Extended Data: Nasopharyngeal Dysbiosis Precedes the Development of Lower Respiratory Tract Infections in Young Infants, a Longitudinal Infant Cohort Study

Figure 1 – Healthy Infants.

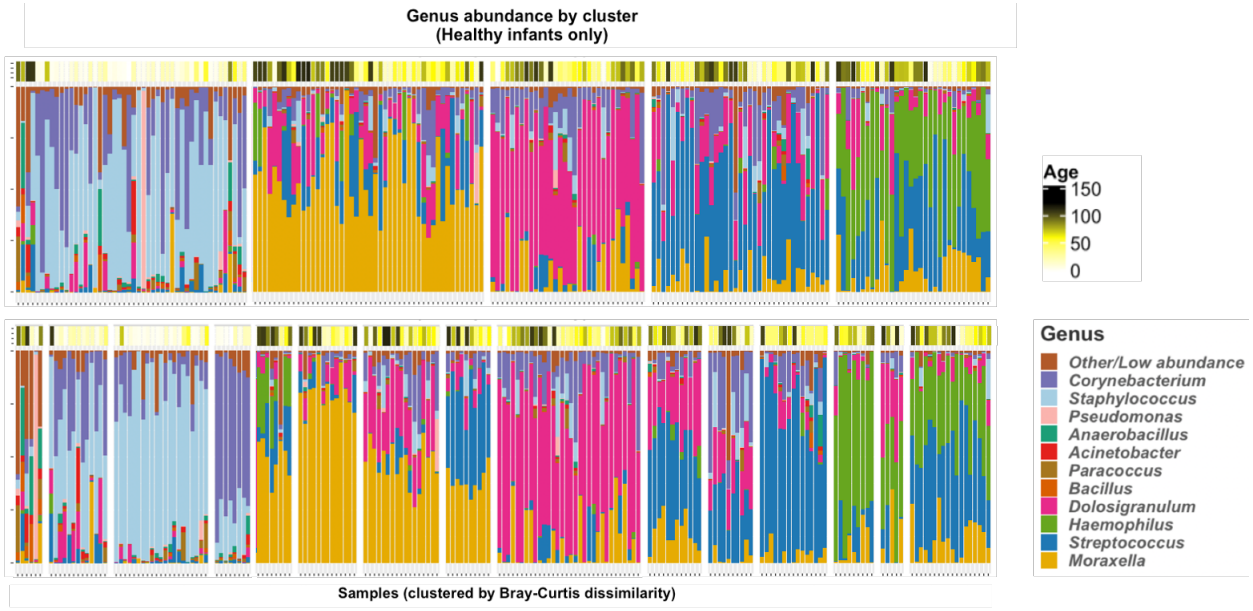
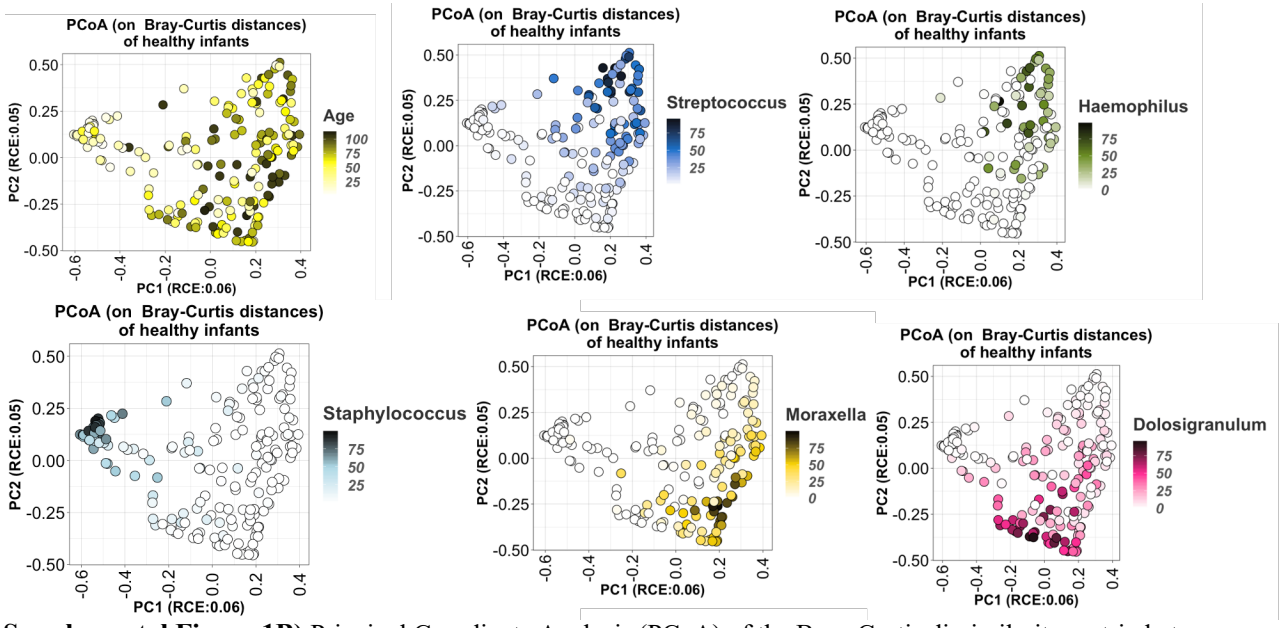


