Unravelling Metagenomics Approach for Microbial Biofuel Production

Km Sartaj 1, Alok Patel 2, Leonidas Matsakas 2 and Ramasare Prasad 1,*

1 Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India
2 Biochemical Process Engineering, Division of Chemical Engineering, Department of Civil, Environmental, and Natural Resources Engineering, Luleå University of Technology, SE-971 87 Luleå, Sweden
* Correspondence: ramasare.prasad@it.iit.ac.in; Tel.: +91-01332-285791 or +91-01332-285282

Abstract: Renewable biofuels, such as biodiesel, bioethanol, and biobutanol, serve as long-term solutions to fossil fuel depletion. A sustainable approach feedstock for their production is plant biomass, which is degraded to sugars with the aid of microbes-derived enzymes, followed by microbial conversion of those sugars to biofuels. Considering their global demand, additional efforts have been made for their large-scale production, which is ultimately leading breakthrough research in biomass energy. Metagenomics is a powerful tool allowing for functional gene analysis and new enzyme discovery. Thus, the present article summarizes the revolutionary advances of metagenomics in the biofuel industry and enlightens the importance of unexplored habitats for novel gene or enzyme mining. Moreover, it also accentuates metagenomics potentials to explore uncultivable microbiomes as well as enzymes associated with them.

Keywords: biofuels; metagenomics; enzyme; biomass; microbes

1. Introduction

To address the inadequate global energy supply and climate change concerns, much effort has been made towards minimizing the world’s dependency on conventional fossil fuel reserves as well as the generation of advanced technologies for sustainable fuel production [1]. Consequently, microbial-derived biofuels, such as biodiesel, ethanol, and butanol, have emerged as a promising solution and have attained growing interest from researchers, industry and policymakers due to their eco-friendly nature [2]. Owing to the capability to diminish these menacing concerns to a large extent, the interest in the aforementioned biofuel production using microbes has been steadily increasing in recent years. For example, U.S. and EU members both committed to using more renewable energy over conventional fuels to achieve net zero emissions by 2050 [3]. Even with the recent growth in global biofuel production, considerable technological bottlenecks still exist in the production processes to efficiently convert biomass into biofuels. The metabolic diversity of microorganisms enables them to produce biofuel via fermenting various substrates obtained from food crops (sugars, starch, canola seeds), lignocellulosic biomass from agricultural waste (sugar cane bagasse, corn stalks, switchgrass, poplar) and nutrient rich waste materials, etc. However, the food crops have sparked the “food versus fuel debate”, and, as such, non-food or feed plant biomass and waste materials should be used as feedstock for the microbial production of biofuels. Lignocellulosic biomass and waste material often require a pre-treatment step to break complex sugar molecules into fermentable ones or to minimize the toxicants in waste materials [4]. Nowadays, several thermochemical pre-treatment technologies are being successfully applied for this purpose, though implementation of environmentally hazardous organic solvents and high processing energy led to a sharp decline in their use. Thus, along with these technologies, industries...
decidedly consider enzymes a vital tool for biomass utilization and biofuel development because of their efficiency and selectivity in reaction chemistry [5]. Enzyme production has scaled up correspondingly; however, it has been found incompetent at meeting the rising demand for sustainable biofuels due to the unavailability of more robust strains, a complex production process, and the lack of infrastructure and expertise to manufacture beyond laboratory scale. To date, about 99.9% of microbes could not be discovered or isolated from the natural environment—a major limitation of the classical method of enzyme production (involve bacterial isolates culturing). Thus, tremendous research is being progressed worldwide to exploit unexplored microbial communities for identification and potent enzymes isolation [6].

Metagenomics is an advanced methodology that bypasses the requirement of the cultivation and isolation procedures allied with traditional microbial methods, and drastically broadens the spectrum of microbial resource utilization. It has become one of the powerful research tools for genetic engineering, environmental science, microbiology and biotechnology [7]. Countless reviews are available on every aspect of metagenomics, including sequencing strategies, functional metagenomics and biotechnological potential; nevertheless, they confer more towards varied frameworks for metagenomics-assisted isolation methods. Thus, the present article emphasizes metagenomics’ contribution as an emerging molecular technique for sustainable biofuel production (bioethanol and biodiesel both) concerning future technical and strategic improvements on a single platform, alongside deciphering advances in microbial metabolic engineering with the aid of metagenomics. Further, the article discusses the biotechnological implications as well as the metagenomic impact in shaping the socioeconomic landscape of the modern world.

