

RESEARCH ARTICLE

Human pharyngeal microbiota in age-related macular degeneration

Eliza Xin Pei Ho¹*, Chui Ming Gemmy Cheung², Shuzhen Sim¹, Collins Wenhan Chu¹, Andreas Wilm¹, Clarabelle Bitong Lin¹, Ranjana Mathur², Doric Wong², Choi Mun Chan², Mayuri Bhagarva³, Augustinus Laude⁴, Tock Han Lim⁴, Tien Yin Wong², Ching Yu Cheng², Sonia Davila¹, Martin Hibberd^{1,5}

1 Genome Institute of Singapore, Singapore, Singapore, **2** Singapore Eye Research Institute, Singapore National Eye Center, Duke-NUS Medical School, National University of Singapore, Singapore, Singapore, **3** Department of Ophthalmology, National University of Singapore and National University Health System, Singapore, Singapore, **4** National Healthcare Group Eye Institute, Tan Tock Seng Hospital, Singapore, Singapore, **5** Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

* These authors contributed equally to this work.

* hoxpe@gis.a-star.edu.sg



OPEN ACCESS

Citation: Ho EXP, Cheung CMG, Sim S, Chu CW, Wilm A, Lin CB, et al. (2018) Human pharyngeal microbiota in age-related macular degeneration. PLoS ONE 13(8): e0201768. <https://doi.org/10.1371/journal.pone.0201768>

Editor: Simon J. Clark, University of Manchester, UNITED KINGDOM

Received: April 20, 2018

Accepted: July 20, 2018

Published: August 8, 2018

Copyright: © 2018 Ho et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw sequence data are available from ENA, study accession number PRJEB15613.

Funding: This project is funded by grant number 10/1/35/19/671 from the Biomedical Research Council, Singapore. The funders had no role in study design, data collection, analysis, and interpretation, manuscript preparation, or decision to publish.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Background

While the aetiology of age-related macular degeneration (AMD)—a major blinding disease—remains unknown, the disease is strongly associated with variants in the complement factor H (CFH) gene. CFH variants also confer susceptibility to invasive infection with several bacterial colonizers of the nasopharyngeal mucosa. This shared susceptibility locus implicates complement deregulation as a common disease mechanism, and suggests the possibility that microbial interactions with host complement may trigger AMD. In this study, we address this possibility by testing the hypothesis that AMD is associated with specific microbial colonization of the human nasopharynx.

Results

High-throughput Illumina sequencing of the V3-V6 region of the microbial 16S ribosomal RNA gene was used to comprehensively and accurately describe the human pharyngeal microbiome, at genus level, in 245 AMD patients and 386 controls. Based on mean and differential microbial abundance analyses, we determined an overview of the pharyngeal microbiota, as well as candidate genera (*Prevotella* and *Gemella*) suggesting an association towards AMD health and disease conditions.

Conclusions

Utilizing an extensive study population from Singapore, our results provided an accurate description of the pharyngeal microbiota profiles in AMD health and disease conditions. Through identification of candidate genera that are different between conditions, we provide preliminary evidence for the existence of microbial triggers for AMD.

Ethical approval for this study was obtained through the Singapore Health Clinical Institutional Review Board, reference numbers R799/63/2010 and 2010/585/A.

Introduction

The human throat harbors a complex bacterial community located at the intersection of the digestive and respiratory tracts. This pharyngeal microbiome is dominated by the phyla Firmicutes and Bacteroidetes [1] and is thought to protect against respiratory tract infections and invasive disease by preventing the outgrowth of potentially pathogenic bacteria (reviewed in [2]).

The human complement system is a major innate immune defense against meningococcal disease. Briefly, the binding of complement proteins to antigen-antibody complexes (classical pathway) or to the pathogen surface (alternative pathway) activates a triggered-enzyme cascade, resulting in an amplified response that brings about pathogen lysis, pathogen opsonization, and the recruitment and activation of phagocytes.

To prevent damage to host cells, this potent response is kept under regulatory control. This host cell protection mechanism nonetheless can be exploited by microbes developing complement evasion mechanisms that aid in their pathogenicity [3]. These pathogens include *Neisseria meningitidis* and *Streptococcus pneumoniae*, which are carried asymptotically in the nasopharynx [4,5]. These microbes exploit this dampening mechanism by factor H sequestration, via binding to domains 6–7 or through carboxyl-terminal domains 19–20 on factor H [5]. The complement factor H (*CFH*) gene encodes a negative regulator of the alternative complement pathway. Individuals homozygous for risk variants of *CFH*, when exposed to *Chlamydia pneumoniae*, were found to face an increased risk of age-related macular degeneration (AMD) progression [6]. Similarly, variants in *CFH* were found to show strong associations with susceptibility to meningococcal disease [7].

Intriguingly, the *CFH* gene is also associated with susceptibility to AMD [8–9], a blinding disease that is an important cause of vision loss in the elderly. It is thought that the disease will afflict 288 million people globally by 2040 [10].

Early-stage AMD, which presents minimal symptoms and is usually detected through routine eye examinations, is characterized by the presence of drusen deposits within the ocular sub-retinal space. Most vision loss occurs during advanced stage disease, which takes two main forms: geographic atrophy ("dry" AMD), characterized by degeneration of large, confluent regions of the retinal pigmented epithelium, and neovascular ("wet") AMD, characterized by invasion of the retina by leaky choroidal blood vessels and macrophages (reviewed in [11]).

Despite extensive research, the etiology of AMD is unknown. It is well known, however, that *CFH* risk variants encode proteins that bind with lower affinity to Bruch's membrane (the multilayered extracellular matrix separating the retina from choroidal blood vessels) and to other complement pathway components, potentially resulting in an impaired inhibitory effect [12–14]. In addition, increased alternative pathway activity has been reported for several AMD-associated risk variants in other complement pathway components and regulators, with protective variants having the opposite effect [15]. Thus, one hypothesis for AMD etiopathogenesis is complement over-activation in response to injury and debris in the retina, rendering increased exposure to complement-mediated damage to host cells. There are postulations that interactions between microbial colonizers and host complement could provide a trigger for AMD [6,16].

In this study, we used 16S rRNA sequencing to comprehensively characterize the pharyngeal microbiomes of 260 AMD cases and 386 controls, thus allowing us to investigate possible roles for specific microbial genera in the pathogenesis of AMD.

Materials and methods

Study population and sample collection

Following protocols described previously, we recruited 311 AMD cases at three clinical sites: Singapore National Eye Centre, Tan Tock Seng Hospital and National University Hospital [17]. Ethical Approval was obtained through the Singapore Health Clinical Institutional Review Board (R799/63/2010; 2010/585/A), with written informed consent from participants. AMD diagnosis was made according to standard definitions based on ophthalmic examinations, including dilated fundus photography, fluorescein angiography, indocyanine green angiography and optical coherence tomography. Grading of fluorescein angiograms for presence of choroidal neovascularization was done using a modification from the Macular Photocoagulation Study [18]. Indocyanine green angiography was done to diagnose definitive polypoidal choroidal vasculopathy (PCV), using the Japanese Study Group guidelines [19]. Similarly, 421 healthy participants from the Singapore Chinese Eye Study without AMD nor carrying risk factors such as hypertension, hyperopic refraction and chronic kidney disease, were recruited as control subjects [17, 20]. Throat swabs were stored immediately at -80°C upon collection until deoxyribonucleic acid (DNA) template extraction [21].