Figure 1A) The taxonomic profiles of healthy infants' NP samples (n=30), clustered by pairwise Bray-Curtis dissimilarity. Clusters were defined by performing hierarchical clustering on the beta diversity matrix and then cutting the resulting dendrogram into an optimal number of clusters according to the Silhouette index (5) and Frey indx (15). The color bars above the stacked bar plots indicate the infants' ages at the time of each sample.



Supplemental Figure 1B) Principal Coordinate Analysis (PCoA) of the Bray-Curtis dissimilarity matrix between healthy infants' samples, colored by age and relative abundance of the dominant genera.

Figure 2 – LRTI infants only.

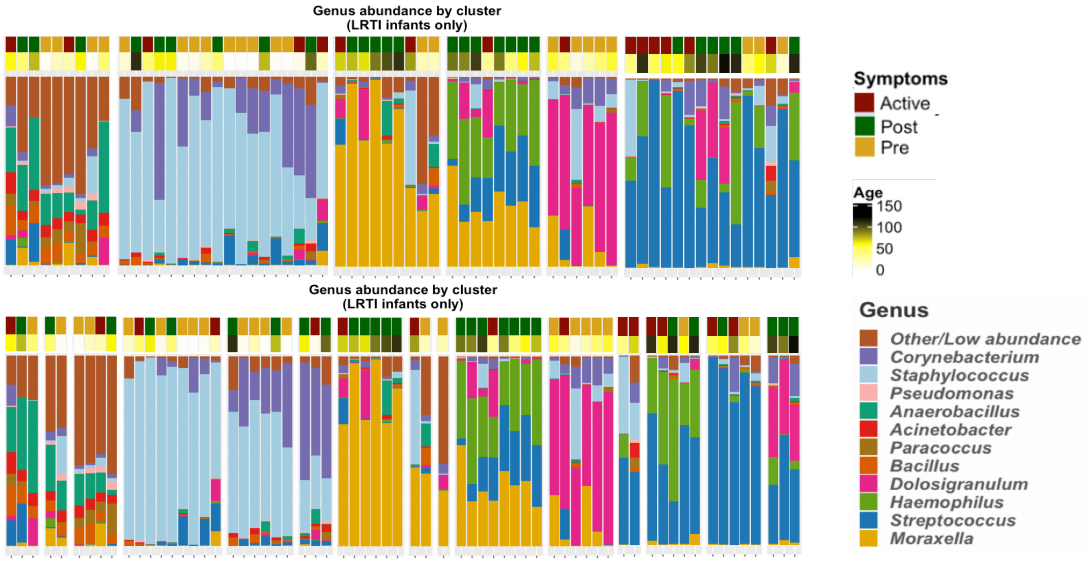


Figure 2: The taxonomic profiles of LRTI infants' NP samples (n=10), clustered by pairwise Bray-Curtis dissimilarity. Clusters were defined by performing hierarchical clustering on the beta diversity matrix and then cutting the resulting dendrogram into an optimal number of clusters according to the Silhouette index (6) and Frey index (15). The color bars above the stacked bar plots indicate the infants' ages and symptom status at the time of each sample.

