008). Effects of routine infant vaccination with the 7-valent pneumococcal conjugate vaccine on nasopharyngeal colonization with *Streptococcus pneumoniae* in children in Calgary, Canada. *Pediatr Infect Dis J* **27**: 526–532.
- Kelly D, Conway S, Aminov R. (2005). Commensal gut bacteria: mechanisms of immune modulation. *Trends Immunol* **26**: 326–333.
- Lathrop SK, Bloom SM, Rao SM, Nutsch K, Lio CW, Santacruz N *et al.* (2011). Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478**: 250–254.
- Laupland KB, Davies HD, Low DE, Schwartz B, Green K, McGeer A. (2000). Invasive group A *Streptococcal* disease in children and association with varicella-zoster virus Infection. *Pediatrics* **105**: e60–e60.
- Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Osteras M *et al.* (2009). Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* **79**: 266–271.
- Lemon KP, Klepac-ceraj V, Schiffer HK, Brodie EL, Lynch SV, Kolter R. (2010). Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. *mBio* **1**: e00129–10.
- Lijek RS, Weiser JN. (2012). Co-infection subverts mucosal immunity in the upper respiratory tract. *Curr Opin Immunol* **24**: 417–423.
- Liu H, Wang H, Huang M, Mei Y, Gu B, Wu R *et al.* (2013). Analysis of antimicrobial resistance and class 1 integrons among strains from upper respiratory tract of healthy adults. *J Thor Dis* **5**: 149.
- Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE *et al.* (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**: 2151–2161.
- Liu WT, Marsh TL, Cheng H, Forney LJ. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516–4522.
- Loucks OL. (1970). Evolution of diversity, efficiency, and community stability. *Am Zool* **10**: 17–25.
- Lynch SV. (2013). Probiotic bacteria and their effect on human health and well-being. In: Guarino A, Quigley E, & Walker W (eds). *World Rev Nutr Diet*, Vol 107. Karger: Basel, Switzerland, pp 64–71.
- Margolis E, Yates A, Levin BR. (2010). The ecology of nasal colonization of *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus*: the role of competition and interactions with host's immune response. *BMC Microbiol* **10**: 59.
- Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10.
- Masella AP, Bartram A, Truszkowski JM, Brown DG, Neufeld JD. (2012). Pandaseq: paired-end assembler for Illumina sequences. *BMC Bioinformatics* **13**: 31.
- Matsen FA, Kodner RB, Armbrust EV. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics* **11**: 538.
- McCann KS. (2000). The diversity-stability debate. *Nature* **405**: 228–233.
- McMurdie PJ, Holmes S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One* **8**: e61217.
- Mizgerd J. (2006). Lung infection—a public health priority. *PLoS Med* **3**: e76.
- Mizgerd J. (2008). Acute lower respiratory tract infection. *N Engl J Med* **358**: 716–727.
- Nasidze I, Li J, Quinque D, Tang K, Stoneking M. (2009). Global diversity in the human salivary microbiome. *Genome Res* **19**: 636–643.
- O'Brien K, Nohynek H. (2003). Report from a WHO working group: standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*. *Pediatr Infect Dis J* **22**: e1–e11.
- Pelton SI. (2012). Regulation of bacterial trafficking in the nasopharynx. *Paediatr Res Rev* **13**: 150–153.
- Rambaut A, Drummond ATracer v1.6 (2009). Available at <http://beast.bio.ed.ac.uk/software/Tracer>.
- Regev-Yochay G, Dagan R, Raz M, Carmeli Y, Shainberg B, Derazne E *et al.* (2004). Association between carriage of *Streptococcus pneumoniae* and *Staphylococcus aureus* in children. *JAMA* **292**: 716–720.
- Ricketson LJ, Wood ML, Vanderkooi OG, MacDonald JC, Kellner JD. (2014). Trends in asymptomatic nasopharyngeal colonization with *Streptococcus pneumoniae* after introduction of the 13-valent pneumococcal

- conjugate vaccine in Calgary, Canada. *Pediatr Infect Dis J* **33**: 724–730.
- Ronquist F, Huelsenbeck JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Round JL, Mazmanian SK. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**: 313–323.
- Schloss PD, Gevers D, Westcott SL. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* **6**: e27310.
- Segata N, Izard J, Walron L, Gevers D, Miropolsky L, Garrett W *et al*. (2011). Metagenomic biomarker discovery and explanation. *Gen Biol* **12**: R60.
- Shannon CE, Weaver W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press: Urbana 5.
- Sibley CD, Church DL, Surette MG, Dowd SE, Parkins MD. (2012). Pyrosequencing reveals the complex polymicrobial nature of invasive pyogenic infections: microbial constituents of empyema, liver abscess, and intracerebral abscess. *Eur J Clin Microbiol Infect Dis* **31**: 2679–2691.
- Sibley CD, Grinwis ME, Field TR, Eshaghurshan CS, Faria MM, Dowd SE *et al*. (2011). Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS One* **6**: 11.
- Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC, Surette MG. (2008). A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *PNAS* **105**: 15070–15075.
- Silvestro D, Michalak I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org Divers and Evol* **12**: 335–337.
- Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Starkey MR, Nguyen DH, Kim RY, Nair PM, Brown AC, Essifie AT *et al*. (2013). Programming of the lung in early life by bacterial infections predisposes to chronic respiratory disease. *Clin Obstet Gynecol* **56**: 566–576.
- Stewart EJ. (2012). Growing unculturable bacteria. *J Bacteriol* **194**: 4151–4160.
- Vaneechoutte M, Verschraegen G, Claeys G, Weise B, Van den Abeele AM. (1990). Respiratory tract carrier rates of Moraxella (Branhamella) catarrhalis in adults and children and interpretation of the isolation of M. catarrhalis from sputum. *J Clin Microbiol* **28**: 2674–2680.
- Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta Z *et al*. (2013). Global burden of childhood pneumonia and diarrhoea. *Lancet* **381**: 1405–1416.
- Wang Q, George M, Tiedje J, Cole J. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *App Environ Microbiol* **73**: 5261–5267.
- Whelan F, Verschoor C, Stearns J, Rossi L, Johnstone J, Surette M *et al*. (2014). The loss of topography in the microbial communities of the upper respiratory tract in the elderly. *Ann Am Thor Soc* **11**: 513–521.
- Ye Y. (2011). Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 153–157; doi: 10.1109/BIBM.2010.5706555.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)