Novelty Journals

ISSN 2394-966X

International Journal of Novel Research in Life Sciences
Vol. 8, Issue 4, pp: (5-16), Month: July - August 2021, Available at: www.noveltyjournals.com

Metagenomic and Metabolomic Technologies: Species Identification & Discovery of Novel Biocatalysts from Marine Organisms

SAMAH SULAIMAN[1], BAESHEN, M. N.[2], HAYAM S. ABDELKADER[3]

Department of Biology, College of Science, University of Jeddah, Jeddah, Saudi Arabia.

Abstract: **Marine metagenomics is the study of the genomes of these microorganisms, as sampled directly from the aquatic environment. The great biodiversity of marine natural products offers exciting prospects for drug development, a critical domain of marine biotechnology. In this review, we will highlight how high-throughput techniques such as metagenomics and metabolomics are extremely useful for exploring novel biocatalysts much faster than conventional approaches. Understanding the ecological and biological factors that influence the production of a specific metabolite may be extremely useful for boosting compound yields, optimizing compound extraction, and selecting bioactive compounds. Metabolomics techniques enable the analysis of vast numbers of metabolites, including those generated in trace amounts. Whether the emphasis of microbiome research is biological or environmental, it is obvious that the various Omics approaches provide crucial data. Though each approach provides significant information on its own, we demonstrate that when they are integrated, they generate a full picture. We believe that network-based approaches can allow extensive in-depth analyses of microbiomes, especially when applied with integrated work, and thus offer valuable insight into the microbial ecosystem.**

Keywords: **microbiome, marine metagenomics, metabolomics, bioactive compounds, Omics, novel biocatalysts.**

1. INTRODUCTION

Background:

Miscellaneous microbial communities (bacteria, viruses, archaea and single-celled eukaryote) have vital roles in environment and in human health. Microbes are often difficult for laboratory cultures that can confuse listing members and understanding how communities' function. High-throughput (meta-omics) sequencing techniques and a suite of computational pipelines are combined into shotgun metagenomic sequencing (SMS) ways that transmuted microbiology (Quince et al., 2017). Not long ago, the wide majority of global microbial variations were secluded, obscure and largely underestimated by culture-dependent methods; where the estimated cultivated fraction of the 4-6x1031 prokaryotic genomes moving around the biosphere is about 1-5% (Costa et al., 2020) while the remainder percent of microbes cannot be cultivated easily, thus the use of traditional microbiological culturing methods to study microbes had limited success (Mikaela, 2018). To access and explore unexplored microbiota, the genetic material of the collective cells from an environmental sample is directly extracted. This microbial DNA community (the metagenome) is further analyzed, via modern technologies; as screens of constructed expression libraries and direct high-throughput sequencing. The molecular analysis strategies that examine microbial metagenomes were called metagenomics techniques (Katherine and Pollard, 2020). Also, the broad field may be referred as environmental genomics, ecogenomics or community genomics (Ghosh et al., 2019). Screening of metagenomic libraries allows studying genes and functions from previously inaccessible microbes, opening up exiting new possibilities for the development of novel products (Ekkers et al., 2012).

Metagenomics; the field of functional pure-culture independent and sequence-based analysis of collective genomic analyses of microbial communities contained in an environmental sample where its composition and dynamics depend on external factors (pH, temperature, and salinity) (Srivastava et al., 2013), was developed as a breakthrough for the

weakness of culture-based method to overcome these difficulties; so access and study the genomes of the earlier unavailable microbes (Prayogo et al., 2020). This is as the functional insights gained from to date studies on environmental DNA (eDNA) is largely relied on PCR- or activity-based screening of eDNA fragments cloned in Fosmid or Cosmid libraries. The alternative shotgun meta-genomics holds underexplored potential to discover new enzymes directly from eDNA by avoiding common biases introduced through PCR- or activity-guided functional metagenomics workflows. Without direct links between genotype and phenol-type, inferring new enzyme functions directly from eDNA is alike with searching for a needle in a haystack (Robinson et al., 2021). This review will explore and discuss marine metagenomics, a powerful tool for discovering enzymes and bioactive metabolites.

2. METAGENOMIC TECHNIQUES

The term metagenomics describes both a field of scientific research and a set of techniques that enables the culture-independent analysis of a microbial community in any environmental sample (Culligan et al., 2014). It is a tool for exploring the genetically rich resources of uncultured microbiota, without using conventional culturing methods, and is based on the principle of direct isolation of DNA from a complex environmental sample, containing diverse microbiota; to reveal the true microbial composition of the env-t (Ahmed et al., 2021). It was first coined by Handelsman and his collagues in 1998; when cloning the metagenome of an environmental soil sample to access the collective genomes and biosynthetic machinery of the microbial microflora (Neelakanta and Sultana, 2013). Historically, in 1985, a new branch of microbial ecology was created by Pace and coworkers, via the analysis of 5S and 16S rRNA in environmental samples; to describe the microbial diversity without pure-culturing. As RNA must be directly sequenced and after using PCR technology to amplify the entire genes, discovering various taxa for habitats from the whole universe was accelerated (Mikaela, 2018). These earlier PCR-based techniques have limitations; because of the challenged information details needed to make deep description of the whole complex microbial communities. By major advances in sequencing of whole genomes, chiefly the development of next generation sequencing (NGS), large scale sequencing is revolted and simplified sequence library preparation; by skipping cloning step, and largely drop costs (Jünemann et al., 2017), i.e. NGS made meta-genomic studies more reachable via targeted metagenomics (fixed chosen amplified regions of genomic DNA; as 16S amplicon sequencing) (Amrane and Lagier, 2018). After sequencing, data are gathered into progressively longer vicinal sequences and finally a whole genome is assembled (Klumpp et al., 2012). Although it comes so far since the first complete genomic sequencing, this still has some problems. When sequencing a single organism, assembly is easier and computational gene localization is possible; because there is only one organism to consider. When an entire microbial community is studied, assembling becomes more difficult and not every domain of the entire genome of every single species of the existing microorganism can be sequenced (Land et al., 2015). With the advanced high-throughput technology, sequencing of 16S rRNA gene marker can highlight the taxonomic description of the amazing microbial diversity inhabiting the earth. 16S rRNA-based screening of microbial environmental niches is currently carried by several technologies; including large-scale clonal Sanger sequencing, oligonucleotide microarrays and 454 pyrosequencing that targets specific regions or is linked to bar-coding approach. Interestingly, the short read length, produced by next-generation sequencing, led to new computational efforts in the taxonomic sequence function process (Otlewska et al., 2014).

