

# Methods and Tools Used in Metagenomics and its Applications in Biotechnology

Safaa M. Ali\*

**Nucleic Acid Research Department, Genetic Engineering and Biotechnology Research Institute (GEBRI), City of Scientific Research and Technological Applications, Alexandria, Egypt  
Universities and Research Centers District, P.O. Box: 21934 New Borg El-Arab, Alexandria, Egypt**

(Received 8 May, 2021; Accepted 27 June, 2021)

## ABSTRACT

Metagenomics is a collection of techniques and methods for analyzing genetic material (genes) found in the environment lacking of the microbial elements that live there. This field is regarded as a combined study of Microbiology and Genomics. This is as opposed to traditional microbiology approaches, which rely on the examination of the component to comprehend the whole. Metagenomics goes the other direction, focusing on recognizing the environmental microbial potential and determining its forms by distinguishing strains. It can also be used to extract specific proteins directly from the natural world. Microbiologists were able to gain a deeper comprehension of how microbial communities function and how species communicate with one another thanks to this genetic knowledge. It would also be beneficial to both the economy and the environment. It also helps to improve food security by improving sustainable agricultural practices and ensuring the quality and provision of ecosystem services, in addition providing insight into the environment of microorganisms that are beneficial or threaten the growth of agricultural crops. Genomic data is also being used to see if there are any non-arable microbial representations that could be used as alternative energy sources to address pressing environmental issues such as global warming and the growth of renewable energy sources like hydrogen and methane. Metagenomics can also be applied to diagnose infectious diseases, classify intestinal microbes, develop biofuels, and treat the climate. Environmental genomes have the potential to advance understanding in a variety of fields (medicine, engineering, agriculture, sustainability, and environmental science).

**Key word:** Metagenomics, Function, Environmental, Medicine, Engineering, Agriculture

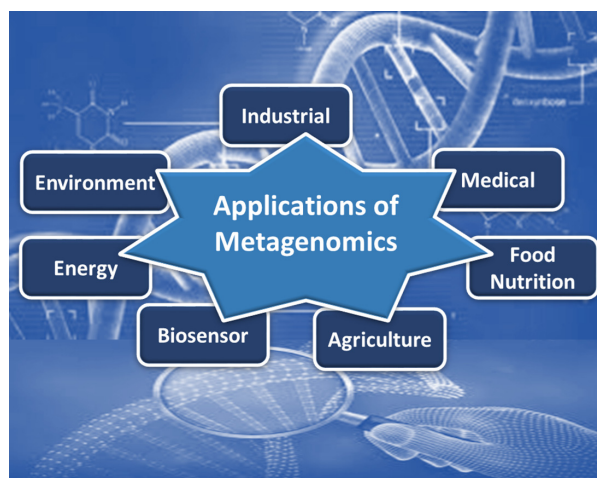
## Introduction

### History of Metagenomics

Metagenomics refers to a society that isn't focused on genomics. Metagenomics has become a common method of educating viruses and prokaryotes in the location by analyzing DNA from ecological samples. This data is descriptive of all types of microbial DNA. Conventional genomics started with the cultivation of identifiable cells as the source of DNA for

analysis. The first study, available by Pace and colleagues in 1991, on the isolation and cloning of DNA bulk from an environmentally friendly sample (Schmidt *et al.*, 1991).

Metagenomics has developed modern Biomedical Science and Industry research. A significant development of the late 20<sup>th</sup> and early 21<sup>st</sup> centuries was the detection of new microbes and numerous microbial products bypassing culture methods. Modern tools and techniques, such as



**Fig. 1.** Various aspects of metagenomics applications in different fields of biological science

bioinformatics, NGS technology, and methods of data analysis, prove to be facilitators of the field of trending research. Biological data is increasing its size continuously; thus researchers have a golden opportunity to more effectively solve or retrieve the hidden information present in assembled or unassembled reads using modern analytical tools. The bacteria-specific 16S rRNA sequence was the main focus of early metagenomics in order to overcome this hurdle in the field of genomics, since this sequence was relatively short, often preserved within a species and generally different between species (Chakravorty *et al.*, 2007).

Metagenomics has increased in popularity in latest years as a novel detection field and strategy for avoiding the genomic diversity of most microorganisms and their unculturability, two of the most significant hurdles to progress in environmental and clinical microbiology (Gu *et al.*, 2019). Population genomics research at the individual level of microbes is known as metagenomics, referring to the idea that a set of genes derivative from a certain environment is analyzed using a method similar to that used to analyze a genome, providing a lens control in the microorganisms community with the prospective to transform clinical sciences (Duan *et al.*, 2021).

### Types of metagenomics

Major Metagenomics themes are functional metagenomics that proceeds the whole environmental DNA; marker metagenomics that studies structure of microbial community by directing the actual

conserved 16S rRNA gene; and identification of novel enzymes. Targeted metagenomics is beneficial in recognizing the variety of only gene of interest, on the other hand it is fractional by the primers category of PCR used for analysis (Parada *et al.*, 2016; Tas *et al.*, 2021).

