

Alternative hosts for functional (meta)genome analysis

Wolfgang Liebl · Angel Angelov · Julia Juergensen · Jennifer Chow · Anita Loeschcke ·
Thomas Drepper · Thomas Classen · Jörg Pietruzka · Armin Ehrenreich ·
Wolfgang R. Streit · Karl-Erich Jaeger

Received: 30 May 2014 / Revised: 14 July 2014 / Accepted: 15 July 2014 / Published online: 5 August 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Microorganisms are ubiquitous on earth, often forming complex microbial communities in numerous different habitats. Most of these organisms cannot be readily cultivated in the laboratory using standard media and growth conditions. However, it is possible to gain access to the vast genetic, enzymatic, and metabolic diversity present in these microbial communities using cultivation-independent approaches such as sequence- or function-based metagenomics. Function-based analysis is dependent on heterologous expression of metagenomic libraries in a genetically amenable cloning and expression host. To date, *Escherichia coli* is used in most cases; however, this has the drawback that many genes from heterologous genomes and complex metagenomes are expressed in *E. coli* either at very low levels or not at all. This review emphasizes the importance of establishing alternative microbial expression systems consisting of different genera and species as well as customized strains and vectors optimized for heterologous expression of membrane proteins, multigene clusters encoding protein complexes or entire

metabolic pathways. The use of alternative host-vector systems will complement current metagenomic screening efforts and expand the yield of novel biocatalysts, metabolic pathways, and useful metabolites to be identified from environmental samples.

Keywords Metagenomics · Metagenomic libraries · Functional genome analysis · Functional metagenome analysis · Heterologous expression hosts

Introduction

In response to their biotic and abiotic surroundings, microorganisms in the course of evolution have developed a broad variety of genetic and physiological traits which allow them to survive and proliferate successfully in their habitats. In natural habitats, microorganisms usually do not exist as pure clonal populations but rather in communities of varying complexity consisting of several to thousands of taxonomically different organisms. Taken together, the number of prokaryotic taxa on earth has been estimated to amount to 10^6 – 10^8 distinct genospecies (Simon and Daniel 2011) which outnumbers the list of described isolated species (roughly 10^4) by several orders of magnitude. This vast microbial diversity represents a huge natural resource for the isolation of new genes, enzymes, and metabolic pathways which await investigation and eventually exploitation for biotechnological applications. However, the use of molecular marker techniques has demonstrated that, depending on the habitat investigated, only a very small fraction of the organisms (for example from soil samples approximately 0.1–1 %) can be cultivated using standard techniques (Amann et al. 1995). Within the past years, new high-throughput–*omics* methods have become available which allow the characterization of genes, gene transcripts, proteins, and metabolites from environmental

W. Liebl (✉) · A. Angelov · A. Ehrenreich
Lehrstuhl für Mikrobiologie, Technische Universität München,
Emil-Ramann-Str. 4, 85654 Freising, Germany
e-mail: wliebl@wzw.tum.de

J. Juergensen · J. Chow · W. R. Streit
Abteilung für Mikrobiologie und Biotechnologie,
Universität Hamburg, Hamburg, Germany

A. Loeschcke · T. Drepper · K.-E. Jaeger
Institut für Molekulare Enzymtechnologie,
Heinrich-Heine-Universität Düsseldorf und Institut für
Bio- und Geowissenschaften IBG-1, Forschungszentrum Jülich,
52426 Jülich, Germany

T. Classen · J. Pietruzka
Institut für Bioorganische Chemie, Heinrich-Heine-Universität
Düsseldorf und Institut für Bio- und Geowissenschaften IBG-1,
Forschungszentrum Jülich, 52426 Jülich, Germany

samples without prior cultivation of the respective organisms. The collection of methods focused on isolation, cloning, and sequencing of environmental DNA has been coined metagenomics (Handelsman et al. 1998). Extraction of bacterial DNA from environmental samples was reported more than three decades ago (Torsvik and Goksøyr 1978) followed by its use to study microorganisms (Olsen et al. 1986; Pace et al. 1986) and to construct and perform screenings on environmental gene libraries, and to functionally express metagenomic genes (e.g., Healy et al. 1995; Handelsman et al. 1998).

Today, metagenomic methods are often used to characterize the composition and the dynamics of changes within microbial communities, e.g., by amplification, cloning, and sequence analysis of conserved marker genes or gene fragments, such as 16S rRNA gene sequences derived from environmental DNA samples. On the other hand, metagenomic approaches are not restricted to phylogenetic biodiversity analysis but also allow the retrieval of functional information directly from environmental samples; for example, the identification of novel enzymes for biotechnological applications including biocatalytic reactions implemented within chemical synthesis routes (see Lorenz and Eck 2005; Steele et al. 2009). Along these lines, various metagenomic studies using soil, arctic sediment, hot spring, termite intestine, or cow rumen samples were reported (Wang et al. 2012; Fu et al. 2013; Graham et al. 2011; Nimchua et al. 2012; Ferrer et al. 2012).

The retrieval of genes encoding new and biotechnologically relevant enzymes from the environment can follow two different approaches. (1) Sequence-based strategies depend on methods like PCR-based screening or hybridization of clone DNA against degenerated probes or massive random sequencing of environmental DNA. As a priori sequence information is needed, this approach will deliver only new enzyme variants belonging to already known enzyme families, but fails to identify genes encoding truly novel enzymes unrelated to known enzymes (Liebl 2011; Steele et al. 2009). (2) Sequence-independent functional screening bears the potential to uncover genes for enzymes and enzyme classes with little or even no homology to already known enzyme families (e.g., Delavat et al. 2012). However, success of function-based screening is dependent on the functional heterologous expression of metagenomic genes in a given host organism and the availability of appropriate screening assays (Leis et al. 2013; Steele et al. 2009). In the light of the recent decay of sequencing costs, it becomes increasingly attractive to use a combined approach of both sequence and function-based searches.

Previous reviews have dealt with various aspects of metagenomic studies such as methods of sampling, library construction, high-throughput sequencing, in silico sequence analysis and interpretation of metagenomic data, functional screenings, and the use of metagenomics for discovery of industrial biocatalysts and pharmaceutically relevant compounds (Daniel 2005; Delmont et al. 2011a, b; Ekkers et al.

2012; Leis et al. 2013; Lorenz et al. 2002; Lorenz and Eck 2005; Shokralla et al. 2012; Simon and Daniel 2011; Streit and Schmitz 2004; Taupp et al. 2011; Thomas et al. 2012; Wooley et al. 2010). Here, we will look at metagenomics from a different angle and describe novel expression tools and alternative host organism(s) which we consider useful for metagenomic library construction and screening.

