Biotechnology Advances xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis

Mei-Ning Xing [a,1], Xue-Zhu Zhang [b,1], He Huang [a,*]

[a] Department of Biological Engineering and Key Laboratory of Systems Bioengineering of Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China
[b] The First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China

ARTICLE INFO

Available online xxxx

Keywords:
Metagenomic sequencing
Biofuel synthesis
Critical enzymes
Genetic modification
Biomass

ABSTRACT

Feedstock for biofuel synthesis is transitioning to lignocelluosic biomass to address criticism over competition between first generation biofuels and food production. As microbial catalysis is increasingly applied for the conversion of biomass to biofuels, increased import has been placed on the development of novel enzymes. With revolutionary advances in sequencer technology and metagenomic sequencing, mining enzymes from microbial communities for biofuel synthesis is becoming more and more practical. The present article highlights the latest research progress on the special characteristics of metagenomic sequencing, which has been a powerful tool for new enzyme discovery and gene functional analysis in the biomass energy field. Critical enzymes recently developed for the pretreatment and conversion of lignocellulosic materials are evaluated with respect to their activity and stability, with additional explorations into xylanase, laccase, amylase, chitinase, and lipolytic biocatalysts for other biomass feedstocks.

Contents

* Corresponding author at: Department of Bioengineering, school of Biochemical Engineering, Tianjin University, Tianjin 300072, China. Fax: +86 22 2740 9598.
E-mail address: huang@tju.edu.cn (H. Huang).
[1] Mei-Ning Xing and Xue-Zhu Zhang are equal contributors.

0734-9750/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.biotechadv.2012.01.021

1. Introduction

Derived from biomass, biofuels offer an environmentally benign and cost-effective solution for fossil fuel depletion. Of this alternative, renewable sources of energy, biodiesel and bioethanol have attracted growing attention from policy makers, industry and researchers for their economic, environmental and social benefits. Significant incentive exists for the development of efficient biofuel technology, with both the U.S. and members of the EU committing to increasing the proportion of renewable energy in their primary energy supply to 10% and 20% by 2010 and 2020, respectively. Other governments, such as Sweden's, have adopted even more ambitious targets, attempting to replace all fossil fuels with biofuels after 2020 to eliminate their dependence on oil.

Despite recent growth in global production of biofuels, significant technological bottlenecks still exist in the production processes to efficiently convert biomass into biofuels. While thermochemical conversion technologies can be applied for biomass conversion, industry is increasingly considering enzymes as a key technology for biofuels development and utilization, citing their efficiency and selectivity in the reaction chemistry (Jaeger et al., 1999). However, the enzymes currently employed for biomass conversion cannot meet the growing demand for economically viable biofuels due to their high cost, low activity and poor stability under the required operating conditions. Thus, continued development of novel enzymes for use in the production of advanced biofuels is required (Barnard et al., 2010).

Acceleration of the novel enzyme development process is primarily dependent on two factors: (1) efficiency and sensitivity of the screening strategy, and (2) diversity of candidate genes (microbial). The fact that traditionally enzymes could only be obtained from bacterial isolates was one of the main limitations to the widespread application of enzymes in industry (Leresche and Meyer, 2006). More than 99% of microorganisms from natural environments cannot be efficiently cultivated using current isolation and culture methods, severely reducing the microbial resources which can be utilized (Torsvik and Ovreas, 2002). Scientists have thus focused on the development of new methods capable of utilizing the genes of these microorganisms in biotechnology which are independent of routine culture techniques.

Metagenomics is an advanced methodology which emerged in the late 1990s, by means of extracting all microbial genomic DNAs in a certain environmental habitat, constructing metagenomic libraries, and screening to seek novel functional genes and/or biologically active compounds (Ferrer et al., 2005; Wang et al., 2009). Metagenomics overcomes the disadvantages of isolation and cultivation procedures of the traditional microbial method, and thus greatly broadens the space of microbial resource utilization. It has become one of the powerful research tools for microbiology, biotechnology, soil and environmental sciences, and a new field of genetic engineering.

At present, with the help of the rapid development of high-throughput sequencing methods, metagenomics has been employed to identify enzymes for use in biofuels production. Many novel enzymes have been found by means of this technology, including lignases, xylanase, endoglucanase, amylolytic enzymes, β-glucosidase for bioethanol, and lipolytic enzymes for biodiesel. Some of these have multiple functions and can catalyze a number of different reactions (Kim et al., 2008; Nam et al., 2009, 2010; Palackal et al., 2007; Zhao et al., 2010). Some exhibit high activities (Fang et al., 2009), specificities (Wong et al., 2010) and stability (Pottkamper et al., 2009), and can work under a wide range of pH (Duan et al., 2009), temperature (Sharma et al., 2010) or ionic conditions (Ilmberger and Streit, 2010). These enzymes may have potential for new application in biofuels production.

2. Access to novel biocatalysts from the metagenome

Research strategies for accessing novel biocatalysts from the metagenome include: pretreatment of genes of interest, extraction of nucleic acid, selection of vector and host system, and metagenomic library screening (Fig. 1). Multidisciplinary developments in the areas of microbiology, molecular biology and bioinformatics have enabled metagenomic technologies within each of these stages, contributing significantly to the development of novel biocatalysts.

2.1. Pretreatment of environmental samples

Pretreatment for nucleic acid extraction processing operations includes non-enrichment or enrichment of interested genes. In most experimental research, non-enriched methods are used due to their improved ability to maintain the diversity of microbial communities. However, enrichment methods are known to improve the specificity of a sample's genomic DNA, benefiting sequencing-based screening of novel genes. Stable-isotope probing (SIP) (Radajewski et al., 2002), suppression subtractive hybridization (SSH) (Galbraith et al., 2004), differential display (Liang, 2002), phage-display (Crameri and Suter, 1993), affinity capture (Demidov et al., 2000), and microarrays (Wu et al., 2001) are all methods of enrichment.

2.2. Extraction of nucleic acid

Construction of a metagenomic library requires a sufficient number of high quality DNA samples, making the extraction and purification of DNA from the environmental samples a critical step (Wilkinson et al., 2002). Two types of extraction methods are commonly applied, according to the size of target genes and different screening strategies: direct extraction and indirect extraction. Direct extraction methods use detergents and enzymes to process the test samples without the cultivation of microorganisms, followed by phenol or chloroform-based extraction and separation of the DNA. Although this method has a greater DNA recovery rate, the smaller extracted DNA fragments (general 1–50 kb) and elevated impurity content due to destructive mechanical forces makes this method inappropriate for constructing large inserts libraries (Desai and Madamwar, 2007). Nonetheless, direct extraction methods have been successfully used to extract DNA from microbial communities (Bey et al., 2010). Indirect extraction methods (cell separation and extraction method) employ physical means to separate the microorganisms from the sample followed by lysis extraction, thus obtaining larger DNA fragments by avoiding high mechanical strength actions directly on the DNA. The recovery rate of indirect extraction is 10–100 times lower than direct extraction (Parachin et al., 2010). Thus, in specific experiments, the extraction method should be selected by weighing the various requirements for product recovery, including: operational