Nucleic acid extraction, 16S rRNA amplification and DNA sequencing

Extraction of total DNA from throat swabs was performed via a combination of mechanical and chemical lysis using FastPrep Instrument (MP Biomedicals) and Qiagen DNeasy EZ1 Advanced XL system (Qiagen). Extraction of samples was conducted without differentiating between Case/Control statuses. An approximate 750bp region (encompassing V3-V6) on prokaryotic 16S rRNA was amplified from total DNA as described previously [22,23]. In summary, we purified DNA amplicons from 35 cycles of PCR using Qiagen Hotstar (Germany), which yielded approximately $1\mu\text{g}$ of nucleic acids. Amplicons were sheared and constructed into sequencing libraries using GeneRead DNA Library I Core Kit (Qiagen) according to the manufacturer's protocol. DNA libraries were multiplexed using Illumina 48-plex barcodes and 76bp paired-end sequenced on the Illumina HiSeq 2500. After demultiplexing (with the Illumina bcl2fastq 2.17.1.14 software) and removal of reads that failed Illumina's purity filters (PF = 0), reads were converted to FASTQ format. Reads were then trimmed by removal of trailing bases with quality score ≤ 2 . Read pairs containing reads shorter than 60bp were also removed [24].

Reconstruction and classification of 16S rRNA amplicon sequences

To reconstruct the original amplicon (V3-V6) sequences, we followed the workflow described previously [22,23], which is largely based on EMIRGE [24], an expectation maximization method not only reconstructs amplicons but also provides estimates of relative taxon abundances. In brief, quality-trimmed reads (see above) were input into EMIRGE (GIT version 98787b5). EMIRGE performs template-guided "assembly" based on a modified SILVA SSU (version 102) database that contains sequences between 1200 – 1900bp, using an expectation-maximization algorithm to iterate, align, and assign reads to candidate 16S sequences [24]. This iterative mapping of paired-end reads also prevents chimeric sequences from mapping. The EMIRGE-based reconstruction methodology has been evaluated and benchmarked

against modQIIME and RTAX, and was found to be comparatively more robust and to produce highly concordant estimates of taxonomic abundance [22], thus clearly demonstrates its utility for profiling microbiomes in a precise, high-resolution manner.

To reduce computational requirements and runtime, we applied EMIRGE to the top (in terms of average quality) 500,000 reads in each sample [22]. We found this number to be robust enough to accurately reflect the 16S composition in each sample (S1 Fig and S1 Table.) [22]. De facto outputs indicate OTU abundance; where sequences with relative abundance below 0.1% were removed for data quality control.

EMIRGE-reconstructed sequences were trimmed to the primer-amplified section and searched using BLAST against the chimera-checked Greengenes 16S rRNA database (current_GREENGENES_gg16S_unaligned.fasta) [25]. BLAST hits were sorted by (in consecutive order) smallest E-value, highest bit score, highest percent identity, and longest alignment length; only the top hit after this sorting was used for classification. Minimum percent identity levels were set at 80%, 90%, and 95% for the phylum, family, and genus levels respectively; hits below these percentages were dropped and not considered for classification purposes.

Characterization of microbial community composition

EMIRGE assigns abundance estimates to reconstructed sequences, allowing us to directly use these results to characterize community composition at various taxonomic levels. A relative abundance profile of OTUs was generated for each sample, excluding those that failed to meet the minimum percent identity levels after BLAST (laid out above).

Throat microbiome profiles were successfully obtained (i.e. sufficient 16S PCR product for library building, and >500,000 high quality reads such that EMIRGE reconstruction succeeded) from 245 out of 311 AMD-positive throat swabs, and from 386 out of 421 control throat swabs. Out of the 245 cases, 80 were classified as early-stage AMD and 165 as late-stage AMD (Table 1). Smoking status was available for 126/245 cases and 386/386 controls; only these microbiomes were used for the analysis of the impact of smoking on the pharyngeal microbiome.

Data visualization and statistical analyses

Abundance estimates from EMIRGE were converted to relative abundances at the genus level for each sample. Clinical data, OTU relative abundance tables, Shannon's and Simpson's α -diversity indices were processed and calculated using customized R script in R version 3.3.2. OTU relative abundance tables were input for Principal Component Analyses (PCA). The R package "DESeq2" was used to determine differential abundance for case/control conditions (in all samples), and early/late AMD conditions (in case samples). Age and gender were included as covariates. Only taxa found to be significant ($P < 0.05$) were reported.

Generalised Linear Mixed Model (GLMM) was performed to determine genera with significantly different relative abundances between conditions. P-values were corrected for multiple testing using Benjamini-Hochberg correction.

We used "SVA" R package to examine batch effects across samples. Additionally, Principal Coordinate Analyses (PCoA) and Guided Principal Component Analyses (gPCA) [26] were performed to visually inspect for batch effects using "stats" R package.

Quantitative-PCR (qPCR) of total bacteria and selected microbial genera

Absolute gene counts of total *Prevotella*, *Gemella* and *Streptococcus* were determined with quantitative PCR (Q-PCR). Twenty samples were randomly selected from among the AMD case population for 16S qPCR gene copy determination; this was matched by an equal number

Table 1. Clinical parameters of study subjects.

	Cases	Controls	Total
No. of individuals with swabs	311	421	732
No. of sequenced individuals w/ clinical data	245	386	631
	Mean ± Standard Deviation (Range) or n (%)		
Age	Case	Control	Chi Square (p-value)
<60	41	196	1.58 x 10 ⁻¹⁷
>60	204	190	
Mean age ± Standard Deviation	67.46 ± 8.57 (32–85.26)	60.76 ± 9.95 (45.82–84.39)	
Gender			
Female	88	197	2.76 x 10 ⁻⁴
Male	157	189	
Smoking Status			
Current or Previous smoking	48	94	4 x 10 ⁻³
Non-smoking	78	292	
Late AMD			
Geographic Atrophy (GA)	7		
Polypoidal Choroidal Vasculopathy (PCV)	80		
Typical Neovascular AMD (tAMD)	71		
Mixed Atrophy	7		
Early AMD	80		
Control		386	

<https://doi.org/10.1371/journal.pone.0201768.t001>

of randomly selected control samples. To quantify total bacteria, primers were selected that amplify a uniform 16S region in these genera (S2 Table) [27]. *Prevotella* and *Streptococcus* primers were sourced from Matsuki et al. and Picard et al. respectively [28, 29], while *Gemella* primers were established in-house via multiple sequence alignments of *Gemella* 16S sequences obtained from the NCBI *nr* repository. Primers were tested to ensure specificity to corresponding genera.