Microbial cloning and expression hosts play an important role in metagenomics for gene library construction, gene amplification, and, in the case of functional metagenomics, also for gene expression. Metagenomic libraries often comprise several Gbp of environmental DNA harboring millions of genes. Strikingly, however, most of this enormous biodiversity remains unused for the discovery and utilization of new proteins and metabolites due to drawbacks inherent to the most widely used expression host *Escherichia coli*. To date, only very few reports are available which describe attempts to use shuttle vectors to broaden the host-range for functional screening purposes (Aakvik et al. 2009; Angelov et al. 2009; Courtois et al. 2003; Craig et al. 2010; Kakirde et al. 2011; Martinez et al. 2004; Troeschel et al. 2010). Recent advances in “recombineering”, i.e., homologous recombination methods based on targeted recombination using phage-derived recombinase enzymes, now offer interesting perspectives for establishing further multihost systems for comparative functional screening in different hosts (Leis et al. 2013).

In the following, we discuss drawbacks of more traditional microbial expression systems and report about recently developed strategies to further improve already existing or establish new systems for functional metagenome analysis to demonstrate the need for alternative screening hosts and to show the potential which such hosts bear for metagenomic library exploitation.

Drawbacks of established expression systems

The vast majority of genes in any metagenomic library constructed from a complex microbial community is routinely cloned and screened in standard cloning and expression hosts, particularly in *E. coli* because of its ease of handling, efficient genetic manipulation, and the availability of sophisticated genetic tools. However, it should be noted that the functional expression of genes and operons is a complex molecular biological process, which involves transcription, translation, protein folding, and sometimes export or secretion. Thus, it is highly unlikely that one single expression host like *E. coli* will be able to functionally express all the heterologous genes and operons contained in a complex metagenomic library (Fig. 1). It can be assumed that close phylogenetic relatedness of the DNA donor strain and the expression host increases the probability of functional expression (Leis et al. 2013). However, in

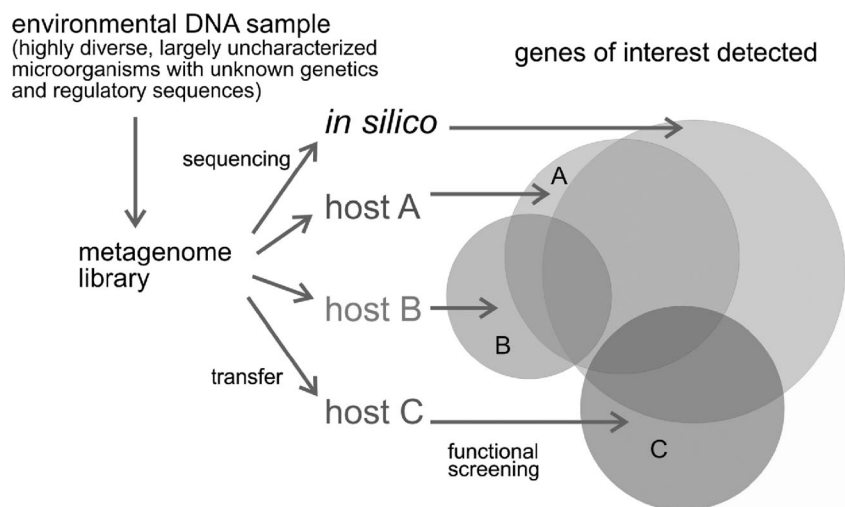


Fig. 1 Strategy to maximize the yield of genes of interest from functional screening of metagenomic libraries. Not all genes encoding an enzymatic function can be identified with sequence analysis because *in silico* detection relies on similarity to known sequences. The same holds true for other sequence-based methods such as PCR or hybridization strategies using degenerate primers (not depicted in the figure). Therefore, completely novel genes for a given function will be missed with purely sequence-based search efforts. Function-based screening on the other hand also allows the detection of genes which lack similarity to previously known sequences

encoding the function of interest. Due to (1) the diverse nature of the cloned environmental DNA with respect to promoter sequences, G+C content etc., and (2) differences in the gene expression machineries of the hosts it can be expected that the detection frequency as well as the nature of the genes identified will differ for the different host organisms used in functional screening. Therefore, since no single host organism is able to express all genes present in a complex metagenome, combining the use of several hosts can improve the yield

complex metagenomic libraries which can contain DNA from many different deep-branching phyla, the majority of genes originate from organisms unrelated to the screening host. Hence, functional metagenomic screenings for new enzymes can presently access only a small fraction of the tremendous genetic biodiversity. The main reasons include (1) a lack of efficient methods to construct and stably maintain large (meta)genomic gene libraries in prokaryotic organisms other than *E. coli* (Leis et al. 2013) and (2) the limited expression capacity of commonly used screening hosts due to missing promoter recognition, different codon usage, failure to correctly fold and assemble enzyme proteins, missing capacity for cofactor synthesis, etc.

In functional screening approaches in practice today, large libraries often containing millions of cloned metagenomic genes must be expressed in a high-throughput manner to identify only one or a few enzymes of interest although there are differences concerning the enzyme type sought for (e.g., see Lorenz and Eck 2005; Lorenz et al. 2002). Because this is a highly labor- and cost-consuming process, the question arises on how current metagenomic approaches can be improved to increase the yield of retrieval of new genes for a desired function from metagenomes. Strategies that aid to achieve this goal include (1) the (parallel) use of different host organisms for metagenomic library screening, (2) the improvement of the expression of heterologous genes within a given host, and (3) the development of rapid, more sensitive screening assays for enzymatic functions of interest. In particular, alternative cloning and expression hosts need to be

established and sophisticated genetic tools and methods for their efficient genetic modification must be constructed.

In the following, a selection of specific examples is presented demonstrating the value of using novel microbial expression systems for the functional expression of metagenome-derived genes and operons.