### Functional metagenomics

Functional metagenomics was important in appreciative the microbiological population's role in both geochemical cycles and microbial ecology around the world. As such, essential to recognize new enzymes from the environmental sample (Uchiyama and Miyazaki, 2009). Thus the metagenomics of function played a major role in the databank of proteins and nucleic acids by adding new functional annotation. This strategy has however a major drawback with a low hit rate of successful copies, poor performance and time consuming screening (Hosokawa *et al.*, 2015).

### Shotgun metagenomics

Shotgun metagenomics method used for generate whole or almost complete pathogen genome associations and to detect of type of pathogen (Seth-Smith *et al.*, 2013; Nayfach *et al.*, 2021). These results arrange for an estimation of microbial genotypes and microbial phenotypes by defining whether antimicrobial resistance and epidemic dynamics are present or not (Bertelli and Greub, 2013; Piombo *et al.*, 2021).

### Currently metagenomics

Currently metagenomics is a powerful technique to have industrial applications in identification of novel biocatalysts, discovering novel antibiotics, and bioremediation. The application of metagenomics is increasing rapidly, and these are being listed below. Metagenomics' earliest applications were in the finding of microbial enzymes and antibiotics (Gillespie *et al.*, 2002) and its benefits are existence predictable every day. Metagenomic library screening has improved a view of microbial ecology and provided us new insight into gene cataloging as it has opened up avenues for gaining access to the vast the microbiological universe's diversity and huge reservoirs of genes and researching their importance in different metabolic and immune response processes (Pindjakova *et al.*, 2017), overcoming the drawbacks of culture-based methods. A recent study by Yutin *et al.*, 2018 documented the discovery

of the bacteriophage family which includes crass-phage based viruses. Most of these viruses are associated with human intestinal bacteria belonging to phylum bacterioidetes that includes some of the most common bacteria in the human intestine and in other habitats. The same community has carried out comprehensive silico research of genomic and metagenomic datasets to evaluate and forecast phage protein functions (Yutin *et al.*, 2018). Some researchers are trying to introduce the concept to the general population. Hackuarium's neighborhood library, a non-profit organization, provided people with the intended metagenomic profile of 39 bottled beers from 5 countries based on the fungal species sequencing of internal transcribed spacers (ITS) (Sobel *et al.*, 2017). Metagenomics until that time used to differentiate microbiome between cultural groups (Chen *et al.*, 2016). Metagenomics has shown to be a very useful tool in linking microbiome and cancer (Mukherjee *et al.*, 2017).

### Metagenomics Techniques and Tools

Experimental design act as a significant role in obtaining high quality data, reliable and precise. Researchers in the subject of metagenomics need to accuracy of methods, sequencing cost-effectiveness and concentrate on data replication numbers used to conduct metagenomics data analysis. The sample position should be clearly defined relating to certain parameters (Cooke *et al.*, 2017).

#### a. Sampling

Samples are obtained from various sources (soil, air, water, biopsy, plants, etc.) known as sampling. Sampling relies on the data's availability obtained from metagenomics (Thomas *et al.*, 2012). The study will represent the entire population when defining biodiversity (Wooley *et al.*, 2010) and should also represent habitat. When collecting the samples, the time (day, date and year of collection), the samples quantity and the amount of samples needed to define the samples should be known. Sample fractionation could be prepared of lysing the cell to extricate the genomic DNA. So, it's exceptionally critical to deactivate or inactivate the nucleases by including solid denaturing operators to keep our genomic DNA secure (Yatsunenکو *et al.*, 2012). Cell lysis can be performed by warm, chemical, mechanical, and enzymatic strategies (Felczykowska *et al.*, 2015).

#### b. DNA Extraction

DNA extraction may be a vital step for analyzing the genome of uncultivable organism. So, it's exceptionally critical to choose a subjective and quantitative DNA extraction strategy for getting tall abdicate and great quality of DNA (Felczykowska *et al.*, 2015). The test contains DNA in different bundles like infection particles, eukaryotic DNA, and prokaryotic DNA counting free DNA. This could be suspended in fluid, bound to strong, or caught within the biofilm or tissue. The bacterial differing qualities of DNA recuperated by roundabout implies was unmistakably higher (Berry *et al.*, 2003).

