International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Taxonomical Annotation of Whole Genome Metagenomic Data to identify Microbiome in Indian TB Patients

Tanusree Chaudhuri, Vidya N, A H Manjunatha Reddy, Varun S

Abstract*: Till date, Tuberculosis is a major health problem especially in India. Though, Mycobacterium tuberculosis is the causative agent of tuberculosis (TB) and it kills approximately about 1.3 million people every year according to WHO, yet the role of complex microbial community favoring the growth of Mycobacterium complex is immense. There are several studies that have already been reported that there exists a strong association between the complex microbial interactions in the community and the disease condition for majority of the diseases that are caused by different microorganisms.Our present analysis aims to characterize up-till species level abundance in Mycobacterial tuberculosis metagenome. In our study we have taken a total of 100 whole genome sequenced Indian Tuberculosis patient samples. These 100 samples were analyzed to obtain taxonomical abundance. It was seen that South Indian samples had 54 unique species whereas North Indian samples had only 13 unique species. Our study revealed that Achromobacter, Nocardia, Chromobacterium, Staphylococcus, Xanthomonas, Bacillus, Sanguibacter and Bordetella are mostly abundant in Indian tuberculosis patients and we are able to classify all these till species level, which other 16s rRNA studies have failed to do. Knowing the species present during the tuberculosis infection it will be of great importance to treat the patients with right antibiotics for the microbiome present.

Keywords* : *M.Tuberculosis; Metagenome; WGS; Taxonomical abundance; Microbiome

I. INTRODUCTION

Tuberculosis remains as a major global health problem [1]. Mycobacterium tuberculosis is the causative agent of tuberculosis (TB) and it kills approximately about 1.3 million people every year [2]. In India approximately 2.790 million persons get affected by this disease [3]. Here, basically, TB occurs at high rates because of high air pollution rates that causes many effects in the air people breathe, poor built environments including hazards in the workplace, poor ventilation, and overcrowded homes, and least chances of early detection [4].So, at present, India belongs to one of such countries, which has one of the highest burden of TB and the situation has become more adverse as the effective diagnosis of TB patients is quite challenging for the public health communities [5]. The majority of the affected people are poor and cannot access TB care financially .Though it has been claimed by different organization that, the success rate in microbiologically confirmed new TB patients are consistently above 85% and in previously treated microbiologically confirmed TB patients are almost more than70%, yet, there are many cases in India which has not been listed at all [6]. So, it is an urgent need to analyse Indian TB cases more intriguingly to search for the methods of early detection as well as to provide affordable treatment to all.

Most of the research articles, that tried to shed a light towards understanding the effect of tuberculosis infection on public health, have not taken into account the members of the normal flora in a complex ecological community [7]. Studies have revealed that there exists a strong association between the complex microbial interactions in the community and the disease condition for several diseases that are caused by different microorganisms. Tuberculosis is also no exception [8]. Hence it is important to study the microbiota during tuberculosis infection, which might give us an insight into how different organisms present in the host affect the disease. It is seen that during tuberculosis, Mycobacterium tuberculosis complex varies with respect to geographical locations, patient condition and even with other existing diseases. In India the food habits, lifestyle, and also environment changes diversely according to different geographical locations. As for example, Delhi, Chandigarh, Sikkim, Himachal Pradesh and Gujarat are the states that show the highest occurrence of Tb cases in India [3]. Hence it is expected that the complex bacterial community associated during the Mycobacterium infection will vary from place to place as well as from person to person. Many researchers have used 16s rRNA sequencing strategy as a method to detect the bacterial community present, but it has a limitation of taxonomically classifying the data not up-to species level, which according to us are the urgent need for the understanding of tuberculosis on a broader aspect(9).

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Tanusree Chaudhuri *, Department of Biotechnology, RV College of Engineering, RV Vidyanikethan Post, Mysuru Road, Bengaluru, Karnataka,India,560059,Email: tanusree.ch3@gmail.com
Vidya N, Department of Biotechnology, RV College of Engineering, RV Vidyanikethan Post, Mysuru Road, Bengaluru, Karnataka, India, 560059,Email: vidya.n@rvce.edu.in
A H Manjunatha Reddy, Department of Biotechnology, RV College of Engineering, RV Vidyanikethan Post, Mysuru Road, Bengaluru, Karnataka, India, 560059,Email: ahmanjunatha@rvce.edu.in

This will not only help us to eradicate TB by 2025 but also help us to repurpose existing drugs that are used as medication for other bacterial diseases. The researchers like Krishna et.al.,[7] have already classified 16s rRNA of Mycobacterium tuberculosis till the genus level abundance.

