Review

Biofpr Biofuels Bioproducts & Biorefining

From lignocellulosic metagenomes to lignocellulolytic genes: trends, challenges and future prospects

Ramón Alberto Batista-García and María del Rayo Sánchez-Carbente, Universidad Autónoma Estado Morelos (UAEM), Cuernavaca-Morelos, México
Paola Talia, Instituto de Biotecnología, INTA-Castelar, Argentina and Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina
Stephen A. Jackson, Niall D. O`Leary and Alan D. W. Dobson, University College Cork, Ireland
Jorge Luis Folch-Mallol, Universidad Autónoma Estado Morelos (UAEM), Cuernavaca-Morelos, México

Received April 17, 2016; revised July 31, 2016; and accepted August 1, 2016
View online at Wiley Online Library (wileyonlinelibrary.com); DOI: 10.1002/bbb.1709;
Biofuels, Bioprod. Bioref. **(2016)**

Abstract: Lignocellulose is the most abundant biomass on Earth with immense potential to act as a primary resource for the production of a range of compounds currently obtained from fossil fuel sources. However, lignocellulosic feedstocks remain largely underexploited due to the complex mixture of recalcitrant polymers present, whose structural features hinder access to the utilizable monosaccharide reservoir within cellulose. Various fungi and bacteria have been identified that can enzymatically decompose lignocellulose to its monomeric compounds for use as carbon sources. The investigation of such lignocellulolytic organisms has proven very useful in gaining primary insights into degradation processes and key microbial enzymes, but the established limitations of culture-based approaches suggest that we have yet to understand the full range of lignocellulolytic mechanisms, likely expressed within natural systems. In this review, we focus on metagenomic approaches to study lignocellulose degradation from structural and functional perspectives, which may provide novel insights into this process in order to rationally design methods for the extraction of compounds from biomass that could enhance biorefinery efficiencies. © 2016 Society of Chemical Industry and John Wiley & Sons, Ltd

Keywords: metagenomics; lignocellulolytic enzymes; guts; lignocellulose degradation; biorefineries

Correspondence to: Ramón Alberto Batista-García, PhD, Centro de Investigación en Dinámica Celular (CIDC)-UAEM, Av. Universidad 1001, Chamilpa. Cuernavaca-62209, Mexico. E-mail: rbatista25@yahoo.com, rabg@uaem.mx;
Jorge Luis Folch-Mallol, PhD, Centro de Investigación en Biotecnología (CEIB)- UAEM, Av. Universidad 1001, Chamilpa Cuernavaca 62209, México. E-mail: jordi@uaem.mx

SCI

Introduction

The current energy crisis requires urgent solutions to satisfy the increasing demands for fossil fuels. Total world petroleum consumption in 2014 was 92.42 million barrels/day (mb/d) (https://www.eia.gov/forecasts/steo/report/global_oil.cfm), while demands for crude petroleum are projected to increase by ~25% to 116 mb/d by 2030.[1] With respect to world energy consumption, 78.3% is derived from fossil fuels and 19.1% is based on renewable energy sources, of which only 0.8% is attributable to established biofuels (e.g. bioethanol and biodiesel).[2] Despite the lack of clarity regarding accessible oil reserves and the number of decades for which we can rely on same, society has clearly reached a watershed in resource dependency.[3] Adaptation and diversification have thus emerged as key goals of academic and industrial research, as we seek feasible routes toward sustainable fuel resources.

Against this backdrop, bioethanol has become firmly established as an alternative renewable energy source, capable of displacing petrochemical fuels through blending with gasoline.[4] Indeed, bioethanol world production increased by more than 300% between 2004 and 2014 (28.5 *vs* 94 billion liters).[5] Bioethanol is mainly produced from sugarcane and corn, (78% of world production), while only 4.2% is derived from biomass.[1] Bioethanol production from lignocellulosic material is currently quite challenging however, because of a dearth of cost-effective break-through technologies to facilitate the conversion of plant biomass into alcohol.[6] Biomass saccharification is a complex process, typically leading to quite low yields, and is often regarded as the critical conversion step.[7]

Plant biomass is the most abundant and widespread material on Earth (10^9 tons/annum) and represents a significantly under-utilized resource at present.[8] It is predominantly composed of cellulose, hemicellulose, pectin, and lignin and requires a number of enzymes to be completely decomposed in its primary components (mainly monosaccharides, organic acids, and phenolic alcohols).[9] In fact, we may still be missing some of the necessary activities involved in complete decomposition of lignocellulose and the order in which these enzymes act it is still not well defined.[10–12] However, the main enzymes in lignocellulose decomposition have been widely studied.

Lignin is probably the most recalcitrant component of plant biomass. It is an amorphous polymer composed of three aromatic alcohols: ρ-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, which may vary in their relative abundance depending on the plant species. In nature, degradation of lignin takes place by random oxidation of aryl-aryl bonds by oxidases produced by fungi and bacteria. The main enzymes in lignin degradation are several types of peroxidases (lignin peroxidases, manganese peroxidases, and versatile peroxidase) and laccases. The first group uses hydrogen peroxide produced by specific enzymes, while laccases take molecular oxygen from the environment to accomplish lignin oxidation.[13, 14]

Cellulose is a linear glucose polymer joined by β-1,4-glycosidic bonds, which can adopt several structures with different degrees of order that affect its susceptibility to be degraded by cellulases.[15] In its crystalline form (a very ordered structure stabilized by intra and inter molecular hydrogen bonds) it is insoluble in water and practically inaccessible to enzymatic degradation. There are three main types of cellulases: endo glucanases, which can hydrolyze the glycosidic bonds in internal regions of disorganized cellulose fibers (called amorphous or non-crystalline cellulose); exo-glucanases, that can degrade cellulose by the reducing or non-reducing ends of the cellulose fibres liberating the disaccharide cellobiose, and β-glycosidases, which finally produce glucose from cellobiose.[11] Recently, a novel kind of enzyme has been involved in cellulose degradation called polysaccharide monooxygenases. These are metalloproteins, which can generate cupper radicals that can cleave crystalline cellulose and are proposed to act synergistically with canonical cellulases.[16, 17]

Hemicelluloses are complex polymers that include xylans, xyloglucans, glucomannans, and mannans, which are composed of several kinds of sugars of which xylose is a principal component (except for mannans and glucomannas, where mannose is the main backbone sugar). Usually a xylose backbone joined by β-1,4-glycosidic bonds is then branched with other sugars as arabinose, glucose, galactose, fructose, etc., in β-1 or β-6 positions. Organic acids such as ρ-coumaric, ferulic or acetic acid may be found esterified to some hemicelluloses in different degrees.[18] For hemicellulose degradation thus, a number of enzymes are needed which include glycosyl hydrolases (like endo and exo 1–4 β-xylosidases, α-D-galactosidase, α-D-glucuronidase, etc.) and esterases (feruloyl esterase, acetyl xylan esterase, etc.).[19]

Pectins are also branched polymers of α 1–4 linked galacturonic acid in which arabinose and galactose are the most abundant sugars. In these polymers, ferulic acid can also be found esterified to the sugars. Enzymes involved in pectin degradation include endo- and exo-polygalacturonases that hydrolyze the galacturonic acid backbone, while the branched regions are degraded by endo- and

exo-rhamnogalacturonases, α-rhamnosidases and xylogalacturonases, among others.[19]

A number of microbial enzymatic formulations (e.g. Celluclast® 1.5L, HTec2® Enzymes, Novozyme 188, and Cellic® CTec series from Novozymes, Accellerase® TRIO from Genencor, CMAX3® and 4® both from Dyadic) are currently applied in commercial industries to release fermentable sugars from cellulose, hemicellulose and pectin in order to facilitate bioethanol, biogas and biopolymer production.[9, 20] However, the overall recalcitrant, crystalline and amorphous structure of lignocellulose impedes access of these enzymes to their target substrates.[9, 21] Some examples are the use of these commercial preparations for pulp bleaching in the paper industry; in the food industry to release antioxidants and to improve of yields in starch and protein extraction from fruit and clarification of fruit juices; in laundry by using cellulase-based detergents with improved cleaning action; in the textile industry for removal of excess dye from fabrics and restoration of color brightness, etc.[22]

Our understanding of how to efficiently transform lignocellulose through technologically sustainable processes is currently incomplete. In order to fully deliver on the goals of low cost, sustainable production of bioethanol from biomass, it is essential that we continue to screen for robust, bio-catalytic enzymes for lignocellulose degradation.[20] In an effort to address this issue several research groups are currently focused on exploiting the genetic diversity of microbial communities inhabiting natural environments associated with lignocellulosic biomass disposal and biodegradation.

Metagenomic-based approaches

Due to their immense metabolic diversity, micro-organisms possess the ability to colonize a wide variety of natural and anthropogenic environments, contributing to critical biogeochemical processes of organic and inorganic nutrient cycling.[23, 24] Current estimates suggest that approximately 4–6 x 10^{30} bacteria may inhabit the earth, with around 2.6 x 10^{29} micro-organisms proposed to reside in soil and 1.2 x 10^{29} in the open oceans.[25] With respect to terrestrial habitats, it is recognized that circa 99% of bacteria cannot be cultured, while studies in marine ecosystems suggest that as few as 0.001–0.1% of microbes are currently cultivable.[26] Culture-based techniques therefore only facilitate a limited recovery of <1% of total microbial biodiversity.[20, 27] As a result our current exploitation of pure-culture derived microbial enzymes for biofuel production, is highly unlikely to reflect the full biocatalytic potential encapsulated within microbial biodiversity.

In an effort to address this, metagenomic-based, and culture-independent approaches have been developed over the past few decades to access and analyse the biodiversity in different environments (Fig. 1). In addition to enabling microbial community diversity profiling, metagenomics also provides an opportunity to investigate novel genes/proteins of biotechnological value, while circumventing the traditional limitation of species cultivation. Other *omics*-based approaches are also now being routinely employed to analyse microbial metabolic biodiversity (Fig. 1).