Reactions were carried out in 384-well qPCR plates with duplicate 10 µL reaction volumes containing the KAPA SYBR FAST qPCR Master Mix (2X) for LightCycler 480 (Sigma-Aldrich, Inc.), the primers (15 µM) and 2.0 µL of sample DNA. The LightCycler® 480 Instrument II (Roche) was used for amplification and detection with the following thermocycling parameters: 1 cycle of 95°C for 20s followed by amplification at 95°C for 15s, 60°C for 20s for 40 cycles and 1 cycle of 95°C for 15s, 60°C for 15s and 95°C for 15s with readings collected at final step for melting curve analysis. Ct values were recalculated for log copy number/ul DNA from raw data exported into Excel.

Calibration standards for converting Ct values to bacterial 16S rRNA gene copy numbers were generated as follows. Double-stranded DNA oligomers were synthesised (gBLOCK, Integrated DNA Technologies, Inc.) to span respective regions covered by 0020corresponding *Prevotella/Gemella/Streptococcus* forward and reverse primers. A 10 ng/µL stock solution was prepared from the lyophilised oligomers as per manufacturer’s instruction. Copy number/µL of stock solution were determined from calculations (DNA concentration / (fragment length x weight of base pair)) and serially diluted to 6 standards with final concentrations spanning 1.0

$\times 10^3$ to 1.0×10^8 copy numbers. A plot of Ct versus $\log_{10}(\text{copy number})$ produced linear calibration curves with typical R^2 values of 0.99.

Copy numbers for each genera were expressed as percentage of total microbial copy number. Pearson's correlation coefficient between qPCR and 16S sequencing techniques was determined using Microsoft Excel 2013.

Results

Pharyngeal microbiome structure in AMD cases and controls

We successfully obtained bacterial 16S rRNA sequence data from 631 out of 732 individual throat swabs. Of these, 245 were derived from AMD cases, and 386 from controls. Clinical parameters, including age, gender, smoking status, and disease stage are shown in [Table 1](#). With the exception of disease-type variable for AMD cases, pairwise statistical comparisons (PERMANOVA) of microbial community abundance among case and control samples revealed no obvious stratification of pharyngeal microbiome profiles by gender and age variables ([S3 Table](#)). No significant surrogate variables were identified from "SVA" analysis (data not shown). Visual inspection for batch effects using Principal Coordinate Analysis (PCoA), similarly indicated an absence of batch effects among samples ([S2 Fig](#)). Finally, application of Guided Principal Components Analysis (gPCA) revealed that proportion of variance due to batch effects was not significantly greater than would be expected ($p < 0.05$) ([S3 Fig](#)).

In total, nine phyla and 63 genera were detected across all 631 subjects. The mean number of identified genera were 17 ($SD = 6$) and 18 ($SD = 5$) in 245 cases and 386 controls respectively. This reflects some degree of similarity in the number of identified genera from samples collected from a specific anatomical niche, albeit with varied disease status. Consistent with previous reports [1], two dominant phyla—Firmicutes and Bacteroidetes—accounted for more than half of all OTUs. Members of the microbial community in case and control samples were similar, however, relative abundances of each member differed between case ($n = 245$) and control ($n = 386$) groups ([Fig 1](#)). In terms of genera prevalence among study population, *Streptococcus*, *Prevotella*, *Veilonella* and *Gemella* were most prevalent in both case ($n = 245$) and control ($n = 386$) samples, being detected in $>85\%$ of all subjects ([S4 Fig](#)). Shannon and Simpson diversity indices did not differ significantly between cases and controls ($p < 0.05$, Mann-Whitney U test) ([S5 Fig](#)).

Principal Components Analysis (PCA) revealed the lack of clustering across conditions, implying that maximum abundance variances are insufficient to distinguish between conditions. It also suggests that genera relative abundances between conditions are subtle ([S3 Fig](#)).

Association of specific microbial genera with AMD

245 case and 386 control samples were used for DESeq2 to evaluate differential abundance of bacterial genera. DESeq2 analyses found the case samples to be enriched for *Gemella* ($Adj-p = 3.15 \times 10^{-5}$), while *Prevotella* had reduced relative abundance ($Adj-p = 1.89 \times 10^{-5}$) ([Fig 2](#)). While, in the comparisons between early versus late AMD cases, no genera was detected to be significantly different in relative abundance among the groups. Likewise was in the smoking status comparisons ($Adj-p < 0.05$). Using Generalised Linear Mixed Model analysis, we observed the robust associations of *Prevotella* and *Gemella* with case/control status, as similarly shown in the DESeq2 analysis. In addition, we detected *Streptococcus* and *Leptotrichia* to associate with case/control status as well ([Fig 3](#) and [S4 Table](#)). *Streptococcus* and *Prevotella*, the two most abundant genera, showed an opposite trend: *Streptococcus* was present at a significantly higher abundance in cases than in controls (mean relative abundance of 23.4% vs 18.6%; $Adj-p = 0.002$), while *Prevotella* was present at a significantly lower abundance (12.7% vs 19.3%; $Adj-$

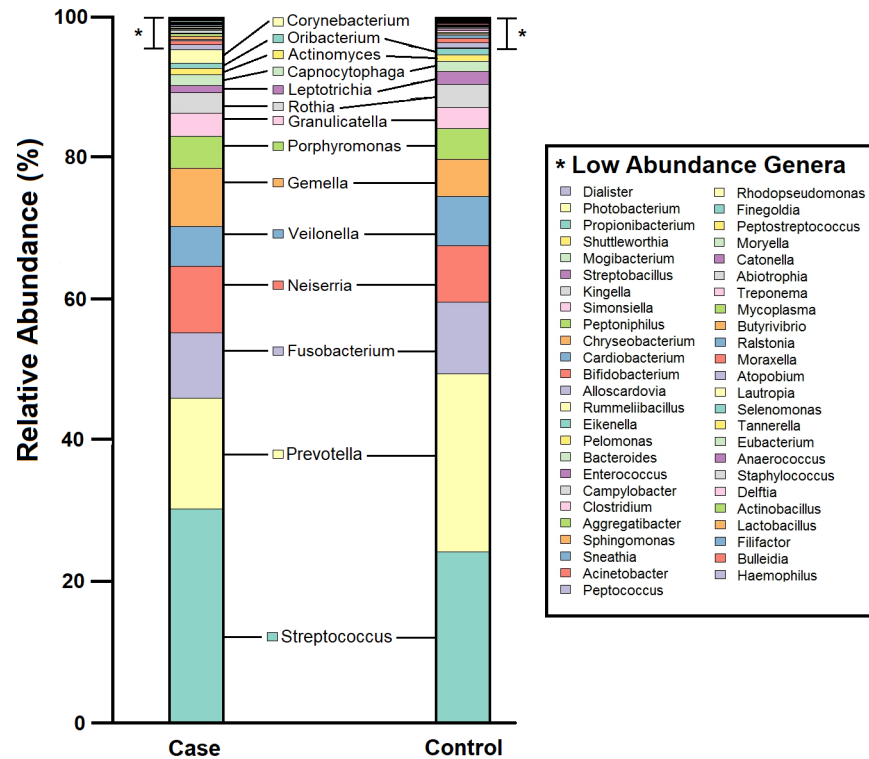


Fig 1. Pharyngeal microbiome profile of the study population, at genus level. Cumulative relative abundance of each of the 63 bacterial genera in case and control samples. Microbial community members with relative abundance <0.1% are listed as Low Abundance Genera (indicated by “*”).

