



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)



### Data Article

# Molecular investigation of bacterial communities: Data from two frequently used surfaces in the São Paulo Institute of Tropical Medicine



Tairacan Augusto Pereira da Fonseca<sup>a</sup>, Rodrigo Pessôa<sup>a</sup>,  
Sabri Saeed Sanabani<sup>a,b,\*</sup>

<sup>a</sup> *Clinical Laboratory, Department of Pathology, LIM 03, Hospital das Clínicas, School of Medicine, University of São Paulo, São Paulo 05403 000, Brazil*

<sup>b</sup> *São Paulo Institute of Tropical Medicine, University of São Paulo, São Paulo 05403 000, Brazil*

#### ARTICLE INFO

##### Article history:

Received 21 March 2016

Received in revised form

18 May 2016

Accepted 25 May 2016

Available online 2 June 2016

#### ABSTRACT

This article contains data on the bacterial population of two frequently used surfaces in the São Paulo Institute of Tropical Medicine (ITM) using the Illumina sequencing for massive parallel investigation of the bacterial 16S ribosomal RNA gene. Surface samples were obtained from restroom surfaces and the fingerprint door clock system. Mothur package and Shannon-ace-table.pl software programs (Chunlab Inc.: Seoul, Korea) were used to compute the diversity indices of bacterial community. The sequencing data from both surfaces have been uploaded to Zenodo: <http://dx.doi.org/10.5281/zenodo.47709>

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\* Correspondence to: Universidade de São Paulo, Faculdade de Medicina Instituto de Medicina, Tropical de São Paulo, LIM 56- Av. Dr. Enéas Carvalho de Aguiar, 470-2º andar-Cerqueira Cesar, 05403-000 Sao Paulo, SP-Brasil. Tel.: +55 11 3061 8699; fax: +55 11 3061 7020.

E-mail address: [sabyem\\_63@yahoo.com](mailto:sabyem_63@yahoo.com) (S.S. Sanabani).

<http://dx.doi.org/10.1016/j.dib.2016.05.064>

2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	<i>Biology</i>
More specific subject area	<i>Microbiome</i>
Type of data	<i>Text files: sequences</i>
How data was acquired	<i>Culture-independent Illumina massively parallel sequencing approach of the 16S rRNA genes using the Illumina sequencing-by-synthesis method (MiSeq platform).</i>
Data format	<i>Sequencing</i>
Experimental factors	<i>Bacterial genomic DNA was extracted and used as a template to amplify the V4 region of the 16S rRNA gene. The amplicons were molecularly barcoded, pooled, and sequenced by a paired-end protocol (Illumina).</i>
Experimental features	<i>Illumina massively parallel sequencing of the 16S rRNA gene libraries and subsequent analysis.</i>
Data source location	<i>São Paulo, Brazil</i>
Data accessibility	<i>Data is within this article and have been uploaded to Zenodo (<a href="http://dx.doi.org/10.5281/zenodo.47709">http://dx.doi.org/10.5281/zenodo.47709</a>).</i>

## Value of the Data

- Exploration of bacterial communities on these surfaces in our institute and other establishments will advance our knowledge of microbial trafficking and colonization in these systems.
- These data are publically available and their access permits researchers to conduct further analyses based on their own studies with newly developed bioinformatics packages.
- The data presented here underscores the need to increase public awareness on the importance of personal hygiene and proper hygienic practices of inanimate surfaces to reduce the transmission of disease-causing organisms.

## 1. Data

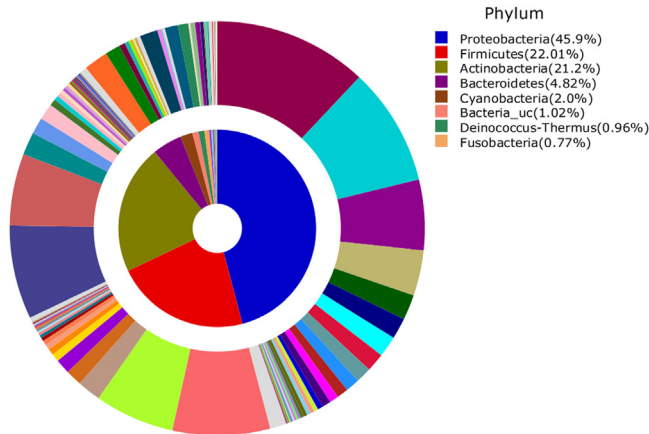
All sequence data described here are available in the online Zenodo repository: <http://dx.doi.org/10.5281/zenodo.47709>. The two sets of massively parallel sequence (MPS) data represent the bacterial communities from the restroom surfaces and the fingerprint door clock system (FDLS) in the São Paulo Institute of Tropical Medicine (ITM).

