Available online on 15.04.2022 at http://jddtonline.info

Journal of Drug Delivery and Therapeutics

Open Access to Pharmaceutical and Medical Research

Open Access Full Text Article

Research Article

Metagenomics Analysis of Breast Cancer to Study Bacterial Diversity

Muskan Jain and Ruchi Yadav*

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, UP, India

Article Info:

Article History:

Received 21 February 2022
Reviewed 29 March 2022
Accepted 07 April 2022
Published 15 April 2022

Cite this article as:

Jain M, Yadav R, Metagenomics Analysis of Breast Cancer to Study Bacterial Diversity, Journal of Drug Delivery and Therapeutics. 2022; 12(2-s):119-126

DOI: http://dx.doi.org/10.22270/jddt.v12i2-s.5433

*Address for Correspondence:

Ruchi Yadav, Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, UP, India

ORCID ID: *https://orcid.org/0000-0002-0971-2667*

Abstract

Introduction: Breast cancer is the swelling structures in the breast cells. After skin cancer, breast cancer is the most broadly accepted malignancy analysed in ladies in the United States. Breast cancer can happen in both males and females, yet it's undeniably more normal in ladies. Indications of mammary glands cancer may involve a ball or lumps in the breast, an alteration of breast shape, scratching of the skin, fluid coming out from areola, a recently reversed areola, or a red or textured fix of skin. In those with removed lay out of the illness, there can be bone agony, swollen lymph hubs, windedness, or yellow skin. There are certain microorganisms that are involved in breast cancer and have function thereof. The study of these microorganisms can be done through metagenomics study and comparing with normal samples.

Methods: For current study pairwise metagenomics data for breast cancer sample was retrieved from ENA site. Four FASTQ files were retrieved with the Accession no PRJEB25419 and complete metagenomics analysis was done using GALAXY Server. Tools like FASTQC, Trim galore, KRAKEN2, Krona pie chart etc. were used for classification and identification of microorganisms in the samples.

Results: The taxonomic and functional analysis of the metagenome of breast cancer identifies different organisms like. *Actinobacteria*, *Bifidobacterium cuniculi*, *Murnine leukemia virus*, *Human endogenous retrovirus*, *Pezizomycotina*. Different microbes and their percentage in the breast cancer have been predicted that are 4% of *Nematocera*, 2% of *Brachycer*a, 3% of *Galegeae*, 2% of *Nelumbo nucifera*.

Conclusion These results give insights into the metagenome of breast cancer that can be used for clinical perspectives. Detailed study is required to establish the association of these microbes with development and progression of breast cancer.

Keywords: - Breast cancer, Metagenomics, Galaxy, ENA Database, Krona pie chart.

INTRODUCTION

Metagenomics is the inspection of hereditary material recuperated straightforwardly from environmental instance[1]. This field may also be alluded to as natural genomics, Eco genomics or community genomics. Metagenomics gives a culture independent perception about microbial communities[2]. It has come out as a powerful core since most microorganisms can't grow in a pure culture and that culturing can't capture that full spectrum of microbial variegation[3].

Metagenomics provides the powerful tool of research and study about prokaryotes and viruses that are existing in the adjoining through the analysis of their DNA of microbes[4]. It can not only recognize the microbes but also gives perception into the practical and metabolic part of the microbes[5]. There are also many biotechnological applications of metagenomics as it gives the great impact on the industrial production. Metagenomics has also been useful in finding the diseases that are caused by microorganism[6]. Shotgun metagenomics is used in the discovery and diagnosis of pathogens in clinical samples. For RNA viruses, firstly RNA is isolated from the sample which is converted to cDNA. Besides bacterial pathogen and viruses, metagenomics also helps in the detection of parasitic infection[7].

Metagenomics also helps to determine gut microbial species it gives insight into the development of probiotics[8]. Observation of *homosapiens* associated bacterial communities permits to establish ways to modulate them, to optimize human health[9]. Metagenome alludes to an assortment of complete hereditary material of a blended community of microbes, for example, natural metagenome, human metagenome, and so on it contains genomes of both cultivatable and uncultivatable microorganisms[10]. Metagenomics is an atomic apparatus used to investigate the blended genomic materials separated from ecological examples, which gives definite data of species variety and abundancy[11].

Breast cancer is the swelling structures in the breast cells. After skin cancer, breast cancer is the most broadly accepted malignancy analysed in ladies in the United States. Breast cancer can happen in both males and females, yet it's undeniably more normal in ladies[12]. Indications of mammary glands cancer may involve a ball or lumps in the breast, an alteration of breast shape, scratching of the skin, fluid coming out from areola, a recently reversed areola, or a red or textured fix of skin. In those with removed lay out of the

illness, there can be bone agony, swollen lymph hubs, windedness, or yellow skin[13]. Breast cancer most likely takes place in cells from the covering of milk conduits and the lobules that stockpile these pipes with milk. Cancers taking place in the pipes are referred to as ductal carcinomas, weather those take place from lobules are referred to as lobular carcinomas[14].

