GENOMIC ATTRIBUTES OF AIRWAY COMMENSAL BACTERIA AND MUCOSA
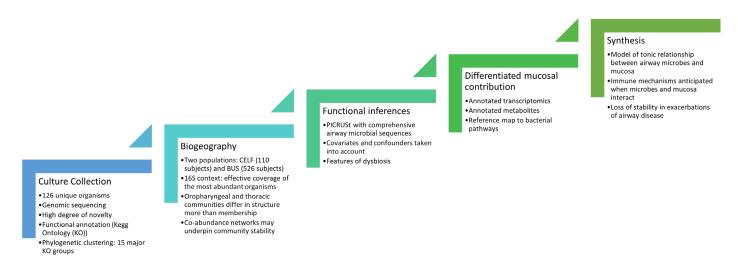
Supplementary Figures

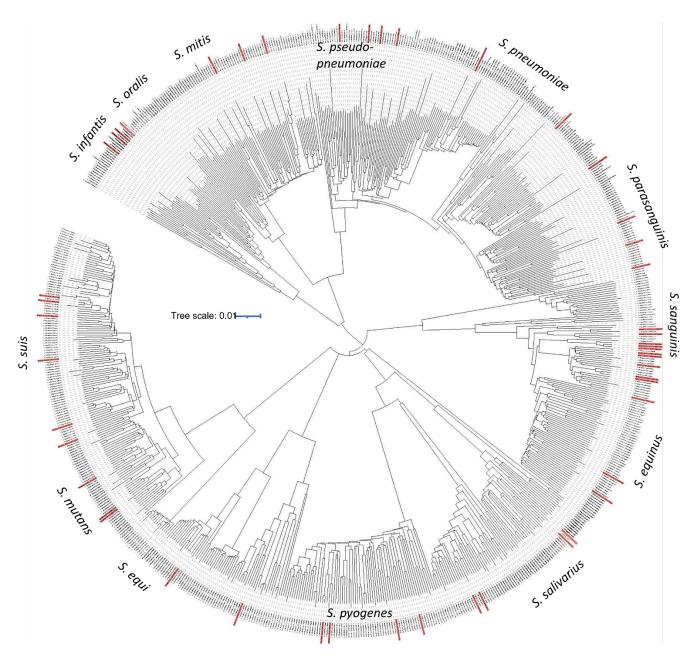## CONTENTS

SUPPLEMENTARY FIGURE 1. STUDY DESIGN.

Culture Collection

- 126 unique organisms
- Genomic sequencing
- High degree of novelty
- Functional annotation (Kegg Ontology (KO))
- Phylogenetic clustering: 15 major KO groups

Biogeography

- Two populations: CELF (110 subjects) and BUS (526 subjects)
- 16S context: effective coverage of the most abundant organisms
- Oropharyngeal and thoracic communities differ in structure more than membership
- Co-abundance networks may underpin community stability

Functional inferences

- PICRUSt with comprehensive airway microbial sequences
- Covariates and confounders taken into account
- Features of dysbiosis

Differentiated mucosal contribution

- Annotated transcriptomics
- Annotated metabolites
- Reference map to bacterial pathways

Synthesis

- Model of tonic relationship between airway microbes and mucosa
- Immune mechanisms anticipated when microbes and mucosa interact
- Loss of stability in exacerbations of airway disease

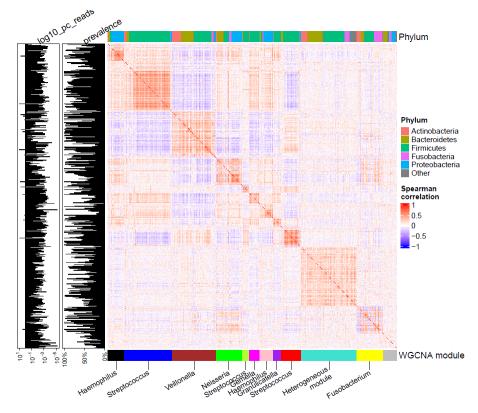The figure reads from left to right, ascending.

Comparison of the full sequences of airway streptococcal isolates from this study (shown in red) in a pan-genome analysis of 2477 public *Streptococcus* spp. genomes. Airway isolates are each distinct and are widely distributed within the pan-genomes Areas within the tree are labelled according to their most common members, although designations of *S. unknown* are found throughout the tree. A high-resolution image will be available online.