2. Microbial Biofuel-an Overview

Fuels derived from biological materials, most often from plants or plant-based resources, with the use of microbes (bacteria, cyanobacteria, microalgae, yeast, fungi) and their enzymes are generally known as microbial biofuels. With the ability to combat against global climate change by reducing CO2 and other greenhouse gas emissions, microbial biofuels are being actively pursued as a potential renewable substitute for petroleum diesel. Thus, to enhance the security of national fuel supplies, the European Union, Brazil, and the United States have been implemented in a number of biofuel programs [8]. Depending on the feedstock sources, biofuels are classified as either first, second, third or fourth generation. Conventional or first-generation biofuel mainly includes the processing of food crops grown on arable land including starch, sugar, vegetable oil etc. Bioethanol is the most popular example of this generation that is produced from sugars (derived via enzymatic digestion of corn starch) or sucrose (obtained from sugarcane or sugar beets) fermentation and is widely applied in diverse sectors such as pharma, solvent industries, transportation etc. However, its economic incompetency and competition for land and water used for food production (food vs. fuel) are often listed as primary disadvantages [9]. To overcome these drawbacks, researchers turned their focus towards lignocellulosic biomass (grown on land not appropriate for food crops), waste agricultural residues, and organic fractions of municipal or household solid waste that give rise to second generation biofuel, which provides additional benefits for waste management improvement as well as fuel security. Algae-derived biofuels such as biodiesel are classified as the third generation while electro and photobiological solar fuels come under the category of fourth generation biofuels [10]. These two technologies are under development and not yet ready to be marketed at the industrial scale.

In Europe, biodiesel is the most common biofuel generated by the transesterification of algal, vegetable oil and animal fats. Chemically, it encompasses long chain fatty acid methyl/alkyl esters and is used as a fuel for vehicles in its pure form (B100). Irrespective of standard diesel, biodiesel being oxygenated contains high oxygen and hydrogen content that leads to lower particulates and carbon monoxide emissions. However, utilization in its pure form may increase emissions of nitrous oxide, a greenhouse gas [11]. In terms
of gaseous biofuels, biogas is a potential renewable energy source that can be upgraded to meet natural gas standards. Its production is a cyclic and continuous process including anaerobic digestion by methanogens. Besides zero carbon dioxide emissions, residual digestate (solid by product) generated during the process can be used as fertilizers. Biogas largely comprises carbon dioxide (CO₂) and methane (CH₄) which releases energy by combusting or oxidizing with oxygen thus biogas produced by advanced waste treatment technologies comprises 55–75% of methane [10]. Biobutanol is another promising biofuel, cited as a potential replacement for gasoline due to its direct utilization in internal combustion engines. It is mainly produced by an anaerobic fermentation process carried out by several microbial strains belonging to the genus *Clostridium* that yields acetone: butanol: ethanol in the ratio of 1:6:1. Moreover, unlike ethanol biobutanol, it shows high fuel density and less corrosivity, can be mixed in a high blend ratio called BUT16 and compatibility with existing petroleum distribution systems, including fuel pumps. Industrial investment in biobutanol is becoming more profitable as a genetically modified metabolic pathway enabled *Clostridium* to ferment agricultural, forest residues and suggests the feasibility of the second generation biobutanol industry [10,12].

Furthermore, amongst the above-discussed sustainable fuels, the focus towards microbes-derived second generation bioethanol and biodiesel (mostly obtained from lignocellulosic biomass and waste agricultural residues) production has been gradually increasing in the present-day. Thus, in line to advance their production process, new approaches are being developed continually. Genetic and metabolic engineering in microbes, multi-omics (transcriptomics, proteomics and metabolomics) tactics, nanotechnology-based biocatalyst synthesis (assist in transesterification reaction of biodiesel production), synthetic and system biology-based (includes metagenomics) approaches are few examples [13]. Metagenomics, used to study the structural and functional aspects of the entire nucleotide sequence isolated from varied environmental samples, is a new and thriving area and is likely to be an important contributor to resolving problems with renewable energy. Therefore, the next section discusses metagenomics and their foremost benefits for mainstream biofuel production.