#### c. DNA Sequencing

By and large, there are three sorts of sequencing strategies, viz., amplicon sequencing, shotgun sequencing, and metagenomics sequencing. Amplicon sequencing is utilized for characterization of microbiota differences and it is the foremost commonly utilized strategy. It targets the little subunit of ribosomal RNA (16s) locus, which acts as marker which gives data around phylogeny and scientific categorization (Hugenholtz and Pace, 1996). This sequencing strategy is utilized to characterize an expansive range of microbial differing qualities within the human intestine (Yatsunenکو *et al.*, 2012), Arabidopsis thaliana roots (Lundberg *et al.*, 2012), sea warm vents (McCliment *et al.*, 2006), hot springs (Bowen *et al.*, 2013), and Antarctic spring of gushing lava mineral soils (Soo *et al.*, 2009). Due to certain limitations of amplicon sequencing, shotgun sequencing came within the picture. Shotgun sequencing has capability to overcome the confinements of past approach. This approach depends on extricating DNA from cells in community and dividing it into minor parts (i.e., peruses) that are utilized to adjust against the known genome and 16S rRNA. Thus, it gives opportunity to investigate microbiota community with two viewpoints (Sharpton, 2014). Shotgun sequencing has moreover impediment like huge information dealing with, peruses may not display within the entirety genome, and some of the time two peruses of the same quality don't cover (Sharpton *et al.*, 2011). Headway in shotgun sequencing empowers it to reply the above-raised questions and has been utilized for recognizable proof of modern infections (Yozwiak *et al.*, 2012) as well as characterization of uncultured microscopic organisms (Wrighton *et al.*, 2012). This progressed

metagenomics sequencing has been utilized to characterize the organisms related with roots (Bulgarelli *et al.*, 2013) additionally utilized for recognizable proof of taxa that are related with the human intestine (Morgan *et al.*, 2012). The sequencing information gotten from NGS innovation is to begin with subjected to quality control considers. It is the method of sorting out and screening low-quality peruses, which influence the downstream investigation (Zhou *et al.*, 2014). The precision of microbial biodiversity can be made strides by quality sifting (Handelsman, 2004).

#### d. Assembly

Fundamentally, there are two sorts of congregations, i.e., *de novo* gathering in which the genome is built from peruses information and the moment is comparative get together which is utilized to recreate the genome employing a closely related living being (Medvedev *et al.*, 2007). For the *de novo* gathering, three algorithm-based procedures are utilized named as *eager* (Pop and Salzberg, 2008), *cover* format agreement (Myers, 1995), and *De Bruijn* chart (Zerbino and Velvet, 2008). Progressed *de novo* gatherings have been produced with the assistance of a known reference genome to create a comparative get together like *OSLay* (ideal syntenic format of unfinished congregations) (Richter *et al.*, 2007), *Projector 2* (Van *et al.*, 2005) and *ABACAS* (algorithm-based programmed contiguation of collected arrangements) (Assefa *et al.*, 2009).

### The Strategy for Metagenomic Analysis

#### Bioinformatics

One of the most grounded possibilities of metagenomics when compared to ordinary genomics lies is the capacity to identify patterns and relationships characteristic of intuitive between the microbial world and the environment. In display day, metagenomic investigations are reasonable and open to the normal microbiology extend, permitting for the era of gigantic grouping yields (Bertelli and Greub, 2013). The primary step a metagenomic examination after procuring of a test includes the sequencing of DNA.

Information estimate created after sequencing can be decreased by metagenome get together by utilizing coordinates computational approach (Howe *et al.*, 2014). Day by day, innovation is moving forward which leads to lessening in sequencing fetched; subsequently analysts can get to the natural

metagenome, and bioinformatics instruments can be coordinates with metagenome information to deliver valuable comes about and discoveries (Albertsen *et al.*, 2013). Structural and useful explanation of microbial community can be done by utilizing collected peruses and unassembled peruses as well.

#### - Metagenomics Databases and Online Resources

There are numerous databases and online instruments for analyzing and recovering metagenomics information. Table 3 appears the title together with interface of such databases/servers. The European Bioinformatics Founded (EBI) Metagenomics empowers us to yield, analyze, visualize, and compare our information (Mitchell *et al.*, 2016). *MG-RAST* may be a metagenomics examination server for explanation of grouping parts, their phylogenetic classification, useful classification of tests, and comparison between different metagenomes. It moreover computes a starting metabolic reproduction for the metagenome and permits comparison of metabolic recreations of metagenomes and genomes (Wilke *et al.*, 2016). *MEGAN* (Huson *et al.*, 2011) may be a comprehensive tool stash for analyzing microbiome information. One can perform the distinctive analytics utilizing this device like ordered investigation, utilitarian investigation, etc. *QIIME* (Quantitative Bits of knowledge Into Microbial Environment) may be an openly accessible bioinformatics instrument for performing microbiome investigation from crude DNA sequencing information. One can perform *DE* multiplexing and quality sifting, *OTU* (operational ordered unit) picking, ordered task, phylogenetic remaking, and differences investigations and visualizations (Caporaso *et al.*, 2010). *RDP* (ribosomal database) gives quality-controlled, adjusted, and clarified archaeal and bacterial 16S rRNA groupings, contagious 28S rRNA groupings, and a collection of examination instruments to the logical community. *RDP* is an internet apparatus which is utilized to ponder the modern contagious 28S rRNA grouping collection. *RDP* devices are presently openly accessible in bundles for clients to consolidate in their neighborhood workflow (Cole *et al.*, 2009). *SILVA* is an internet unreservedly available apparatus to check the quality of peruses and adjusted (16S/18S, little subunit ribosomal RNA) and huge subunit (23S/28S, LSU) rRNA grouping information of microbes, archaea, and eukaryote (Quast *et al.*, 2013). *Real Time Metagenomics* is an internet