But, very few analysis are there where the classification is till species level abundance. In our present work, we used whole genome data to find the taxonomical abundance till the species level. Whole genome sequences undergo assembly and binning steps which utilizes best algorithms to classify the reads to species level with higher confidence [9].

For our present study we have considered 100 Indian Tuberculosis patient samples classified based on their geographic location in north as well as South India. We performed assembling of reads to obtain contigs which were binned to produce whole genome of organism. For this data further analysis revealed the taxonomy annotation and average abundance of microorganisms present at species level.

II.MATERIALS AND METHODS

2.1 Data retrieval, Quality analysis and Assembly

Tuberculosis metagenomics samples were obtained from SRA [10]. SRA database hosted by NCBI consists of sequencing data and other details regarding the sequences. A total of 945 whole genome sequenced Indian samples were found in the database of which 50 good samples each of north Indian and South Indian respectively were classified for this project. 'SRAToolkit 2.9.4 [11] was used for retrieving the data from SRA Database. prefetch module was used to download the .sra files from the database and fastq-dump module was used to convert the .sra files to fastq format. Next, As a preliminary step, samples were analyzed for quality and samples with low quality reads were preprocessed. Quality was analyzed using FastQC [12]. Those samples with low quality are scrutinized using CutAdapt [13]. To get an overview on the number of organism present kraken tool was used [14]. The result revealed a huge number of unique species with less abundance. To minimize the false positive results and increase the confidence score of the reads aligning to particular organism we subjected the samples for assembly and binning. These would provide binned files containing whole genome of organism which would reduce the number of shorter reads hitting to different organism. While performing genome assembly, we considered paired end reads which is produced when the fragment size is much longer (typically 250 - 500 bp long) and the ends of the fragments are read towards the middle. This produces two "paired" reads. One from the left-hand end of a fragment and one from the right with a known separation distance between them. The goal of a sequence assembler is to produce long contiguous pieces of sequence (contigs) from these reads. The contigs are sometimes then ordered and oriented in relation to one another to form scaffolds. The distances between pairs of a set of paired end reads is useful information for this purpose. MetaSPADES being the widely used assembler for shotgun metagenomics sequences, the same was used for our study [15]. We used all possible kmers like 21, 33, 55, 77, 99, 127. The MetaSPADES output provided a single contig file containing best contigs from each kmer chosen. The assembly results were evaluated using Quast [16].

2.2 Bining, taxonomy annotation and taxonomy binning

Binning algorithms attempt to group contigs or scaffolds from the same or closely related organisms, so that taxonomic assignment and functional analysis can be done on them instead of individual contigs. Binning has shown to cluster contigs even from rare species and can recover draft genomes from previously uncultivated bacteria [17]. MaxBin tool was used for our present analysis, where the contig file was given as an input [18]. Completeness and contamination of binned files were assessed used checkm [19]. PhymmBL, a classifier for metagenomic data has been used for phylogenetic classification purpose. Where, we have trained 539 complete, curated genomes and can accurately classify reads as short as 100 bp, representing a substantial leap forward over previous composition-based classification methods [20]. We used MEGAN for analysis of large metagenomic data sets. In a preprocessing step, the set of DNA sequences is compared against databases of known sequences using BLAST or another comparison tool, and then MEGAN is used to compute and explore the taxonomical content of the data set, employing the NCBI taxonomy to summarize and order the results. A simple lowest common ancestor algorithm assigned reads to taxa such that the taxonomical level of the assigned taxon reflects the level of conservation of the sequence. The raw BLAST output is given as input to the MEGAN software to do taxonomy binning and generate taxonomical name to percentage abundance file [21].

2.6 The obtained abundance file was analyzed statistically to confidently characterize the microbiome diversity in the TB samples.

III.RESULT AND DISCUSSION

3.1 Human respiratory tract is a region like human GI tract which is heavily exposed to microorganisms, specially influenza virus, Mycobacterium tuberculosis, and respiratory syncytial virus. But, with the current advancement of the sequencing techniques, it is now possible to capture the alteration of the microbiota profile in the host in response to the infection of these dangerous pathogens [22]. Presently, the next-generation sequencing technology is being used as a technique to determine the number of microbes present at certain stages of infection from the sputum samples as a comprehensive view of the microbiota contemporaneous to the lower respiratory tract.