## 2. Experimental design, materials, and methods

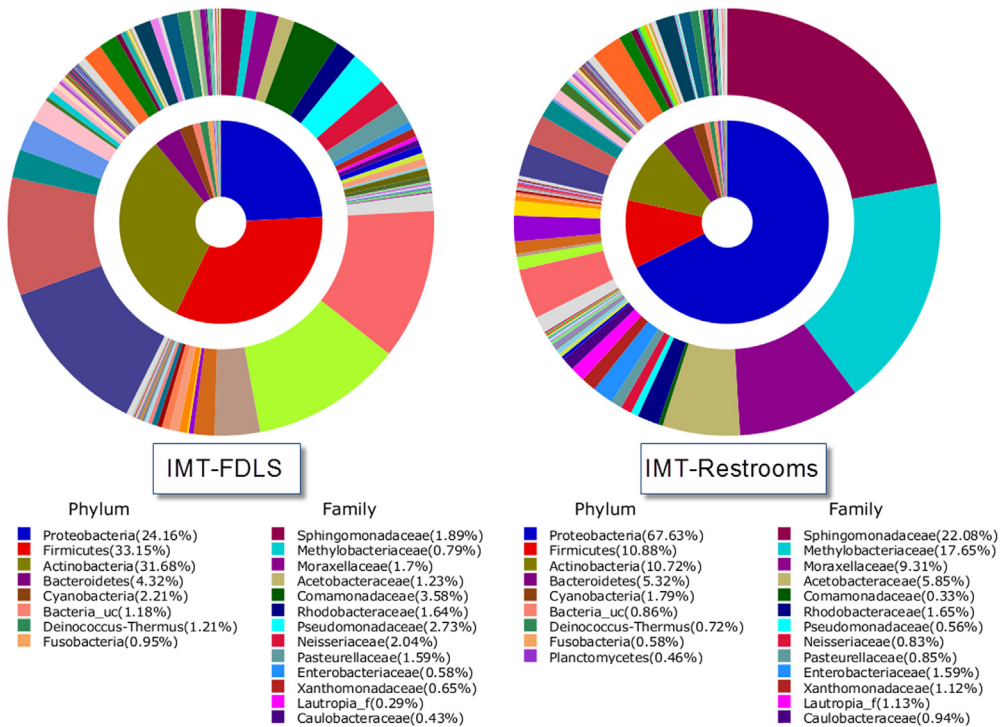
Surfaces of four IMT-FDLS that are frequently used by the institute's professionals were swabbed and included in this study. Samples were also obtained from 12 surfaces in two male and two female restrooms including door handles into and out of the restroom, toilet seats, faucet handles, and hands-toilet flush handles. All surfaces were sampled and treated as previously described [1].

Genomic DNA was extracted from each swab using the PowerSoil DNA kit (MO BIO Laboratories™: Carlsbad, CA, USA) as per the manufacturer's instructions. After the extraction, DNA from each surface group was pooled together, and the pooled DNA was submitted to the amplification of the V4 region of the 16S rRNA gene using the previously published primers Bakt\_341F/Bakt\_805R [2] according to the conditions previously described [1,3]. Purification of the amplified products were performed

using the Freeze N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad: Hercules, CA, USA), following the recommendations of the manufacturer. After purification and quantification on a Qubit 2.0 Fluorometer (Life Technologies: Carlsbad, CA, USA), the amplicons from each surface group were pooled at equimolar concentration, and diluted to 4 nM. Indexing of DNA and preparation of libraries were



**Fig. 1.** Average composition of bacteria from all samples (inner area: Phylum, outer area: Family). Phyla and Families with more than 1% of their proportion were represented.



