Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/15257
DOI URL: http://dx.doi.org/10.21474/IJAR01/15257

RESEARCH ARTICLE

METAGENOMICS: DNA SEQUENCING OF UNCULTURED MICROORGANISMS

Aarushi Ravi Shankar

Manuscript Info

Manuscript History
Received: 26 June 2022
Final Accepted: 28 July 2022
Published: August 2022

Abstract

Microbiology has completely transformed in the last 25 years and the realization that most microorganisms cannot be grown in a pure culture resulted in new methods of analyzing genomic diversity. One such method, Metagenomics, refers to the study of microorganisms on the basis of their genomic diversity, acquired directly from environmental samples. It facilitates the study of physiology and ecology of environmental microorganisms through a vast array of sequencing and functional techniques. Over the years, many sequencing techniques like Sangers Sequencing, Shotgun Sequencing, Whole Metagenomic Sequencing and 16s rRNA gene sequencing were used, which revolutionized the portrait of the microbial world. Microbiome profiles of these samples can be generated and analysed through various statistical tools like functional annotations, marker data profiling and biomarker detection which can give a description of community membership, and provide insight into the genetics and biochemistry of the members. The application of metagenomic sequence information will result in better culturing strategies and help link genomic analysis with pure culture studies.

Introduction:-

Microbes are everywhere. They take part in many biochemical cycles that convert key elements of life (carbon, nitrogen, oxygen, sulfur) to biologically assembled forms. Microbes function as multicellular organisms causing metabolic versatility which can regulate the majority of matter on Earth either as an individual or in a symbiotic association.

Microbes were first discovered in the late 18th century due to the invention of microscopes. Thus began the molecular biology and genomics revolutions to gain a thorough understanding of its underlying genetic basis. From the discovery of 'nuclein' (now known as DNA) by Friedrich Miescherin in 1871 to the 'Human Genome Project' in 1990, genomics became a field of increasing interest. Almost all knowledge of microbes is largely laboratory knowledge and is attained in unusual and unnatural circumstances of growing them optimally in artificial media in a pure culture. Although it greatly contributed to understanding genomics, on the other hand, it limits the ability to elucidate the dynamics of microbial communities.

It is estimated that only 0.1-1.0% of the living bacteria present in soil can be cultured in standard conditions. So, it became important to find non-traditional approaches through which microorganisms could be studied without the necessity of culturing. Therefore, the role of metagenomics outshines as an approach to investigate microbes from their natural environment without the need of culturing.

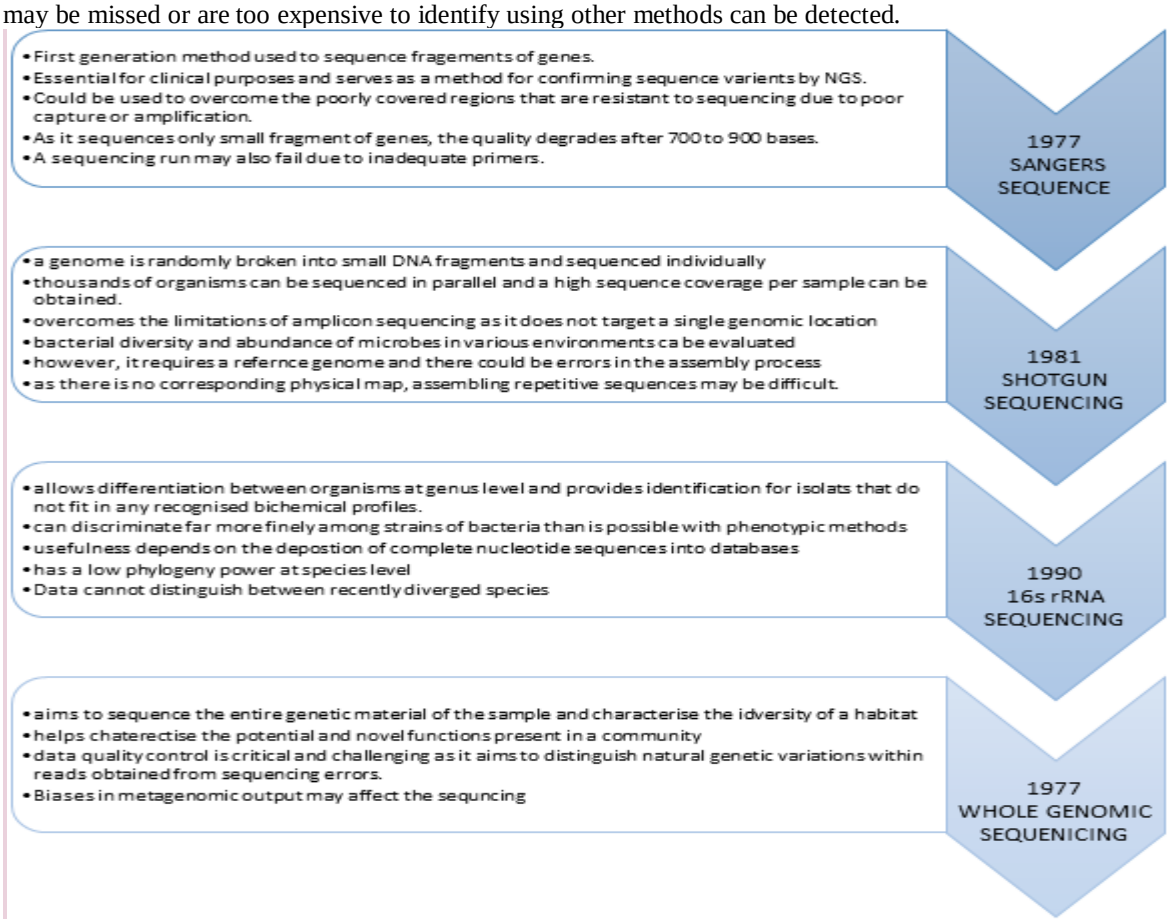
Corresponding Author:- Aarushi Ravi Shankar