3. Introduction of Metagenomics

Metagenomics is a cutting-edge technology that emerged in the late 1990s and usually involves complete microbial genomic DNA extraction from specific environmental habitats, constructing libraries, and screening to seek novel functional genes or biologically active compounds [2,5]. This methodology plays a significant role in processing microbial genetic information that has not been cultivated before or that is difficult to culture in a laboratory. Similarly, available data can be utilized for new species discovery and to uncover earlier-identified biochemical pathways and functions. Thus, metagenomics offers a powerful tool for tapping the genetic and metabolic diversity of complex ecosystems [14,15].

Metagenomic studies typically follow a multistep strategy including the pre-treatment of samples containing environmental microorganisms, DNA extraction, proper host and vector selection for gene cloning as well as protein expression and further metagenomic library construction (functional metagenomic approach), sequencing, post metagenome sequencing (sequence-driven metagenomic approach), and data analysis. Amongst all steps, the isolation of pure, undegraded DNA from environmental samples is considered to be the most important and highly critical step. In view of this, numerous physical, chemical, mechanical and enzymatic methods for DNA material extraction are available. Those methods are categorized into the direct and indirect extraction methods. Direct methods include onsite cell lysis followed by the extraction and purification of metagenomic DNA, whereas microbial cells’ separation prior to lysis is vital for indirect methods. Direct methods are generally used for high DNA yield whereas the indirect method results in the high purity of DNA. Additionally, the overall metagenomic process relies on the development of molecular biology, microbiology and bioinformatics [16]. Through
facilitating detailed learning of the microbial genome, metagenomics provides a platform for new gene discoveries—thus novel enzyme isolation that can be further screened for their different functional aspects like potential in antimicrobial agents, biochemically active and therapeutic compounds etc. (Figure 1).

**Figure 1.** Overview of metagenomic approach for biodiesel production (image of biofuel is acquired from https://es.dreamstime.com/ciclo-de-vida-del-combustible-biol%C3%B3gico-image (accessed on 10 September, 2022) 112,357,052).

4. **Metagenomics—an Advanced Tool in Biofuel Sector**

Recently in the biofuel sector, metagenomics is significantly contributing by discovering novel cell wall degrading enzymes that are key factors for lignocellulosic biomass
Depolymerization for subsequent conversion into biofuel and hence providing a cost-effective way to biofuel technology (Table 1).