openly accessible instrument which performs comment of metagenomes by relating the person grouping peruses with a database of known arrangements and relegating a one of a kind work to each examined. They produced a new methodology to clarify metagenomes utilizing one of a kind k-meroligo peptide arrangements from 7 to 12 amino acids long (Edwards *et al.*, 2012).

### **Application of Metagenomics and the Impact on Biotechnology**

The discipline of metagenomics is mostly unexplored is anticipated to bring productive result for the analysts working within the zone of microbiology in basically two ways: in to begin with application it'll give information approximately those microbes which are still not cultivated so distant (approximately 99% are uncultured within the unadulterated culture). Furthermore it'll give get to entire organism community dwelling in variety of characteristic environment. Coordinate get to the hereditary cosmetics of organisms of the complete environment community will give unused premise for essential inquire about and unused instrument for application in environment, agribusiness, human wellbeing, bio-industry, etc.

### **Industrial Enzymes and Bioactive Compounds**

There's an expanding request of novel chemicals for mechanical applications, and metagenomics is playing a critical part in giving these biomolecules (Schloss and Handelsman, 2003) extraordinarily proteins that are utilized in wide extend of applications (Kirk *et al.*, 2002). These are required in diminutive sum to synthesize tremendous sum of key particles that are utilized in creating dynamic pharmaceuticals as these are the major building piece of those items (Patel *et al.*, 1994). There are numerous mechanical enzymes which have an awfully wide application in businesses and act as their spine like cellulases, xylanases, lipases, amylases, etc. A novel amylase was confined from a soil metagenome that appeared 90% movement at moo temperature which demonstrated its potential for mechanical misuse (Sharma *et al.*, 2010).

In recent times a main international health-associated hassle involves treating infections which are proof against antibiotics. Earlier these antibiotics have been used for treating human contamination, but they became popular in agriculture and meals industry as well as many different associated sec-

tors, thus eventually implementing high effect on human health (Radhouani *et al.*, 2014). Today metagenomics is playing a completely essential function in discovery of bioactive compounds and antibiotics. It's far taken into consideration as an alternative way of keeping apart antibiotics from environmental samples in addition to to hint the mechanism of bacterial gene resistance. Metagenomics is putting attempt to type out the drug resistance genes in microorganisms against various magnificence of antibiotics. It's any other utility is identification of bioactive molecules having antimicrobial properties (Lim *et al.*, 2005). Nowadays, antibiotic resistance of microbes is an alarming worldwide trouble and emerging as a main threat (Èivljak *et al.*, 2014) as those microbes are developing resistance in opposition to many traditional antibiotics, and alternatively, many researchers are discovering many novel antimicrobial compounds from distinctive environmental resources inclusive of microorganisms, flowers, and animals likewise (de Souza *et al.*, 2014). Its miles suggested that uncultivated soil microbes have ability of novel biomolecules which could be thoroughly exploited in any biotechnological utility (Wilson and Piel, 2013).

### **Applications in Human Health**

Humans are constantly surrounded by using microbes as they not only floor over them but additionally live within their frame. The microbes which might be dwelling inside the human plants are not absolutely characterized (much less than 1%). Furthermore there are sure microbes in our surroundings which can be causative marketers of many infectious diseases. These infectious microbes are particularly characterized by means of laboratory-primarily based surveillance and syndromic surveillance which are strictly counting on the non-laboratory statistics. Detecting these causative marketers of infected illnesses is failed in approximately forty% gastroenteritis cases and 60% in encephalitis instances when conventional approach is used (Finkbeiner *et al.*, 2008; Ambrose *et al.*, 2011). In the current situation of affairs, metagenomics is gambling a totally important role in investigating novel species and traces (Wan *et al.*, 2013), outbreaks (Loman *et al.*, 2013), and complicated illnesses (Wang *et al.*, 2012). Finally, tendencies in infection research have recommended that interactions between organisms in a community can have an effect

on ailment consequences and in some instances it might even be suitable to treat a whole microbial network as a pathogenic entity, in contrast to the belief that a single pathogen reasons an unmarried disorder (Wang *et al.*, 2017).