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

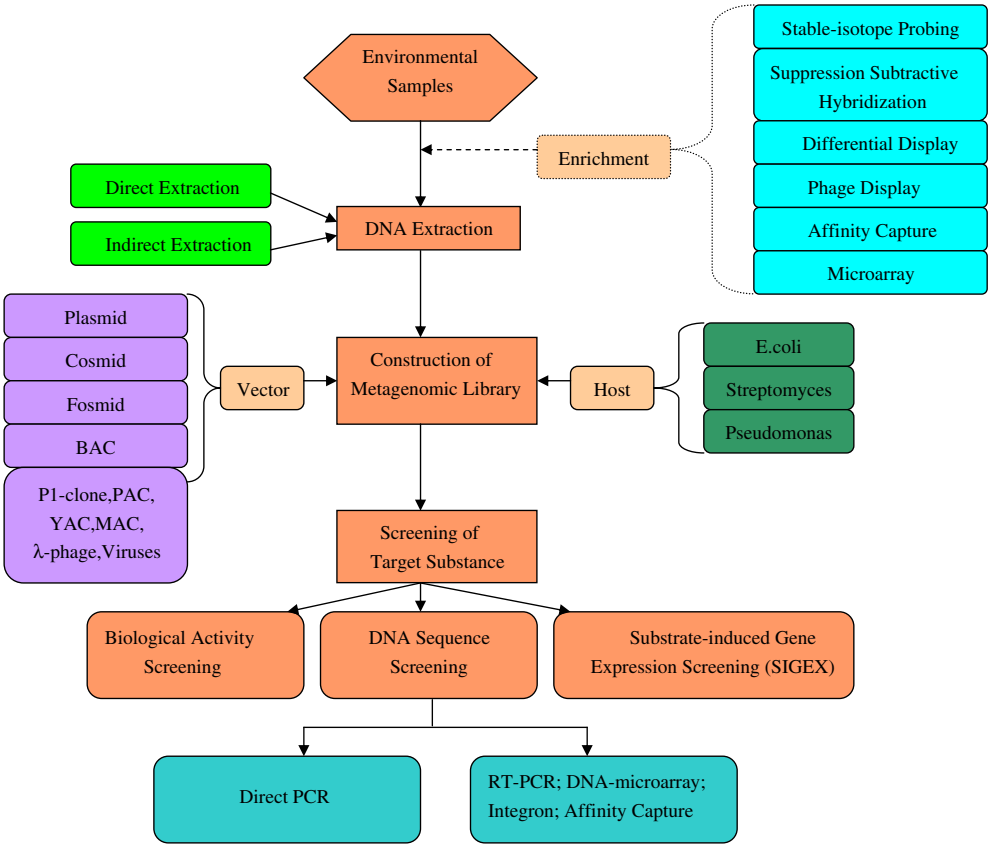

Fig.1. General process of metagenomie strategies for mining novel enzyme genes.

complexity and the recovery rate, integrity and purity of the resultant DNA.

2.3. Construction of metagenomic library

The key to metagenomic library construction is the selection of appropriate vector and host systems depending on the characteristics of environmental samples and the intended purpose of the constructed database.

2.3.1. Vector selection

Appropriate vector selection plays an important role in metagenomic technology, determining whether genome or gene clusters can be transuded into host cells or highly expressed. The selection of vector systems depends on the quality of extracted DNA samples and the research objectives, requiring consideration of the size of insert fragments, the copy number of vector needed, the host used, and the screening method. Plasmid, bacterial artificial chromosomes (BAC), cosmid, and fosmid are examples of frequently used vectors.

When the purpose is the separation of independent genes or small size of operons using coding novel functions, plasmid can be used as a vector insert small fragment (15 kb) (Henne et al., 1999). The purity and recovery of DNA fragments must then be considered when extracting and purifying the samples. When large gene fragments encoding the complex biosynthetic pathway are desired, cosmid (35–45 kb) (Henne et al., 1999) or BAC (about 200 kb) (Entcheva et al., 2001) is used to build a library containing higher lengths of DNA. The fosmid vector has also been used for the construction of a large insert library similar to cosmid, although the former has a higher cloning efficiency and stability when expressed in *Escherichia coli*. (Beja, 2004).

In addition, some other vectors have been used to build metagenomic libraries. P1-clone was used in the map constructed for further understanding of the fundamental genomic architecture of *Leishmania* (Zhou et al., 2004). P1-derived artificial chromosome (PAC) was used to facilitate the molecular classification of gliomas (Roerig et al., 2005). Cloning of large chunks of human genomic DNA in recombinant systems such as yeast or bacterial artificial chromosomes (YAC) has greatly facilitated the construction of physical maps (Weier et al., 2009), while mammalian artificial chromosomes (MAC) present a promising clinical strategy for numerous diseases (Katona et al., 2011).

2.3.2. Host selection

The choice of host strain is one of the prerequisites for efficient cloning or expression of recombinant genes. When selecting the host strain, the efficiency of the conversion process, gene expression, plasmid stability in the host cells and screening of target traits should be taken into account. Currently, *E. coli* is the most widely used host. However, many eukaryotic genes cannot express functional proteins of biological activity in *E. coli* due to its prokaryotic nature, necessitating the development and establishment of a novel alternative host system (Ward, 2006).

Other microbes, such as *Streptomyces* and *Pseudomonas*, can also be used as the host for library construction (Courtois et al., 2003; Wang et al., 2000; Wilkinson et al., 2002). The efficiency of genetic screening will be greatly improved as the technology continues to mature and new host bacteria are developed, enabling the detection of more functional genes and gene clusters of interest, and the discovery of novel active substances.

2.4. Metagenomic library screening

A wealth of genetic resources and novel active substances can be obtained from the metagenomic library. However, methods of effectively screening functional genes from the large number of microbial species in environmental samples are still evolving, necessary for the continued development of a sophisticated, large capacity library. At present, there are

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

four screening programs: Biological activity screening; DNA sequence screening; Compound configuration screening; and substrate-induced gene expression screening (SIGEX).

2.4.1. Biological activity screening

Biological activity screening, or function-based screening, involves the identification of positive clones expressing desirable characteristics, followed by the use of sequence or biochemical analysis to verify the active clones obtained through high-throughput screening technologies. Because the biological activity screening is not dependent on sequence information or sequence similarity to known genes, this methodology has led to the development of novel natural products and protein genes, including lyase (Solbak et al., 2005), amylase (Yun et al., 2004), esterase and protease.

Two kinds of function-based approaches have primarily been employed for screening of metagenomic libraries. One is direct detection of specific phenotypes of individual clones by adding chemical dyes and insoluble or chromophore-containing derivatives of enzyme substrates into the growth medium. Examples for this simple activity-based approach are the detection of positive clones with β-glucosidase activity on LB agar plate containing esuculin hydrate and ammonium ferric (Fang et al., 2010), or the detection of lipolytic activity by employing indicator agar containing tributyrin (Rashamuse et al., 2009). The second approach is the use of host strains that require heterologous complementation by foreign genes for growth under selective conditions. Only recombinant clones harboring the targeted gene and producing the corresponding gene product in an active form are able to grow. A lysine racemase (lyr) gene was isolated from a soil metagenome by functional complementation for the first time by using *E. coli* BCRC 51734 cells as the host and D-lysine as the selection agent (Chen et al., 2009).

However, the expression of heterologous genes in a particular host bacterium is hampered by various limitations including inefficient transcription of target genes as well as improper assembly of the corresponding enzymes. Therefore, it is subject to high workload requirements, low efficiencies, and limitations in detection methods. Troeschel et al. (2010) have established an efficient expression and screening system by using a shuttle vector that allows comparative expression and screening of metagenomic DNA in different bacteria, and a recombinant transposon MuExpress to enhance the expression capacity of *E. coli*. Screening efficiency can also be improved by enrichment of target bacteria, use of screening-sensitive substrates, or the development of novel substrates (Streit and Schmitz, 2004).

2.4.2. DNA sequence screening

DNA sequence screening, or sequence-driven screening, may involve the design and engineering of primers or probes based on a known conserved sequence, via PCR amplification or hybridization, to filter an objective clone. The analysis of complete genomes of a symbiotic microbial community by this method has been reported (Tyson et al., 2004). Metagenomic DNA are primarily from the acidophilic biofilm, whereby anglicizing these DNA sequences and functional details enables the identification of a large number of biocatalysts capable of performing their functions under extreme conditions (Schmeisser et al., 2003). However, this strategy is limited to the isolation of new members from known gene family or genes that contain a highly conserved region, since an understanding of the relevant gene sequences is required.

