Next-Generation Sequence: A Review on Metagenomic Approach to Discovery of Novel Enzymes from the Soil Environment

Rita-Cindy Aye-Ayire Sedjoah[1*]

[1]*College of Food Science and Technology, Nanjing Agricultural University, Nanjing, P. R. China.*

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Mini-review Article

ABSTRACT

Next-generation sequencing (NGS) makes a large mass of sequences. As a technology that allows the sequence of deoxyribonucleic acid (DNA) molecules larger than one million base pairs, it has been applied in the food research and medical fields. In the food sector, NGS has been used in food safety for the detection of species authenticity of food products and for mostly discovering novel industrial enzymes. The soil ecosystem houses a great number of non-culturable microbes thus novels enzymes can still be discovered to date. The conventional methods used in enzyme discovery have less chances to identify novel gene clusters and bioactivities. Therefore, there is a dire need for high-throughput technology, together with advanced bioinformatics for the search of novel enzymes or biocatalysts from soil metagenomes. This review article thus gives a summary of the progress in the application of next-generation sequencing in the identification and characterization of novel enzymes with a special focus on enzymes from the soil environment.

Keywords: Soil metagenome; high throughput; novel enzymes; next-generation sequencing.

_____

*Corresponding author: E-mail: ritacindy19@gmail.com;

1. INTRODUCTION

1.1 Soil Microbial Diversity and Metagenomics Overview

Soil is a reliable source for direct sequencing of new microbial resources hence there have been efforts to find microbial resources from the soil metagenome [1]. The soil diversity is known to be the highest community compared to other microbial communities but then only about 1% is known to be specified by traditional cultivation-based sequencing methods [2]. The microbial population of the soil, which is mostly dominated by some members of the bacteria phyla, is affected by chemical features, soil richness, pH and physical features [3,4]. The types and number of microbes in the soil vary from one type of soil to another. For instance, forest soils have a higher number of the phylum Acidobacteria but a lower amount of β-proteobacterial [5]. Also, a high number of members of β-proteobacteria are found in agricultural soils [6]. The Amazon forest soil is highly incorporated with plant organic matter (cellulose) which is degraded by microorganisms, therefore, microorganisms are supposed to be able to express a variety of enzymes that degrade cellulose, like β-glucosidases [7]. Ring hydroxylating dioxygenases (RHDs) which catalyze the biodegradation of hydrocarbons are expected to be found in polluted soils [8]. A gram of soil habits about 10 billion microorganisms [9], thus the soil is seen as a great source for direct sequencing of novel microbial resources, mostly enzymes.

Enzymes are known as catalysts, as such, they serve as biocatalysts for biotechnological applications and are commonly used in cleaning detergents, medical cleaning devices, textile and food manufacturing industries. In the food industry, enzymes serve as biocatalysts and are used for processing uncooked substances for the manufacturing of extraordinary sorts of products. For this reason, a good number of researchers have been successful in finding novel enzymes from the soil metagenome [6,10]. In addition, enzymes are found in plant roots and residues, living and dead microbes, and soil animals. The types of soils that contain enzymes could be humus or clay or humus-clay. Lipolytic enzymes (esterase/lipase) are one of the most popular recent soil metagenome enzymes [1,11,12], and they are very useful in organic synthesis. Enzymes used in the industries mostly include amylolytic enzymes, lipolytic enzymes, xylanase and β-glucosidase. The use of an enzyme in industrial processes focuses on its attributes, such as the specificity of its substrate, main reaction products, the pattern of action, effective temperature and pH [13].