Metagenomics to be utilized within a medical context not as it were requires the information of the Manual of Clinical Microbiology but moreover all of natural microbiology, Worldwide Diary of Orderly and Developmental Microbiology and the aggregate of the NCBI scientific classification database. In any case, when the usage of clinical symptomatic metagenomics does happen, then as metagenomics gets more affordable and faster, it'll ended up conceivable to serially characterize human microbiomes to explore for infection affiliations (Miller *et al.*, 2013).

Pathogen detection by sequencing is a feasible choice since almost all infectious agents have DNA or RNA genomes. The cost of high-throughput or next-generation sequencing has decreased by orders of magnitude since its introduction in 2004, making it a viable technical method for the identification and taxonomic analysis of clinical samples microorganisms (Gu *et al.*, 2019). In expansion to bacterial pathogens and infections, metagenomics has so distant seen a small utilize within the discovery of parasitic contamination. Plasmodium and Toxoplasma configurations, for example, were discovered in the metagenomes of Egyptian mummies (Khairat *et al.*, 2013).

Because the urine microbiota is so important to an individual's health and disease appearance, metagenomic research into the urine and gut microbiomes is warranted and can lead to new insights into how microbiomes affect human health and urologic infections (Whiteside *et al.*, 2015). Clinical use of symptomatic metagenomics in the future may enable physicians to discover a wide range of picky, anaerobic, and even microorganisms that aren't cultivable in solid people's "sterile" bladder urine as well as in patients with various urological clutters (Smelov *et al.*, 2016).

### Agricultural Applications

The productiveness of agriculture is severely laid low with presence of natural and inorganic anthropogenic pollution that play a totally good sized role in abiotic pressure. These types of abiotic stresses are chargeable for reduction in crop yield. To improve the quality of such soil contaminated through

anthropogenic pollution, bioremediation is needed. Microorganisms of soil metagenome are quite able to producing bio surfactants which could dispose of many anthropogenic pollutants which can be both hydrocarbons and heavy metals (Sun *et al.*, 2006). Bio surfactants are able to removing heavy metals and hydrocarbons through the combination of soil washing and cleanup generation (Kang *et al.*, 2010). Bio surfactants that are produced by means of rhizobacteria have antagonistic houses (Nihorimbere *et al.*, 2011). A common plant pathogen *Pseudomonas aeruginosa* is found to be inhibited by the biosurfactants of staphylococcus of oil-infected soil metagenome (Eddouaouda *et al.*, 2012).

### Environmental Applications

Different sorts of organisms are living in our situations which are accommodating in numerous ways. They play an awfully vital part in decomposing dead fabric show within the environment and making it free from poisons. There are certain organisms which are able to corrupt oil at whatever point it spills over water surface. Numerous organisms moreover have the capacity of cleaning the ground water. Here metagenomics may play exceptionally imperative part in recognizing specific species which are concerned with water treatment reason. Oil-consuming organisms that are display in ocean are reasonable illustrations of microbial bioremediation of water. Numerous other microbes that are show within the soil have qualities of expending overwhelming metals and may be accommodating in decreasing soil poisonous quality. Distinguishing confirmation of these species might be a significant step forward in encouraging research and exploration in this area. As a result, metagenomics and natural scientists may be interested in this area.

### Bioremediation

Bioremediation is the process of degrading and detoxifying natural pollutants via microbe-mediated preparation (Chakraborty *et al.*, 2012). It entails the expulsion of organic and anthropogenic impurities using a specific procedure, making it the most convincing technique (Lovley, 2003). In characteristic weakening local life forms are utilized for detoxifying contaminants through utilizing normal prepare. Many industries are mindful for expanded level of hydrocarbons within the environment due to the deficient combustion of fossil fuel. Metagenomics

can be accommodating in debasement of fragrant compounds by screening and recognizing appropriate living beings in a metagenomic library gotten from oil source (Sierra-García *et al.*, 2014). Numerous qualities and their pathways were distinguished for the corruption of phenol and fragrant compound by utilizing metagenomic approach (Silva *et al.*, 2013).