So, in our present analysis, a total of 945 sputum samples from Tb patients with different age groups were collected from NCBI SRA database pertaining only to Indian cases of tuberculosis. As our major goal is to classify the microorganisms present in the tuberculosis cases from India, we collected the metadata only from Indian patients. As we know that the population differences such as diet, geography, animal facility, antibiotics etc is highly related to with microbiome shifts we did not select our samples based on one geographical location, rather, we distributed our samples into two groups , such as samples from north India and samples from south India [23]. Among all samples, we have selected 50 good samples randomly with respect to south India and north India respectively, so that the analysis produces an unbiased data.

All whole genome samples, that we selected for our analysis was sequenced by Illumina sequencer and having paired end reads as it allows to sequence both ends of a fragment and generate high-quality, align able sequence data. This indicates that, the results that are obtained from our data should be having high precision.

The 50 samples, taken for North Indian data analysis had an origin from Mumbai, Agra, Delhi, Pune and Punjab. The average spot length of North Indian samples was 300. All north Indian samples were submitted to SRA between years 2016-2018. The 40 GB North Indian samples were retrieved from SRA database and the compressed .sra files were converted to fastq format.

The 50 South Indian samples were downloaded from SRA database had origin mostly from Tamil Nadu. There were no other samples from other regions in South India that were deposited to SRA at the time of our analysis. The average spot length of South Indian samples was 200. Samples submitted to SRA database was in between years 2014- 2018. The 45 GB South Indian samples were retrieved from SRA database and compressed .sra files were converted to fastq format same like north Indian sample.

3.2 Presently, due to the advancement of current techniques, the number of next-generation related whole genome datasets is being increasing very rapidly, but all of them are not having good quality of data. Hence, the datasets have to be processed by Quality Control analysis procedures before they could be utilized for downstream analysis purpose. Otherwise the results that we obtained from the analysis might be misleading. These procedures usually include identification and filtration of sequencing artifacts such as low-quality reads and contaminating reads [24]. For our study, quality analysis of our data was done using FastQC tool. The important parameters to access the quality of the read is per base sequence quality and per base sequence content. We have measured per base sequence quality by Phred score (Q).

Per base sequence quality of North Indian samples were low and had an average Q value of 28. Therefore, we used CutAdapt to preprocess the reads and trim the 3' end. The average read length obtained after preprocessing was 280 bp. As, the per base sequence quality increased to an average Q value of 30 after trimming, these sequences were used for further analysis. South Indian Samples had a good per base sequence quality with an average Q value of 32. Hence, they were used for further analysis without any preprocessing.

After running the Kraken tool on the preprocessed reads, the results revealed that there were a large number of unique sequences with lesser hits present in case of most of the sequences. Hence, we assembled the reads into larger contigs and performed binning for them to the whole genome. This method annotated the read with higher confidence.

3.3 After quality control, the reads that are being obtained can either be assembled into longer contiguous sequences called contigs or passed directly to taxonomic classifiers [25]. We decided to assemble our reads into contigs before we proceed for the taxonomical analysis. We used MetaSPADES for assembly [26]. MetaSPAdes showed the overall best assembly size statistics while also capturing a relatively large fraction of the expected diversity. Though the usage of this was quite identical to that of SPAdes, but it was quite flexible, while regarding the format of the input data [27]. The assessment of the contigs showed that the North Indian samples had lesser N50 value as well as lesser total length of contig and is explained in the following figures (Figure 1 a and b). These results indicate that, south Indian samples are more likely to show more abundance of different microorganisms rather than north Indian samples. As our major objective is to completely classify the microorganisms present in case of Indian tuberculosis infections, hence we decided to perform binning to get large sequences.

Figure1. a) Average total contig length b) Average N50 values of South India and North India samples

a) Total length is the total number of bases in the assembly1. It was observed that North Indian samples had an average contig total length of 3817034 and South Indian samples had an average contig total length of 97,152.b)The N50 is the length for which the collection of all contigs of that length or longer covers at least half an assembly18.

It was observed that North Indian samples had an average N50 value of 28,641 and South Indian samples had an average N50 value of 97,152.