Metagenomics is an advanced method for analyzing microbial genomic DNAs from unculturable microorganisms in the natural environment [14]. It allows for the identification and characterization of microbes without prior knowledge of them. Metagenomics was introduced in response to the limitations of cultured microorganisms to obtain enzymes using conventional methods [15]. Metagenomics solves the drawbacks of the conventional microbial method's isolation and cultivation processes, and in so doing, enhances the spectrum of exploitation of microbial resources. This helps to understand the various characteristics and functions of microbes present in the soil environment. Most enzymes are unculturable microbes hence the application of metagenomics aids to discover novel enzymes. Metagenomics has been mostly used to explore novel enzymes for biotechnological applications in the food and biofuel industries through the use of highly advanced genome sequencing methods. For instance, two novel β-glucosidases (AmBGL$_{17}$ and AmBGL$_{18}$) were reported after functional screening and sequencing by Illumina from Amazon forest soil [7]. This enzyme has biotechnological applications in the food industry such as detoxification of cassava cyanogenic glucosides, and enhance aroma in wine-making.

Since conventional approaches are time-consuming and high-priced in identifying and characterizing novel biocatalysts, it is necessary to design new effective detection strategies for novel soil metagenomic enzymes. Metagenomic procedures have made a significant contribution to a deeper understanding of a microbiome via a species-level/strain-level description [16].

Therefore, this review concisely presents the progress of next-generation sequencing as a metagenomic method in the determination and characterization of novel enzymes with a special focus on enzymes found in the soil environment.

2. NEXT-GENERATION SEQUENCING

2.1 Overview of Metagenomic Sequencing Techniques

Besides screening, particularly based on sequence and function, metagenomic

sequencing is another way to isolate novel enzymes. Owing to the low frequency of clones with desired traits, the screening methods are cumbersome [17]. Other popular sequencing technologies used to identify and characterize enzymes includes Sanger sequencing, PCR cloning, 16S rDNA sequencing, and whole-genome sequencing. Conventional culturing techniques for DNA sequencing of novel enzymes that were being used some years ago are usually followed by phenotypic screening, compound isolation, and characterization. With the non-culturing genome sequence approach, there is direct cloning of the targeted genes using the polymerase chain reaction (PCR) [18]. Early DNA sequencing methods are usually called large-scale sequencing since they are used to sequence very long DNA pieces. Notable among them is the Shotgun sequencing approach. These methods are faced with various issues such as being time-consuming, slow analysis which does not allow sequencing of multiple or complex genes, limitation of primers, less specificity, and technical biases [19]. There are different methods in metagenomic sequencing approaches based on DNA cloning but, the desire for rapid advancement and high preference for low-priced sequencing has prompted the deployment of high-throughput sequencing methods that run a massive number of sequencing tasks simultaneously [20,21].

2.2 What is Next-Generation Sequencing?

Next-generation sequencing is an advanced technology that has allowed millions of DNA molecules to be sequenced simultaneously. NGS has, however, yielded significant benefits in less than a decade to address the shortcomings of traditional methods [22]. There are three sequencing generation groups; the first, second, and third-generation sequencing methods. In 2005, the first-generation technology was launched which consist of Sangers sequencing [23,24]. Prior to the sequence of amplified DNA clones, the sequencing methods of the 1st and 2nd generations are expected to put together sequencing libraries. Meanwhile, the third-generation referred to as the technology of single-molecule sequencing can be sequenced without the construction of the amplification libraries, this technique known as the third-generation sequencing techniques (long-read sequencing).

The term 'Next-Generation Sequencing' is another name for the second-generation sequencing, and it is often referred to as a short-read or high-performance sequencing method [25]. This is because, the entire genome is first sequenced by splitting the whole genome into shorter fragments and then sequencing the various fragments together. Also, next generation sequencing facilitates the recovering of further details regarding the taxonomy of the sample to be retrieved [26]. The NGS is sequenced either by synthesis or ligation [27]. There are 3 commonly used NGS techniques in discovering novel enzymes from the soil microbiome which consists of 454-pyrosequencing procedure, Ion Torrent PGM method, and Illumina MiSeq. For metagenomic research, NGS can sequence and classify nucleic acids from many different taxa, so it can identify genes from several species that belong to completely different kingdoms [28], consequently, the technique is widely applied in molecular biology. Second-generation sequencing has many benefits over traditional sequencing methods, including, (1) time reduction, (2) sequencing results are instantly detectable without electrophoresis, (3) increased efficiency, and (4) lower operating cost [29].