## References

- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W. and Nielsen, P.H. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol.* 31 (6): 533–538.
- Ambrose, H.E., Granerod, J., Clewley, J.P., Davies, N.W.S., Keir, G., Cunningham, R., Zuckerman, M., Mutton, K.J., Ward, K.N., Ijaz, S., Crowcroft, N.S. and Brown, D.W.G. 2011. Diagnostic strategy used to establish etiologies of encephalitis in a prospective cohort of patients in England. *J Clin Microbiol.* 49 (10) : 3576–3583.
- Assefa, S., Keane, T.M., Otto, T.D., Newbold, C. and Berriman, M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics.* 25 (15): 1968–1969.
- Berry, A.E., Chiocchini, C., Selby, T., Sosio, M. and Wellington, E.M. 2003. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiol Lett.* 223 : 15–20.
- Bertelli, C. and Greub, G. 2013. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clin Microbiol Infect.* 19: 803–813.
- Bowen, D.L.K., Gerlach, R., Peyton, B.M. and Fields, M.W. 2013. Archaeal and bacterial communities in three alkaline hot springs in heart Lake Geyser Basin, Yellowstone National Park. *Front Microbiol.* 4 : 330.
- Bulgarelli, D., Schlaeppi, K., Spaepen, S., Ver Loren, Van Themaat, E. and Schulze-Lefert, P. 2013. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Bio.* 4 : 807–838.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D. and Costello, E.K. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 7 : 335–336.
- Chakraborty, R., Wu, C.H. and Hazen, T.C. 2012. Systems biology approach to bioremediation. *Curr Opin Biotechnol.* 23 : 483–490.
- Chakravorty, S., Helb, D., Burday, M., Connell, N. and Alland, D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods.* 69(2) : 330–339.
- Chen, L., Zhang, Y.H., Huang, T. and Cai, Y.D. 2016. Gene expression profiling gut microbiota in different races of humans. *Sci Rep.* 6 : 23075.
- Ėivljak, R., Giannella, M., Di Bella, S. and Petrosillo, N. 2014. Could chloramphenicol be used against ESKAPE pathogens? Are view of in vitro data in the literature from the 21st century. *Expert Rev Anti-Infect Ther.* 12(2) : 249–264.
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B. and Farris, R.J. 2009. The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37: D141–D145.
- Cooke, S.J., Birnie-Gauvin, K., Lennox, R.J., Taylor, J.J., Rytwinski, T., Rummer, J.L., Franklin, C.E., Bennett, J.R. and Haddaway, N.R. 2017. How experimental biology and ecology can support evidence-based decision-making in conservation: avoiding pitfalls and enabling application. *Conserv Physiol.* 5(1): cox043.
- de Souza, C.E., e Silva Cardoso, M.H., Sousa, D.A., Viana, J.C., de Oliveira-Júnior, N.G., Miranda, V. and Franco, O.L. 2014. The use of versatile plant antimicrobial peptides in agribusiness and human health. *Peptides.* 55 : 65–78.
- Duan, H., Li, X., Mei, A., Li, P., Liu, Y., Li, X. and Xie, S. 2021. The diagnostic value of metagenomic nextC generation sequencing in infectious diseases. *BMC Infectious Diseases.* 21(1); 1–13.þ
- Eddouaouda, K., Mnif, S., Badis, A., Younes, S.B., Cherif, S., Ferhat, S., Mhiri, N., Chamkha, M. and Sayadi, S. 2012. Characterization of a novel biosurfactant produced by *Staphylococcus* sp. strain 1E with potential application on hydrocarbon bioremediation. *J Basic Microbiol.* 52(4) : 408–418.
- Edwards, R. A., Olson, R., Disz, T., Pusch, G. D., Vonstein, V., Stevens, R., Overbeek, R. 2012. Real time metagenomics: using k-mers to annotate metagenomes. *Bioinformatics.* 28(24) : 3316–3317.
- Felczykowska, A., Krajewska, A., Zielińska, S. and Łoć J.M. 2015. Sampling, metadata and DNA extraction important steps in metagenomic studies. *Acta Biochim Pol.* 62(1) : 151–160.
- Finkbeiner, S.R., Allred, A.F., Tarr, P.I., Klein, E.J., Kirkwood, C.D., Wang, D. and Holmes, E.C. 2008. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog.* 4(2) : e1000011.
- Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman R.M. and Handelsman, J. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol.* 68(9): 4301–4306.
- Gu, W., Miller, S. and Chiu, C. Y. 2019. Clinical metagenomic next-generation sequencing for pathogen detection. *Annual Review of Pathology: Mechanisms of Disease.* 14 : 319–338.
- Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol*

