36. Casjens, S. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* **49**, 277–300 (2003).
37. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
38. Angly, F. *et al.* PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6**, 41 (2005).
39. Bak, P. *How Nature Works: The Science of Self-Organized Criticality* (Springer–Verlag, New York, 1996).
40. Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114. (2000).
41. Paul, J., Jeffrey, W. & DeFlaun, M. Dynamics of extracellular DNA in the marine environment. *Appl. Environ. Microbiol.* **53**, 170–179 (1987).
42. Paul, J., Jiang, S. & Rose, J. Concentration of viruses and dissolved DNA from aquatic environments by vortex flow filtration. *Appl. Environ. Microbiol.* **57**, 2197–2204 (1991).
43. Simon, M. & Azam, F. Protein content and protein synthesis rates of planktonic marine bacteria. *Mar. Ecol. Prog. Ser.* **51**, 201–213 (1989).
44. Rohwer, *F. Construction and analyses of linker-amplified shotgun libraries (LASLs)* [online], <http://www.sci.sdsu.edu/PHAGE/LASL/index.htm> (2002).
45. Blaisdell, B. E., Campbell, A. M. & Karlin, S. Similarities and dissimilarities of phage genomes. *Proc. Natl Acad. Sci. USA* **93**, 5854–5859 (1996).

Acknowledgements
Support from the National Science Foundation and Moore Foundation is gratefully acknowledged.

Competing interests statement
The authors declare no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez
ΦX147 | *Streptococcus pyogenes* strain MGAS315 | *Xylella fastidiosa* strain Temecula1

FURTHER INFORMATION
Forest Rohwer's laboratory: http://phage.sdsu.edu
BLAST: http://www.ncbi.nlm.nih.gov/BLAST
FASTA: http://www.ebi.ac.uk/fasta33
GenBank: http://www.ncbi.nlm.nih.gov/Genbank/index.html
PHACCS: http://phage.sdsu.edu/phaccs
Phrap computer program: http://www.phrap.com
Phred computer program: http://www.phrap.com/phred
Sequencher: http://www.genecodes.com/sequencher
Access to this interactive links box is free online.

OUTLOOK

Metagenomics and industrial applications

Patrick Lorenz and Jürgen Eck

Abstract | Different industries have different motivations to probe the enormous resource that is uncultivated microbial diversity. Currently, there is a global political drive to promote white (industrial) biotechnology as a central feature of the sustainable economic future of modern industrialized societies. This requires the development of novel enzymes, processes, products and applications. Metagenomics promises to provide new molecules with diverse functions, but ultimately, expression systems are required for any new enzymes and bioactive molecules to become an economic success. This review highlights industrial efforts and achievements in metagenomics.

Metagenomics[1] has the potential to substantially impact industrial production. The dimensions of the enormous biological and molecular diversity, as shown by Torsvik[2], Venter[3] and their co-workers, are truly astonishing. A pristine soil sample might contain in the order of 10^4 different bacterial species. More than one million novel open reading frames, many of which encode putative enzymes, were identified in a single effort that sampled marine prokaryotic plankton retrieved from the Sargasso Sea.

An industrial perspective

In this perspective, the discussion is limited to prokaryotes, as their genomes are most easily targeted by the functional screening tools available in metagenomics and because it is assumed, based on published literature, that the largest biodiversity occurs in the bacterial lineages[4–6]. Different industries are interested in exploiting the resource of uncultivated microorganisms that has been identified through large-scale environmental genomics for several reasons detailed below.

The ideal biocatalyst. For any industrial application, enzymes need to function sufficiently well according to several application-specific performance parameters (FIG. 1). With the exception of yeasts and filamentous fungi, access to novel enzymes and biocatalysts has largely been limited by the comparatively small number of cultivable bacteria. A corollary of this limitation is, however, that any application has to be designed with enzymatic constraints in mind, leading to suboptimal process and reaction conditions. Instead of designing a process to fit a mediocre enzyme, it is conceivable that the uncultivated microbial diversity, together with *in vitro* evolution technologies, might be used to find a suitable natural enzyme(s) that can serve as a backbone to produce a designer enzyme that optimally fits process requirements that are solely dictated by substrate and product properties[7].

Novelty. For industries that produce bulk commodities such as high-performance detergents, a single enzyme backbone with superior functionality that has an entirely new sequence would be useful to avoid infringing competitors' intellectual property rights. This problem is illustrated by the fact that substitutions at nearly every position in the mature 275 amino acid BPN (bacillus protease Novo type, from *Bacillus amyloliquefaciens*) subtilisin have been claimed in patents[8].

Maximum diversity. The pharmaceutical and supporting fine-chemicals industries often seek entire sets of multiple diverse biocatalysts to build in-house toolboxes for biotransformations[9]. These toolboxes need to be rapidly accessible to meet the strict timelines of a biosynthetic-feasability evaluation in competition with traditional synthetic chemistry.

Elusive metabolites. Many pharmacologically active secondary metabolites are produced by bacteria that live in complex CONSORTIA (see Glossary) or by bacteria that inhabit niches that are difficult to reconstitute *in vitro*[10]. So, although there are reports on how to circumvent this general problem of microbial cultivation either by mimicking natural habitats[11] or by allowing for interspecies communication after single cell micro-encapsulation[12], the cloning and heterologous expression of biosynthetic genes that encode secondary metabolites (usually present as gene clusters) is the most straightforward and reproducible method of accessing their biosynthetic potential.