Other DNA sequence screening methods published in literature include: reverse transcription PCR (RT-PCR) (Wilson et al., 1999), DNA microarrays (Park et al., 2008; Park et al., 2010; Wu et al., 2001), integron (Rowe-Magnus and Mazel, 2001), affinity capture (Stull and Pisano, 2001) and subtractive hybridization magnetic bead capture (Meiring et al., 2010; Meyer et al., 2007).

2.4.3. Substrate-induced gene expression screening (SIGEX)

SIGEX is a new method used for genetic screening, based on the selective expression by metabolism-related genes or enzyme genes in the presence of certain substrates. This principle is used to filter the target genes that do not express (Uchiyama and Miyazaki, 2010; Uchiyama and Watanabe, 2008; Uchiyama et al., 2005; Yun and Ryu, 2005).

2.4.4. Compound configuration screening

Compound configuration screening relies on the identification of chromatographic peaks which differ from those of the host materials. Clones are screened based on their ability to produce novel structural compounds which yield diverse chromatographic peaks relative to the host cell. While effective, this method is cost prohibitive and labor intensive. By using rapid HPLC-ESIMS screening method and downstream database treatment, Wang et al. (2000) identified two transformants that contain novel compounds and five novel compounds.

3. Research progress of metagenomic sequencing

Each of the aforementioned screening methods has advantages and disadvantages governing their application, whereby selection of the appropriate method is highly dependent on the characteristics of the test sample extracted. While precedent examples exist of metagenomics applied to the screening of novel biological catalysts, high-throughput and sensitive screening methods are desiderated due to the high library capacity and gene diversity.

3.1. Traditional sequencing and second-generation sequencing

DNA sequencing technology has undergone a long and tortuous course of development. Using conventional technology, the DNA sequences obtained were primarily capillary-based, semi-automated implementations of the Sanger biochemistry (Hunkapiller et al., 1991). Since the emergence of the human genome project, attempts to establish the reference sequence database of human and other major model organisms has enhanced short fragment mapping capabilities. The newly emerging molecular biology techniques have promoted the development of high-throughput DNA sequencing methods, while co-development of complementary technology has provided greater support for DNA sequencing technology.

Improvements in technology have led to the appearance of second-generation sequencing products on the market, referring to the implementations of cyclic-array sequencing capable of analyzing dense arrays of DNA features by iterative cycles of enzymatic manipulation and imaging-based data collection (Shendure et al., 2005).

3.2. Second-generation DNA sequencing

Compared with the Sanger sequencing, the second-generation or cyclic-array strategies break the bottlenecks of restricting the massively parallel sequencing, which rely on the in vitro construction of sequencing libraries and in vitro amplification of target DNA fragments. Using the array-based sequencing, hundreds of millions of sequencing reads can potentially be obtained in parallel by raster imaging of a reasonably sized surface area as the effective size of sequencing features can be less than 1 μm. A single reagent volume can enzymatically manipulate array features as they immobilized to a planar surface. Collectively, these differences translate into dramatically lower costs for DNA sequence production.

The leaders of second-generation sequencing products currently on the market are the 454 Genome Sequencer from Roche Applied Science, the Illumina Genome Analyzer from Illumina and Solexa technology, and the SOLiD platform from Applied Biosystems.

3.2.1. 454 pyrosequencing

The 454 sequencer has greatly increased the volume of sequencing conducted by the scientific community and expanded the range

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

of problems that can be addressed by the direct readouts of DNA sequence. Higher throughput, the miniaturization of sequencing chemistries and simplified in vitro sample preparation are all key breakthroughs in the development of the 454 sequencing platform. By pioneering solutions to the three bottlenecks faced by the research community (library preparation, template preparation and sequencing), 454 initiated the second-generation movement. It enabled massively parallel sequencing reactions to be carried out at improved scales and cost.

The 454 sequencer was the first non-Sanger technology to assemble bacterial genomes *de novo* (Margulies et al., 2005) and sequence an individual human genome (Wheeler et al., 2008). 454 sequencing was also employed in other notable studies, including: uncovering the potential cause of the disappearance of the honeybee (Cox-Foster et al., 2007), revealing the complexity of rearrangements between individual human genomes (Korbel et al., 2007), providing new approaches to understand infectious diseases (Palacios et al., 2008), and sequencing the first million base pairs of a Neanderthal (Briggs et al., 2007; Green et al., 2006; Noonan et al., 2006).

3.2.2. Illumina Genome Analyzer

The Illumina Genome Analyzer platform has its origins in the works by Turcatti et al., commonly referred to as 'the Solexa' (Fedurco et al., 2006; Turcatti et al., 2008). Features of Illumina Genome Analyzer include the detection of homopolymers with increased certainty relative to other platforms (i.e. 454), average raw error rates of 1–1.5%, with higher accuracy bases with error rates of 0.1% or less achievable through quality metrics associated with each base-call. Compared to other systems, modifications have recently enabled mate-paired reads, with each sequencing feature yielding 2×36 bp independent reads derived from each end of a given library molecule several hundred bases in length.

However, the Illumina Genome Analyzer possesses shortcomings in regards to sequence length, including problems related to mixing with the fluorescent marker gene on the nucleotide, or the termination of incomplete gene excision causing the sequencing signal to attenuate and phase shift. These limitations introduce the greatest propensity for error when the Illumina Genome Analyzer is used for the detection of base substitution, followed by the detection of insertion or deletion of genes.

3.2.3. AB SOLiD

This platform has its origins in the system described by J.S. and colleagues (Shendure et al., 2005) and in the works by McKernan and colleagues at Agencourt Personal Genomics.

The AB SOLiD system enables the use of highly control beads enriched with the template fragments in any order in the chip. An additional feature of this platform is the use of two-base encoding, which is an error-correction scheme employing two adjacent bases, rather than a single base, correlating with the label. Each base position is then queried twice (once as the first base, and once as the second base, in a set of 2 bp interrogations on a given cycle) such that miscalls can be more readily identified (Rothberg and Leamon, 2008; Shendure and Ji, 2008).

In recent years, the new generation of sequencing technology has undergone rapid development. Projects which, previously, could only be carried out by large-scale sequencing centers are now practical in small laboratory settings. Sequencing research has been accelerating, with a multitude of novel biological catalysts identified using high-throughput and low-cost genetic screening.

4. Mining enzymes for biofuel production from metagenomes

4.1. Significance of metagenome-derived novel enzymes in biofuel production

Biofuel products currently obtained from organic substrates contain bioethanol, biodiesel, biobutanol and biogas, all of which rely on the use of substrates such as sugars, starch and oil crops, agricultural and animal wastes, and lignocellulosic biomass (Barnard et al., 2010).

As an important renewable energy alternative to fossil fuels, bioethanol can facilitate the combustion of fuel and can reduce the release of carbon monoxide and hydrocarbons. The production process of bioethanol involves the pretreatment and bioconversion from polysaccharides (starch, cellulose or hemicelluloses) to pentoses and hexose, and subsequent or simultaneous fermentation. The feedstock used for first generation bioethanol fermentation is primarily from the edible portion of the crops, potentially contributing to a shortage of food supplies. Second-generation technology utilizes lignocellulosic materials, derived from agricultural residues, dedicated energy crops, wood residues and municipal paper waste, suitable for application in the large-scale industrial production of biofuel.

