Metagenomic screening strategies for bioprospecting enzymes from environmental samples

A K Wani¹, F Rahayu², F T Kadarwati², C Suhara², R Singh¹*, D S Dhanjal¹, N Akhtar¹, T G Mir¹ and C Chopra¹

¹School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab-144411, India.
²Indonesian Sweetener and Fiber Crops Research Institute, Jl. Raya Karangploso, KM-4, Malang 65100, Indonesia.

*Corresponding author email: reena.19408@lpu.co.in | ORCID ID : 000-0003-4986-3547

Abstract. Globally, there is a growing demand for biocatalysts because of the associated efficacy and efficiency. The applications of enzymes in food, paper, pulp, textile, and chemical industries have prompted enzyme exploration. Microbes, being the natural reservoirs of enzymes, have gained researchers' attention, and the quest for microbial enzymes has increased in past years. This review provides insights about metagenomics techniques and their applicability in obtaining microbial-origin enzymes from diverse environmental samples besides highlighting their importance. The metagenomic approach has emerged as a promising way towards replacing conventional microbial techniques with culture-independent methods involving direct isolation of DNA environmental samples. There are two primary methodologies, i.e., functional-based and sequence-based, to identify and characterize industrially valuable biocatalysts from the environmental microcosms. Many of the obtained enzymes are successfully used in diverse food, cosmetics, and pharmaceutical industries. However, there are some complications associated with it that can be minimized only by further investigations. The paper focuses on the advancement of metagenomics for bioprospecting to stress on its significance in microbial characterization and exploration. This will also ensure the in-depth analysis of several unexplored and unknown microbial communities and/or members from complex niches.

Keywords: environmental DNA; microorganisms; biocatalysts; metagenomic screening

1. Introduction
In recent years, the collection of useful enzymes from different environmental samples has become the focus of many researchers. Owing to the global demand for enzymes, which has risen to 6.9 billion USD in 2017, it is expected to grow up to 10.7 billion USD by 2024 [1]. The eco-friendly nature of these biocatalysts has increased their demand in contrast to the chemical catalysts. It is believed that the catalysts produced by microorganisms will replace 40% of the chemically mediated processes by the year 2030 [2]. The usefulness of enzymes is associated with quality production, low cost, lesser energy consumption, and eco-friendly nature. The applicability of biocatalysts is also considered by their stereospecificity and higher region-
specific nature in the synthesis of organic compounds like alcohols (1°) and carboxylic acids. These associated properties have pushed market demand in the cosmetic, textile, biofuel, pharmaceutical, and food industries. A large fraction of Earth's biomass is represented by microorganisms [3].

The cell estimation of the prokaryotes is estimated to be around $4.6 \times 10^{30}$. Various microorganisms are found in different niches of this Earth, and numerous microorganisms like *E. coli*, *Bacillus subtilis*, *Hansenula polymorpha*, *Pichia pastoris*, *Pseudomonas fluorescens*, *Ralstonia eutropha*, *Saccharomyces cerevisiae*, and *Sorangium cellulosum* have been comprehended for synthesizing different types of enzymes [4]. Microbes inhabit almost every ecosystem like hot springs, deep seas, arid and fertile lands, salt-saturated brines, and acid water mines. The unique member of planctomycetes and myxobacteria families have been isolated from these diverse niches. The study of their DNA, popularly known as e-DNA (Environmental), has shown 0.1% -1.0 % of prokaryotes are cultivable using conventional microbial techniques [5]. In recent times, the collection of biological samples has become more focused on culture-independent techniques, which includes metagenomics and metaproteomics. These are seen as an excellent source of useful novel enzymes.