2.3 The Workflow of Next-Generation Sequencing

Certain procedures are carried out to be able to arrive at a successful sequenced genome. These steps are explained below:

1. DNA preparation: This first and critical step entails the extraction of complete DNA from the soil surrounding and reduced into small fragments. There are two strategies of DNA isolation; direct and indirect methods. Direct extraction techniques rely on cell lysis of test samples using enzymes, detergents, or acoustic agents, inside the soil environmental matrix [30]. In the quest for high molecular-weight DNA from soil, advanced methods for the direct isolation of DNA from soil have been established by enriching the soil preceding DNA isolation to improve the size [31]. Fragmentation can be done by enzymatic method, sonication, or by chemicals [32,33]. To generate the 'sequencing library', the fragments are connected to adapters. Isolation of excessive molecular mass DNA lets in the characterization of large areas of the genomes and it also enables the characterization of broad genome areas [34].

2. Construction of metagenomic library: The process involves making ready the DNA in a manner that is consistent with the method to be used for sequencing [32]. Soils with a high level of humic acid are co-purified before cloning. The concentrated soil metagenomic DNA is further cloned into multiple separate plasmids. According to the literature, the fosmid vector is the most efficient cloning vector for building a library of soil metagenomes [35], although, certain soil metagenomic libraries have been created using cosmid vectors or bacterial artificial chromosome (BACs) vectors. To complete the process, a substitute host bacterium, like *Escherichia coli*, is needed [36]. There are two methods of amplifying DNA samples for library preparations after fragmentation; either by enlargement of sequence sections by PCR amplification usually called amplicon-based amplification, or by using probes to capture regions of interest. By attachment to solid substrates or tiny beads, individual fragments are then naturally separated in the case of Illumina sequencing technique [37]. Besides Illumina's Nextera prep, which is a commercial kit for making sequencing libraries without using PCR [38], the preparation of the library mainly includes: (a) splitting, (b) end-repair, (c) phosphorylation of the 5' prime ends, (d) A-tailing of the 3' ends to allow adapter sequencing ligation, (e) adapter ligation, and (f) PCR amplification to enhance a product with adapters connected to both sides. The whole library or a selected fragment of the generated library is used to prepare templates for the sequencing of DNA fragments. To be able to identify a sequenced fragment, each added nucleotide is labeled with a different colored fluorescent probe [37]. After library preparation, it is subjected to either metagenomic screening methods or high-throughput sequencing [7,39] for biocatalyst prospecting.

3. Sequencing: This review focuses on only those NGS high-throughput methods that have been used in discovering novel enzymes from the soil metagenome over the years, which mostly happens to be the methods belonging to the second-generation sequencing. Each fragment is sequenced several times in NGS sequencing, allowing multiple contigs to cover the whole regions of sequenced DNA [37]. These methods are increasingly being used for the sequencing of soil metagenomes, such as discussed below:

2.3.1 Roche GS-FLX 454 genome sequencing

Also known as "pyrosequencing", amplifies DNA using the emulsion PCR style (emPCR) [40], which is then immobilized on beads. Just like all second-generation sequencing methods have in common, pyrosequencing makes use the construction of solid surfaces in library amplification, and DNA template library is achieved without cloning in a host cell [23]. This approach can have a mean read length of over 400 nucleotides per run by the use of picotiter volume wells containing one bead, and, sequencing enzymes that facilitate the sequencing of large amounts of DNA [37]. It works by the principle of adding one nucleotide after another in a cycle, and the release of pyrophosphate (PPi) to generate the device's identified light signals that are then converted into nucleotide sequences, thus, the addition of nucleotides will make chain elongation possible [41]. The sequencing machine contains many picolitre-volume wells [24]. Compared to Sanger sequencing, this technique yields shorter reading spans, but at a higher speed [42]. Being the first of the NGS, its assessment showed some limitations including; repeating the same nucleotides leading to low accuracy in detection, and, prolonged sample preparation protocols based on emPCR [16,17,24].