Metagenomics becomes one of the most important research fields for microbial ecology during the last decade. Starting from 16S rRNA marker gene analysis to identify the community compositions to whole metagenome shotgun sequencing which in turn allows functional analysis and metagenomics applications in a wide spectrum of research areas. The reduced cost along with the increased amount of data, due to the advent of next-generation sequencing, showed rapidly growing demands for bioinformatic software in metagenomics. Nowadays, a large number of the used tools to analyze metagenomic datasets are developed (Jünemann et al., 2017). Hence, metagenomics come a long way since the term was first introduced. The enormous potential of metagenomics to promote both bio exploration and understanding of ecosystems become clear and the efficacy for using the obtained genomic information to search the genomes for novel biocatalysts is of great interest and potential (Culligan et al., 2014).

Also, metagenomics presents an emerging molecular tool to study microorganisms in a population acquired directly from an environmental sample, without the need of a pure culture, via analyzing their DNA as a whole, and has potential applications in industrial settings (Coughlan et al., 2015). Metagenomics helped in understanding the complex diversity of prokaryotes and also helped in understanding the evolution process along with the ecology of microbes. Sequencing and

analysis of total metagenomic DNA informs about several aspects of the sample, providing better identification of microbial life in a given environment. Metagenomics not only indicates identity of the present species, but also insights into the metabolic activities and functional roles of the present microbes in a given population (Langille et al., 2013). It is a powerful technique with a big limit and the potential for novel gene discovery that accelerates basic research by investigating the fundamentals of microbiology, and provides opportunity for to develop novel biotechnological processes and products. The expanded discovery of novel biocatalysts from environmental samples is a result of successful implementation of metagenomics (Ahmad et al., 2019). As this technique is in the infancy stage, a large number of technical issues, related to direct screening of the potential enzyme coding genes and the availability of suitable host-vector combinations for successful expression of metagenome- derived enzymes, limit the great potential of functional metagenomics (Coughlan et al., 2015). Efficiency and efficacy of metagenomics to discover industrially strong biocatalysts or enzymes with novel substrate specificities, biological activities, and molecular stabilities with a strong potential of subsequent product development is enhanced by discovering more efficient high-throughput screening methods (Wang et al., 2019).

High-throughput sequencing approaches enable genomic analyses of all microbes in a sample, not just those that are amenable to cultivation (Quince et al., 2017). One such method, shotgun metagenomics, is the untargeted (shotgun) sequencing of all (meta-) microbial genomes (genomics) present in a sample. Shotgun sequencing is used to profile taxonomic composition and functional potential of microbial communities and to recover whole genome sequences. Approaches; as high-throughput 16S rRNA gene sequencing, which profile selected organisms or single marker genes, are sometimes referred: metagenomics, but this is a misnomer; as they do not target the entire genomic content of a sample (Olson et al., 2019).

Discoveries aided by metagenomics technology include the identification of environmental bacterial phyla with endosymbiotic behavior and species that can carry out complete nitrification of ammonia as same as the widespread presence of antibiotic genes in commensal gut bacteria' tracking of human outbreak pathogens' the strong association of the viral and bacterial fractions of microbiome with inflammatory bowel diseases and the ability to monitor strain-level changes in the gut microbiota after perturbations; as those induced by fecal microbiome transplantation (Ahmad et al., 2021). Metagenomics enables the study of all microorganisms, regardless of whether they can be cultured or not, through the analysis of genomic data obtained directly from an environmental sample, providing knowledge of the present species and allowing the extraction of information regarding the functionality of microbial communities in their natural habitat (Coughlan et al., 2015).

A typical shotgun metagenomics technique (**Fig. 1**) comprises the following five steps, after the initial study design and experimental protocol (Quince et al., 2017; Prayogo1 et al., 2020). These are: i) collection, processing and sequencing of the samples, ii) preprocessing of the sequencing reads, iii) sequence analysis to profile taxonomic, functional and genomic features of the microbiome (should comprise a combination of read- and assembly-based approaches; depending on the experimental objectives); iv) statistical and biological postprocessing analysis (various multivariate statistical techniques can be used to interpret data), and v) validation; where conclusions from high-dimensional biological data are susceptible to study-driven biases; so follow-up analyses are vital. Numerous experimental and computational approaches are available to carry out each step. Despite apparent simplicity, shotgun metagenomics has limitations, due to potential experimental biases and the complexity of computational analyses and interpretations (Bolger et al., 2014).

Otherwise, metagenomic analysis involves the basic steps: 1) selection of an environmental niche, 2) isolation of genetic material directly from an environmental sample, 3) manipulation of the genetic material, 4) library construction, and 5) analysis of genetic material in the metagenomic library (**Fig.2**). The screening of clones is done for phylogenetic markers, by hybridization or multiplex PCR (for other con-served genes or for expression of specific traits; as enzyme activity, or antibiotic production, or can be randomly sequenced) (Ahmad et al., 2019).

Marine organisms can produce enzymes with unique properties, compared to those of the same enzymes of terrestrial organisms. Metagenomics, or the culture-independent cloning of environmental DNA obtained directly from a sample, has revealed a plethora of data on the uncultured microbial world. It facilitated the discovery of novel marine biocatalysts by allowing researchers to probe directly into a huge diversity of enzymes within natural microbial communities. By metagenomics, the community structure; including species richness and distribution, and also metabolic pathways (**Fig. 1**)

can be analyzed (Kumar et al., 2015). In comparison with construction of metagenome libraries, screening strategies are more complex and technically demanding. The commonly employed methods for metagenomics library construction involve: i) function-based, ii) sequence-based, iii) substrate-induced gene-expression (SIGEX), and iv) compound configuration screenings (Ahmad et al., 2019); of which function- and sequence-based are which are based on sequence or function of the enzymes. Each approach has certain strengths and limitations, but has enriched the knowledge of the uncultured microbial community by unmasking information about the groups of prokaryotes that are otherwise a mystery (Ngara and Zhang, 2018). The importance of genome mining and discovering novel enzymes from marine microbes and their biotechnological relevance have been highlighted by Bramhachari et al., (2020). Metagenomics has proven to be a new field of research by elucidating uncultured microbes which represent the vast majority of organisms in marine environments. By applying metagenomics, it is possible to investigate resources for the development of novel enzymes that have biotechnological applications. It represents an alternative approach to conventional microbial screening; allowing meticulous screening of microbial genomes in marine environments. Metagenomic libraries are essential tools to discover novel enzymatic activities, facilitating genetic tracking for innovative biotechnological applications of interest. Although its many limitations, metagenomics can provide scientific insights along with economic benefits, especially in industry.