Engineered *E. coli* strains equipped with foreign sigma factors

A recent transcriptome analysis of *E. coli* Epi 300 carrying different metagenome fosmid suggested that the transcription of metagenome-derived genes is a limiting step. The *E. coli* Epi 300 strain (Epicentre, Madison, USA) is the strain that is most commonly used as a host for metagenome fosmid libraries as it is part of a commercial fosmid cloning kit that allows rather efficient fosmid cloning and thus has been used by many labs. Using RNAseq technology in the background of this strain, we observed that many of the genes that were derived from bacteria phylogenetically distantly related to *E. coli* were less frequently transcribed compared to those genes that originated from closely related species (W.R.S., unpublished data). In bacteria, promoter recognition is carried out by the initiation factor σ , which recruits RNA polymerase core enzyme to the promoter for transcription initiation. Bacteria encode a single housekeeping σ -factor and a variable number of accessory σ -factors that turn on transcription of specific sets of genes in response to environmental stimuli

(Wösten 1998). In *E. coli*, seven σ -factors are known of which *rpoD* encodes for the housekeeping σ -factor. *E. coli rpoD* is responsible for the majority of transcription of essential genes during exponential growth and it recognizes a typical -10 and -35 binding motif (Gruber and Gross 2003). Thus, the addition of heterologous σ -factors to the *E. coli* genome may aid to enlarge its transcription proficiency. As a first step towards this direction, we have constructed *E. coli* strains harboring an additional *rpoD* gene of the phylogenetic distant bacterium *Clostridium cellulolyticum*. One of these strains was designated *E. coli* UHH01. Initial tests with this strain showed an increase in detection frequency by 20–30 % in functional metagenome screens (W.R.S., unpublished data). This was done by screening for hydrolytic enzymes (i.e., lipases and amylases) on agar plates and in liquid media. Thereby, we observed that both, the parent and the engineered strain, resulted in the detection of different fosmid clones suggesting that the additional *rpoD* gene resulted in the elevated transcription of functional genes that differed from the parent strain. Although we cannot exclude that the foreign *rpoD* gene causes an increased stress in the engineered strain, the concept of using additional *rpoD* genes appears to be promising and it has so far allowed the detection of functional genes that would not have been detected using the nonmodified screening host. Besides the functional metagenomic screening purposes, transcription factor modification and the use of exogenous sigma factors in expression host strains, which has also been called “transcriptional engineering”, can be useful for the improvement of the productivity of valuable compounds in recombinant bacteria and even industrial production hosts (Wang et al. 2014; Yu et al. 2008).

Heterologous expression of large gene clusters

Bacteria produce numerous metabolites with high-value activities such as antibiosis, cytotoxicity, and immunosuppression (Newman and Cragg 2012). Their biosynthesis is genetically encoded by clustered genes which are difficult to target by metagenome screening, as their synthesis in a heterologous production host is hampered by various limitations: (1) The co-expression of all relevant genes has to be achieved, (2) the respective host needs to produce functional enzymes which may be assembled to higher-order enzyme complexes, and (3) an appropriate screening method must exist allowing to identify the produced metabolite.

(1) *The expression of clustered genes is challenging* The concerted functional expression of many genes located in large gene clusters is often limited, since the original promoters are not necessarily recognized by the host RNA polymerases. Furthermore, the use of flanking host-specific promoters only rarely allows complete transcription of

metagenomic genes because premature transcription termination frequently occurs due to large DNA template length or transcription termination signals. Moreover, gene clusters are often composed of several transcriptional units arranged in different orientations (Fischbach and Voigt 2010) inevitably rendering genes inaccessible to a single flanking host-specific promoter. As an alternative, the use of T7 RNA polymerase (T7RP) for the expression of clustered genes has been suggested as it was reported to be highly processive and to ignore bacterial transcription termination sites (Zhang et al. 2011; Ongley et al. 2013). In nonmetagenome studies, the T7 system has already proven useful for directed heterologous expression of polyketide and other gene clusters in *E. coli* and *Rhodobacter capsulatus* (Zhang et al. 2011; Stevens et al. 2013; Ongley et al. 2013; Arvani et al. 2012). An expression tool named TREX was recently established (Loeschcke et al. 2013) which utilizes convergent T7RP-dependent expression of a given DNA fragment thereby enabling full transcription of all cluster genes irrespective of their orientation and operon structure. As a proof of concept, bidirectional transcription and metabolite production was shown using a carotenoid (6.9 kb) and a prodigiosin gene cluster (21.8 kb). Comparative expression studies demonstrated that the TREX system is applicable in different host organisms such as *E. coli*, *Pseudomonas putida*, and *R. capsulatus*. Thus, in order to overcome the mentioned limitations at the transcriptional level, genetic tools allowing to tune the host RNA polymerases for recognizing metagenomic promoters (see above) or to use alternative viral promoter/polymerase systems for the concerted expression of metagenomic genes can help to adapt standard and alternative bacterial expression hosts for functional metagenome analysis.

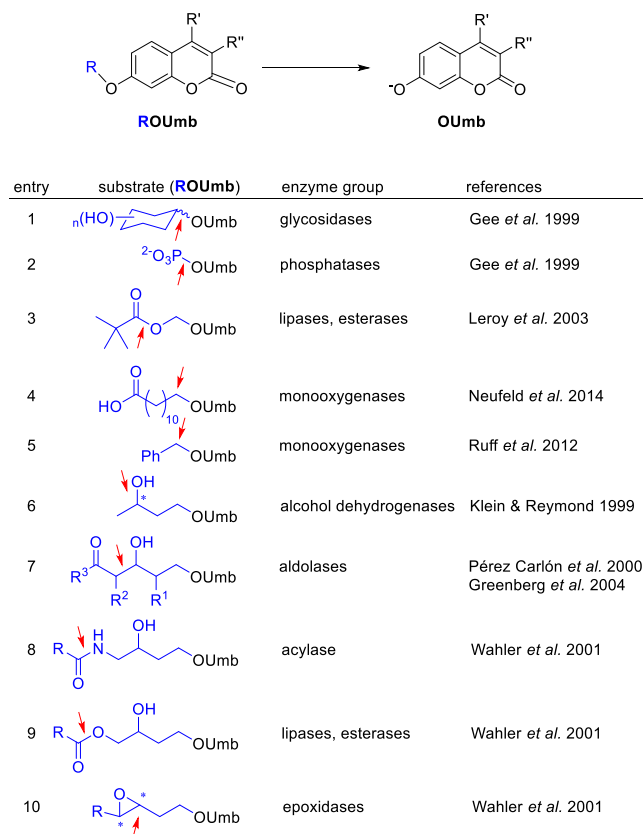
(2) *Suitability of the host metabolic background is largely unpredictable* The host organism provides the critical background for successful metabolite production by expression of metagenome-derived metabolic pathways. Functional enzymes must be synthesized requiring appropriate codon usage and a folding machinery, supply of suitable precursor molecules, and persistence of intermediates and end products, which finally should not be toxic to the host. These highly complex processes necessarily produce completely different outcomes from different pathway/host combinations. Accordingly, from directed heterologous expression of known pathway-encoding gene clusters such as enterocin AS-48 from *Enterococcus faecalis*, isomigrastatin from *Streptomyces platensis*, or violacein cluster from a species of *Duganella* sp., it is known that results are differential depending on host organisms (Fernández et al. 2007; Feng et al. 2009; Yang et al. 2011; Jiang et al. 2010). In addition, comparative screenings of metagenomic libraries with different hosts led to significantly more positive hits. For example, Craig et al. screened for phenotypes such as pigmentation and antibiosis using

Agrobacterium tumefaciens, *Burkholderia graminis*, *Caulobacter vibrioides*, *E. coli*, *P. putida*, and *Ralstonia metallidurans*, each displaying different results (Craig et al. 2010). In another study, a metagenome library was functionally screened in *Streptomyces lividans* focusing on phenotypes such as hemolytic activity and pigment production. The positive clones found were also tested in *E. coli*, where none of them produced the screened phenotype (McMahon et al. 2012). In this context, again, it is worth to mention that the highly variable results of function-based screening approaches are usually caused by host-specific differences at the expression as well as the metabolic level.