2. Metagenomics and screening approaches
Metagenomics is a broad term that includes the genetic analysis of the microbial world by initially extracting DNA, followed by cloning and expression. This starts with the genomic DNA isolation from the microbial environment followed by the cloning of e-DNA into an appropriate vector to construct the clone libraries [6]. This subsequently leads to the DNA analysis by experimental modes and in silico (Bioinformatics) [7]. *E. coli* is considered the most widely studied organism concerning metagenomics besides *Pseudomonas putida*, *Streptomyces lividans*, and *Bacillus subtilis* [8]. Most often than not, the cloning step of e-DNA is bypassed. It is subjected directly to the sequencing (Denovo) and then data analysis (In silico) [9].

Considering the source of biocatalysts via metagenomic pathways, it is possible to develop metagenomic libraries from diverse ecosystems. However, the lesser success of metagenomics is associated with the low hit rate of positive clones obtained during the cloning [10]. The pre-selection of the samples to be extracted from the environment to upgrade the libraries is another issue associated with the metagenomic analysis [11]. In the year 2019, chitinase, a metagenome originated biocatalyst, was characterized, and identified from a soil sample having the inhibitive property against chitin [8].

A more enhanced approach to this concept is the targeted metagenomics that involves the artificial manipulation of the microbial world to boost the target function prevalence in situ [12]. The typical examples of this technique include the chitin addition to the sample spots (agricultural land), hemicellulase derivation from the cellulose modified sample places. Incorporating a well-defined isotope called the probe is also an enhanced and innovative technique for library construction [13]. A typical illustration includes the incubation of soil samples with an isotope that is stable and well-characterized. This enriches the eDNA of biomass-degrading bacteria. The metagenomic libraries with abundant genetic reservoirs can be used in two ways: screening which is sequence-driven, and function-driven [14]. It is essential to mention that the screening approach, which is sequence-driven, is used more continuously than the other because of the easily accessible available data of metagenomic sequencing. However, this way of screening is also influenced by genome annotation accuracy, which is seen as an inherent limitation associated with it. Thus, if the sequence similarity does not coincide with the functional domain, it may not give the desired results. Therefore, the screening driven by functional aspect is preferable while taking gene discovery and associated functions into account [15].

2.1. Sequence-based screening
Sequence-based screening depends on the primer designing (probes) for the PCR reaction set up. These may include the genetically coded and/or 16s rRNA enzymes followed by comparing and sequencing with the
stable and know homologous biocatalyst [16]. The sequencing of genomes from the microbial world is always preferred to get an insight into the diversity of microorganisms, associated genes, and the related metabolic pathways. But this is a very tedious process because of the genome length of bacteria (0.5 -1.0 Mbps) [17]. To overcome this shortcoming, the genome is sequenced via 'genome reads', which are small fragments of the genome. Since sequencing of these reads is done randomly, they are assembled and arranged computationally in a library to provide the desired and meaningful results. Also, the sequencing procedures used nowadays are very robust. As a result, sequencing is no longer a time-consuming process [18]. Another difficulty associated with discovering new enzymes is the lesser similarity percentage between the already known enzymes and newly sequenced ORF's, yet this has ensured the discovery of numerous novel enzymes [19].

2.2. Function directed screening
The function-directed metagenomic screening is the method that allows the screening of a specific function. This method has been put into practice in the identification of enzymes involved in vitamin production, antibiotic resistance, and degradation of pollutants [20]. A common example is the discovery of the thermostable Taq enzyme that has been isolated from Thermus aquaticus, originally in one of the famous national parks of the USA, popularly known as Yellowstone national park. Thus, it helps scientists to explore and identify various diversified genomes without having any information about the gene sequence, protein structure, or the origin of the microorganism [21]. For example, using the fosmid library in functional genomics, a researcher may simultaneously look for the gene that a bacterium uses in the degradation of Styrofoam or look for the antibiotic resistance gene [22]. Therefore, this is a complement screening technique to the sequence-based method that helps to identify and characterize novel biocatalysts considering functional aspects. This has further enhanced metagenomics as it permits scientists to have an accessible microbial genome which in turn ensures the screening of the active gene obtained from e-DNA samples. This ensures the easy detection and screening of the previously unknown clones. Even though this approach has been in practice widely in recent times, yet it has certain limitations associated with it like time consumption. This is perhaps because the genes are initially expressed in a specific and suitable host, which is followed by the folding of proteins and analysis of functions [23].