2.3.2 Illumina sequencing

DNA amplification is by bridge PCR style [43]. This sequencing method works by using the principle of the reversible dye-terminators and polymerases, in addition, clonal amplification of DNA on a surface is involved [44]. To form local clonal DNA colonies, DNA fragments and primers are connected to a flow cell and enhanced with a polymerase. The dye is chemically extracted, along with the terminal 3' blocker, facilitating the subsequent cycle. Four kinds of reversible terminator bases (RT-bases) are inserted, and non-incorporated nucleotide sequences are been washed to decide a sequence. A digital camera captures pictures of the fluorescent-labeled nucleotides. By inserting each nucleotide one after the other, DNA chains are extended, but then the photo can be captured at a later stage. This permits sequential images taken from a sole camera, capturing

large arrays of DNA colonies. Bergman et al. [7], used sequencing by Illumina technology to discover two novel putative glycosyl hydrolases; AmBGL17 and AmBGL18 from two positive fosmid metagenomic clones found in the Amazon soil. To investigate the capacity of polycyclic aromatic hydrocarbons (PAHs) at a polluted site, a blend of pyrosequencing and Illumina metagenomic sequence evaluation was applied [8]. A notable limitation of Illumina is, not deleting a nucleotide blocker after signal detection which could hinder the bonding of a fresh nucleotide to the DNA fragment in the next step [43].

2.3.3 Ion Torrent semiconductor sequencing

Being similar to pyrosequencing technology, ion torrent sequencing yields a similar output and a read length. Also, both Illumina and the Ion Torrent platforms make use of sequencing by synthesis technology. The generation of hydrogen ions causes the release of a proton identified by a supersensitive ion sensor and converts into nucleotide sequences each time a nucleotide corresponding to the leading template is introduced to the cycle. The sequencing reaction occurs in millions of microwells [45]. Ions are produced in the process of DNA polymerization, and the template strand to be sequenced is saturated with a single nucleotide form. The Ion Torrent will sense the shift in hydrogen, like that of a potentiometer. With a mode span of 400 bp, its highest value is ~500 million reads [46]. Moreover, Ion Torrent sequencing is less expensive as it turns out to be one-tenth of the pyrosequencing's price [47]. Seven α-L-fucosidase-encoding genes were identified when Lezyk and others sequenced soil metagenomic DNA using Ion Torrent [48]. In bioprocesses such as inflammation, cell signaling, and metastasis, Fucose (6-deoxy-L-galactose) is employed. This method is also faced with difficulty in sequencing homopolymer regions. Detailed descriptions of studies that use NGS approaches to classify novel soil-related enzymes are given in Table 1.
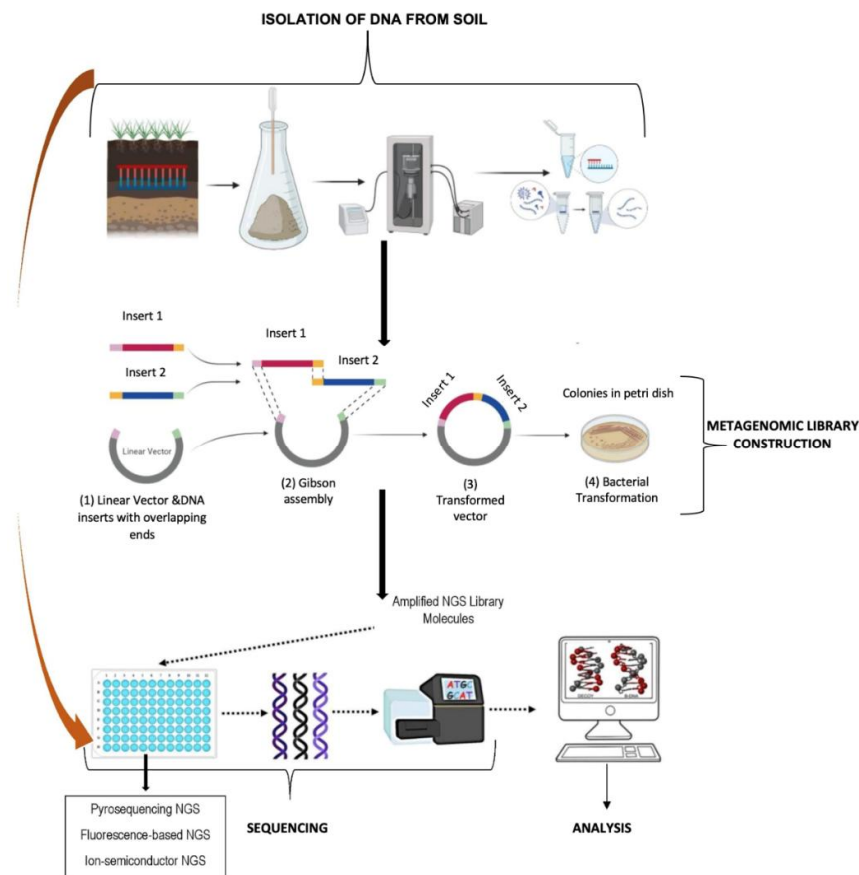
Fig. 1. A schematic workflow of next-generation sequencing

