

## Minireview

# Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening

María-Eugenia Guazzaroni,<sup>1</sup> Rafael Silva-Rocha<sup>2</sup> and Richard John Ward<sup>1\*</sup>

<sup>1</sup>Departamento de Química, FFCLRP,

<sup>2</sup>Departamento de Bioquímica e Imunologia, FMRP, University of São Paulo, Ribeirão Preto, SP, Brazil.

## Summary

**There is a growing demand for enzymes with improved catalytic performance or tolerance to process-specific parameters, and biotechnology plays a crucial role in the development of biocatalysts for use in industry, agriculture, medicine and energy generation. Metagenomics takes advantage of the wealth of genetic and biochemical diversity present in the genomes of microorganisms found in environmental samples, and provides a set of new technologies directed towards screening for new catalytic activities from environmental samples with potential biotechnology applications. However, biased and low level of expression of heterologous proteins in *Escherichia coli* together with the use of non-optimal cloning vectors for the construction of metagenomic libraries generally results in an extremely low success rate for enzyme identification. The bottleneck arising from inefficient screening of enzymatic activities has been addressed from several perspectives; however, the limitations related to biased expression in heterologous hosts cannot be overcome by using a single approach, but rather requires the synergetic implementation of multiple methodologies. Here, we review some of the principal constraints regarding the discovery of new enzymes in**

**metagenomic libraries and discuss how these might be resolved by using synthetic biology methods.**

## Introduction

Biotechnology takes advantage of processes that utilize biological components or living organisms, and plays an increasing role in industry, agriculture, medicine and energy generation. In the industrial context, many manufacturing processes that previously depended strictly on complex (and frequently harmful) chemical reactions have been superseded by much simpler and safer enzyme-based catalysis (Kirk *et al.*, 2002). The introduction of biotechnology in industrial processes generates not only a reduction in the final amount and toxicity of effluents, but can also considerably reduce costs (Herrera, 2004). The number of biotechnology applications has expanded in recent years, and this has created a growing demand for biocatalysts with superior performance or tolerance to extreme application-specific conditions (Lorenz *et al.*, 2002; Schloss and Handelsman, 2003). This is particularly true in those industries that produce bulk commodities such as detergents (Maurer, 2004). Similarly, fine-chemical industries require multiple biocatalysts in order to perform highly diverse transformations for the production of new compounds (Homann *et al.*, 2004).

The use of enzymes in industrial applications has grown considerably, and represents a market of approximately US\$4 billion in 2011 (GIA, 2012), and a number of different categories of enzymes have been used in a wide variety of applications (Schoemaker *et al.*, 2003). For example, proteases have been used in detergents, in pharmaceutical and chemical synthesis industries to degrade proteins into amino acids (Gupta *et al.*, 2002). Lipases hydrolyse fats to fatty acids and glycerol, and are useful for effluent treatment, detergent production and synthesis of fine chemicals among others (Hasan *et al.*, 2006). Glycosyl hydrolases (GHs) catalyse the hydrolysis of carbohydrates to sugars and have found applications in many processes in the textile, pulp and paper, and food production industries (Kirk *et al.*, 2002). The food industry also takes advantage of amylases, enzymes that hydrolyse starch into sugars (Kirk *et al.*, 2002).

Received 30 April, 2014; revised 22 June, 2014; accepted 28 June, 2014. \*For correspondence. E-mail rjward@ffclrp.usp.br; Tel. (+55) 16 3602 4384; Fax (+55) 16 3633 2660.

*Microbial Biotechnology* (2015) 8(1), 52–64  
doi:10.1111/1751-7915.12146

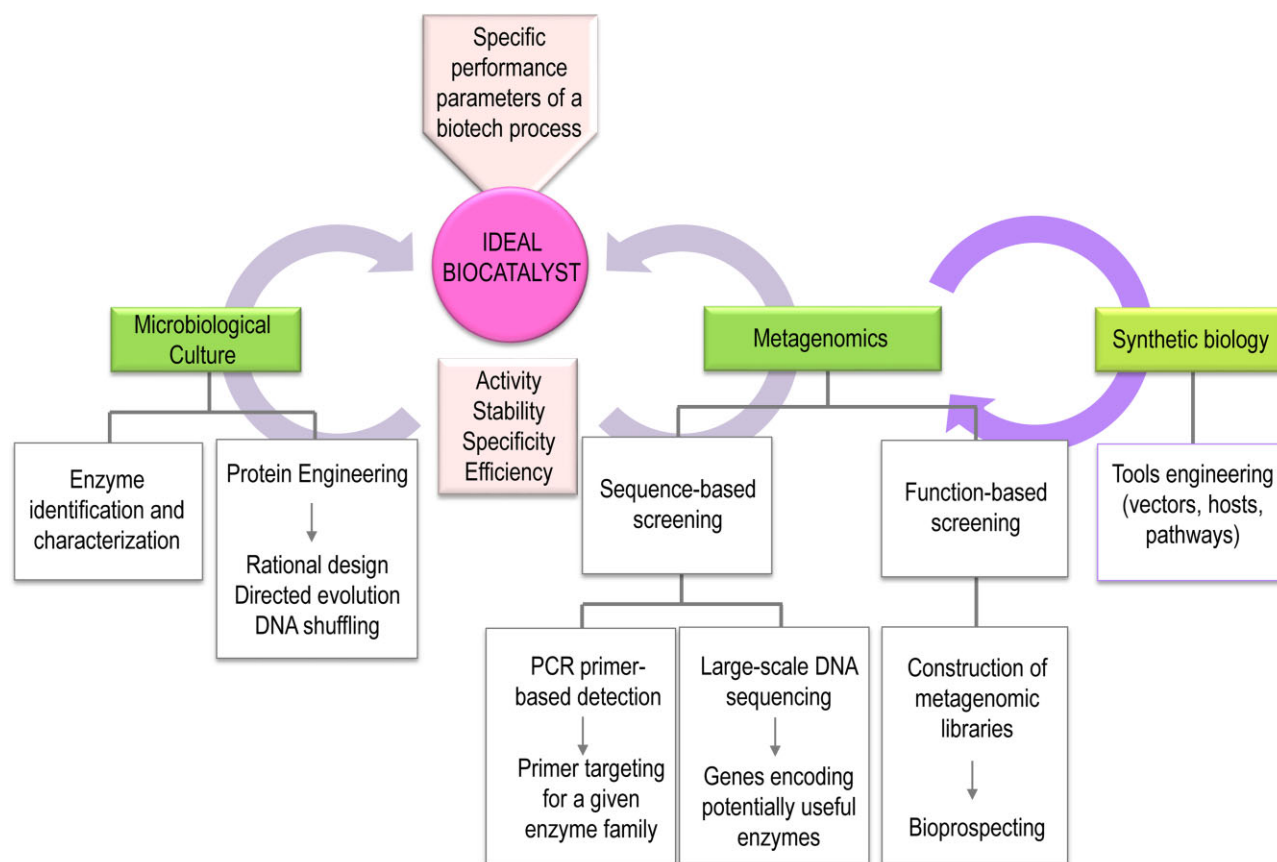
**Funding Information** The study was supported by CNPq (CNPq 472893/2013-0 and 307795/2009-8) and FAPESP (2010/18850-2). MEG was a beneficiary of a CNPq Young Talent Fellowship (CNPq 370630/2013-0). RSR was supported by a FAPESP Post-Doctoral Fellowship (FAPESP 2013/04125-2).

Furthermore, cellulases are not only useful for fuel production but are also applied in food processing, chemical synthesis and textile industries (Bhat, 2000).