The major components of lignocellulosic materials are cellulose, hemicellulose and lignin. Cellulose and hemicelluloses are the source of fermentable polysaccharides, while lignin is primarily aromatic, having an inhibiting effect on cellulose hydrolysis. Glycohydrolases currently available for industrial bioethanol production are incapable of meeting future demand due to the recalcitrant nature of lignocellulosic materials and the relatively low activity of currently available hydrolytic enzymes. In order to enhance the biodegradation of lignocellulosic materials for biomass conversion and industrial biofuel production, enzymes capable of efficiently catalyzing the hydrolysis of plant cell-wall components are the subject of intense research (Gray et al., 2006; Lin and Tanaka, 2006).

As a substitute of high quality petrol-based diesel, biodiesel is non-toxic, sulphur-free and biodegradable. Biodiesel production relies on the catalyst-driven chemical reaction (transesterification) between an oil feedstock (vegetable oil, animal fats and waste cooking oil included) and an alcohol (methanol, ethanol), methyl esters and glycerol. Transesterification has traditionally been catalyzed by a strong base (sodium or potassium hydroxide) under high temperature conditions (230 to 250 °C). However, significant disadvantages arise in this process due to treatment requirements of alkaline wastewater generated during removal of the alkaline catalyst and glycerol from the final product. Alternative methods used to overcome these obstacles include enzymatic transesterification, whereby biocatalysts such as lipases and esterases are used in the place of a strong base to catalyze the hydrolysis and synthesis of ester compounds. The biosynthesis pathway possesses several advantages due to the mild reaction conditions, convenience of operation, and a reduction in wastewater treatment. However, currently available lipolytic biocatalysts are inefficient in the use of short-chain alcohols, where low activities and product inhibition contribute to low yields and increased cost of production. In order to satisfy the industrial need of biodiesel production from biomass, new lipolytic biocatalysts with stability, catalytic selectivity and high reaction activity are urgently needed.

Considering the processes of industrial biofuel production, enzymes play a critically important role and are of growing industrial interest. But only a small percentage microorganism can be cultured, so the type and quantity of enzymes got from traditional cultivation-based method is not enough and cannot meet the industry demand. Metagenomics has recently appeared as an alternative approach to conventional microbial screening and the analysis of DNA from environmental samples, representing a strategy for discovering diverse enzymes encoded in nature (Ferrer et al., 2005; Wang et al., 2009). Following the rapid development of sequencing and high-throughput screening techniques, various novel enzymes with unique activities and/or sequences from different environments have been identified using metagenomics. These novel enzymes display high activities, specificities and stability, and can work under a wide range of pH, temperature or ionic conditions, all of which are needed for effective biomass utilization and biofuel development. To date, correlative biomass-degrading enzymes screened from environmental samples include: carboxyl-

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

hydrolases (esterases, lipases), polysaccharide-modifiying enzymes (cellulases, α-amylases, xylanases, 1, 4-α-glucan branching enzymes), oxidoreductases, dehydrogenases, and oxygenases. A list of recently discovered representative enzymes is provided in Table 1.

4.2. Lignocellulose degrading enzymes for biofuel production

In order to efficiently utilize plant biomass for biofuel production, lignocellulose degrading enzymes need to be widely developed and utilized. Metagenomics have been used to identify a number of enzymes, including lignase, xylanase and cellulases.

4.2.1. Lignin degrading enzymes

The accessability cellulose and hemicellulose for biofuel production is limited by crosslinking between lignin, cellulose and hemicellulose via ester and ether linkages. The efficiency of enzymatic hydrolysis is thus dependent on the ability of pretreatment to remove lignin from the feedstock. In addition to traditional physical, physical–chemical and chemical pretreatment methods, biological treatments are increasingly considered in biofuel biotechnologies (Alvira et al., 2010). Metagenomic approaches have been developed to isolate enzymes for this purpose, including peroxidases, such as lignin peroxidase, manganese peroxidase and versatile peroxidase, and many phenol-oxidases of laccase type (Beloqui et al., 2006; Fang et al., 2011; Ferrer et al., 2010; Ye et al., 2010).

4.2.2. Xylanase

As the second most abundant renewable polymer in lignocellulosic material after cellulose, hemicellulose consists of a complex matrix of polysaccharides constructed from xylan (β-1, 4-linked xylose) and mannan (β-1, 4-linked mannose), which is rapidly hydrolyzed into monosaccharides under mild catalytic conditions. The principle enzyme

Table 1
Examples of recently identified biocatalysts from metagenomic libraries.

Enzyme name	Sample source	Library vector/ host	Average insert size	Total size	Number of clones screened	Positive clones	Screening technology	Sequencing techniques	Reference
Laccase	Surface water of the sea	pIndigoBAC-5 vector/*E. coli* EPI300	50–150 kb	1.4Gb	20,000	1gene	Sequence-based(PCR)	Sanger sequencing	Fang et al., 2011
Xylanase	Soil	pHBM803/ *E. coli* DE3	7–10 kb	–	24,000	1	Function-based	Sanger sequencing	Hu et al., 2008
Xylanase	Holstein cows rumen	pCCIBAC/ *Escherichia coli* EPI300	54.5 kb	837 Mb	15,360	18	Function-based	Sanger sequencing	Zhao et al., 2010
Endoglucanase; Exoglucanase; β-glucosidase	Contents of buffalo rumen	Cosmid pWEB/ *E. coli* EPI100	35 kb	525 Mb	15,000	11; 2; 48	Function-based	Sanger sequencing	Duan et al., 2009
Endoglucanase; β-glucosidase	Forest soil, elephant dung, cow rumen, rotted tree	Lambda ZAP Expression vector/ *E. coli*	5.3 kb	212 Mb	40,000	5; 2	Function-based	Sanger sequencing	Wang et al., 2009
Endoglucanase	Compost	Cosmid pWEB/ *E. coli* EPI100	33 kb	3.3Gb	100,000	4	Function-based	Sanger sequencing	Pang et al., 2009
Endoglucanase	Human gut	pCC1FOS fosmid/ *E. coli*	30–40 kb	5.46Gb	704,000	11	Multi-step Functional screening	Pyrosequencing −454 GS FLX	Tasse et al., 2010
β-glucosidase	Sludge from biogas reactor	Cosmid/*E. coli*	35 kb	1.05Gb	30,000	1	Function-based	Sanger sequencing	Jiang et al., 2010
β-glucosidase	Alkaline polluted soil	pGEM-3Zf (+)/ *E. coli* DH5α	3.5 kb	105 Mb	30,000	2	Function-based	Sanger sequencing	Jiang et al., 2011
Exoglucanase	Casts of two earthworm species	pCCFOS Fosmid vector/*E. coli* EPI300	–	400 Mb	5,760	1	Function-based; PCR	Sanger sequencing	Beloqui et al., 2010
Mannanase–xylanase–glucanase	Cow rumen fluid	Lambda ZAP vector/*E. coli*	3.0 kb	–	50,000	2	Function-based	Sanger sequencing	Palackal et al., 2007
Glucanase–xylanase	Soil (Korean)	Fosmid pSuperCosI/ *E. coli* DH5α	–	–	70,000	1	Function-based	Sanger sequencing	Kim et al., 2008; Nam et al., 2009
Esterase/lipase	Deep sea sediment	Fosmid pCC1FOS/ *E. coli*	32.3 kb	284 Mb	8,823	1	Function-based	Sanger sequencing	Jeon et al., 2009
Lipase	Dairy cow rumen	BAC/*E. coli* TransforMax EPI300	54.5 kb	837 Mb	15,360	18	Function-based	Sanger sequencing	Zhao et al., 2009
Lipases/ esterases	Peat–swamp forest soil	Fosmid pCC1FOS/ *E. coli* EPI300	20–40 kb	–	15,000	25	Functional and sequence-based	Pyrosequencing −454 GS FLX	Bunterngsook et al., 2010
Esterase	Marine sediment	Fosmid/*E. coli*	36 kb	756 Mb	40,000	19	Function-based	Sanger sequencing	Hu et al., 2010
Esterase	Composting	Fosmid pCC2FOS/*E. coli*	35 kb	–	23,400	19	Function-based	Sanger sequencing	Kim et al., 2010
Esterase	Marine sponge Hyrtios erecta	vector pHSG398/ *E. coli* DH10B	3–5 kb	79.5–180 Mb	26,496	1	Function-based	Sanger sequencing	Okamura et al., 2009
Esterase	Arctic soil	Fosmid/*E. coli*	–	–	60,000	121	Function-based	Sanger sequencing	Yu et al., 2011
Amylase	Human gut	Fosmid/*E. coli*	30–40 kb	5.46Gb	156,000	6	Multi-step Functional screening	Pyrosequencing on −454 GS FLX	Tasse et al., 2010
Amylase	Soil (Astaka)	Cosmid SuperCos1/ *E. coli*	32 kb	272 Mb	35,000	1	Function-based	Sanger sequencing	Sharma et al., 2010