<https://doi.org/10.1371/journal.pone.0201768.g001>

$p = 6.95 \times 10^{-5}$). In addition, *Gemella* was significantly more abundant in cases than in controls (6.0% vs 4.0%; *Adj-p* = 0.007). *Leptotrichia* which were present at low relative abundance (<1%), also differed significantly (*Adj-p* = 0.007) between cases and controls (Fig 3 and S4 Table). Analysing the >60 years subset revealed only *Prevotella*, *Leptotrichia* and *Streptococcus* to be significantly associated (*Streptococcus Adj-p* = 1.85×10^{-5} ; *Leptotrichia Adj-p* = 0.005; *Streptococcus Adj-p* = 0.035) among cases and controls (S5 Table).

Stratifying the AMD cases by disease into early and late stages, *Prevotella* and *Leptotrichia* relative abundance was significantly lower in late AMD samples than controls (Fig 4 and S6 Table). Conversely, late AMD samples were revealed to have greater relative abundances of *Streptococcus* and *Gemella* (*Streptococcus Adj-p* = 1.19×10^{-4} ; *Gemella Adj-p* = 0.028) (Fig 4). Comparing early AMD samples to controls did not reveal significant differences in genera relative abundances between both conditions. Further analysis to compare microbiomes between subgroups of late AMD disease (as described in Table 1) was under powered by reduced numbers of individuals, but remains an area of future interest.

Sampling rarely yields individual microbes in isolation; instead, microbes exist in communities together with other bacterial genera. To understand bacterial communities associated with *Prevotella* and *Gemella*, we identified subjects with high relative abundance of each genus, defined as relative abundance greater than two standard deviations above the population mean. Comparing the microbiota of these high abundance groups with the rest of the study population revealed distinct subsets of bacterial genera that are associated with each genus (Tables 2 and 3).

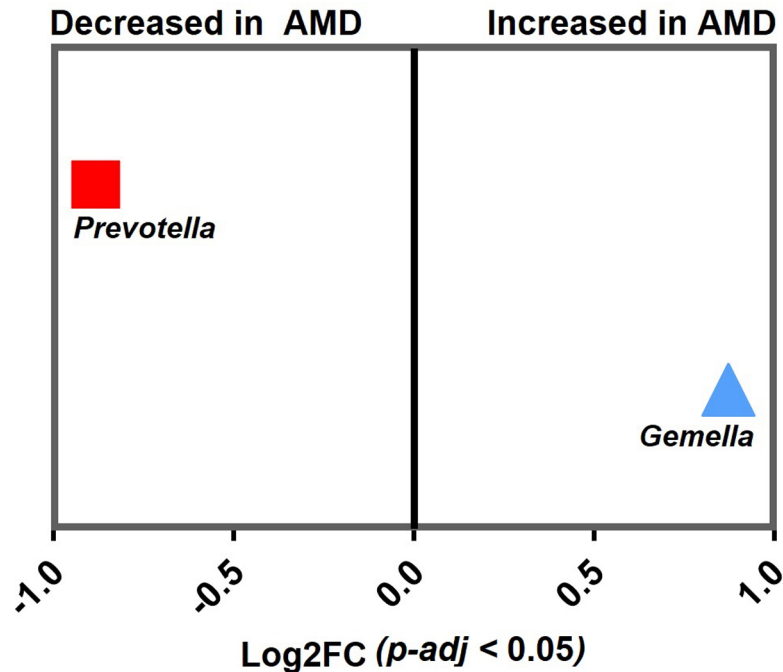


Fig 2. DESeq2 differential abundance analysis expressed as Log₂FC comparison of AMD-positive samples and control samples. Negative fold change scores (log₂) indicate genera with decreased abundance in AMD-positive samples, and positive fold change scores indicate genera with increased abundance in AMD-positive samples. Each point represents a genus. Genera detected to have significant difference in abundance (*Adj-p* < 0.05) are shown.

<https://doi.org/10.1371/journal.pone.0201768.g002>

High *Prevotella* abundance was uniquely associated with reduced *Gemella*, *Granulicatella*, *Porphyromonas* and *Streptococcus* abundance, and high *Gemella* abundance was uniquely associated with higher *Filifactor*, *Haemophilus* and *Aggregatibacter* abundances. Using PCA to map

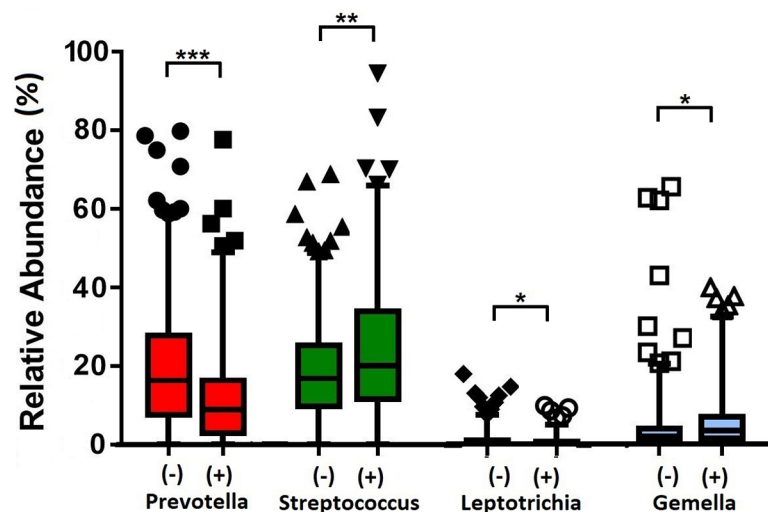


Fig 3. Relative abundances of significant genera between 245 case and 386 control samples. AMD-positive and control samples are denoted by (+) and (-) respectively. Statistical significance is indicated by (**Adj-p* < 0.05), (***Adj-p* < 0.005) or (***)*Adj-p* < 0.0005). Mean relative abundances, standard deviations and P-values are presented in S4 Table.