An advantage of these high throughput sequencing is that this experiment doesn't need cloning the DNA or prior information about genes or vectors for sequencing. This key feature of NGS technology has made RNA sequencing or Genomic sequencing easier, cheaper and faster with ability to identify mutation in single bases and prediction of novel genes[15]. Traditionally metagenomics sequencing was done using high-throughput sequencing of marker genes like rRNA genes or expressed genes using highly parallel 454 pyrosequencing[16]. Along with this technology other second-generation sequencing techniques like ion-torrent segueing, virtual terminator sequencing, illumina sequencing are widely used. These second-generation sequencing technologies for sequencing DNA produces shorter read fragments as compared to sangers sequencing. Ion Torrent and pyrosequencing typically produces ~400 bp reads fragments, Illumina technique produces 400-700bp reads. These reads sequencing are stored in from of Fastq files that have information about read sequence along with quality score of each base of all read sequences. These Fastq files are further used for metagenomics analysis like taxonomic classification, phylogenetic analysis, mutation analysis etc. [17]

The information created by metagenomics tests are both huge and innately loud, containing divided information addressing upwards of 10,000 species[18]. The sequencing of the cow rumen metagenome produced 279 gb, or 279 billion base sets of nucleotide succession data, while the human gut microbe's quality list recognized 3.3 million qualities amassed from 567.7 gigabases of grouping data[19]. Many tools have been created to integrate metadata and sequence data, permitting downstream comparative examination of different datasets using several ecological indices[20].

Comparative interpretation between metagenomes can give additional perception into the purpose of complex microbes and their functions in host[21]. Metagenomic analysis can be done by comparing one or more fastq files that is Pairwise or more than one fastq files that is multiple comparison. These comparisons of metagenomes between different samples files can be used for the comparing read sequences, alignment with reference microbial database, taxonomic classification and diversity, or functional analysis. Metagenomics data be inferred from complete microbial sequence, read sequences or small marker sequences like rRNA or 16s rRNA sequence. Variation in metagenome can be used to infer differentially expressed genes, pathway expression and function of each microbe in the samples. Conditions where there is low level of microbial population or diversity complete analysis can be done by the comparison with microbial databases[23].

Metagenomic sequencing is especially valuable in the analysis of viral communities and the best way to get to the hereditary range of the viral local area from an environmental example is through metagenomics[24]. As viruses come up short on a common all-inclusive phylogenetic marker (as 16S RNA for microscopic organisms and archaea, and 18S RNA for eukaryote), Viral metagenomes (likewise called viromes) should hence give increasingly more data about viral variety and development[25]. Advances in bioinformatics, refinements of DNA enhancement, and the expansion of computational force have significantly helped the investigation of DNA successions recuperated from ecological examples, permitting the transformation of shotgun sequencing to metagenomic samples[26].

MATERIALS AND METHODS

Metagenomic data retrieval: Metagenomic data of breast cancer was retrieved from ENA database (European Nucleotide Database) *https://www.ebi.ac.uk/ena/browser/home* with the accession number ERP107332 and project Id PRJEB25419 https://www.ebi.ac.uk/ena/browser/view/PRJEB25419. Four FASTQ format files were downloaded for two different breast cancer metagenomics sample as shown in table 1. In this experiment Illumina HiSeq 2000 paired end sequencing was done of Breast cancer metagenome. For current research these two samples files were selected for metagenomics analysis of breast cancer samples.

Table 1: Table of study accession, sample accession, experiment accession, run accession and sample name.

S. No	Sample Accession	Run Accession	Sample name	SRA files: FTP
1	SAMEA104667884	ERR2368791	Breast Cancer Metagenomics	1.ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR236/001/ERR2368791/ERR2368791_1.fastq.gz 2.ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR236/001/ERR2368791/ERR2368791_2.fastq.gz
2	SAMEA104667885	ERR2368792	Breast Cancer Metagenomics	1.ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR236/002/ERR2368792/ERR2368792_1.fastq.gz 2.ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR236/002/ERR2368792/ERR2368792_2.fastq.gz

Metagenomics tools used analysis: Galaxy server (https://usegalaxy.org/) and metagenomics tools at Galaxy server (https://metagenomics.usegalaxy.eu/) were used for metagenomics analysis and annotation of samples data. **FastQC tool[27]** was used for quality control analysis and identification of statistical parameters and reads information in each file. **Trimmomatic[28]** tool was used for trimming fastq file to filter the adapter sequence and to increase the quality of raw files so that it can be used for further analysis.

Kraken 2[29] tool was used for taxonomic analysis of all samples and bacterial community along with their abundance was predicted. Further **Convert Kraken[30]** tool was used to convert taxonomic file generated from Kraken 2 tool into Krona compatible file so that it can be used in **Krona tool[31]** for visualization of bacterial diversity in all the sample files taken into consideration.

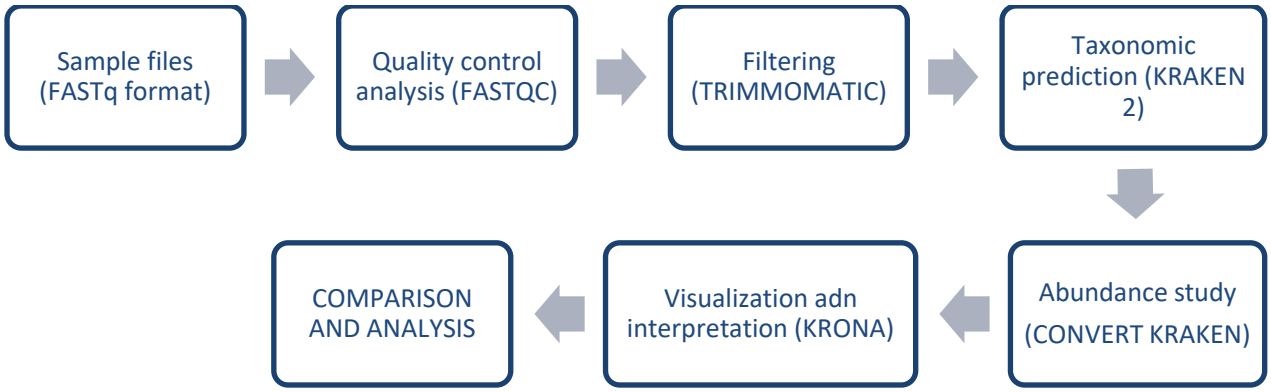

Figure 1: Steps used for metagenomics analysis and tools used for each step has been shown in brackets. These tools were used form metagenomics galaxy server.