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

in the progressive breakdown of xylan is endo-β-1, 4-xylanase (EC3.1.2.8), which attacks the non-hydrolyzed polymer (Brennan et al., 2004).

Based on functional assay, several novel xylanase have been screened from environmental DNA libraries (Brennan et al., 2004). In 2006, successful cloning of a cold-active xylanase Xyn8 was reported (Lee et al., 2006). Another xylanase was later obtained from a soil-derived metagenomic library, displaying high activity at reduced temperatures, under weakly alkaline conditions (Hu et al., 2008). A highly active, substrate-specific, and endo-acting alkaline xylanase (Xyl6E) with considerably wide pH tolerance and stability has also been retrieved from the metagenomic library of microbiome extracted from fungus-growing termites (Liu et al., 2011), but with limited industrial applicability due to it's extremely low thermostability.

4.2.3. Cellulase

Cellulose is the most abundant biopolymer in nature, and has been long-recognized as a potential source for biofuel production. It consists of linear β-1, 4-linked glucose chains, necessitating a combination of multiple classes of cellulase for biodegradation. High cellulase activity levels are determined by endoglucanases (EC 3.2.1.4), which randomly sever internal amorphous sites in the cellulose chain; exoglucanases (cellobiohydrolase) (EC 3.2.1.91 and EC 3.2.1.74), which progressively act on the reducing or non-reducing ends of cellulose chains, releasing either cellobiose or glucose; and β-glucosidases (EC 3.2.1.21), which hydrolyze soluble cellodextrins and cellobiose to glucose.

Cellulases applied in the industrial generation of bioethanol often have to function under extreme conditions, requiring cellulolytic activity under a range of pH, temperature, or ionic conditions (Ilmberger and Streit, 2010).

In recent years, many novel cellulases with special properties have been screened from different metagenomic libraries. By functional screening for cellulase activities from a ruminal metagenomic library, one novel endoglucanase was found with optimal activity at low pH (4.5) and stability for a broad pH range (3.5 to 10.5) (Duan et al., 2009). Another endoglucanase (CelA) was reported with high tolerance in ionic liquids, having sufficient stability and activity to create novel options for cellulose degradation in homogenous catalysis (Pottkamper et al., 2009). By functional screening and recombinant expression, a β-glucosidase (Bgl1A) with excellent glucose tolerance from marine metagenome libraries was identified, having great potential for industrial applications due to its low sensitivity to product inhibition (Fang et al., 2010). Another β-glucosidase (Bgl1B) was also obtained from the same library, exhibiting relatively higher activity and stability at pH values between 7.0 and 9.0 (Fang et al., 2009). A novel β-glucosidase (Bgl1D) retrieved from an uncultured soil metagenomic library exhibited higher activity in lower temperature and at high ionic concentrations, and displayed remarkable activity across a broad pH range (5.5–10.5) (Jiang et al., 2011). The efficient performance of these cellulases in special conditions makes them potential candidates for industrial applications for cellulose degradation and saccharification.

Many other cellulases have been mined from various metagenomic libraries by function-based screening. Examples include: carbohydrate-modifying enzymes (2β-Glucosidase, 3 Endoglucanase, and 1 β-Cellobiohydrolase) retrived from guts and casts of earthworms (Beloqui et al., 2010); 5 endoglucanases and 2 β-glucosidases from four different environmental DNA libraries (Wang et al., 2009); endoglucanase uncovered in the transcriptome of Formosan subterranean termite (Zhang et al., 2011), or the metagenome of the compost soils (Pang et al., 2009); β-glucosidase from uncultured microorganisms in contents of a bioreactor (Jiang, et al., 2010), a marine metagenome (Fang et al., 2009), or a soil metagenome (Jiang et al., 2009). A novel exo-glucanase with exclusive specificity towards only xyloglucan and oligoxyloglucan substrates was also isolated from a rumenal microbial metagenomes (Wong et al., 2010).

High-throughput functional screening was first applied to the human gut metagenomic library, encoding the CAZymes and isolating 310 clones showing beta-glucanase, hemicellulase, galactanase, amylase, and pectinase activities (Tasse et al., 2010). Apart from activity-based functional screening, the non-functional screening methods began to rapidly develop, especially the sequence-derived screening based on high-throughput analysis. Hess et al. (2011) recently sequenced and analyzed 268 gigabases of metagenomic DNA from microbes adherent to plant fiber incubated in cow rumen, identifying 27,755 putative carbohydrate-active genes and expressing 90 candidate proteins, of which 57% were enzymatically active against cellulosic substrates. These data sets provide a substantially expanded catalog of genes participating in the deconstruction of cellulosic biomass.

4.2.4. Multifunctional enzymes

Some of the multifunctional enzymes mined using metagenomic approaches are of great interest for industrial applications in plant processing, such as biomass saccharification.

A unique multifunctional glycosyl hydrolase was discovered by screening an environmental DNA library prepared from a microbial consortium collected from cow rumen. It consisted of two adjacent catalytic domains: one (GH family) catalytic site hydrolyzing β-1,4-linked mannan substrates, and a second (GH family 26) catalytic site hydrolyzing β-1,4-linked xylan and β-1,4-linked glucan substrates (Palackal et al., 2007).

Another novel mutifuntional endo-type cellulase (CelM2) was screened from metagenomic libraries (Kim et al., 2008). Its crystalline structure and two domains, which are able to effectively hydrolyze glucan and xylan, were described by Nam et al. (2009). The novel structural characteristics of the metal-binding site and the structure of the complex formed between glucanase–xylanase CelM2 and its substrate were later reported (Nam et al., 2010). It showed that CelM2 is attractive as an industrial enzyme and that the structural results provided insights relevant to the genetic engineering of multifunctional enzymes

A novel endo-β-1, 4-xylanase with two catalytic domains of family GH43 and two cellulose-binding modules (CBMs) of family IV was recently identified through functional screening of a rumen BAC library. Partial characterization showed that this endo-xylanase has a greater specific activity than a number of other xylanases over a wide temperature range at neutral pH, with potential industrial applications (Zhao et al., 2010).

4.3. Lipolytic enzymes for biodiesel production

Many lipolytic enzymes with special characteristics, including lipases that hydrolyze long-chain acylglycerols, and esterases that hydrolyze short-chain acylglycerols, have been identified by function-based or sequence-based screening approaches from different environmental samples, such as soil, marine sediment, fermented compost, and animal rumen (Table 1).