<https://doi.org/10.1371/journal.pone.0201768.g003>

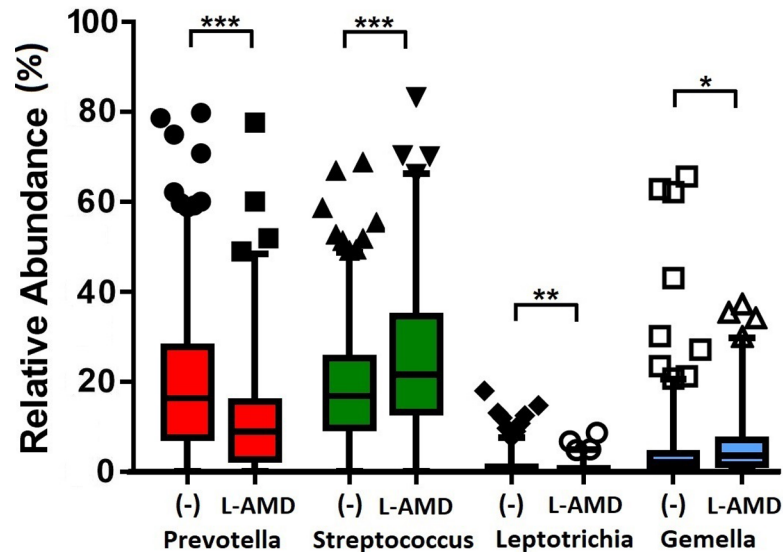


Fig 4. Relative abundances of significant genera between 386 controls and 165 individuals with late AMD. (-) and L-AMD denote controls and late AMD samples respectively. Statistical significance is indicated by (**Adj-p* < 0.05), (***Adj-p* < 0.005) or (***)*Adj-p* < 0.0005). Mean relative abundances, standard deviations and P-values are presented in S6 Table.

<https://doi.org/10.1371/journal.pone.0201768.g004>

samples harbouring either high *Prevotella* or high *Gemella* abundances, a distinct cluster of 27 samples was observed with high *Prevotella* abundance, of which six were case samples (S6 Fig).

Quantification-PCR (qPCR) validation of selected genera

To test the ability of our 16S sequencing to detect differences in abundance, we utilised qPCR assays as a validation platform to quantify microbial loads of three highly prevalent genera exhibiting narrow difference in mean abundance among controls and cases. Pearson’s correlation showed both techniques were positively correlated to some extent ($R = 0.322$). Microbial loads reported by qPCR demonstrated a similar trend as relative abundance by sequencing for case/control status. Using qPCR, *Streptococcus* was significantly enriched in case samples and *Prevotella* in controls. *Gemella* was increased in case samples, however, this difference was not significant by qPCR (Fig 5). While we found that 16S sequencing and qPCR are consistent in capturing overall microbial loads, the choice of different technology will result in different outputs and data formats. In addition to abundance estimates, 16S sequencing is able to identify members of a microbial community and their respective quantities in relation to other members.

Table 2. Microbial genera significantly associated with high *Prevotella* relative abundance.

	Mean relative abundance ± Standard Deviation		Adj. P-value
	Rest (n = 604)	High <i>Prevotella</i> (n = 27)	
<i>Streptococcus</i>	0.204 ± 0.144	0.068 ± 0.069	6.11 × 10 ⁻⁵
<i>Porphyromonas</i>	0.035 ± 0.045	0.003 ± 0.006	0.005
<i>Granulicatella</i>	0.025 ± 0.026	0.007 ± 0.01	0.008
<i>Gemella</i>	0.05 ± 0.073	0.005 ± 0.006	0.042

<https://doi.org/10.1371/journal.pone.0201768.t002>

Table 3. Microbial genera significantly associated with high *Gemella* relative abundance.

	Mean relative abundance ± Standard Deviation		Adj. P-value
	Rest (n = 610)	High <i>Gemella</i> (n = 21)	
<i>Aggregatibacter</i>	2.1 x 10 ⁻⁴ ± 7 x 10 ⁻⁴	0.001 ± 0.003	1.9 x 10 ⁻⁴
<i>Prevotella</i>	0.182 ± 0.156	0.042 ± 0.055	0.001
<i>Filifactor</i>	0.002 ± 0.009	0.016 ± 0.057	0.002
<i>Haemophilus</i>	0.005 ± 0.0122	0.016 ± 0.03	0.004
<i>Veillonella</i>	0.054 ± 0.053	0.01 ± 0.019	0.006

<https://doi.org/10.1371/journal.pone.0201768.t003>

Discussion

We have comprehensively and accurately described the human pharyngeal microbiome, at genus level in this a large case-control study of 631 AMD subjects, by shotgun sequencing the V3-V6 region of the microbial 16S rRNA gene. Consistent with previous studies [1,30–33], pharyngeal microbiomes in our study population were dominated by relatively few bacterial genera, of which *Streptococcus* and *Prevotella* were most abundant and prevalent. These "core" genera may play roles in maintaining the stability and complexity of individual microbiomes over time, properties thought to be associated with health [1].

Investigating community relative abundances using PCA revealed that the microbial drivers in the pharyngeal community are not vastly different between cases and controls. Rather, they suggest that the pharyngeal community is overall stable in terms of community genera members and their respective relative abundances across both disease and health conditions. Overall, it is interesting to note that pharyngeal community relative abundances in early AMD were more similar to healthy controls, while late AMD condition revealed significant dissimilarity in community relative abundance when compared to controls, which potentially suggests some dysbiosis related to disease progression. AMD is a progressive disease, and this observation suggests that alterations in the pharyngeal microbiota becomes more evident as the disease progresses. Current treatment of “wet” AMD is based on suppression of vascular endothelial growth factor (VEGF) with intraocular injections of anti-VEGF agents. While this has significantly improved the outcomes in the management of AMD, this treatment does not

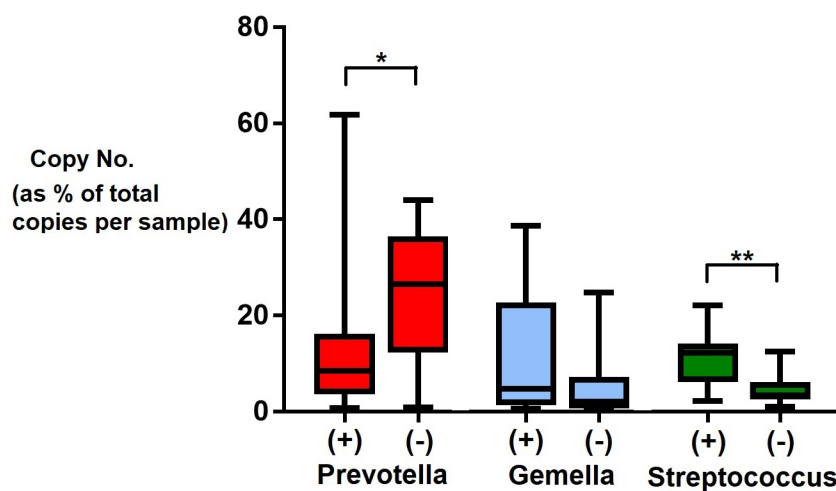


Fig 5. Microbial loads from qPCR. qPCR-derived copy numbers of each genera were expressed as a percentage of total microbial copy number within each sample. Twenty samples were randomly selected for each disease status. Statistical significance is indicated by (*P < 0.01) or (**P < 0.001) as determined by Mann-Whitney U test.