**Table 1.** Enzymes discovered through metagenomic approach along with isolation site/screening method and industrial applications.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Source</th>
<th>Special Feature/Screening Method</th>
<th>Application</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellulase</td>
<td>Buffalo rumen</td>
<td>Thermotolerant</td>
<td>Bioethanol, biopolishing of textile fibers</td>
<td>[17]</td>
</tr>
<tr>
<td>Cellulase</td>
<td>Anaerobic beer</td>
<td>Halotolerant</td>
<td>Bioethanol, biodiesel, extraction of flavoring materials and essential oils</td>
<td>[18]</td>
</tr>
<tr>
<td>Hemicellulase</td>
<td>Degrading wheat straw</td>
<td>Thermotolerant, halotolerant</td>
<td>Bioethanol, biodiesel, bioplastic from wood biomass</td>
<td>[19]</td>
</tr>
<tr>
<td>Xylanase</td>
<td>Sea-bottom</td>
<td>Sequence based</td>
<td>Bioethanol</td>
<td>[20]</td>
</tr>
<tr>
<td>Xyloglucanase</td>
<td>Bovine rumen</td>
<td>Activity based</td>
<td>Bioethanol</td>
<td>[21]</td>
</tr>
<tr>
<td>Acidobacterial-laccase like copper oxidase</td>
<td>Marsh bog soil</td>
<td>Thermostable, halotolerant, stable in the presence of organic solvents</td>
<td>Polyphephenol removal from wine, bioethanol production</td>
<td>[22]</td>
</tr>
<tr>
<td>Esterase</td>
<td>Antarctic soil</td>
<td>Function based (fosmid vector)</td>
<td>Bioethanol production</td>
<td>[23]</td>
</tr>
<tr>
<td>Glycoside hydrolases</td>
<td>Anaerobic digestion sludge</td>
<td>Sequence based</td>
<td>Bioethanol production</td>
<td>[24]</td>
</tr>
<tr>
<td>Cyclodextrinase</td>
<td>Bovine rumen</td>
<td>Function based</td>
<td>Biodiesel production</td>
<td>[25]</td>
</tr>
<tr>
<td><strong>Enzymes involved in bioremediation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naphthalene dioxygenase</td>
<td>Oil-contaminated soil</td>
<td>Function based</td>
<td>Bioremediation of oil-contaminated soil/water</td>
<td>[26]</td>
</tr>
<tr>
<td>Oxygenases</td>
<td>Artificially polluted soil</td>
<td>Function based</td>
<td>Remediation of oil-contaminated soil/water</td>
<td>[27]</td>
</tr>
<tr>
<td>Laccase</td>
<td>Water from South China Sea</td>
<td>Sequencing based</td>
<td>Dye degradation</td>
<td>[28]</td>
</tr>
<tr>
<td>Phenol hydroxylases and catechol 2,3-dioxygenases</td>
<td>Wastewater treatment plant</td>
<td>Function based</td>
<td>Potential use in aromatic compound degradation</td>
<td>[29]</td>
</tr>
<tr>
<td><strong>Enzymes used in other applications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipase/protease/hemolysins/biosurfactants</td>
<td>Slaughterhouse drain</td>
<td>Function based</td>
<td>Applied as antimicrobial agent</td>
<td>[30]</td>
</tr>
<tr>
<td>β-Glucosidase</td>
<td>Hydrothermal spring water</td>
<td>Function based</td>
<td>Thermotolerant and heat-stable enzyme</td>
<td>[31]</td>
</tr>
</tbody>
</table>

Lignocellulosic biomass is mainly composed of cellulose, hemicellulose (fermentable polysaccharides) and lignin and can be hydrolyzed by various glycohydrolases on an industrial scale [2]. The recalcitrant nature of lignocellulosic materials and the constant need to further improve the efficiency of the hydrolysis enzymes necessitate microbial bio-prospecting on extremophilic sites to identify unique cellulolytic microbes possessing high efficiency in biomass hydrolysis, required for biofuels commercialization. In this direction, Dimitra Zarafeta et al. (2016) introduced a robust GH5 cellulase termed CelDZ1 whose catalytic profile (activity at low pH, good thermostability, high halotolerance, resistance for metal ions and other denaturing agents) affirms its relevance towards ‘enzyme cocktails,’ mainly applied for the second-step processing of biomass in the biofuel industry. Briefly, the polysaccharide degrading microbe, *Thermoanaerobacterium*, was isolated...
via samples collected from Icelandic hot springs through an enrichment approach (anaerobic condition at pH 7, 55 °C with carbon source xylan and 0.01% yeast extract). Whole genomic DNA was extracted from these samples, sequenced via NGS (next-generation sequencing) and consecutively subjected to bioinformatic analysis in order to find sequences encoding for presumed cellulolytic enzymes. After ensuring enzyme sequence, amplified celDZ1a (via PCR, polymerase chain reaction) constructed into a plasmid pET-CelDZ1a that was cloned and overexpressed in Escherichia coli BL21(DE3) cells. Protein was purified, visualized by SDS-PAGE and Western blotting. Zymography, an electrophoretic technique for hydrolytic enzyme detection, applied further for qualitative analysis of cellulolytic activity in which 0.25% CMC (carboxymethylcellulose) enriched 12% SDS-PAGE gel, was used. Enzyme activity was quantified by measuring the reducing sugar amount released from substrate (CMC) using the 3,5-dinitrosalicylic acid (DNS) method.