Industrial enzyme applications

Enzymes are used in a wide range of applications and industries[13]. They are required in only minute quantities to synthesize kilograms of stereochemically challenging chiral SYNTHONS that are used as building blocks to produce highly active pharmaceuticals[14], and at a kiloton/year scale as active ingredients for bulk products such as high-performance

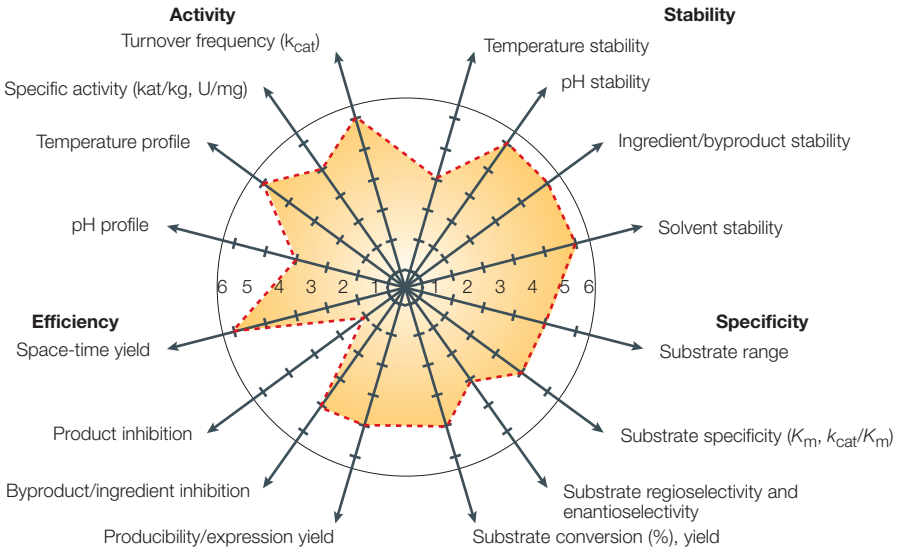

Figure 1 | **Multi-parameter footprint analysis.** This figure illustrates the ideal biocatalyst concept. Each enzyme candidate from the metagenome is ranked, from low (rating of 1) to high (rating of 6) using a specific set of criteria, to produce a multi-parameter fingerprint (shown in yellow). Criteria include *in vitro* enzyme activity, efficiency, specificity and stability. This decision matrix reveals the strengths and weaknesses of every candidate enzyme, so that the most promising candidate enzymes from diverse enzyme libraries can be selected for further process development by re-screening, protein engineering or directed evolution methods. kat, catalytic reaction rate; k_{cat}, catalytic constant; K_m, Michaelis constant; U, unit.

laundry detergents[8]. Their versatility allows their use in many applications, including processes to degrade natural polymers such as starch, cellulose and proteins, as well as for the REGIOSELECTIVE or ENANTIOSELECTIVE synthesis of asymmetric chemicals.

Global industrial enzyme sales were estimated at $2.3 billion in 2003 (REF. 15). The main profits were divided among detergents ($789 million), food applications ($634 million), agriculture/feed ($376 million), textile processing ($237 million), and pulp/paper, leather and other applications including enzymes for the production of fine and bulk chemicals ($222 million). In the face of soaring energy costs, dwindling fossil resources, environmental pollution and a globalized economy, the large-scale use of biotechnology instead of, or to complement, traditional industrial production processes, particularly in the chemical sector, is viewed as both an opportunity and a necessity. In the future, novel biotechnological applications should boost the market for industrial enzymes.

White biotechnology

The notion that environmentally sound, commercially viable biotechnological processes can take their respected place in a global industrial environment has been acknowledged for some time[16]. Companies from Europe, Canada, Japan, South Africa and the USA reported on their experiences in processes as diverse as the production of acrylamide (Mitsubishi Rayon, Japan) and the use of enzymes in oil-well completion (British Petroleum Exploration, UK). Currently, the movement towards implementing sustainable technologies and processes is gaining momentum, particularly in Europe.

'Industrial' or WHITE BIOTECHNOLOGY is currently a buzzword in the European biobusiness community. The term was coined in 2003 by the European Association for Bioindustries (EuropaBio), based on a case study report, and it denotes all industrially harnessed bio-based processes that are not covered by the RED BIOTECHNOLOGY (medical) or GREEN BIOTECHNOLOGY (plant) labels[17]. White biotechnology has its roots in ancient human history and its products are increasingly part of everyday life, from vitamins, medicines, biofuel and bioplastics to enzymes in detergents or dairy and bakery products. It is the belief of industrial promoters, analysts and policy makers that white biotechnology has the potential to impact industrial production processes on a global scale. The main long-term applications of white biotechnology will be focused on replacing fossil fuels with renewable resources (biomass conversion), replacing conventional processes with bioprocesses (including metabolic engineering) and creating new high-value bioproducts, including nutraceuticals, performance chemicals and bioactives.

Besides the involvement of the food, feed, detergent and politically heavily-promoted biofuel industries, it is the globally operating chemical and pharmaceutical industries that are riding this 'third wave of biotechnology'[18] (a term illustrating the chronology of developments in which red and green biotechnology come first and second, respectively). As industries face increasing low-cost competition, particularly from East Asia, and political pressure to reduce their environmental impact and resource consumption to improve sustainability, it is felt that there is a strong need for smart and innovative technologies, processes and products to remain competitive.