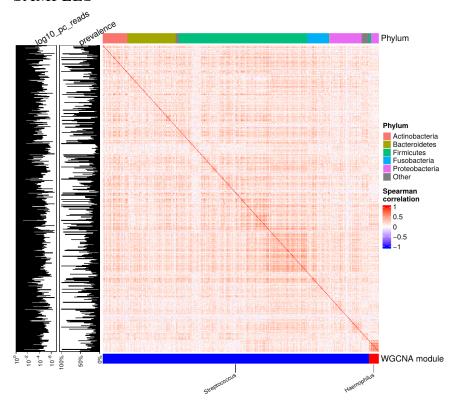# SUPPLEMENTARY FIGURE 3. CORRELATION STRUCTURE OF BUSSELTON PTOP SAMPLES

Spearman correlation for the Busselton ptOP samples. The 646 OTUs included in the WGCNA analysis are shown with their Phylum (top colour bar) and WGCNA module (bottom colour bar).

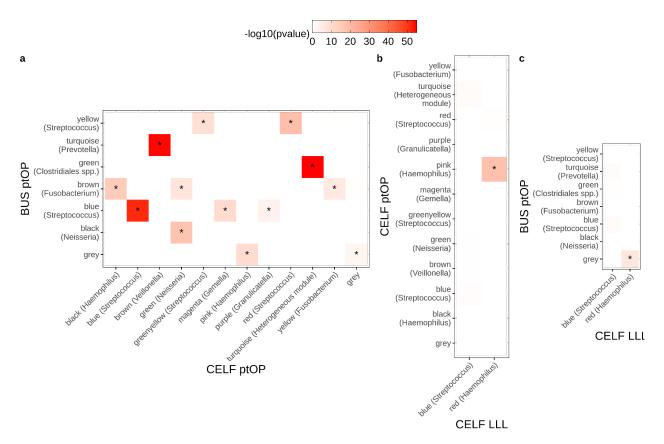# SUPPLEMENTARY FIGURE 4. CORRELATION STRUCTURE OF CELF ptOP SAMPLES

Spearman correlation for the Celtic Fire ptOP samples. The 646 OTUs included in the WGCNA analysis are shown with their Phylum (top colour bar) and WGCNA module (bottom colour bar).