What Is Metagenomics

Metagenomics refers to the study of microorganisms on the basis of their genomic diversity, acquired directly from environmental samples. It overcomes the unculturability of many microbes, resulting in advances in clinical and environmental microbiology. The traditional cultivation based methodologies, although proven successful in the identification and analysis of microorganisms, fails to recognise a population of highly similar yet distinguishable individual genotypes which cannot be cultured. Thus, metagenomics uses various bioinformatic tools and statistical pipelines to analyse the genetic information of a microbiome. It is a new and evolving field and the application of next generation technologies further increases the amount of sequence data observed through various databases like NCBI, HOMD, eHOMD, greengenes.

Many different techniques like Sangers sequencing, shotgun sequencing, 16s rRNA sequencing and whole metagenome sequencing are used in metagenomics.

Sanger sequencing, also known as the 'chain termination method', refers to the process of determining the nucleotide sequence of DNA. It is the oldest sequencing technique, developed in 1977 by Frederick Sanger, and is the basis for next generation sequencing. Unlike the PCR, which is used to amplify and duplicate DNA in its entirety, Sanger sequencing is used to generate every possible length of DNA up to the full length of the target DNA and so, dideoxynucleotides (ddNTP) are necessary. The three main steps to Sanger's sequencing include DNA sequence for chain terminator, size separation by DNA electrophoresis and gel analysis and determination on gene sequence. Although Sangers sequencing is a first generation method, it is essential for clinical purposes. However, sangers sequence can only sequence small strands of DNA, there could be errors in ddNTP, and the DNA sequence cannot be randomly cut. In 1981, Sanger developed the shotgun sequencing method, through which data from different sequences is compiled to form a larger genome. In this process, a genome is randomly broken into small fragments which are then sequenced individually. This type of metagenomic sequencing allows researchers to sequence thousands of organisms in parallel. As many samples could be combined in a single sequencing run, high sequence coverage per sample can be obtained and thus, low abundance members of the microbial community that may be missed or are too expensive to identify using other methods can be detected.



Sources of Samples in Metagenomics

Only about 1% of the microbes present in different environments can be cultured thus leaving a vast array of microorganisms unknown. This diversity can be estimated through culture independent approaches using environmental samples, some of them includes:

1. SOIL:

Soil holds a quarter of the total biodiversity on the planet and is one of the main reservoirs of taxonomy. It has a high level of heterogeneity and many microorganisms, including insects, mollusks, protozoa, algae, bacteria, fungi, etc. The diversity of microorganisms depends on a number of factors like organic matter content, soil mineral composition, pH and soil management practices. Bacteria in the soil play a key role in nutrient recycling (Nitrosomonas), development, soil structuring (rhizobia), regulation of diseases and the depollution of contaminated soils. They can form symbiosis with plants by interacting in the rhizosphere and can also be pathogenic, for both plants and animals. Studying such bacteria helps generate more information about biodiversity in soils.