Table 1. NGS methods for discovering soil-related novel enzymes

Source	Enzymes	Sequencing approach	Amplification for library construction
Amazon Forest Soil	β -glucosidase	Illumina [7]	Yes
Dump Soil	α-L-fucosidases	Ion Torrent PGMs [48]	Yes
Peat-Swamp Forest Soil	Lipases/esterases (EstPS2)	Pyrosequencing [49]	Yes
Polluted Soil	Dioxygenase	Pyrosequencing [8]	Yes
Organic Field Soil	Esterase/lipase (abgT)	Pyrosequencing [50]	Yes
Soil	Carboxylesterase	Pyrosequencing [51]	Yes
Soil	Chitinase	Pyrosequencing [52]	Yes
Grassland Soil	Cellulase, Xylanase	Pyrosequencing[53]	Yes

2.4 Data Analysis

The next step after sequencing is the assembly and annotation of 'gene' sequences. Analysis and interpretation of NGS data are usually operated using bioinformatics. After analyzing the novel enzymes, they may be improved or modified to suit certain industrial applications [54]. Software or computational tools are used to analyze the resulting sequence data sometimes, or mostly by comparing it with a 'normal reference' genome (comparative metagenome). Examples include; using BLAST to compare with NCBI nucleotide database, or using MEGAN and Mothur Softwares [54,55]. NGS' benefits can be summarized in the following points: (i) amplification of DNA fragments in the laboratory; (ii) generates multiple short reads in parallel sequencing reactions; (iii) generates a high volume of data; (iv) construction of the sequencing library in vitro; (v) immobilization of DNA on a solid substrate; (vi) low operational cost; and (vii) a wide range of detection without interferences [16,56,57].

2.5 Limitations of NGS

Meanwhile, NGS technologies are faced with some limitations such as; the GC bias as a result of PCR amplification is that leads to mishaps in the sequencing [51]. Also, fluorophore-dependent errors usually arrive from reads from the pyrosequencing and Illumina sequences. Furthermore, objects can be produced if nucleic acids are fragmented or of low quantity/quality [58].