The impact of the use of enzymes in industrial processes has stimulated an increased interest both in identifying new variants with enhanced kinetic parameters and in modifying previously characterized enzymes to increase their suitability for industrial applications (Lorenz *et al.*, 2002; Schloss and Handelsman, 2003). Parameters such as activity, efficacy, specificity and stability are used to characterize and select enzymes for different applications (Lorenz and Eck, 2005). Enzymes used in industry have been identified from different sources through a combination of two major strategies: (i) the identification of novel enzymes from cultured microorganisms and (ii) molecular evolution by DNA shuffling and rational design (Lynd *et al.*, 2002; Percival Zhang *et al.*, 2006; Krogh *et al.*, 2010; Ward, 2011; Chen *et al.*, 2012)

(see Fig. 1). However, enzymes suitable for a given biotechnology application need to work efficiently within specified parameters (Fig. 1), and since those currently used are frequently not optimized, the industrial processes have to be adjusted to accommodate these sub-optimal catalysts (Warnecke and Hess, 2009). As a consequence, there is an increasing demand for new biocatalysts with improved properties for industrial applications, such as higher catalytic efficiency on insoluble substrates (as in the case of cellulases used in the production of second generation bioethanol), increased stability at elevated temperature and at defined pH, and higher tolerance to end-product inhibition (Ward, 2011; Singhania *et al.*, 2013).

Microorganisms play a central role in biotechnology, not only as tools in molecular biology techniques, but also as the major source of biocatalysts for industrial applications (Fernandez-Arrojo *et al.*, 2010). Bearing in mind that the



**Fig. 1.** Overall diagram for the identification strategy of an ideal biocatalyst. The identification of enzymes from cultured microorganisms and metagenomics are the two principal approaches currently employed for recovering of genes encoding the desired enzymatic activity for industrial processes. The genes encoding enzymes identified from cultured microorganisms may be cloned and expressed, and parameters such as activity, stability, specificity and efficiency improved using protein rational design and *in vitro* evolution techniques. Metagenomics strategies are based on either activity-based approaches, which involve the construction of expression libraries and its posterior activity screening, or sequence-based approaches. Sequence-based approaches involve either the design of DNA primers for conserved regions of known protein families or data mining of genes encoding potential biocatalysts identified in sequences from next generation sequencing projects. Synthetic biology can provide solutions to the current limitations in activity-based metagenomic approaches. Development of methods for the engineering of new bacterial hosts and molecular biology tools promise to increase the efficiency of discovery of biotechnologically relevant enzymes.

total number of microbial cells on Earth is estimated to be  $10^{30}$  (Turnbaugh and Gordon, 2008), the huge natural wealth regarding protein diversity can be envisaged. Although prokaryotes represent the largest proportion of individual living organisms with an estimated  $10^3$ – $10^5$  microbial species in 1 g of soil (Schloss and Handelsman, 2006), less than 1% can be cultured using existing methodologies (Sleator *et al.*, 2008). If we assume that a single genome encodes 4000 proteins (as is the case for the typical bacteria *Escherichia coli*), then  $4 \times 10^8$  potential proteins might be expected in just 1 g of soil. Supposing that 40% of these proteins display catalytic activity (Dinsdale *et al.*, 2008), we might expect to find  $1.6 \times 10^8$  biocatalysts, which highlights the vast inventory of biological functions available in nature. Metagenomics avoids the necessity of isolation and laboratory cultivation of individual microorganisms, and has become a powerful tool for accessing and exploring the biological and molecular biodiversity present in different natural environments. Over the past decades, many studies using metagenomic approaches have proven to be successful for the recovery of novel enzymes with potential use in industrial applications (Lorenz and Eck, 2005; Fernandez-Arrojo *et al.*, 2010). Despite these successes, metagenomic strategies typically have low rates of target identification, and a number of issues need to be addressed in order to improve the screening efficiency of metagenomic libraries. The limits include: (i) bias imposed by host organism expression; (ii) vector performance in particular hosts; and (iii) suitable screening strategies in relation to the specific properties required in the target enzymes. As has already been demonstrated in several recent studies, all these issues may be addressed using a synthetic biology approach (Williamson *et al.*, 2005; Uchiyama and Watanabe, 2008; Uchiyama and Miyazaki, 2010). In addition to the application of existing synthetic biology approaches, the development of new methodologies is imperative for the next generation of metagenomic studies that aim to recover ideal biocatalysts for given industrial processes. This review focuses on how the interplay between synthetic biology and functional metagenomics can yield novel strategies to obtain ideal candidate enzymes with specific characteristics. We also discuss how innovative synthetic biology applications could help relieve current limitations on metagenomic screening.

### **Bioprospecting metagenomes to identify new biocatalysts**

#### *Recent advances and major limitations*

Although most environmental bacteria are refractory to cultivation, the biotechnological potential of the uncultivated bacteria can be realized by directly cloning the DNA

retrieved from the microbial community (Guazzaroni *et al.*, 2010a). The construction and subsequent screening of metagenomic libraries allows identification of the targeted genes encoding the desired catalytic activities (Fig. 1). Accordingly, a well-planned strategy should take into consideration the vector to be used, the host organism for transformation and the screening strategy in order to maximize the rate of identification of the target activities. For example, if single genes or small operons are of interest, the best option is to use a small-insert metagenomic library instead of a large-insert library (Guazzaroni *et al.*, 2010a). Small-insert expression libraries, especially those using lambda phage vectors and plasmids, are especially suitable for activity-based screening. The small size of the cloned fragments (up to 8 kb) means that most genes that are present in the appropriate orientation will be under the influence of strong vector promoters, and thus have a good chance of being expressed and detected in activity screens (Ferrer *et al.*, 2009). On the other hand, if the goal is biosynthetic pathway mining or functional expression of large multi-enzyme assemblies (for example, in the case of polyketide synthases clusters), the preferred option is library construction using cosmids or fosmid, which can harbour DNA inserts of up to 40 kb in size (Guazzaroni *et al.*, 2010a). Since fosmids use both the F-plasmid origin of replication and the partitioning mechanisms present in *E. coli* (Kim *et al.*, 1992), the use of these vectors are limited to this host organism. It is noteworthy that the bacterium *E. coli* is the most commonly used host strain for library screening, since many of the currently available genetic tools have been developed in this organism. One example is the use of modified phages that undergo the lytic cycle under controlled conditions, which allows the identification of proteins that would otherwise be toxic for the cell (Guazzaroni *et al.*, 2010a). Furthermore, a number of protocols for obtaining high rates of efficiency of transformation in *E. coli* are well established. In addition, kits that facilitate library construction and efficient transformation with different vectors are commercially available (Guazzaroni *et al.*, 2010a). However, it is important to appreciate that significant differences in expression modes exist between different taxonomic groups of prokaryotes, and that only 40% of enzymatic activities may be detected by random cloning in *E. coli* (Gabor *et al.*, 2004). Therefore, it is likely that also performing metagenomic library screening in hosts other than *E. coli* will expand the range of detectable activities, although achieving this goal will require further optimization of the conditions for high transformation efficiency. Indeed, promising results of metagenomic library screening have been reported in *Streptomyces* spp. (Wang *et al.*, 2000), *Rhizobium leguminosarum* (Wexler *et al.*, 2005) and diverse Proteobacteria (Craig *et al.*, 2010). For example, Wexler and collaborators constructed a library in

the broad host-range cosmid pLAFFR3 using metagenomic DNA obtained from the microbial community of an anaerobic digester in a wastewater treatment plant (Wexler *et al.*, 2005). After screening the metagenomic libraries in *R. leguminosarum*, a single cosmid that enabled *R. leguminosarum* to grow on ethanol as the sole carbon and energy source was recovered. Further analysis identified the presence of a gene encoding an atypical alcohol dehydrogenase that did not confer ethanol utilization ability to either *E. coli* or to *Pseudomonas aeruginosa*, even though the gene was transcribed in both these hosts (Wexler *et al.*, 2005). These results show that the use of broad host-range vectors enhances the flexibility of metagenomic library screening. Furthermore, a recent functional metagenomic study showed that recovery of genes conferring acid resistance to *E. coli*, and the subsequent transfer of some of these genes to *P. putida* and *B. subtilis*, expanded the capabilities of these two bacteria to survive harsh acid conditions (Guazzaroni *et al.*, 2013). However, in agreement with previous studies (Craig *et al.*, 2010), variable gene doses were present due to the use of different cloning plasmids in each host organism, and no quantitative comparison could be made regarding gene expression or activity levels between the hosts. Thus, the developing of robust broad-host-range vectors capable of replication in several different hosts is one of the major challenges in metagenomics, and one to which synthetic biology may make a significant contribution.