- Mol Biol Rev.* 68 : 669–685.
- Hosokawa, T., Kaiwa, N., Matsuura, Y., Kikuchi, Y. and Fukatsu, T. 2015. Infection prevalence of *Sodalis* symbionts among stinkbugs. *Zool Lett.* 1(1).
- Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M. and Brown, C.T. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. *Proc Natl Acad Sci U S A.* 111(13) : 4904–4909.
- Hugenholtz, P. and Pace, N.R. 1996. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol.* 14(190–197) : 10025–10021.
- Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N. and Schuster, S.C. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21(9) : 1552–1560.
- Kang, S.W., Kim, Y.B., Shin, J.D. and Kim, E.-K. 2010. Enhanced biodegradation of hydrocarbons in soil by microbial biosurfactant, sophorolipid. *Appl Biochem Biotechnol.* 160(3) : 780–790.
- Khairat, R., Ball, M., Chang, C. C. H., Bianucci, R., Nerlich, A. G., Trautmann, M. and Pusch, C. M. 2013. First insights into the metagenome of Egyptian mummies using next-generation sequencing. *Journal of Applied Genetics.* (54) : 309–325.
- Kirk, O., Borchert, T.V. and Fuglsang, C.C. 2002. Industrial enzyme applications. *Curr Opin Biotechnol.* 13(4) : 345–351.
- Lim, H.K., Chung, E.J., Kim, J.C., Choi, G.J., Jang, K.S., Chung, Y.R., Cho, K.Y. and Lee, S.W. 2005. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl Environ Microbiol.* 71(12) : 7768–7777.
- Loman, N.J., Constantinidou, C., Christner, M., Rohde, H., Chan, J.Z. M., Quick, J., Weir, J.C., Quince, C., Smith, G.P., Betley, J.R., Aepfelbacher, M. and Pallen, M.J. 2013. A culture-independent sequence-based Metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA.* 309(14) : 1502.
- Lovley, D.R. 2003. Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Microbiol.* 1(1) : 35–44.
- McCliment, E.A., Voglesonger, K.M., O'Day, P.A., Dunn E.E., Holloway, J.R. and Cary, S.C. 2006. Colonization of nascent, deep-sea hydrothermal vents by a novel archaeal and Nanoarchaeal assemblage. *Environ Microbiol.* 8 : 114–125.
- Medvedev, P., Georgiou, K. and Myers, G. 2007. Computability of models for sequence assembly. *Gene.* 4645: 289–301.
- Mitchell, A., Bucchini, F. and Cochrane, G. 2015. EBI metagenomics in 2016 – an expanding and evolving resource for the analysis and archiving of metagenomic data. In: *Nucleic acids research.* <https://doi.org/10.1093/nar/gkv1195>.
- Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L. and Ward, D.V. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13 : R79.
- Mukherjee, P.K., Wang, H., Retuerto, M., Zhang, H., Burkey, B. 2017. Bacteriome and mycobiome associations in oral tongue cancer. *Oncotarget.* 8 : 97273–97289.
- Myers, E.W. 1995. Toward simplifying and accurately formulating fragment assembly. *J Comput Biol.* 2 : 275–290.
- Nayfach, S., Camargo, A. P., Schulz, F., Eloie-Fadrosch, E., Roux, S. and Kyrpides, N.C. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology.* 39(5): 578–585.
- Nihorimbere, V., Marc Ongena, M., Smargiassi, M. and Thonart, P. 2011. Beneficial effect of the rhizosphere microbial community for plant growth and health. *Biotechnol Agron Soc Environ.* 15 : 327–337.
- Parada, A.E., Needham, D.M. and Fuhrman, J.A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol.* 18(5) : 1403–1414.
- Pindjakova, J., Sartini, C., Lo Re, O., Rappa, F., Coupe, B. 2017. Gut Dysbiosis and Adaptive Immune Response in Diet-induced Obesity vs. Systemic Inflammation. *Front Microbiol.* 8 : 1157.
- Piombo, E., Abdelfattah, A., Droby, S., Wisniewski, M., Spadaro, D. and Schena, L. 2021. Metagenomics Approaches for the Detection and Surveillance of Emerging and Recurrent Plant Pathogens. *Microorganisms.* 9(1) : 188.
- Pop, M. and Salzberg, S.L. 2008. Bioinformatics challenges of new sequencing technology. *Trends Genet.* 24 : 142–149.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41 : D590–D596.
- Radhouani, H., Silva, N., Poeta, P., Torres, C., Correia, S. and Igrejas, G. 2014. Potential impact of antimicrobial resistance in wildlife, environment and human health. *Frontiers in Microbiology.* 5 : 23.
- Richter, D.C., Schuster, S.C. and Oslay, H.D.H. 2007. Optimal syntenic layout of unfinished assemblies. *Bioinformatics.* 23(13) : 1573–1579.
- Schloss, P.D. and Handelsman, J. 2003. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol.* 14: 303–310.
- Schmidt, T.M., DeLong, E.F. and Pace, N.R. 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol.* 173(14) : 4371–8.
- Seth-Smith, H.M.B., Harris, S.R., Skilton, R.J., Radebe, F.M., Golparian, D., Shipitsyna, E., Duy, P.T., Scott,