The enzyme was further screened for optimal pH, temperature, substrate and salt concentrations and revealed structural features through X-ray crystallography provide potential explanations behind its biochemical characteristics [32]. New extremophilic cellulolytic eubacterium (Caldicelluloseruptor bescii) was also isolated from hot springs in the Valley of Geysers by implementing a metagenomic approach. A detailed study on its cell-free extracellular cellulase (CEC) represented its efficiency for the degradation of crystalline cellulose, various polysaccharides like xylan, non-pretreated plant biomass, and mannose-based substrates at high temperatures (≥75 °C–85 °C). Unlike Trichoderma reesei CEC (commonly used commercial cellulase) the degradation activities of C. bescii CEC for rice straw, timothy grass, and Avicel (microcrystalline cellulose) were found to be more effective, which suggests its usefulness in the processes operated at high temperatures > 75 °C [33,34]. Intending to identify novel bio-catalytic enzymes encompassing great efficiency to degrade cellulolytic biomass gut microbiota are likely serving as a main target. For instance, bioinformaticians in NIAS (National Institute of Advanced Studies) explored goat rumen liquid and identified unique bifunctional cellulolytic gene sequences. From the hindgut of wood-feeding higher termite (Nasutitermes corniger), 886 proteins were identified among which 197 show known enzymatic functions. To perform this metagenomic study, collected termites from the Costa Rica were first morphologically classified on the basis of cytochrome oxidase gene sequence and dissected in an anterior P3 hindgut compartment within 36 h of collection. Luminal contents (about 165 specimens) were pooled, diluted with buffered saline and frozen immediately. Further, DNA extraction was performed and inserted containing whole genome shotgun libraries prepared as well as sequenced with BigDye Terminator v3.1. From these libraries, fosmid vectors were prepared which were completely sequenced using traditional and pyrosequencing methods. All collected data were incorporated into the Integrated Microbial Genomes with Microbiome Samples (IMG/M) system. Aliquots encompassing centrifuged luminal content and protein from soluble fraction were heat-treated, reduced, alkylated, digested and analyzed by LC-MS/MS and were searched against the termite metagenome database with SEQUEST (program for mass spectrometry data analysis). Identified glycosyl hydrolases (GHs) were named as CAZy (carbohydrate-active enzymes) and were thoroughly screened for their activity with microcrystalline and phosphoric-acid-swollen cellulose (PASC) [33,35]. Moreover, functional screening of metagenomic libraries from distinct microorganisms also directed the purification and characterization of several cellulolytic enzymes from three distinct cellulolytic sites; extreme environmental samples, rabbit’s cecum and lake water microbiomes [33]. Additionally, a novel endoglucanase showing stability on a broad range of pH (3.5–10.5), enzymes such as β-glucosidase (Bgl1A, Bgl1D) were also reported from marine and soil metagenomes, respectively. Bgl1A finds significant industrial application owing to its high glucose tolerance and low sensitivity to product inhibition, and Bgl1D was found to be highly active at low temperatures, pH (5.5–10.5) and high ionic liquid titer. In the case of Bgl1A, microbial samples were preferably collected from South China Sea’s surface water. The BAC (Bacterial Artificial Chromosome) library using vector pIndigoBAC-5 was prepared with extracted DNA and transformed.
into the E. coli EPI300. After screening of positive clones (forms black halo around colonies on LB agar plate supplemented esculin hydrate, ammonium ferric, chloramphenicol), plasmid accommodating β-glucosidase gene retransformed into E. coli EPI300 by electroporation. From these positive clones, BAC DNA was extracted and recombinant plasmid vector pUC19 constructed and transformed into E. coli DH5α with the next steps being isolation and the sequencing of inserted fragments. Further, the β-glucosidase gene bgl1A was PCR-amplified and subjected to overexpression using recombinant plasmid pET22b-bgl1A encompassing E. coli BL21(DE3) as the host. Protein purified and enzyme activity were tested by pNPG (p-Nitrophenyl β-D glucopyranoside) as a substrate. Likewise, for Bgl1D, DNA was extracted from alkaline-polluted soil in Southern China and a metagenomic library was constructed by the pGEM-3Zf (+) cloning vector with E. coli DH5α as a host. Further, DNA was sequenced and sequences allied with the gene bgl1D were identified and overexpressed in E. coli Tuner (DE3)pLacI by forming a recombinant plasmid pGEXC106. Lineweaver–Burk plot (Enzyme kinetic program) was used to determine enzymes kinetic parameters Vmax and Km [36–38]. Several carbohydrate-modifying enzymes (three endoglucanase, one β-cellobiohydrolase, 2β-glucosidase) were mined in metagenomic libraries from casts and guts of earthworms (Lumbricus terrestris and Aporrectodea caliginosa) collected from Moscow State University, Russia. Casts of each species were obtained on wet sterile filter paper through which suspension was prepared with sterile distilled water. The aqueous phase of this suspension was used to inoculate the Getchinson medium (details provided by Beloqui et al. (2010)). Only the liquid phase was considered for DNA extraction. A fosmid library was prepared with pCCFOS vector and E. coli EPI300-T1 as host. GHS sequences were identified, amplified by PCR and cloned into the pET-30-Ek/LIC vector and then introduced into E. coli BL21 (DE3)pLysS for expression [39].