A number of novel lipolytic enzymes belonging to novel families of the bacterial lipolytic enzyme have recently been reported in literature. Two lipolytic active proteins, FLS18C and FLS18D, could not be assigned to any known family (Hu et al., 2010), while esterase (EstD2), exhibiting increased enzymatic activity in the presence of 15% butanol and 15% methanol, was shown to be a member of a novel lipolytic family by phylogenetic analysis (Lee et al., 2010). Using a highly efficient activity-based cluster screening, EstGK1 and EstZ3 having only minor overall sequence similarity to known esterases were found (Bayer et al., 2010). A new lipase (LipEH166) belonging to a novel lipolytic family was isolated from an intertidal flat metagenome and was characterized as a novel cold-adapted alkaline lipase (Kim et al., 2009). Other cold-active alkaline lipolytic enzymes, EML1 (Jeon et al., 2009) and lipCE (Elend et al., 2007), were also obtained. Another novel, low-temperature-active alkaliphilic esterase from an Antarctic desert soil metagenomes has also

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

been identified as a potential candidate for industrial application (Heath et al., 2009).

Many other acidic or alkaline lipolytic enzymes are suitable for commercial development. The EstPS2 enzyme retrieved from peat–swamp forest soil possesses lipolytic activity under acidic conditions, exhibiting highest activity toward p-nitrophenyl butyrate at 37 °C and a pH 5 (Bunterngsook et al., 2010). A novel alkaline esterase (Est2K) from a compost metagenomic library was optimally active at pH 10.0 and 50 °C, was stable in the presence of 30% methanol, and preferred short to medium length p-nitrophenyl esters as its substrate (Kim et al., 2010).

The stability and activity of lipolytic biocatalysts with unique substrate specificities is important for industrial applications. A new thermostable esterase (EstE1) was cloned from thermal environment samples, showing activity towards short-chain acyl derivatives at temperatures of 30–90 °C, and displaying a high thermostability above 80 °C (Byun et al., 2006). Another esterase (Est1) from a hot spring metagenomic library exhibited more than 50% of its maximum activity at alkaline pH and in the temperature range of 50–75 °C, stable at 70 °C for at least 120 min (Tirawongsaroj et al., 2008). Three lipases derived from dairy rumen microflora had different substrate specificity and good thermal stability, providing a basis for large-scale industrial applications (Zhao et al., 2009). A novel esterase (EstHE1) exhibited activity against C2, C4, and C6 substrates, and had the thermal stability and salt tolerance necessary for industrial utilization (Okamura et al., 2009). EstMY isolated from an activated sludge metagenomic library exhibited the highest activity at 35 °C and pH 8.5 by hydrolysis of p-NP caprylate, maintaining comparable activity over wide temperature and pH spectra and in the presence of metal ions and detergents (JunGang et al., 2010). The high level of stability of these esterases makes them valuable for downstream biotechnological applications.

The esterase pFLS10 was also reported to have good thermal stability and high lipolytic activity at low temperatures and under basic conditions (Xu et al., 2010). Functional screening for lipolytic enzymes at low temperatures resulted in the isolation of the novel cold-active esterases, EstM-N1 and EstM-N2, from a metagenomic DNA library of arctic soil samples. Both enzymes exhibited very narrow substrate specificity and were expected to be useful for potential biotechnological applications as interesting biocatalysts (Yu et al., 2011).

Though metagenomic approaches have identified a number of novel genes encoding lipolytic enzymes, and many interesting novel lipolytic enzymes have been discovered, the fact remains that the majority of biocatalysts are still uncharacterized.

4.4. Other enzymes involved in biofuel production

Following the rapid development of sequencing and screening techniques, metagenomics has been used to obtain a plethora of enzymes with special characteristics, with increasing potential for the discovery new enzymes for industrial use. While many of the enzymes discussed have potential applications for biofuel production from lignocellulosic material, there remains a broad range of accessory enzymes and alternative feedstock which have yet to be discussed.

While starches have traditionally been acid-hydrolysed to yield starch-derived glucose, largely utilized by industry for ethanol production through fermentation, consideration is being given to the transition to microbial enzymes.

Amylases have been the focus of many metagenomic studies with several reports available on the isolation of novel amylolytic enzymes from metagenomes DNA libraries (Lammle et al., 2007; Tasse et al., 2010; Yun et al., 2004). The enzyme α-amylase (1 EC 3.2.1.1) is currently used in a broad array of industrial applications. Recently, a gene (pAMY) found encoding for amylase from a soil-derived metagenomic library was discovered, and the amylase was observed to have a maximal activity at 40 °C under nearly neutral pH conditions. The amylase retained 90% of its activity at low temperatures, making it unique among existing reports in literature (Sharma et al., 2010).

Cyclodextrinase can be used to hydrolyze cyclodextrin into maltoses or glucoses. A novel cyclodextrinase (RA 04) from metagenomic DNA of rumen fluid exhibited maximum activity at 70 °C and was active within an unusually broad pH range. It hydrolyzed alpha-D-(1, 4) bonds 13-fold faster than alpha-D-(1, 6) bonds and exhibited transglycosylation activity (Ferrer et al., 2005, 2007)

Metagenomics has also been responsible for the discovery of many other enzymes applicable to biofuel production, such as alcohol/aldehyde dehydrogenase (Wexler et al., 2005), alcohol oxidoreductase (Knietsch et al., 2003).

5. Conclusion

The potential of long-term energy shortages with the depletion of fossil fuels continues to drive exploration of alternative fuel resources. As an essential renewable energy source, biofuels have attracted considerable interest and government support throughout the world. Enzymes play an important role in the process for efficient microbial conversion of biomass into biofuels. The inefficiency and low activities of currently available enzymes for producing two of the most important biofuels, bioethanol and biodiesel, has limited industrial-scale production. Therefore, mining for enzymes, especially novel enzymes with special characteristic, is a key step in the development of industrial biofuel production. The appearance of metagenomic technology offers an alternative approach to conventional microbial screening, enabling the detection of novel enzymes from environmental samples independent of routine culture techniques. With the help of high-throughput sequence screening techniques, metagenomics has been used to identify various novel enzymes with unique activities, specificities and stabilities, with the potential for subsequent development into industrial biocatalysts.

Nonetheless, metagenomics is still a new technology, with a number of pending issues to be resolved for the construction of metagenomic libraries. It is necessary to explore more suitable eukaryotic expression vectors and host cells for the construction of metagenomic libraries, while also strengthening the combination of metagenomic techniques and bioinformatics. Access to the complex metagenomic sequence information is a huge project, where traditional methods of genetic analysis are labour intensive, subject to low efficiencies and high cost. Bioinformatics has the potential to be a powerful tool offering researchers increased convenience. With the continuous development of biology, improvements in library construction and screening strategies, and enhanced heterologous gene expression, metagenomics technology will be one of the most important means of studying microbial diversity in natural environment and screening for new genes and biologically active substances.

Acknowledgements

This research was partially supported by Tianjin Nature Science Foundation (No. 09JCZDJC17100) and Project 111 of the Ministry of Education of China (No. B06006).

References

Alvira P, Tomas-Pejo E, Ballesteros M, Negro MJ. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. Bioresour Technol 2010;101:4851–61.

Barnard D, Casanueva A, Tuffin M, Cowan D. Extremophiles in biofuel synthesis. Environ Technol 2010;31:871–88.

Bayer S, Kunert A, Ballschmiter M, Greiner-Stoeffele T. Indication for a new lipolytic enzyme family: isolation and characterization of two esterases from a metagenomic library. J Mol Microbiol Biotechnol 2010;18:181–7.

Beja O. To BAC or not to BAC: marine ecogenomics. Curr Opin Biotechnol 2004;15:187–90.