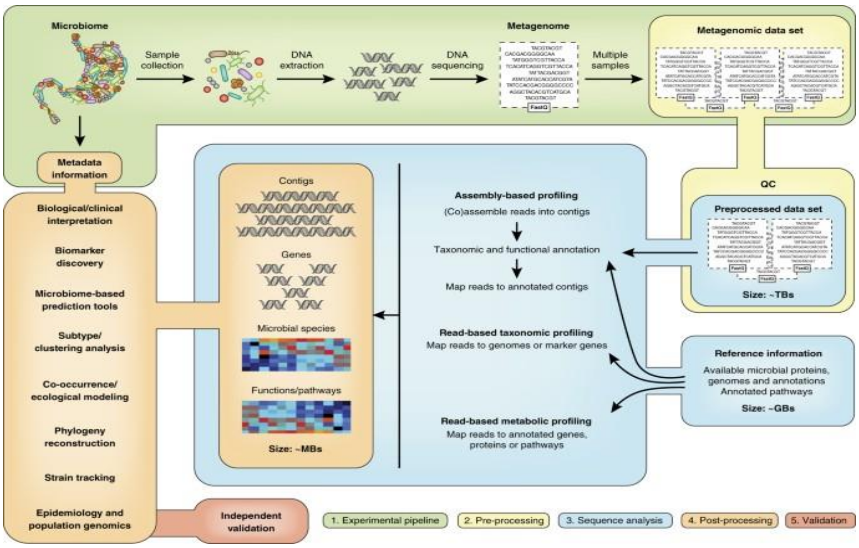

Fig. 1. Summary of a metagenomics workflow.

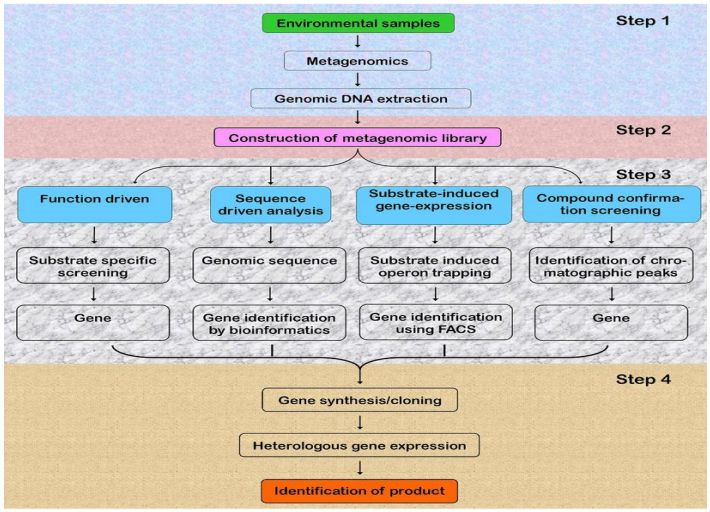

Fig. 2. Schematic representation of metagenomic analysis and identification of novel biomolecules.

(Prayogo1 et al., 2020)

International Journal of Novel Research in Life Sciences
Vol. 8, Issue 4, pp: (5-16), Month: July - August 2021, Available at: www.noveltyjournals.com

3. METABOLOMIC SCREENING

In the function/homology-based approach, screening is done by expressing a particular enzyme or product; according to principle of discriminating biochemical activities of enzyme catalyzed reactions (formation of zone of clearance around the colonies), and detecting a variety of novel genetic elements with functions that are discrete from the known enzymes. Function-driven screening is the preferable, when it comes to discovering genes with novel functions or exploring the sequence diversity of protein families with certain functions (Ngara and Zhang, 2018), i.e. The sequence-driven approach identifies genes; based on homology in databases and is classically performed by using PCR-primers constructed from known conserved regions (usually enzymes ´catalytic sites). With the rapid technology of NGS it is now possible to screen whole metagenomes; using bioinformatic approaches, called in silico screening that helps to avoid the tiring library constructions (Berini et al., 2017). Though it secures the entire form of gene or gene cluster essential for the desired trait, it has several limitations; as to satisfy the expression of the functionality of the gene of interest (transcription), translation, folding and secretion in the heterologous hosts (*E. coli*), and gene clustering (Ahmad et al., 2019). Although, the major limitation of impossibility to correctly annotate a novel gene or function; as if it is truly novel, it will not be present in a database, using this approach makes the dependency of gene expression in a host to be alleviated (Cheng et al., 2017).

Function-based screenings, following cloning and expression of metagenomic DNA in a heterologous host, are used to find out novel proteins of industrial interest encoded with the aid of using the genes of previously inaccessible microorganisms (Mikaela, 2018). Functional metagenomics (**Fig. 3**) is a great potential in the food and pharmaceutical industries; because it aids in: 1) identity of enzymes with desirable technological properties, capable of catalyzing novel reactions or replacing current chemically synthesized catalysts which can be difficult or expensive to provide and capable of work beneath neath an extensive range of environmental conditions encountered in food and pharmaceutical processing cycles which include extreme situations of temperature, pH, osmolarity, etc, 2) discovering novel bio-actives which include antimicrobials active against microorganisms of concern both in food and clinical settings, and 3) investigation of industrial and societal problems along with antibiotic resistance development (Coughlan et al., 2015). Badie and his coworkers confirmed useful metagenomics to be an effective experimental technique for analyzing gene function in uncultivated microbial communities; because it includes isolation of environmental DNA immediately from microbial communities and examination of the functions of the corresponding encoded proteins. This technique allowed the invention of novel enzymes for which functional prediction couldn't be carried out primarily based only on DNA sequence. Using such a technique, they isolated a gene encoding a novel thioredoxin reductase enzyme from a water sample taken from the LCL of the Red Sea Atlantis II brine pool. Thioredoxin reductase (TrxR) performs an essential function in retaining redox stability and counteracting oxidative stress inside cells (Badie et al., 2019).