Figure 3

Figure 3: Each facet represents the relative abundance of bacterial genera in one infant's nasopharynx over time. Samples are ordered chronologically from left to right, with labels on the x-axis displaying the infant's age at the time of the NP swab. Time points where infants were experiencing severe respiratory symptoms are marked with a *. Time points where infants were experiencing mild symptoms which did not qualify as LRTI are marked with a +.

Note that the order of these infants (reading rows left to right, top row first) is the same as the order in Table 2 (Clinical symptoms)

Figure 4

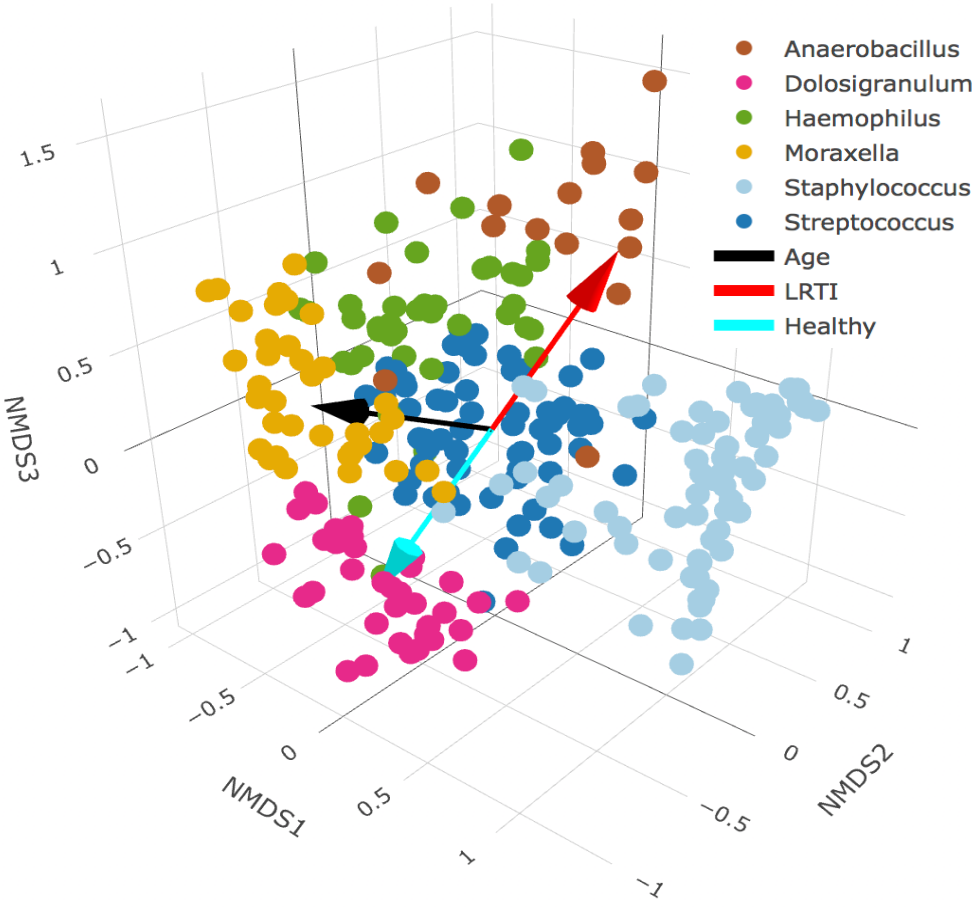


Figure 4: A 3-D scatterplot of infant nasopharyngeal samples projected into nonmetric multidimensional scaling (NMDS) ordination space. We applied 3-dimensional NMDS ordination to the Bray-Curtis dissimilarity matrix between all infants' NP swabs, and projected vectors into that ordination space representing the best fit correlations for the age at sampling (the black vector) and LRTI status (the cyan vector represent control infants, the red vector represent LRTI infants). Samples are colored by their primary taxonomic profile cluster membership (see Figure 3 for details). Age is highly correlated with the x-axis, and samples on the young end of the age vector mostly belong to the *Staphylococcus*-dominated profile, whereas samples on the older end tend to belong more to the *Haemophilus* and *Moraxella*-dominated profiles. The *Dolosigranulum*-dominated profile is associated with the healthy end of the vector for LRTI status, while the *Anaerbacillus*-dominated profile is associated with disease.