3. Enzyme exploration using metagenomic methodology
There is an utmost interest growing among researchers about the exploration of natural resources to advance genetic assets. Enzymes are one of the common and industrially important compounds that are now being explored from nature via a metagenomic approach. The derived enzymes are very beneficial in the pharmaceutical industry, soap industry, clinical and forensic testing. However, the derivation of biocatalysts from the microbial world is not new [8]. In 1985, Pace and his teammates investigated the direct cloning approach using samples from the environment [24]. Nacke et al. were the first to study function-directed screening who studied the genes encoding cellulase enzymes [25]. This was followed by the work of Rondon et al., who used the BAC vector and created a stable metagenomic library using the soil sample. These enzymes include nucleases, lipases, and amylases [26]. The widely investigated enzymes using metagenomic studies that are industrially important include oxidoreductases, acylases, hydrolases, and lipolytic enzymes. Table1 provides detailed information about the metagenome-derived enzymes.

3.1. Some efficacious enzymes and associated applications
With the advent of metagenomics, many of the enzymes (table 1) derived have been explored for their wide range of applications. The enzyme classes mentioned have a tremendous range of applicability in diverse fields, as discussed below:
Table 1: Novel biocatalysts derived by metagenomic strategies via function-based screening (FBS) and sequence-based screening (SBS).

| Enzyme Class     | Enzyme name     | DNA source         | Host  | Screening Approach | Reference |
|------------------|-----------------|--------------------|-------|--------------------|-----------|
| Lipase/Esterase  | Carboxylesterase| Marine mud         | E. coli| FBS                | [27]      |
|                  | Esterase        | Sediment           | E. coli| FBS                | [28]      |
|                  | Lipase          | Soil               | /     | SBS                | [29]      |
|                  | Sulfatase       | Soil and Rumen     | E. coli| FBS                | [30]      |
|                  | Phosphodiesterase| Coalfield        | E. coli| FBS                | [31]      |
| Cellulase/Hemicellulase | Xylanase      | Compost            | /     | SBS                | [32]      |
|                  | Glucosidase     | Bovine rumen       | E. coli| FBS                | [33]      |
|                  | Endoglucanase   | Algae              | E. coli| FBS                | [34]      |
|                  | Fucosidase      | Soil               | E. coli| FBS                | [35]      |
| Amylase          | Amylopullulnase | Insect gut         | E. coli| FBS                | [36]      |
|                  | α-amylase       | Feces              | E. coli| FBS                | [37]      |
| Protease         | Metalloprotease | Sludge             | E. coli| FBS                | [38]      |
|                  | Serine Protease | Hot Spring         | E. coli| FBS                | [39]      |
|                  | Cysteine Protease| Ovine Rumen      | E. coli| SBS                | [40]      |
| Phytase/Phosphatase | Phytase        | Peat Soil          | /     | SBS                | [41]      |
|                  | Phytase         | Fungus garden      | /     | SBS                | [42]      |

3.1.1. Lipase/esterase.
In the last decades, many libraries of eDNA samples have been analyzed and screened to characterize the hydrolytic lipases/esterases that also help in the ester bond synthesis between alcohol and carboxyl groups [43]. Popovic et al. characterized more than 50 enzymes of this class by functional screening from different terrestrial and marine microcosms. Enzymes that belong to the family of esterase and lipase have certain unique biochemical properties that add to their value concerning industrial applications. They are extensively used in food, cosmetics, detergents, bioremediation, and biodiesel industries. The most common producers of lipases include bacterial Bacillus spp. such as B. licheniformis, B. pumilus, B. alcalophilus and B. subtilis [44].