There are two principal objectives within metagenomics studies which can be performed in isolation or in parallel, namely structural and functional (Fig. 2). The first seeks to describe the major genera and species that inhabit an ecosystem, providing an ecological profile by which to propose potential roles in biogeochemical cycles, possible ecological interactions and evolutionary aspects. Functional metagenomics, on the other hand, seeks to explore genomic diversity within an environmental sample to isolate novel genes/pathways encoding functional enzymes and/or synthesizing novel biomolecules.[28] Successful applications of functional metagenomic strategies to date have resulted in the isolation and identification of entirely novel protein families forming deeply branched phylogenetic lineages. This is particularly true in the case of lignocellulolytic enzymes, including cellulases, xylanases, esterases and lipases.[29–31]

While a large number of environments have been studied using metagenomics, few have involved lignocellulosic rich ecosystems.[32–35] It is possible that the structural complexity and the chemical composition of lignocellulosic materials limit the extent to which microbes can colonize these recalcitrant environments, while also hindering the essential, primary step of high-quality DNA extraction. However, lignocellulosic materials represent the best options to study lignocellulosic microbes and to explore the catabolic potential of associated, non-culturable populations. Environments containing, for example, sugarcane bagasse (SCB), wheat straw, corn stover, agave fibres or rice straws are therefore ideal for functional metagenomic based approaches for the isolation and identification of novel cellulases, xylanases, ligninases, and esterases/lipases from as yet uncharacterized microbial populations within these different lignocellulosic rich biomasses. SCB, one of the most recalcitrant wastes in agriculture, is composed of 35–50% cellulose and 20–30% each of hemicellulose and lignin.[35] The increasing interest in bagasse-based

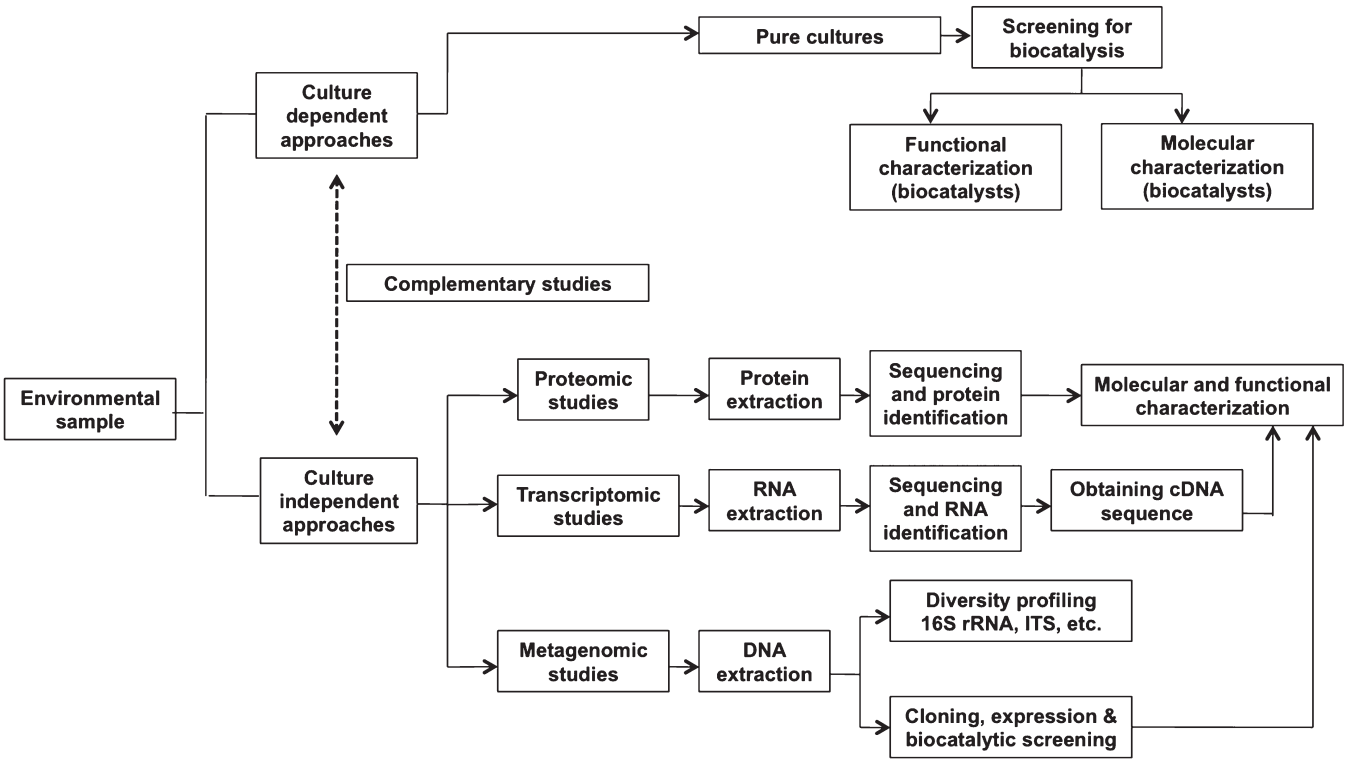

Figure 1. Culture dependent and culture independent approaches for biocatalytic screenings.

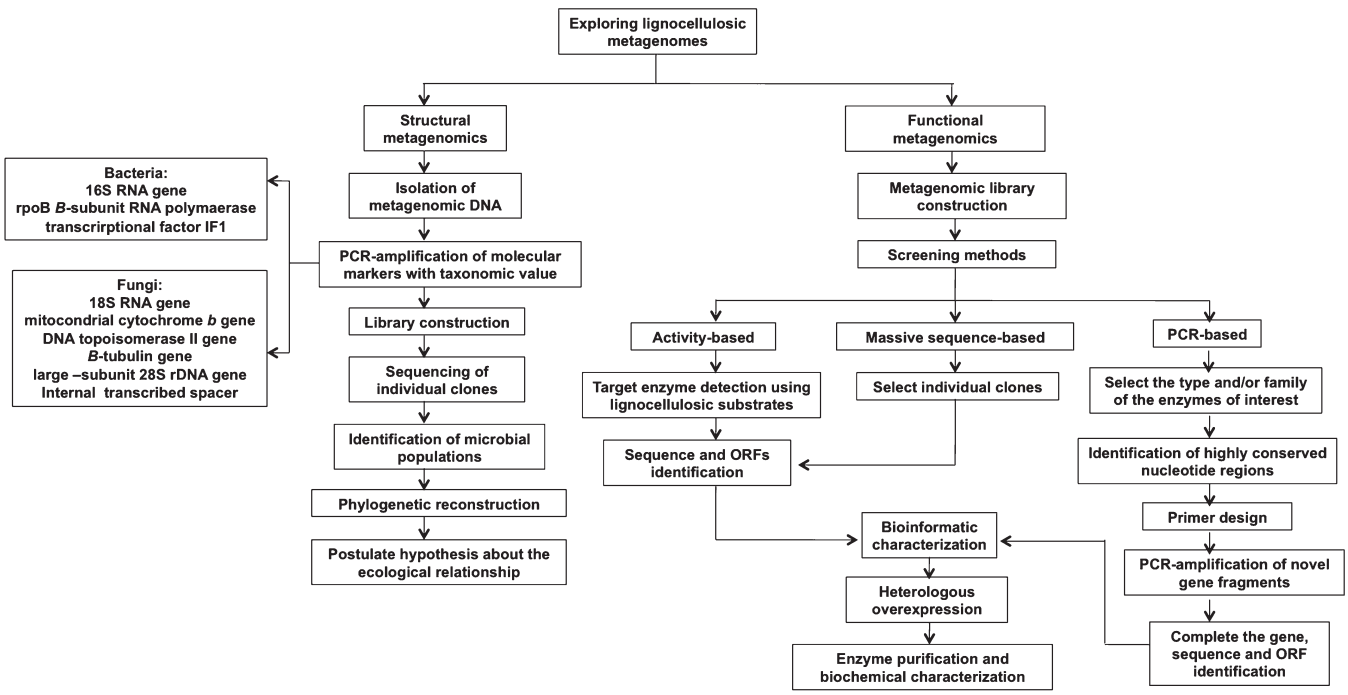

Figure 2. Structural and functional metagenomic to investigate lignocellulosic materials.

biorefineries provides an obvious driver for metagenomics derived-studies to maximize access to, and mining of, the enzymatic resources within the unculturable lignocellulose degraders colonizing such material.[34]

Challenges to metagenomic applications in lignocellulosic rich environments

The steps involved in the construction and functional screening of metagenomic libraries are well established (Fig. 3), but there are a number of specific, technical limitations associated with lignocellulose rich ecosystems. High-quality DNA is critical in the construction of libraries but lignocellulosic materials present many contaminants during extraction. Acids, furan derivatives, and phenolic compounds are often concomitantly extracted with the DNA.[36] These can contribute to the denaturation of nucleic acids, interfere with DNA transformation and inhibit numerous enzymes required during library preparation.[37] Furthermore, plant biomass may also contain

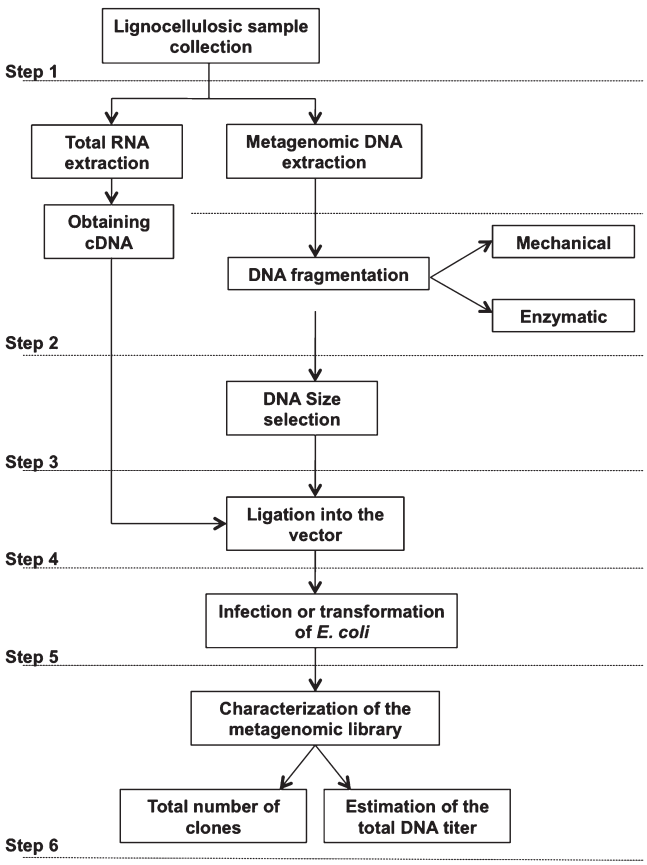

Figure 3. General workflow to obtain a metagenomic library.

fertilizers, preservatives and/or stabilizers from industrial processes; which can also affect the integrity of the DNA isolated. The impact of these contaminants is further compounded by the low yields of metagenomic DNA (mDNA) reported from lignocellulosic materials; attributed to limited microbial colonization of this recalcitrant substrate. In addition, it has been reported that variations in sample granularity may also affect yields.[33,36] As a result, standardized methods have not been established to date for mDNA extraction from lignocellulosic rich sources, with modified versions of existing protocols routinely being employed.[33,35]

One solution to the problem of limiting mDNA yields due to low microbial loads within lignocellulosic samples is to incorporate pre-enrichment strategies. Pre-enrichment via cellulose addition to samples has been employed to increase the potential of finding genes encoding cellulases, xylanases and lipases/esterases within subsequent mDNA libraries.[38] In addition, it has been reported that prokaryotic or eukaryotic enrichment by size-selective filtration or centrifugation processes can be successfully incorporated prior to metagenomic analyses.[39,40] Such procedures could be particularly beneficial where a study sought to exclusively investigate bacterial or fungal lignocellulolytic enzymes. Furthermore, Mori *et al*. have demonstrated that it is also possible to combine pre-enrichment and prokaryotic/eukaryotic pre-selection strategies in metagenomic studies of cellulolytic communities.[39] It must be noted however that when microbial populations undergo enrichment, the structural studies lose relevance due to the exogenous alteration imposed on the natural biomass ecosystem.

