Microbiome Diversity on Materials

Chandrima Bhattacharya*, Pinaki Chakraborty, Rohit Pandey and Malay Bhattacharyya
Indian Institute of Engineering Science and Technology, Shibpur
Howrah - 711103, West Bengal, India
E-mail: malaybhattacharyya@it.iiests.ac.in
*** The presenting author**

Abstract

The study of microbiome is promising in understanding the higher-level organisms to a broader extent. In this paper, we consider to analyze the microbiome diversity across different materials. We particularly focus on the metagenomics data from the MetaSUB International Consortium for the said purpose. With an emphasis on the stationary materials and fluids available in subways and its vicinity, we demonstrate how diverse the microbiome community might appear in multiple cities.

1 INTRODUCTION

Microbiome analysis has recently drawn attention of the worldwide researchers due to its broader scope of understanding higher-level organisms. The microbiome refers to the microorganisms dwelling in a particular environment. Studying the microbiome data, which is increasingly becoming open for analysis, provides us a better picture of the system level activities surrounding an organism. However, it is also interesting to understand the interaction and activities within this microbiome community. The study of diversity in microbiome across the different localities has rarely been studied until recently. The current paper aims to address this issue with the support from MetaSUB International Consortium data.

The study of microbiome diversity has been carried out in a limited previous attempts. Some of them are based on specific organisms and some are phylogenetic attempts [1]. Given the diverse landscape of studies initiated with microbiome data, there is still the scarcity of material-specific analysis of microbiome ecology based on the environments they belong to. Our principle motivation in this paper is to analyze the diversity of microbiome habitat on different materials. We plan to understand the miscellany of different microorganisms in different materials and collected from different sources. The MetaSUB datasets have recently been employed for the purpose of characterizing microbiome community available in the locations like subways [2, 3]. To the best of our knowledge, the study on materials are however new that we are attempting in the current paper.

2 DATASET DETAILS

We have basically taken the data of MetaSUB International Consortium for the current analysis with the aforementioned motivation. The MetaSUB International Consortium aims to create the world's only

longitudinal metagenomic map of mass-transit systems and other public spaces across the globe [4]. An early release of the multi-city analysis data from MetaSUB consists of the microbiome details collected from the three major cities New York City (New York, USA), Boston (Massachusetts, USA), and Sacramento (California, USA), although having data from a total 40 locations in the entire dataset to be published soon. We took a subset of this dataset for the current analysis. A statistical overview of this subset is shown in Table 1.

Table 1. Details about the subset of MetaSUB dataset that we analyzed. The bracketed values denote the number of samples available for each case.

	Boston	**New York City**
Data Size	~ 32 GB	~ 1 TB
Number of Samples	141	1572
Sources of Collection	Station touchscreen (36), Train Car (105)	Subway (1536), Park (15), Ferry (2), Airport (1), Canal (14), Fish Gut (1)
Materials Collected From	Polyester (33), PVC (28), Steel (32), Glass (34), Vinyl (7), Plastic (6)	Wood (355), Metal (788), Plastic (34), Metal/Plastic (329), Concrete (9), Water (17), Pool of samples (1), Animal (1), Culture (2), Stone (1)

Each file in the Boston dataset consists of data containing 96 sequences from train cars and subway stations across the Boston subway system: 72 sequences consist of the V4 amplicons of the 16S rRNA genes and 24 sequences are shotgun metagenomic sequences. A survey of the New York City subway system to study the microbiome of a large, dynamic, metropolitan area. Most data from Boston includes urban metagenome while those from New York City are terrestrial metagenome.

The data is provided with FASTQ sequences of the microbiomes which are compressed by DSRC or DNA Sequence Reads Compression format. The platform used for sequencing is ILUMINA for data collected in Boston and New York City. The smallest decompressed data is in the size of few hundred KB while the largest is more than 5GB. Due to the limitation of computational infrastructure, we selected a subset of the data, where the file size of each sample is no more than 1GB in compressed form.