In addition to viable library construction, an additional bottleneck in metagenomic screening is related to the low frequency of positive clones that are typically recovered (Vieites *et al.*, 2009). Common screening methods are based on the degradation of specific enzyme substrates that result either in the appearance of halos surrounding the positive clones or alternatively on the use of chromogenic substrates (Guazzaroni *et al.*, 2010b). Depending on the type of substrate used and on the enzyme screened, detection of positive clones can be assayed directly in solid or in liquid media (Guazzaroni *et al.*, 2010b), and the choice of medium will have consequences on the throughput level of the screening method. Similarly, adequate selection of specific activity-driven substrates plays a key role in the success of recovering of the desired enzymatic activity and decreasing the frequency of false positives.

Although methods are available to the direct screening of many enzymes, there is an increasing need to expand the tools available for enzyme detection. An alternative approach to this problem using synthetic biology is the design and implementation of *in vivo* biosensors capable of generate a detectable output in response to the degradation or production of a particular metabolite (Galvao and de Lorenzo, 2006). In this context, substrate-induced gene expression and product-induced gene expression

approaches have been developed and successfully applied to the detection of enzymatic activities associated with the metabolic modification of compounds of interest (Uchiyama *et al.*, 2005; Uchiyama and Miyazaki, 2010). In a broader approach, the concept of genetic traps has been used to guide the construction of synthetic circuits containing engineered regulators to control gene expression responses to metabolites generated by enzymatic activities present in cloned metagenomic fragments (Uchiyama and Watanabe, 2008). In general, these strategies are particularly useful for screening of libraries where processing of small molecules is targeted (Williamson *et al.*, 2005).

#### *Activity-based versus sequence-based screening strategies*

There are two distinct strategies in functional metagenomics for the recovery of sequences encoding the desired enzymatic activity (Fig. 1). First, activity-based approaches involve construction of small- to large-insert expression libraries that are suitable for direct activity screening, such as lambda phage, plasmid, cosmid or copy-controlled fosmid vectors (Lorenz and Eck, 2005). Once a library has been constructed, a critical step is the screening of a large number of clones, and in the case of activity-based screening, thousands of clones may be analysed in a single screen, with the advantage that sequence information is not required. Therefore, this is the strategy that has the potential to identify entirely novel classes of genes encoding known or novel functions (Handelsman, 2004; Daniel, 2005; Gloux *et al.*, 2010; Bhat *et al.*, 2013). Furthermore, activity-driven screening strategies can potentially provide a means to reveal undiscovered genes or gene families that cannot be detected by sequence-driven approaches.

The most widely used activity-driven screening approach is phenotype detection of the desired activity using chemical dyes and insoluble or chromophore-bearing derivatives of enzyme substrates. When incorporated into the growth medium, these compounds allow the detection of specific catalytic capabilities of individual clones (Ferrer *et al.*, 2009). In an elegant example using this approach, the diversity of rumen enzymes was characterized by screening for hydrolase activity in a metagenomic phage library from the rumen content of a dairy cow (Ferrer *et al.*, 2005b). In total, 22 clones with distinct hydrolytic activities were identified and characterized, among which four hydrolases exhibited no sequence similarity to enzymes deposited in the public databases, and the putative catalytic residues in the sequences from four other clones showing esterase and cellulase-like activities did not match any known enzymes. Similarly, activity-based screening of metagenomic fosmid libraries

from cellulose-depleting microbial communities found in the fresh casts of two different earthworm species retrieved 10 carbohydrate-modifying enzymes (Beloqui *et al.*, 2010). Interestingly, two of the GHs identified represented two novel families of  $\beta$ -galactosidases/ $\alpha$ -arabinopyranosidases (Beloqui *et al.*, 2010). In another example, metagenomic DNA extracted from activated sludge from industrial wastewater was cloned into fosmids and the resulting *E. coli* library was screened for extradiol dioxygenase activity using catechol as the substrate (Suenaga *et al.*, 2007). Analysis of the sequences from a total of 38 clones identified various clusters encoding putative metabolic pathways that were dissimilar to previously reported pathways for degradation of the compound (Suenaga *et al.*, 2009a,b). Additionally, several other recent examples have shown that the phenotype detection approach can be successfully applied to the identification of novel targeted enzymes such as dehalogenases (Sharma *et al.*, 2013), *meta*-cleavage product hydrolases (Alcaide *et al.*, 2013), GHs (Ferrer *et al.*, 2012; Ko *et al.*, 2013), xylanases (Gong *et al.*, 2013b) carboxyl esterases and lipases (Martinez-Martinez *et al.*, 2013; 2014).

An alternative activity-driven screening approach is based on heterologous complementation of host strains that require the targeted genes for growth under selective conditions. In this strategy, functional complementation permits growth of a clone transformed with a metagenomic DNA insert containing the necessary genes for survival (Wenzel and Muller, 2005). This technique allows the straightforward and rapid screening of complex metagenomic libraries comprising millions of clones, and since false positives are rare, this approach is highly selective for the targeted genes (Simon *et al.*, 2009). However, this strategy is limited to the detection of enzymes that catalyse the synthesis of an essential product and for which an auxotrophic host is either available or can readily be constructed. Several examples in the literature have shown that this approach has been successful in the detection of different enzymes such as racemases (Chen *et al.*, 2010), DNA polymerases (Simon *et al.*, 2009),  $\beta$ -lactamases (Riesenfeld *et al.*, 2004), alcohol dehydrogenases (Wexler *et al.*, 2005) and enzymes involved in poly-3-hydroxybutyrate metabolism (Wang *et al.*, 2006).

Sequence-based approaches have also led to the effective identification of genes encoding enzymes, such as dimethylsulfoniopropionate-degrading enzymes (Varaljay *et al.*, 2010), glycerol dehydratases (Knietsch *et al.*, 2003), dioxygenases (Morimoto and Fujii, 2009; Sul *et al.*, 2009; Zaprasis *et al.*, 2010), nitrite reductases (Bartossek *et al.*, 2010), hydrazine oxidoreductases (Li *et al.*, 2010), chitinases (Hjort *et al.*, 2010), glycoside hydrolases (Lee *et al.*, 2013), nitrilases (Gong *et al.*,

2013a), prephenate dehydrogenases (Jiang *et al.*, 2013) and hemicellulases (Yan *et al.*, 2013). The application of sequence-based approaches involves the design of polymerase chain reaction (PCR) primers for the target sequences that are derived from conserved regions of known protein families, and this dependence on prior knowledge limits the possibility for identifying new protein families (Ferrer *et al.*, 2009). The large-scale sequencing of bulk DNA or metagenomic libraries through deep sequencing techniques provides the raw data for mining sequences encoding potentially useful enzymes. Since homology-based methods are effective only when the information regarding the reference sequences is accurate, a further disadvantage of this approach is its reliance both on existing genome annotations and on the quality and completeness of current databases (Hallin *et al.*, 2008). It is worth a cautionary note, since a significant number of genomes in the current databases contain misannotations (Schnoes *et al.*, 2009). Considering that the classification of protein families is based on amino acid similarity, novel enzyme families could not be detected by database searching with sequences from metagenomic sequencing or PCR-based detection methods, and might be annotated as hypothetical proteins. A review of prokaryotic protein diversity in different shotgun metagenome studies indicated that 30–60% of the proteins could not be assigned known functions using current public databases (Vieites *et al.*, 2009). The advantage of using activity-based rather than sequence-based screening was highlighted in a recent report in which a novel cold-tolerant esterase with low sequence similarity (< 29% amino acid identity) was identified by functional screening of an Antarctic soil sample (Heath *et al.*, 2009). This esterase had no identity to any GenBank nucleotide sequence, and it is unlikely that it would have been detected by low stringency PCR-based screening methods or by deep sequencing techniques.