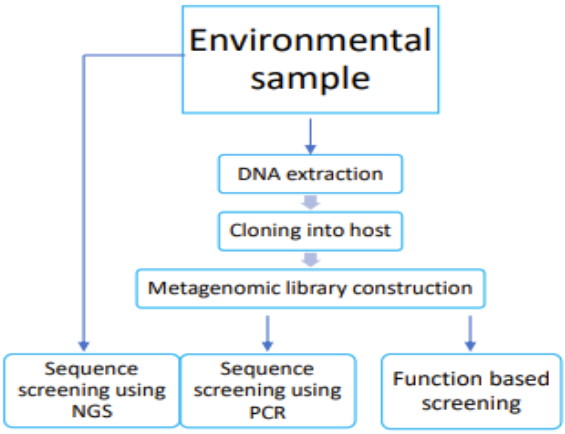

Fig. 3. Schematic metagenomic approaches; sequence- and functional-driven.

Sequence/activity-based screening is based on the principle of polymerase chain reaction; PCR or Southern hybridization; this involves the use of DNA probes isolated from known samples encoding protein families of interest to detect the desired gene of interest (Kumar et al., 2015). Limiting heterologous expression is dominant, and DNA analysis does not require a specific database. Sequence identity does not guarantee a complete gene or gene cluster (Ahmad et al., 2019). Both function- and sequence-driven screening strategies are exhausting and time-consuming. To overcome this,

Uchiyama and Watanabe (2007) proposed an opportunity screening technique (SIGEX) for a groundwater metagenome library containing fragrant hydrocarbon brought about genes; to enhance the frequency of screening (Kumar et al., 2015). SIGEX; the used method to expose novel catabolic operons from metagenomes, works through a promoter trap technique brought about through a substrate and single-cellular sorting of clones the use of fluorescence-activated cellular sorting (FACS); on foundation of the reality that the substrate-inducible gene expression is managed through regulatory factors placed proximately; so, it monitors the clones harboring desired substrate-brought about catabolic genes (Meier and Goker 2019). An operon-trap vector; p18GFP, became designed by Uchiyama and co-workers in 2010; for a high-throughput process, for the construction of metagenomic libraries. The catabolic genes expression of cloned metagenomic DNA became evaluated by green fluorescent protein; GFP, co-expression in presence of appropriate enzyme substrate, and the positive clones have been differentiated and selected by using FACS (Taupp et al., 2011). This is a very rapid and low-cost method; however, it is very perceptive in terms of co-orientation of the preferred genes and bioavailability of the enzyme substrates inside the cell cytoplasm, and it gives the technical know-how to set up and analyze FACS (Ahmad et al., 2019).

The compound configuration screening (CCS) method; the most advanced and indirect selection technique for specific gene products, helps to identify specific metabolites; through chromatographic analysis during mass spectral analysis, which is different from that of host materials. Screening relies on the capability of clones to produce novel compounds in various chromatographic peaks in comparison with the host cell. Although effectual, CCS of clone screening and identification is cost prohibitive and labor intensive (Zhang and Huang, 2012).

Various methods had been devised in metagenomics for gene finding and discovery of many biocatalysts. Various efforts and achievements were made in the industries for finding new molecule with diverse functions. Recently, drug discovery has been done based on the metagenomic approaches and also on finding the microbial diversity in the marine organisms (Wang et al., 2020).

One of the major advantages of metagenomics is the access of genetic information about the uncultivated majority of microbes that still largely lack functional characterization, hence metagenomics can reshape our view of the tree of life (Parks et al., 2018) and led to the identification of deeply rooted and metabolically-diverse line-ages; as the DPANN archaea and candidate phyla radiation (Jaffe et al., 2020). Further, metagenomics revealed remarkable biosynthetic potential of many unculti-vated microbial phyla; including *Candidatus tectomicrobia*, *Eelbacter* and *Angelo-bacter*. In case of *Ca. tectomicrobia*, heterologous expression enabled the experi-mental characterization of new biosynthetic pathways and products, while, the tantalizing promises of discovering new enzymology from metagenomes goes hand-in-hand with the challenges of working with DNA from organisms that have eluded laboratory cultivation (Crits-Christoph et al., 2020).

The new technology known as metagenomics is now being utilized to study novel enzyme genes, and which biocatalysts yield better biocatalysts in respect to the known enzymes. Its potential is assured by the enormous biodiversity of both culturable and unculturable microbes in all environments of the earth (Berini et al., 2017). Screening gene-specific Amplicons from metagenomes (S-GAM) is a powerful technique to efficiently isolate target genes from metagenomes, and can overcome a major disadvantage of previous techniques; the low efficiency in obtaining target genes from metagenomes. Also, S-GAM can omit the time-consuming subcloning and expression optimization procedures (Itoh et al., 2016). S-GAM is a short-cut way to locate the treasure map of undiscovered natural proteins. S-GAM generates chimeric and artificial gene products, pointing to the likelihood of significant advancements via directed evolution. Otherwise, homologous sequences; as HPARs, which can share more than 97% amino acid identity with the original PAR enzyme and were considered unimportant, are very useful for engineering superior enzymes, hence metagenomics approach including S-GAM, will be the next-generation method for biocatalysis (Itoh, 2017).

To date, only a limited number of metagenome-derived enzymes are optimized through directed evolution techniques, indicating a vast scope for further improvement of enzyme attributes by protein engineering to fulfill the requirements of various industrial processes (Mazurenko et al., 2020). Therefore, it is necessary to identify more suitable host-vector expression systems to construct functionalized metagenomics libraries and the discovery of fast, economical, and efficient direct selection methods for strengthening biocatalyst discovery (Sebastian et al., 2013). With these developments, metagenomics may emerge as a powerful means of exploring microbial biodiversity present in the natural econiche with

previously unrecognized biological attributes, and identifying novel enzymes with robust properties and a broad substrate specificities that may pave their way to tremendous biotechnological potential (Ahmad et al., 2019).