After purification of mDNA, it is typically size fractionated and subsequently cloned either into plasmids (<20 kb insert size), cosmids and fosmids (< 40 kb insert size) or Bacterial Artificial Chromosomes (BACs) (>40 kb insert size), depending on what is being targeted in the functional screening of the metagenomic libraries. Given that genes involved in related metabolic pathways are typically clustered in microbial genomes (e.g. operons or super-operonic clusters), the preferred choice is to clone the mDNA into cosmids or fosmids for functional screening.[41,42] This approach has been successfully used to study the genetic organization, diversity and function of denitrification genes organized into clusters/operons.[43, 44] In contrast, short inserts cloned into plasmid do not allow the recovery of large clusters of genes and can be less productive in functional metagenomic strategies.[45] Plasmids with promoters on both sides of a multiple clone site enabling bidirectional transcription can

increase the number of positive clones in plasmid-based libraries. The use of this approach in a metagenomic screen for lipolytic-esterases yielded similar numbers of positive clones when compared with yields from equivalent cosmid libraries.[46] In addition, as gene expression is heavily host-dependent, it is advantageous to employ broad-host range systems to maximize the potential for successful expression and detection of the genes being targeted.[47] *Escherichia coli* is a well-established and commonly employed host for the efficient, cost-effective, high-level production of many heterologous proteins.[48] Despite this, the use of *E. coli* as a heterologous host may have somewhat limited the number of lignocellulolytic enzymes that have to date been isolated from metagenomic libraries.[49] For example, the probability of finding lignocellulolytic enzymes of fungal origin is markedly reduced when bacterial systems such as *E. coli* are employed. Divergences in codon usage, promoter regulation/activation and RNA processing/translation inherently limit the effective functional expression of eukaryotic genes in prokaryotic systems. In addition, essential post-translational modifications such as glycosylation of eukaryotic cellulases and xylanases to facilitate secretion is deficient in prokaryotic hosts.[49–51] It is therefore not surprising perhaps that if one analyzes lignocellulolytic enzymes characterized to date from mDNA, the vast majority belong to prokaryotic proteins.[49]

Given the limitations in using *E. coli* as a heterologous expression host, then additional hosts should be considered such as *Pseudomonas putida, Burkholderia graminis, Bacillus subtilis, Ralstonia metallidurans, Caulobacter vibrioides, Thermus thermophilus, Sulfolobus solfataricus*, and *Streptomyces*; for which expression systems have been developed.[52–54] For example, *T. thermophilus* has been successfully employed as a metagenomic library host for the detection of esterases, yielding higher numbers of active esterase clones when compared with clone yields following library transformations into *E. coli*.[55] The system was also shown to be better than *E. coli* for the recombinant expression of xylanases.[56] Ongoing eukaryotic host system development will also play a key role in the full exploitation of metagenomic approaches for biomass degradation enzymes. It is anticipated that fungal hosts will facilitate post-transcriptional and post-translational modifications, function-dependent tertiary structure formation and signal peptide recognition/protein secretion to export cellulases, xylanases and esterases.[57] Some authors have proposed yeast species as ideal hosts to detect fungal xylose isomerase activity by function-based screening,[58] while others have reported the characterization of phosphatases by metatranscriptomic library expression in *Saccharomyces cerevisiae* host strains.[59]

Functional metagenomics: methodologies and challenges

Substrate selection and sensitivity

One of the most widespread functional screening approaches is to employ activity-based strategies.[28] These are based on the degradation/transformation of a substrate and usually rely on a color change or the development of a halo around the positive clone screened.[52] In these methods the suitability of the substrate(s) is the main bottleneck, while pH, temperature and salts can be considered as secondary restrictions, with the success rate typically correlated to the number of analyzed clones. Substrates must be carefully selected and should be broad spectrum. Tributyrin, xylan, and cellulose are excellent substrates for esterase, cellulase and xylanase detections. With respect to glycosidases more than 15 distinct chromogenic and fluorimetric substrates have been employed in functional screens, with varying levels of success. Azurine hydroxyethyl cellulose, a unique substrate for endo-celluloses, demonstrates a high degree of sensitivity capable of yielding high positive clone hit rates of 1 per 108 clones screened.[60] In contrast, bromo-4-chloro-3-indolyl-β-D-galactopyranoside, the common substrate for β-galactosidases, produces the lowest positive hit rate of 1:700,000.[61] An alternative approach is to incorporate more than one substrate into the screening media, which can yield higher hit rates when compared with the use of single substrates. This methodology has been successfully reported in a functional screen for (hemi)cellulase-degrading enzymes (beta-galactosidase, beta xylosidase and glycosyl hydrolases) from wheat straw-degrading microbial consortia. Seventy-one positive clones were detected, with two clones expressly requiring the presence of multiple substrates for detectable activity.[62]

A recent meta analysis of functional metagenomic screens informed over the last two decades reported that only 861 glycosidases has been discovered among a total of 6100 clones purportedly containing novel enzymes:[63] 4034 of the clones represented esterases/lipases; 859 were oxido-reductases; while 73 contained genes encoding proteases, amidases, nitrilases, phosphatases, dehalogenases, glycosyltransferases, penicillin G acylases and trehalose synthases, respectively.[63] The study also cited the incidence rate of positive clones among functional screens for six industrially relevant type of enzymes (acylases,

phosphatases, oxidoreductases, proteases, esterases/lipases, and glycosidases). Interestingly, glycosidases demonstrated the lowest overall incidence rate at 1 positive hit per 31 190 screened clones. By comparison, oxidoreductase functional screening revealed significantly higher rates of one positive hit per 6670 screened clones.[63]

In summary, functional metagenomic screens to date have revealed a consistently low incidence rate for the isolation of lignocellulolytic enzymes, with hit rates varying from 1:440 to 1:1,047 for glycosyl hydrolases,[62] and 1:40,000 to 3:40,000 for xylanases and cellulases, respectively.[39] The reported incidence rates are heavily dependent on multiple contributing factors including the enzyme activity being sought, the sensitivity of the substrate used and screening conditions employed, the metagenomic source and efficiency of the DNA extraction method and, the choice of cloning vector and expression host strains.

Sequence-driven metagenomic screens

Sequence-driven screening approaches are also used in functional metagenomics, as these are not dependent on the expression of cloned genes in foreign hosts. The targeting of specific classes of enzymes can be directly incorporated into PCR-based analyses of metagenomic DNA (Fig. 4), via highly conserved domains in particular enzymes.[28, 64] In these strategies primer design represents the critical step and can introduce intrinsic bias through a marked influence on the types and relative novelty of genes that may be amplified. Consensus and degenerate primers, and primers combining both regions, (CODEHOPs: Consensus-Degenerate Hybrid Oligonucleotide Primers), may be used to improve the success rate. Degenerate primers increase the probability of finding sequences that code for proteins with lower percentage identity than those used to design the primers. Conversely, consensus primers may introduce a bias for the amplification/detection of gene sequences encoding proteins very closely related to those used during primer design (Fig. 4). An additional challenge of targeting conserved sequences is the necessity to reconstruct the full gene sequence of any consensus hits, in order to attempt to characterize the encoded enzyme activity. It may not be easy to complete the sequence considering a complete library but methods such as Genome Walker and 5'-3' RACE have been successfully employed in the

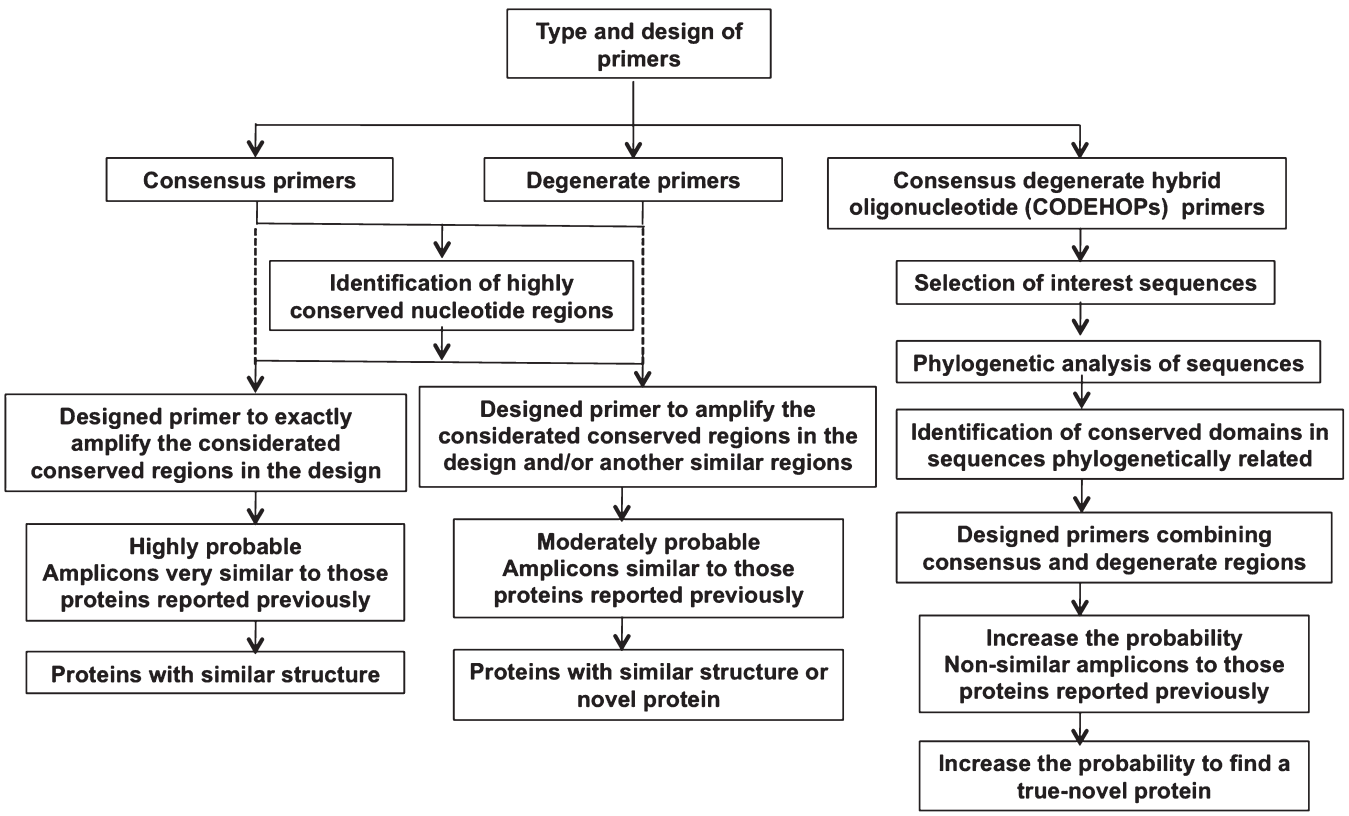

Figure 4. Consensus, degenerate and CODEHOPs primers for functional screenings.

past.[65] CODEHOPs have been successfully used to detect new genes (carboxypeptidase, xylanase, esterase, cellulase, cytochrome P450, etc.) in plants, animals and bacteria.[66–69]

The specific application of PCR-based metagenome screenings to identify new lignocellulolytic enzymes is hampered by an additional challenge. Cellulases, lipases/esterases and xylanases are grouped in more than 267 families between glycosidases, transferases, lyases, and esterases (CAZy: Carbohydrate-Active Enzyme, http://www.cazy.org/). Highly conserved amino acid regions are identifiable within certain families, which appear suitable for primer design. However, codon usage variations frequently result in low levels of conservation among the corresponding nucleotide sequences, hindering optimal primer design. The most useful sequence targets to date arise within domains that are important for binding to polysaccharides, as these enzymes share homology only in small regions of catalytic motifs. Moreover, it is impossible to consider all families or even all representatives of one family in the primers design. Despite these difficulties our own group has employed the use of CODEHOPs to successfully identify xylanase and lipase/esterase homologues in genomic investigations of *Bjerkandera adusta* (Sánchez-Carbente *et al*.; Batista-García *et al.,* unpublished).