Figure 1 shows the steps and tools used for metagenomics analysis. First step is **quality control**. This process is performed for the accuracy, completeness, relevancy, validity, timeliness of the data.

Second step is **trimming of data** it executes a wide range of useful trimming tasks for illumina paired end and single ended data. To perform the trimming process of data search Trimmomatic tool from the galaxy tool panel. FastQC output file was used to perform trimmomatic tool.

Third step is taxonomic prediction which was done using **Kraken 2** tool it is a taxonomic sequence classifier that gives taxonomic labels to small DNA reads. Kraken 2 tool analyses the k-mers within a read and aligns the taxonomic database with these query k-mers. Output of trimmomatic tool was used as an input in Kraken 2 tool. **Convert Kraken** tool was used to create taxonomic abundance file that can used further for visualization purpose**.** This tool converts Kraken metagenomic classifier to the full portrayal of NCBI taxonomy. Kraken 2 output file was used as input in Convert Kraken tool.

Lastly, **Krona** pie chart tool was used for visualization of metagenomic result in a zoomable pie-chart. Result of convert kraken tool was used as input file in krona tool.

RESULTS AND DISCUSSION

Quality control of all sample's files used for this research as mentioned in table 1 was done using FASTQC tool . Table 2 shows the comparison of total number sequence and GC content in all four samples that was generated from quality control analysis.

Table 2: Sample title, total sequence, and GC content for all samples.

S.N.	Sample Title	Total seq.	%GC
1	ERR2368791_1_fastq_gz	747728	42
2	ERR2368791_2_fastq_gz	7472728	42
3	ERR2368792_1_fastq_gz	7893195	43
4	ERR2368792_2_fastq_gz	7893195	43

FASTQC generates different statistical graphs that can be used to analyse the quality of each sample file. Quality control for all sample file was done but FASTQC result of first file that is ERR2368791_1_fastq_gz has been shown her demonstration purpose. Figure 2 shows the Per base sequence quality graph more the graph is in green region means sequence is of good quality.