The McKinsey consultancy predicts that, by 2010, biotechnology could be applied in the production of between 10% and 20% of all chemicals sold (amounting to a value of $160 billion) and that up to 60% of all fine chemicals (medium-volume products used as intermediates in the manufacturing of products such as pharmaceuticals, flavours, fragrances, agro-chemicals and detergents) might be produced using biotechnology[17]. Even for the traditional mainstay of the chemical industries, the polymer market (typical bulk-volume products), McKinsey predicts that biotechnology might account for up to 10% of the output.

The European Chemical Industry Council and EuropaBio have responded to a call from the European Commission, which asked for suggestions for ten central technological platforms, by proposing a 'European Technology Platform for Sustainable Chemistry'. White biotechnology will be one of three pillars of the platform, together with materials science, and reaction and process design. As a corollary, this triggered the nomination of sustainable chemistry as one of the thematic priorities in the European Union's seventh framework research programme, with the aim of strengthening the scientific and technological bases of European industry and encouraging its international competitiveness. Clearly, this joint European initiative does not come without 'stimulating' precedent from forward-looking international peers in the United States or Japan, and leading European stakeholders (among them DSM[19], Degussa, BASF, Henkel, Novozymes and Genencor) have urged policy makers to act to ensure a favourable competitive position. The promises of industrial biotechnology cannot be realized without coordinated and well-funded research and development (R&D) efforts, a supportive and guiding political framework and early public involvement to pre-empt concerns over new

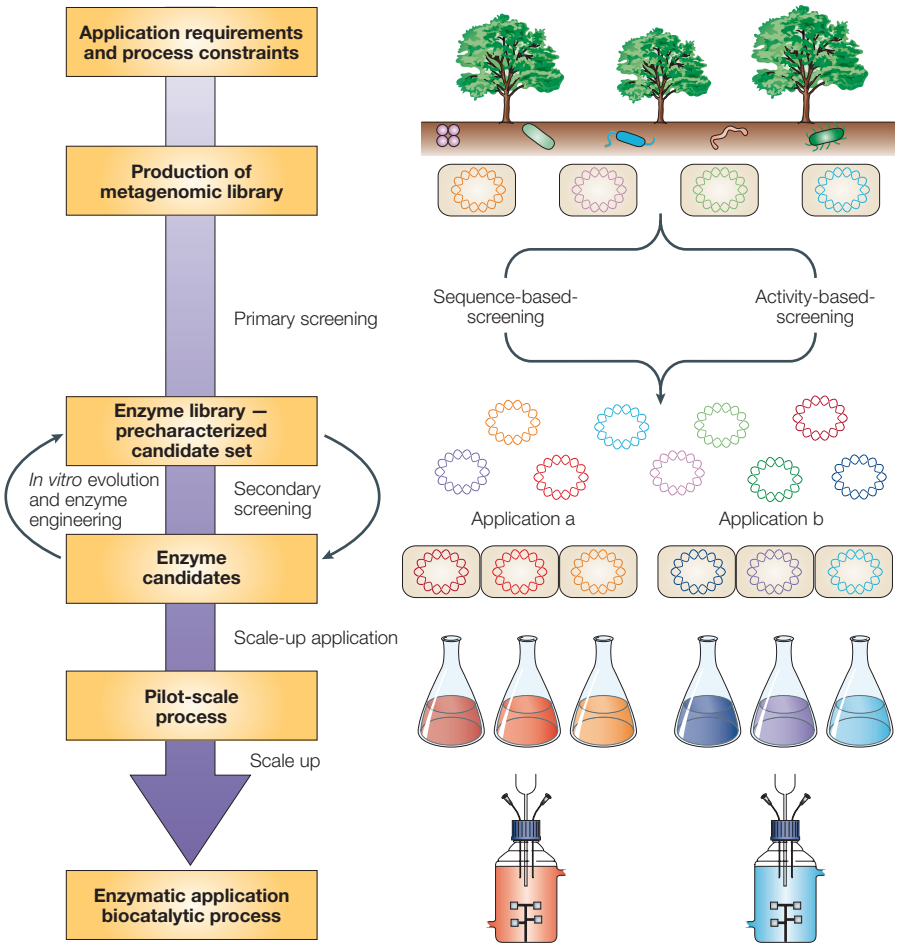

Figure 2 | **Industrial enzymes — from the metagenome to applications and processes.** This figure illustrates industrial rational bioprospecting of the metagenome. A library of cloned DNA is produced and primary screening, based on application requirements such as the conversion of an indicator substrate, produces enzyme libraries (different clones, which encode different enzymes, are indicated by different colours) that serve as platforms for subsequent development. Secondary screening of the enzyme library can identify process-specific properties such as substrate specificity, activity or stability. Primary and secondary screening both involve several stages based on different criteria, therefore the screening stages are multi-layered. Enzyme engineering and *in vitro* gene evolution can form a part of the development process. Subsets of cloned enzymes (red clones represent one subset, blue clones represent another subset) are then used in scale-up application or process testing to identify suitable enzyme candidates. Finally, process improvement by enzyme optimization and process engineering are carried out. Economic feasibility must be proven in a pilot-scale process environment as a prerequisite for scale-up to production or final application scale. Examples of the use of enzyme libraries in an industrial context include nitrilases[36] (Diversa), alcohol dehydrogenases[9] (Schering Plough) and glycosylhydrolases[39] (BRAIN AG).

technologies and generate the demand that industry requires for economic success[20].