(3) *Detection of novel metagenomic activities* Naturally, the assay determines the outcome of any screening. Hence, function-based metagenome screenings for novel metabolites have focused on defined and easily detectable phenotypes, such as antibiotic resistances, antibiotic activities, morphological changes, or pigmentations. To expand the group of targeted compounds, screening hosts can genetically be modified to apply simple high-throughput screening methods. For example, an elegant *E. coli*-based colorimetric screening method for terpene synthases was developed recently which can be used to identify terpenoid pathways (Furubayashi et al. 2014). Another strategy combining sequence- and function-based methods of screening metagenomic libraries led to the identification of tryptophan dimer biosynthesis clusters in *E. coli* (Chang and Brady 2011; Chang and Brady 2014). Nevertheless, new screening approaches are needed to uncover more of the microbial chemical world.

Metagenomic strategies can be employed successfully to identify novel enzymes with biocatalytic potential. To this end, hydrolases and oxidoreductases are of special interest. Appropriate screening assays are needed for their detection (Franken et al. 2010; Reymond 2006). Among them, fluorimetric assays have the advantage of higher sensitivity as compared to chromogenic ones, which is of particular importance with respect to the usually moderate expression levels observed in metagenomic libraries. The most common fluorimetric probe is umbelliferone, which has been used in various compositions for screening (; Reymond 2006, 2009). The fluorogenic moiety of the respective substrate molecules (ROUmb) is either located close to site of the enzymatic reaction (entries 1–5 in Scheme 1) or remote from it (entries 6–10) with the latter type of substrates being significantly more stable thus also reducing the frequency of false-positive hits drastically.

These fluorogenic screenings are highly parallelisable using microtiter plates and robotic liquid handling systems. The step from high- to ultrahigh-throughput has recently been demonstrated (Ruff et al. 2012) with an umbelliferone-based monooxygenase screening system using fluorescence-activated cell sorting (FACS). This method allows testing of



Scheme 1 Umbelliferone (Umb) can be used for various screenings maintaining a high sensitivity. The umbelliferone (Umb) can be released directly by hydrolysis (entries 1+2) or by spontaneous (entries 3–7) or triggered (entries 8–10) secondary reactions. The primary enzymatic reaction taking place is indicated by the arrow

10,000 of single cells per minute thus enabling the screening of large metagenomic libraries and further highlights the exquisite signal-to-noise-ratio of fluorescence probes.

High-throughput screening can also be carried out conveniently by using agar plates containing chromogenic substrates; such assays have been described for various hydrolases (see, e.g., Topakas and Christakopoulos 2014; Jaeger and Kovacic 2014) and also for laccases which can degrade, e.g., xenobiotics in waste water and lignin. Golyshin and co-workers demonstrated the use of colorimetric assays for identifying an unknown laccase from mammal ruminal metagenome (Beloqui et al. 2006).

The above-mentioned screening systems allow for the functional identification of novel biocatalytic activities within metagenomic libraries. However, it should be noted that the detailed biochemical characterization of an enzyme still requires time and often high-end specialized equipment, especially when addressing the issue of enantioselectivity (Franken et al. 2010).

Besides the modification of *E. coli* strains and the development of new tools and detection systems, the expansion of the available expression systems beyond this traditional host via the establishment of phylogenetically diverse new

expression hosts and the use of more than one host for screening of metagenomic libraries can help to overcome the drawbacks mentioned above. If sophisticated genetic tools and methods for efficient genetic modification are made available, such hosts can be used in high-throughput functional screening strategies to enhance the detection frequency of the genes of interest (Fig. 1). Examples for novel host bacteria used in the authors' groups (Table 1) are briefly described in the following paragraphs.

***Thermus thermophilus*, a thermophilic host bacterium**

In the past years, the extremely thermophilic bacterium *T. thermophilus* has been developed as a host for large-insert library construction and functional screening of genomic and metagenomic libraries at elevated temperatures. *T. thermophilus* is a heterotrophic aerobic Gram-negative representative of the *Deinococcus-Thermus* phylum that grows at temperatures up to 85 °C. Genome sequences of *T. thermophilus* strains have been reported, and some genetic tools for cloning, selection and counterselection, genome modification, and inducible gene expression are available (see Cava et al. 2009; Angelov et al. 2009; Angelov et al. 2013; Liebl 2004). Importantly, *T. thermophilus* cells are highly and constitutively competent for natural transformation and are not discriminatory with respect to the source of the externally added DNA, which enables efficient introduction of heterologous DNA and genetic modification.

An important issue in functional metagenomic screening approaches is the decision about the insert sizes used for library construction. Small fragments (<15 kb) are cloned into plasmid vectors while high-molecular weight DNA can be

used for cloning into fosmids (up to 40 kb) or BACs (up to 200 kb). In small-insert libraries, each clone carries only a few genes, but by using high-copy number replication origins and sometimes strong vector promoters the detection of even weakly expressed genes and weakly active enzymes can be enhanced. In contrast, large-insert libraries carry many genes on each insert but heterologous expression must be driven mainly by native promoters located on the insert. For *E. coli*, various cosmid, fosmid, and BAC vectors with single-copy origin or alternatively inducible multicopy origins are available (see Leis et al. 2013).

Large DNA fragments of course bear more metagenomic information than small inserts; therefore, theoretically less clones from a library must be screened with functional screening assays, and, in addition, complete gene clusters can be expressed. However, the tradeoff for less assays is the risk that probably not all promoters on the metagenomic inserts will be active in the host's transcriptional background (Liebl 2011). In addition, other factors such as G+C content, Shine Dalgarno sequences, codon usage, etc. can cause a more or less pronounced bias on heterologous expression of metagenomic DNA, but unfortunately, such effects elicited by the host's expression apparatus have not been studied systematically.