Hybridization-driven screens have also been applied to functional metagenomics, based on the re-association of nucleic acid probes with homologous sequences within metagenomic DNA from environmental samples.[70] Gene detections can be achieved via the use of fluorophore-labelled nucleic acid probes in conjunction with fluorescence *in situ* hybridization (FISH) to reveal structural and/or functional insights within microbial communities.[71] Suppressive subtractive hybridization (SSH), provides an additional approach, which has been employed in the past to analyse the presence of closely related bacteria in animal rumen.[72] In addition to providing structural metagenomic insights, SSH has also been successfully utilized to identify unique functional genes present in rumen metagenomic samples in conjunction with particular dietary regimes.[73] Colony hybridization strategies (CH) can be also employed in functional metagenomic prospecting for novel enzymes. For example, Schwarz *et al.* applied a CH approach to successfully detect 36 positive clones for glycerol-dependent dehydratases using dehydratase-specific 1000-bp PCR amplicons as probes.[74] A similar CH approach has also been reported in the identification of novel P450 monooxygenases from a soil metagenome library.[75] The use of hybridization methodologies in functional-metagenomic screens to describe new lignocellulolytic enzymes is poorly described in the literature to date. However, some studies describe the relative abundance of cellulose degrading bacteria via *in situ* hibridisation.[76] In our laboratory, we have designed several probes with highly conserved carbohydrate-binding motifs for use in CH experiments. However, this strategy frequently yielded numerous false positive clones with respect to the enzymatic activity of interest (unpublished data). In conclusion, PCR-based and hybridization-based screenings require primers and/or probes, which are designed in basis on known genes, consequently limiting these approaches to identifying closely related genes with those previously known.

The ongoing development of low-cost, high-throughput sequencing platform technologies have also facilitated the application of massive sequencing projects to functional metagenomics. However major concerns persist with respect to fragment assembly and its practicality when compared with other methods for libraries with more than 500 000 clones. The approach detects genes on the basis of known conserved sequences, potentially limiting the capacity of this strategy to deliver truly novel biocatalysts. Despite these challenges, a recent application of this approach has successfully identified CAZy families useful in biofuel production, including cellulases, hemicellulases, amylases, cyclomaltodextrin glucanotransferase, galactosidases, mannosidases, glucosaminidases, sucrose phosphorylase, arabinofuranosidase, among others.[49]

Stable-isotope probing is a powerful tool in microbial ecology and it is a useful method to search for lignocellulolytic enzymes in metagenomes. This method can describe the ecological niche of the lignocellulosic communities and their interactions and even identify microorganisms involved in specific metabolic processes under conditions, which approach those occurring *in situ*.[77] Stable-isotope probing is also a technique with taxonomic value because it is a culture-independent procedure that allows the isolation of DNA from micro-organisms involved in specific degradation process. In stable-isotope probing techniques an appropriate substrate, for example cellulose if one is screening for cellulases, is enriched with a stable isotope ($(13)CH_3OH$ or $(13)CH_4$) prior to incorporation into the screen. In subsequent steps the DNA of active microorganisms is selectively recovered through density-gradient centrifugation.[78] Finally, the active microorganisms are identified via rRNA subunit sequencing and their genetic/metabolic versatility assessed via genomic approaches. The technique is more informative when combined with microarrays and metagenomics data.[78] With respect to future strategies to identify novel

lignocellulosic genes, it is clear from the above that a myriad of sequence driven strategies are available to maximize the opportunities for successful biocatalyst prospecting.

Once a sequence of interest is identified, it can be expressed in heterologous hosts by cloning the open reading frame by conventional PCR techniques. However, the selection of the host, promoter, codon usage, and post-translational modifications should be carefully taken into account. Eukaryotic genes in particular, which contain introns and their products, suffer post-translational modifications and are difficult to express successfully. If a metatranscriptome is isolated (instead of DNA), the problem with introns may be surpassed.[79] A successful example of the sequence-driven approach was achieved by Komori *et al*.[80] who cloned and expressed a metagenomic laccase gene and described the crystal structure of the protein expressed in *E. coli*. Another interesting alternative to overcome this problem is the use of synthetic genes; several companies sell this with accessible costs. In addition, codon optimization would be useful for enhancing protein expression.[81]

Metagenomic investigations of lignocellulose-associated microbial communities

Structural metagenomic diversity analyses

Several metagenome studies involving a variety of different sample sources such as forest soils,[82] tropical peat swamp forest,[83] switch grass-adapted compost community,[84] biogas reactors,[85] yak rumen,[86] and air-metagenome[87] have described the microbial diversity and potential metabolic capabilities of natural biomass decomposing populations. The composite picture emerging from such studies is a complex one, whereby community composition and relative taxonomic abundances are heavily influenced by niche environmental factors. Such factors include plant biomass chemistries, aerobic/anaerobic/anoxic conditions, temperature and/or pH ranges, availability of metals (e.g. Fe), enrichment activities and, spatial distribution within biomass systems. The scenario presents a plethora of opportunities for metagenomic/metaproteomic investigations as it suggests that novel discoveries are likely to continue to be delivered through the exploration of novel habitats. Several emerging issues should be considered in the promotion and execution of such exploration. Berlemont *et al*. reported that the distribution of glycosyl hydrolases among 8133 sequenced microbial genomes tended to cluster taxonomically at genus and species level.[88] The finding supports the diverse community compositions currently linked with varying habitats, but also raises the question of potential conservation of key species within same.

Proteobacterial genera have been reported as significant community constituents across several plant-biomass degrading systems, together with *Actinobacteria*, *Bacteroidetes,* and *Firmicutes*. Ventorino *et al*. identified the dominance of *Proteobacteria* among the microbiota associated with biomass piles of *Arundo donax*, *Eucalyptus camaldulensis,* and *Populus nigra*.[89] *Proteobacteria* also dominate bacterial inhabitants within compost[90] and within the 'fungal' gardens of leaf-cutter ants, in which fungi grow on the leaf biomass and support the ant colony as a food source.[91] A comparison of 12 taxonomic metagenome profiles from lignocellulosic habitats, (bagasse, compost, farm soil, peat swamp, termite gut, and wallaby gut) and non-lignocellulosic habitats, (fresh water, human and mouse gut, whale carcass, sludge, and seawater), was also recently reported.[34] It was noted that *Proteobacteria* dominated the community profiles in the open systems, but were remarkably absent among the 'closed' gut systems, potentially as a result of the dominant anaerobic nature of gut environs. Overall, the 12 communities were found to contain varying bacterial phyla abundances with the authors proposing that lignocellulosic degradation capacities may not necessarily be directly linked with a defined core community. Indeed comparison of the glycosyl hydrolase gene abundances in leafcutter ant fungus gardens and bovine rumen metagenomes were reported to be strikingly similar, while their taxonomic profiles were strongly divergent. The authors proposed an evolutionary convergence of enzymatic strategies for plant biomass degradation may have arisen on the premise that similarities in plant-biomass polymer compositions are likely encountered in these habitats.[91] Brossi *et al.* posited a similar theory following metagenomic analyses of a forest soil microbial community, enriched with wheat straw, corn stover, or switch grass, respectively.[92] A consortium of generalists comprising members of the genera *Sphingobacterium, Raoultella, Pseudomonas,* and *Stenotrophomonas* were identified across the three substrates, despite differing rates of lignin, cellulose, and hemicellulose degradation by the respective consortia. Thus, a core consortium targeting common saccharides/intermediates appears likely to be complemented with more specialist genera required to deconstruct unique chemical compositions, including bond structures and relative solubility of substrate specific polymers.

However, despite the importance of lignocellulosic material as a source of second generation biofuels, very few structural metagenomic studies on microbial communities specifically growing on lignocellulosic substrates (as the sole carbon source) have been reported (Table 1). The exception is sugarcane bagasse where such studies have identified aerobic and facultative anaerobic bacteria, together with cellulolytic and hemicellulolytic Ascomycota.[33, 35,93] Interestingly, Kanokratana *et al*. reported that sugarcane bagasse piles have distinctive micro-environmental conditions which markedly influence the microbial community structures within the piles.[33] These differences were dependent on oxygen limitation with the result that *Proteobacteria, Bacteroidetes* and *Acidobacteria* formed the representative phyla at the exterior base of the pile, while *Spirochaetes* were identified as the most abundant phyla in the anaerobic interior.

Spirochaetes have also been reported as a dominant species, (~50% of sequenced reads), within termite gut metagenomes while being largely absent from a wallaby gut sample (lignocellulosic), and in human and mouse gut equivalents (non-lignocellulosic). Thus, in addition to oxygen concentrations, niche variation in temperature and pH, which is typically high in termite guts, appears to distinguish community compositions further.[34] The significance of temperature and the enrichment of communities of thermophillic species capable of lignocellulose deconstruction are of considerable biotechnological significance. Recent studies suggest that cellulolytic enzymes from hyperthermophilic communities adapted to switch-grass substrate demonstrate co-resistance to high temperatures and ionic liquid exposures.[94,95] Imidazolium based ionic liquids are promising pre-treatment compounds for recalcitrant lignocellulosic substrates; however, they can inhibit cellulase cocktails for downstream saccharification, particularly fungal-derived cellulases. Metagenomic screening of a switch grass adapted community for thermotolerant enzymes yielded 21 suitable candidates, of which those with a T_{opt} >70% correlated with a 78% chance of being co-tolerant to high strength ionic liquids.[94] Further metagenomics-driven investigations of thermophilic systems are likely to progress this area, particularly when coupled with directed evolution strategies for enzyme optimisation as reviewed recently.[96] The findings highlight the importance of careful sampling of target environments, such as, for example sugarcane bagasse, not only from a substrate perspective re lignocellulosic

Table 1. Phyla distribution in lignocellulosic substrates.