3. CONCLUSION

Several works of literature indicate that mining of industrial enzymes from the soil environment using high throughput metagenomics also called NGS has been on the rise even though its use is not much in soil metagenome bioprospecting. Compared to the previously used methods, this advanced technology aids to sequence DNA on a wide scale more quickly, yet at a very minimal price. NGS metagenomics approaches also contain more detailed information. To exploit such complex microbiome as soil for new enzymes, the combination of functional screening of metagenomes, NGS methods and proper bioinformatics are essential. NGS sequencing has redefined genomics research and molecular biology.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Lee MH, Lee SW. Bioprospecting potential of the soil metagenome: novel enzymes and bioactivities. Genomics inform. 2013;11(3):114.
DOI: 10.5808/GI.2013.11.3.114

2. Torsvik V, Sørheim R, Goksøyr J. Total bacterial diversity in soil and sediment communities—a review. J. Ind. Microbiol. Biotechnol. 1996;17(3-4).
DOI: 10.1007/bf01574690

3. Hassink J, Bouwman LA, Zwart KB, Bloem J, Brussaard L. Relationships between soil texture, physical protection of organic

matter, soil biota, and C and N mineralization in grassland soils. Soil Structure/Soil Biota Interrelationships; 1993.

DOI: 10.1016/0016-7061(93)90150-J

4. Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM. Accessing the soil metagenome for studies of microbial diversity. Appl. Environ. Microbiol. 2011;77(4):1315-24.

DOI: 10.1128/AEM.01526-10

5. Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on Escherichia coli. Appl. Environ. Microbiol. 2005;71(12):7768-77.

DOI: 10.1128/AEM.71.12.7768-7777.2005

6. Simon C, Daniel R. Metagenomic analyses: Past and future trends. Appl. Environ. Microbiol. 2011;77(4):1153-61.

DOI: 10.1128/AEM.02345-10

7. Bergmann JC, Costa OY, Gladden JM, Singer S, Heins R, D'haeseleer P, Simmons BA, Quirino BF. Discovery of two novel β-glucosidases from an amazon soil metagenomic library. FEMS Microbiol. Lett. 2014;351(2):147-55.

DOI: 10.1111/1574-6968.12332

8. Chemerys A, Pelletier E, Cruaud C, Martin F, Violet F, Jouanneau Y. Characterization of novel polycyclic aromatic hydrocarbon dioxygenases from the bacterial metagenomic DNA of a contaminated soil. Appl. Environ. Microbiol. 2014;80(21):6591-600.

DOI: 10.1128/AEM.01883-14

9. Jansson J. Towards tera terra: Terabase sequencing of terrestrial metagenomics. Microbe Mag. 2011;6(7):309-315.

DOI: 10.1128/microbe.6.309.1

10. Iqbal HA, Feng Z, Brady SF. Biocatalysts and small molecule products from metagenomic studies. Curr. Opin. Chem. Biol. 2012;16(1-2):109-16.

DOI: 10.1016/j.cbpa.2012.02.015

11. Lee SW, Won K, Lim HK, Kim JC, Choi GJ, Cho KY. Screening for novel lipolytic enzymes from uncultured soil microorganisms. Appl. Microbiol. Biotechnol. 2004;65(6):720-6.

DOI: 10.1007/s00253-004-1722-3

12. Lee MH, Hong KS, Malhotra S, Park JH, Hwang EC, Choi HK, Kim YS, Tao W, Lee SW. A new esterase EstD2 isolated from plant rhizosphere soil metagenome. Appl. Microbiol. Biotechnol. 2010;88(5): 1125-34.

DOI: 10.1007/s00253-010-2729-6

13. Homaei AA, Sariri R, Vianello F, Stevanato R. Enzyme immobilization: An update. J. Chem. Biol. 2013;6(4):185-205.

DOI: 10.1007/s12154-013-0102-9

14. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. 1998;5(10):R245-9.

DOI: 10.1016/S1074-5521(98)90108-9

15. Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH. Are uncultivated bacteria really uncultivable?. Microbes Environ. 2012;27(4):356–366.

DOI: 10.1264/jsme2.ME12092

16. Cao Y, Fanning S, Proos S, Jordan K, Srikumar S. A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. Front. Microbiol. 2017;8:1829.