The possibility to conveniently construct large-insert libraries from high molecular weight metagenomic DNA, i.e., using commercial fosmid vectors with cos sites for packaging of the ligated DNA into λ phage particles prior to infection of *E. coli* host cells, is a large advantage of the *E. coli* cloning system which is not available for other host bacteria. For the thermophilic host, *T. thermophilus* tools are now available which allow the transfer of recombinant fosmid inserts from *E. coli* to *T. thermophilus*. To this end, a fosmid library is first constructed in *E. coli* using the two-host fosmid vector

Table 1 Characteristics of alternative host bacteria described in this work

Organism	Phylum/Class Order Family	Growth temp Range or temp. used in lab (T_{opt})	Genome size ^a (accession no.)	G+C content (%)	Genetic accessibility
<i>Thermus thermophilus</i> HB27	<i>Deinococcus- Thermus</i> <i>Thermales</i> <i>Thermaceae</i>	40–85 °C (73 °C)	2.12 Mbp (AE017221, AE017222)	69.4	Natural transformation
<i>Pseudomonas antarctica</i>	γ - <i>Proteobacteria</i> <i>Pseudomonadales</i> <i>Pseudomonadaceae</i>	4–30 °C (22 °C)	6.27 Mbp	59.6	Electroporation heat shock transformation
<i>Rhodobacter capsulatus</i>	α - <i>Proteobacteria</i> <i>Rhodobacterales</i> <i>Rhodobacteraceae</i>	20–30 °C	3.87 Mbp (CP001312, CP001313)	66.6	Conjugation
<i>Gluconobacter oxydans</i> 621H	α - <i>Proteobacteria</i> <i>Rhodospirillales</i> <i>Acetobacteraceae</i>	25–30 °C	2.92 Mbp (CP000009, CP000004, CP000005, CP000006, CP000007, CP000008)	60.8	Electroporation conjugation

^a Total genome size, including chromosome and plasmids if present

pCT3FK (Angelov et al. 2009) which carries an antibiotic resistance marker which can be selected for in the thermophilic host, and DNA fragments that flank the chromosomal *pyrE* gene of *T. thermophilus* HB27. Recombinant fosmids isolated from *E. coli* library clones are introduced into *T. thermophilus* by natural transformation where site-specific integration into the *T. thermophilus* chromosome occurs via homologous recombination at the *pyrE* locus. In proof-of-principle studies, this two-host fosmid vector system has been used for the comparative functional screening of fosmid libraries constructed from chromosomal DNA from two thermophilic species, *Spirochaeta thermophila* and *Thermus brockianus*, in *T. thermophilus* as well as *E. coli*. Corresponding clones of both hosts (in *E. coli* each cloned insert is carried on the recombinant fosmid whereas in *T. thermophilus* the identical insert is integrated into the host's chromosome) were subjected to screening for hydrolase activities using plate assays. In both cases, more active clones were found with the host *T. thermophilus* than with *E. coli* (Angelov et al. 2009; Leis et al. 2013).

***Pseudomonas antarctica*, a psychrophilic host bacterium**

Heterologous expression of proteins is often hampered due to instability or toxicity of proteins in the mesophilic host *E. coli*. As many enzymes have a relatively low activity within the psychrophilic temperature range, expression at low temperatures can be an advantage when enzymes show a harmful effect on the metabolism, cell wall, or membrane of the host. Therefore, a psychrophilic expression host was developed in the laboratory of one of the authors (W.R.S.). The *P. antarctica* strain Shivaji CMS 35 is a nonpathogenic, free-living Gram-negative bacterium phylogenetically related to *P. fluorescens* and other *Pseudomonads*. The psychrophilic bacterium is able to grow in common Luria-Bertani (LB) and CASO broth between 4 and 30 °C with an optimum growth temperature at 22 °C (Reddy et al. 2004). It possesses only weak endogenic lipase activity and can utilize adonitol, meso-erythritol, D-galactose, D-glucose, glycerol, meso-inositol, and D-mannitol as carbon sources in contrast to, e.g., D-cellobiose, lactose, D-maltose, and sucrose (Reddy et al. 2004). Fortunately, it is sensitive to most antibiotics commonly used for cloning and it accepts and replicates commonly used broad host range vectors such as pBBR1MCS-5 (Kovach et al. 1995). In order to transform plasmids into the cells, protocols were established that allow easy transformation of *P. antarctica* by heat shock and electroporation. The uptake of the vector was confirmed by molecular methods, in particular plasmid isolation and PCR. In order to investigate the expression of functional enzymes within the psychrophilic bacterium, six different genes of metagenomic lipases and esterases within pBBR1MCS-5 under control of the *lac*

promoter were transformed into the strain that was grown at 22 °C without further induction. Activity assays on agar plates containing tributyrin and olive oil/rhodamine B as substrate revealed lipolytic activity of the crude cell extracts that was considerably higher than the weak endogenic lipase activity of the wild type (unpublished work).

The genome of *P. antarctica* was sequenced. It has an overall size of ~6.3 Mb with a G+C content of 59.6 %; it encodes a number of secretory systems including a complete set of genes for the assembly of a type 2 secretion machinery (Chow 2012, and own unpublished data). With its ability to grow at low temperatures, its easy transformability, and physiological properties, *P. antarctica* has the potential to become a promising expression and screening host for a variety of proteins that cannot be easily expressed in *E. coli*.

***R. capsulatus*, a facultative phototrophic host bacterium**

The heterologous expression of membrane proteins is a major concern, in particular for biomedical research since nearly 70 % of the available drugs are either directly or indirectly targeting human membrane proteins (Lundstrom 2007). Furthermore, many enzymes of microbes, plants, and mammals that are involved in the synthesis and functionalization of hydrophobic natural compounds, such as fatty acids and terpenes, are either peripheral or intrinsic membrane proteins. However, the intricate nature of membrane proteins often hampers their structural and functional studies because commonly used expression hosts like *E. coli* are in general optimized for the production of soluble proteins (Schlegel et al. 2010). Consequently, the activity of the membrane protein folding and translocation machinery as well as the intrinsic storage capacity of the host's membrane is commonly not appropriate or sufficient for foreign membrane proteins produced at high amounts (e.g., Wagner et al. 2006; Nannenga and Baneyx 2011). As a result of these limitations, heterologous expression of membrane proteins often leads to the formation of inclusion bodies consisting of misfolded membrane proteins or it is toxic to the host cell. Therefore, the development of alternative expression hosts is key for the function-based identification and production of novel membrane-bound proteins and enzymes.