#### *Mining biocatalysts from extreme environments*

The maximum yield of industrial processes is achieved by optimization of physicochemical parameters, and most currently available enzymes are incompatible with these conditions, and the use of enzymes requires that these processes be adapted, which can result in reduced production levels. Extreme environments, such as solfataric hot springs (Rhee *et al.*, 2005), Urania hypersaline basins (Ferrer *et al.*, 2005a), acid mine drainage biofilms (Guazzaroni *et al.*, 2013), glacier soil (Yuhong *et al.*, 2009), glacial ice (Simon *et al.*, 2009) and Antarctic soil (Cieslinski *et al.*, 2009; Heath *et al.*, 2009) represent a rich and largely unexploited reservoirs of novel biocatalysts with biotechnologically valuable properties. Although the diversity of microbial communities present in

many extreme habitats is likely to be low, samples from these environments are still a valuable source of novel enzymes that are active under extreme conditions (Steele *et al.*, 2009), and as might be expected, the properties of the enzymes retrieved are consistent with conditions of the source environments (Feng *et al.*, 2007; Heath *et al.*, 2009; Jiang *et al.*, 2009; Hu *et al.*, 2012).

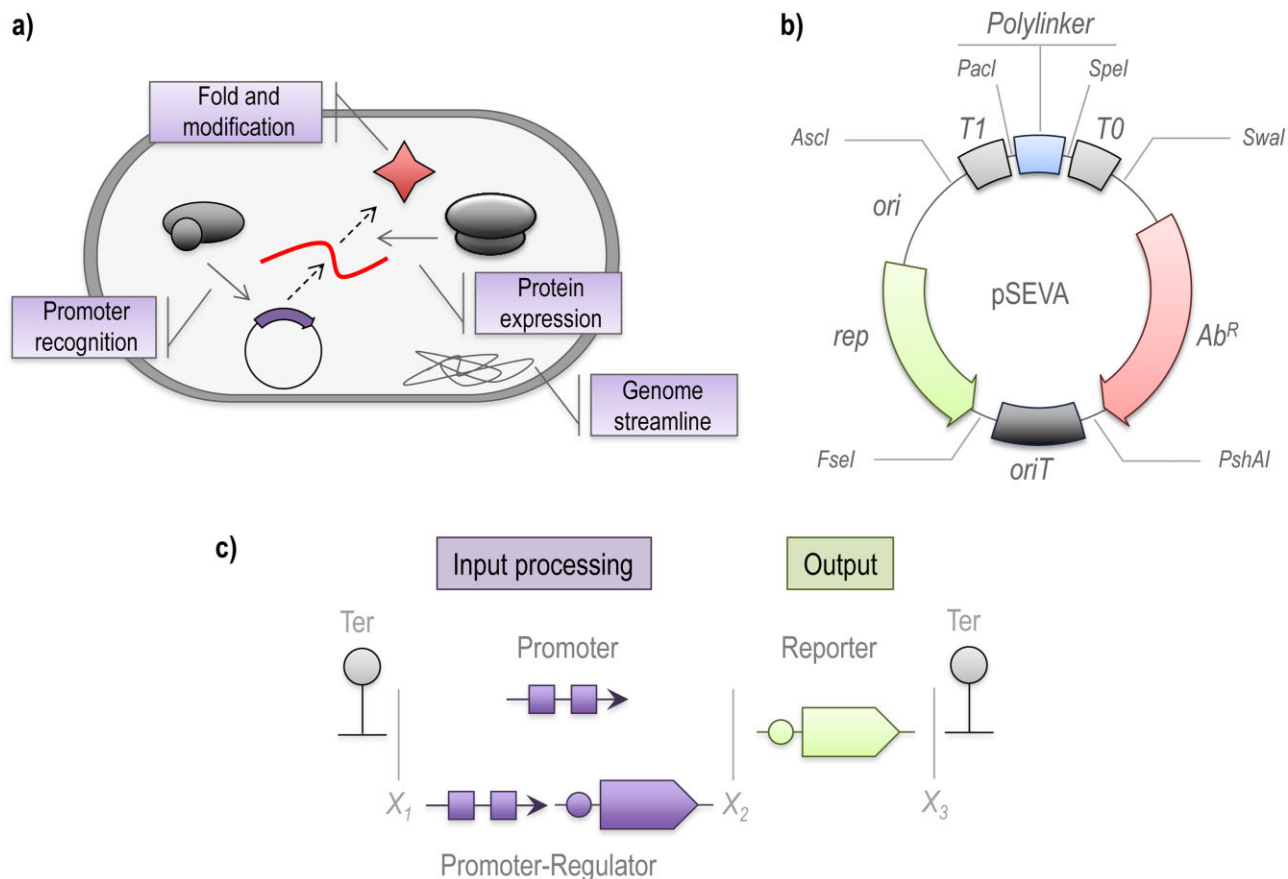
Metagenomic libraries derived from extreme habitats have been constructed and enzymes displaying environment compatible properties have been recovered. A relevant consideration is the compatibility of the host organism growth with the optimum physicochemical characteristics used in the screening of the catalytic activity. *Escherichia coli* is highly sensitive to the conditions that are typically present in extreme environments, and experimental conditions should therefore be adjusted to perform library screening under the conditions of interest (for example, low pH) while still allowing growth of this organism. Alternatively, screening could be performed without subjecting the host to any selective pressure, for example in cases where the host organism is used to clone total metagenomic DNA and the screening is performed under regular growth conditions. In both cases, after the initial identification of the enzymes, characterization of the desired catalytic properties of the source clones can be made with crude culture extracts and selected enzymes may then be purified for a more detailed characterization. For instance, a novel alkaliphilic esterase active at 7°C was identified by screening a metagenomic library from Antarctic desert soil (Heath *et al.*, 2009). In a separate study, another cold-active esterase was also isolated from the same library using a similar approach (Hu *et al.*, 2012). Their low-temperature activity and alkaliphilic properties make these enzymes interesting candidates for industrial applications. Esterases are of particular industrial interest since they serve as useful biocatalysts in the detergent and food industries, and for the production of fine chemicals and in bioremediation processes (Aurilia *et al.*, 2008). Cold-adapted esterolytic enzymes could be of further value with regard to savings in energy, as compared with their mesophilic counterparts, due to their high catalytic activities at low temperatures (Hu *et al.*, 2012). At the opposite extreme, Rhee and colleagues (2005) constructed fosmid libraries using environmental samples from solfataric hot springs, and function-based screening identified a novel esterase that exhibited a high temperature optimum activity and high thermal stability.

Cellulases are enzymes with a wide range of applications such as in the textile industry for cotton softening and denim finishing; in the food industry for mashing; and in the cellulose pulp and paper industries for de-inking, fiber refining and fiber modification (Bhat, 2000). Recent interest has focussed on the use of cellulases in the hydrolysis of pretreated lignocellulosic materials into

sugars, which can then be fermented to produce second-generation ethanol and other bio-products. Metagenomic approaches have been widely employed to isolate cellulases from environmental samples in which lignocellulosic materials were under intensive degradation, including soil (Kim *et al.*, 2008; Jiang *et al.*, 2009), hindgut contents of higher termites (Warnecke *et al.*, 2007), compost (Pang *et al.*, 2009), contents of rabbit cecum (Feng *et al.*, 2007), fresh cast from earthworms (Beloqui *et al.*, 2010), Brazilian mangroves (Thompson *et al.*, 2013), calf rumen (Ferrer *et al.*, 2012) and cow rumen (Ferrer *et al.*, 2005b; Wang *et al.*, 2013). As anticipated, most cellulases retrieved from uncultured microbes from the rumen of herbivores were acidic and mesophilic, properties which are compatible with the conditions of yeast fermentation for the production of second-generation ethanol. Although a number of cellulase active clones have been isolated from metagenomic libraries by functional screening, the frequencies of true positive identifications were low due to low and biased expression of the majority of cellulase genes in *E. coli*. Moreover, most of the enzymes identified were endoglucanases (despite the use of specific substrates for screening other types of cellulases), which was probably a consequence of the expression of the correctly folded enzymes in *E. coli*. No cellobiohydrolases active against crystalline cellulose substrates have as yet been isolated from metagenomic libraries (Duan and Feng, 2010), which may be due to the fungal source of the majority of cellobiohydrolases and, to date, there are no reports using activity-based screening of cellulases from a cDNA metagenomic library in fungi. Moreover, since extreme environments are colonized mainly by microorganisms highly adapted to such harsh conditions, the majority of species are likely to be phylogenetically distant from *E. coli*. The recurring limitation of the metagenomic approach for certain enzyme families is therefore the lack of genetic tools for library screening in phylogenetically diverse hosts, and the development of such tools would significantly increase the probability of successful enzyme retrieval from a wide range of environmental samples.