Whereas the food, feed and detergent industries typically concentrate on a limited number of enzyme reactions and substrates, the chemical and pharmaceutical industries deal with thousands of chemically and structurally diverse molecules, and production of each of these requires individual enzymatic solutions. Consequently, owing to a wealth of potentially useful biocatalysts, biotechnological applications of microbial resources are increasingly popular with the chemical industries and are viewed as indispensable for the modern organic chemist[21]. With many successfully implemented processes operating worldwide[22,23] and the number of industrialized biotransformations having doubled every decade since 1960 (REF. 24), it is estimated that biocatalysis might provide a superior synthetic solution over classical chemistry in 10% of processes[25]. In fact, the unavailability of an appropriate biocatalyst is thought to be a limiting factor for any biotransformation process[22].

So, there is ample demand for novel enzymes and biocatalysts, and metagenomics is currently thought to be one of the most likely technologies to provide the candidate molecules required[26,27]. The diversity of potential substrates for enzymatic transformations in the fine-chemical industry and the short time that is usually available to propose viable synthetic routes (particularly for the pharmaceutical industry) make it useful to produce pre-characterized enzyme libraries using generic substrates, before screening for a specific enzyme that is required for biotransformation of a particular substrate of interest (FIG. 2).

Screening for industrial enzymes

After Torsvik's seminal report in 1980 on the extraction and digestion of genomic DNA from bacterial samples prepared from soil[28] and the proof of the concept for generating gene libraries directly from environmental DNA[29], it was only in the next decade that the first such metagenome libraries were reported. These were constructed from non-(PCR-)amplified genomic DNA fragments, isolated directly from marine plankton[30] and from enriched consortia sampled from cellulose digesters[31]. The research on the enriched consortia was particularly important, in that it describes the first successful expression screening of metagenomic DNA. Once this was achieved, the usefulness of the metagenome approach for biotechnology was clear. It was already evident that most environmental bacteria were recalcitrant to cultivation[32], but metagenomics meant that the biotechnological potential of these uncultivated bacteria might be accessible by directly cloning DNA sequences retrieved from the environment. To our knowledge, Diversa (San Diego, California, USA) and TerraGen Discovery (Vancouver, Canada, taken over in 2000 by Cubist Pharmaceuticals, Cambridge, Massachusetts, USA) were the first companies to file patents on specific aspects of metagenome technology in 1996. Diversa claimed methodologies to isolate and normalize environmental DNA prior to cloning and screening and described a resulting fosmid library created from marine picoplankton[33]. TerraGen's claims were focused on the retrieval of environmental DNA sequences from soil to express industrially relevant xylanases[34]. Since then, several small- to medium-sized enterprises (SMEs) have used uncultivated environmental microbial resources to retrieve relevant enzymes and biocatalysts. Most industrial metagenomic discoveries reported so far have been made by SMEs or academic groups collaborating

Table 1 | **Activity-based screening for industrially relevant enzymes and biocatalysts from metagenomic libraries**

Function	Habitat	Library type	Average insert size (kb)	Number of clones screened	Library size (Mb)	Substrate	Number of hits	Hit rate (hit per Mb)	Ref.
Esterase/lipase	Forest soil	Plasmid	8	67,000	536	Tributyrin	98	1/5.5	*
Esterase/lipase	Forest soil	Fosmid	40	19,968	799	Tributyrin	47	1/17	*
Esterase/lipase	Sandy ecosystem	Fosmid	30	29,884	903	Tributyrin	49	1/18	*
Esterase/lipase	Sandy ecosystem	Fosmid	40	25,344	1,014	Tributyrin	29	1/35	*
Esterase/lipase	Soil	Plasmid	6	286,000	1,716	Tributyrin	3	1/572	63
Esterase/lipase	Soil	Plasmid	6	730,000	4,380	Triolein	1	1/4,380	63
Esterase/lipase	Soil	BAC	27	3,648	100	Bacto Lipid	2	1/50	64
Oxidation of polyols	Soil	Plasmid	3	900,000	2,700	1,2-ethanediol; 1,2-propanediol; 2,3-butanediol	15	1/180	65
Alcohol oxidoreductase	Soil/enrichment	Plasmid	4	400,000	1,600	Glycerol/1,2-propanediol	10	1/160	66
Amidase	Soil/enrichment	Plasmid	5	193,000	965	D-phenylglycine-L-leucine	7	1/138	67
Amylase	Soil	Plasmid	5	80,000	400	Starch	1	1/400	68
Amylase	Soil	BAC	27	3,648	100	Starch	8	1/12	64
Biotin production	Soil/excrement enrichment	Cosmid	35	50,000	1,750	Biotin-deficient medium	7	1/250	69
Protease	Soil	Plasmid	10	100,000	1,000	Skimmed milk	1	1/1,000	70
Cellulase	Sediment enrichment	λ phage	6	310,000	1,860	Carboxymethyl-cellulose	3	1/620	71
Chitinase	Seawater	λ phage	5	825,000	4,125	Methylumbelliferyl-diacetylchitobioside	11	1/375	72
Dehydratase	Soil/sediment enrichment	Plasmid	4	560,000	2,240	Glycerol	2	1/1,120	38
4-Hydroxybutyrate conversion	Soil	Plasmid	6	930,000	5,580	4-Hydroxybutyrate	5	1/1,116	73
β-Lactamase	Soil	Plasmid	5	80,000	400	Ampicillin	4	1/100	68

The screening host for all studies was *Escherichia coli*. *Unpublished results, K. Liebeton *et al.*, BRAIN AG. BAC, bacterial artificial chromosome.

with larger companies, mainly in the chemical industry. Different methods for cloning environmental DNA and screening for enzymes or bioactive molecules have been devised by different companies. Only a few selected examples per company are mentioned in the following section.