<https://doi.org/10.1371/journal.pone.0201768.g005>

reverse damage related to AMD, is invasive, expensive, and incurs significant healthcare resources. Thus, a search for the underlying cause of AMD with strategies to prevent its development and progression remains a key research priority.

Gemella and *Prevotella* were detected to have significant differential abundance between case and control conditions. Difference in mean relative abundance of both genera in disease and health conditions also reached statistical significance. This could either suggest that health/disease conditions could have an impact on community structure, or community abundances play a role towards disease propagation. Our findings provide preliminary insight into the composition of pharyngeal microbiota, as well as candidate genera with their abundance alterations in AMD health and disease conditions. Metagenomic sequencing and translational studies can be performed in future to provide more robust evidence to elucidate relevant biological factors leading to this observation.

We were also able to investigate potential roles for pharyngeal microbiome members in the initiation and pathogenesis of AMD. While overall diversity and composition of pharyngeal microbiomes from AMD cases and controls were relatively similar, we observed clear differences in specific bacterial genera between the two groups. *Gemella* and *Prevotella* (also the two most abundant genera), showed opposite trends, with *Gemella* being more abundant in cases and *Prevotella* more so in controls. *Gemella* is a dominant member of the laryngeal mucosal community and has been identified to be present at higher abundances in patients with laryngeal squamous cell carcinoma than healthy subjects [34, 35]. *Prevotella* species are not known to be protective; some such as *Prevotella intermedia* are known to cause periodontal disease [36], and *Prevotella copri* in the gut has also been associated with rheumatoid arthritis [37]. While there is a possibility that the protective effect we observed here for AMD is due to corresponding lower levels of *Gemella* in individuals with high *Prevotella* loads, nevertheless it begets the question of why some community members show differential abundance in health or disease conditions.

To better understand the composition of communities harbouring high *Prevotella* and *Gemella* loads, we looked for community members associated with high *Prevotella* and *Gemella*. Most associations were negative, i.e. high driver abundance correlated with reduced abundance of the genus in question. This could reflect the fact that the drivers were the most abundant genera (such that other genera are correspondingly low), or alternatively could indicate the existence of competitive mechanisms. Conversely, positive associations (for example, high *Gemella* was associated with high *Haemophilus*, *Filifactor* and *Aggregatibacter*) could indicate cooperative or symbiotic interactions. These patterns suggest that intra-community interactions are likely to affect the overall contribution of the individual microbiota to AMD pathogenesis. PCA clustering revealed high *Prevotella* samples ($n = 27$) to form an exclusive cluster, suggesting the high *Prevotella* subset to possess unique microbial profile against the rest of the samples. Among this subset were six case samples, indicating that while increased *Prevotella* abundance is associated with control status, it is not the sole scenario. It suggests greater complexity between association of the pharyngeal microbiota and AMD to be beyond the sheer abundance of a single genus.

Host genetic differences between cases and controls may also influence pharyngeal microbiome composition and colonization efficiency. A recent study revealed the rs3006458-T allele (associated with lower abundance of *Aerococcus* and *Micrococcaceae* in the nasopharynx) [38] was associated with increased PGLYRP4 expression in mucosal tissues supporting antimicrobial innate immunity [39]. These data suggest that host genotype regulates PGLYRP4 expression on the respiratory mucosa that leads to modulatory effect on bacterial abundance [38]. A number of pathogenic bacteria, including *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, and *Borrelia burgdorferi* encode surface proteins that bind to and recruit

CFH, thus mediating evasion of complement-mediated killing [40, 41]. Genetic variants in CFH associated with increased AMD risk have been shown to alter binding to *S. pyogenes* surface proteins, and consequently impact complement activation, opsonization, and phagocytosis [42,43]. In addition, individuals carrying the Tyr402His CFH risk variant who also harbored high levels of antibodies against *Chlamydia pneumoniae* showed an increased risk of AMD progression [6]. Our work provides further evidence that these factors play a role in AMD disease and progression.

Conclusions

Through 16S sequencing of an extensive population cohort, this study has provided accurate representation of the population as well as considerable confidence in a role for the pharyngeal microbiota profile in AMD health and disease conditions. Pharyngeal microbiota members and their relative abundances are highly stable among populations. Subtle differences were identified in relative abundances of *Prevotella*, *Gemella*, *Streptococcus* and *Leptotrichia* in the pharyngeal composition of AMD cases when compared to controls, a phenomenon similarly observed in comparisons between controls and patients with late-stage AMD. Prospective studies investigating pharyngeal microbiomes prior to disease, longitudinal studies tracking microbiomes over the course of disease progression, and analysis of genetic factors influencing microbiome composition will be necessary to fully characterize microbial triggers of AMD.

Supporting information

S1 Fig. Rarefaction curves indicating the number of genera detected with less than <500,000 reconstructed reads in 30 randomly picked samples. Samples are labelled with “PHT” prefix and corresponding sample number.
(DOCX)

S2 Fig. Visual inspection for batch effects using Principal Coordinate Analysis (PCoA) based on relative abundance at genus level. Each point represents an individual.
(DOCX)

S3 Fig. Guided Principal Component Analysis (gPCA) based on relative abundance on (A) Gender, (B) Disease status (C) Age and (D) Disease progression factors. Each point represents an individual. Case samples has a similar community composition similar to that of control samples. Additionally, microbial community composition is highly similar among early/late AMD status.
(DOCX)

S4 Fig. Prevalence (percentage of individuals in which the genus was detected) of each microbial genus in case and control samples.
(DOCX)

S5 Fig. Measures of pharyngeal microbiome community diversity. Shannon and Simpson diversity indices for pharyngeal microbiomes in AMD cases and controls. Each data point represents an individual pharyngeal microbiome; boxes indicate the mean and 25th and 75th percentiles; whiskers indicate interquartile ranges. Shannon and Simpson diversity indices did not differ significantly between cases and controls ($p < 0.05$, Mann-Whitney U test).
(DOCX)

S6 Fig. PCA plots colored by samples harbouring (A) high *Prevotella* (n = 27) and (B) high *Gemella* (n = 21) relative abundances. High *Prevotella* samples clustered at the tip of the plot

(A).
(DOCX)

S1 Table. Comparison of identified genera in 30 samples using 500,000 and <500,000 reconstructed reads.

(DOCX)

S2 Table. Quantitative-PCR (qPCR) primers and their corresponding amplicon sizes.

(DOCX)

S3 Table. Gender, age and disease-type Pairwise Statistical Comparisons (PERMANOVA) of microbial community abundance among case and control samples.

(DOCX)

S4 Table. Association between microbial genera and AMD status. Genera with significantly different relative abundances in case/control conditions are shown.

(DOCX)

S5 Table. Association between microbial genera and AMD status in individuals >60 years. Genera with significantly different relative abundances in case/control conditions are shown.