2. WATER:

Many water bodies, including water springs, deep ocean vents, drinking water supply, groundwater, lakes etc provide an important ecosystem to microbial communities that remain largely unknown. Bacteria like E. coli, Campylobacter jejuni, Giardia lamblia, Legionella pneumophila, etc are all pathogenic bacteria found in water.

3. AIR:

Microbes in the air mainly consist of bacteria that are airborne, often attached to dust particles or droplet dispersion from coughs and sneezes or breezes over land or bodies of water. The spore formation in the aerosols that persist in the environment includes a number of bacteria like Bacillus, Micrococcus, Staphylococcus, etc. Microbes in the air are in low concentrations and can be collected by sampling large volumes of air through air handling units.

Different Techniques of Metagenomics:

Metagenomic DNA Sequencing:

Sequencing metagenomics refers to the study of microbial and functional diversity of an environmental sample using reference databases. It relies on the identification of comparison of sequences with previously annotated sequences uploaded to databases. The sequences obtained can be aligned to generate a phylogenetic tree in which the diversity of microorganisms in the sample is revealed. However, sequences with a low similarity to the given databases are classified as unidentified and thus, novel genes are often disregarded and entire gene sequences are rarely revealed. This leaves ambiguity around the gene products which prevents comprehensive biochemical and functional annotation.

Functional Metagenomics

Functional metagenomics refers to the study of the functions of encoded proteins in microorganisms by isolating DNA from microbial communities. In this process, DNA fragments are cloned, their genes are expressed in a surrogate host, and are screened for enzymatic activities. The construction and screening of metagenomic libraries and cosmid/fosmid libraries are preferred due to their large and consistent insert size and high cloning efficiency. To construct such libraries, DNA is first extracted from its environmental sample, size selected, and end repaired followed by ligation to a cos-based vector. This library is used to transform a bacterial host which can be manipulated in the lab. A number of novel enzymes could be identified on the basis of their DNA sequences. Information from the following approach annotates the genome from sequenced based analysis while functional metagenomics complements the prior ones.

16S rRNA Sequencing:

The 16s rRNA sequencing is a method used to identify, classify, and quantify microbes within complex biological mixtures (like environmental samples or gut samples). It was first proposed in 1977 by Carl Woese and George E. and the first microbial community study was conducted in 1990 by Giovannoni. The comparison of the 16s rRNA gene sequence allows differentiation between organisms at the genus level across all major phyla of bacteria. This gene is a highly conserved component of the transcriptional machinery of all microorganisms and so, is a suitable target gene for sequencing DNA in samples containing thousands of species. It is about 1500 bp long and thus, with sufficient polymorphisms of the 16s rRNA gene, can make distinguishing and statistically valid measurements. As the 16s rRNA has both conserved and variable regions, the PCR targets the conserved regions of 16s while sequencing the variable regions allows discrimination between specific different microorganisms. With the coupling

of 16s rRNA PCR and the next generation sequencing, samples from comprehensive time series can be analysed to quantify microbial community dynamics across many sites or produce detailed 3D maps of microbial communities. Moreover,

1. it's present in almost all bacteria,
2. its function has not changed overtime (thus random sequence changes are a more accurate measure of time)
3. The gene is large enough for informatic purposes

The 16S rRNA gene sequencing has been determined for a large number of strains. Genbank (the largest databank of nucleotide sequences) has over 20 million deposited sequences, out of which over 90000 are of 16S rRNA gene. These are the most common genetic markers to study bacterial phylogeny and taxonomy as the 16S rRNA gene sequence analysis can discriminate far more finely among strains of bacteria than is possible with phenotypic methods. It provides genus and species identification for isolates that don't fit in any recognised biochemical profiles (those strains that generate only a 'low likelihood' or 'acceptable' identification according to the databases.) The usefulness of 16s rRNA gene sequencing as a tool for identification is dependent on the deposition of complete unambiguous nucleotide sequences into databases and applying the correct 'label' to each sequence. Studies show that 16S rRNA sequencing provides genus identification in more than 90% of the cases but in none of the studies has the definition of species matched more than 99%, i.e. there has never been a perfect match established.