R. capsulatus is a photosynthetic Gram-negative α -proteobacterium that has been used as model organism over decades to study the regulation and function of anoxygenic photosynthesis as well as CO₂ and N₂ fixation (e.g., Wu and Bauer 2008; Gregor and Klug 2002; Tichi and Tabita 2002; Masepohl and Hallenbeck 2010). Beside the photoautotrophic growth mode, where *R. capsulatus* uses carbon dioxide and dinitrogen as sole C and N sources, its metabolic versatility further enables this bacterium to grow under a broad range of different conditions in the light and dark. Because of its

facultative phototrophic nature, *R. capsulatus* is a promising alternative expression host that is particularly suited for the functional expression of heterologous membrane-bound enzymes and, in turn, for the catalytic conversion and storage of hydrophobic substrates and products: Phototrophic growth conditions induce an intracellular differentiation of the inner membrane, leading to the formation of membrane vesicles that house the photosynthetic apparatus. These membrane vesicles also provide an intrinsically high folding and incorporation capacity for recombinant membrane proteins and can further serve to accumulate catalytically converted hydrophobic compounds. These properties form the prerequisite for a rapid identification and functional overexpression of novel membrane proteins of biotechnological interest.

For the heterologous expression of single and multiple target genes, a set of different broad-host-range tools has been developed allowing comparative expression studies in *R. capsulatus* under different growth conditions as well as in other Gram-negative bacteria including *E. coli* and *P. putida* (Katzke et al. 2010; Katzke et al. 2012; Arvani et al. 2012; Loeschcke et al. 2013). The expression toolbox comprises replicative broad host range expression vectors (termed pRho plasmids) and cassettes for chromosomal integration (Ω Sp-PT7 and TREX, see above) carrying either the bacterial *aphII* promoter for constitutive and moderate expression or the viral T7 promoter for inducible T7RP-mediated high-level expression of target genes. Because of its inducer- and T7RP-independent activity, the *aphII* promoter is primarily useful for parallelized high-throughput screening approaches in various Gram-negative expression hosts. In contrast, the utilization of T7-RNA polymerase-dependent promoters requires appropriate host strains but allows, as outlined above, the concerted expression of multiple target genes, which are located on a metagenomic DNA fragment or a cluster of functionally coupled genes.

The *R. capsulatus* T7 expression system could already be used to express soluble recombinant proteins such as the yellow and the flavin-binding fluorescent proteins under heterotrophic and phototrophic conditions achieving protein yields of up to 80 mg l⁻¹ of culture (Drepper et al. 2007; Katzke et al. 2010) and the light-operating protochlorophyllide reductase from the marine phototrophic bacterium *Dinoroseobacter shibae* (Kaschner et al. 2014). Furthermore, the functional expression in *R. capsulatus* of microbial and human membrane proteins including membrane-bound enzymes (e.g., P450 monooxygenases) and receptors (e.g., rhodopsins) was successfully demonstrated (Malach, Özgür, Heck, Jaeger & Drepper, unpublished data). Finally, the T7 expression toolbox was also employed to facilitate the concerted expression of naturally clustered genes including the [NiFe] hydrogenase encoding gene cluster from *R. capsulatus*

(Arvani et al. 2012) and the *crt* gene cluster from *Pantoea ananatis* (Loeschcke et al. 2013).

***Gluconobacter oxydans*, a special host for the expression of membrane dehydrogenases**

An interesting case where the prerequisites for the in vivo expression and screening of membrane-bound enzymes from metagenomic DNA have recently been established is the case of the acetic acid bacterium *G. oxydans*. Acetic acid bacteria are acid-tolerant aerobic bacteria known for their special metabolic lifestyle of utilizing membrane-bound, pyrroloquinoline quinone (PQQ)- or flavin adenine dinucleotide (FAD)-dependent dehydrogenases for the incomplete oxidation of alcohols, aldehydes, polyols, sugars, and sugar derivatives. Their membrane dehydrogenases oxidize their substrates on the outer surface of the cytoplasmic membrane in a stereo- and regio-specific manner, feeding the electrons directly into the respiratory electron transport chain. These bacteria are currently used in various efficient whole-cell biocatalytic processes for the production of bulk and speciality chemicals such as organic acids, erythritol, dihydroxyacetone, pharmaceuticals, etc. After the establishment of efficient genetic tools for chromosomal gene insertion and replacement (Peters et al. 2013a; Kostner et al. 2013), *G. oxydans* strains have been constructed via step-by-step markerless deletion of all major membrane-bound dehydrogenases (Peters et al. 2013b). These *G. oxydans* multideletion strains have been successfully used for the expression of heterologous membrane dehydrogenase genes isolated from metagenomes of acetic acid bacteria-containing mother of vinegar microbial communities (Peters, Liebl and Ehrenreich, unpublished work). Functional expression of such metagenomic membrane dehydrogenases in the multideletion strain allows for the rapid and detailed in vivo characterization of their substrate specificity using a sensitive whole-cell activity assay (Peters et al. 2013b).

Conclusion

Today, metagenomic techniques are applied to characterize the composition of microbial communities from environmental samples and to investigate the abundance of marker genes indicative of certain physiological traits. From the biotechnological perspective, metagenomics represents the most important methodology to identify novel genes encoding single biocatalysts or entire biochemical pathways allowing to produce novel enzymes and valuable metabolites. The availability of advanced and high-throughput-compatible gene expression tools, including alternative and broadly applicable microbial expression systems, which can be combined to increase the yield of genes of interest from functional screening of

(meta)genomic libraries, will be essential to access the vast natural biodiversity.

Acknowledgments This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the program GenoMik (Genomforschung an Mikroorganismen; FKZ 0315586). Work in the laboratory of TD and KEJ was funded by the Deutsche Forschungsgemeinschaft through EXC 1028.