Diversa, the largest and most prominent specialist biotech company for the commercialization of metagenome technologies, has described several approaches to access 'uncultivable' microorganisms. Applying a classical growth-selection-based expression strategy, diverse environmental libraries were constructed in *Escherichia coli* using phage λ and were screened for nitrile use. After growth in media containing nitriles as the sole nitrogen source, more than 100 new and diverse nitrilase genes were recovered[35]. The resulting enzyme library is marketed to serve the fine-chemical and pharmaceutical industries[36]. Besides target-gene-specific molecular enrichment strategies (see below), a technical focus is also placed on the sensitivity and throughput of screening by using cell-sorting technology[37]. Diversa has partnered with several companies for the discovery and development of enzymes, among them DSM, Syngenta and BASF. Genencor International (Palo Alto, California, USA) teamed up with researchers from Göttingen University (Germany) to search for new dehydratases of potential use in the production of 1,3-propanediol from glucose, using a metabolically engineered strain of *E. coli*[38]. Both sequence-based screens, targeting new genes that were related to known reference genes, and growth selection protocols, to screen for expressed and functional enzyme activity, were used. The German biotech company BRAIN AG (Zwingenberg, Germany) collaborated with the detergent manufacturer Henkel (Düsseldorf, Germany) to screen for novel glycosyl hydrolases for use in laundry applications[39], and with chemicals company Degussa (Düsseldorf, Germany) to use sequence-based screening strategies to clone novel nitrile hydratases[40].

Because of the ease of genetic manipulation of *E. coli*, combined with the availability of elaborate cloning and expression systems, most of the screens to date have used gene libraries constructed in this bacterium. However, the reported hit rates for the required targets vary between enzyme classes (TABLE 1). This might be due to difficulties in expressing specific genes in *E. coli*, their scarcity in the metagenome pool or a combination of both. To overcome the potential limitations of expressing genes in *E. coli*,

Table 2 | **Activity-based screening for antimicrobial metabolites from metagenomic libraries**

Habitat	Library type	Average insert size (kb)	Number of clones screened	Library size (Mb)	Target organism	Screening host	Positive hits	Hit rate (hit per Mb)	Ref.
Soil	BAC	44.5	24,546	1,092	Several	*	3	1/364	74
Soil	BAC	63	12,000	756	*B. subtilis, S. aureus*	*E. coli*	4	1/189	56
Soil	BAC	27	3,648	100	*B. subtilis*	*E. coli*	1	1/100	64
Soil	Fosmid	30	13,440	403	*S. aureus*	*E. coli*	3	1/134	‡
Soil	Cosmid	nd	5,000	nd	*B. subtilis*	*E. coli*	1	nd	55
Soil	Cosmid	30–45§	nd	Unknown	*B. subtilis*	*E. coli*	10	1/300–900	75

*Producer colonies identified by production of a pigment. Isolated pigments were shown to have antimicrobial activity. ‡Unpublished results, R. Schulze and G. Meurer, BRAIN AG. §Estimate based on specifications of the pWEB cosmid cloning system (Epicentre Technologies). BAC; bacterial artificial chromosome; *B. subtilis*, *Bacillus subtilis*; *E. coli*, *Escherichia coli*; nd, not disclosed; *S. aureus*, *Staphylococcus aureus*.

alternative cloning hosts such as *Bacillus subtilis, Streptomyces* spp., *Pseudomonas* spp. or eukaryotic expression systems could be used. However, this option is limited by the low throughput that is associated with alternative cloning hosts and serves mainly as an option for large-insert (>30 kb) cloning, which is used in bioactive molecule discovery (see below and TABLE 2). A different strategy, whereby the abundance of target genes in the metagenomic DNA pool is increased through microbial enrichment using low concentrations of externally added nutrients prior to DNA extraction, has been pursued by the Icelandic biotech company Prokaria[41] (Reykjavik, Iceland). Sequence-specific enrichments ('biopanning'), carried out either on the metagenomic input DNA prior to cloning or on the library after cloning in plasmid vectors, have been reported by Diversa[42].

In addition to new technologies to amplify DNA from limited resources using random primers and strand-displacing DNA polymerase from phage Φ29 (REF. 43), a strategy promoted by Lucigen (Middleton, Wisconsin, USA), it is clear that current mass-sequencing efforts in several laboratories will facilitate the *in silico* identification of open reading frames that might encode potentially useful enzymes[44]. Another interesting approach for a focused, high throughput screening system, substrate-induced gene expression screening (SIGEX), was recently applied to isolate catabolic enzymes from a groundwater metagenome[45]. Environmental DNA was cloned upstream of a plasmid-borne promoter-less *gfp* (green fluorescent protein) reporter gene in *E. coli*. Substrate (benzoate or naphthalene)-induced expression of GFP in recombinant cells was used to identify and isolate (by fluorescence activated cell sorting, FACS) single cells carrying metagenomic inserts (catabolic operons) with substrate-responsive promoters and associated downstream enzymes that were subsequently shown to convert the inducer substrates to products.

Once new genes are cloned and screened for activity, the main stumbling block is the expression of pure protein in sufficient amounts at reasonable costs. A cheap and efficient enzyme, usually produced in efficient expression systems like bacilli or filamentous fungi, is a key factor for success, particularly when the enzyme functions as part of the final (bulk) product such as in detergents. In the fine-chemical industry, this might be similar for bulk product synthesis. Particularly in the pharmaceutical industry, the time taken for a target compound to come to the market is decisive, and in these applications it might be even more important for a company to have a large collection of biochemically diverse catalysts, even if these molecules are not expressed in large amounts.