Although this type of sequencing is very useful in regards to bacterial classification, it has a low phylogenetic power at the species level and poor discriminatory power for some genera. Moreover, when the 16s rRNA sequences have very high similarity scores, the data cannot distinguish between recently diverged species and thus, cannot provide a definitive answer. Thus, to differentiate a species within a particular genus, a better gene may be found to identify the species.

Whole Metagenomic Sequencing:

Whole metagenomic sequencing, first employed in 1977 to sequence the phiX174 virus, refers to the sequencing of all genomes existing in an environmental sample to analyse the biodiversity and functional capabilities of the microbial community studied. It aims to sequence the entire genetic material of the sample and characterise the complete diversity of a habitat, including archaea, bacteria, eukarya, and viruses. This method of sequencing offers not only the possibility to characterise the genomic diversity of the community but also the potential and novel functions that are present in a community. Through correct sequencing depth, it is possible to assemble full genomes from metagenome data to gain insights into the genomic diversity of microbial communities and to obtain draft genomes of uncultured microorganisms. Data quality control is a critical and often challenging step in whole metagenomic sequencing as it aims to distinguish natural genetic variations within reads obtained from sequencing errors.

This approach allows taxonomy to be assigned at the species and strain levels. Whole metagenomic sequencing is easily achieved in habitats like the human skin, or the lungs, characterised by low biomass and high host DNA contamination although better results are generated for simpler organisms, such as bacteria. However, biases in metagenomic output may affect whole metagenomic sequencing. It is less affected by biases associated with the PCR necessary for amplifying the marker genes. Rather it is affected by biases in metagenomic output due to the whole metagenomic protocol.

Marker Data Profiling

The process by which microbiome profiles can be generated through quality filtered data is known as marker data profiling. All raw sequencing data is processed before microbiome profiles are generated and biologically relevant information is obtained. Evaluation of data quality and filtering of low quality data is carried out for analysing taxonomic microbiome profiles based on the relative abundance of taxa. The functional capacity of the genetic expression individually amongst the community explains the relevance amongst various functional pathways. The feature of community profiling which is majorly studied includes alpha-beta diversity and core microbiome analysis and characterisation through LEfSe and random forests.

Alpha diversity refers to the variability of species in a sample. It measures the richness, the number of species in an environment, and evenness, the homogeneity in abundance of different species in the sample. Some indices include the Chao1 index and Shannon index. Beta diversity measures the differences in microbiome diversity in different samples. For example, Bray-Curtis, UniFrac, weighted Unifrac distance, Jaccard index and the aitchison distance.

PCoA, standing for principal coordinate analysis, refers to the diagrammatic representation of microbial communities in a sample. Points that are closer together have microbial communities similar in sequence composition. Core microbiome analysis helps to distinguish between the different types of functions carried out by host associated microbes.

LEfSe is an algorithm for high dimensional biomarker discovery. It identifies genomic features, characterising the differences between two or more biological conditions.

Functional Annotations
Functional annotation refers to the process of collecting information about and describing a gene's biological identity i.e its various aliases, molecular function, biological roles, subcellular location, etc. It is the method of attaching biological information to sequences of genes or proteins.

The most basic level of functional annotation is using sequence alignment tools like BLAST for finding similarities, and then annotating genes or proteins based on that. Due to technological advancements, additional information of biological functions is added to the annotation systems. This allows hand operated annotation to distinguish genes that have the same annotation. Thus, computational annotation methods to characterise genes from their sequence are increasingly important.

Metadata Correlation
Information associated with DNA sequences is known as metadata. Metadata correlation refers to the analytical approach of identifying patterns in ecological properties of microbial communities. This could be done by sequencing community structure and function and defining the physical, chemical and biological parameters of the ecosystem.