3.1.2. Cellulases.
Cellulases are known to catalyze the decomposition of cellulose by the process of cellulolysis into the related polysaccharides. The known optimal temperature concerning the catalytic activity of cellulase ranges from 50-55 °C [45]. It is utilized in numerous industries like paper, textile, animal feed, and food. Since it has been seen that cellulase helps to decrease the greying and piling of fabrics with the soil, so it has an imperative role in cleaners and detergents [46]. While exploring the enzyme cellulase through conventional methods, Aspergillus sp. holds the top place for showing high cellulase activity. However, the novel metagenomics methodology has shown the presence of cellulase in numerous organisms. One of the
researchers found cellulolytic microbes like *Cloacibacterium*, *Exiguobacterium*, *Tolumonas*, *Paludibacter* and *Clostridium* with an ability to produce cellulase biocatalysts. Earlier researchers have studied the existence of genes that encode the cellulose biocatalyst in the human intestinal microcosm. The study is further extended to the exploration of the cellulase enzyme from higher temperature places with the hope of identifying the enzymes with thermostable attributes \[1, 47\].

### 3.1.3. Amylases

Amylases are biocatalysts with tremendous applications in the detergent, food, paper, and textile industries. In the last decade or so, the microbial world-derived amylases have replaced the conventionally chemically synthesized α-amylases that hydrolyze starch into sugars \[48\]. This enzyme is used in fruit juices, digestive aids, baking, brewing, and starch syrups. One of the researches carried out on the soil sample of western ghats employing metagenomics for enzymatic studies showed the amylolytic property. At an increased temperature (optimum 60 °C), the enzyme exhibited stability as well \[49\].

### 3.1.4. Proteases

Proteases are the biocatalysts that hydrolyze the peptide bonds which join the amino acid in chains. These are continuously utilized in pharmaceutical, food, detergent, and beverage industries \[50\]. There is a wide spectrum of protease sources available, which include plants, animals as well as microorganisms. One of the best protease-producing microorganisms currently utilized in industries is obtained from *Bacillus sp* \[51\]. The principles of metagenomics have been extended to find more efficient and effective protease-producing organisms in recent times. Biver et al. characterized a gene encoding protease isolated from an organism having similarities with *Desulfobacter postgatei*. Another group of researchers found a gene that is known to encode protease enzyme obtained from sludge. The Prt1A gene-derived protease was found to show optimal activity at 55°C. Till now, protease derived from environmental microcosm includes serine proteases obtained both from groundwater as well as activated tannery sludge \[52\].

### 3.1.5. Phosphatases

Phosphatases are a sub-category of hydrolases that utilize water to cleave monoester phosphoric acid into alcohol and phosphate ions. Phosphate of alkaline nature, known as alkaline phosphatases, are utilized as indicators for pasteurization in the dairy industry \[53\]. Many of the phytases having application in agriculture and animal feed have been recently characterized using the metagenomic methodology. Tan et al. identified and studied phytase from the fungus garden using the metagenomic technique. The characterized enzyme showed a longer half-life at a higher temperature. Besides, alkaline phosphatase of psychrophilic nature has also been studied in sediments of tidal flats via an induced gene expression \[42\].

### 4. Complications in metagenome analysis

One of the fundamental complications associated with the investigation of novel enzymes from environments is the selection of sample locations. The microbial niches play a significant role in the identification as well as characterization of the novel biocatalysts. A peculiar microcosm niche is driven by the environmental interaction with a particular microorganism \[54\]. The presence of cellulolytic bacteria in bagasse sugarcane samples was characterized by Mhuantong et al. and its quantity is more significant than in cellulolytic bacteria \[55\]. The presence of the genes that are known to encode hydrocarbon-degrading enzymes has been reported by Nie et al. in the oily microbial world \[56\]. However, it is very difficult to separate such biocatalysts from the environments of oily nature. Moreover, the region with extreme environmental conditions like hot temperature, and cold weather, demand more accuracy, prevention and caution while exploring the metagenome. The problem of DNA extraction from environmental samples is
also quite challenging and complicated because of the degree of contaminants like humic acid, carbohydrates, and proteins associated with e-DNA [57,58].