3 METHODS

In this paper, we primarily focus upon the data from New York City collected from three types of materials, namely Metal, Metal/Plastic and Wood from the sources like Subway and Park. Also, we note the overall comparative taxonomic distribution in the respective subset of data of Boston and New York City. For most of the cases, we studied multiple number of samples for each data resource. Firstly, we

have uncompressed the DSRC [5] files to FASTQ/FASTA using DRSC or SRA Toolkit [6]. We have then used the metagenomics RAST server (MG-RAST) for the analysis of the shotgun metagenome data as described above [7]. At first dereplication i.e. removal of artificial replicate sequences which are produced by sequencing artifacts were performed and then these sequences were screened for species specific host sequence(human) using DNA level matching with Bowtie. After that the steps involved in processing with MG-RAST are RNA Detection, RNA Clustering, RNA Slims BLAT, Gene Calling, AA Filtering, AA Clustering, AA Sims BLAT, RNA Sims Annotation and Index Sim Sequencing.

We computed the α-diversity of the samples taken from different materials and sources of data collection [8]. The α-diversity denotes the diversity of organisms in a sample and is computed as follows.

$$ {}^{q}D_{\alpha} = \sqrt[1-q]{\sum_{j=1}^{N} w_j ({}^{q}D_{\alpha j})^{1-q}} $$

Here, N denotes the total number of subunits in the dataset, w_j is the nominal weight of the subunit j, ${}^{q}D_{\alpha j}$ represents the effective species density in the subunit j, and q is a constant.

4 RESULTS

We have studied the microbiome data separately for the materials from different sources and analyzed the data. We consider the materials like metal, plastic and wood as stationary materials and water as the only type of fluid. The observations are reported hereunder.

4.1 Preliminary Insights

We noted that a vast majority of the samples have a 'failed QC' percentage value of no more than 4%, thereby establishing the high-standard of datasets. We have compared our datasets with clustering hierarchical protein databases like COG, KO and NOG, and interestingly found that majority of the proteins covered by these samples are in fact annotated proteins mainly from bacteria, However, some of them include proteins from eukaryotes, viruses, archaea, etc. These proteins are mainly associated to the functionalities like metabolism, cellular processes and signaling, information storage and processing, etc. We see that the major part of the metagenome population (around 40%) belongs to carbohydrates, clustering-based subsystems and amino acids and its derivative subsystems for both the cities.

4.2 Study on Stationary Materials

The material wise analysis of the diversity values are highlighted in Table 2. It becomes evident from this table that the diversity of microbiome community is high across the materials (metal, plastic and wood) across different sources, which we considered. Interestingly, the diversity varies a lot (as evident from the high standard deviation values) for the different samples for all the materials studied in the Subway and its surrounding Park environment.

Table 2. The comparative α-diversity values (in terms of Mean and Standard Deviation, denoted as SD) of the microbiome community in the materials considered from different sources. The microbiome diversity appears to be consistent across the sources.

	Subway	**Park**
Metal	Mean = 47.9, SD = 25.97 (20 samples)	Mean = 67, SD = 25.59 (4 samples)
Metal/Plastic	Mean = 47.25, SD = 27.57 (12 samples)	Mean = 48.5, SD = 2.12 (2 samples)
Wood	Mean = 42.3, SD = 24.67 (10 samples)	Mean = 33.75, SD = 12.36 (8 samples)

4.3 Study on Fluid

Due to the non-stationary nature of water, the microbiome landscape is expected to differ a lot if taken as a material of comparison between multiple sources. We analyzed the microbiome data for different water samples collected from the sources like subways and canal and found interesting diversity of microorganisms. The α-diversity values are found to be in the order of several hundreds for all the samples we considered. The α-diversity values vary significantly (p-value < 0.0001) between the water samples collected from subways (Mean = 183.33, SD = 100.15) and canal (Mean = 603.1, SD = 55.96), having 3 samples and 10 samples, respectively. As well as, the class, order, family and genus of the observed microorganisms in different water samples were found to vary significantly. The interaction between species can put light to interesting relationships between different classes of organisms and their growth pattern. While comparing the α-diversity values between the stationary materials and fluid, highly significant differences of microbiome community were observed.

