The evolving interface between synthetic biology and functional metagenomics

Eric van der Helm1, Hans J. Genee2 and Morten O. A. Sommer1*

Nature is a diverse and rich source of bioactive pathways or novel building blocks for synthetic biology. In this Perspective, we describe the emerging research field in which metagenomes are functionally interrogated using synthetic biology. This approach substantially expands the set of identified biological activities and building blocks. In reviewing this field, we find that its potential for new biological discovery is dramatically increasing. Functional metagenomic mining using genetic circuits has led to the discovery of novel bioactivity such as amidases, NF-κB modulators, naphthalene degrading enzymes, cellulases, lipases and transporters. Using these genetic circuits as a template, improvements are made by designing biosensors, such as in vitro-evolved riboswitches and computationally redesigned transcription factors. Thus, powered by the rapidly expanding repertoire of biosensors and streamlined processes for automated genetic circuit design, a greater variety of complex selection circuits can be built, with resulting impacts on drug discovery and industrial biotechnology.

Nature is composed of a rich and diverse set of microbiomes that can be mined for biological activity and novel building blocks for synthetic biology. It is estimated that there are 10^{30} microbial cells on earth1 and approximately 10^2–10^5 microbial species in 1 g of soil2. However, a majority of these organisms remain recalcitrant to culturing under common laboratory conditions3–5. For harnessing the biological activities and building blocks of this difficult-to-grow biodiversity, culture-independent methods are advantageous. Advances in metagenomic sequence analysis have led to the discovery of novel biomass-degrading enzymes from cow rumen6, the identification of new CRISPR systems7, and the establishment of a reference gene catalog of the human microbiome8, which are reviewed elsewhere9. In this Perspective, we focus on recent developments within functional metagenomics whereby environmental DNA is directly cloned into bacterial strains, enabling phenotypic characterization (Fig. 1)10. In particular, we focus on the emerging research field in which functional metagenomic libraries are interrogated using advanced genetic circuits, substantially expanding the set of biological activities and building blocks that can be identified.

Synthetic biology has developed to a point where complex biological circuits are designed to control specific cellular programs in response to specific inputs with reasonable predictability11. Substantial research has focused on designing suitable input sensors, defined as components that can sense an input and relay this information downstream; examples include transcription factors12, riboswitches13 and protein-based sensors14. The input signal(s) can be propagated and coupled to an output such as fluorescence or even survival under selective conditions.

In this Perspective we analyze the current state of the rapidly evolving interface between synthetic biology and functional metagenomics. We highlight two general requirements for successful functional metagenomic mining using synthetic biology: (1) genetic circuits need to be rationally designed, and (2) genetic circuits need to enable a high-throughput workflow. Emerging solutions to these requirements are supplied by computational protein design and genetic circuit design automation, which have been substantially improved within the last few years. Accordingly, the potential utility and impact of this emerging field has dramatically increased.

Interrogation of metagenomes using phenotypic screens

Early implementation of functional metagenomics was based on classic molecular biology and relied on screens with a phenotypic readout, such as halo formation and changes in colony color or morphology, in order to detect specific activities. Such screening has been employed to discover hemolytic15, lipolytic16 and antimicrobial activities17–20. Phenotypic screening of metagenomic libraries remains useful, but the approach is normally limited to a few biological activities and rarely includes anabolic activities.

Biological activities may also be identified using complementation assays that circumvent screening by knocking out one or more essential host gene and then using a functional metagenomic library to rescue this deficiency. An example is the biotin-auxotrophic Escherichia coli strain ΔbiouvB, which was used to identify genes from the environmental metagenome that are involved in biotin synthesis21. Using a similar approach, novel DNA polymerases in an E. coli polA mutant22 were identified. More recently, a phosphopantetheine transferase knockout in E. coli enabled enrichment of nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) genes from a metagenomic library23.

In a manner similar to that of a complementation assay, genes conferring resistance toward a given compound can be identified from the metagenome by exposing a metagenomic library to inhibitory concentrations of a particular toxin. This approach has been used to discover antibiotic resistance genes in soil24, cow manure25 and human microflora26, as well as to discover salt25,26 and acid tolerance genes27, CO₂-fixing enzymes28 and benzoylformate decarboxylases29. This strategy has also been used to identify genes that confer resistance to biomass-derived inhibitors, potentially enabling the construction of strains with increased robustness for the fermentation of lignocellulose-derived sugar streams30,31. Though powerful, such classical functional metagenomic methods suffer from limitations in the spectrum of phenotypes that can be assessed.

1Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark. 2Biosyntia ApS, Copenhagen, Denmark.
*e-mail: msom@bio.dtu.dk

NATURE CHEMICAL BIOLOGY | www.nature.com/naturechemicalbiology
Metagenome mining for catabolic pathway components: SIGEX

Biological engineering was first applied to screen a metagenomic library in 2005 (ref. 32). The approach made use of the fact that catabolic gene expression is often regulated by regulatory elements in proximity to the catabolic genes. The method, substrate-induced gene expression screening (SIGEX), relies on an operon trap that employs green fluorescent protein (GFP) as a fluorescence readout to identify catabolic pathways (Fig. 2a). In the SIGEX workflow, metagenomic DNA is shotgun cloned upstream of gfp. The library is then incubated with the substrate of interest, leading to the induction of the operon and thus resulting in a fluorescence signal. Using SIGEX, a metagenomic library of 152,000 clones with an average size of 7 kb was screened, and 4 clones that are regulated by naphthalene and 58 clones that are regulated by benzoate were identified. The clones consisted largely of catabolic genes, transcription regulators, and transporters, as expected based on the spatial distribution of operons. The SIGEX method has been applied to identify components that are regulated by aromatic compounds34, constitutive promoters35 and naringenin-inducible promoters36. Furthermore, by shotgun sequencing the initial environmental DNA, contextual information on the original pathways was recovered37. Though powerful, the SIGEX method is subject to the following limitations38: (1) transcriptional regulators may not be in spatial proximity to the catabolic genes, (2) catabolic genes may be constitutively expressed instead of being regulated by a transcriptional regulator, and (3) transcriptional regulators may be modulated by a substrate that is not part of the pathway that they regulate.

Detecting pathways that induce quorum sensing

To identify metagenomic fragments that inhibit or induce quorum sensing, the metabolite-regulated expression (METREX; Fig. 2b) method was developed38. METREX consists of a genetic circuit with a quorum-sensing promoter that is activated by an acylhomoserine lactone or another small molecule, resulting in expression of the reporter gene gfp. By screening eight metagenomic libraries that, in total, contained 53,000 clones with inserts ranging from 1 to 190 kb in size, 11 clones were identified that modulate the output of the circuit (gfp). One of these clones harbored a luxl homolog (62% sequence homology to AamfI from Pseudomonas fluorescens). The LuxI protein is involved in the synthesis of an N-acyl homoserine lactone (AHL) quorum-sensing signal molecule. METREX was subsequently used to identify AHL signal-molecule-modulating pathways in gypsy moth gut microbiotas39 as well to screen for AHL compounds in forest soil and activated sludge from a coke plant40. The METREX method was modified to identify genes that modulate quorum sensing by integrating the luxR homolog traR and luxI homolog traI into the genetic circuit (Fig. 2c)41. Three metagenomic clones were identified by co-culturing E. coli (harboring the metagenomic library) and the plant pathogen Agrobacterium tumefaciens (harboring the genetic circuit). Introduction of the metagenomic DNA into Pseudomonas aeruginosa led to reduced biofilm formation caused by the lactone-degrading activity derived from the metagenomic fragment.

Product-induced mining of metagenomic libraries

Amidases are an industrially relevant class of enzymes that catalyze the hydrolysis of an amide bond. Product-induced gene expression (PIGEX) was introduced to discover novel amidases in a sequence-independent manner42. Although the name implies similarities to the aforementioned SIGEX method, its mechanism is conceptually different from that of SIGEX (Fig. 2d) and similar to that of METREX40 (Fig. 2b). The PIGEX system relies on induction of a preselected transcription regulator (for example, encoded by benR) that can be activated by the product of interest (for example, benzoate) to result in expression of GFP. A 96,000-membered E. coli fosmid library was co-cultivated with benzamide (a substrate for amidases) and the sensor E. coli strain that contained the benzoate-responsive circuit. After two rounds of screening in 96-well plates, the authors found eleven amidase genes, of which three were novel and contained no sequence homology to known bacterial amidases. The transcription regulator and cognate operator sequence can readily be replaced to screen for a wide range of enzymatic activities (provided that a transcriptional repressor for the product molecule is known).