Searching for bioactive molecules

Despite large increases in the R&D expenditure of big and medium-sized Pharma, the number of pharmacologically active 'new chemical entities' (NCEs) has not increased proportionally[46], and this is most notable in the paucity of new antibiotics. Big Pharma currently find antibiotics economically unattractive to develop for various reasons when compared with drugs to treat long-term chronic conditions like obesity and high cholesterol[47]. This could have dramatic consequences for the global antibiotic resistance problem. Natural compounds and their derivatives will continue to have an important role in drug discovery, as they have done in the past[48] — natural products constituted 63% of all newly approved anti-infectives between 1983 and 1994 (REF. 49). Many natural products are bacterial and fungal secondary metabolites, but as most microorganisms cannot easily be cultivated, it is probable that many potentially active compounds have never been characterized. Metagenomics might have an invaluable role in the discovery process of new bioactive molecules[50,51]. For this reason, metagenomes have attracted the attention of academia and specialized pharmaceutical companies.

TerraGen Discovery[52,53] first incorporated PCR-amplified microbial-soil-metagenome genes into incomplete polyketide synthase clusters present in a recipient *Streptomyces* strain to show the usefulness of functionally cloning and expressing environmental genes as part of a host cell's secondary metabolite biosynthetic machinery. In a second step, the same researchers reported the production

Glossary

CONSORTIUM
Physical association between cells of two or more types of microorganisms. Such an association might be advantageous to at least one of the microorganisms.

ENANTIOSELECTIVE
Selection of a particular enantiomer, one of a pair of stereospecific isomers.

GREEN BIOTECHNOLOGY
Biotechnology applied to agriculture, for example, production of transgenic crop plants with genetically engineered improved pest resistance.

RED BIOTECHNOLOGY
Biotechnology in health care that uses substances produced in the human body to fight disease — medicines, vaccines, diagnostics and gene therapy.

REGIOSELECTIVE
Selection of a change that occurs with a greater frequency at one site than at several other sites, usually involving a structural or positional isomer.

SYNTHON
A molecule used as a chemical building block to synthesize complex compounds.

WHITE BIOTECHNOLOGY
Industrial use of biotechnology, for example, to produce fine chemicals, biofuel, bioplastics, enzymes for use in detergents, food and feed. The boundaries between white, green and red biotechnology are blurred.

of novel bioactive metabolites (which were patented as terragines) after cloning metagenomic DNA fragments in a *Streptomyces lividans* host using *E. coli* shuttle cosmid vectors. This approach addressed two important issues of using heterologous expression systems to obtain useful amounts of pharmacologically active, naturally occurring secondary metabolites. First, it is important to choose an appropriate expression host. *S. lividans* is an actinomycete, a family of high GC-content Gram-positive bacteria that devote large proportions of their genomes to biosynthetic genes that synthesize considerable amounts of secondary metabolites. Therefore, an actinomycete expression host is more likely than *E. coli* to provide the necessary genetic, biochemical and physiological background to produce bioactive molecules from genes that are derived from uncharacterized bacteria which are also thought to be members of this taxonomic lineage. This, however, might change in the future, as successful attempts to metabolically engineer *E. coli* to improve its use as a cloning and production host for complex secondary metabolites (for example, 6-deoxyerythronolide B, an antibiotic precursor) have been reported[54]. Second, this study showed the importance of developing high-capacity vectors and suitable transformation protocols to accommodate the large gene clusters that are involved in secondary metabolite biosynthesis (TABLE 2).

Similar strategies were used by researchers from Aventis Pharmaceuticals[55] (Cambridge, Massachusetts), which addressed the cloning-vector capacity and expression host issues by designing bacterial artificial chromosomes that could be used for cloning metagenome DNA in *E. coli*[56], *Streptomyces* and *Pseudomonas* strains[57] (key members of this research team have since joined ActivBiotics of Lexington, Massachusetts, USA). These endeavours were paralleled by efforts to extract large clonable DNA fragments (mainly from soil). This work was supported by European biotech company BRAIN AG, collaborating with researchers from the University of Darmstadt, Germany[58], and BioSearch Italia (Gerenzano, Italy; recently rebranded as Vicuron Pharmaceuticals), collaborating with the University of Warwick, UK[59]. Through examination of the patent literature, it seems that Diversa is using similar methods for the discovery of small molecules using direct cloning[60].

So, the obvious potential of the metagenomic approach for drug discovery has been realized by companies around the world. Whereas in the recent past, SMEs engaged in metagenomic research for the sake of anti-infectives discovery, other uses for bioactive molecules in cancer therapy and other medical fields will probably lead companies that are currently engaged in red biotechnology to use metagenomics as its potential unfolds[61].

Metagenomics — an industrial future

As the excitement about genetic access to the boundless realms of microbial diversity slowly gives way to the reality of tapping into this diversity, the usual challenge of heterologous gene expression needs to be addressed to turn metagenomic technologies into commercial successes, particularly in applications for which bulk enzyme or product quantities have to be produced at competitive prices. Metagenomics, together with *in vitro* evolution and high-throughput screening technologies, provides industry with an unprecedented chance to bring biomolecules into industrial application. That this is possible has been shown by Diversa. A full cycle from the discovery of novel molecular scaffolds from multiple resources, including metagenomes, to *in vitro* evolution technology to generate an improved biocatalyst for a specific application was realized to produce α-amylases for applications in the hot and acidic process of starch liquefaction[62], showing the feasibility of the 'ideal biocatalyst' concept.