#### *Using synthetic biology to improve metagenomic screening strategies*

The previous sections have highlighted a number of bottlenecks facing metagenomic screening that need to be resolved in order to improve the discovery rate of target enzymes. These limitations can be grouped in three main categories (as shown in Fig. 2). First, there is a need for improvement in host organism capabilities with the aim of improving the expression of the target enzymes. Second, the development of new genetic tools is necessary in order to improve the construction of metagenomic



**Fig. 2.** Synthetic Biology may overcome limiting steps in activity-based metagenomic library screening. Current bottlenecks in functional metagenomics are related to (A) limitations in the host capabilities, (B) the performance of the genetic tools and (C) the availability of efficient screening methods.

A. In the case of the host, critical steps related to the recognition of transcriptional and translational signals, as well as the folding and modification of the expressed enzyme need to be enhanced. Host performance might be improved by reducing the metabolic burden related to the expression of unnecessary genes.

B. The use of semi-synthetic, high-efficiency genetic tools is essential for the construction of metagenomic libraries that can be maintained and screened in a wide number of microorganisms. The example shows the pSEVA bacterial vector, which is endowed with several functional features such as terminators, origin for transfer and an extensive polylinker optimized for use in several bacterial hosts.

C. Genetic circuits constructed by combining input modules (e.g. promoters and regulators) and output devices (such as reporter proteins) assembled with a standard format that uses the same sets of restriction enzymes (represented by  $X_1$ ,  $X_2$ , etc.). Such circuits facilitate the screening of enzymatic activities in metagenomic libraries. The standardization of the assembly process facilitates the combination of several independent modules to construct sophisticated activity-triggered biosensors.

libraries suitable for screening in different hosts. Finally, continuation of ongoing research to elaborate novel screening strategies that enhance the discovery rate of the enzymes of interest is needed. The advances in synthetic biology over the past decade could provide the framework to address these constraints, and a particularly promising approach is the analysis of biological systems in an analogous way as electronic devices, whereby cells can be reprogrammed to perform new tasks with high efficiency (Purnick and Weiss, 2009; Weber and Fussenegger, 2010). Synthetic biology relies on a conceptual framework more closely related to engineering than biology, such as design, modelling, implementation and debugging (Canton *et al.*, 2008; Purnick and Weiss, 2009; Weber and Fussenegger, 2010). The design aspect

focuses on the planning and construction of new gene circuits for the desired application (Canton *et al.*, 2008). Modelling involves computational simulation of the proposed gene circuits in order to both evaluate performance and capabilities and to guide the selection of the suitable molecular components necessary for its construction (Koide *et al.*, 2009). The implementation step encompasses the physical assembly of the DNA elements encoding the appropriate components (such as promoters, regulators, terminators, enzymes, transporters, etc.), and follows a specific assembly standard (Arkin, 2008). Finally, the debugging step requires the testing and validation of the circuit *in vivo*, and includes the correction of undesirable traits that have their origin in the emergent properties of biology systems (Gardner *et al.*, 2000; Moon

*et al.*, 2011; Siuti *et al.*, 2013). Several examples of new biological circuits that have been successfully designed and implemented are currently available (Gardner *et al.*, 2000; Cox *et al.*, 2007; 2010; Moon *et al.*, 2011; Silva-Rocha and de Lorenzo, 2011), and in recent years, the field has developed at a remarkable speed (Weber and Fussenegger, 2010; Zhan *et al.*, 2010; Regot *et al.*, 2011; Siuti *et al.*, 2013).

As mentioned in previous sections, failure of heterologous gene expression in host cells is among the main causes of the low recovery rates of enzymes of interest from metagenomic libraries (Handelsman, 2004; Lorenz and Eck, 2005). A combination of different strategies could be applied to optimize this critical event, (Fig. 2A). In the initial transcription step, a reduced affinity of the RNA polymerase (RNAP) for intrinsic promoters derived from metagenomic fragments represents an important limitation for heterologous protein expression in the host organism. Promoter recognition in prokaryotes is strongly biased among the different phylogenetic groups (Gabor *et al.*, 2004), and the ideal host for heterologous protein expression should be endowed with transcriptional machinery with broad promoter recognition capability. Such a goal could be attained, for example, by coexpressing heterologous sigma factors with different promoter specificities (Osterberg *et al.*, 2011; Rhodius *et al.*, 2013), thereby permitting protein expression from promoters derived from different bacterial sources. In a more direct approach, the expression of foreign genes in the host organism could be driven by a high-efficiency expression system such as the T7 RNAP (Terron-Gonzalez *et al.*, 2013), and has the advantage that genes controlled by complex signal transduction mechanisms could be easily expressed in response to a single inducer such as isopropyl-beta-D-thiogalactopyranoside (Tabor, 2001). The same line of reasoning may also be applied to mRNA translation, where poor recognition of the ribosome-binding site (RBS) can reduce protein expression levels (Zelcbuch *et al.*, 2013). In this case, the coexpression of additional proteins related to the mRNA recognition step could expand the host capability for foreign gene expression (Uchiyama and Miyazaki, 2009). While these strategies aim to improve the level of expression of target proteins, an additional factor is related to the activity levels of expressed enzymes. This is particularly true in the case of enzymes displaying either complex folding or requiring additional processing steps (such as cleavage, secretion, or peptide modification) (DeSantis and Jones, 1999; Bhat, 2000). For instance, in the case of protein folding, the coexpression of molecular chaperons (Dobson, 2003) has been reported to enhance the expression of heterologous proteins in *E. coli* (Ferrer *et al.*, 2004).

Among the additional approaches to improve host organism capabilities, genome edition (or streamlining)

(Cambray *et al.*, 2011) deserves special attention. Since many of the genes in bacteria such as *E. coli* are not essential for growth under standard laboratory conditions (Medini *et al.*, 2005), maintaining these genes and the consequent redundant expression of unnecessary proteins represents a significant energetic cost and a metabolic burden to the host cell (Posfai *et al.*, 2006). The reduction of genome size in bacteria by the removal of non-essential genes has been shown to endow bacteria with renewed metabolic vigor that enhances the production level of heterologous proteins (Posfai *et al.*, 2006; Martinez-Garcia *et al.*, 2014). Due to their superior expression capabilities, bacterial strains with minimized genomes are therefore promising host organisms for screening of metagenomic libraries.

In addition to genetic engineering of the more common hosts such as *E. coli*, several attempts have been focused on the use of alternative hosts to increase the rate of enzyme identification in metagenomic screening (Craig *et al.*, 2010; Guazzaroni *et al.*, 2013). Screening in alternative hosts requires either library construction in non-optimal, broad-host-range vectors (Craig *et al.*, 2010) or subcloning of target genes in appropriate vectors (Guazzaroni *et al.*, 2013). Universal and standardized tools are required in order to facilitate this type of multi-organism approach for the screening of metagenomic libraries in a larger number of hosts, and synthetic biology approaches are particularly suited to the development and use of such tools for the construction of adequate gene circuits in an increasing number of host cell platforms (Andrianantoandro *et al.*, 2006; Arkin, 2008; Shetty *et al.*, 2008). Among these new multihost tools, the pSEVA vectors are of particular interest as synthetic broad-host-range vectors that are expected to work in about 100 different bacterial species (Fig. 2B; Silva-Rocha *et al.*, 2013). Using this system as a starting point, new genetic tools could be developed for cloning and screening of environmental DNA from phylogenetically diverse bacteria.