Inflammatory response modulators from the gut microbiota

The transcription factor NF-kB is involved in host immune and inflammatory response, and the gut microbiota is known to modulate NF-kB. To identify microbial factors that modulate NF-kB, a reporter cell line in which a secreted alkaline phosphatase is controlled by NF-kB was constructed (Fig. 2e). Factors that modulate NF-kB would trigger expression of alkaline phosphatase, which can be readily assayed using a chemiluminescent substrate. Cell lysate from 2,460 individual metagenomic E. coli fosmid clones was added to this reporter cell line43. Using this method, an ABC-transport system and a putative lipoprotein were identified as potential modulators of NF-kB. In 2014 this protocol was further optimized for high-throughput screening44.

Subsequently, the circuit was modified to express GFP instead of secreted alkaline phosphatase45. This alteration enabled the identification of commendamide, an NF-kB-modulating compound...
Fig. 2 | Genetic circuits used to mine metagenomic libraries. a, Schematic representation of the SIGEX procedure\(^3\). The platform plasmid is digested by BamHI and ligated with metagenomic DNA. Clones that self-ligate or contain a short insert are sorted out in a preround of FACS by inducing the Plac promoter with IPTG, resulting in a fluorescence signal for the false positives. Next, the cells are incubated with a substrate of interest, thereby activating transcription of potential catabolic gene(s). In the original SIGEX study\(^3\), benzoate served as this substrate of interest. b, The quorum-sensing genetic circuit used in the METREX strategy\(^5\) is based on the luxR-encoded transcriptional activator. Upon addition of a metagenomic fragment that expresses a gene of interest (GOI), a signal molecule is formed. This signal molecule activates the LuxR transcriptional regulator, resulting in gfp transcription. c, Genetic circuit used in an A. tumefaciens reporter strain\(^6\). The reporter lacZ is fused on the transcriptional level to the traR-traI operon. Upon induction by a lactone signal molecule, the P_{tra} promoter is activated, and the lacZ gene is expressed. When a lactonase is expressed from a metagenomic library, the signal molecule is degraded, and the reporter lacZ gene is not expressed. d, The PIGEX genetic circuit used to identify amidases in a metagenomic library\(^4\). The transcription regulator, encoded by benR, responds to the presence of benzoate by activating transcription from the P_{benA} promoter. In the PIGEX circuit, P_{benA} drives the transcription of gfp, thus resulting in fluorescence when benzoate is formed. e, Genetic circuit in HT-29 cells\(^4\); the NF-κB promoter is activated when the NF-κB complex responds to a signal molecule, resulting in expression of secreted alkaline phosphatase (SEAP), a reporter gene. The signal molecule is generated by an expressed gene or pathway encoded in the metagenomic library. f, Genetic circuit in Acinetobacter baylyi that responds to the presence of salicylate by activating the SalR transcription regulator protein\(^5\). The activated SalR protein in turn initiates transcription of the luxCDABE reporter from the Psal promoter, resulting in fluorescence. g, Genetic circuit in E. coli that responds to the presence of a compound such as phenol by activating the transcriptional regulator protein DmpR from P. putida\(^6\). The activated transcriptional regulator DmpR subsequently binds the P_{tra} promoter (also from P. putida) to initiate transcription of the reporter gfp. Phenolic compounds are converted from a substrate with an added phenyl tag by a metagenomic library-encoded gene or pathway. The terminators flanking dmpR and gfp genes and the ribosome-binding site upstream of gfp have been engineered to increase the dynamic range of the circuit and were not part of the wild-type operon in P. putida. h, The ThIM19 riboswitch occludes the ribosome binding site (RBS) from translation initiation in the absence of thiamine pyrophosphate (TPP)\(^6\). Upon the addition of TPP, the riboswitch changes confirmation and facilitates recognition of the RBS, resulting in translation of the resistance genes cat and aadA. To reduce the number of false-positive hits, the circuit has a redundant layout in which two individual riboswitches control the translation of the resistance genes.
secreted by a metagenomic clone. Further work showed that this N-acyl-amide acts as a mammalian signaling molecule by activating G-protein-coupled receptors (GPCRs). Inspired by these findings a bioinformatic analysis was carried out to identify genes in the human microbiota that could encode GPCR-active N-acyl syntheses. A commensal GPR119 agonist was identified that has structural mimicry of human GPCR ligands and has also been shown to regulate the metabolic and glucose homeostasis as efficiently. These studies provide an interesting case for applying functional metagenomics to identify an effect of human–microbe interactions. While much work in the microbiome field currently relies on sequence-based metagenomics, the use of such synthetic biology–powered functional metagenomics can yield new bioactive molecules that might be missed by searches based on sequence homology.