Figure 5

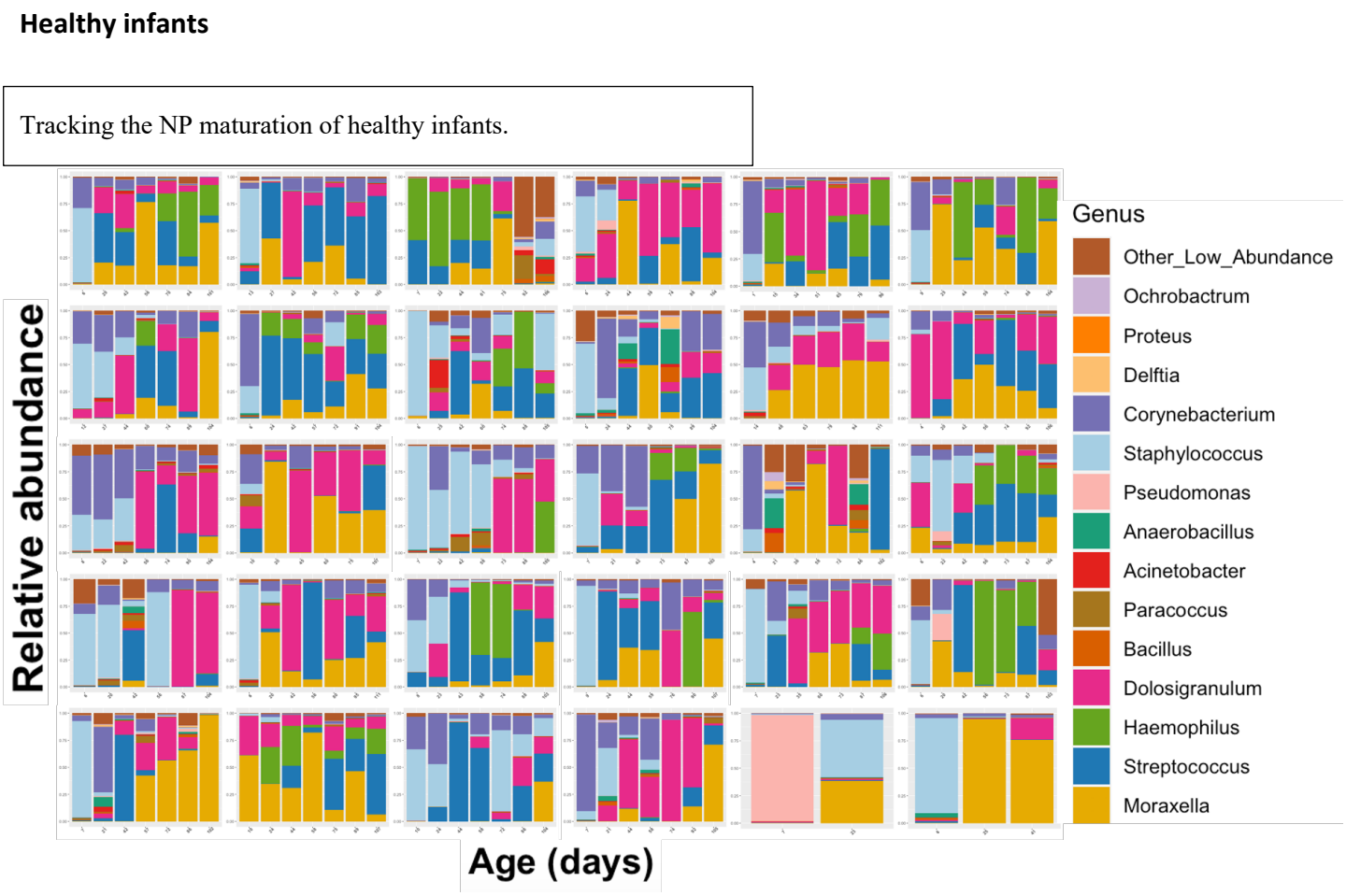


Figure 5: Tracking the NP maturation of healthy infants. Each facet represents the relative abundance of bacterial genera in one infant's nasopharynx over time. Samples are ordered chronologically from left to right, with labels on the x-axis displaying the infant's age at the time of the NP swab.