(DOCX)

S6 Table. Association between microbial genera with control and Late AMD status. Genera with significantly different relative abundances in control and Late AMD conditions are shown.

(DOCX)

Acknowledgments

We thank all patients and their families for their participation in this study, as well as hospital staff who assisted in patient enrolment and sample collection. We also thank the GERMS Microbial Genomics platform at the Genome Institute of Singapore for sequencing and bioinformatics expertise.

Author Contributions

Conceptualization: Tien Yin Wong, Ching Yu Cheng, Sonia Davila, Martin Hibberd.

Data curation: Eliza Xin Pei Ho, Ching Yu Cheng.

Formal analysis: Eliza Xin Pei Ho, Collins Wenhan Chu.

Funding acquisition: Tien Yin Wong, Martin Hibberd.

Investigation: Eliza Xin Pei Ho, Tien Yin Wong, Ching Yu Cheng, Martin Hibberd.

Methodology: Eliza Xin Pei Ho, Martin Hibberd.

Project administration: Eliza Xin Pei Ho, Chui Ming Gemmy Cheung, Ranjana Mathur, Doric Wong, Choi Mun Chan, Mayuri Bhagarva, Augustinus Laude, Tock Han Lim, Sonia Davila.

Resources: Chui Ming Gemmy Cheung, Andreas Wilm, Clarabelle Bitong Lin, Ranjana Mathur, Doric Wong, Choi Mun Chan, Mayuri Bhagarva, Augustinus Laude, Tock Han Lim, Tien Yin Wong, Sonia Davila.

Software: Andreas Wilm.

Supervision: Chui Ming Gemmy Cheung, Andreas Wilm, Sonia Davila.

Validation: Eliza Xin Pei Ho, Clarabelle Bitong Lin.

Writing – original draft: Eliza Xin Pei Ho, Shuzhen Sim.

Writing – review & editing: Shuzhen Sim, Collins Wenhan Chu, Andreas Wilm, Sonia Davila, Martin Hibberd.

References

1. Consortium THMP. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012 Jun 14; 486(7402):207–14. <https://doi.org/10.1038/nature11234> PMID: 22699609
2. Gao Z, Kang Y, Yu J, Ren L. Human Pharyngeal Microbiome May Play A Protective Role in Respiratory Tract Infections. *Genomics Proteomics Bioinformatics*. 2014 Jun; 12(3):144–50. <https://doi.org/10.1016/j.gpb.2014.06.001> PMID: 24953866
3. Freeley S, Kemper C, Le Friec G. The “ins and outs” of complement-driven immune responses. *Immunol Rev*. 2016 Nov 1; 274(1):16–32. <https://doi.org/10.1111/immr.12472> PMID: 27782335
4. Yazdankhah SP and Caugant DA. *Neisseria meningitidis*: an overview of the carriage state. *J Med Microbiol*. 2004 Sep; 53(Pt 9):821–32. <https://doi.org/10.1099/jmm.0.45529-0> PMID: 15314188
5. Meri T, Amdahl H, Lehtinen MJ, Hyvärinen S, McDowell JV, Bhattacharjee A, et al. Microbes Bind Complement Inhibitor Factor H via a Common Site. *PLOS Pathog*. 2013 Apr 18; 9(4):e1003308. <https://doi.org/10.1371/journal.ppat.1003308> PMID: 23637600
6. Baird PN, Robman LD, Richardson AJ, Dimitrov PN, Tikellis G, McCarty CA, et al. Gene-environment interaction in progression of AMD: the CFH gene, smoking and exposure to chronic infection. *Hum Mol Genet*. 2008 May 1; 17(9):1299–305. <https://doi.org/10.1093/hmg/ddn018> PMID: 18203751
7. Davila S, Wright VJ, Khor CC, Sim KS, Binder A, Breunis WB, et al. Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. *Nat Genet*. 2010 Sep; 42(9):772–6. <https://doi.org/10.1038/ng.640> PMID: 20694013
8. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science*. 2005 Apr 15; 308(5720):385–9. <https://doi.org/10.1126/science.1109557> PMID: 15761122
9. Edwards AO, Ritter R, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. *Science*. 2005 Apr 15; 308(5720):421–4. <https://doi.org/10.1126/science.1110189> PMID: 15761121
10. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng C-Y, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta analysis. *Lancet Glob Health*. 2014 Feb; 2(2):e106–116. [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1) PMID: 25104651
11. Ambati J, Atkinson JP, Gelfand BD. Immunology of age-related macular degeneration. *Nat Rev Immunol*. 2013 Jun; 13(6):438–51. <https://doi.org/10.1038/nri3459> PMID: 23702979
12. Laine M, Jarva H, Seitsonen S, Haapasalo K, Lehtinen MJ, Lindeman N, et al. Y402H Polymorphism of Complement Factor H Affects Binding Affinity to C-Reactive Protein. *J Immunol*. 2007 Mar 15; 178(6):3831–6. PMID: 17339482
13. Clark SJ, Perveen R, Hakobyan S, Morgan BP, Sim RB, Bishop PN, et al. Impaired Binding of the Age-related Macular Degeneration-associated Complement Factor H 402H Allotype to Bruch’s Membrane in Human Retina. *J Biol Chem*. 2010 Sep 24; 285(39):30192–202. <https://doi.org/10.1074/jbc.M110.103986> PMID: 20660596
14. Sjöberg AP, Trouw LA, Clark SJ, Sjölander J, Heinegård D, Sim RB, et al. The Factor H Variant Associated with Age-related Macular Degeneration (His-384) and the Non-disease-associated Form Bind Differentially to C-reactive Protein, Fibromodulin, DNA, and Necrotic Cells. *J Biol Chem*. 2007 Apr 13; 282(15):10894–900. <https://doi.org/10.1074/jbc.M610256200> PMID: 17293598
15. Tortajada A, Montes T, Martinez-Barricarte R, Morgan BP, Harris CL, de Córdoba SR. The disease protective complement factor H allotypic variant Ile62 shows increased binding affinity for C3b and enhanced cofactor activity. *Hum Mol Genet*. 2009 Sep 15; 18(18):3452–61. <https://doi.org/10.1093/hmg/ddp289> PMID: 19549636
16. Khor CC, Hibberd ML. Host–pathogen interactions revealed by human genome-wide surveys. *Trends Genet*. 2012 May; 28(5):233–43. <https://doi.org/10.1016/j.tig.2012.02.001> PMID: 22445588
17. Cheung CMG, Li X, Cheng C-Y, Zheng Y, Mitchell P, Wang JJ, et al. Prevalence, Racial Variations, and Risk Factors of Age-Related Macular Degeneration in Singaporean Chinese, Indians, and Malays.