Beloqui A, Pita M, Polaina J, Martinez-Arias A, Golyshina OV, Zumarraga M, et al. Novel polyphenol oxidase mined from a metagenome expression library of bovine

rumen: biochemical properties, structural analysis, and phylogenetic relationships. J Biol Chem 2006;281:22933–42.
Beloqui A, Nechitaylo TY, Lopez-Cortes N, Ghazi A, Guazzaroni ME, Polaina J, et al. Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Appl Environ Microbiol 2010;76:5934–46.
Bey BS, Fichot EB, Dayama G, Decho AW, Norman RS. Extraction of high molecular weight DNA from microbial mats. Biotechniques 2010;49:631–40.
Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S, et al. Unusual microbial xylanases from insect guts. Appl Environ Microbiol 2004;70:3609–17.
Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A 2007;104:14616–21.
Bunterngsook B, Kanokratana P, Thongaram T, Tanapongpipat S, Uengwetwanit T, Rachdawong S, et al. Identification and characterization of lipolytic enzymes from a peat–swamp forest soil metagenome. Biosci Biotechnol Biochem 2010;74:1848–54.
Byun JS, Rhee JK, Kim DU, Oh JW, Cho HS. Crystallization and preliminary X-ray crystallographic analysis of EstE1, a new and thermostable esterase cloned from a metagenomic library. Acta Crystallogr Sect F Struct Biol Cryst Commun 2006;62:145–7.
Chen IC, Lin WD, Hsu SK, Thiruvengadam V, Hsu WH. Isolation and characterization of a novel lysine racemase from a soil metagenomic library. Appl Environ Microbiol 2009;75:5161–6.
Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, et al. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 2003;69:49–55.
Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. Science 2007;318:283–7.
Crameri R, Suter M. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. Gene 1993;137:69–75.
Demidov VV, Bukanov NO, Frank-Kamenetskii D. Duplex DNA capture. Curr Issues Mol Biol 2000;2:31–5.
Desai C, Madamwar D. Extraction of inhibitor-free metagenomic DNA from polluted sediments, compatible with molecular diversity analysis using adsorption and ion-exchange treatments. Bioresour Technol 2007;98:761–8.
Duan CJ, Xian L, Zhao GC, Feng Y, Pang H, Bai XL, et al. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. J Appl Microbiol 2009;107:245–56.
Elend C, Schmeisser C, Hoebenreich H, Steele HL, Streit WR. Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (R)-ibuprofen esters. J Biotechnol 2007;130:370–7.
Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl Environ Microbiol 2001;67:89–99.
Fang W, Fang Z, Liu J, Hong Y, Peng H, Zhang X, et al. Cloning and characterization of a beta-glucosidase from marine metagenome. Sheng Wu Gong Cheng Xue Bao 2009;25:1914–20.
Fang Z, Fang W, Liu J, Hong Y, Peng H, Zhang X, et al. Cloning and characterization of a beta-glucosidase from marine microbial metagenome with excellent glucose tolerance. J Microbiol Biotechnol 2010;20:1351–8.
Fang Z, Li T, Wang Q, Zhang X, Peng H, Fang W, et al. A bacterial laccase from marine microbial metagenome exhibiting chloride tolerance and dye decolorization ability. Appl Microbiol Biotechnol 2011;89:1103–10.
Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res 2006;34:e22.
Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VA, et al. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ Microbiol 2005;7:1996–2010.
Ferrer M, Beloqui A, Golyshina OV, Plou FJ, Neef A, Chernikova TN, et al. Biochemical and structural features of a novel cyclodextrinase from cow rumen metagenome. Biotechnol J 2007;2:207–13.
Ferrer M, Beloqui A, Golyshin PN. Screening metagenomic libraries for laccase activities. Methods Mol Biol 2010;668:189–202.
Galbraith EA, Antonopoulos DA, White BA. Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. Environ Microbiol 2004;6:928–37.
Gray KA, Zhao L, Emptage M. Bioethanol. Curr Opin Chem Biol 2006;10:141–6.
Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. Analysis of one million base pairs of Neanderthal DNA. Nature 2006;444:330–6.
Heath C, Hu XP, Cary C, Cowan D. Isolation and characterisation of a novel, low-temperature-active alkaliphilic esterase from an Antarctic desert soil metagenome. Appl Environ Microbiol 2009. AEM.02597-08.
Henne A, Daniel R, Schmitz RA, Gottschalk G. Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl Environ Microbiol 1999;65:3901–7.
Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 2011;331:463–7.
Hu Y, Zhang G, Li A, Chen J, Ma L. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. Appl Microbiol Biotechnol 2008;80:823–30.
Hu Y, Fu C, Huang Y, Yin Y, Cheng G, Lei F, et al. Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. FEMS Microbiol Ecol 2010;72:228–37.
Hunkapiller T, Kaiser RJ, Koop BF, Hood L. Large-scale and automated DNA sequence determination. Science 1991;254:59–67.
Ilmberger N, Streit WR. Screening for cellulase-encoding clones in metagenomic libraries. Methods Mol Biol 2010;668:177–88.
Jaeger KE, Dijkstra BW, Reetz MT. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. Annu Rev Microbiol 1999;53:315–51.
Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, et al. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. Appl Microbiol Biotechnol 2009;81:865–74.
Jiang C, Ma G, Li S, Hu T, Che Z, Shen P, et al. Characterization of a novel beta-glucosidase-like activity from a soil metagenome. J Microbiol 2009;47:542–8.
Jiang C, Hao ZY, Jin K, Li SX, Che ZQ, Ma GF, et al. Identification of a metagenome-derived β-glucosidase from bioreactor contents. J Mol Catal B: Enzym 2010;63:11–6.
Jiang C, Li SX, Luo FF, Jin K, Wang Q, Hao ZY, et al. Biochemical characterization of two novel beta-glucosidase genes by metagenome expression cloning. Bioresour Technol 2011;102:3272–8.
JunGang L, KeGui Z, WenJun H. Cloning and biochemical characterization of a novel lipolytic gene from activated sludge metagenome, and its gene product. Microb Cell Fact 2010;9:83.
Katona RL, Vanderbyl SL, Perez CF. Mammalian artificial chromosomes and clinical applications for genetic modification of stem cells: an overview. Methods Mol Biol 2011;738:199–216.
Kim SJ, Lee CM, Han BR, Kim MY, Yeo YS, Yoon SH, et al. Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett 2008;282:44–51.
Kim EY, Oh KH, Lee MH, Kang CH, Oh TK, Yoon JH. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. Appl Environ Microbiol 2009;75:257–60.
Kim YH, Kwon EJ, Kim SK, Jeong YS, Kim J, Yun HD, et al. Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library. Biochem Biophys Res Commun 2010;393:45–9.
Knietsch A, Waschkowitz T, Bowien S, Henne A, Daniel R. Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. Appl Environ Microbiol 2003;69:1408–16.
Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science 2007;318:420–6.
Lammle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, et al. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol 2007;127:575–92.
Lee C, Kibblewhite-Accinelli R, Wagschal K, Robertson G, Wong D. Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. Extremophiles 2006;10:295–300.
Lee MH, Hong KS, Malhotra S, Park JH, Hwang EC, Choi HK, et al. A new esterase EstD2 isolated from plant rhizosphere soil metagenome. Appl Microbiol Biotechnol 2010;88:1125–34.
Leresche JE, Meyer H-P. Chemocatalysis and biocatalysis. (Biotransformation): some thoughts of a chemist and of a biotechnologist. Org Proc Res Dev 2006;10:572–80.
Liang P. A decade of differential display. Biotechniques 2002;33:338–44. 46.
Lin Y, Tanaka S. Ethanol fermentation from biomass resources: current state and prospects. Appl Microbiol Biotechnol 2006;69:627–42.
Liu N, Yan X, Zhang M, Xie L, Wang Q, Huang Y, et al. Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. Appl Environ Microbiol 2011;77(1):48–56.
Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437:376–80.
Meiring T, Mulako I, Tuffin MI, Meyer Q, Cowan DA. Retrieval of full-length functional genes using subtractive hybridization magnetic bead capture. Methods Mol Biol 2010;668:287–97.
Meyer QC, Burton SG, Cowan DA. Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome. Biotechnol J 2007;2:36–40.
Nam KH, Kim SJ, Hwang KY. Crystal structure of CelM2, a bifunctional glucanase–xylanase protein from a metagenome library. Biochem Biophys Res Commun 2009;383:183–6.
Nam KH, Lee WH, Rhee KH, Hwang KY. Structural characterization of the bifunctional glucanase–xylanase CelM2 reveals the metal effect and substrate-binding moiety. Biochem Biophys Res Commun 2010;391:1726–30.
Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, et al. Sequencing and analysis of Neanderthal genomic DNA. Science 2006;314:1113–8.
Okamura Y, Kimura T, Yokouchi H, Meneses-Osorio M, Katoh M, Matsunaga T, et al. Isolation and characterization of a GDSL esterase from the metagenome of a marine sponge-associated bacteria. Mar Biotechnol (NY) 2009;12:395–402.
Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 2008;358:991–8.
Palackal N, Lyon CS, Zaidi S, Luginbuhl P, Dupree P, Goubet F, et al. A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. Appl Microbiol Biotechnol 2007;74:113–24.
Pang H, Zhang P, Duan CJ, Mo XC, Tang JL, Feng JX. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. Curr Microbiol 2009;58:404–8.
Parachin NS, Schelin J, Norling B, Radstrom P, Gorwa-Grauslund MF. Flotation as a tool for indirect DNA extraction from soil. Appl Microbiol Biotechnol 2010;87:1927–33.