3.4 Our present study, aims to characterize the taxonomic and functional diversity of the microbiome present during Mycobacterium tuberculosis infection through analysis of DNA fragments. That's why we have adapted binning approach. For our present analysis, we used MaxBin tool for binning. This yielded binned files with more than 95% complete genome and less than 10% redundancy. We checked the completeness and contamination of the binned files using Checkm, where completeness is measured as the estimation of presence/absence of marker genes and the expected co-localization of these genes and contamination is determined by the presence of multi-copy marker genes and the expected co-localization of these genes which is explained in figure 2.

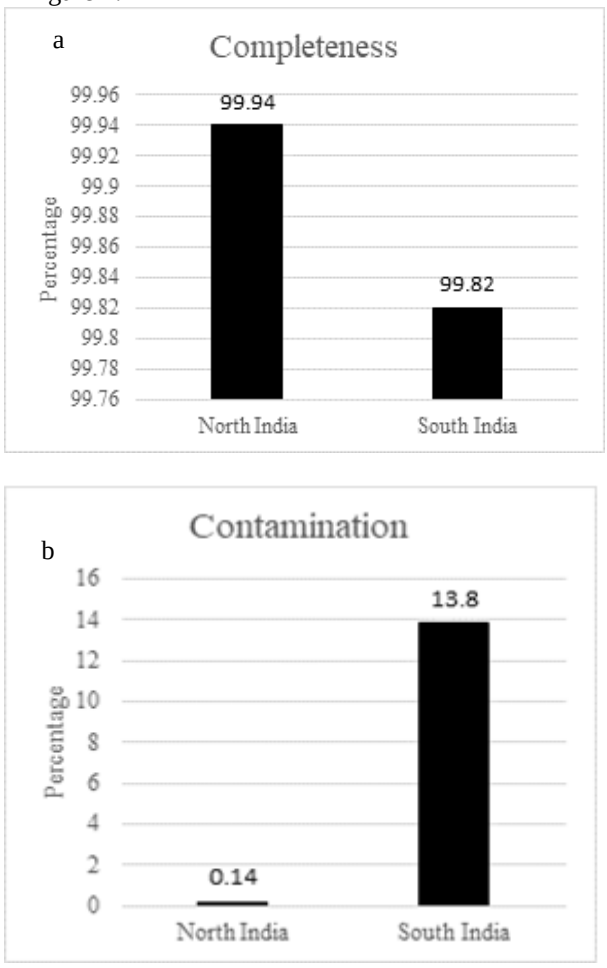
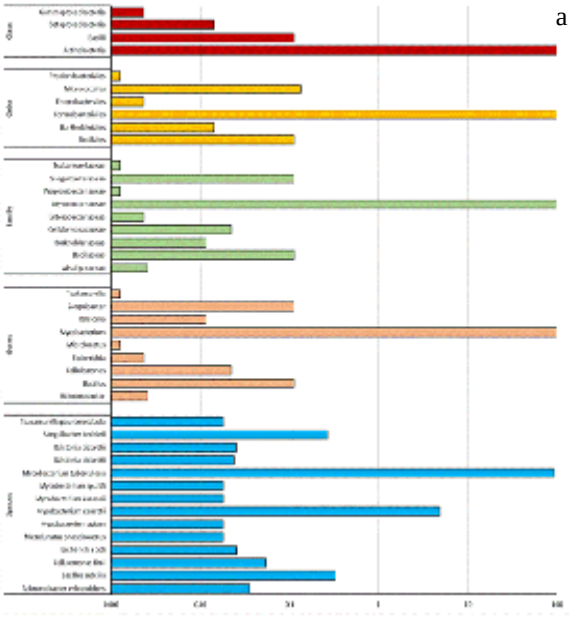


Figure2. a) Average Completeness and b) average contamination of binned files of South India and North India samples

a) It was observed that North Indian sample had 99.94% of completeness. South Indian sample was seen to have 99.82% completeness of genome.b) North Indian sample had 0.14% of contamination. South Indian sample was seen to have 13.8% contamination of genome.

It was seen that, in case of North Indian sample most of the reads were hitting to only Mycobacterium Genus. This is the major cause for observation of least contamination in these samples. On the contrary South Indian Samples had reads hitting to different genus which increased the contamination with respect to marker gene. This idea has lead us to conclude that, multi species abundance is more in case of south Indian sample rather than north Indian samples.