A library created from insect gut microbiomes through the NGS approach as later screened for xylanase activity. In addition to the successful cloning of cold-active xylanase (Xyn8), highly active, substrate specific endo-acting alkaline xylanase (Xyl6E) has also been retrieved from a library of microbiome extracted from fungus-growing termites [40–42]. Moreover, lignin degradation has been achieved by several peroxidases; lignin, manganese and phenol-oxidases, isolated by the metagenomic-derived approach [43,44].

Metagenomic libraries prepared from intertidal flat sediments of coastal regions, South Korea revealed the presence of lipase—LipEH166—a major contributor in biodiesel production and also employed for lipid-rich wastewater bioremediation obtained from several industries such as dairy, cosmetics, food, detergents etc. Additionally, to search cold-active lipases, cold-sea sediment samples were also screened. Subsequently, the fosmid clone was selected and the whole sequence was determined by shotgun sequencing. Data analysis reveals the presence of 25 open reading frames (ORF) among which ORF20 was identified and enzyme activity determined. Acidic (EstPS2, lipolytic activity at 37 °C, pH 5) and alkaline (Est2K, lipolytic activity at 50 °C, pH 10.0) esterase were recovered from peat swamp forest soil and a compost metagenomic library, respectively [46,48]. Various thermostable (Est1–70 °C; EstE1–80 °C), as well as cold-adapted (EstMN2; EstM-N1) esterase, have been derived from thermal environments and arctic soil samples, respectively. In detail, samples from the Jæ Sawon hot spring at 70 °C temperature and pH 7 were taken for thermostable esterase Est1 isolation. Post-extraction and purification, the DNA was fractioned (1–10 kb) with Sau3AI, ligated to pZErO-2 vector and transformed into E. coli TOP10 by electroporation. Transformed colonies were selected and screened for esterase activity (on LB agar plate encompassing 1% trioleoylglycerol and Rhodamine B dye). Bacterial colonies displaying positive results on screening were selected for plasmid DNA isolation and sequencing. A comparative study was performed
Among several cellulosic biomass necessitate coconversion to 40% in lignocellulosic hydrolysate. However, to attain upscaling. For instance, native enzymes are highly valuable in downstream application of biocatalysis because of their ability to resist metal ions, detergents and other organic solvents. These aforementioned enzymes are highly valuable in downstream applications, hence they revolutionized second-generation biofuel production by cutting down the costs of pre-treatment processes and thus helping to commercialize the technology.