Biomarker Detection
A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological and pathogenic processes. It is a measurable diagnostic indicator for assessing the risk or presence of disease. The discovery of biomarkers has proven to be one of the most broadly applicable and successful means of translating molecular and genomic data into clinical practice. The detection of biomarkers in bodily fluids such as blood and urine is a powerful tool for early diagnosis and treatment of diseases. Biomarker detection can be formulated as determining the most revealing features that can differentiate multiple sets of samples. It is challenging for metagenomic biomarkers to identify the most biologically informative features differentiating two or more phenotypes. Powerful statistical tools are required to ensure the reproducibility of conclusions drawn from metagenomic data as they are crucial for the clinical application of biological findings.

Bibliography:-

[1] Alex Nord, Stephen J. Salipante, Colin Pritchard, Shashikant Kulkarni (E), John Pfeifer (E), Chapter 11 - Copy Number Variant Detection Using Next-Generation Sequencing,Clinical Genomics, Academic Press 2015, Pages 165-187, ISBN 9780124047488, https://doi.org/10.1016/B978-0-12-404748-8.00011-3. (https://www.sciencedirect.com/science/article/pii/B9780124047488000113)
[2] Alice Risely, Applying the core microbiome to understand host–microbe systems, Journal of Animal Ecology, volume 89, issue 7, 2020, https://doi.org/10.1111/1365-2656.13229,
[3] Alshawaqfeh M, Bashaireh A, Serpedin E, Suchodolski J. Reliable Biomarker discovery from Metagenomic data via RegLRSD algorithm. BMC Bioinformatics. 2017;18(1):328. Published 2017 Jul 10. doi:10.1186/s12859-017-1738-1
[4] Berardini TZ, Mundodi S, Reiser L, et al. Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol. 2004;135(2):745-755. doi:10.1104/pp.104.040071
[4] C. A. Petti, C. R. Polage, P. Schreckenberger, The Role of 16S rRNA Gene Sequencing in Identification of Microorganisms Misidentified by Conventional Methods, DOI:https://doi.org/10.1128/JCM.43.12.6123-6125.2005,
[5] Calle M. L. (2019). Statistical Analysis of Metagenomics Data. Genomics & informatics, 17(1), e6. https://doi.org/10.5808/GI.2019.17.1.e6
[6] Chistoserdovai L. Functional metagenomics: recent advances and future challenges. Biotechnol Genet Eng Rev. 2010;26:335-52. PMID: 21415887.