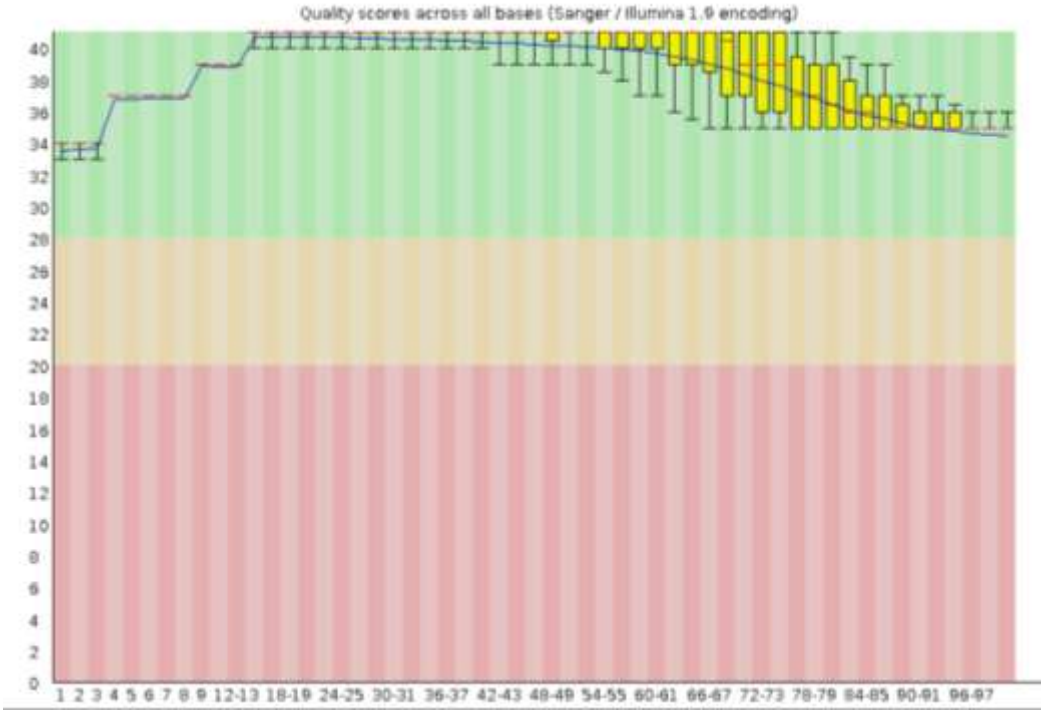

Figure 2: Per base sequence quality graph of ERR2368791_1_fastq_gz

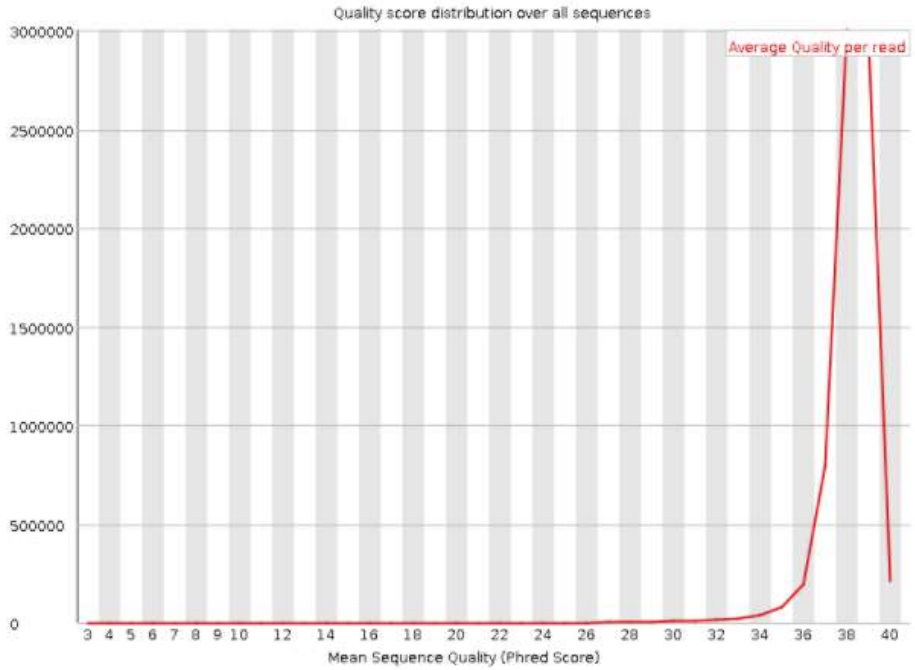

Figure 3: Per sequence quality scores

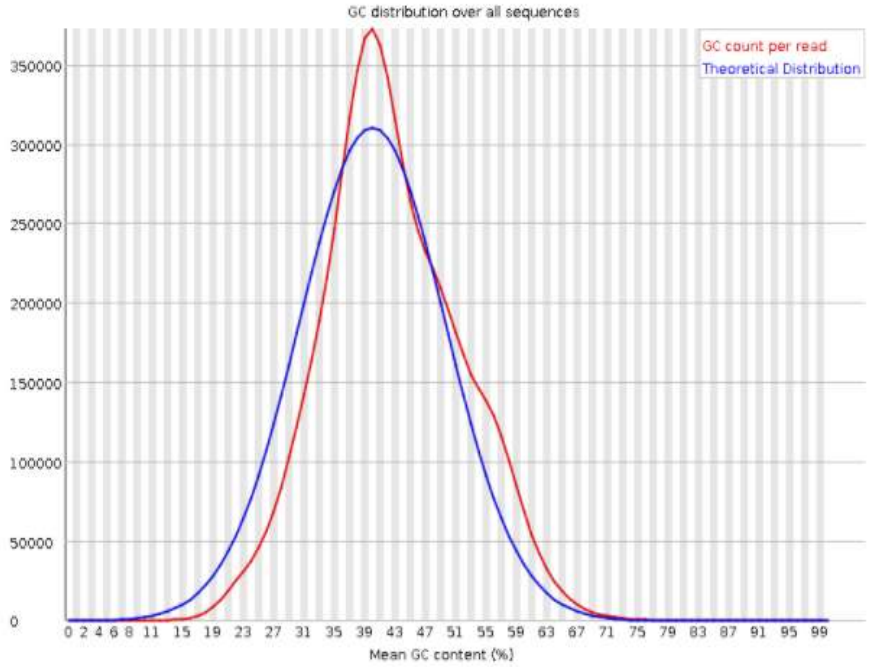

Figure 4: per sequence GC content

Figure 3 shows the Per sequence quality scores that is used to analyze the quality of reads sequence in each Fastq file. Average quality of reads in fastq file is calculated according to Phred score. Figure 4 shows the per sequence GC content in sample file (red line) in comparison with standard GC content (blue line). GC content graph of ERR2368791_1_fastq_gz shows that file can be further used for metagenomics analysis.

Kraken2 which is a taxonomic sequence classifier that gives taxonomic labels to small DNA reads. This tool aligns each k-mer in Kraken's genomic library to lowest common ancestor (LCA) during a taxonomic data analysis in comparison with all genomes that contain that k-mer[32]. The set of LCA taxa that correspond to the k-mers of reads in sample files are then analysed to make one taxonomic label for the reads[33]. It classifies each sequence in a single line of output this file have five tab-delimited fields that are "C/U" which is depicting the whether the sequence is classified or not, sequence ID, taxonomic ID, length of sequence in base pairs, space-delimited list showing LCA mapping of each k-mer in sequence.

Convert Kraken tool convert output of KRAKEN 2 into table format. This table includes Read name, Tax ID field, and appended 22 columns having taxonomic ranks from Super kingdom to Subspecies.

Visualization of taxonomy files by Krona tool

Krona visualization tool renders results of a metagenomic profiling as a zoomable pie chart. It allows hierarchical data, here taxonomic levels, to be explored with zooming, multi-layered pie charts.