Currently available methods that are used to screen for new catalytic activities in metagenomic libraries frequently rely either on the use of chromogenic enzyme substrates (Hu *et al.*, 2012; Ko *et al.*, 2013) or substrates that when degraded leave a clear halo around positive colonies (Wong *et al.*, 2009). More labor-intensive screening procedures such as colony PCR are also used for enzyme discovery (Hrvatín and Piel, 2007). The rapid expansion of synthetic biology has already resulted in the construction of regulatory circuits using well-characterized parts (Voigt, 2006), and there is a tremendous potential to use this accumulated knowledge for further advances in the design of biosensors to screen for enzymatic activities (Williamson *et al.*, 2005; Nasuno *et al.*, 2012). The use of biosensors is an *in vivo* strategy that has been used to



identify specific enzymatic activities using engineered regulatory circuits coupled to a reporter gene, such as *lacZ*, green fluorescent protein or luciferase (Mohn *et al.*, 2006; de Las Heras *et al.*, 2010). This genetic trap approach (Uchiyama *et al.*, 2005; Uchiyama and Watanabe, 2008; Uchiyama and Miyazaki, 2010) eliminates the necessity for extensive manipulation during screening, yet allows the identification of positive clones in metagenomic libraries. Synthetic biology approaches have also been used to modulate enzyme levels in biosynthetic pathways by combinatorially pairing genes with a defined set of RBS (Zelcbuch *et al.*, 2013), resulting in the modulation of protein abundance by several orders of magnitude, which showed that engineering of metabolic pathways relies on precise control of enzyme levels (Zelcbuch *et al.*, 2013). These examples demonstrate how synthetic biology approaches can improve the ability to interconnect regulatory components (e.g. promoters and regulators) to generate new circuits with reliable performance characteristics (Fig. 2C; Schmidt and de Lorenzo, 2012). The combination of the available assembly platforms for circuit engineering (Arkin, 2008; Zhan *et al.*, 2010; Nikel and de Lorenzo, 2013) together with approaches for the redesign of regulatory systems to recognize new molecules (de Las Heras and de Lorenzo, 2012) is likely to lead to new concepts and allow the implementation of new genetic traps for the identification of enzymatic activities of interest (Fernandez-Arrojo *et al.*, 2010), and will have a significant impact in the detection rate of new biotechnologically relevant biocatalysts (Ferrer *et al.*, 2009).

### Concluding remarks

Classic approaches for isolating industrial enzymes involve their identification from cultured microorganisms, and more recently, this strategy has been expanded to include *in vitro* evolution and rational design techniques for the improvement of their catalytic properties. However, the diversity of proteins identified using this strategy is restricted, the methods are time-consuming, costly and in the case of *in vitro* evolution, it is impossible to test all variants (Sommer *et al.*, 2010). The metagenomics approach encompasses the idea that the desired biocatalyst may already exist in nature, and is an alternative strategy that is used to explore the inherent diversity of natural environments. As a result, development of a broad repertoire of culture-independent techniques have been developed and advances towards the identification of novel and potent biocatalysts from metagenomic libraries have been made. However, the available metagenomic approaches require further refinement to achieve the goal of identifying industrially relevant biocatalysts. Existing limitations with respect to host

expression, vector availability and specific screening restrictions cannot be solved by using a single approach, and requires the synergic implementation of multiple methodologies. A growing number of studies have shown that synthetic biology may significantly improve metagenomic library screening and allow exploitation of the rich biochemical potential present in natural environments.

### Conflict of Interest

None declared.

### References

- Alcaide, M., Tornes, J., Stogios, P.J., Xu, X., Gertler, C., Di Leo, R., *et al.* (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the alpha/beta hydrolase family. *Biochem J* **454**: 157–166.
- Andrianantoandro, E., Basu, S., Karig, D.K., and Weiss, R. (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* **2**: 2006. 0028.
- Arkin, A. (2008) Setting the standard in synthetic biology. *Nat Biotechnol* **26**: 771–774.
- Aurilia, V., Parracino, A., and D'Auria, S. (2008) Microbial carbohydrate esterases in cold adapted environments. *Gene* **410**: 234–240.
- Bartossek, R., Nicol, G.W., Lanzen, A., Klenk, H.P., and Schleper, C. (2010) Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* **12**: 1075–1088.
- Beloqui, A., Nechitaylo, T.Y., Lopez-Cortes, N., Ghazi, A., Guazzaroni, M.E., Polaina, J., *et al.* (2010) Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. *Appl Environ Microbiol* **76**: 5934–5946.
- Bhat, A., Riyaz-Ul-Hassan, S., Ahmad, N., Srivastava, N., and Johri, S. (2013) Isolation of cold-active, acidic endocellulase from Ladakh soil by functional metagenomics. *Extremophiles* **17**: 229–239.
- Bhat, M.K. (2000) Cellulases and related enzymes in biotechnology. *Biotechnol Adv* **18**: 355–383.
- Cambray, G., Mutalik, V.K., and Arkin, A.P. (2011) Toward rational design of bacterial genomes. *Curr Opin Microbiol* **14**: 624–630.
- Canton, B., Labno, A., and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* **26**: 787–793.
- Chen, H.L., Chen, Y.C., Lu, M.Y., Chang, J.J., Wang, H.T., Ke, H.M., *et al.* (2012) A highly efficient beta-glucosidase from the buffalo rumen fungus *Neocallimastix patriciarum* W5. *Biotechnol Biofuels* **5**: 24.
- Chen, I.C., Thiruvengadam, V., Lin, W.D., Chang, H.H., and Hsu, W.H. (2010) Lysine racemase: a novel non-antibiotic selectable marker for plant transformation. *Plant Mol Biol* **72**: 153–169.
- Cieslinski, H., Bialkowska, A., Tkaczuk, K., Dlugolecka, A., Kur, J., and Turkiewicz, M. (2009) Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library. *Pol J Microbiol* **58**: 199–204.

- Cox, R.S., 3rd, Surette, M.G., and Elowitz, M.B. (2007) Programming gene expression with combinatorial promoters. *Mol Syst Biol* **3**: 145.
- Cox, R.S., 3rd, Dunlop, M.J., and Elowitz, M.B. (2010) A synthetic three-color scaffold for monitoring genetic regulation and noise. *J Biol Eng* **4**: 10.
- Craig, J.W., Chang, F.Y., Kim, J.H., Obiajulu, S.C., and Brady, S.F. (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl Environ Microbiol* **76**: 1633–1641.
- Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**: 470–478.
- DeSantis, G., and Jones, J.B. (1999) Chemical modification of enzymes for enhanced functionality. *Curr Opin Biotechnol* **10**: 324–330.
- Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., *et al.* (2008) Functional metagenomic profiling of nine biomes. *Nature* **452**: 629–632.
- Dobson, C.M. (2003) Protein folding and misfolding. *Nature* **426**: 884–890.
- Duan, C.J., and Feng, J.X. (2010) Mining metagenomes for novel cellulase genes. *Biotechnol Lett* **32**: 1765–1775.
- Feng, Y., Duan, C.J., Pang, H., Mo, X.C., Wu, C.F., Yu, Y., *et al.* (2007) Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol* **75**: 319–328.
- Fernandez-Arrojo, L., Guazzaroni, M.E., Lopez-Cortes, N., Beloqui, A., and Ferrer, M. (2010) Metagenomic era for biocatalyst identification. *Curr Opin Biotechnol* **21**: 725–733.
- Ferrer, M., Chernikova, T.N., Timmis, K.N., and Golyshin, P.N. (2004) Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain. *Appl Environ Microbiol* **70**: 4499–4504.
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins Dos Santos, V.A., Yakimov, M.M., *et al.* (2005a) Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* **12**: 895–904.
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A., *et al.* (2005b) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* **7**: 1996–2010.
- Ferrer, M., Beloqui, A., Timmis, K.N., and Golyshin, P.N. (2009) Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol* **16**: 109–123.
- Ferrer, M., Ghazi, A., Beloqui, A., Vieites, J.M., Lopez-Cortes, N., Marin-Navarro, J., *et al.* (2012) Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS ONE* **7**: e38134.
- Gabor, E.M., Alkema, W.B., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**: 879–886.
- Galvao, T.C., and de Lorenzo, V. (2006) Transcriptional regulators a la carte: engineering new effector specificities in bacterial regulatory proteins. *Curr Opin Biotechnol* **17**: 34–42.
- Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**: 339–342.
- GIA (2012) *Global Industry Analysts. Industrial Enzymes – A Global Multi-Client Market Research Project*. California, USA: GIA.
- Gloux, K., Berteau, O., El Oumami, H., Beguet, F., Leclerc, M., and Dore, J. (2010) A metagenomic beta-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc Natl Acad Sci USA* **108** (Suppl. 1): 4539–4546.
- Gong, J.S., Lu, Z.M., Li, H., Zhou, Z.M., Shi, J.S., and Xu, Z.H. (2013a) Metagenomic technology and genome mining: emerging areas for exploring novel nitrilases. *Appl Microbiol Biotechnol* **97**: 6603–6611.
- Gong, X., Gruninger, R.J., Forster, R.J., Teather, R.M., and McAllister, T.A. (2013b) Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Appl Microbiol Biotechnol* **97**: 2423–2431.
- Guazzaroni, M.-E., Golyshin, P.N., and Ferrer, M. (2010a) Analysis of complex microbial communities through metagenomic survey. In *Metagenomics: Theory, Methods and Applications*. Marco, D. (ed.). Norfolk, UK: Caister Academic, pp. 55–77.
- Guazzaroni, M.E., Beloqui, A., Vieites, J.M., Al-ramahi, Y., Cortés, N.L., Ghazi, A., *et al.* (2010b) Metagenomic mining of enzyme diversity. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, K. (ed.). Heidelberg, Germany: Springer, pp. 2911–2927.
- Guazzaroni, M.E., Morgante, V., Mirete, S., and Gonzalez-Pastor, J.E. (2013) Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environ Microbiol* **15**: 1088–1102.
- Gupta, R., Beg, Q.K., and Lorenz, P. (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol* **59**: 15–32.
- Hallin, P.F., Binnewies, T.T., and Ussery, D.W. (2008) The genome BLASTatlas—a GeneWiz extension for visualization of whole-genome homology. *Mol Biosyst* **4**: 363–371.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669–685.
- Hasan, F., Shah, A.A., and Hameed, A. (2006) Industrial applications of microbial lipases. *Enzyme Microb Technol* **39**: 235–251.
- Heath, C., Hu, X.P., Cary, S.C., and Cowan, D. (2009) Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. *Appl Environ Microbiol* **75**: 4657–4659.
- Herrera, S. (2004) Industrial biotechnology—a chance at redemption. *Nat Biotechnol* **22**: 671–675.
- Hjort, K., Bergstrom, M., Adesina, M.F., Jansson, J.K., Smalla, K., and Sjoling, S. (2010) Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiol Ecol* **71**: 197–207.
- Homann, M.J., Vail, R.B., Previte, E., Tamarez, M., Morgan, B., Dodds, D.R., and Zaks, A. (2004) Rapid identification of enantioselective ketone reductions using targeted microbial libraries. *Tetrahedron* **60**: 789–797.