5. Conclusion
Metagenomics is a novel approach to genomic studies of the microbial world. This is primarily categorized into two basic approaches: functional metagenomics and structural metagenomics. The former approach involves the construction of genes, screening, expression (heterologous), In-silico analysis, and characterization of the product. The latter approach involves binning, assembly, profiling, prediction of the gene and pathways. Many of the enzymes like lipases, cellulases, and phosphatases have been explored for their numerous benefits. This gained the interest of many researchers to further explore the microbial world to replace the conventional chemical catalysts that are costly and not eco-friendly. This review gives a deeper insight into the basic metagenomic methodologies, associated benefits and the confronted complications associated with these strategies.

Undoubtedly, the metagenomic approach has the potential to unveil the information of novel enzymes and harness the true potential of natural biodiversity that exists in the environment. The major challenge in enzyme industries is to look for an enzyme with improved compatibility, activity against diverse substrates, and enhanced tolerance against stress. Concerning this, there is a need for parallel development and involvement of directed evolution approaches to bring the leap in this field and untangle its real potential. Moreover, an association of bioinformatics, smart screening approaches, sequencing technology and synthetic biology will also provide better insight into enzymes in near future. To attain this, researchers from all the fields need to cooperate and coordinate in the direction, so that, the future platform could be developed for natural product and enzyme discovery.

References
[1] Duan, C.-J.; Feng, J.-X. 2010. Biotechnol. Lett. 32 (12), 1765–75. https://doi.org/10.1007/s10529-010-0356-z.
[2] Bunge, J.; Willis, A.; Walsh, F. 2014. Annu. Rev. Stat. Its Appl. 1, 427–45. https://doi.org/10.1146/annurev-statistics-022513-115654.
[3] Wilmes, P.; Heintz-Buschart, A.; Bond, P. L. 2015. Proteomics. 15 (20), 3409–3417. https://doi.org/10.1002/pmic.201500183.
[4] Katzke, N.; Knapp, A.; Loeschecke, A.; Drepper, T.; Jaeger, K.-E. 2017. Methods Mol. Biol. Clifton NJ. 1539, 159–196. https://doi.org/10.1007/978-1-4939-6691-2_10.
[5] Culligan, E. P.; Sleator, R. D.; Marchesi, J. R.; Hill, C. 2014. Virulence. 5 (3), 399–412. https://doi.org/10.4161/viru.27208.
[6] Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. 1998. Chem. Biol. 5 (10). https://doi.org/10.1016/S1074-5521(98)0108-9.
[7] Lombard, N.; Prestat, E.; van Elsas, J. D.; Simonet, P. 2011. FEMS Microbiol. Ecol. 78 (1), 31–49. https://doi.org/10.1111/j.1574-6941.2011.01140.x.
[8] Hjort, K.; Presti, I.; Elväng, A.; Marinelli, F.; Sjöling, S. 2014. Appl. Microbiol. Biotechnol. 98 (6), 2819–28. https://doi.org/10.1007/s00253-013-5287-x.
[9] Nagayama, H.; Sugawara, T.; Endo, R.; Ono, A.; Kato, H.; Ohtsubo, Y.; Nagata, Y.; Tsuda, 2015. Appl. Microbiol. Biotechnol. 99 (10), 4453–70. https://doi.org/10.1007/s00253-014-6322-2.
[10] Wommack, K. E.; Bhavsar, J.; Ravel, J. 2008. Appl. Environ. Microbiol. 74 (5), 1453–63. https://doi.org/10.1128/AEM.02181-07.
[11] Bharti, R.; Grimm, D. G. 2019. Brief. Bioinform. https://doi.org/10.1093/bib/bbz155.
[12] Emiola, A.; Zhou, W.; Oh, J. 2020. Sci. Adv. 6 (17), eaaaz2299. https://doi.org/10.1126/sciadv.aaz2299.
Knietsch, A.; Waschkowitz, T.; Bowien, S.; Henne, A.; Daniel, R. 2003. *Appl. Environ. Microbiol.* 69 (3), 1408–16. https://doi.org/10.1128/aem.69.3.1408-1416.2003.