5. Metagenomics and Enzyme Engineering Assisted Uplifting of Biofuel Production

Metabolic engineering strategies including carbon flux modulation towards target pathways, co-factor engineering for an adequate supply of NADH/NADPH, codon optimization, incorporating heterologous pathways, promoter engineering, competitive pathway deletion, consolidated pathway insertion, directed enzyme evolution, tailoring toxic metabolite tolerance and respiratory pathway manipulation, are currently being employed to refine microbial metabolic pathways towards advanced biofuel production. As an example, native Saccharomyces cerevisiae, the most apposite host for bioethanol production, was not able to metabolize atypical substrates such as xylene (found up to 40% in lignocellulosic hydrolysate). However, to attain a high ethanol yield from lignocellulosic biomass necessitate co-fermentation of all sugar types present in hydrolysate. Among several attempts made in this direction, metagenomics-derived isolation of gene
encoding Xylose isomerase (XI) enzyme is the one. Bingyin et al. (2015) extracted metagenomic DNA from soil, snail manure from Singaporean natural forests and XI genes from mammal gut microbes, Alistipes sp. HGB5, actively expressed in a respiration-deficient S. cerevisiae strain that results in maximal xylose conversion and high, specific ethanol productivity (0.283 g g⁻¹ h⁻¹) [55]. Two other XI genes (xym1, xym2) were also isolated from the soil metagenomic library. However, their enzymatic activities expressed in S. cerevisiae were comparable to Pirromyes XI (known for high XI activity) but the aerobic growth rate was found to be very low. In this regard, the activity of XI genes (Ru-xylA) isolated from the bovine rumen metagenomics library was improved by mutagenesis (site-directed mutation in K11T/D220V) and functionally expressed in S. cerevisiae [56,57]. The substitution of asparagine residue at position 337 of novel XI, RsXI-C1 (from protist residing in hindgut of termite Reticulitermes speratus) with cysteine improved xylose utilization ability of yeast 2.5 times more than without mutation. As the recombinant strain allows much lower xylitol accumulation, this provides the additional benefits of cofactor imbalance that could impede metabolic flux [58]. A psychrophilophilic Xyn-2, novel endo-1,4-xylanase from the camel rumen metagenome, is a successful hydrolysis agricultural waste; wheat bran-100%, rice straw-56%, wheat straw 56%, rice bran 41%. Ethanologenic Bacillus subtilis AP is able to produce 5.5 g/L ethanol with 22.6% yield in consolidated bioprocessing (CBP) alongside this enzyme. Additionally, in the SSF system (simultaneous saccharification and fermentation), this may have a great impact with 7.3 g/L concentration and 26.8% ethanol yield [59]. The hydrolysis of autoclaved pre-treated rice straw (cellulose rich most abundant agricultural waste) in the presence of PersiCel4, endo-β-1,4-glucanase from the sheep rumen metagenome increased the reducing sugar content by up to 260% after 72 h in very harsh conditions (temperature 85 °C, pH 8.5) [60]. Another cellulase, Cel-5M from the rumen metagenome identified as the GH5 family endoglucanase, was tested for lignocellulosic biomass hydrolysis especially wheat straw. An ethanol yield of 0.46 g/g and 0.43 g/g cellulase was obtained after alkali- and steam-treated wheat straw respectively. This multifunctional enzyme also showed activity on various substrates such as CMC, Xylan, Avicel, Glucopyranoside, β-Mannan and filter paper [61]. Moreover, the heterologous expression of endoglucanase Cel776 (from hot spring of Spain) in S. cerevisiae enabled the strain to degrade cellulose and xylan both [62]. In a very new approach for cellulose hydrolysis, magnetic nanobiocatalyst was prepared by way of the covalent immobilization of cellulase CelDZ1 (described in Section 4), horseradish peroxidase (HRP), β-glucosidase (bgl) and glucose oxidase (G Ox) on the surface of amino-functionalized magnetic nanoparticles. In a cascade of reaction, CelDZ1 hydrolyze cellulose into cellobiose converts into glucose via bgl. Further, glucose oxidation forms gluconic acid and H₂O₂ that act as a substrate for HRP, reduced to form H₂O [63]. Metagenomically isolated cellulase from pig ordure and rice straw, Exo2b was applied to improve cellulase activity in T. reesei a cellulase-producing fungi. In brief, the fragment-encoding Exo2b catalytic domain fused with the cbh1 gene from T. reesei and was successfully expressed in T. reesei Rut-C30. Compared to the parental strain, CMCase activity in secreted proteins was increased by 18% and high glucose concentration by 19.8% was attained by secreted protein-mediated hydrolysis of corn stover [64]. A crude extract of LipC12, lipase obtained by metagenomics immobilized on Immobead 150 (Ibead-CELipC12) and applied for the enzymatic synthesis of biodiesel as 85% oleic acid conversion was observed in 48 h [65]. In order to form immobilized lipase artificial body AOB-sole-lip-1, a new gene encoding lipase (Lip-1) extracted from metagenomic BAC library constructed via water samples from hot springs, China was cloned and fused with the oleosin gene, and overexpressed in E. coli. The monomeric Sole-lip-1 fusion protein showed transesterification activity at a 45 °C optimal temperature. Saturated and polyunsaturated methyl esters were obtained after enzymatic transesterification of olive oil and methanol [66].