Although several novel enzymes of industrial importance have been discovered using this method it is still primarily used for mapping functionality and metabolic pathways. Function-based screening is the only approach which is useful for finding truly novel gene classes or functions since sequence information is not needed for comparative reasons (Mikaela, 2018). Additionally, while metagenomics has led to discover and synthesize numerous biologically significant compounds; as polyketide synthase, nonribosomal peptide synthetase, antibiotics, and biocatalyst and offers different advantages over conventional sequencing techniques, they also have certain limitations including bias classification, non-availability of quality DNA samples, heterologous expression, and host selection (Dhanjal et al., 2020). The assimilation of advanced amplification and screening methods; as φ29 DNA polymerase, NGS, Cosmids and recent bioinformatics´ tools; as automated genome mining, anti-SMASH, provide promising results to overcome these constrains. Consequently, functional genomics and bioinformatics along with synthetic biology are crucial for the success of the metagenomic approach and for exploring new possibilities among the microbial consortia for the future drug discovery process (Mahapatra1 et al., 2019).

The use of metagenomics for discovery of novel enzymes is of great interest and is growing. There are several approaches to use metagenomics for enzyme discovery (Kamble et al., 2018). However, there are, also still a lot of limitations; as the very low hit rates making the work tedious, time consuming and costly and the gain often very low. Metagenomics also has the disadvantage of looking into the genome, thus, only studying the genetic potential, not which proteins the members of a microbial population actually express (Ahmad et al., 2021). Metagenomics does not provide any information about what genes are actually active under a given condition or metabolic need; thus, by only extracting and analyzing the genome from microorganisms thriving in a habitat which have similarities to the conditions needed in industrial processes, some crucial information is not obtained. Simply because a gene coding for an enzyme of a certain activity is present and identified in a metagenome, this does not mean that the gene is actually expressed and used under the conditions under which the metagenome was collected. To find the proteins that are expressed under certain conditions and needs the proteome or more specifically the metaproteome is needed (Roux et al., 2021). The potential degradation and utilization of chitin "the polymer of (14)-linked N-acetylglucosamine; GlcNac and its derivatives by marine sponges, octocorals, sediments, and seawater biotopes microbiomes have been studied by Raimundo et al., (2021), and determine whether chitin degrading assemblages in these microbiomes are taxonomically and metabolically distinct. Forty-one bacterial isolates from two previously published culture collections were utilized to investigate the chitin degrading capabilities of culturable symbionts of octocorals and marine sponges, *Sarcotragus spinosulus*, and *Ircinia variabilis*. The isolates were subjected to chitin degradation and chitinase activity bioassays, PCR amplification of chitinase-encoding genes, and Pfam-based annotations to deduce the protein domains involved in chitin and chitooligosaccharides (COSs) degradation, chitin deacetylation, and GlcNac utilization (when genome sequences were available) that underpin. The sponge metagenome dataset contains four microbiomes from the interior body of four separate *Spongia officinalis* specimens, as well as three samples from saltwater and sediments. The dataset containing the octocoral metagenome comprises 13 microbial metagenomes, all from healthy tissues of three octocoral species: (3 healthy tissues, 3 necrotic tissues of *Eunicella gazella*, 3 healthy *Eunicella verrucosa*, and 3 healthy *Leptogorgia sarmentosa*). All samples, from both datasets, were taken in the same area near the coast of Algarve, Portugal. The results showed that the Aquimarina, Enterovibrio, Microbulbifer, Pseudoalteromonas, Shewanella, and Vibrio decreased colloidal chitin in vitro. The different chitin breakdown routes, use, and turnover suggest that microniches that have been fed upon by many species of chitin-utilizing organisms degrade chitin to enable them to coexist. Additionally, chitin metabolism may indicate the integrity of the microbiome in corals and possible novel chitinolytic enzymes in the species Aquimarina may have been discovered. Vibrio and Aquimarina species contain numerous endochitinase-encoding genes beneath their capacity to cleave massive chitin polymers into di- and oligomers. N-acetylglucosamine-using Alphaproteobacteria were specified (a chitin monomer). Numerous full-length endochitinase genes were found in both Aquimarina and Vibrio strains, providing evidence of highly capable chitin catabolism. Chitin metabolism-related genes are heavily associated with polysaccharide deacetylases in endosymbiotic consortia of *Spongia officinalis*, showing that the marine sponge microbiome may transform chitin into its deacetylated- and biotechnologically-valuable-chitosan. Endochitinase and chitin-binding protein-encoding genes were abundant in healthy octocorals but were drastically decreased in necrotic octocorals. Uncultivated Gammaproteobacteria and Chloroflexi symbionts have been found to be

significant for chitin digestion in marine sessile invertebrates. Robinson et al. (2021) studied a roadmap for metagenomic enzyme discovery. The authors cover both computational and experimental strategies to invest metagenomes and explore protein sequence space with a spot-light on natural product biosynthesis. Specifically, they compare in silico methods for enzyme discovery; including phylogenetics, sequence similarity net-works, genomic context, 3D structure-based approaches and machine learning techniques. Also, they discuss various experimental strategies to test computational predictions including heterologous expression and screening. Additionally, they provide an outlook for future directions in the field with an emphasis on Meta omics, single-cell genomics, cell-free expression systems and sequence-in-dependent methods. Surprisingly, there was a paucity of studies of de novo and reference-based enzyme discovery studies that have used shotgun metagenomics rather than functional metagenomics. Even while gathering the petabytes of Meta omics data in public databases, there was a disconnection between the relative ease of next-gene-ration sequencing and the difficulty of gaining insights into new protein families and their functions. Costa et al., (2020) discovered prospective glycoside hydrolases (GHs) microbial producers and GHs genes with biotechnological potential by the application of a microbial exopolysaccharide (EPS). GHs are highly desired and essential in the manufacturing of products and biofuels, as well as in the eradication of hazardous biofilms and EPS. The EPS-complex structure (WH15EPS) of *Acidobacteria granulicella* sp. strain WH15 was investigated in both cultivation-independent and cultivation-dependent approaches as an enrichment factor. The results revealed that cultivation-independent methods; like metagenomics, are tools to explore and discovery of biotechnological compounds produced by microbes in natural environments. SIP metagenome study of soil litter revealed that the most common taxa were *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, and *Planctomycetes* in both modified and unamended WH15EPS treatments. The enrichment cultures in solid culture medium coupled to metagenomics elucidated an enrichment in *Proteobacteria* and the metagenome assembly of this enrichment cultures resulted in 4 metagenome-assembled microbial genomes (MAGs), with low identity (42-86%) to known microorganisms. The glycoside transferase (GT) family was the most prevalent among all carbohydrate-active enzymes (CAZymes) genes retrieved, both in culture-independent and culture-based metagenome datasets. Among all carbohydrate-active enzymes (CAZymes) genes recovered, the glycoside transferase (GT) family was the most frequent, both in culture-independent and culture-based metagenome datasets. In both metagenome datasets, the glycoside hydrolase (GH) family GH13 was the most abundant. Those GH families are affiliated to microorganism that can degrade WH15EPS and potentially applicable for biofilm deconstruction. According to culture-based metagenomic analysis, the assembled four MAGs (unclassified Proteobacteria) comprised GH-related families such as mannosidases, lysozymes, galactosidases, and chitinases. Consequently, WH15EPS functional diversity, in both culture-dependent and -independent approaches was enriched in GHs; as amylases and endoglucanases that have chemical, pharmaceutical and food industrial applications. Furthermore, WH15EPS may be used to explore and identify previously unknown taxa, such as unclassified Proteobacteria and Planctomycetes, therefore expanding the number of currently cultured bacteria with biotechnological potential.