Detection of biodegradation pathways using DNA enrichment

Naphthalene contamination in ground water can be remediated by naphthalene-metabolizing organisms. To identify active naphthalene degraders, [13C]naphthalene was incubated with a contaminated groundwater sample, and the [13C]-incorporated DNA was subsequently isolated using ultracentrifugation. After the DNA was sequenced and annotated, potential naphthalene-degrading operons were identified. These operons were amplified using a polymerase chain reaction (PCR), and the PCR products were cloned into an expression vector. The resulting plasmids were transformed into a strain containing a salicylate (an intermediate metabolite in naphthalene catabolism) biosensor coupled to luciferase (Fig. 2f). Screening with this system enabled the discovery of a novel nag operon that is responsible for naphthalene biodegradation and highlighted the power of combining an enrichment strategy with a genetic circuit.

Generalized genetic circuits detect enzymatic byproducts

Most of the previous examples are based on visual screens that are carried out on solid media or microtiter plates, which limits the throughput of the screen. However the throughput can be increased by using fluorescence-activated cell sorting (FACS) to determine the circuit’s state. As over 300 enzyme species have been reported to generate phenols or phenyl compounds as byproducts of their reactions, it was hypothesized that sensing these byproducts would provide a general platform for screening enzyme activity. Accordingly, a genetic circuit (Fig. 2g) that was read out using FACS was used to find enzymes, including cellulases, lipases or alkaline phosphatases, that degrade phenyl-tagged compounds encoded in a metagenomic library. The general circuits consist of the dmpR gene from Pseudomonas putida whose gene product responds to the presence of phenolic compounds, resulting in a change in GFP fluorescence output. Using this method, termed the genetic enzyme screening system (GESS), a ~20,000-membered fosmid library was screened for enzymes that degrade phenyl-tagged compounds. This strategy resulted in the identification of a novel phosphatase gene that shares 59% homology with a Spingomonas alkaline phosphatase. Further characterization of this alkaline phosphatase identified its thermolabile properties, making it a potentially useful tool in molecular biology.

Interestingly, this study highlights the need to optimize genetic circuits before employing metagenomic screens. In this case, two terminators and a strong E. coli ribosome-binding site were inserted in a circuit (Fig. 2g) that originated from P. putida. This refactoring resulted in an ~11-fold more sensitive circuit than the circuit previously constructed by the same group; the need for refactoring has also been observed in the optimization process of other biosensors.

Dual-selection circuit to select for novel transporters

The functions of bacterial transporters are poorly annotated, and to discover novel thiamine and xanthine uptake transporters, we constructed a synthetic selection circuit to mine metagenomic libraries. Placing the thiamine pyrophosphate (TPP, the biologically active form of thiamine)-responsive ThiM19 riboswitch upstream of an antibiotic resistance cassette created a synthetic selection circuit that renders the growth of the selected strain dependent on an excess of intracellular TPP (Fig. 2h). By applying this circuit to a metagenomic library, 26 phylogenetic distant members of a novel class of thiamine transporters (encoded by pnuT) were discovered and functionally validated. Improving the robustness to decrease false-positive colony formation was instrumental to this success. This robustness was achieved by introducing redundancy into the circuit; it contains chloramphenicol and spectinomycin resistance genes, both with an upstream ThiM19 riboswitch. The generality of this method was further demonstrated by deploying a fully synthetic riboswitch that was responsive to xanthine alkaloids to functionally mine xanthine importers from metagenomes. Continued application of this method should permit deep mining of transporters from entire microbial communities and allow exhaustive elucidation of transporters in individual organisms.