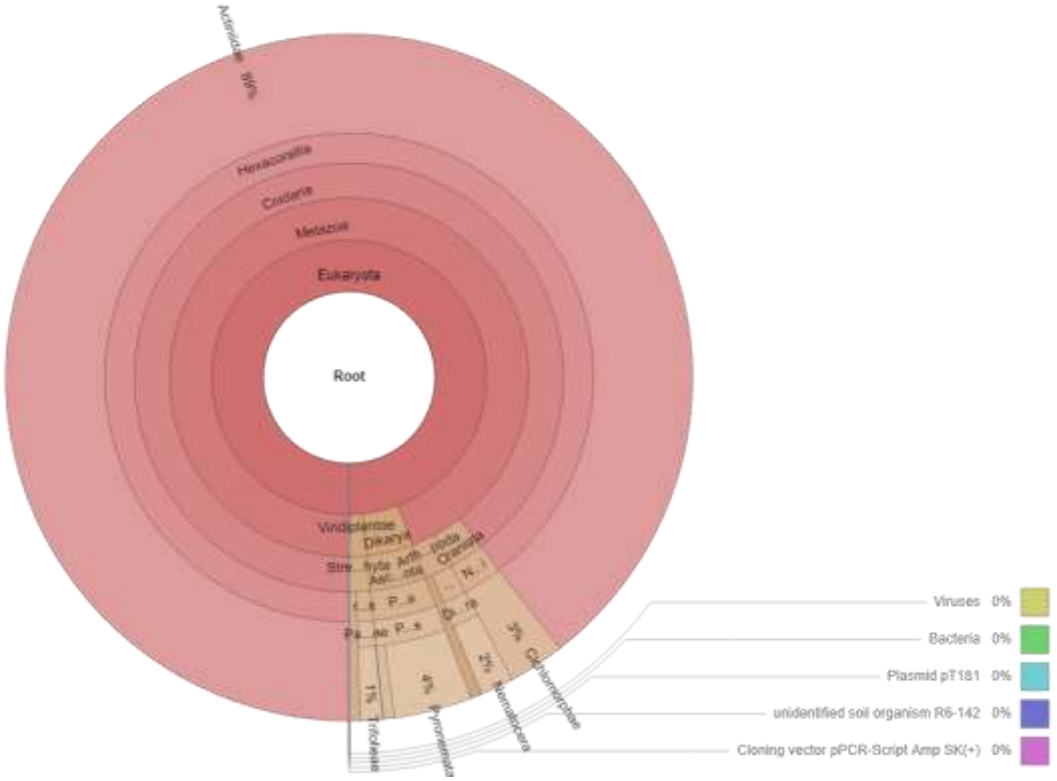

Figure 5: Result of KRONA pie chart for the paired data ERR2368791_1 and ERR2368791_2

Figure 5 shows all the bacteria present in ERR2368791_1 and ERR2368791_2 breast cancer metagenomics samples. A different colour in the figure shows the different bacteria and also their ratio in the sample. Species which are shown here with different ratios like 3% of *Cichlomorphae*, 2% of *Nematocera*, 4% of *pyronemataceae*, 2% of *streptophyta*, 89% of *Actinidiae*.

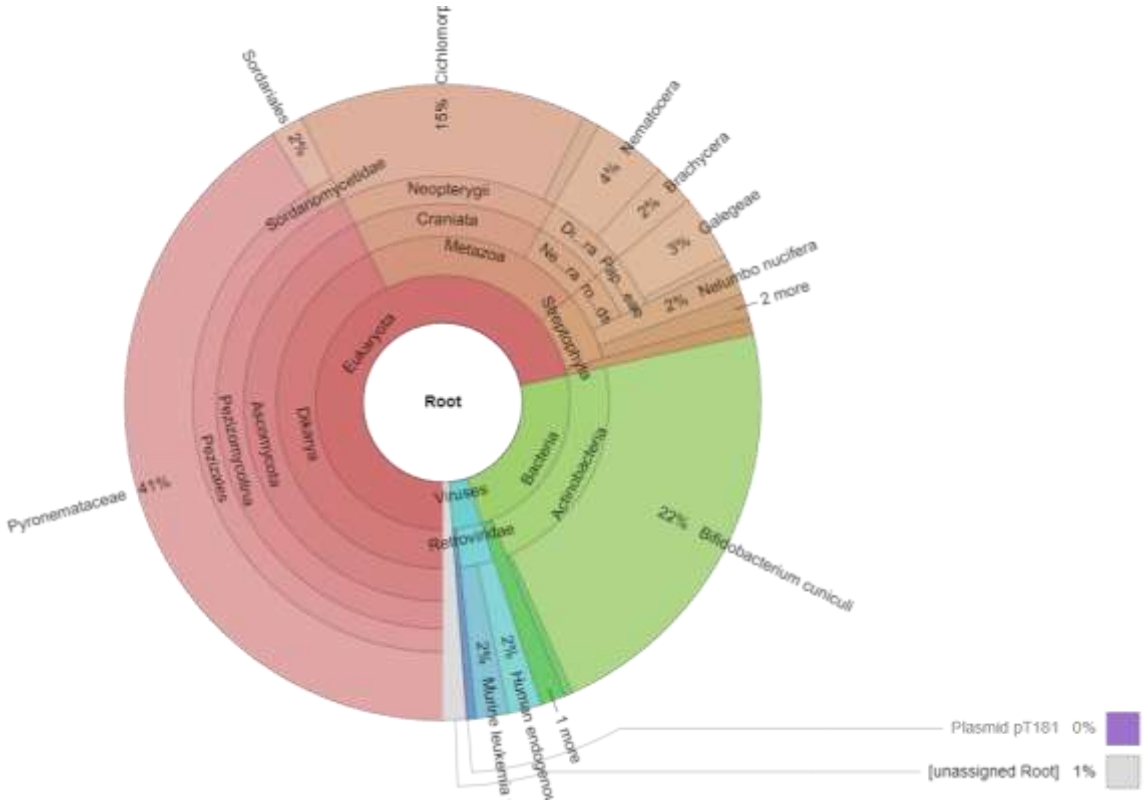

Figure 6: Result of KRONA pie chart for the paired data ERR2368792_1 and ERR2368792_2

Figure 6 shows all the bacteria present in the paired data of ERR2368792_1 and ERR2368792_2. Species which are shown in this figure are 4% of *Nematocera*, 2% of *Brachycer*a, 3% of *Galegeae*, 2% of *Nelumbo nucifera*, 22% of *Bifidobacterium cuniculi*, 2% of *Human endogenous virus*, 2% of *Murine leukemia*, 41% of *Pyronemetaceae*, 15% of *Cichlomorphae* and so on.