4. CHEMICAL CLASSES OF NATURAL PRODUCTS ISOLATED FROM MARINE ORGANISMS

The metagenomic diversity of the Red Sea organisms and their natural products repertoire and pharmacological activities have been studied by El-Hossary et al., (2020). The literature search covering the time period until December 2019, with the inclusion of extra five years (from 2015 to 2019). According to the study design, the marine organisms were collected from the Red Sea coasts of seven countries (Egypt, Saudi Arabia, Palestine, Eritrea, Djibouti, Jordan, and Yemen). The results showed that the Red Sea is a rich and diverse ecosystem with 2000 km of coral reef extending along its coastline. Over 1000 invertebrate species and 200 soft and hard corals make this habitat. Due to this high biodiversity, the Red Sea is a promising underexplored habitat for the discovery of new bioactive marine natural products. Out of these, a total of 677 marine natural products were isolated from Red Sea marine organisms, with almost 60% of these compounds coming from reports over the last decade (2011-2019). This significant increase of reported compounds highlights the enormous potential of marine organisms, particularly from the Red Sea. Further, the diversity of these chemical structures is quite remarkable. Majorly and more interesting of these compounds are terpenes and alkaloids, but other classes as sterols, peptides and other nitrogenous compounds, polyketides, fatty acids, macrolides, quinones, polyacetylenes, and flavonoids are included (**Fig. 4**). The structures and biological activities of selected 111 marine natural products, due to their potent biological activities, are categorized according to their isolation source (marine organism); including micro-organisms, invertebrates, algae, and seagrasses. These compounds exhibited a wide range of potent biological activities; as antioxidant, anticonvulsant, anticancer, and anti-infective activities. However, these biologically active marine natural

products were found to be not toxic on normal cells; thus, could be promising compounds to develop new drugs after further investigations. Among the organisms collected from the Red Sea, marine sponges remain the most widely sampled. Also, gram-positive bacteria as actinomycetes and fungi isolated from various marine organisms from the Red Sea were shown to be the prolific producers of these bioactive natural products. Furthermore, most of the reported marine organisms were collected from Egypt and Saudi Arabian coasts, accounting for 58% and 16% respectively; due to the high biodiversity of organisms in their environments compared to other locations. Moreover, more research should be directed to explore other environments.

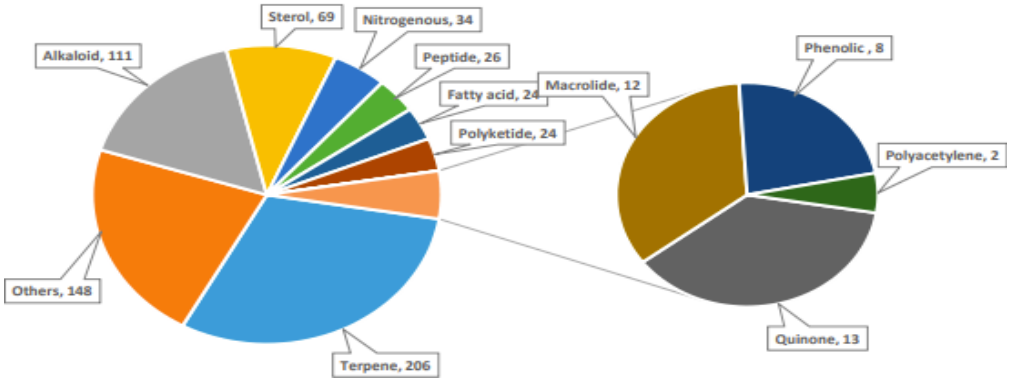

Fig. 4. Chemical classes of natural products isolated from the Red Sea marine organisms.

As shown in **Fig. (5)**, metagenomics is divided into two primary studies; structural and functional. While the basic structural metagenomics method consists of assembly, binning, and microbial community analysis; as taxonomic profiling, gene prediction and metabolic pathways, functional metagenomics consists of gene construction, screening, heterologous overexpression, bioinformatic analysis and protein products identification. To understand the applications and basic methods of metagenomics and its impact on increasing discoveries about the information of the microbial community and utilization to novel enzymes exploration from nature, especially cellulases, proteases and lipases Prayogo et al., (2020) deeply discussed the metagenomic applications and development of novel enzymes from nature. Novel enzymes exploration with a metagenomic technique revealed several novel enzymes from nature (reaching of 99% and various types of genes encoding an enzyme that still not identified); like cellulases, proteases, lipases, trans-aminases, chitinases, AHL-lactonase, bleomycin resistance dioxygenase (BRPD) and oxoflavin-degrading enzyme, Metagenomics helped researchers uncover novel enzymes from nature that are beneficial to the industries, thus understanding meta-genomic and its application impact the useful technological developments.