References

- Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Völker U, Ellingsen TE, Valla S (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiol Lett* 296:149–158
- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169
- Angelov A, Mientus M, Liebl S, Liebl W (2009) A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst Appl Microbiol* 32:177–185
- Angelov A, Li H, Geissler A, Leis B, Liebl W (2013) Toxicity of indoxyl derivative accumulation in bacteria and its use as a new counterselection principle. *Syst Appl Microbiol* 36:585–592
- Arvani S, Markert A, Loeschcke A, Jaeger K-E, Drepper T (2012) A T7 RNA polymerase-based toolkit for the concerted expression of clustered genes. *J Biotechnol* 159:162–171
- Beloqui A, Pita M, Polaina J, Martínez-Arias A, Golyshina OV, Zumárraga M, Yakimov MM, García-Arellano H, Alcalde M, Fernández VM, Elborough K, Andreu JM, Ballesteros A, Plou FJ, Timmis KN, Ferrer M, Golyshin PN (2006) Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen. Biochemical properties, structural analysis, and phylogenetic relationships. *J Biol Chem* 281:22933–22942
- Cava F, Hidalgo A, Berenguer J (2009) *Thermus thermophilus* as biological model. *Extremophiles* 13:213–231
- Chang F-Y, Brady SF (2011) Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017. *J Am Chem Soc* 133:9996–9999
- Chang F-Y, Brady SF (2014) Characterization of an environmental DNA-derived gene cluster that encodes the bisindolylmaleimide methylarcyriarubin. *Chembiochem* 15:815–821
- Chow J (2012) Doctoral thesis, Universität Hamburg. <http://ediss.sub.uni-hamburg.de/volltexte/2013/6001/>
- Courtois S, Cappellano CM, Ball M, Francou F-X, Normand P, Helync G, Martínez A, Kolvek SJ, Hopke J, Osburne MS, August PR, Nalin R, Guérineau M, Jeannin P, Simonet P, Pernodet J-L (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* 69:49–55
- Craig JW, Chang F-Y, Kim JH, Obiajulu SC, Brady SF (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
- Delavat F, Phalip V, Forster A, Plewniak F, Lett M-C, Lièvreumont D (2012) Amylases without known homologues discovered in an acid mine drainage: significance and impact. *Sci Rep* 2:354
- Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM (2011a) Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* 77:1315–1324
- Delmont TO, Robe P, Clark I, Simonet P, Vogel TM (2011b) Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J Microbiol Methods* 86:397–400
- Drepper T, Eggert T, Circolone F, Heck A, Krauß U, Guterl J-K, Wendorff M, Losi A, Gärtner W, Jaeger K-E (2007) Reporter proteins for in vivo fluorescence without oxygen. *Nat Biotechnol* 25:443–445
- Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD (2012) The great screen anomaly - a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* 93:1005–1020
- Feng Z, Wang L, Rajska SR, Xu Z, Coeffet-LeGal MF, Shen B (2009) Engineered production of iso-migrastatin in heterologous *Streptomyces* hosts. *Bioorg Med Chem* 17:2147–2153
- Fernández M, Martínez-Bueno M, Martín MC, Valdivia E, Maqueda M (2007) Heterologous expression of enterocin AS-48 in several strains of lactic acid bacteria. *J Appl Microbiol* 102:1350–1361
- Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, Nechitaylo TY, Guazzaroni M-E, Polaina J, Waliczek A, Chernikova TN, Reva ON, Golyshina OV, Golyshin PN (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
- Fischbach M, Voigt CA (2010) Prokaryotic gene clusters: a rich toolbox for synthetic biology. *Biotechnol J* 5:1277–1296
- Franken B, Jaeger K-E, Pietruszka J (2010) Screening for enantioselective enzymes. *Handbook of hydrocarbon and lipid microbiology*. Springer, New York, pp 2859–2876
- Fu J, Leiros H-KS, Pascale D, Johnson KA, Blencke H-M, Landfald B (2013) Functional and structural studies of a novel cold-adapted esterase from an Arctic intertidal metagenomic library. *Appl Microbiol Biotechnol* 97:3965–3978
- Furubayashi M, Ikezumi M, Kajiura J, Iwasaki M, Fujii A, Li L, Saito K, Umeno D (2014) A high-throughput colorimetric screening assay for terpene synthase activity based on substrate consumption. *PLoS ONE* 9:e93317
- Gee KR, Sun W-C, Bhalgat MK, Upson RH, Klaubert DH, Latham KA, Haugland RP (1999) Fluorogenic substrates based on fluorinated umbelliferones for continuous assays of phosphatases and β -galactosidases. *Anal Biochem* 273:41–48
- Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala HA, Rowland SE, Blanch HW, Clark DS, Robb FT (2011) Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat Commun* 2:375
- Greenberg WA, Varvak A, Hanson SR, Wong K, Huang H, Chen P, Burk MJ (2004) Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proc Natl Acad Sci U S A* 101:5788–5793
- Gregor J, Klug G (2002) Oxygen-regulated expression of genes for pigment binding proteins in *Rhodobacter capsulatus*. *J Mol Microbiol Biotechnol* 4:249–253
- Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* 57:441–466
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5:R245–249
- Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT (1995) 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
- Jaeger K-E, Kovacic F (2014) Determination of lipolytic enzyme activities. *Methods Mol Biol* 1149:111–134