[7] Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17(4):840-862. doi:10.1128/CMR.17.4.840-862.2004
[8] Damien Ennis, Sergiu Dascalu, Frederick C. Harris Jr.. (2016) Leveraging Clustering Techniques to Facilitate Metagenomic Analysis. Intelligent Automation & Soft Computing 22:1, pages 153-165.
[9] Diseases,Genomics, Circuits, and Pathways in Clinical Neuropsychiatry, Academic Press 2016, Pages 179-199, ISBN 9780128001059, https://doi.org/10.1016/B978-0-12-800105-9.00012-3. (https://www.sciencedirect.com/science/article/pii/B9780128001059000123)
[10] Durmaz, A. A., Karaca, E., Demkow, U., Toruner, G., Schoumans, J., & Cogulu, O. (2015). Evolution of genetic techniques: past, present, and beyond. BioMed research international, 2015, 461524. https://doi.org/10.1155/2015/461524
[11] Escobar-Zepeda Alejandra, Vera-Ponce de León Arturo, Sanchez-Flores Alejandro, The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics, Frontiers in Genetics, volume 6 2015, ISSN 1664-8021 10.3389/fgene.2015.00348 https://www.frontiersin.org/article/10.3389/fgene.2015.00348
[12] Evolution of Infectious Diseases (Second Edition), Elsevier 2017, Pages 553-572, ISBN 9780127999425, https://doi.org/10.1016/B978-0-12-799942-5.00024-X.
(https://www.sciencedirect.com/science/article/pii/B978012799942500024X)
[13] Genetics and Evolution of Infectious Diseases (Second Edition), Elsevier 2017, Pages 421-434, ISBN 9780127999425, https://doi.org/10.1016/B978-0-12-799942-5.00018-4. (https://www.sciencedirect.com/science/article/pii/B9780127999425000184)
[14] George E. Fox, Jeffrey D. Wisotzkey, Peter Jurtshuk JR., How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity , https://doi.org/10.1099/00207713-42-1-166, https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-42-1-166
[15] Gilbert J. (2013) Metagenomics, Metadata, and Meta-analysis. In: Nelson K. (eds) Encyclopedia of Metagenomics. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6418-1_17-1
[16] Goodrich JK, Di Rienzi SC, Poole AC, et al. Conducting a microbiome study. Cell. 2014;158(2):250-262. doi:10.1016/j.cell.2014.06.037
[17] Griesemer, M., Kimbrel, J.A., Zhou, C.E. et al. Combining multiple functional annotation tools increases coverage of metabolic annotation. BMC Genomics 19, 948 (2018). https://doi.org/10.1186/s12864-018-5221-9
[18] Haley Abel, John Pfeifer, Eric Duncavage, Shashikant Kulkarni (E), John Pfeifer (E), Chapter 10 - Translocation Detection Using Next-Generation Sequencing, Clinical Genomics, Academic Press 2015, Pages 151-164, ISBN 9780124047488, https://doi.org/10.1016/B978-0-12-404748-8.00010-1. (https://www.sciencedirect.com/science/article/pii/B9780124047488000101)
[19] Ian S. Hagemann, hashikant Kulkarni (E), John Pfeifer (E), Chapter 1 - Overview of Technical Aspects and Chemistries of Next-Generation Sequencing, Clinical Genomics, Academic Press 2015,Pages 3-19, ISBN 9780124047488, https://doi.org/10.1016/B978-0-12-404748-8.00001-0.
(https://www.sciencedirect.com/science/article/pii/B9780124047488000010)
[20] Izenman, A.J. (2013). Linear Discriminant Analysis. In: Modern Multivariate Statistical Techniques. Springer Texts in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-78189-1_8
[21] J.H. Leamon, J.M. Rothberg, Moselio Schaechter(E), DNA Sequencing and Genomics,Encyclopedia of Microbiology (Third Edition), Academic Press 2009, Pages 148-161, ISBN 9780123739445, https://doi.org/10.1016/B978-012373944-5.00024-9.
(https://www.sciencedirect.com/science/article/pii/B9780123739445000249)
[22] Lam Kathy N., Cheng Jiujun, Engel Katja, Neufeld Josh D., Charles Trevor C, Current and future resources for functional metagenomics, Frontiers in Microbiology,2015, ISSN 1664-302X, 10.3389/fmicb.2015.01196, https://www.frontiersin.org/article/10.3389/fmicb.2015.01196
[23] M.M.H. Abdelbary, P. Basset, D.S. Blanc, E.J. Feil, Michel Tibayrenc (E), 24 - The Evolution and Dynamics of Methicillin-Resistant Staphylococcus aureus, Genetics and
[24] Mukesh Kumar Malviya, Manoj Kumar Solanki, Chang-Ning Li, Reemon Htun, Rajesh Kumar Singh, Pratiksha Singh, Yang-Rui Li, Manoj Kumar Solanki (E), Prem Lal Kashyap (E), Rizwan Ali Ansari, Baby Kumari(E), Microbiomes and Plant Health, Academic Press 2021, Pages 225-242, ISBN 9780128197158, https://doi.org/10.1016/B978-0-12-819715-8.00007-0.
(https://www.sciencedirect.com/science/article/pii/B9780128197158000070)
[25] National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications. The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. Washington (DC): National