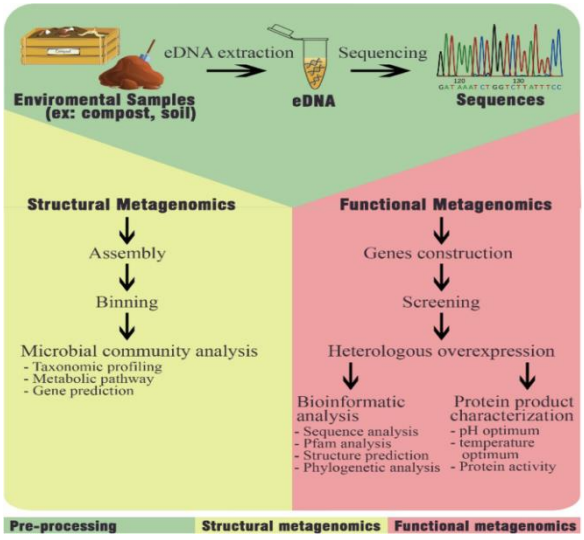

Fig. 5. Framework for metagenomics with two primary studies; structural and functional.

In conclusion, metagenomics technology has become increasingly important for mining the enzymatic diversity found in a variety of natural habitats, enabling the rapid discovery of novel enzymes for biomass breakdown and biofuel production. Functional genomics and bioinformatics, as well as synthetic biology, are key components of the success of metagenomic techniques and the investigation of novel possibilities within microbial consortia throughout the future drug discovery process. Thus, functional metagenomics enables the development of novel innovative products that add value to industrial processes.

REFERENCES

[1] **Ahmad, T.,SarupSingh, R., Gupta, G., AnshulaSharma, A., BaljinderKaur, B. (2019).** Metagen-omics in the search for industrial enzymes. Chap. 15, In: Biomass, Biofuels, Biochemicals: Advances in enzyme technology, Singh, R.S.,Singhania, R.R., Pandey, A., Larroche, C. (edns.), Elsevier, 419-52.

[2] **Ahmad, T., Gupta, G., Sharma, A., Kaur, B., El-Sheikh, M.A., Alyemeni, M.N. (2021)**. Metagenomic analysis exploring taxonomic and functional diversity of bacterial communities of a Himalayan urban fresh water lake. PloS one, 16 (3), e0248116, 19 pages.

[3] **Amrane, S. and Lagier, J. C. (2018).** Metagenomic and clinical microbiology. Hum. Microbio.J. 9, 1-6.

[4] **Badiea, E.A., Sayed, A.A., Maged, M., Fouad, W.M., Said, M.M. and Esmat, A.Y. (2019).** A novel thermostable and halophilic thioredoxin reductase from the Red Sea Atlantis II hot brine pool. PLoS One. 2019 May 31;14(5): e0217565.

[5] **Berini, F., Casciello, C., Marcone, G.L. and Marinelli, F. (2017).** Metagenomics: Novel enzymes from non-culturable microbes. FEMS Microbiol. Lett., 364 (21), 19 pages.

[6] **Bolger, A.M., Lohse, M. and Usadel, B. (2014).** Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform., 30, 2114-20.

[7] **Bramhachari, P.V., Sheela, G.M. and Satish, T. (2020).** Perspectives for novel enzyme discovery from marine environments through genome mining and metagenomics. Chap. 109, In: Encyclopedia of marine biotechnology, Kim, S.K. edn., John Wiley & Sons Ltd., Wiley Online Library 2423-42.

[8] **Costa, O.Y.A., de Hollander, M. and Pijl, A. (2020).** Cultivation-independent and cultivation-dependent metagenomes reveal genetic and enzymatic potential of microbial community involved in the degradation of a complex microbial polymer. Microbiome, 8, 76, 19 pages.

[9] **Coughlan, L.M., Cotter, P.D., Hill, C. and Alvarez-Ordóñez, A. (2015).** Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Front. microbiol., 6, 672, 22 pages.

[10] **Crits-Christoph, A., Olm, M. R., Diamond, S., Bouma-Gregson, K., and Banfield, J. F. (2020).** Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. ISME J., 14 (7):1834-46.

[11] **Culligan, E.P., Sleator, R.D., Marchesi, J.R., and Hill, C. (2014).** Metagenomics and novel gene discovery: promise and potential for novel therapeutics. Virulence, 5 (3), 399-412.

[12] **Dhanjal, D.S., Chopra, C., and Chopra, R.S. (2020).** Metagenomic DNA sequencing: Technological advances and applications. Chap. 3, In: Metagenomics: Techniques, applications, challenges and opportunities, Chopra, R.S., Chopra, C., Sharma, N.R. (edns.), Springer, 37-53.

[13] **Ekkers, D.M., Cretoiu, M.S., Kielak, A.M. and van Elsas, J.D. (2012).** The great screen anomaly: A new frontier in product discovery through functional metageno-mics. Appl. Microbiol. Biotechnol., 93, 1005-20.

[14] **El-Hossary, E.M., Abdel-Halim, M., Ibrahim, E.S., Pimentel-Elardo, S.M., Nodwell, J.R., Handoussa, H., Abdelwahab, M.F., Holzgrabe, U., and Abdelmohsen, U.R. (2020).** Natural products repertoire of the Red Sea. Mar. Drugs, 18, 457, 28 pages

[15] **Ghosh, A., Mehta, A., and Khan, A.M. (2019).** Metagenomic analysis and its application. In: Encyclopedia of bioinformatics and computational biology, Shoba Ranganathan, Michael Gribskov, Kenta Nakai, Christian Schönbach (edns.), Academic Press,184-93.

[16] **Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. and Goodman, R.M. (1998).** Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 5, R245– R249.

[17] **Itoh, N., Kazama, M., Takeuchi, N., Isotani, K., and Kurokawa, J. (2016).** Gene-specific amplicons from metagenomes as an alternative to directed evolution for enzyme screening: A case study using phenylacetaldehyde reductases. FEBS Open Bio., 6, 566-75.

[18] **Jaffe, A.L., Castelle, C.J., Matheus Carnevali, P.B. Carnevali, M., Gribaldo, S., and Ban-field, J.F. (2020).** The rise of diversity in metabolic platforms across the Can-didate Phyla radiation. BMC Biol., 18 (1): 69.