- P., Cutcliffe, L.T., O'Neill, C., Parmar, S., Pitt, R., Baker, S., Ison, C.A., Marsh, P., Jalal, H., Lewis, D.A., Unemo, M., Clarke, I.N., Parkhill, J. and Thomson, N.R. 2013. Whole-genome sequences of chlamydia trachomatis directly from clinical samples without culture. *Genome Res.* 23(5) : 855–866.
- Sharma, S., Khan, F.G. and Qazi, G.N. 2010. Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. *Appl Microbiol Biotechnol.* 86(6) : 1821–1828.
- Sharpton, T.J. 2014. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci.* 5 : 209.
- Sharpton, T.J., Riesenfeld, S.J., Kembel, S.W., Ladau, J., O'Dwyer, J.P. and Green, J.L. 2011. PhylOTU: a high-throughput procedure quantifies microbial community diversity and resolves novel taxa from metagenomic data. *PLoS Comput Biol.* 7:e1001061.
- Sierra-García, I.N., Alvarez, J.C., de Vasconcellos, S.P., de Souza, A.P., Neto, E.V. d S, de Oliveira, V.M. and Mormile, M.R. 2014. New hydrocarbon degradation pathways in the microbial metagenome from Brazilian petroleum reservoirs. *PLoS One.* 9(2) : e90087.
- Silva, C.C., Hayden, H., Sawbridge T, Mele P, Paula SOD, Silva, L.C.F., Vidigal, P.M.P., Vicentini, R., Sousa M.P., Torres, A.P.R., Santiago, V.M.J. and Oliveira, V.M. 2013. Identification of genes and pathways related to phenol degradation in metagenomic libraries from petroleum refinery wastewater. *PLoS One.* 8(4) : 1–11.
- Smelov, V., Naber, K., Bjerklund Johansen, T.E. 2016. Letter to the Editor: Diagnostic Criteria in Urological Diseases do not Always Match with Findings by Extended Culture Techniques and Metagenomic Sequencing of 16S rDNA. *Open Microbiol J.* 10 : 23–26.
- Sobel, J., Henry, L., Rotman, N. and Rando, G. 2017. BeerDeCoded: the open beer metagenome project. *F1000 Research.* 6 : 1676.
- Soo, R.M., Wood, S.A., Grzymalski, J.J., McDonald, I.R. and Cary, S.C. 2009. Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus. *Antarctica Environ Microbiol.* 11: 715–728.
- Sun, X., Wu, L. and Luo, Y. 2006. Application of organic agents in remediation of heavy metals- contaminated soil. *Ying Yong Sheng Tai Xue Bao.* 17(6) : 1123–1128.
- Ta°, N., de Jong, A. E., Li, Y., Trubl, G., Xue, Y. and Dove, N. C. 2021. Metagenomic tools in microbial ecology research. *Current Opinion in Biotechnology.* 67 : 184–191.
- Thomas, T., Gilbert, J. and Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microb Inf Exp.* 2(1) : 3.
- Uchiyama, T. and Miyazaki, K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol.* 20(6) : 616–622.
- Van Hijum, S.A.F.T., Zomer, A.L., Kuipers, O.P. and Kok, J. 2005. Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. *Nucleic Acids Res.* 33 : 560–566.
- Wan, X.F., Barnett, J.L., Cunningham, F., Chen, S., Yang, G., Nash, S., Long, L.P., Ford, L., Blackmon, S., Zhang, Y., Hanson, L. and He, Q. 2013. Detection of African swine fever virus-like sequences in ponds in the Mississippi Delta through metagenomic sequencing. *Virus Genes.* 46(3) : 441–446.
- Wang, B., Yao, M., Lv, L., Ling, Z. and Li, L. 2017. The human microbiota in health and disease. *Engineering.* 3: 71–82.
- Wang, L., Wang, S. and Li, W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics.* 28(16): 2184–2185.
- Whiteside, S. A., Razvi, H., Dave, S., Reid, G. and Burton, J. P. 2015. The microbiome of the urinary tract: a role beyond infection. *Nat Rev Urol.* 12(2) : 81–90.
- Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P. and Chaterji, S. 2016. The MG-RAST metagenomics database and portal. *Nucleic Acids Res.* 44(D1) : D590–D594.
- Wilson, M.C. and Piel, J. 2013. Metagenomic approaches for exploiting uncultivated bacteria as are source for novel biosynthetic enzymology. *Chem Biol.* 20(5) : 636–647.
- Wooley, J.C., Godzik, A. and Friedberg, I. 2010. A primer on metagenomics. *PLoS Comput Biol.*
- Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., Long, P.E. and Banfield, J.F. 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science.* 337(6102) : 1661–1665.
- Yatsunenkov, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M. and Heath, A. C. 2012. Human gut microbiome viewed across age and geography. *Nature.* 486 : 222–227.
- Yozwiak, N.L., Skewes-Cox, P., Stenglein, M.D., Balmaseda, A., Harris, E., DeRisi, J.L. and Rico-Hesse, R. 2012. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis.* 6(2) : e1485.
- Yutin, N., Makarova, K. S., Gussow, A. B., Krupovic, M., Segall, A., Edwards, R. A. and Koonin, E. V. 2018. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat Microbiol.* 3 : 38–46.
- Zerbino, D.R. and Velvet, B.E. 2008. Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18 : 821–829.
- Zhou, Q., Su, X. and Ning, K. 2014. Assessment of quality control approaches for metagenomic data analysis. *Sci Rep.* 4 : 6957.