	Metagenome source	Sugarcane Bagasse pile (Rattanachomsri *et al.*[35])	Poplar chips (Van der Lelie *et al.*[98])	Sugarcane Bagasse pile (Wongwilaiwalin *et al.*[93])	Sugarcane Bagassse soil (Mhuantong *et al*.[34])
	Phyla				
Bacteria	Acidobacteria	Yes (10.9%)	Yes (minor %)	NI	Yes (minor %)
	Actinobacteria	Yes (minor %)	Yes (minor %)	Yes (minor %)	Yes (7.9%)
	Bactereoidetes	Yes (15.3%)	Yes (9.9%)	Yes (23.8%)	Yes (10.2%)
	Chloroflexi	NI	NI	NI	Yes (minor %)
	Cyanobacteria	NI	Yes (minor %)	NI	Yes (minor %)
	Deinococcus-Thermus	Yes (minor %)	NI	NI	NI
	Firmicutes	Yes (35.5%)	Yes (45.9%)	Yes (50.1%)	Yes (minor %)
	Planctomycetes	Yes (minor %)	Yes (minor %)	Yes (minor %)	NI
	Proteobacteria	Yes (24.6%)	Yes (32.3%)	Yes (14.1%)	Yes (66.1%)
	Spirochaetes	Yes (minor %)	NI	Yes (5.1%)	NI
	Synergistetes	NI	NI	Yes (minor %)	NI
	Thermotogae	NI	NI	Yes (minor %)	NI
Fungi	Ascomycota	Yes	NI	ND	ND
	Basidiomycota	ND	NI	ND	ND
	Chytridiomycota	ND	NI	ND	ND
	Glomeromycota	ND	NI	ND	ND
	Zygomycota	ND	NI	ND	ND

ND=not determined, NI=non identified

material abundance, but also to take account of variations in micro-environmental conditions.[34,35] Members of the *Actinobacteria* and *Firmicutes* phyla have been reported in sugarcane bagasse, as well as in the metagenomes of soils and waste streams containing a high lignocellulose content.[97] A recent metagenomic characterization of a compost-derived lignocellulolytic community revealed a significant, and previously unrecognized, contribution (46.1%) of glycosyl hydrolase enzymes from *Actinobacteria*.[90] The versatile suite of enzymes contributed included cellobiohydrolases, β-glucosidases, acetyl xylan esterases, arabinofuranosidases, pectin lyases, and ligninase genes.

Additionally, a separate metagenomic study monitoring changes in microbial community structure during the anaerobic digestion of poplar chips identified quite different microbial phyla than those from other substrate-associated communities (Table 1).[98] The more abundant phyla were *Firmicutes* and *Proteobacteria* (45.9 and 32.3%, respectively), followed by *Bacteroidetes* (9.9%). Among the dominant members found in the phyla of *Protobacteria* is a bacterium similar to *Magnetospirillum*, which has been reported to play a role in the anaerobic breakdown of aromatic compounds.[98]

The majority of these metagenomic studies targeted the 16S rRNA gene to characterize the prokaryotic populations and, in some cases the Internal Transcribed Spacer (ITS) regions to monitor fungal populations; however the latter marker is known not to detect all fungal phyla.[99] Moreover, in other metagenomic studies fungal characterization was under reported due to a lack of specific fungal markers or because the number of reads obtained was too low to allow accurate further identification.[98] Wongwiliawilan *et al.* report that fungal species are not likely to be highly prevalent within the samples due to ecological niche conditions such as high temperature and relatively low oxygen potentially limiting their growth.[100] However, additional metagenomic analyses have indicated the clear presence of fungi, which could play a crucial role in the lignocellulose degradation.[34] At present it is difficult to posit any definitive conclusion regarding the contribution/significance of fungi and their lignocellulolytic enzymes as relevant metagenomic studies to date have likely underestimated both their presence and role.

Functional metagenomic identification of enzymes for biomass feedstock optimization

While a number of reviews have focused on functional screens for ligninocellulolytic enzymes from diverse environmentally sourced metagenomes, there has been limited data related to metagenomes from microbial populations growing on lignocellulosic-rich substrates.[101,102] In this section, we focus on the major discoveries over the last five years from the biomass-degrading enzymes belonging particularly to the GHases family. Cellulases and hemicellulases have long been recognized as very useful biocatalysts because of their wide-ranging versatility in industrial applications, including food technology and textile production and in paper and biofuel production. The composition of crystalline cellulose is quite homogenous in different types of plants; however, hemicellulose and lignin are polymers of quite diverse composition and/or linkages between monomers. This diversity is likely to have generated quite a high level of evolutionary pressure and, as a consequence, a wide range of enzyme diversification amongst microbes with the ability to use lignocellulose substrates as carbohydrate sources.[103] Table 2 provides a summary of the ligninocellulolytic enzymes recently characterized from a wide range of metagenome studies. Metagenomic libraries from sugarcane bagasse samples appear to be a particularly good source of GHases with rare properties, such as glucanases and xylanases.[104,105] Recently, an alkaliphilic xylanase has been characterized from a compost-soil metagenome, with the capacity to substitute the use of toxic chemicals in pulp-biobleaching and xylo-oligosaccharides generation in paper and textile industries, respectively.[106] In addition, an endoglucanase was isolated which exhibited sustained enzymatic activity over a wide pH range and temperatures of up to 80 °C; both useful properties for biorefinery processes.[107]

With respect to bioethanol production from lignocellulosic substrates, efficiencies can be increased via the consumption of xylose. Ethanol producing strains of *Saccharomyces cerevisae,* while lacking the ability to convert xylose into ethanol, can utilize its isomer D-xylose. The recent discovery of a diverse range of bacterial isomerase genes from soil metagenomes,[108] represent a potential opportunity to utilize these isomerase genes to increase the metabolic versatility of fermentative *S. cerevisiae* strains. Indeed, it has been demonstrated that the expression in *Saccharomyces cerevisiae* of isomerases isolated from a bovine rumen metagenomic library and a mammalian gut *Bacteroidetes* cluster allowed the yeast to consume xylose, and increased ethanol production up to 16.67 g/L.[109, 110]

Metagenomic mining of cellulolytic enzymes from insects

In selecting environments for lignocellulolytic communities, recognition has been given to the symbiotic relation-

Table 2. Hydrolytic enzymes of bacteria and fungi identified by culture-independent methods.

Culture-Independent Methods	Candidate glycoside hydrolase genes/clone	Substrate used	Number of Clones with Assayed and characterized Activity. Reference
Metagenome sequencing from switchgrass-adapted compost community (548 733 reads)	800 candidates (25 full length)	carboxymethyl-cellulose	1 Endoglucanase (GH9)[84]
Pyrosequencing (1 283 902 reads)	37 candidate genes	Avicel	4 Endoglucanases (GH5)[122]
Metagenomic library from sugarcane field land soil (26 900 clones of 1–8 kb size)	1 candidate gene	Carboxymethyl-cellulose	1 Endoglucanase (GH5)[29]
Metagenomic DNA from sugarcane bagasse compost	Several genes 1 candidate gene	Sugarcane bagasse and filter paper Xylan beechwood	1 Endoxylanase (GH10)[104] 1 Endoxylanase (GH10)[106]
Metagenomic DNA pyrosequencing (144 253 raw reads)	57 candidate genes	Carboxymethyl-cellulose, birchwood xylan or β-glucan	General presence of cellulases, and xylanases[93]
Metagenome cosmid library of yak rumen	4000 candidate clones	4-nitrophenyl-β-d-glucopyranoside (pNPG), 4-nitrophenyl-β-d-xylopyranoside (pNPX) Fluorescent 4-methylumbelliferyl-β-d-xylopyranoside (MuX)	1 β-glucosidase (GH3), 1 β-xylosidase (GH3)[123] 1 β-xylosidase (GH 43), 1 α-l-arabinofuranosidase (GH 30)[124]
Metagenome DNA library of cow rumen	Not mentioned (referred to Wong *et al*.)[125]	p-nitrophenyl ferulate	1 feruloyl esterase[126]
Metagenome fosmid DNA library of bovine rumen	70 000 clones	Carboxymethyl-cellulose and birchwood xylan	2 Bifunctional Cellulase–Xylanase (GH5)[127]
Metagenome fosmid library from sugarcane bagasse sample	7 candidates	AZCL-Xylane, AZCL-HE-Cellulose, AZCL-β-glucan, AZCL-xylo-glucan	1 GH-11 endoxylanase 1 GH-9 endoglucanase[105]

ships such communities may play within insect species capable of lignocellulose consumption. Termites secrete enzymes which allow them to use lignocellulosic substrates as energy sources.[111] In fact, termites are the most efficient decomposers of wood on Earth,[112] and the most studied with respect to their cellulolytic systems. Insects, therefore, represent a unique resource from which to isolate novel and efficient cellulolytic enzymes although they are not ubiquitous. In certain insects, such as *Drosophila melanogaster*, *Anopheles gambiae,* and *Bombyx mori*, these types of enzymes are absent and they are unable to degrade lignocellulosic substrates.[113] In addition, where cellulolytic activities are detected in insects these can be attributed to endogenous enzymes and/or enzymes from symbiotic microorganisms in their gut. *Limnoria* species (wood borers), do not contain microorganisms in their digestive tracts, but produce solely endogenous enzymes necessary for lignocellulose degradation.[114] Indeed, for many years the endosymbionts in termites were believed to be the predominant lignocellulose degraders; however, more recent evidence has demonstrated the important role that endogenous insect enzymes play in this process. Tartar *et al*. and Scharf *et al.* reported that several endogenous termite enzymes are expressed at higher levels than enzymes of their intestinal microbiota.[115, 116] Despite this changing perspective on the complexity of insect degradation of lignocellulose, culture independent approaches and molecular analyses have been employed to assess their biodiversity and lignocellulolytic potential (Table 3).[115] Culture independent approaches involving the screening of cDNA libraries have also been used to identify both endogenous and symbiotic insect cellulases.[117]

In termites, most reports have focused on their endogenous cellulases, with some being expressed in heterologous host systems such as *Pichia pastoris* and *Aspergillus oryzae*.[112] Nimchua *et al.* identified 14 clones with cellulase and xylanse activities from a metagenomic fosmid library of *Microcerotermes* sp., a wood-feeding higher termite. The enzymatic activities of three of these clones have been characterized following heterologous expression in *E. coli*

Table 3. Hydrolytic enzymes of insect and their endosymbionts identified by culture-independent methods.