Academies Press (US); 2007. 1, Why Metagenomics? Available from: https://www.ncbi.nlm.nih.gov/books/NBK54011/
[26] Patel, J.B. 16S rRNA Gene Sequencing for Bacterial Pathogen Identification in the Clinical Laboratory. Molecular Diagnosis 6, 313–321 (2001). https://doi.org/10.1007/BF03262067
[27] Pedron, R., Esposito, A., Bianconi, I. et al. Genomic and metagenomic insights into the microbial community of a thermal spring. Microbiome 7, 8 (2019). https://doi.org/10.1186/s40168-019-0625-6
[28] Prakash T, Taylor TD. Functional assignment of metagenomic data: challenges and applications. Brief Bioinform. 2012;13(6):711-727. doi:10.1093/bib/bbs033
[29] Prayogo, F.A., Budiharjo, A., Kusumaningrum, H.P. et al. Metagenomic applications in exploration and development of novel enzymes from nature: a review. J Genet Eng Biotechnol 18, 39 (2020). https://doi.org/10.1186/s43141-020-00043-9
[30] Quince, C., Walker, A., Simpson, J. et al. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35, 833–844 (2017). https://doi.org/10.1038/nbt.3935
[31] Rodrigues RR, Rodgers NC, Wu X, Williams MA. COREMIC: a web-tool to search for a niche associated CORE MICrobiome. PeerJ. 2018;6:e4395. Published 2018 Feb 15. doi:10.7717/peerj.4395
[32] Sabale, S. N., Suryawanshi, P. P.,& P.U., K. (2019). Soil Metagenomics: Concepts and Applications. In (Ed.), Metagenomics - Basics, Methods and Applications. IntechOpen. https://doi.org/10.5772/intechopen.88958
[33] Saumya S. Jamuar, Alissa M. D'Gama, Christopher A. Walsh, Thomas Lehner (E), Bruce L. Miller (E), Matthew W. State (E), Chapter 12 - Somatic Mosaicism and Neurological
[34] Segata, N., Izard, J., Waldron, L. et al. Metagenomic biomarker discovery and explanation. Genome Biol 12, R60 (2011). https://doi.org/10.1186/gb-2011-12-6-r60
[35] Sharpton Thomas J, An introduction to the analysis of shotgun metagenomic data, Frontiers in Plant Science, VOLUME 5 2014, https://www.frontiersin.org/article/10.3389/fpls.2014.00209, DOI 10.3389/fpls.2014.00209, ISSN 1664-462X
[36] Sunil Nagpal, Rashmi Singh, Deepak Yadav, Sharmila S Mande, MetagenoNets: comprehensive inference and meta-insights for microbial correlation networks, Nucleic Acids Research, Volume 48, Issue W1, 02 July 2020, Pages W572–W579, https://doi.org/10.1093/nar/gkaa254
[37] Ten Hoopen P, Finn RD, Bongo LA, et al. The metagenomic data life-cycle: standards and best practices. Gigascience. 2017;6(8):1-11. doi:10.1093/gigascience/gix047
[38] Thomas, T., Gilbert, J. & Meyer, F. Metagenomics - a guide from sampling to data analysis. Microb Informatics Exp 2, 3 (2012). https://doi.org/10.1186/2042-5783-2-3
[39] Tremblay Julien, Singh Kanwar, Fern Alison, Kirton Edward, He Shaomei, Woyke Tanja, Lee Janey, Chen Feng, Dangl Jeffery, Tringe Susannah, Primer and platform effects on 16S rRNA tag sequencing , Frontiers in Microbiology, VOLUME 6 2015, DOI 10.3389/fmicb.2015.00771, ISSN 1664-302X
https://www.frontiersin.org/article/10.3389/fmicb.2015.00771,
[40] Tringe, S., Rubin, E. Metagenomics: DNA sequencing of environmental samples. Nat Rev Genet 6, 805–814 (2005). https://doi.org/10.1038/nrg1709
[41] Tringe SG, Zhang T, Liu X, et al. The airborne metagenome in an indoor urban environment. PLoS One. 2008;3(4):e1862. Published 2008 Apr 2. doi:10.1371/journal.pone.0001862https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2270337/
[42] Wagg, C., Bender, S. F., Widmer, F., & van der Heijden, M. G. (2014). Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proceedings of the National Academy of Sciences of the United States of America, 111(14), 5266–5270. https://doi.org/10.1073/pnas.1320054111
[43] W. Florian Fricke, Thomas A. Cebula, Jacques Ravel, Bruce Budowle (E), Steven E. Schutzer (E), Roger G. Breeze (E), Paul S. Keim (E), Stephen A. Morse (E), Chapter 28- Genomics, Microbial Forensics, second edition, Pages 479-492, ISBN 9780123820068, https://doi.org/10.1016/B978-0-12-382006-8.00028-1.
(https://www.sciencedirect.com/science/article/pii/B9780123820068000281)
[44] Wooley JC, Ye Y. Metagenomics: Facts and Artefacts, and Computational Challenges*. J Comput Sci Technol. 2009;25(1):71-81. doi:10.1007/s11390-010-9306-4
[44] Xanthopoulos, P., Pardalos, P.M., Trafalis, T.B. (2013). Linear Discriminant Analysis. In: Robust Data Mining. SpringerBriefs in Optimization. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9878-1_4.