Expanding the repertoire of new biosensors

Tailoring input systems to respond to a compound of interest requires a compound-specific biosensor that can be functionally implemented in a genetic circuit. Availability of relevant biosensors is therefore critical for the use of such genetic systems in metagenomic mining. In nature, gene expression is often regulated in response to intracellular metabolite pools, and accordingly, most biosensors are derived from natural systems (see Table 1). More than 17 small-molecule classes can be detected by riboswitches, and natural allosteric transcription factors have been identified for 243 compound classes. This broad palette of compounds includes vitamins, amino acids and antibiotics, and recent studies suggest that many more riboswitches are yet to be found. Because natural biosensors have evolved in vivo, they respond to physiologically relevant concentration ranges and display high specificity toward their target ligands. These properties make them ideal candidates for in vivo applications and functional metagenomics.

---

**Table 1 | Application of biosensors in the context of high-throughput functional metagenomic screens**

<table>
<thead>
<tr>
<th>Type of biosensor</th>
<th>Examples</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription-based</td>
<td>Transcription factor</td>
<td>Can be used modularly</td>
<td>Slow response, relies on gene expression</td>
</tr>
<tr>
<td>Translation-based</td>
<td>Riboswitch, ribozyme</td>
<td>Natural riboswitches can have high sensitivity and specificity</td>
<td>Not broadly applicable as design of novel riboswitches is challenging</td>
</tr>
<tr>
<td>Post-translational-based</td>
<td>Fluorescence, FRET, protease</td>
<td>Fast response</td>
<td>Absolute signals often low</td>
</tr>
</tbody>
</table>

Input sensors can be classified into three categories: (1) transcription-based, i.e., a regulatory protein or RNA aptamer interacts with a stimuli of interest (for example, a small molecule), which results in a modulation of transcriptional activity; (2) translation-based, i.e., riboswitches and aptamers that upon compound binding interfere either in cis or in trans with translation by acting upon translation initiation, mRNA stability or mRNA splicing; and (3) post-translational-based sensors in which a reporter protein generates a signal upon interaction with a molecule of interest.
When no biosensor is available for a compound of interest, biosensors can be synthetically created. Ligand-binding domains of riboswitches can be selected from large RNA pools in vitro, and selected aptamers can be integrated into expression platforms to enable small-molecule-responsive genetic circuits (Fig. 3a). This approach has led to the creation of fully synthetic riboswitches for a handful of compounds. Furthermore, progress has been made by using a statistical thermodynamic model that predicts the sequence–structure–function relationship: 62 synthetic riboswitches were computationally designed to sense a set of diverse chemicals. Using another de novo design approach, synthetic aptamers were fused to a designed transcription termination element, resulting in a tetracycline-dependent riboswitch modulating downstream transcription. A fully de novo design remains absent for transcription factors, but substantial progress is being made in repurposing the specificity of existing transcription factors. Such efforts include pipelines that can broaden availability using computational design. As an example, bacterial allosteric transcription factors were redesigned using the Rosetta framework (Fig. 3b). The recognition profile of the LacI transcription factor from *E. coli* was altered using single-residue saturation mutagenesis and random mutagenesis to create mutant proteins that recognize fucose, gentiobiase, lactitol and sucralose. Another method to broaden the range of available biosensors relies on ligand-binding domains that contain engineered destabilizing mutations. Ligand-binding domain accumulation only occurs in cells in which the target ligand is also present; in other cells, the sensor is degraded by the ubiquitin proteasome system. The ligand-binding domain was fused in-frame with a fluorescent domain or transcription activator to establish a readout. The previously constructed digoxin-binding domain was redesigned to be conditionally stable upon ligand binding and fused to either *gfp* or a transcriptional activator domain. When the binding domain is coupled to the transcriptional activator domain, activation by the ligand resulted in a 100-fold increase in fluorescence.

Another strategy to expand the repertoire of biosensors is to extend the metabolic network from compounds for which no biosensor is available to molecules for which a biosensor is available. The minimum amount of enzymes that are necessary to do so can be identified using a database containing 9,319 biochemical reaction rules. Equipping *E. coli* with these pathways and sensors resulted in strains that fluorescently respond to cocaine, para-thion, 2-chloro-4-nitrophenol, hippuric acid, and nitroglycerin. Considering the available enzymes and the known biosensors, this strategy should substantially expand the number of compounds that can be sensed. These examples highlight the fact that novel transcription factors and circuits can be constructed, thus substantially widening the scope of functional metagenomic mining.