- Jiang P, Wang H, Zhang C, Lou K, Xing X-H (2010) Reconstruction of the violacein biosynthetic pathway from *Duganella sp. B2* in different heterologous hosts. *Appl Microbiol Biotechnol* 86:1077–1088
- Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W, Liles MR (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene* 475:57–62
- Kaschner M, Loeschcke A, Krause J, Minhc BQ, Heck A Svensson V, Wirtz A, von Haeseler A, Jaeger K-E, Drepper T, Krauss U (2014) The first light-dependent protochlorophyllide oxidoreductase in anoxygenic phototrophic bacteria (in press)
- Katzke N, Arvani S, Bergmann R, Circolone F, Markert A, Svensson V, Jaeger K-E, Heck A, Drepper T (2010) A novel T7 RNA polymerase dependent expression system for high-level protein production in the phototrophic bacterium *Rhodobacter capsulatus*. *Protein Expr Purif* 69:137–146
- Katzke N, Jaeger K-E, Drepper T (2012) High-level gene expression in the photosynthetic bacterium *Rhodobacter capsulatus*. *Methods Mol Biol* 824:251–269
- Klein G, Reymond J-L (1999) Enantioselective fluorogenic assay of acetate hydrolysis for detecting lipase catalytic antibodies. *Helv Chim Acta* 82:400–407
- Kostner D, Peters B, Mientus M, Liebl W, Ehrenreich A (2013) Importance of *codB* for new codA-based markerless gene deletion in *Gluconobacter* strains. *Appl Microbiol Biotechnol* 97:8341–8349
- Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, Peterson KM (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166:175–176
- Leis B, Angelov A, Liebl W (2013) Screening and expression of genes from metagenomes. *Adv Appl Microbiol* 83:1–68
- Leroy E, Bensel N, Reymond J-L (2003) A low background high-throughput screening (HTS) fluorescence assay for lipases and esterases using acyloxymethylethers of umbelliferone. *Bioorg Med Chem Lett* 13:2105–2108
- Liebl W (2004) Genomics taken to the extreme. *Nat Biotechnol* 22:524–525
- Liebl W (2011) Metagenomics. In: Reitner J, Thiel V (eds) *Encyclopedia of geobiology*. Springer, Dordrecht, pp 553–558
- Loeschcke A, Markert A, Wilhelm S, Wirtz A, Rosenau F, Jaeger K-E, Drepper T (2013) TREX—a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synth Biol* 2:22–33
- Lorenz P, Eck J (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* 3:510–516
- Lorenz P, Liebeton K, Niehaus F, 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
- Lundstrom K (2007) Structural genomics and drug discovery. *J Cell Mol Med* 11:224–238
- Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA, Osburne MS (2004) 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
- Masepohl B, Hallenbeck PC (2010) Nitrogen and molybdenum control of nitrogen fixation in the phototrophic bacterium *Rhodobacter capsulatus*. *Adv Exp Med Biol* 675:49–70
- McMahon MD, Guan C, Handelsman J, Thomas MG (2012) Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression. *Appl Environ Microbiol* 78:3622–3629
- Nannenga BL, Baneyx F (2011) Reprogramming chaperone pathways to improve membrane protein expression in *Escherichia coli*. *Protein Sci* 20:1411–1420
- Neufeld K, zu Berstenhorst SM, Pietruszka J (2014) Evaluation of coumarin-based fluorogenic P450 BM3 substrates and prospects for competitive inhibition screenings. *Anal Biochem* 456:70–81
- Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 75:311–335
- Nimchua T, Thongaram T, Uengwetwanit T, Pongpattanakitsote S, Eurwilaichitr L (2012) Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *J Microbiol Biotechnol* 22:462–469
- Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol* 40:337–365
- Ongley SE, Bian X, Neilan BA, Müller R (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep* 30:1121–2138
- Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986) The analysis of natural microbial populations by ribosomal RNA sequences. *Adv Microb Ecol* 9:1–55
- Pérez Carlón R, Jourdain N, Reymond JL (2000) Fluorogenic polypropionate fragments for detecting stereoselective aldolases. *Chem Eur J* 6:4154–4162
- Peters B, Junker A, Brauer K, Mühlthaler B, Kostner D, Mientus M, Liebl W, Ehrenreich A (2013a) Deletion of pyruvate decarboxylase by a new method for efficient markerless gene deletions in *Gluconobacter oxydans*. *Appl Microbiol Biotechnol* 97:2521–2530
- Peters B, Mientus M, Kostner D, Liebl W, Ehrenreich A (2013b) Characterization of membrane-bound dehydrogenases from *Gluconobacter oxydans* 621H via whole-cell activity assays using multi-deletion strains. *Appl Microbiol Biotechnol* 97:6397–6412
- Reddy GS, Matsumoto GI, Schumann P, Stackebrandt E, Shivaji S (2004) Psychrophilic pseudomonads from Antarctica: *Pseudomonas antarctica* sp. nov., *Pseudomonas meridiana* sp. nov. and *Pseudomonas proteolytica* sp. nov. *Int J Syst Evol Microbiol* 54:713–719
- Reymond JL (2006) *Enzyme assays*. Wiley, New York
- Reymond JL (2009) Colorimetric and fluorescence-based screening. *Protein engineering handbook*, Volume 1 & Volume 2, pp. 669–711
- Ruff AJ, Dennig A, Wirtz G, Blanus M, Schwaneberg U (2012) Flow cytometer-based high-throughput screening system for accelerated directed evolution of P450 monooxygenases. *ACS Catal* 2:2724–2728
- Schlegel S, Klepsch M, Gialama D, Wickström D, Slotboom DJ, de Gier JW (2010) Revolutionizing membrane protein overexpression in bacteria. *Microb Biotechnol* 3:403–411
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* 21:1794–1805
- Simon C, Daniel R (2011) Metagenomic analyses: past and future trends. *Appl Environ Microbiol* 77:1153–1161
- Steele HL, Jaeger K-E, Daniel R, Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* 16:25–37
- Stevens DC, Hari TP, Boddy CN (2013) The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat Prod Rep* 30:1391–1411
- Streit WR, Schmitz RA (2004) Metagenomics—the key to the uncultured microbes. *Curr Opin Microbiol* 7:492–498
- Taupp M, Mewis K, Hallam SJ (2011) The art and design of functional metagenomic screens. *Curr Opin Biotechnol* 22:465–472
- Thomas T, Gilbert J, Meyer F (2012) Metagenomics—a guide from sampling to data analysis. *Microb Inform Exp* 2:3
- Tichi MA, Tabita FR (2002) Metabolic signals that lead to control of CBB gene expression in *Rhodobacter capsulatus*. *J Bacteriol* 184:1905–1915

- Topakas E, Christakopoulos P (2014) Screening and purification of recombinant lignocellulolytic enzymes. *Methods Mol Biol* 1129: 517–526
- Torsvik VL, Goksøyr J (1978) Determination of bacterial DNA in soil. *Soil Biol Biochem* 10:7–12
- Troeschel SC, Drepper T, Leggewie C, Streit WR, Jaeger KE (2010) Novel tools for the functional expression of metagenomic DNA. *Methods Mol Biol* 668:117–139
- Wagner S, Bader ML, Drew D, de Gier JW (2006) Rationalizing membrane protein overexpression. *Trends Biotechnol* 24:364–371
- Wahler D, Badalassi F, Crotti P, Reymond JL (2001) Enzyme fingerprints by fluorogenic and chromogenic substrate arrays. *Angew Chem* 113:4589–4592
- Wang G, Meng K, Luo H, Wang Y, Huang H, Shi P, Yang P, Zhang Z, Yao B, Tang H (2012) Phylogenetic diversity and environment-specific distributions of glycosyl hydrolase family 10 xylanases in geographically distant soils. *PLoS ONE* 7:e43480
- Wang H, Yang L, Wu K, Guanghui L (2014) Rational selection and engineering of exogenous principal sigma factor (σ^{HrdB}) to increase teicoplanin production in an industrial strain of *Actinoplanes teichomyeticus*. *Microb Cell Fact* 13:10
- Wooley JC, Godzik A, Friedberg I (2010) A primer on metagenomics. *PLoS Comput Biol* 6:e1000667
- Wösten MMSM (1998) Eubacterial sigma-factors. *FEMS Microbiol Rev* 22:127–150
- Wu J, Bauer CE (2008) RegB/RegA, a global redox-responding two-component system. *Adv Exp Med Biol* 631:131–148
- Yang D, Zhu X, Wu X, Feng Z, Huang L, Shen B, Xu Z (2011) Titer improvement of iso-migrastatin in selected heterologous *Streptomyces* hosts and related analysis of mRNA expression by quantitative RT-PCR. *Appl Microbiol Biotechnol* 89:1709–1719
- Yu H, Tyo K, Alper H, Klein-Marcuschamer D, Stephanopoulos G (2008) A high-throughput screen for hyaluronic acid accumulation in recombinant *Escherichia coli* transformed by libraries of engineered sigma factors. *Biotechnol Bioeng* 101:788–96
- Zhang H, Boghigian B, Armando J, Pfeifer B (2011) Methods and options for the heterologous production of complex natural products. *Nat Prod Rep* 28:125–151