**Fig. 2.** Average composition of bacteria from each sample (inner area: Phylum, outer area: Family). Only bacterial phyla and families that had a relative abundance of 1% or greater are presented.

**Table 1**  
Identities of the six most abundant OTUs in the bacterial communities.

Abundance order	Taxa (% of abundance)	
	IMT-FDLS	IMT-restrooms
1	<i>Propionibacterium acnes</i> (7.9%)	<i>Moraxella osloensis</i> (5.6%)
2	<i>Streptococcus dentisani</i> (3.5%)	<i>Methylobacterium phyllostachyos</i> (5.4%)
3	<i>Staphylococcus epidermidis</i> (3%)	EU440723_s (4.2%)
4	<i>Staphylococcus capitis</i> (2.4%)	<i>Sphingomonas ginsenosidimitans</i> (3.8%)
5	<i>Streptococcus_uc m</i> (1.9%)	<i>Roseomonas mucosa</i> (3.7%)
6	Streptococcaceae_uc_s (1.9%)	Sphingomonadaceae_uc_s (3.4%)

performed as previously reported (our references). The prepared library was finally loaded on an Illumina MiSeq cartridge for paired end 300 sequencing.

Image analysis, base calling, and data quality assessment were initially performed on the MiSeq instrument (San Diego, CA, USA). Sequence cleaning and analysis was conducted according to the recently published methods by our group [1,3]. Briefly, the PANDAseq v.2.9 software was used to pair the MiSeq forward and reverse reads [4] with default parameters. The UCHIME algorithm [5] was used to detect and remove any potential recombinant sequences from the generated data. The EzTaxon-e database [6] adjusted at a 97% cutoff of sequence similarity identity was used to define the taxonomic classification of each read. The Chao1 estimation and the Shannon diversity index at a 3% distance were used to calculate the richness and diversity of samples. The Mothur and Shannon-acceptable.pl software programs (Chunlab Inc.: Seoul, Korea) were used to compute the diversity indices of bacterial community. The Fast UniFrac [7] was used to estimate the overall phylogenetic distance between communities. This distance was visualized using principal coordinate analysis (PCoA). Shared OTUs operational taxonomic units (OTUs) were identified using XOR analysis of CL community program v3.43 (Chunlab Inc.: Seoul, Korea).

The most abundant OTUs at phylum and family levels that accounted for more than 1% of all sequences, IMT-FDLS and, IMT-restrooms are shown in Figs. 1 and 2. The six OTUs of the most abundance species associated with the two sample libraries are provided in Table 1.

## Acknowledgments

This work was supported by grants 2011/12297-2 and 2014/26983-3 from the Fundação de Amparo à Pesquisa do Estado de São Paulo.

## Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.05.064>.

## References

- [1] T.A. Pereira da Fonseca, et al., Diversity of bacterial communities on four frequently used surfaces in a large Brazilian teaching hospital, *Int. J Environ. Res. Public Health* 13 (2) (2016).
- [2] D.P. Herlemann, et al., Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea, *ISME J.* 5 (10) (2011) 1571–1579.
- [3] T.A. Pereira da Fonseca, R. Pessoa, S.S. Sanabani, Molecular analysis of bacterial microbiota on Brazilian currency note surfaces, *Int. J Environ. Res. Public Health* 12 (10) (2015) 13276–13288.
- [4] A.P. Masella, et al., PANDAseq: paired-end assembler for illumina sequences, *BMC Bioinform.* 13 (2012) 31.

- [5] R.C. Edgar, et al., UCHIME improves sensitivity and speed of chimera detection, *Bioinformatics* 27 (16) (2011) 2194–2200.
- [6] O.S. Kim, et al., Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species, *Int. J. Syst. Evol. Microbiol.* 62 (Pt. 3) (2012) 716–721.
- [7] M. Hamady, C. Lozupone, R. Knight, Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data, *ISME J.* 4 (1) (2010) 17–27.