- Ophthalmology. 2014 Aug; 121(8):1598–603. <https://doi.org/10.1016/j.ophtha.2014.02.004> PMID: 24661862
18. Laser photocoagulation of subfoveal recurrent neovascular lesions in age-related macular degeneration. Results of a randomized clinical trial. Macular Photocoagulation Study Group. Arch Ophthalmol Chic Ill 1960. 1991 Sep; 109(9):1232–41.
 19. Japanese Study Group of Polypoidal Choroidal Vasculopathy. [Criteria for diagnosis of polypoidal choroidal vasculopathy]. Nippon Ganka Gakkai Zasshi. 2005 Jul; 109(7):417–27. PMID: 16050460
 20. Lavanya R, Jegathan VSE, Zheng Y, Raju P, Cheung N, Tai ES, et al. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. Ophthalmic Epidemiol. 2009 Dec; 16(6):325–36. <https://doi.org/10.3109/09286580903144738> PMID: 19995197
 21. Hang J, Desai V, Zavaljevski N, Yang Y, Lin X, Satya RV, et al. 16S rRNA gene pyrosequencing of reference and clinical samples and investigation of the temperature stability of microbiome profiles. Microbiome. 2014 Sep 16; 2:31. <https://doi.org/10.1186/2049-2618-2-31> PMID: 25228989
 22. Ong SH, Kukkillaya VU, Wilm A, Lay C, Ho EXP, Low L, et al. Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences. PLoS ONE. 2013 Apr 8; 8(4):e60811. <https://doi.org/10.1371/journal.pone.0060811> PMID: 23579286
 23. The HC, Florez de Sessions P, Jie S, Pham Thanh D, Thompson CN, Nguyen Ngoc Minh C, et al. Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children. Gut Microbes. 2017 Aug 2;0.
 24. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol. 2011; 12(5):R44. <https://doi.org/10.1186/gb-2011-12-5-r44> PMID: 21595876
 25. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a Chimera Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol. 2006 Jul 1; 72(7):5069–72. <https://doi.org/10.1128/AEM.03006-05> PMID: 16820507
 26. Reese SE, Archer KJ, Therneau TM, Atkinson EJ, Vachon CM, de Andrade M, et al. A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. Bioinforma Oxf Engl. 2013 Nov 15; 29(22):2877–83.
 27. Chua MC, Ben-Amor K, Lay C, Neo AGE, Chiang WC, Rao R, et al. Effect of Synbiotic on the Gut Microbiota of Caesarean Delivered Infants: A Randomized, Double-Blind, Multicenter Study. J Pediatr Gastroenterol Nutr. 2017 May 3;
 28. Picard FJ, Ke D, Boudreau DK, Boissinot M, Huletsky A, Richard D, et al. Use of tuf sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. J Clin Microbiol. 2004 Aug; 42(8):3686–95. <https://doi.org/10.1128/JCM.42.8.3686-3695.2004> PMID: 15297518
 29. Matsuki T, Watanabe K, Fujimoto J, Takada T, Tanaka R. Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces. Appl Environ Microbiol. 2004 Dec; 70(12):7220–8. <https://doi.org/10.1128/AEM.70.12.7220-7228.2004> PMID: 15574920
 30. Boutin S, Graeber SY, Weitnauer M, Panitz J, Stahl M, Clausnitzer D, et al. Comparison of Microbiomes from Different Niches of Upper and Lower Airways in Children and Adolescents with Cystic Fibrosis. PLoS ONE. 2015 Jan 28; 10(1):e0116029. <https://doi.org/10.1371/journal.pone.0116029> PMID: 25629612
 31. Bogaert D, Keijsers B, Huse S, Rossen J, Veenhoven R, van Gils E, et al. Variability and Diversity of Nasopharyngeal Microbiota in Children: A Metagenomic Analysis. PLoS ONE. 2011 Feb 28; 6(2): e17035. <https://doi.org/10.1371/journal.pone.0017035> PMID: 21386965
 32. Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, et al. Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE. 2010 Dec 20; 5(12):e15216. <https://doi.org/10.1371/journal.pone.0015216> PMID: 21188149
 33. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the Respiratory Microbiome in Healthy Nonsmokers and Smokers. Am J Respir Crit Care Med. 2013 May 15; 187(10):1067–75. <https://doi.org/10.1164/rccm.201210-1913OC> PMID: 23491408
 34. Gong H, Shi Y, Zhou X, Wu C, Cao P, Xu C, et al. Microbiota in the Throat and Risk Factors for Laryngeal Carcinoma. Appl Environ Microbiol. 2014 Dec; 80(23):7356–63. <https://doi.org/10.1128/AEM.02329-14> PMID: 25239901
 35. Gong H-L, Shi Y, Zhou L, Wu C-P, Cao P-Y, Tao L, et al. The Composition of Microbiome in Larynx and the Throat Biodiversity between Laryngeal Squamous Cell Carcinoma Patients and Control Population. PLoS ONE [Internet]. 2013 Jun 18 [cited 2015 Jun 15]; 8(6).

36. Potempa M, Potempa J, Kantyka T, Nguyen K-A, Wawrzonek K, Manandhar SP, et al. Interpain A, a Cysteine Proteinase from *Prevotella intermedia*, Inhibits Complement by Degrading Complement Factor C3. *PLoS Pathog.* 2009 Feb 27; 5(2):e1000316. <https://doi.org/10.1371/journal.ppat.1000316> PMID: 19247445
37. Scher JU, Szczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, et al. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife.* 2013 Nov 5; 2:e01202. <https://doi.org/10.7554/eLife.01202> PMID: 24192039
38. Igartua C, Davenport ER, Gilad Y, Nicolae DL, Pinto J, Ober C. Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome.* 2017 Feb 1; 5:16. <https://doi.org/10.1186/s40168-016-0227-5> PMID: 28143570
39. Consortium TGt. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* 2015 May 8; 348(6235):648–60. <https://doi.org/10.1126/science.1262110> PMID: 25954001
40. Schneider MC, Prosser BE, Caesar JJE, Kugelberg E, Li S, Zhang Q, et al. *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature.* 2009 Apr 16; 458(7240):890–3. <https://doi.org/10.1038/nature07769> PMID: 19225461
41. Józsi M, Zipfel PF. Factor H family proteins and human diseases. *Trends Immunol.* 2008 Aug; 29(8):380–7. <https://doi.org/10.1016/j.it.2008.04.008> PMID: 18602340
42. Haapasalo K, Vuopio J, Syrjänen J, Suvilehto J, Massinen S, Karppele M, et al. Acquisition of Complement Factor H Is Important for Pathogenesis of *Streptococcus pyogenes* Infections: Evidence from Bacterial In Vitro Survival and Human Genetic Association. *J Immunol.* 2012 Jan 1; 188(1):426–35. <https://doi.org/10.4049/jimmunol.1102545> PMID: 22140259
43. Haapasalo K, Jarva H, Siljander T, Tewodros W, Vuopio-Varkila J, Jokiranta TS. Complement factor H allotype 402H is associated with increased C3b opsonization and phagocytosis of *Streptococcus pyogenes*. *Mol Microbiol.* 2008 Nov; 70(3):583–94. <https://doi.org/10.1111/j.1365-2958.2008.06347.x> PMID: 18627465