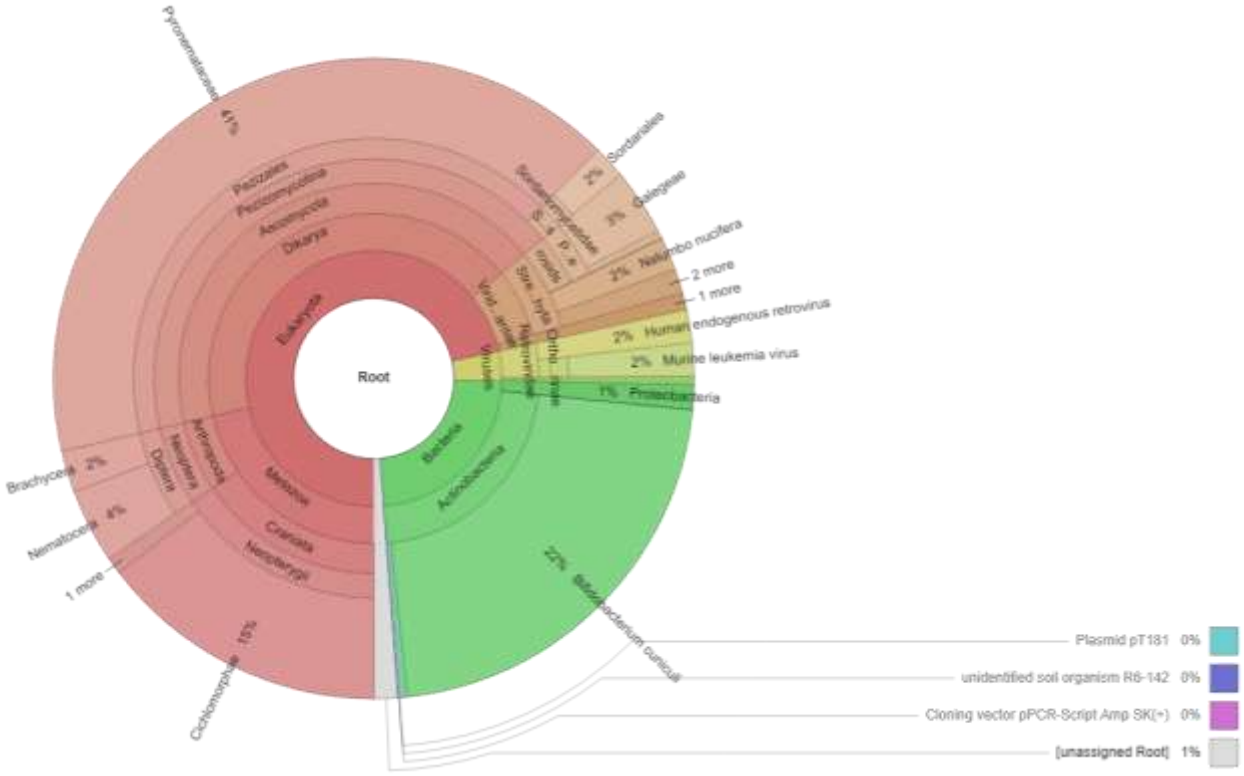

Figure 7: Combined result of KRONA pie chart for the all the paired ERR2368791_1, ERR2368791_2 and ERR2368792_1 and ERR2368792_2 for comparison

Figure 7 shows the combined taxonomic profile in all samples ERR2368791_1, ERR2368791_2 and ERR2368792_1 and ERR2368792_2. Comparative analysis between all sample files was done. Species which are shown in this figure 7 are 4% of *Nematocera*, 2% of *Brachycer*a, 3% of *Galegeae*, 2% of *Nelumbo nucifera*, 22% of *Bifidobacterium cuniculi*, 2% of *Human endogenous virus*, 2% of *Murine leukemia*, 41% of *Pyronemataceae*, 15% of *Cichlomorphae,* 1% of *Proteobacteri*a, 2% of *Sordariales*.

Comparative metagenomics analysis identifies the microbes present in breast cancer data these are *Actinobacteria, Bifidobacterium cuniculi, Murnine leukemia virus, Human endogenous retrovirus, Pezizomycotina*

Actinobacteria are a group of Gram-positive bacteria which has high guanine and cytosine content in their DNA, which can be terrestrial or aquatic. Although they are single celled like bacteria, they don't have particular cell wall, but they generate a mycelium which is nonseptate and more slender[34].

Bifidobacterium cuniculi is a class of gram-positive, nonmotile, frequently fanned anaerobic microscopic organisms. They are universal occupants of the gastrointestinal parcel, vagina and mouth (B. dentium) of well evolved creatures, including people. *Bifidobacteria* are one of the significant genera of microorganisms that make up the gastrointestinal parcel microbiota in vertebrates. Some *bifidobacteria* are utilized as probiotics[35].

Murnine leukemia virus (MLVs or MuLVs) are retroviruses named for their capacity to cause disease in murine (mouse) has. Some MLVs may taint different vertebrates. MLVs incorporate both exogenous and endogenous infections. Duplicating MLVs have a positive sense, single-abandoned RNA (ssRNA) genome that repeats through a DNA halfway by means of the cycle of opposite record[36].

Human endogenous retrovirus these are endogenous viral components in the genome that intently look like and can be gotten from retroviruses. They are plentiful in the genomes of jawed vertebrates, and they contain up to 5–8% of the human genome[37].