**Signal processing: automation and improving robustness** Current circuits used to interrogate metagenomic libraries are based only on a single input signal. However, to advance selection methods multiplexing more than one input signal can be advantageous; for example, the conversion facilitated by a metagenomic gene of substrate to product could be wired using an N-IMPLY circuit. Such a circuit has been built previously to detect the presence of the small molecule erythromycin only in the absence of philoretin. To combine more signals (for example, to detect multiple pathway intermediates) or construct more complex conditions, more elaborate genetic circuit design tools are required. Recently, a tremendous advance was reported in the field of automated genetic circuit design (Fig. 4). Using the aforementioned standardized parts and isolators and with the help of a new software, the authors designed 60 genetic programs that were automatically converted into logic gates and then finally into DNA. Of the 60 designed circuits, 45 performed according to their initial design. These results are encouraging, as they exemplify the power of genetic circuit design. However, it should be noted that the construction of these circuits required extensive characterization of the used parts.

Most of the previously discussed examples are screening approaches, not for selection (except for the thiamine and xanthine transport mining circuits shown in Fig. 2b). Several sources have stated that a more efficient readout is required to advance biosensor genetic circuits. Although selection systems present a possible solution from an engineering perspective, surmounting the evolutionary properties of biological systems is a major challenge when engineering selection systems. One approach to overcoming such evolutionary decay of the selection circuits is to introduce redundancy. Redundancy can be achieved by having more than one independent selection system operating simultaneously in the cell. This approach was used to reduce the false-positive levels in the thiamine selection system previously developed by our group. from 10^8 in 10^7 cells to less than 1 in 10^7 cells. Another example of the necessity to combat the effects of evolutionary pressure on genetic circuits became apparent upon the optimization of the naringenin pathway using multiplex automated genome engineering (MAGE). To increase the throughput of the genome engineering process, the researchers constructed a genetic circuit that relied on two copies of the regulator gene *ttgR*. The second copy of the regulator caused the escape rate to drop from 10^-5 CFU/total cells to 10^-4 CFU/total cells. The authors hypothesized that this change occurs because the selective pressure in the case with two copies can only be released by incorporating inactivating mutations in both of the *ttgR* genes, compared to the ease of introducing a single mutation in one copy of the gene. In spite of these solutions, the challenges of evolutionary escape must be assessed in the context of...
each specific circuit: the mode of failure depends substantially on the logic of the circuit, as highlighted in recent examples from the field of biocontainment. These modes include point mutations, deletions and insertion sequence (IS) elements causing the circuit to fail. However as a general rule these can be mitigated by introducing redundancy into the genetic circuit, for example by introducing two copies of a selection marker. These advances should be used in the design of synthetic biology–enabled functional metagenomic selections, as they will dramatically reduce circuit failure and thus lower the false-positive rate of the selections.

Considerations when applying functional metagenomics

When designing genetic circuits to mine metagenomic libraries it is important to consider the fundamental limitations of metagenomic expression libraries. Indeed, functional expression of metagenomic DNA is dependent on (i) propagation of metagenomic DNA in the heterologous host, (ii) transcription of the metagenomic DNA by either a vector-borne promoter or recognition of heterologous regulatory elements, and (iii) efficient translation of mRNA and possibly post-translational modification into the functional protein.

Maintenance of metagenomic DNA is routinely performed on plasmids, cosmid or fosmids, but the choice depends on the bioactivity under investigation. For example, antibiotic resistance genes are often contained within a single open reading frame (ORF). Antibiotic resistance genes can therefore be captured with a metagenomic library insert size of 1–3 kb (ref. 57) and are easily maintained on a plasmid backbone. However, the DNA found when functionally selecting for secondary metabolite clusters can be propagated only rarely on a plasmid, as the size of these clusters can range from 5 (ref. 43) to 150 kb (ref. 82). A suitable way to capture and maintain larger fragments is by using cosmids or fosmids, which allow the construction of 25–40-kb inserts. It is therefore key when performing functional synthetic metagenomic selections to determine the most suitable vector for finding the desired activity.