3.5 Further, we tried to check the unique species present in each of North Indian and south Indian samples. For this purpose, we adapted hybrid binning strategies. Where, we wanted to do binning by incorporating a combination of alignment and composition-based strategies for taxonomic classification. We selected PhymmBL tool which uses a composition-based methodology of Phymm along with an alignment-based step (BLAST) to improve the confidence of taxonomic assignments, for this purpose. The output has been used for taxonomy binning using MEGAN tool. This time we followed least common ancestor method to generate phylogenetic tree for each sample. We set the maximum hit for each read as to 100. At the same time, we provided the binned fasta file as reference for taxonomy binning. The taxonomy name as well as percentage ratio of each species assigned, were extracted from MEGAN in tab delimited format is explained by figure 3. Where each bar represents the abundance value for each of the phyla identified from both north Indian as well as south Indian samples.



Retrieval Number: A4570119119/2019©BEIESP
DOI: 10.35940/ijitee.A4570.119119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

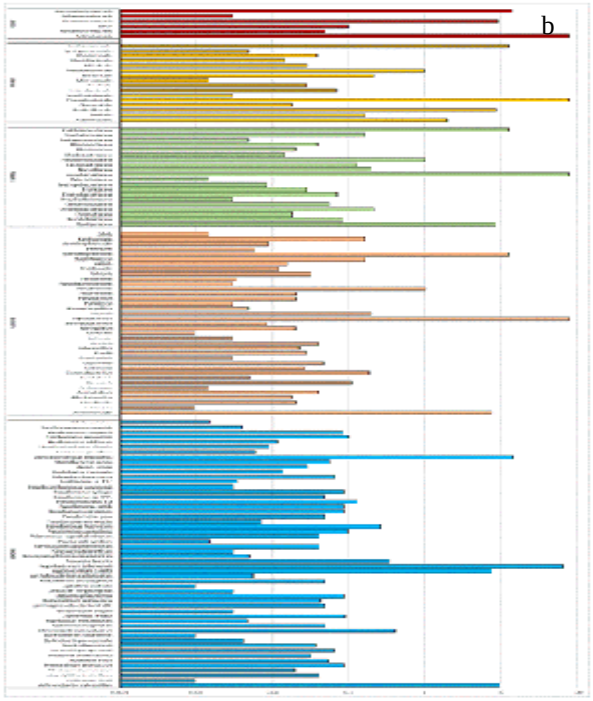
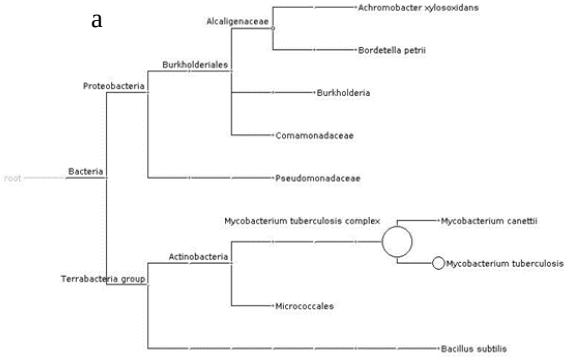


Figure3. Relative abundance of Organisms in North Indian as well as South Indian samples

From figure 3 a and b is have been observed that in case of north Indian sample although Mycobacterium tuberculosis is the most abundant species, yet Achromobacter xylosoxidans, Bacillus subtilis, Cellulomonas fimi, Escherichia coli, Microlunatus phosphovorus, Mycobacterium avium, Mycobacterium kansasii, Mycobacterium sp. JLS, Ralstonia pickettii, Sanguibacter keddieii, Tsukamurella paurometabola are also present with very less abundance. But in case of south Indian sample, other than Mycobacterium tuberculosis, Stenotrophomonas maltophilia , Achromobacter Xylosoxidans and Mycobacterium canetii was found with a good abundance range from 8-16% along with other species with relatively less abundance as expected.

3.6 To identify the interspecies relationship among the species that has been identified from the north as well as south Indian samples, we plotted phylogenetic tree through Megan. For this purpose we have used, the binned output from PhymmBL. The phylogenetic tree is being displayed in the figure 4. We used the cut off value for a read to be assigned to the taxon as 0.1% .