Pezizomycotina make up the vast majority of the Ascomycota growths and incorporate most lichenized parasites as well. *Pezizomycotina* contains the filamentous ascomycetes and is a region of the Ascomycota (growths that structure their spores in a sac-like ascus). It is pretty much inseparable from the more established taxon *Euascomycota*. These growths recreate by splitting instead of sprouting and this region incorporates practically all the ascus organisms that have fruiting bodies noticeable to the unaided eye[38].

CONCLUSION

Metagenomics study of breast cancer was done to predict the bacterial diversity in the samples using computational methods. Complete metagenomics analysis was done to stud the taxonomic and functional classification metagenomes in breast cancer data. Metagenomics analysis identifies the microbes in breast cancer data these are *Actinobacteria, Bifidobacterium cuniculi, Murnine leukemia virus, Human endogenous retrovirus, Pezizomycotina* . Current study signifies the important microorganisms that are present in breast cancer samples. These results can be used understanding the function and relation of these bacteria with the breast cancer development and progression. Detailed study is required to establish the association of microbes in breast cancer that can be further used for drug development.

Acknowledgment

We would like to acknowledge Amity Institute of biotechnology, Amity University Uttar Pradesh, Lucknow campus for providing us facilities to conducting this study. This research project is not funded by any specific grant from

funding agencies in the public, commercial, or non-profit sectors.

Conflict of Interest

Authors do not have any conflict of interests.

REFERENCES

1. Chiu CY, Miller SA. Clinical metagenomics. Nature Reviews Genetics. 2019 Jun; 20(6):341-55. https://doi.org/10.1038/s41576-019-0113-7

2. Zhang YZ, Shi M, Holmes EC. Using metagenomics to characterize an expanding virosphere. Cell. 2018 Mar 8; 172(6):1168-72. https://doi.org/10.1016/j.cell.2018.02.043

3. Simon HY, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. Cell. 2019 Aug 8; 178(4):779-94. https://doi.org/10.1016/j.cell.2019.07.010

4. Zhang YZ, Chen YM, Wang W, Qin XC, Holmes EC. Expanding the RNA virosphere by unbiased metagenomics. Annual Review of Virology. 2019 Sep 29; 6:119-39. https://doi.org/10.1146/annurev-virology-092818-015851

5. Fadiji AE, Babalola OO. Metagenomics methods for the study of plant-associated microbial communities: a review. Journal of microbiological methods. 2020 Mar 1; 170:105860. https://doi.org/10.1016/j.mimet.2020.105860

6. Kumar Awasthi M, Ravindran B, Sarsaiya S, Chen H, Wainaina S, Singh E, Liu T, Kumar S, Pandey A, Singh L, Zhang Z. Metagenomics for taxonomy profiling: tools and approaches. Bioengineered. 2020 Jan 1; 11(1):356-74. https://doi.org/10.1080/21655979.2020.1736238

7. Laudadio I, Fulci V, Stronati L, Carissimi C. Next-generation metagenomics: methodological challenges and opportunities. Omics: a journal of integrative biology. 2019 Jul 1; 23(7):327-33. https://doi.org/10.1089/omi.2019.0073

8. Sharma P, Kumar S, Pandey A. Bioremediated techniques for remediation of metal pollutants using metagenomics approaches: a review. Journal of Environmental Chemical Engineering. 2021 Aug 1; 9(4):105684. https://doi.org/10.1016/j.jece.2021.105684

9. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. Evaluating the information content of shallow shotgun metagenomics. Msystems. 2018 Nov 13; 3(6):e00069-18. https://doi.org/10.1128/mSystems.00069-18

10. Ramachandran PS, Wilson MR. Metagenomics for neurological infections-expanding our imagination. Nature Reviews Neurology. 2020 Oct; 16(10):547-56. https://doi.org/10.1038/s41582-020-0374-y

11. Yukgehnaish K, Kumar P, Sivachandran P, Marimuthu K, Arshad A, Paray BA, Arockiaraj J. Gut microbiota metagenomics in aquaculture: Factors influencing gut microbiome and its physiological role in fish. Reviews in Aquaculture. 2020 Aug; 12(3):1903-27. https://doi.org/10.1111/raq.12416

12. Zhu J, Liao M, Yao Z, Liang W, Li Q, Liu J, Yang H, Ji Y, Wei W, Tan A, Liang S. Breast cancer in postmenopausal women is associated with an altered gut metagenome. Microbiome. 2018 Dec; 6(1):1-3. https://doi.org/10.1186/s40168-018-0515-3

13. Huang T, Chen Y, Zhang J, He R, Qu D, Ye Q, Chen X. Rapid and accurate diagnosis of brain abscess caused by Nocardia asiatica with a combination of Ziehl-Neelsen staining and metagenomics next-generation sequencing. European Journal of Neurology. 2021 Jan; 28(1):355-7. https://doi.org/10.1111/ene.14533

14. Mahale P, Nomburg J, Song JY, Steinberg M, Starrett G, Boland J, Lynch CF, Chadburn A, Rubinstein PG, Hernandez BY, Weisenburger DD. Metagenomic analysis to identify novel infectious agents in systemic anaplastic large cell lymphoma. Infectious Agents and Cancer. 2021 Dec; 16(1):1-6. https://doi.org/10.1186/s13027-021-00404-0

15. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Elife. 2019 Sep 10; 8:e46923. https://doi.org/10.7554/eLife.46923

16. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. Journal of Infection. 2018 Mar 1; 76(3):225-40. https://doi.org/10.1016/j.jinf.2017.12.014

17. Clarke EL, Taylor LJ, Zhao C, Connell A, Lee JJ, Fett B, Bushman FD, Bittinger K. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. Microbiome. 2019 Dec; 7(1):1-3. https://doi.org/10.1186/s40168-019-0658-x