Order/Species	Culture-Independent Methods	Origin	Candidate genes/clone	Number of assayed and character-ized clone activity. Reference
Termite/*Nasutitermes aphratea*	454 pirosequencing	Hindgut Symbiont	33 symbiotic bacterial	NP[119]
Termite/*Reticulitermes flavipes*	Metatranscriptomic cDNA library	Endogenous and Symbiont	171 candidate gene encod-ing lignocellulases	NP[116]
Termite/*Reticulitermes speratus*	cDNA libraries	Endogenous	1 endoglucanase	1 enzyme[128]
Termite/*Microcerotermes sp.*	Metagenomic fosmid libraries	Endogenous	2 cellulases 12 xylanases	3 enzymes[118]
Termites/*Macrotermes annandalei*	Metagenomic fosmid libraries	Gut Symbiont	13 gene encoding cellulases	1 xylanase [129]
Termite/*Reticulitermes flavipes*	cDNA libraries and macroarrays	Endogenous and Symbiont	4 cellulases	NP[130]
Termite/*Pseudacanthotermes militaris*	Two Fosmid libraries	Gut symbiont	101 positive clones	6 enzymes candidates[131]
Termites/fungus-growing termite, *Macrotermes annandalei*	Fosmid libraries 454 pyrosequencing	Gut Symbiont	10 gene encoding putative β-glucosidase	3 enzymes[132]
Mastotermitidae/*Mastotermes darwinenesis* Termopsidae/*Hodotermopsis sjoestedti* Kalotermitidae/*Neotermes koshunensis* Rhinotermitidae/*Reticulitermes speratus* Cryptocercidae/*Crytocercus puntulatus*	cDNA libraries	Symbiotic protists	74 cellulases/ 24 xylanases 89 cellulases/ 17 xylanases 90 cellulases/ 73 xylanases 74 cellulases/ 11 Xylanases 31 cellulases/ 14 xylanases	NP[120]
Grasshopper/*Acrida cinerea* Cutworm/*Agrotis ípsilon* *Termite/Nasutitermes sp.*	Metagenomic sequencing. Illumina genome analyzer II	Gut Symbiont	31 (*A. cinerea*), 40 (*A. ípsilon*) and 52 (*Nasutitermes sp.*)	4 enzymes[121]
Termite/*taxonomic identification was not conducted*	Genomic libraries	Gut Symbiont	1 xylanase 3 xylanases	4 enzymes[133]
Coleoptera/*Apriona germari*	cDNA libraries	Endogenous	1 endoglucanase	1 enzyme[134]
Coleoptera/*Phaedon cochleariae*	cDNAs gut library	Endogenous and Symbiont	7 genes encoding putative alpha-amylase, cysteine proteinase, trypsin, chymo-trypsin, cellulase, pectinase and xylanase	NP[135]
Limnoriidae/*Limnoria quadripunctata*	Transcriptome	Endogenous	4 cellulases	NP[114]
Coleotera/*Chrysomela tremulae*	454 pyrosequencing	Endogenous	20 gene encoding cellulases	NP[117]
NP = not provided by the authors or enzymatic activity has not confirmed				

and were found to have optimal activities at pH 8.0 and 50 to 55°C and to have potential utility in pulp biobleaching and in biostaining of denim.[118] Metatranscriptomic analyses of both endogenous and symbiotic cDNA libraries has been performed on the gut of *Reticulitermes flaviceps,* from which 171 candidate genes were identified encoding lignocellulases, and from which functional analysis of phenoloxidase activity was subsequently performed.[116] Phenoloxidases obtained in the previous study were induced by lignin and identified as laccases or xenobiotic

metabolism/detoxification associated enzymes, which were found in *R. flavipes* but not in the symbionts. These activities, both phenoloxidases/laccases and xenobiotic metabolism/detoxification associated enzymes, are useful in lignocellulose deconstruction for bioethanol production.[116]

The metagenomic analysis of hindgut microbiota of the higher termite *Nasutitermes ephratae* also revealed a large number of genes involved in cellulose and hemicellulose degradation.[119] Genes associated with other relevant symbiotic functions such as H_2 production, reductive acetogenesis and N_2 fixation were also identified.[119] Similar work in this area has focused on analyzing the gut symbionts of four representative lower termite species and a wood-feeding cockroach by metatranscriptomics in order to compare the lignocellulose-degrading system in these species.[120] This has resulted in the isolation of around 4000 clones as meta-expressed sequence tags from cDNA libraries, and the identification of cellulases and hemicellulases in more than 10% of the clones from each library. More recently, metagenomic sequencing analysis of the gut symbionts of the grasshopper (*Acridacinerea*) and the cutworm (*Agrotisipsilon*) involving a comparative analysis of previously reported metagenomic analysis of the termite gut microbiome concluded that the grasshopper could be a good candidate for the discovery of biocatalysts due to the high cellulolytic activities in its gut.[121] Thus with the abundance of ligninocellulolytic activities in insects, it is clear that these enzymes possess a very high potential from a biotechnological perspective in bioenergy production from renewable plant material. For this reason, an increase in the number of studies focusing on endogenous and symbiotic cellulolytic enzymes can be expected in the near future.

Conclusions and future perspective

In reviewing the state of the art of metagenomic investigations into lignocelluloytic microbial communities and their related enzymes, the authors have sought to highlight the clear biotechnological potential of such systems in tandem with the underexploited nature of same. Despite the abundance of lignocellulosic biomass on our planet and the essential biogeochemical recycling of same within ecosystems, there exists an apparent, recalcitrant lag in the rate of lignocellulolytic enzyme biodiscovery and industrial deployment. What has begun to emerge from successive metagenomic studies, demonstrating varying degrees of success, is the multivariate complexity associated with comprehensively screening lignocellulosic environments and their associated communities. The authors note that the approaches described were typically performed in isolation, confining their degree of relative success within the inherent limitations linked with each methodology. Future iterations of functional metagenomic screens could therefore benefit from strategically combined approaches, specifically selected to offset or complement the limitations imposed by any singular protocol. In this review we have identified several critical structural and functional metagenomic screen variables including; low levels of community diversity/colonization, the recalcitrance of such systems to yield high quality DNA, sub-optimal enzyme substrate sensitivities and heterologous host system expression, impacts of ecological niche microenvironments, narrow spectrum nucleic acid probes, reliability of sequence reassembly and an obvious under representation to date of fungal contributions. Despite these challenges, it is our view that significant opportunities remain for lignocellulosic community based metagenomic approaches to deliver the biocatalytic potential demanded for rapid industrial expansion of sustainable bioethanol production from lignocellulosic biomass.

References

1. Balan V, Current challenges in commercially producing biofuels from lignocellulosic biomass. *ISRN Biotechnology* **2014**: 463074 (2014).
2. Ren21, Key Findings. *Global Status Report* REN, Paris, pp. 1–33 (2015).
3. Owen NA, Inderwildi OR and King DA, The status of conventional world oil reserves-Hype or cause for concern? *Energy Policy* **38**:4743–4749 (2010).
4. Hansen AC, Zhang Q and Lyne PWL, Ethanol-diesel fuel blends – a review. *Bioresource Technol* **96**:277–285 (2005).
5. Foley T, Thornton K, Hinrichs-Rahlwes R, Sawyer S, Sander M, Taylor R *et al., REN21-Renewables 2015 Global Status Repor*t. REN, Paris, pp. 251 (2015).
6. Chen H and Fu X, Industrial technologies for bioethanol production from lignocellulosic biomass. *Renew Sustain Energ Rev* **57**:468–78 (2016).
7. Chen C, Ding S, Wang D, Li Z and Ye Q, Simultaneous saccharification and fermentation of cassava to succinic acid by *Escherichia coli* NZN111. *Bioresource Technol* **163C**:100–105 (2014).
8. Guerriero G, Hausman JF, Strauss J, Ertan H and Siddiqui KS, Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. *Eng Life Sci* **16**:1–16 (2016).
9. Liu S, Lu H, Hu R, Shupe A, Lin L and Liang B, A sustainable woody biomass biorefinery. *Biotechnol Adv* **30**:785–810 (2012).
10. Arantes V and Saddler JN, Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol Biofuel* **3**:4 (2010).
11. Quiroz-Castañeda RE and Folch-Mallol JL, Plant Cell wall degrading and remodeling proteins: Current perspectives. *Biotechnol Appl* **28**:194–215 (2011).

12. Saini JK, Saini R and Tewari L, Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* **5**:337–353 (2014).
13. Baldrian P, Fungal laccases – occurrence and properties. *FEMS Microbiol Rev* **30**:215–242 (2006).
14. Welinder KG, Superfamily of plant, fungal and bacterial peroxidases. *Curr Op Struct Biol* **2**:388–393 (1992).
15. Arantes V and Saddler JN, Cellulose accessibility limits the effectiveness of minimum cellulase loading on the efficient hydrolysis of pretreated lignocellulosic substrates. *Biotechnol Biofuel* **4**: 3 (2011).
16. Horn SJ, Vaaje-Kolstad G, Westereng B and Eijsink VG, Novel enzymes for the degradation of cellulose. *Biotechnol Biofuel* **5**:45 (2012).
17. Morgenstern I, Powlowski J and Tsang A, Fungal cellulose degradation by oxidative enzymes: from dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family. *Briefings in Functional Genomics* **13**:471–481 (2014).
18. Scheller HV and Ulvskov P, Hemicelluloses. *Ann Review Plant Biol* **61**:263–289 (2010).
19. De Souza WR, Microbial degradation of lignocellulosic biomass, in *Sustainable Degradation of Lignocellulosic Biomass-Techniques, Applications and Commercialization,* ed by Chadel AK and da Silva SS. INTECH, Brazil (2013).
20. Gilbert J, Li L-L, Taghavi S, McCorkle SM, Tringe S and van der Lelie D, Bioprospecting metagenomics for new glycoside hydrolases. *Meth Molec Biol (Clifton, NJ)* **908**:141–151 (2012).
21. Geng W, Huang T, Jin Y, Song J, Chang H-M and Jameel H, Comparison of sodium carbonate-oxygen and sodium hydroxide-oxygen pretreatments on the chemical composition and enzymatic saccharification of wheat straw. *Bioresource Technol* **161**:63–68 (2014).
22. Kuhad RC, Gupta R and Singh A, Microbial cellulases and their industrial applications. *Enzyme Res* DOI: 10.4061/2011/280696 (2011).
23. Falkowski PG, Fenchel T and Delong EF, The microbial engines that drive Earth's biogeochemical cycles. *Science (New York, NY)* **320**:1034–1039 (2008).
24. Leis B, Angelov A and Liebl W, Screening and expression of genes from metagenomes. *Adv Appl Microbiol* **83**:1–68 (2013).
25. Whitman WB, Coleman DC, Wiebe WJ, Prokaryotes: the unseen majority. *Proc Nat Acad Sci USA* **95**:6578–6583 (1998).
26. Kennedy J, Marchesi JR and Dobson ADW, Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microbial Cell Factories* DOI: 10.1186/1475-2879-7-27 (2008).
27. Nazir A, Review on metagenomics and its applications. *Imperial Journal of Interdisciplinary Research* **2**:277–286 (2016).
28. Simon C and Daniel R, Metagenomic analyses: past and future trends. *Applied and environmental microbiology* **77**:1153–1161 (2011).
29. Alvarez TM, Paiva JH, Ruiz DM, Cairo JPLF, Pereira IO, Paixão DAA *et al.*, Structure and function of a novel cellulase 5 from sugarcane soil metagenome. *PloS One* DOI: 10.1371/journal.pone.0083635 (2013).
30. Coughlan LM, Cotter PD, Hill C and Alvarez-Ordoñez A, Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontier Microbiol* **6**:1–22 (2015).
31. Couto GH, Glogauer A, Faoro H, Chubatsu LS, Souza EM and Pedrosa FO, Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genet Mol Res:GMR* **9**:514–523 (2010).
32. Batista-García RA, Casasanero R, Alvárez-Castillo A, Stephen A, Dobson ADW and Folch-Mallol JL, Prokaryotic diversity from the culture independent taxonomic analysis of a sugarcane bagasse metagenome. **4**:22–38 (2016).
33. Kanokratana P, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Eurwilaichitr L, Pootanakit K *et al.*, Phylogenetic analysis and metabolic potential of microbial communities in an industrial bagasse collection site. *Microb Ecol* **66**:322–334 (2013).
34. Mhuantong W, Charoensawan V, Kanokratana P, Tangphatsornruang S and Champreda V, Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. *Biotechnol Biofuel* **8**:197–212 (2015).
35. Rattanachomsri U, Kanokratana P, Eurwilaichitr L, Igarashi Y and Champreda V, Culture-independent phylogenetic analysis of the microbial community in industrial sugarcane bagasse feedstock piles. *Biosci Biotechnol Biochem* **75**:232–239 (2011).
36. Chandel AK, Antunes FAF, Anjos V, Bell MJV, Rodrigues LN, Polikarpov I *et al.*, Multi-scale structural and chemical analysis of sugarcane bagasse in the process of sequential acid–base pretreatment and ethanol production by *Scheffersomyces shehatae* and *Saccharomyces cerevisiae*. *Biotechnol Biofuel* DOI: 10.1186/1754-6834-7-63 (2014).
37. Lim HK, Chung EJ, Kim J-C, Choi GJ, Jang KS, Chung YR, *et al.*, Characterization of a Forest Soil Metagenome Clone That Confers Indirubin and Indigo Production on *Escherichia coli*. *Applied and Environmental Microbiology* **71**: 7768–77 (2005).
38. Vester JK, Glaring MA and Stougaard P, Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. *Extremophiles: Life under Extreme Conditions* **19**:17–29 (2015).
39. Mori T, Kamei I, Hirai H and Kondo R, Identification of novel glycosyl hydrolases with cellulolytic activity against crystalline cellulose from metagenomic libraries constructed from bacterial enrichment cultures. *SpringerPlus* **3**: 365 (2014).
40. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.*, Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, NY)* **304**:66–74 (2004).
41. Cheng J, Pinnell L, Engel K, Neufeld JD and Charles TC, Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. *J Microbiol Meth* **99**:27–34 (2014).
42. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR *et al.*, Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**:2541–2547 (2000).
43. Demaneche S, Philippot L, David MM, Navarro E, Vogel TM and Simonet P, Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. *Appl Environ Microbiol* **75**:534–537 (2009).
44. Suenaga H, Targeted metagenomics: A high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol* **14**:13–22 (2012).