DOI: 10.3389/fmicb.2017.01829

17. Daniel R. The soil metagenome–a rich resource for the discovery of novel natural products. Curr. Opin. Biotechnol. 2004;15(3):199-204.

DOI: 10.1016/j.copbio.2004.04.005

18. Zaparucha A, De Berardinis V, Vaxelaire-Vergne C. Chapter 1. Genome mining for enzyme discovery. 2018;1–27.

19. Ryding S. Shotgun metagenomic sequencing. 2013;1-5.

Acessed 10 December 2020.

Available:https://www.news-medical.net/life-sciences/Shotgun-Metagenomic-Sequencing.aspx P

20. Grada A, Weinbrecht K. Next-generation sequencing: Methodology and application. J. Invest. Dermatol. 2013; 133(8):e11.

DOI: 10.1038/jid.2013.248

21. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. T Am. J. Hum. Genet. 2009;85(2):142-54.

DOI: 10.1016/j.ajhg.2009.06.022

22. Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, Arvanitidis C, Iliopoulos L. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. Bioinform. Biol. Insights. 2015;9:BBI-S12462.

DOI: 10.4137/BBI.S12462

23. Kulski JK. Next-generation sequencing—an overview of the history, tools, and "omic" applications. Next generation sequencing-advances, Applications and challenges. 2016;14:3-60.

24. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057): 376-80.

DOI: 10.1038/nature03959

25. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. Nat. Rev. Microbiol. 2015;13(12):787-94.

DOI: 10.1038/nrmicro3565

26. Song ZQ, Wang FP, Zhi XY, Chen JQ, Zhou EM, Liang F, et al. Bacterial and archaeal diversities in Y unnan and T ibetan hot springs, China. Environ. Microbiol. 2013;15(4):1160-75.

DOI: 10.1111/1462-2920.12025

27. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. 2016;17(6):333.

DOI: 10.1038/nrg.2016.49

28. Metagenomic Next Generation Sequencing: How Does It Work and Is It Coming to Your Clinical Microbiology Lab. American Society for Microbiology. 2019;11:4.

Acessed 30 November 2020.

Available:https://asm.org/Articles/2019/November/Metagenomic-Next-Generation-Sequencing-How-Does-It

29. Prayogo FA, Budiharjo A, Kusumaningrum HP, Wijanarka W, Suprihadi A, Nurhayati N. Metagenomic applications in exploration and development of novel enzymes from nature: A review. J. Genet. Eng. Biotechnol. 2020;18(1):39.

DOI: 10.1186/s43141-020-00043-9

30. Harris AD. Soil metagenomics: A prospective approach for novel enzyme discovery. Int. J. Curr. Res. 2012;4: 88-92.

31. Bertrand H, Poly F, Lombard N, Nalin R, Vogel TM, Simonet P. High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. J J. Microbiol. Methods. 2005;62(1):1.

DOI: 10.1016/j.mimet.2005.01.003

32. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. Library construction for next-generation sequencing: Overviews and challenges. Biotechniques. 2014;56(2):61-77. DOI: 10.2144/000114133

33. Marine R, Polson SW, Ravel J, Hatfull G, Russell D, Sullivan M, Syed F, Dumas M, Wommack KE. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. Appl. Environ. Microbiol. 2011; 77(22):8071-9.

DOI: 10.1128/AEM.05610-11

34. Berry AE, Chiocchini C, Selby T, Sosio M, Wellington EM. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol. Lett. 2003; 223(1):15-20.

DOI: 10.1016/S0378-1097(03)00248-9

35. Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res. 1992;20(5):1083-5.

DOI: 10.1093/nar/20.5.1083

36. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, et al. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 2000;66(6):2541-7.

DOI: 10.1128/AEM.66.6.2541-2547.2000

37. Horn-Saban S, Amann-Zalcenstein D. Frontiers in DNA sequencing: The (R) evolution of sequencing technologies. Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation. Chichester: John Wiley & Sons, Ltd; 2006.