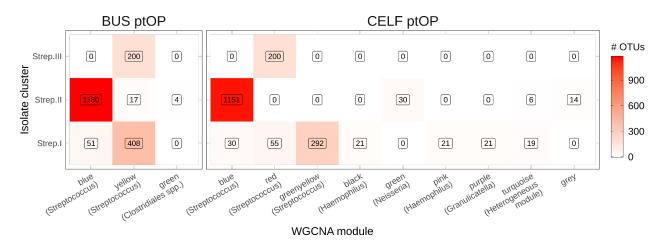# SUPPLEMENTARY FIGURE 5. CORRELATION STRUCTURE OF CELF LLL SAMPLES

Spearman correlation for the CELF LLL samples. The 646 OTUs included in the WGCNA analysis are shown with their Phylum (top colour bar) and WGCNA module (bottom colour bar). In general, correlations are weaker than in the BUS ptOP (3a above) and CELF ptOP (Figure 3b above) samples.
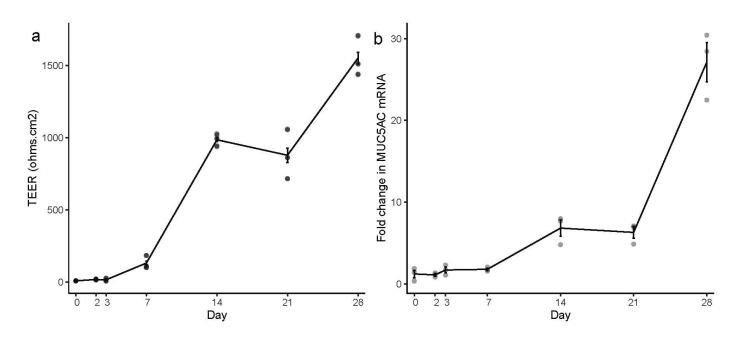
WGCNA network structure is more conserved between the two sets of ptOP samples from the different hemispheres (panel a) than within the same study between the lower airway (panel b) or between the BUS ptOP samples and CELF LLL samples (panel c). Every module in the CELF ptOP network is observed in the BUS ptOP network (*P*<0.05 denoted using an asterisk). In contrast, only one of the CELF ptOP modules is conserved in the CELF LLL network. Overlap between module assignments tested using Fisher's exact test and corrected using false discovery rate.
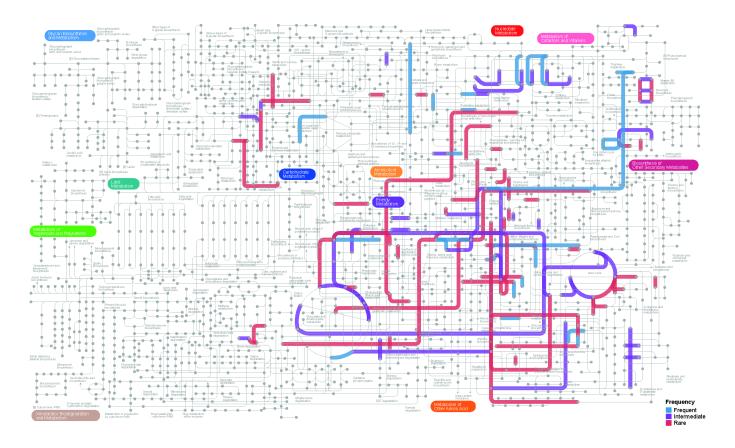
The co-abundance patterns of *Streptococcus* in the ptOP samples mirrors the phylogenetic similarity according to the KO analysis. Plotted is the number of *Streptococcus* OTUs with at least 99% 16S rRNA gene sequence similarity to an isolate, split into the three phylogenetic clusters. In the BUS ptOP samples the blue WGCNA module contains OTUs with higher similarity to isolates in the Strep. II cluster, while the yellow WGCNA module contains OTUs with higher similarity to isolates in Strep. I and Strep. III. In the CELF ptOP samples the blue WGCNA module corresponds to Strep. II, red to Strep. III and greenyellow to Strep. I. In the CELF ptOP network small numbers of *Streptococcus* OTUs are found in other modules.

Supplementary Figure 8. Trans-epithelial resistance and mucin production during airway cell differentiation in an ALI model
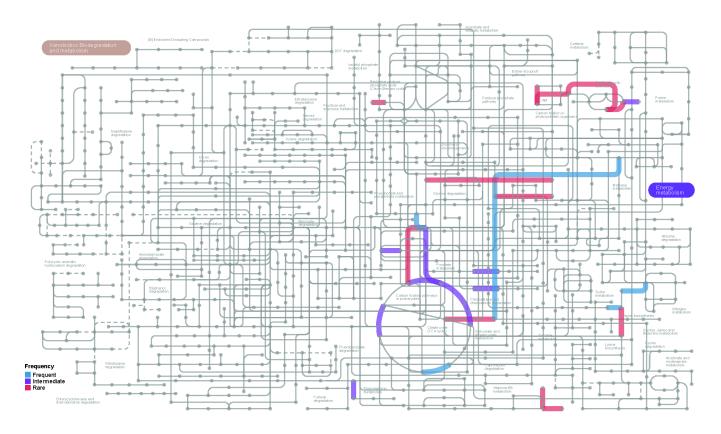

Human Airway Epithelial Cell (HAEC) from a single donor were grown in triplicate and harvested on days 0, 2, 3, 7, 14, 21 and 28. Significant increases were observed in a) trans-epithelial resistance (TEER) b) MUC5AC mucin mRNA production over the time course, indicating full HAEC differentiation. Error bars indicate standard errors of the mean.

Isolate genes were mapped to human KEGG pathways with the frequency of the genes indicated: 'frequent' for genes in >75% of isolates (blue), 'intermediate' for genes in 25-75% of isolates (purple) and 'rare' for those in <25% of isolates (red). The Plot was produced using iPath 3.0.

Isolate genes were mapped to microbial KEGG pathways with the frequency of the genes indicated: 'frequent' for genes in >75% of isolates (blue), 'intermediate' for genes in 25-75% of isolates (purple) and 'rare' for those in <25% of isolates (red). The Plot was produced using iPath 3.0.