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021

Park SJ, Kang CH, Chae JC, Rhee SK. Metagenome microarray for screening of fosmid clones containing specific genes. FEMS Microbiol Lett 2008;284:28–34.
Park SJ, Chae JC, Rhee SK. Application of DNA microarray for screening metagenome library clones. Methods Mol Biol 2010;668:313–24.
Pottkamper J, Barthen P, Ilmberger N, Schwaneberg U, Schenk A, Schulte M, et al. Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. Green Chem 2009;11:957–65.
Radajewski S, Webster G, Reay DS, Morris SA, Ineson P, Nedwell DB, et al. Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. Microbiology 2002;148:2331–42.
Rashamuse K, Magomani V, Ronneburg T, Brady D. A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous beta-lactamase activity on nitrocefin. Appl Microbiol Biotechnol 2009;83:491–500.
Roerig P, Nessling M, Radlwimmer B, Joos S, Wrobel G, Schwaenen C, et al. Molecular classification of human gliomas using matrix-based comparative genomic hybridization. Int J Cancer 2005;117:95-103.
Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nat Biotechnol 2008;26:1117–24.
Rowe-Magnus DA, Mazel D. Integrons: natural tools for bacterial genome evolution. Curr Opin Microbiol 2001;4:565–9.
Schmeisser C, Stockigt C, Raasch C, Wingender J, Timmis KN, Wenderoth DF, et al. Metagenome survey of biofilms in drinking-water networks. Appl Environ Microbiol 2003;69:7298–309.
Sharma S, Khan FG, Qazi GN. Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. Appl Microbiol Biotechnol 2010;86:1821–8.
Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008;26:1135–45.
Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005;309:1728–32.
Solbak AI, Richardson TH, McCann RT, Kline KA, Bartnek F, Tomlinson G, et al. Discovery of pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton fabric. J Biol Chem 2005;280:9431–8.
Streit WR, Schmitz RA. Metagenomics — the key to the uncultured microbes. Curr Opin Microbiol 2004;7:492–8.
Stull D, Pisano JM. Purely RNA: New innovations enhance the quality, speed, and efficiency of RNA isolation techniques. Scientist 2001;15:29–31.
Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, et al. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res 2010;20:1605–12.
Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, et al. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. J Biotechnol 2008;133:42–9.
Torsvik V, Ovreas L. Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 2002;5:240–5.
Troeschel SC, Drepper T, Leggewie C, Streit WR, Jaeger KE. Novel tools for the functional expression of metagenomic DNA. Methods Mol Biol 2010;668:117–39.
Turcatti G, Romieu A, Fedurco M, Tairi AP. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res 2008;36:e25.
Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 2004;428:37–43.
Uchiyama T, Miyazaki K. Substrate-induced gene expression screening: a method for high-throughput screening of metagenome libraries. Methods Mol Biol 2010;668:153–68.
Uchiyama T, Watanabe K. Substrate-induced gene expression (SIGEX) screening of metagenome libraries. Nat Protoc 2008;3:1202–12.
Uchiyama T, Abe T, Ikemura T, Watanabe K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol 2005;23:88–93.
Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, et al. Novel natural products from soil DNA libraries in a streptomycete host. Org Lett 2000;2:2401–4.
Wang F, Li F, Chen G, Liu W. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol Res 2009;164:650–7.
Ward N. New directions and interactions in metagenomics research. FEMS Microbiol Ecol 2006;55:331–8.
Weier HU, Greulich-Bode KM, Wu J, Duell T. Delineating rearrangements in single yeast artificial chromosomes by quantitative DNA fiber mapping. Open Genomics J 2009;2:15–23.
Wexler M, Bond PL, Richardson DJ, Johnston AW. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. Environ Microbiol 2005;7:1917–26.
Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452:872–6.
Wilkinson DE, Jeanicke T, Cowan DA. Efficient molecular cloning of environmental DNA from geothermal sediments. Biotechnol Lett 2002;24:155–61.
Wilson MS, Bakermans C, Madsen EL. In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. Appl Environ Microbiol 1999;65:80–7.
Wong DD, Chan VJ, McCormack AA, Batt SB. Cloning and characterization of an exo-xylogucanase from rumenal microbial metagenome. Protein Pept Lett 2010;17:803–8.
Wu L, Thompson DK, Li G, Hurt RA, Tiedje JM, Zhou J. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol 2001;67:5780–90.
Xu S, Hu Y, Yuan A, Zhu B. Cloning, expression and characterization of a novel esterase from marine sediment microbial metagenomic library. Wei Sheng Wu Xue Bao 2010;50:891–6.
Ye M, Li G, Liang WQ, Liu YH. Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. Appl Microbiol Biotechnol 2010;87:1023–31.
Yu EY, Kwon MA, Lee M, Oh JY, Choi JE, Lee JY, et al. Isolation and characterization of cold-active family VIII esterases from an arctic soil metagenome. Appl Microbiol Biotechnol 2011;90:573–81.
Yun J, Ryu S. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microb Cell Fact 2005;4:8.
Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, et al. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Appl Environ Microbiol 2004;70:7229–35.
Zhang D, Lax AR, Bland JM, Allen AB. Characterization of a new endogenous endo-beta-1,4-glucanase of Formosan subterranean termite (*Coptotermes formosanus*). Insect Biochem Mol Biol 2011;41:211–8.
Zhao S, Wang J, Liu K, Zhu Y, Bu D, Li D, et al. Screening and characterization of lipase from a metagenome library of dairy rumen microflora. Sheng Wu Gong Cheng Xue Bao 2009;25:869–74.
Zhao S, Wang J, Bu D, Liu K, Zhu Y, Dong Z, et al. Novel glycoside hydrolases identified by screening a Chinese Holstein dairy cow rumen-derived metagenome library. Appl Environ Microbiol 2010;76:6701–5.
Zhou S, Kile A, Kvikstad E, Bechner M, Severin J, Forrest D, et al. Shotgun optical mapping of the entire *Leishmania major* Friedlin genome. Mol Biochem Parasitol 2004;138:97-106.

Please cite this article as: Xing M-N, et al, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.01.021