A bottleneck in functional metagenomics is the recognition of heterologous DNA by the host machinery enabling efficient transcription and translation. This recognition can be hampered by, for example, codon bias or a lack of regulatory elements. Initial expression efforts began by using E. coli as a heterologous metagenomic expression host, which is estimated to express only 40% of heterologous proteins from a randomly cloned metagenomic library. Over the past decade, more expression hosts were transformed with the same metagenomic library, that the recognition of heterologous regulatory elements (promoter elements in particular) is a key factor in determining expression. Therefore, small insert libraries are very suitable for activity screening, as transcription of the insert can be driven by a vector-borne promoter. In larger insert libraries, transcription is hampered because of limited recognition of heterologous promoter regions. In this context, it was hypothesized that overexpressing one of the six endogenous sigma factors of E. coli (not including housekeeping sigma factor 70) would increase transcription from metagenomic DNA. Indeed, when the cells overexpressed E. coli sigma factor 54, an introduced heterologous gene cluster produced oxytetracycline.

This approach was taken a step further by testing seven sigma factors from a range of organisms to initiate transcription of a metagenomic GFP trap. The trap was constructed by cloning short soil metagenomic inserts in front of a promoterless gfp gene. GFP fluorescence could be measured only when a promoter element on the insert was recognized by an (endogenous) sigma factor. Using this technique, the rpoD sigma factor from Lactobacillus plantarum was found to increase transcription in a soil library by a factor of nine. Another recent development is based on random, transposon-mediated insertion of bidirectional T7 promoters into a metagenomic fosmid library. This ‘forced transcription’ method resulted in a 6.6-fold increase in positive hits. In another example, tunable gene expression across various Gram-positive and Gram-negative bacteria was achieved by constructing a feedback loop that self-regulates transcription of an orthogonal polymerase that is not driven by host-specific promoters.

Challenges due to codon bias, hampering translation, can be solved by using high-throughput DNA synthesis technology, for example, by matching codon frequencies as near as possible to the native expression system.

These examples show that although expression of metagenomic DNA can be challenging, the expression of metagenomic DNA can be improved by making use of a broader range of hosts or by equipping a host with an extended transcriptional toolbox.

Future perspectives

The results of using synthetic biology in functional metagenomics have been encouraging, and this approach has already led to the discovery of several novel bioactivities. With new approaches for expanding the repertoire of biosensors and increasing the robustness and predictability of genetic circuits, the field is expected to substantially grow. In particular, the field will benefit from the application of computational design to expand the substrate spectrum that can be detected by biosensors. Furthermore, the
automatization of genetic circuit design will aid in the construction of a greater variety of complex selection circuits that can be used in drug discovery and industrial biotechnology. Finally, implementing lessons learned from constructing robust, fail-safe circuits that have been applied to the area of biocontainment will increase the power of the selection systems used to mine metagenomics. Combining these developments is expected to accelerate the speed and output of functional metagenomics.

Received: 14 September 2017; Accepted: 13 June 2018; Published online: 16 July 2018

References


The authors used an automated process to design and construct 60 distinct genetic circuits of which 45 were experimentally validated to show the correct function.

Eggeling, L., Bott, M. & Marienhagen, J. Novel screening methods–automatization of genetic circuit design will aid in the construction of a greater variety of complex selection circuits that can be used in drug discovery and industrial biotechnology. Finally, implementing lessons learned from constructing robust, fail-safe circuits that have been applied to the area of biocontainment will increase the power of the selection systems used to mine metagenomics. Combining these developments is expected to accelerate the speed and output of functional metagenomics.
This study is the first to couple a genetic circuit to mine metagenomic library to high-throughput FACS screening.

The study is the first report of applying functional selection (instead of screening) using a genetic circuit to mine metagenomic libraries for bioactivity in a high-throughput format.

By coupling the Rosetta framework to redesign bacterial allosteric transcription factors with a genetic circuit to interrogate the success of a design, this paper reports a promising method to expand the repertoire of biosensors.

This study is neutral with regard to jurisdictional claims in published research. Correspondence is available at www.nature.com/reprints.

Additional information

Acknowledgements

This research was funded by the Novo Nordisk Foundation and the European Union Seventh Framework Programme (FP7/KBRE-2013-7-single-stage) under grant agreement no. 613745. Promys. M.O.A.S. acknowledges additional funding from The Lundbeck Foundation. E.v.d.H. acknowledges funding from the EU FP7–People-2012–ITN BacTory (317058).

Competing interests

H.G. and M.O.A.S. are co-founders of Biosynxia, with commercial interest in the topic of the Perspective.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to M.O.A.S.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.