- Hrvatín, S., and Piel, J. (2007) Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *J Microbiol Methods* **68**: 434–436.
- Hu, X.P., Heath, C., Taylor, M.P., Tuffin, M., and Cowan, D. (2012) A novel, extremely alkaliphilic and cold-active esterase from Antarctic desert soil. *Extremophiles* **16**: 79–86.
- Jiang, C., Ma, G., Li, S., Hu, T., Che, Z., Shen, P., *et al.* (2009) Characterization of a novel beta-glucosidase-like activity from a soil metagenome. *J Microbiol* **47**: 542–548.
- Jiang, C., Yin, B., Tang, M., Zhao, G., He, J., Shen, P., and Wu, B. (2013) Identification of a metagenome-derived prephenate dehydrogenase gene from an alkaline-polluted soil microorganism. *Antonie Van Leeuwenhoek* **103**: 1209–1219.
- Kim, S.J., Lee, C.M., Han, B.R., Kim, M.Y., Yeo, Y.S., Yoon, S.H., *et al.* (2008) Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiol Lett* **282**: 44–51.
- Kim, U.J., Shizuya, H., de Jong, P.J., Birren, B., and Simon, M.I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res* **20**: 1083–1085.
- Kirk, O., Borchert, T.V., and Fuglsang, C.C. (2002) Industrial enzyme applications. *Curr Opin Biotechnol* **13**: 345–351.
- Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and Daniel, R. (2003) Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. *J Mol Microbiol Biotechnol* **5**: 46–56.
- Ko, K.C., Han, Y., Cheong, D.E., Choi, J.H., and Song, J.J. (2013) Strategy for screening metagenomic resources for exocellulase activity using a robotic, high-throughput screening system. *J Microbiol Methods* **94**: 311–316.
- Koide, T., Pang, W.L., and Baliga, N.S. (2009) The role of predictive modelling in rationally re-engineering biological systems. *Nat Rev Microbiol* **7**: 297–305.
- Krogh, K.B., Harris, P.V., Olsen, C.L., Johansen, K.S., Hojer-Pedersen, J., Borjesson, J., and Olsson, L. (2010) Characterization and kinetic analysis of a thermostable GH3 beta-glucosidase from *Penicillium brasilianum*. *Appl Microbiol Biotechnol* **86**: 143–154.
- de Las Heras, A., and de Lorenzo, V. (2012) Engineering whole-cell biosensors with no antibiotic markers for monitoring aromatic compounds in the environment. *Methods Mol Biol* **834**: 261–281.
- de Las Heras, A., Carreno, C.A., Martinez-Garcia, E., and de Lorenzo, V. (2010) Engineering input/output nodes in prokaryotic regulatory circuits. *FEMS Microbiol Rev* **34**: 842–865.
- Lee, S., Cantarel, B., Henrissat, B., Gevers, D., Birren, B.W., Huttenhower, C., and Ko, G. (2013) Gene-targeted metagenomic analysis of glucan-branching enzyme gene profiles among human and animal fecal microbiota. *ISME J* **8**: 493–503.
- Li, M., Hong, Y., Klotz, M.G., and Gu, J.D. (2010) A comparison of primer sets for detecting 16S rRNA and hydrazine oxidoreductase genes of anaerobic ammonium-oxidizing bacteria in marine sediments. *Appl Microbiol Biotechnol* **86**: 781–790.
- Lorenz, P., and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* **3**: 510–516.
- Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* **13**: 572–577.
- Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**: 506–577. table of contents.
- Martinez-Garcia, E., Nikel, P.I., Chavarria, M., and de Lorenzo, V. (2014) The metabolic cost of flagellar motion in *Pseudomonas putida* KT2440. *Environ Microbiol* **16**: 291–303.
- Martinez-Martinez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., *et al.* (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl Environ Microbiol* **79**: 3553–3562.
- Martinez-Martinez, M., Lores, I., Pena-Garcia, C., Bargiela, R., Reyes-Duarte, D., Guazzaroni, M.E., *et al.* (2014) Biochemical studies on a versatile esterase that is most catalytically active with polyaromatic esters. *Microb Biotechnol* **7**: 184–191.
- Maurer, K.H. (2004) Detergent proteases. *Curr Opin Biotechnol* **15**: 330–334.
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005) The microbial pan-genome. *Curr Opin Genet Dev* **15**: 589–594.
- Mohn, W.W., Garmendia, J., Galvao, T.C., and de Lorenzo, V. (2006) Surveying biotransformations with a la carte genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into lacZ-based phenotypes. *Environ Microbiol* **8**: 546–555.
- Moon, T.S., Clarke, E.J., Groban, E.S., Tamsir, A., Clark, R.M., Eames, M., *et al.* (2011) Construction of a genetic multiplexer to toggle between chemosensory pathways in *Escherichia coli*. *J Mol Biol* **406**: 215–227.
- Morimoto, S., and Fujii, T. (2009) A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Appl Microbiol Biotechnol* **83**: 389–396.
- Nasuno, E., Kimura, N., Fujita, M.J., Nakatsu, C.H., Kamagata, Y., and Hanada, S. (2012) Phylogenetically novel LuxI/LuxR-type quorum sensing systems isolated using a metagenomic approach. *Appl Environ Microbiol* **78**: 8067–8074.
- Nikel, P.I., and de Lorenzo, V. (2013) Implantation of unmarked regulatory and metabolic modules in Gram-negative bacteria with specialised mini-transposon delivery vectors. *J Biotechnol* **163**: 143–154.
- Osterberg, S., del Peso-Santos, T., and Shingler, V. (2011) Regulation of alternative sigma factor use. *Annu Rev Microbiol* **65**: 37–55.
- Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L., and Feng, J.X. (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* **58**: 404–408.

