Supplementary Information

Topographical continuity of bacterial populations in the healthy human respiratory tract

Emily S. Charlson[1,2], Kyle Bittinger[1], Andrew R. Haas[2], Ayannah S. Fitzgerald[2], Ian Frank[3], Anjana Yadav[2], Frederic D. Bushman[1*] and Ronald G. Collman[1,2*]

Table E1. Subject characteristics.

Table E2: OTU table.

Table E3. Bacterial lineages only detected in lung samples.

Table E4. Sequence counts and alpha diversity measures for each sample.

Figure E1. Community clustering demonstrates that oral, glottal, and lung samples share the same structure. The weighted UniFrac distances were calculated between all pairs of samples, then plotted in the first three dimensions of a coordinate space determined by PCoA. Each point corresponds to a community colored by the general category of sample type: Nasopharyngeal, purple; Oropharyngeal, Oral Wash, Scope 1 Tip and Scope 1 Post-wash, yellow; BAL and PSB, blue; and Control, green. The percentage of variation explained by each PCoA is indicated on the axes.

Figure E2. Within vs. Between group average UniFrac distance comparisons by sample type. (A) Schematic of within vs. between group weighted UniFrac distance comparisons for 6 hypothetical communities, grouped and colored based on membership to one of two undefined sample types. On the right, the average UniFrac distance among red communities is calculated and then compared to the average distance between red and blue communities. This analysis determines whether red communities are more similar to each other than to blue communities. The reciprocal comparison is demonstrated in the diagram on the left where the average distance among blue communities is calculated and then compared to the average distance between blue and red communities. (B) In each row of the table, the average weighted UniFrac distance is computed for all sample pairs within the indicated sample type. This value is compared to the average distance to samples of the type appearing columnwise, using Student's t statistic. P-values are generated by comparison of the observed test statistic value to a distribution of test statistic value generated from 1,000 Monte-Carlo simulations and are shown in the table. Significance threshold: $p < 0.05$. A significant result indicates that the community structures of the within group sample type are more similar to themselves than to the between group sample type (columnwise).

Figure E3. OTU network analysis of peri-glottic Scope #1 Tip and lung PSB samples from 6 subjects. The network consists of nodes representing PSB and Scope #1 Tip samples (colored by Subject ID, keyed to the right) and nodes representing OTU's (colored gray), which are connected wherever an OTU is present in a sample. OTU nodes connected to only one sample were removed from the network, OTU nodes connected to exactly 2 samples are colored dark gray, and OTU nodes connected to > 2 samples are colored light gray. The nodes were arranged with Cytoscape (38) according to an Edge-Weighted Spring Embedded layout, where the edge weights are taken to be OTU counts per sample. The number of OTU nodes connected to Scope #1 Tip and PSB sample nodes from the same subject is significantly greater than the number of OTU nodes connected to Scope #1 Tip and PSB sample nodes from different subjects ($p=0.0014$, 720 permutations). This indicates that organisms found in each individual's lung are significantly coupled to their own peri-glottic microbiota.