45. López-López O, Cerdán ME and González Siso MI, New extremophilic lipases and esterases from metagenomics. *Current Protein & Peptide Science* **15**:445–55 (2014).
46. 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* **127**:575–592 (2007).
47. Yoon MY, Lee K-M, Yoon Y, Go J, Park Y, Cho Y-J *et al.*, Functional screening of a metagenomic library reveals operons responsible for enhanced intestinal colonization by gut commensal microbes. *Appl Environ Microbiol* **79**:3829–3838 (2013).
48. Hannig G and Makrides SC, Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotechnol* **16**:54–60 (1998).
49. Li L-L, McCorkle SR, Monchy S, Taghavi S and van der Lelie D, Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol Biofuel* **2**:10 (2009).
50. Flipphi M, Fekete E, Ag N, Scazzocchio C and Karaffa L, Spliceosome twin introns in fungal nuclear transcripts. *Fungal Genet Biol* **57**:48–57 (2013).
51. Veeresh J and Chuan WJ, Microbial cellulases: Engineering, production and applications. *Renew Sustain Energ* **33**:188–203 (2014).
52. Kennedy J, O'Leary ND, Kiran GS, Morrissey JP, O'Gara F, Selvin J *et al.*, Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J Appl Microbiol* **111**:787–799 (2011).
53. Li Y, Wexler M, Richardson DJ, Bond PL and Johnston AWB, Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. *Environ Microbiol* **7**:1927–1936 (2005).
54. Liebl W, Angelov A, Juergensen J, Chow J, Loeschcke A, Drepper T *et al.*, Alternative hosts for functional (meta)genome analysis. *Appl Microbiol Biotechnol* **98**:8099–8109 (2014).
55. Leis B, Angelov A, Mientus M, Li H, Pham VTT, Lauinger B *et al.*, Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Frontiers in Microbiology* **6**:1–12 (2015).
56. Angelov A, Mientus M, Liebl S and Liebl W, A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst Appl Microbiol* **32**:177–185 (2009).
57. Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R *et al.*, Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol for Biofuel* **7**:135 (2014).
58. Parachin NS and Gorwa-Grauslund MF, Isolation of xylose isomerases by sequence- and function-based screening from a soil metagenomic library. *Biotechnol Biofuel* **4**:9 (2011).
59. Kellner H, Luis P, Portetelle D and Vandenbol M, Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. *Microbiol Res* **166**:360–368 (2011).
60. Nguyen NH, Maruset L, Uengwetwanit T, Mhuantong W, Harnpicharnchai P, Champreda V *et al.*, Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. *Biosci Biotechnol Biochem* **76**:1075–1084 (2012).
61. Wang SD, Guo GS, Li L, Cao LC, Tong L, Ren GH *et al.*, Identification and characterization of an unusual glycosyltransferase-like enzyme with beta-galactosidase activity from a soil metagenomic library. *Enzyme Microb Tech* **57**:26–35 (2014).
62. Maruthamuthu M, Jiménez DJ, Stevens P and van Elsas JD, A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. *BMC Genomics* **17**: 86 (2016).
63. Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina OV and Golyshin PN, Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microbial Biotechnol* **9**:22–34 (2016).
64. Itoh N, Isotani K, Makino Y, Kato M, Kitayama K and Ishimota T, PCR-based amplification and heterologous expression of *Pseudomonas* alcohol dehydrogenase genes from the soil metagenome for biocatalysis. *Enzyme Microb Technol* **55**:140–150 (2014).
65. Cuervo-Soto LI, Valdés-García G, Batista-García R, Balcazar-López E, Lira-Ruan V, Pastor N *et al.*, Identification of a novel carbohydrate esterase from *Bjerkandera adusta*: Structural and function predictions through bioinformatics analysis and molecular modeling. *Protein: Structure, Function and Bioinformatic* **83**:3 (2015).
66. Jin H, Li B, Peng X and Chen L, Metagenomic analyses reveal phylogenetic diversity of carboxypeptidase gene sequences in activated sludge of a wastewater treatment plant in Shanghai, China. *Annals Microbiol* **64**:689–697 (2014).
67. Morant M, Hehn A and Werck-Reichhart D, Conservation and diversity of gene families explored using the CODEHOP strategy in higher plants. *BMC Plant Biology* DOI: 10.1186/1471-2229-2-7 (2002).
68. Parra LP, Espina G, Devia J, Salazar O, Andrews B and Asenjo JA, Identification of lipase encoding genes from Antarctic seawater bacteria using degenerate primers: Expression of a cold-active lipase with high specific activity. *Enzyme Microb Technol* **68**:56–61 (2015).
69. Zhang F, Chen JJ, Ren WZ, Lin LB, Zhou Y, Zhi XY *et al.*, Cloning, expression, and characterization of an alkaline thermostable GH11 xylanase from *Thermobifida halotolerans* YIM 90462T. *J Ind Microbiol Biotechnol* **39**:1109–1116 (2012).
70. Torsvik V, Daae FL, Sandaa RA and Øvreås L, Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol* **64**:53–62 (1998).
71. Nielsen JL, Christensen D, Kloppenborg M and Halkjær Nielsen P, Quantification of cell-specific substrate uptake by probe-defined bacteria under in situ conditions by microautoradiography and fluorescence in situ hybridization. *Environ Microbiol* **5**:202–211 (2003).
72. Nesbø CL, Nelson KE and Doolittle WF, Suppressive Subtractive Hybridization Detects Extensive Genomic Diversity in *Thermotoga maritima*. *J Bacteriol* **184**:4475–4488 (2002).
73. Galbraith EA, Antonopoulos DA and White BA, Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: The rumen as a model. *Environ Microbiol* **6**:928–937 (2004).
74. Schwarz S, Waschkowitz T and Daniel R, Enhancement of gene detection frequencies by combining DNA-based stable-isotope probing with the construction of metagenomic DNA libraries. *World J Microbiol Biotechnol* **22**:363–367 (2006).
75. Kim BS, Kim SY, Park J, Park W, Hwang KY, Yoon YJ *et al.*, Sequence-based screening for self-sufficient P450