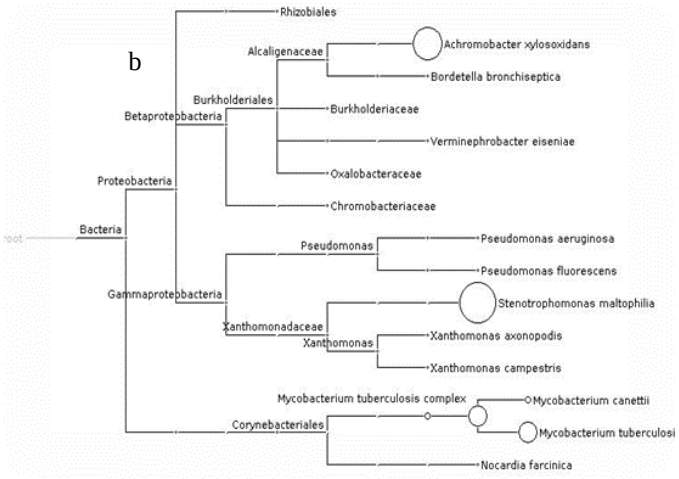


Figure4. a.Phylogenetic tree for North India sample, b. Phylogenetic tree for South India sample

In the phylogenetic tree for the family level to which maximum number of reads were hit, were represented in the species level and rest were kept at family level only.

It has already been reported in several studies, that the microbiota in the sputum sample of pulmonary tuberculosis patients are more diverse than those of healthy participants [28]. Previous studies have already reported many genera that were unique to in the sputum of pulmonary tuberculosis patients, they are Phenylobacterium, Stenotrophomonas, Cupriavidus, Caulobacter, Pseudomonas,Thermus, Sphingomonas ,Pelomonas, Acidovorax, Brevibacillus, Methylobacterium, Diaphorobacter, Comamonas, Mobilicoccus, Fervidicoccus, Serpens, Lactobacillus, Thermobacillus, Auritidibacter, Deinococcus, Lapillicoccus, Devriesea respectively [29]. But, when we performed our analysis only with indian sample, we have identified a wide range of genera. To be very specific, other than Acidovorax, Cupriavidus, Methylobacterium, Pseudomonas, Stenotrophomonas Indian tuberculosis cases are prone to contain a wide range of microorganism, which are specific only to Indian tuberculosis cases. Our study has identified 38 unique genera for a combined set of North as well as south Indian samples .They are Achromobacter, Alicycliphilus, Allochromatium, Aromatoleum, Arthrobacter, Bacillus, Bordetella, Burkholderia, Cellulomonas, Chromobacterium, Collimonas, Desulfovibrio, Escherichia, Frankia, Herbaspirillum, Klebsiella, Laribacter, Leptothrix, Methylibium, Microlunatus, Mycobacterium, Nocardia, Novosphingobium, Paracoccus, Parvibaculum, Polaromonas, Pseudoxanthomonas, Pusillimonas, Ralstonia, Rhodobacter, Sanguibacter, Shigella, Staphylococcus, Tsukamurella, Variovorax, Verminephrobacter, Xanthomonas and Xylella. This explains why Indian population in more prone towards becoming active tuberculosis patients. Moreover, analysis has even in depth to classify till species level of the microorganisms present in the Indian tuberculosis cases.

IV.CONCLUSION

Metagenomics is the advanced and intrigued study of complex microbial communities from varied sources, like sputum, remains of oral and gut microbiota etc. that. The major challenge of metagenomics remains in accurate assignment of taxonomic species. There are many available computational tools that are able to classify microorganisms from whole-genome shotgun sequencing data till species level, so that one can classify microorganisms and map their roles in different aspects of human health. As tuberculosis is one of the major health problems in India, through our study we wanted to provide a special attention while classifying Indian patient's metagenomic data till species level. There are studies that have already explored the potentials of shotgun metagenomics for detection and characterization of strains in sputum samples obtained from The Gambia in West Africa [30]. Our present analysis wanted to follow the same steps to classify strains obtained from only India. In our present analysis, we listed out that the major bacterial species demonstrated among Indian TB cases.Our analysis further reveals that Achromobacter xylosoxidans, Mycobacterium canettii and Mycobacterium tuberculosis, were the most abundant bacterial species in both South Indian as well as North Indian samples. However, statistically, significant differences were observed in the proportion of Achromobacter xylosoxidans being more present in South Indian samples with a larger abundance of 9.63% than in North Indian sample which showed 0.035%. Stenotrophomonas maltophilia was one of the unique species present in south Indian Samples which had a statistically significant abundance of 14.36%, which was completely absent in North Indian samples. When Mycobacteium tuberculosis complex was compared between North Indian and South Indian samples it was seen that Mycobacterium avium, Mycobacterium kansasii, Mycobacterium sp. JLS were seen only in North Indian samples but not in South Indian samples. Mycobacterium tuberculosis was seen as much as 94.32% of abundance in North Indian samples but had only 65.63% in South Indian samples. The main interest of characterizing bacterial community is to determine whether and how active tuberculosis condition is associated with a particular human bacterial community especially in India, and out study highlights that aspect. Though we have tried the best of our knowledge, to characterize the bacterial community till species level, yet we believe, more work should be carried out in this regard.