6. Metagenomic Implications on Biotechnology
Microbial cultivation to produce limitless biological compounds, including antibiotics, enzymes, vitamins, amino acids, various biomaterials, low molecular weight substrates etc., is the most predominant area of industrial biotechnology. Thus, by evolving unculturable microbial species, metagenomics opens new horizons in the development of biotechnology as well as provides new opportunities for searching novel gene families and gene products [67]. As per the previous discussion, the enzymes produced from this technique showed valuable properties, such as stability at low or high temperatures, high activity in acidic or alkaline conditions, resistance to high saline conditions, heavy metals, organic solvents, pressure and radiation are much needed for the industrialization of any biotechnological process. Metagenomics studies working in the direction of pathway-centric approaches have also been applied to provide insight about the microbial community interaction in a specific environment. In this context, an interesting work linked to the functional gene exploration in leaf-cutter ant fungus gardens should be highlighted. By virtue of the metagenomic approach, enzymes and pathways involved in the symbiosis between leaf-cutter ants and their cultivar have been determined with the advent of novel cellulases [35].

7. Future Perspectives and Challenges

Even though metagenomics has been proven an efficient technology in microbial process development, bottlenecks still need to be resolved. For instance, advancement in heterologous gene expression originating from unculturable microorganisms is required, as the majority of enzymes are still uncharacterized. It is challenging to understand the exact role of an enzyme identified from complex environmental microbial communities, which mandates the strengthening of the combination of metagenomics and bioinformatics technologies. The development of high-throughput cloning methods is needed to efficiently express soluble protein, as well as further progression in some novel approaches such as metatranscriptomics, structural genomics must be initiated. Moreover, to access complex metagenomic sequence information, bioinformatic and mining programs such as “metagenome Mapserver, ANASTASIA” need to be established [68,69]. Yet, to assure sustained growth in the metagenomics field, more innovative and sophisticated bioinformatics tools must be developed. With continuous development in all these areas, metagenomics technology will be the most important tool for studying microbial diversity in the natural environment and the screening of new genes or biologically active substances.

8. Conclusions

In day-to-day life, the demand for renewable energy has been continuously increasing, hence breakthrough research is being conducted in this field. Metagenomics allow for the mining of cellulose-degrading enzymes that potentially break down lignocellulosic biomass into fermentable sugars, allowing the identification of more efficient enzymes that, in turn, facilitate the transition to economically feasible biofuel production. Furthermore, the newest development in metagenomics, including metatranscriptomics and metaproteomics, promises the advanced screening of uncultivated microbes and establishes metagenomics as a great tool to access inaccessible organisms. Apart from playing a crucial role in the bioenergy field, metagenomics can also be successfully applied in several other arenas such as in the bioremediation and health sector (known as clinical metagenomics that includes the genomic analysis of samples collected from patients), but their full potential in mentioned fields still remains to be unraveled.

Author Contributions: Writing—original draft preparation, K.S.; writing—review and editing, A.P.; L.M. and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Genes 2022, 13, 1942

Data Availability Statement: Not applicable.

Acknowledgments: Km Sartaj would like to acknowledge CSIR-NET-fellowship, Government of India (Fellowship Grant No. 09/143(0918)2018-EMR-I) and highly obliged to the Indian Institute of Technology, Roorkee.

Conflicts of Interest: The authors declare no conflict of interest.

References