[19] **Jünemann, S., Kleinbölting, N., Jaenicke, S., Henke, C., Hassa, J., Nelkner, J., Stolze, Y., Albaum, S.P., Schlüter, A., Goesmann, A., Sczyrba, A., and Stoye, J. (2017).** Bioinform-atics for NGS-based metagenomics and the application to biogas research. J. Biotechno., 261, 10-23.

[20] **Kamble, P. and Vavilala, S. (2018).** Discovering novel enzymes from marine ecosystems: A metagenomic approach. Botan. Mar., 61 (2), 161-75.

[21] **Klumpp, J., Fouts, D. E. and Sozhamannan, S. (2012**). Next generation sequencing technologies and the changing landscape of phage genomics. Bacteriophage, 2 (3):190-9.

[22] **Kumar, V., Maitra, S.S. and Shukla, R.N. (2015).** Environmental metagenomics: The data assembly and data analysis perspectives. J. Inst. Eng., India Ser. A (Civil, Architectural, Environmental and Agricultural Engineering), 96, 71-83.

[23] **Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S. and Ussery, D.W. (2015).** Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics., 15 (2):141-61.

[24] **Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G. and Huttenhower, C. (2013).** Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 31(9):814-21.

[25] **Mahapatra, G.P., Raman, S., Nayak, S., Gouda, S., Das, G. and Patra, J.K. (2020).** Metagenomics approaches in discovery and development of new bioactive compounds from marine Actinomycetes. Curr. Microbiol., 77, 645-56.

[26] **Mazurenko,S., Prokop, Z. and Damborsky, J. (2020).** Machine learning in enzyme engineering. ACS Catal., 10 (2):1210-23.

[27] **Meier-Kolthoff J. P. and Goker M. (2019).** TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. 10:2182.

[28] **Mikaela, J. (2018).** Metaproteogenomics-guided enzyme discovery: Targeted identification of novel proteases in microbial communities: Linköping Studies in Science and Technology. PhD. Dissertation, Department of Physics, Chemistry and Biology, Faculty of Science and Engineering, Linköping University, LiU-Tryck, Linköping, Sweden. http://www.ep.liu.se.

[29] **Neelakanta, G. and Sultana, H. (2013)**. The use of metagenomic approaches to analyze changes in microbial communities. Microb. Insi.,6, 27-48.

[30] **Ngara, T. R. and Zhang, H. (2018).** Recent advances in function-based metagenomic screening. Genom. Proteom. Bioinform., 16 (6): 405-15.

[31] **Olson,N.D., Treangen, T.J., Hill, C.M., Cepeda-Espinoza, V., Jay Ghurye, Koren, S. and Pop, M. (2019)**. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. Brief Bioinform., 20 (4):1140-50.

[32] **Otlewska, A., Adamiak, J., and Gutarowska, B. (2014).** Application of molecular techniques for the assessment of microorganism diversity on cultural heritage objects. Acta Biochim. Pol., 61 (2): 217-25.

[33] **Pace, G.M., Ivancic, M.T., Edwards, G.L., Iwata, B.A., and Page, T.J. (1985).** Assessment of stimulus preference and reinforcer value with profoundly retarded individuals. J Appl Behav Anal. 18(3):249–255.

[34] **Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P. and Tyson, G.W. (2017).** Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2(11):1533-1542.

[35] **Prayogo, F.A., Budiharjo, A., Kusumaningrum, H.P., Wijanarka,W., Suprihadi, A. and Nurhayati, N. (2020).** Metagenomic applications in exploration and development of novel enzymes from nature: A review. J. Genet. Eng. Biotechnol., 18, 39.

[36] **Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J. and Segata N. (2017).** Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol., 135 (9):833-44.

[37] **Raimundo, I., Silva, R., Meunier, L., Valente, S.M., Lago-Lestón, A., Keller-Costa, T. and Costa, R. (2021).** Functional metagenomics reveals differential chitin degradation and utilization features across free-living and host-associated marine microbiomes. Microbiome 9, 43.

[38] **Robinson, S.L., Piel, J. and Sunagawa, S. (2021).** A roadmap for metagenomic enzyme discovery. Nat. Prod. Rep., 30 pages.

[39] **Roux, S., Matthijnssens, J., and Dutilh, B.E. (2021).** Metagenimics in virology.In: Encyclopedia of virology, Bamford, D.H., Zuckerman, M. (edns.), 4th edn.,Academic press, 133-40.

[40] **Sebastian, R., Kim, J.Y., Kim, T.H., and Lee, K.T. (2013).** Metagenomics: A promising approach to assess enzymes biocatalyst for biofuel production. Asian J. Biotechnol., 5, 33-50.

[41] **Srivastava, S., Ghosh, N. and Pal, G. (2013).** Metagenomics: Mining environmental genomes. Chap. 10, In: Biotechnology for environmental management and resource recovery, Kuhad, R.C. Singh, A. (edns.), 1st edn., Springer, India, 161-89.

[42] **Taupp M., Mewis, K., and Hallam, S.J. (2011).** The art and design of functional metagenomic screens. Curr. Opin. Biotechnol., 22, 465-472.

[43] **Uchiyama and co- Uchiyama, T., and Miyazaki K. (2010)**. Product-induced gene expression (PIGEX): a product-responsive reporter assay for enzyme screening of metagenomic libraries. Appl Environ Microbiol 76(21):7029–7035.

[44] **Wang, H., Hart, D.J. and An, Y. (2019).** Functional metagenomic technologies for the discovery of novel enzymes for biomass degradation and biofuel production. Bioen-erg. Res., 12, 457-70.

[45] **Wang, z., Ren, R. and Yang, Y. (2020).** Mucosa microbiome of gastric lesions: Fungi and bacterial interactions. Chap.6, In: Progress in molecular biology and translational science, Sun, J. edn., Academic press., 171, 195-213.

[46] **Uchiyama, T. and Watanabe, K (2007).** The SIGEX Scheme: High Throughput Screening of Environmental Metagenomes for the Isolation of Novel Catabolic Genes, Biotechnology and Genetic Engineering Reviews, 24:1, 107-116**.**

[47] **Zhang, X.Z. and Huang, H. (2012).** Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. Biotechnol. Adv., 30, 920-929.