REFERENCES

1. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification and assembly. Brief Bioinform. 2017;(September).
2. https://doi.org/10.7717/peerj.585.
3. https://www.who.int/tb/publications/global_report/tb18_ExecSum_web_4Oct18.pdf?ua=1
4. https://www.tbfacts.org/tb-statistics-india/
5. Hargreaves JR, Boccia D, Evans CA, Adato M, Petticrew M, Porter JDH. The social determinants of tuberculosis: from evidence to action. Am J Public Health. 2011;101(4):654–62.
6. Verma R, Khanna P, Mehta B. Revised national tuberculosis control program in India: The need to strengthen. Int J Prev Med. 2013;4(1):1–5.
7. Krishna P, Jain A, Bisen PS. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. Eur J Clin Microbiol Infect Dis [Internet]. 2016;35(7):1205–10. Available from: http://dx.doi.org/10.1007/s10096-016-2654-4
8. Pédron T, Sansonetti P. Commensals, Bacterial Pathogens and Intestinal Inflammation: An Intriguing Ménage à Trois. Cell Host Microbe. 2008;3(6):344–7.
9. Potential Of Microbiome Research In Respiratory Diseases. J Respir Res. 2015;1(1):13–4.
10. Karsch-Mizrachi I, Takagi T, Cochrane G. The international nucleotide sequence database collaboration. Nucleic Acids Res. 2018;46(D1):D48–51.
11. https://ncbi.github.io/sra-tools/
12. Andrews S.FastQC A Quality Control tool for High Throughput Sequence Data, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
13. Martin M.Cutadapt removes adapter sequences from high-throughput sequencing.EMBnet.journal.2015;1–3.
14. Salzberg SL, Wood DE. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol [Internet]. 2014;15. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1093857
15. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler (Supplementary Material). Genome Res. 2017;27(5):824–34.
16. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
17. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144–6.
18. Yu-Wei Wu, Blake AS, Steven WS. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets, Bioinformatics, 2016;32(4):605–607
19. Imelfort M, Skennerton CT, Parks DH, Tyson GW, Hugenholtz P. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
20. Brady A, Salzberg S. Phymm and PhymmBL: Classification with Interpolated Markov Models. Nat Methods. 2009;6(9):673–6.
21. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification and assembly. Brief Bioinform. 2017;(June):1–15.
22. Vollmers J, Wiegand S, Kaster AK. Comparing and evaluating metagenome assembly tools from a microbiologist's perspective - Not only size matters! Vol. 12, PLoS ONE. 2017. 1–31 p.
23. Adami AJ, Cervantes JL. The Microbiome at the Pulmonary Alveolar Niche: How It Affects the Human Innate Response against Mycobacterium tuberculosis HHS Public Access. Tuberc [Internet]. 2015;95(6):651–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4666774/pdf/nihms712460.pdf
24. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377–86.
25. Eshetie S, Van Soolingen D. The respiratory microbiota: New insights into pulmonary tuberculosis. BMC Infect Dis. 2019;19(1):1–7.
26. Manimozhiyan A, Shu Y-Z, Arcuri M, Kozlowski M, Wang R, Lam KS, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174–80
27. Knights D, Ward TL, Mckinlay CE, Miller H. Rethinking "Enterotypes" Dan. Cell Host Microbe. 2017;16(4):433–7.
28. Wood MR, Yu EA, Mehta S. Review article: The human microbiome in the fight against tuberculosis. Am J Trop Med Hyg. 2017;96(6):1274–84.
29. Zhang Y, Lun CY, Tsui SKW. Metagenomics: A newway to illustrate the crosstalk between infectious diseases and host microbiome. Int J Mol Sci. 2015;16(11):26263–79.
30. Stephen Nayfach1 2 and Katherine S. Pollard. Accurate, Toward Metagenomics, Quantitative Comparative. 2016;166(5):1103–16