Patrick Lorenz and Jürgen Eck are at BRAIN AG, Darmstaedterstrasse 34, 64673 Zwingenberg, Germany.
Correspondence to P. L.
e-mail: pl@brain-biotech.de

doi:10.1038/nrmicro1161

1. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 (2004).
2. Torsvik, V., Ovreas, L. & Thingstad, T. F. Prokaryotic diversity-magnitude, dynamics, and controlling factors. *Science* **296**, 1064–1066 (2002).
3. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
4. Curtis, T. P., Sloan, W. T. & Scannell J. W. Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA* **99**, 10494–10499 (2002).
5. Ward, B. B. How many species of prokaryotes are there? *Proc. Natl Acad. Sci. USA* **99**, 10234–10236 (2002).
6. Schloss, P. D. & Handelsman, J. Status of the microbial census. *Microbiol. Mol. Biol. Rev.* **68**, 686–691 (2004).
7. Burton, S. G., Cowan, D. A. & Woodley J. M. The search for the ideal biocatalyst. *Nature Biotechnol.* **20**, 37–45 (2002).
8. Maurer, K. H. Detergent proteases. *Curr. Opin. Biotechnol.* **15**, 330–334 (2004).
9. Homann, M. J. *et al.* Rapid identification of enantioselective ketone reductions using targeted microbial libraries. *Tetrahedron* **60**, 789–797 (2004).
10. Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **21**, 519–538 (2004).
11. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296**, 1127–1129 (2002).
12. Zengler, K. *et al.* Cultivating the uncultured. *Proc. Natl Acad. Sci. USA* **99**,15681–15686 (2002).
13. Kirk, O., Borchert, T. V. & Fuglsang C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351 (2002).
14. Patel, R. N. *et al.* Enzymic preparation of (3R-cis)-3-(acetyloxy)-4-phenyl-2-azetidinone: a taxol side-chain synthon. *Biotechnol. Appl. Biochem.* **20**, 23–33 (1994).
15. Global Industry Analysts. *Industrial enzymes — a global multi-client market research project.* (GIA, San José, California, USA, 2004).
16. *The application of biotechnology to industrial sustainability.* (Organisation for economic co-operation and development (OECD), 2001).
17. Schepens, H. *et al. White Biotechnology: gateway to a more sustainable future* [online] <www.mckinsey.com/clientservice/chemicals/pdf/BioVision_Booklet_final.pdf> (2003).
18. Riese, J. *Surfing the third wave: new value chain creation opportunities in industrial biotechnology* (McKinsey & Company, Frankfurt, Germany, 2003).
19. DSM Position Document, *Industrial (white) biotechnology: an effective route to increase EU innovation and sustainable growth* [online] <http://www.b-basic.nl/documents/DSMIndustrial-white-biotechnology.pdf> (2004).
20. Herrera, S. Industrial biotechnology — a chance at redemption. *Nature Biotechnol.* **22**, 671–675 (2004).
21. Schoemaker, H. E., Mink, D. & Wubbolts, M. G. Dispelling the myths — biocatalysis in industrial synthesis. *Science* **299**, 1694–1697 (2003).
22. Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–268 (2001).
23. Schmid, A., Hollmann, F., Park, J. B. & Buhler, B. The use of enzymes in the chemical industry in Europe. *Curr. Opin. Biotechnol.* **13**, 359–366 (2002).
24. Straathof, A. J., Panke, S. & Schmid, A. The production of fine chemicals by biotransformations. *Curr. Opin. Biotechnol.* **13**, 548–556 (2002).
25. Roberts S. M. Biocatalysts in synthetic organic chemistry. *Tetrahedron* **60**, 499–500 (2004).
26. Lorenz, P., Liebeton, K., Niehaus, F. &, Eck, J. Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr. Opin. Biotechnol.* **13**, 572–577 (2002).
27. Schloss, P. D. & Handelsman, J. Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* **14**, 303–310 (2003).
28. Torsvik V. L. & Goksoyr J. Determination of bacterial DNA in soil. *Soil Biol. Biochem.* **10**, 7–12 (1980).
29. Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J. The analysis of natural microbial populations by ribosomal RNA sequences. *Adv. Microbial Ecol.* **9**, 1–55 (1986).
30. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378 (1991).
31. Healy, F. G. *et al.* Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl. Microbiol. Biotechnol.* **43**, 667–674 (1995).
32. Staley, J. T. & Konopka, A. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* **39**, 321–346 (1985).
33. Short, J. M. & Mathur, E. J. Production and use of normalized DNA libraries. US Patent 5,763,239 (1998).
34. Radomski, C. C. A., Seow, K. T., Warren, R. A. J. & Yap, W. H. Method for isolating xylanase gene sequences from soil DNA, compositions useful in such method and compositions obtained thereby. US Patent US 5,849,491 (1998).
35. Robertson, D. E. *et al.* Exploring nitrilase sequence space for enantioselective catalysis. *Appl. Environ. Microbiol.* **70**, 2429–2436 (2004).
36. DeSantis, G. *et al.* An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. *J. Am. Chem. Soc.* **124**, 9024–9025 (2002).
37. Short, J. M. & Keller, M. Method of high-throughput screening of an environmental library. US Patent 6,806,048 (2004).
38. Knietsch, A., Bowien, S., Whited, G., Gottschalk, G. & Daniel, R. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl. Environ. Microbiol.* **69**, 3048–3060 (2003).

