**Transmission of Oral microbiome and Sequencing (TOS) Project**

**Readme**

**Datasets**

1. **study\_data.csv**: Data on nose/throat samples collected from study participants including index cases and household contact. Rows represent samples collected at time of study enrollment and at the last day of follow up.

Data dictionary for study\_data.csv

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Possible Values** |
| ParticipantID | Unique participant ID | P001 to P717 |
| SampleID | Unique sample ID | T0001 to T1422 |
| Baseline\_Status | Indicates whether participant is 1) an index case or household contact and 2) positive for influenza at time of enrollment | index\_case = index case;  contact\_negative = household contact negative for influenza at enrollment;  contact\_positive = household contact positive for influenza at enrollment |
| Sample\_Time | Indicates whether sample was from enrollment or last day of follow up | Baseline;  Followup |
| Influenza\_Status | Influenza status by RT-PCR | 0 = negative;  1 = positive |
| Influenza\_Subtype | Influenza subtype | No\_Flu = negative for influenza;  H1N1;  H3N2;  B |
| Community\_Type\* | Bacterial community type | 1 to 5 |

\*Dirichlet multinomial mixture modeling, an unsupervised clustering method, was used to assign

nose/throat samples to 5 bacterial community types. Ninety-eight percent of all sequenced samples

were assigned to a community type, after applying a ≥80% posterior probability threshold.

Reference: Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for 529 microbial metagenomics. PloS One. 2012;7: e30126. doi:10.1371/journal.pone.0030126

For a full description of community definitions and methodology, see the associated article and figures.

1. **oligotype\_count.csv**: oligotype count table. For a full description of the oligotyping process, see the associated article and references.
2. **tax\_data.csv:** taxonomic classification of oligotypes. Taxonomic classification obtained by searching representative sequences of each oligotype against the Human Oral Microbiome Database (HOMD) v14.51 (http://www.homd.org/) using blastn v2.2.23 (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For a full description of the taxonomy assignment process, see the associated article and references.

**Data citation**: Lee KH, Gordon A, Foxman B. (2018). Transmission of Oral microbiome and Sequencing - Influenza Susceptibility [Data set]. University of Michigan Deep Blue Data Repository. <https://doi.org/10.7302/Z2736P4B>

**Citation to related material:** Lee KH, Gordon A, Shedden K, Kuan G, Ng S, Balmaseda A, Foxman B. The respiratory microbiome and susceptibility to influenza virus infection. bioRxiv. 2018; online preprint. https://doi.org/10.1101/372649