monooxygenase from a metagenome library. *J Appl Microbiol* **102**:1392–1400 (2007).
76. O'Sullivan C, Burrell PC, Clarke WP and Blackall LL, A survey of the relative abundance of specific groups of cellulose degrading bacteria in anaerobic environments using fluorescence in situ hybridization. *J Appl Microbiol* **103**:1332–1343 (2007).
77. Radajewski S, Ineson P, Parekh NR and Murrell JC, Stable-isotope probing as a tool in microbial ecology. *Nature* **403**:646–649 (2000).
78. Dumont MG and Murrell JC, Stable isotope probing – linking microbial identity to function. *Nature Reviews Microbiology* **3**:499–504 (2005).
79. Guazzaroni M-E, Silva-Rocha R and Ward RJ, Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microb Biotechnol* **8**:52–64 (2015).
80. Komori H, Miyazaki K and Higuchi Y, Crystallization and preliminary X-ray diffraction analysis of a putative two-domain-type laccase from a metagenome. *Acta crystallographica Section F, Structural Biology and Crystallization Communications* **65**:264–266 (2009).
81. Chung BK-S and Lee D-Y, Computational codon optimization of synthetic gene for protein expression. *BMC Systems Biology* DOI: 10.1186/1752-0509-6-134 (2012).
82. Woo HL, Hazen TC, Simmons BA and DeAngelis KM, Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils. *Syst Appl Microbiol* **37**:60–67 (2014).
83. 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* **74**:1848–1854 (2010).
84. Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S *et al.*, Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* DOI: 10.1371/journal.pone.0008812 (2010).
85. Rademacher A, Zakrzewski M, Schluter A, Schonberg M, Szczepanowski R, Goesmann A *et al.*, Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing. *FEMS Microbiol Ecol* **79**:785–799 (2012).
86. Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L *et al.*, Metagenomic insights into the fibrolytic microbiome in yak rumen. *PloS One* **7**:e40430 (2012).
87. Cao C, Jiang W, Wang B, Fang J, Lang J, Tian G *et al.*, Inhalable microorganisms in beijing's PM 2.5 and PM 10 pollutants during a severe smog event. *Environ Sci Technol* **48**:1499–1507 (2014).
88. Berlemont R and Martiny AC, Genomic potential for polysaccharide deconstruction in bacteria. *Appl Environ Microbiol* **81**:1513–1519 (2015).
89. Ventorino V, Aliberti A, Faraco V, Robertiello A, Giacobbe S, Ercolini D *et al.*, Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application. *Scientific Reports* **5**:8161 (2015).
90. Wang C, Dong D, Wang H, Muller K, Qin Y, Wang H *et al.*, Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol Biofuel* **9**:22 (2016).
91. Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomas AA *et al.*, An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genetics* **6**: e1001129 (2010).
92. de Lima Brossi MJ, Jimenez DJ, Cortes-Tolalpa L and van Elsas JD, Soil-derived microbial consortia enriched with different plant biomass reveal distinct players acting in lignocellulose degradation. *Microbial Ecol* **71**:616–627 (2016).
93. Wongwilaiwalin S, Laothanachareon T, Mhuantong W, Tangphatsornruang S, Eurwilaichitr L, Igarashi Y *et al.*, Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. *Appl Microbiol Biotechnol* **97:**8941–8954 (2013).
94. Gladden JM, Park JI, Bergmann J, Reyes-Ortiz V, D'haeseleer P, Quirino BF *et al.*, Discovery and characterization of ionic liquid-tolerant thermophilic cellulases from a switchgrass-adapted microbial community. *Biotechnol Biofuel* DOI: 10.1186/1754-6834-7-15 (2014).
95. Shi J, Gladden JM, Sathitsuksanoh N, Kambam P, Sandoval L, Mitra D *et al.*, One-pot ionic liquid pretreatment and saccharification of switchgrass. *Green Chem* **15**:2579–2589 (2013).
96. Montella S, Amore A and Faraco V, Metagenomics for the development of new biocatalysts to advance lignocellulose saccharification for bioeconomic development. *Crit Rev Biotechnol* DOI: 10.3109/07388551.2015.1083939 (2015).
97. Hanreich A, Schimpf U, Zakrzewski M, Schlüter A, Benndorf D, Heyer R *et al.*, Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation. *Syst Appl Microbiol* **36**:330–338 (2013).
98. van der Lelie D, Taghavi S, McCorkle SM, Li LL, Malfatti SA, Monteleone D *et al.*, The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS One* DOI: 10.1371/journal.pone.0036740 (2012).
99. Prewitt ML, Diehl SV, Mcelroy TC and Diehl WJ, Comparison of general fungal and basidiomycete-specific ITS primers for identification of wood decay fungi. *Forest Prod J* **58**:66–71 (2008).
100. Wongwilaiwalin S, Rattanachomsri U, Laothanachareon T, Eurwilaichitr L, Igarashi Y and Champreda V, Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system. *Enzyme Microbial Technol* **47**:283–290 (2010).
101. Duan C-J and Feng J-X, Mining metagenomes for novel cellulase genes. *Biotechnol Lett* **32:**1765:1775 (2010).
102. Van Dyk JS and Pletschke BI, A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes--factors affecting enzymes, conversion and synergy. *Biotechnol Adv* **30**:1458–1480 (2012).
103. Dashtban M, Schraft H and Qin W, Fungal bioconversion of lignocellulosic residues; Opportunities & perspectives. *Int J Biol Sci* **5**:578–595 (2009).
104. Alvarez TM, Goldbeck R, Santos CRD, Paixão DAA, Gonçalves TA, Franco Cairo JPL *et al.*, Development and biotechnological application of a novel endoxylanase family gh10 identified from sugarcane soil metagenome. *PLoS One* DOI: 10.1371/journal.pone.0070014 (2013).
105. Kanokratana P, Eurwilaichitr L, Pootanakit K and Champreda V, Identification of glycosyl hydrolases from a metagenomic

library of microflora in sugarcane bagasse collection site and their cooperative action on cellulose degradation. *J Biosci Bioeng* **119**:1–8 (2014).

106. Weerachavangkul C, Laothanachareon T, Boonyapakron K, Wongwilaiwalin S, Nimchua T, Eurwilaichitr L *et al.*, Alkaliphilic endoxylanase from lignocellulolytic microbial consortium metagenome for biobleaching of eucalyptus pulp. *J Microbiol Biotechnol* **22**:1636–1643 (2012).
107. Verma D, Kawarabayasi Y, Miyazaki K and Satyanarayana T, Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. *PloS One* DOI: 10.1371/journal.pone.0052459 (2013).
108. Nurdiani D, Ito M, Maruyama T, Terahara T, Mori T, Ugawa S *et al.*, Analysis of bacterial xylose isomerase gene diversity using gene-targeted metagenomics. *J Biosci Bioeng* **120**:174–80 (2015).
109. Hou J, Shen Y, Jiao C, Ge R, Zhang X and Bao X, Characterization and evolution of xylose isomerase screened from the bovine rumen metagenome in *Saccharomyces cerevisiae*. *J Biosci Bioeng* **121**:160–165 (2016).
110. Peng B, Huang S, Liu T and Geng A, Bacterial xylose isomerases from the mammal gut Bacteroidetes cluster function in *Saccharomyces cerevisiae* for effective xylose fermentation. *Microbial Cell Factories* **14**:70 (2015).
111. Brune A, Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**:168–180 (2014).
112. Watanabe H and Tokuda G, Cellulolytic systems in insects. *Ann Rev Entomol* **55**:609–632 (2010).
113. Kunieda T, Fujiyuki T, Kucharski R, Foret S, Ament Sa, Toth AL *et al.*, Carbohydrate metabolism genes and pathways in insects: Insights from the honey bee genome. *Insect Mol Biol* **15**:563–576 (2006).
114. King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ *et al.*, Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proc Natl Acad Sci USA* **107**:5345–5350 (2010).
115. Scharf M, Termites as targets and models for biotechnology. *Ann Rev Entomol* **60**:77–102 (2015).
116. Tartar A, Wheeler MM, Zhou X, Coy MR, Boucias DG and Scharf ME, Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuel* **2**: 25 (2009).
117. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D and Ffrench-Constant RH, Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in *Coleoptera*. *Insect Biochem Mol Biol* **39**:403–413 (2009).
118. Nimchua T, Thongaram T, Uengwetwanit T, Pongpattanakitshote S and Eurwilaichitr L, Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *J Microbiol Biotechnol* **22**:462–469 (2012).
119. Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT *et al.*, Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560–565 (2007).
120. Todaka N, Inoue T, Saita K, Ohkuma M, Nalepa CA, Lenz M *et al.*, Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *PLoS One* **5**:1–10 (2010).
121. Shi W, Xie S, Chen X, Sun S, Zhou X, Liu L *et al.*, Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLoS Genetics* DOI: 10.1371/journal.pgen.1003131 (2013).
122. Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala Ha, Rowland SE *et al.*, Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nature Comm* **2**:375 (2011).
123. Bao L, Huang Q, Chang L, Sun Q, Zhou J and Lu H, Cloning and characterization of two beta-glucosidase/xylosidase enzymes from yak rumen metagenome. *Appl Biochem Biotechnol* **166**:72–86 (2012).
124. Zhou J, Bao L, Chang L, Zhou Y and Lu H, Biochemical and kinetic characterization of GH43 beta-D-xylosidase/alpha-L-arabinofuranosidase and GH30 alpha-L-arabinofuranosidase/beta-D -xylosidase from rumen metagenome. *J Ind Microbiol Biotechnol* **39**:143–152 (2012).
125. Wong DWS, Chan VJ and McCormack AA, Functional cloning and expression of a novel Endo-alpha-1,5-L-arabinanase from a metagenomic library. *Protein Peptide Lett* **16**:1435–1441 (2009).
126. Wong DWS, Chan VJ, Liao H and Zidwick MJ, Cloning of a novel feruloyl esterase gene from rumen microbial metagenome and enzyme characterization in synergism with endoxylanases. *J Ind Microbiol Biotechnol* **40**:287–295 (2013).
127. Rashamuse KJ, Visser DF, Hennessy F, Kemp J, Roux-van der Merwe MP, Badenhorst J *et al.*, Characterisation of two bifunctional cellulase-xylanase enzymes isolated from a bovine rumen metagenome library. *Curr Microbiol* **66**:145–151 (2013).
128. Watanabe A, Yano K, Ikebukuro K and Karube I, Cyanide hydrolysis in a cyanide-degrading bacterium, *Pseudomonas stutzeri* AK61, by cyanidase. *Microbiol (Reading, England)* **144**:1677–1682 (1998).
129. 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 Eviron Microbiol* **77**:48–56 (2011).
130. Scharf ME, Wu-Scharf D, Pittendrigh BR and Bennett GW, Caste- and development-associated gene expression in a lower termite. *Genome Biol* **4**:62 (2003).
131. Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Fauré R *et al.*, Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol Biofuel* **6**:78 (2013).
132. Zhang M, Liu N, Qian C, Wang Q, Wang Q, Long Y *et al.*, Phylogenetic and functional analysis of gut microbiota of a fungus-growing higher termite: bacteroidetes from higher termites are a rich source of β-Glucosidase genes. *Microb Ecol* **68**:416–425 (2014).
133. Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S *et al.*, Unusual microbial xylanases from insect guts. *Appl Environ Microbiol* **70**:3609–3617 (2004).
134. Lee SJ, Kim SR, Yoon HJ, Kim I, Lee KS, Je YH *et al.*, cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. *Compar Biochem Phys – B* **139**:107–116 (2004).
135. Girard C and Jouanin L, Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae*. *Insect Biochem Mol Biol* **29**:1129–1142 (1999).

Ramón A. Batista-García

Ramón A. Batista-García is head of the Laboratory of Extremophile Fungi at the CIDC of the UAEM, Mexico. He received his PhD in Molecular Biology and Biochemistry at UAEM, Mexico. Research in the Batista-García group focuses on the study of extremophile fungi and its biotechnological applications. Batista-García´s team employs omics approaches.

María del Rayo Sánchez Carbente

María del Rayo Sánchez Carbente is associate researcher in the Biotechnology Center at the University of Morelos (UAEM) since 2013. She did a Ph. D. in Biochemical Sciences at Institute of Biotechnology of the National University of México (UNAM) and a postdoctorate in the University of Montreal.

Paola Talia

Paola Talia is head of the Laboratory of Bioenergy at Biotechnology Institute, INTA. She is Professor of Renewable Energy II at the National University of San Martin (UNSAM) and a career research member at CONICET, Argentina. The team of Talia works on microbiome analysis of insects with genomics and metagenomics approaches for lignocellulosic bioethanol production.

Stephen Jackson

Stephen Jackson received his BSc. in Microbiology and his Ph.D in Marine Microbial Ecology, Metagenomics and Biotechnology, at University College Cork, Ireland, and at the Environmental Research Institute, UCC, where he is now a Senior Postdoctoral Researcher in the field of marine biodiscovery.

Dr. Niall O' Leary

Dr. Niall O' Leary is a Principle Investigator in the School of Microbiology and Environmental Research Institute at University College Cork Ireland, where he received his PhD in 2001. Dr. O' Leary has extensive experience in the investigation of microbial systems for bioremediation of waste materials and bioconversions into value added polymers.

Alan Dobson

Alan Dobson is Professor of Environmental Microbiology at University College Cork in Ireland. Research in the Dobson group focuses on gaining a fuller understanding of how microbes survive, grow and interact in their various ecological niches. They employ both genomic and metagenomic approaches to exploit the microbiota associated with marine and terrestrial environments with the goal of identifying novel bioactive compounds and biocatalysts with potential biotechnological applications.

Jorge Folch-Mallol

Jorge Folch-Mallol is head of the Laboratory of Fungal Molecular Biology at the CEIB of the UAEM, Mexico. He obtained his PhD in Spain in collaboration with Leiden University (Netherlands). Recently, he has done metagenomic studies describing microorganism diversity in lignocellulosic environments.