- Percival Zhang, Y.H., Himmel, M.E., and Mielenz, J.R. (2006) Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv* **24**: 452–481.
- Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., *et al.* (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**: 1044–1046.
- Purnick, P.E., and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **10**: 410–422.
- Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellen, J., Peeters, T., *et al.* (2011) Distributed biological computation with multicellular engineered networks. *Nature* **469**: 207–211.
- Rhee, J.K., Ahn, D.G., Kim, Y.G., and Oh, J.W. (2005) New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl Environ Microbiol* **71**: 817–825.
- Rhodijs, V.A., Segall-Shapiro, T.H., Sharon, B.D., Ghodasara, A., Orlova, E., Tabakh, H., *et al.* (2013) Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. *Mol Syst Biol* **9**: 702.
- Riesenfeld, C.S., Goodman, R.M., and Handelsman, J. (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* **6**: 981–989.
- Schloss, P.D., and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14**: 303–310.
- Schloss, P.D., and Handelsman, J. (2006) Toward a census of bacteria in soil. *PLoS Comput Biol* **2**: e92.
- Schmidt, M., and de Lorenzo, V. (2012) Synthetic constructs in/for the environment: managing the interplay between natural and engineered Biology. *FEBS Lett* **586**: 2199–2206.
- Schnoes, A.M., Brown, S.D., Dodevski, I., and Babbitt, P.C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* **5**: e1000605.
- Schoemaker, H.E., Mink, D., and Wubbolts, M.G. (2003) Dispelling the myths—biocatalysis in industrial synthesis. *Science* **299**: 1694–1697.
- Sharma, P., Jindal, S., Bala, K., Kumari, K., Niharika, N., Kaur, J., *et al.* (2013) Functional screening of enzymes and bacteria for the dechlorination of hexachlorocyclohexane by a high-throughput colorimetric assay. *Biodegradation* **25**: 179–187.
- Shetty, R.P., Endy, D., and Knight, T.F., Jr (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* **2**: 5.
- Silva-Rocha, R., and de Lorenzo, V. (2011) Implementing an OR-NOT (ORN) logic gate with components of the SOS regulatory network of *Escherichia coli*. *Mol Biosyst* **7**: 2389–2396.
- Silva-Rocha, R., Martinez-Garcia, E., Calles, B., Chavarria, M., Arce-Rodriguez, A., de Las Heras, A., *et al.* (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* **41**: D666–D675.
- Simon, C., Herath, J., Rockstroh, S., and Daniel, R. (2009) Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* **75**: 2964–2968.
- Singhania, R.R., Patel, A.K., Sukumaran, R.K., Larroche, C., and Pandey, A. (2013) Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour Technol* **127**: 500–507.
- Siuti, P., Yazbek, J., and Lu, T.K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol* **31**: 448–452.
- Sleator, R.D., Shortall, C., and Hill, C. (2008) Metagenomics. *Lett Appl Microbiol* **47**: 361–366.
- Sommer, M.O., Church, G.M., and Dantas, G. (2010) A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. *Mol Syst Biol* **6**: 360.
- Steele, H.L., Jaeger, K.E., Daniel, R., and Streit, W.R. (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* **16**: 25–37.
- Suenaga, H., Ohnuki, T., and Miyazaki, K. (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol* **9**: 2289–2297.
- Suenaga, H., Koyama, Y., Miyakoshi, M., Miyazaki, R., Yano, H., Sota, M., *et al.* (2009a) Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. *ISME J* **3**: 1335–1348.
- Suenaga, H., Mizuta, S., and Miyazaki, K. (2009b) The molecular basis for adaptive evolution in novel extradiol dioxygenases retrieved from the metagenome. *FEMS Microbiol Ecol* **69**: 472–480.
- Sul, W.J., Park, J., Quensen, J.F., 3rd, Rodrigues, J.L., Seliger, L., Tsoi, T.V., *et al.* (2009) DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Appl Environ Microbiol* **75**: 5501–5506.
- Tabor, S. (2001) Expression using the T7 RNA polymerase/promoter system. *Curr Protoc Mol Biol* **16.2.1**–16.2.11.
- Terron-Gonzalez, L., Medina, C., Limon-Mortes, M.C., and Santero, E. (2013) Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Sci Rep* **3**: 1107.
- Thompson, C.E., Beys-da-Silva, W.O., Santi, L., Berger, M., Vainstein, M.H., Guimaraes, J.A., and Vasconcelos, A.T. (2013) A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. *AMB Express* **3**: 65.
- Turnbaugh, P.J., and Gordon, J.I. (2008) An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**: 708–713.
- Uchiyama, T., and Miyazaki, K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**: 616–622.
- Uchiyama, T., and Miyazaki, K. (2010) Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl Environ Microbiol* **76**: 7029–7035.
- Uchiyama, T., and Watanabe, K. (2008) Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nat Protoc* **3**: 1202–1212.

- Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* **23**: 88–93.
- Varaljay, V.A., Howard, E.C., Sun, S., and Moran, M.A. (2010) Deep sequencing of a dimethylsulfoniopropionate-degrading gene (*dmdA*) by using PCR primer pairs designed on the basis of marine metagenomic data. *Appl Environ Microbiol* **76**: 609–617.
- Vieites, J.M., Guazzaroni, M.E., Beloqui, A., Golyshin, P.N., and Ferrer, M. (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiol Rev* **33**: 236–255.
- Voigt, C.A. (2006) Genetic parts to program bacteria. *Curr Opin Biotechnol* **17**: 548–557.
- Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., and Charles, T.C. (2006) Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol* **72**: 384–391.
- Wang, G.Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., *et al.* (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* **2**: 2401–2404.
- Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., and Yu, Z. (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS ONE* **8**: e78507.
- Ward, R. (2011) Cellulase engineering for biomass saccharification. In *Routes to Cellulosic Ethanol*. Buckeridge, M.S., and Goldman, G.H. (eds). New York, USA: Springer, pp. 135–151.
- Warnecke, F., and Hess, M. (2009) A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J Biotechnol* **142**: 91–95.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560–565.
- Weber, W., and Fussenegger, M. (2010) Synthetic gene networks in mammalian cells. *Curr Opin Biotechnol* **21**: 690–696.
- Wenzel, S.C., and Muller, R. (2005) Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr Opin Biotechnol* **16**: 594–606.
- Wexler, M., Bond, P.L., Richardson, D.J., and Johnston, A.W. (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* **7**: 1917–1926.
- Williamson, L.L., Borlee, B.R., Schloss, P.D., Guan, C., Allen, H.K., and Handelsman, J. (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol* **71**: 6335–6344.
- Wong, D.W., Chan, V.J., and McCormack, A.A. (2009) Functional cloning and expression of a novel Endo- $\alpha$ -1,5-L-arabinanase from a metagenomic library. *Protein Pept Lett* **16**: 1435–1441.
- Yan, X., Geng, A., Zhang, J., Wei, Y., Zhang, L., Qian, C., *et al.* (2013) Discovery of (hemi-) cellulase genes in a metagenomic library from a biogas digester using 454 pyrosequencing. *Appl Microbiol Biotechnol* **97**: 8173–8182.
- Yuhong, Z., Shi, P., Liu, W., Meng, K., Bai, Y., Wang, G., *et al.* (2009) Lipase diversity in glacier soil based on analysis of metagenomic DNA fragments and cell culture. *J Microbiol Biotechnol* **19**: 888–897.
- Zaprasias, A., Liu, Y.J., Liu, S.J., Drake, H.L., and Horn, M.A. (2010) Abundance of novel and diverse *tfdA*-like genes, encoding putative phenoxyalkanoic acid herbicide-degrading dioxygenases, in soil. *Appl Environ Microbiol* **76**: 119–128.
- Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., *et al.* (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res* **41**: e98.
- Zhan, J., Ding, B., Ma, R., Ma, X., Su, X., Zhao, Y., *et al.* (2010) Develop reusable and combinable designs for transcriptional logic gates. *Mol Syst Biol* **6**: 388.