39. Breves, R., Maurer, K. H., Eck, J., Lorenz P. & Zinke H. New glycosyl hydrolases. PCT patent application WO 03/054177 (2003).
40. Liebeton K. & Eck J. Identification and expression in *E. coli* of novel nitrile hydratases from the metagenome. *Eng. Life Sci.* **4**, 554–562 (2004).
41. Hobel, C. F. V. *et al.* Use of low nutrient enrichment to access novel amylase genes in silent diversity of thermophiles. *World J. Microbiol. Biotechnol.* **20**, 801–809 (2004).
42. Gray, K. A., Richardson, T. H., Robertson, D. E., Swanson, P. E. & Subramanian, M. V. Soil-based gene discovery: a new technology to accelerate and broaden biocatalytic applications. *Adv. Appl. Microbiol.* **52**, 1–27 (2003).
43. Detter, J. C. *et al.* Isothermal strand-displacement amplification applications for high-throughput genomics. *Genomics* **80**, 691–698 (2002).
44. McDonald, A. E. & Vanlerberghe, G. C. Alternative oxidase and plastoquinol terminal oxidase in marine prokaryotes of the Sargasso Sea. *Gene* **349**, 15–24 (2005).
45. Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnol.* **23**, 88–93 (2004).
46. Service, R. F. Surviving the blockbuster syndrome. *Science* **303**, 1796–1799 (2004).
47. Leeb M. A shot in the arm. *Nature* **431**, 892–893 (2004).
48. Cragg, G. M. & Newman, D. J. Medicinals for the millennia: the historical record. *Ann. N. Y. Acad. Sci.* **953**, 3–25 (2001).
49. Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.* **66**, 1022–1037 (2003).
50. Osburne, M. S., Grossman, T. H., August, P. R. & MacNeil, I. A. Tapping into microbial diversity for natural products drug discovery. *ASM News* **66**, 411–417 (2000).
51. Daniel, R. The soil metagenome — a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* **15** 199–204 (2004).
52. Seow, K. T. *et al.* A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J. Bacteriol.* **179**, 7360–7368 (1997).
53. Wang, G. Y. *et al.* Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* **2**, 2401–2404 (2000).
54. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
55. Courtois, S. *et al.* Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* **69**, 49–55 (2003).
56. MacNeil, I. A. *et al.* Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* **3**, 301–308 (2001).
57. Martinez, A. *et al.* Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* **70**, 2452–2463 (2004).
58. Quaiser, A. *et al.* First insight into the genome of an uncultivated crenarchaeote from soil. *Environ. Microbiol.* **4**, 603–611 (2002).
59. Berry, A. E., Chiocchini, C., Selby, T., Sosio, M. & Wellington E. M. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiol. Lett.* **223**, 15–20 (2003).
60. Short, J. Screening for novel bioactivities. US Patent US 6,057,103 (2000).
61. Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl Acad. Sci. USA* **101**, 16222–16227 (2004).
62. Richardson, T. H. *et al.* A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable α-amylase *J. Biol. Chem.* **277**, 26501–26507 (2002).
63. Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G. & Daniel, R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 3113–3116 (2000).
64. Rondon, M. R. *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).
65. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. & Daniel, R. 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 (2003).
66. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. & Daniel, R. Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl. Environ. Microbiol.* **69**, 1408–1416 (2003).
67. Gabor, E. M., de Vries, E. J. & Janssen, D. B. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ. Microbiol.* **6**, 948–958 (2004).
68. Gabor, E. M. *Harvesting novel biocatalysts from the metagenome.* Thesis, Univ. Groningen (2004).
69. Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W. R. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl. Environ. Microbiol.* **67**, 89–99 (2001).
70. Gupta, R., Beg, Q. K. & Lorenz, P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* **59**, 15–32 (2002).
71. Rees, H. C., Grant, S., Jones, B., Grant, W. D. & Heaphy, S. Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* **7**, 415–421 (2003).
72. Cottrell, M. T., Moore, J. A. & Kirchman, D. L. Chitinases from uncultured marine microorganisms. *Appl. Environ. Microbiol.* **65**, 2553–2557 (1999).
73. Henne, A., Daniel, R., Schmitz, R. A. & Gottschalk, G. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* **65**, 3901–3907 (1999).
74. Gillespie, D. E. *et al.* Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* **68**, 4301–4306 (2002).
75. Brady, S. F., Chao, C. J. & Clardy, J. Long-chain *N*-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* **70**, 6865–6870 (2004).

Acknowledgements

We would like to thank our collegues at BRAIN AG and particularly C. Schleper of the University of Bergen, Norway, and the Zentrum für Evolutionsforschung und Biotechnologie (ZEB), Darmstadt, Germany, for fruitful discussions and sharing ideas and resources.

Competing interests statement

The authors declare competing financial interests; see Web version for details.

Online links

FURTHER INFORMATION

ActivBiotics: www.activbiotics.com
Aventis Pharmaceuticals: www.aventispharma-us.com
BRAIN AG: www.brain-biotech.de
Cubist Pharmaceuticals: www.cubist.com
EuropaBio: www.europabio.org
The European Chemical Industry Council: www.cefic.org
The European Technology Platform for Sustainable Chemistry: www.suschem.org
Degussa: www.degussa.com
Diversa: www.diversa.com
Genencor International: www.genencor.com/wt/home
Henkel: www.henkel.com
Lucigen: www.lucigen.com
Prokaria: www.prokaria.com
Vicuron Pharmaceuticals: www.vicuron.com
Access to this interactive links box is free online.