Table 3: The taxonomic distribution of the metagenomes studied in Boston and New York City. The unclassified entries have not been considered and only the most prominent ones are mentioned.

	BOSTON	**NEW YORK CITY**
Domain	Bacteria, Eukaryota	Bacteria
Phylum	Streptophyta, Actinobacteria, Firmicutes	Proteobacteria, Bacteroidetes
Class	Actinobacteria, Bacilli	Gammaproteobacteria, Deltaproteobacteria
Order	Actinomycetales, Solanales	Enterobacteriales, Pseudomonadales, Actinomycetales, Deltaproteobacteria
Family	Solanaceae, Apocynaceae, Corynebacteriaceae	Enterobacteriaceae, Moraxellaceae, Desulfobacteracea
Genus	Solanum, Nerium, Corynebacterium	Enterobacter, Acinetobacter, Desulfobacterium

4.4 Study on Multiple Cities

We have analyzed a few samples from Boston to enhance our understanding about the microbiome community in New York City. Metagenomic distributions of the samples studied from both the cities are shown in Table 3. As can be seen from the table, the taxonomic domain analysis in New York City highlights that it comprises mostly Bacteria (primarily belonging to the phylum Proteobacteria). In the case of Boston, we observed the presence of eukaryotes as well as bacteria significantly. The complete absence of eukaryotes and archaeae from many data of New York City is noteworthy. Unlike New York City, we saw various other phylum coexisting simultaneously (the most common ones being Streptophyta, Actinobacteria, Firmicutes) in Boston. Conversely in New York City, there is a plethora of Protobacteria, which constitute more than 90% of the total microbiomes.

5 CONCLUSION

The current paper provides some interesting highlights about the diversity of microbiome community in various materials. We observed that the community of microorganisms vary significantly between the materials. Additionally, they can also vary between the materials collected from different locations. It is interesting to further extend this analysis across different cities around the world. The inclusion of β-diversity and changes regarding spatial distribution of the samples can also be done for a better understanding of the microbiome diversity.

REFERENCES

1. M. Bhattacharyya, T. Ghosh, S. Shankar and N. Tomar, “The conserved phylogeny of blood microbiome,” *Molecular Phylogenetics and Evolution*, 109:404-408, 2017.
2. A. D. C. Fernandes, “Characterization of microbiome in Lisbon subway,” 2016.
3. X. Triadó-Margarit, M. Veillette, C. Duchaine, M. Talbot, F. Amato, M. C. Minguillón, V. Martins, E. de Miguel, E. O. Casamayor and T. Moreno. “Bioaerosols in the Barcelona subway system,” *Indoor Air* , 27(3): 564-575, 2017.
4. MetaSUB International Consortium, “The metagenomics and metadesign of the subways and urban biomes (MetaSUB) international consortium inaugural meeting report,” *Microbiome* 4(1):1-14, 2016.
5. Ł. Roguski and D. Sebastian, “DSRC 2 – Industry-oriented compression of FASTQ files,” *Bioinformatics* 30.15: 2213-2215, 2014.
6. S. Sherry, *et al.* “NCBI SRA Toolkit Technology for Next Generation Sequence Data,” *Plant and Animal Genome XX Conference (January 14-18, 2012). Plant and Animal Genome*. 2012.
7. E. M. Glass, J. Wilkening, A. Wilke, D. Antonopoulos and F. Meyer, “Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes,” *Cold Spring Harbor Protocols*, 2010(1):pdb-prot5368, 2010.
8. H. Tuomisto, “A consistent terminology for quantifying species diversity? Yes, it does exist,” *Oecologia*, 164(4), 853-860, 2010.