Singh, J.; Behal, A.; Singla, N.; Joshi, A.; Birbain, N.; Singh, S.; Bali, V.; Batra, N. 2009. *Biotechnol. J.* 4, 480–494. https://doi.org/10.1002/biot.200800201.

Ngara, T. R.; Zhang, H. 2018. *Genomics Proteomics Bioinformatics.* 16 (6), 405–15. https://doi.org/10.1016/j.gpb.2018.01.002.

Handelsman, J. 2004. *Microbiol. Mol. Biol. Rev.* 68 (4), 669 LP – 685. https://doi.org/10.1128/MMBR.68.4.669-685.2004.

Xu, Y.; Zhao, F. 2018. *Higher Education Press.* Vol. 9. https://doi.org/10.1007/s13238-018-0544-5.

Teeling, H.; Glöckner, F. O. 2012. *Brief. Bioinform.* 13 (6), 728–42. https://doi.org/10.1093/bib/bbs039.

Wooley, J. C.; Ye, Y. 2009. *J. Comput. Sci. Technol.* 25 (1), 71–81. https://doi.org/10.1007/s11390-010-9306-4.

Kunin, V.; Copeland, A.; Lapidus, A.; Mavromatis, K.; Hugenholtz, P. 2008. *Microbiol. Mol. Biol. Rev.* 72 (4), 557–78. https://doi.org/10.1128/MMBR.00009-08.

Verma, S. K.; Singh, H.; Sharma, P. C. 2017. *3 Biotech.* 7 (3), 171. https://doi.org/10.1007/s13205-017-0847-x.

Felczykowska, A.; Dydecka, A.; Bohdanowicz, M.; Gc asior, T.; Soboń, M.; Kobos, J.; Blocz, S.; Nejman-Faleńczyk, B. zena; Wegrzyn, G. 2014. *Microb. Cell Factories.* 13 (1), 105. https://doi.org/10.1186/s12934-014-0105-4.

Rondon, M. R.; August, P. R.; Rzepka, A. D.; Brady, S. F.; Grossman, T. H.; Liles, M. R.; Lotacono, K. A.; Lynch, B. A.; MacNeil, I. A.; Minor, C.; Tiong, C. L.; Gilman, M.; Osburne, M. S.; Clardy, J.; Handelsman, J.; Goodman, R. M. 2000. *Appl. Environ. Microbiol.* 66 (6), 2541 LP – 2547. https://doi.org/10.1128/AEM.66.6.2541-2547.2000.

Zhang, Y.; Hao, J.; Zhang, Y.-Q.; Chen, X.-L.; Xie, B.-B.; Shi, M.; Zhou, B.-C.; Zhang, Y.-Z.; Li, P.-Y. 2017. *Front. Microbiol.* Vol. 8. https://doi.org/10.3389/fmicb.2017.00441.
[55] Mhuantong, W.; Charoensawan, V.; Kanokratana, P.; Tangphatsornruang, S.; Champreda, V. 2015. *Biotechnol. Biofuels* 8, 16. https://doi.org/10.1186/s13068-015-0200-8.

[56] Nie, Y.; Zhao, J.-Y.; Tang, Y.-Q.; Guo, P.; Yang, Y.; Wu, X.-L.; Zhao, F. 2016. *Front. Microbiol.* Vol. 7. https://doi.org/10.3389/fmicb.2016.01254

[57] Garmaeva, S.; Sinha, T.; Kurilshikov, A.; Fu, J.; Wijmenga, C.; Zhernakova, A. 2019 *BMC Biol.* 17 (1), 84. https://doi.org/10.1186/s12915-019-0704-y.

[58] Wani, A. K.; Akhtar, N.; Datta, B.; Pandey, J.; Amin-ul Mannan, M. 2021 *Academic Press*, pp 283–303. https://doi.org/10.1016/B978-0-12-824523-1.00003-1.