18. Brumfield KD, Huq A, Colwell RR, Olds JL, Leddy MB. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. PLoS One. 2020 Feb 13; 15(2):e0228899. https://doi.org/10.1371/journal.pone.0228899

19. Deng X, Achari A, Federman S, Yu G, Somasekar S, Bártolo I, Yagi S, Mbala-Kingebeni P, Kapetshi J, Ahuka-Mundeke S, Muyembe-Tamfum JJ. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. Nature microbiology. 2020 Mar; 5(3):443-54. https://doi.org/10.1038/s41564-019-0637-9

20. Latorre-Pérez A, Villalba-Bermell P, Pascual J, Vilanova C. Assembly methods for nanopore-based metagenomic sequencing: a comparative study. Scientific reports. 2020 Aug 12; 10(1):1-4. https://doi.org/10.1038/s41598-020-70491-3

21. Sanderson ND, Street TL, Foster D, Swann J, Atkins BL, Brent AJ, McNally MA, Oakley S, Taylor A, Peto TE, Crook DW. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. BMC genomics. 2018 Dec; 19(1):1-1. https://doi.org/10.1186/s12864-018-5094-y

22. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, Thielebein A, Hinzmann J, Oestereich L, Wozniak DM, Efthymiadis K. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science. 2019 Jan 4; 363(6422):74-7. https://doi.org/10.1126/science.aau9343

23. Cheng AP, Burnham P, Lee JR, Cheng MP, Suthanthiran M, Dadhania D, De Vlaminck I. A cell-free DNA metagenomic sequencing assay that integrates the host injury response to infection. Proceedings of the National Academy of Sciences. 2019 Sep 10; 116(37):18738-44. https://doi.org/10.1073/pnas.1906320116

24. Yadav R. Elucidating Structural, Functional and Phylogenetic Relationship of Large Envelope Protein of Hepatitis B V irus. Indian Journal of Science and Technology. 2022; 15(10):451-456. https://doi.org/10.17485/IJST/v15i10.491

25. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. Evaluating the information content of shallow shotgun metagenomics. Msystems. 2018 Nov 13; 3(6):e00069-18. https://doi.org/10.1128/mSystems.00069-18

26. Deep U, Yadav R. Motif prediction of abemaciclib in a breast cancer cell line using ChIP-Seq data analysis. BMRAT [Internet]. 2022; 9(3):4971-85. https://doi.org/10.15419/bmrat.v9i3.732

27. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FastQC: a quality control tool for high throughput sequence data [Online]. 2010. Avilable from: http://www. bioinformatics. babraham. ac. uk/projects/fastqc/. Accessed. 2020 Jan; 27.

28. Sewe SO, Silva G, Sicat P, Seal SE, Visendi P. Trimming and Validation of Illumina Short Reads Using Trimmomatic, Trinity Assembly, and Assessment of RNA-Seq Data. InPlant Bioinformatics 2022 (pp. 211-232). Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2067-0_11

29. Lu J, Salzberg SL. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. Microbiome. 2020 Dec; 8(1):1-1. https://doi.org/10.1186/s40168-020-00900-2

30. Brisse R, Boche S, Majorczyk F, Lalande JF. KRAKEN: A Knowledge-Based Recommender system for Analysts, to Kick Exploration up a

Notch. In14th International Conference on Security for Information Technology and Communications 2021 Nov 25.

31. Ahmad T, Gupta G, Sharma A, Kaur B, El-Sheikh MA, Alyemeni MN. Metagenomic analysis exploring taxonomic and functional diversity of bacterial communities of a Himalayan urban fresh water lake. PloS one. 2021 Mar 25; 16(3):e0248116. https://doi.org/10.1371/journal.pone.0248116

32. Garcia BJ, Simha R, Garvin M, Furches A, Jones P, Hyatt PD, Schadt C, Pelletier D, Jacobson D. Kraken2 Metagenomic Virus Database. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States). Oak Ridge Leadership Computing Facility (OLCF); 2020 Apr 23.

33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome biology. 2019 Dec; 20(1):1-3. https://doi.org/10.1186/s13059-019-1891-0

34. Binda C, Lopetuso LR, Rizzatti G, Gibiino G, Cennamo V, Gasbarrini A. Actinobacteria: a relevant minority for the maintenance of gut homeostasis. Digestive and Liver Disease. 2018 May 1; 50(5):421-8. https://doi.org/10.1016/j.dld.2018.02.012

35. Mattarelli P, Biavati B. Species in the genus Bifidobacterium. InThe Bifidobacteria and related organisms 2018 Jan 1 (pp. 9-48). Academic Press. https://doi.org/10.1016/B978-0-12-805060-6.00002-8

36. Qu K, Glass B, Doležal M, Schur FK, Murciano B, Rein A, Rumlová M, Ruml T, Kräusslich HG, Briggs JA. Structure and architecture of immature and mature murine leukemia virus capsids. Proceedings of the National Academy of Sciences. 2018 Dec 11; 115(50):E11751-60. https://doi.org/10.1073/pnas.1811580115

37. Küry P, Nath A, Créange A, Dolei A, Marche P, Gold J, Giovannoni G, Hartung HP, Perron H. Human endogenous retroviruses in neurological diseases. Trends in molecular medicine. 2018 Apr 1; 24(4):379-94. https://doi.org/10.1016/j.molmed.2018.02.007

38. Beimforde C, Schmidt AR, Rikkinen J, Mitchell JK. Sareomycetes cl. nov.: A new proposal for placement of the resinicolous genus Sarea (Ascomycota, Pezizomycotina). Fungal systematics and evolution. 2020 Dec 15; 6(1):25-37. https://doi.org/10.3114/fuse.2020.06.02