38. Adey A, Morrison HG, Xun X, Kitzman JO, Turner EH, Stackhouse B, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro

transposition. Genome Biol. 2010;11(12):1-7.

DOI: 10.1186/gb-2010-11-12-r119

39. Kakirde KS, Parsley LC, Liles MR. Size does matter: Application-driven approaches for soil metagenomics. Soil Biol. Biochem. 2010;42(11):1911-23.

DOI: 10.1016/j.soilbio.2010.07.021.

40. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 1996;242(1):84-9.

DOI: 10.1006/abio.1996.0432

41. Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ. Microbiol. 2000;2(5):516-29.

DOI: 10.1046/j.1462-2920.2000.00133.x.

42. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. Front. Genet. 2015;6:348.

DOI: 10.3389/fgene.2015.00348

43. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 2006;34(3):e22.

DOI: 10.1093/nar/gnj023

44. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218): 53-9.

DOI: 10.1038/nature07517

45. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348-52.

DOI: 10.1038/nature10242

46. Glenn TC. Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 2011;11(5):759-69.

DOI: 10.1111/j.1755-0998.2011.03024.x

47. Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, et al. Microbial 16S rRNA Ion tag and community metagenome

sequencing using the Ion Torrent (PGM) Platform. J. Microbiol. Methods. 2012;91(1):80-8.

DOI: 10.1016/j.mimet.2012.07.008

48. Lezyk M, Jers C, Kjaerulff L, Gotfredsen CH, Mikkelsen MD, Mikkelsen JD. Novel α-L-fucosidases from a soil metagenome for production of fucosylated human milk oligosaccharides. PloS one. 2016;11(1):e0147438.

DOI: 10.1371/journal.pone.0147438.

49. Bunterngsook B, Kanokratana P, Thongaram T, Tanapongpipat S, Uengwetwanit T, Rachdawong S, et al. Identification and characterization of lipolytic enzymes from a peat-swamp forest soil metagenome. Biosci. Biotechnol. Biochem. 2010;74(9)1848–1854.

DOI: 10.1271/bbb.100249.

50. Stöveken J, Singh R, Kolkenbrock S, Zakrzewski M, Wibberg D, Eikmeyer FG, et al. Successful heterologous expression of a novel chitinase identified by sequence analyses of the metagenome from a chitin-enriched soil sample. J. Biotechnol. 2015;201:60-8.

DOI: 10.1016/j.jbiotec.2014.09.010.

51. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2011;12(2):1-4.

DOI: 10.1186/gb-2011-12-2-r18

52. Popovic A, Hai T, Tchigvintsev A, Hajighasemi M, Nocek B, Khusnutdinova AN, et al. Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. Sci. Rep. 2017;7(1):1-5.

DOI: 10.1038/srep44103

53. Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett. 2012;34(4):663-75.

DOI: 10.1007/s10529-011-0830-2

54. Huson DH, Weber N. Microbial community analysis using MEGAN. 1st ed. Methods in enzymology. 2013;531:465-85.

55. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing

microbial communities. Appl. Environ. Microbiol. 2009;75(23):7537-41.

DOI: 10.1128/AEM.01541-09

56. Karamperis K, Wadge S, Koromina M, Patrinos GP. Genetic testing. In: Applied Genomics and Public Health 2020;189-207.

57. Meera Krishna B, Khan MA, Khan ST. Next-generation sequencing (NGS) platforms: An exciting era of genome sequence analysis. In: Tripathi V, Kumar P, Tripathi P, Kishore A, Kamle M, editors. Microbial Genomics in Sustainable Agroecosystems: 2nd ed. Singapore: Springer Singapore; 2019.

58. Hsiao SJ. Sources of error in molecular diagnostic analyses. In Accurate Results in the Clinical Laboratory. 2019;337-347.