nature chemical biology

PERSPECTIVE

https://doi.org/10.1038/s41589-018-0100-x

The evolving interface between synthetic biology and functional metagenomics

Eric van der Helm[1], Hans J. Genee[2] and Morten O. A. Sommer[1*]

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 earth[1] and approximately 10^3–10^5 microbial species in 1 g of soil[2]. However, a majority of these organisms remain recalcitrant to culturing under common laboratory conditions[3–5]. 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 rumen[6], the identification of new CRISPR systems[7], and the establishment of a reference gene catalog of the human microbiome[8], which are reviewed elsewhere[9]. 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 predictability[11]. 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 factors[12], riboswitches[13] and protein-based sensors[12]. 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 hemolytic[14], lipolytic[15] and antimicrobial activities[16–18]. 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 Δ*biouvrB*, which was used to identify genes from the environmental metagenome that are involved in biotin synthesis[19]. Using a similar approach, novel DNA polymerases in an *E. coli polA* mutant[20] 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 library[21].

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 soil[22], cow manure[23] and human microflora[24], as well as to discover salt[25,26] and acid tolerance genes[27], CO_2-fixing enzymes[28] and benzoylformate decarboxylases[29]. 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 streams[30,31]. Though powerful, such classical functional metagenomic methods suffer from limitations in the spectrum of phenotypes that can be assessed.

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

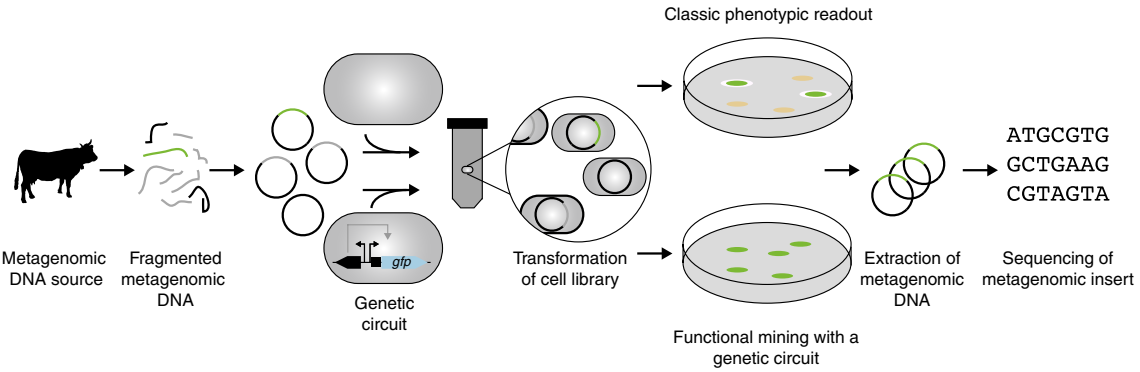

Fig. 1 | Construction of a metagenomic library from sample to sequence. Environmental DNA is extracted, purified, fragmented and cloned into an expression vector. The library of plasmids is then transformed in an expression host, for example, *E. coli*. Finally, the resulting clones can be analyzed, by either genotype or phenotype. The presence of a genetic circuit in the expression host allows a high-throughput interrogation of the metagenomic library.

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 compounds[33], constitutive promoters[34] and naringenin-inducible promoters[35]. Furthermore, by shotgun sequencing the initial environmental DNA, contextual information on the original pathways was recovered[36]. Though powerful, the SIGEX method is subject to the following limitations[37]: (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 developed[38]. 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 *luxI* 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 microbiotas[39] as well to screen for AHL compounds in forest soil and activated sludge from a coke plant[40]. 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 manner[42]. Although the name implies similarities to the aforementioned SIGEX[32] method, its mechanism is conceptually different from that of SIGEX (Fig. 2d) and similar to that of METREX[38] (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-κB is involved in host immune and inflammatory response, and the gut microbiota is known to modulate NF-κB. To identify microbial factors that modulate NF-κB, a reporter cell line in which a secreted alkaline phosphatase is controlled by NF-κB was constructed (Fig. 2e). Factors that modulate NF-κB 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 line[43]. Using this method, an ABC-transport system and a putative lipoprotein were identified as potential modulators of NF-κB. In 2014 this protocol was further optimized for high-throughput screening[44].

Subsequently, the circuit was modified to express GFP instead of secreted alkaline phosphatase[45]. This alteration enabled the identification of commendamide, an NF-κB-modulating compound

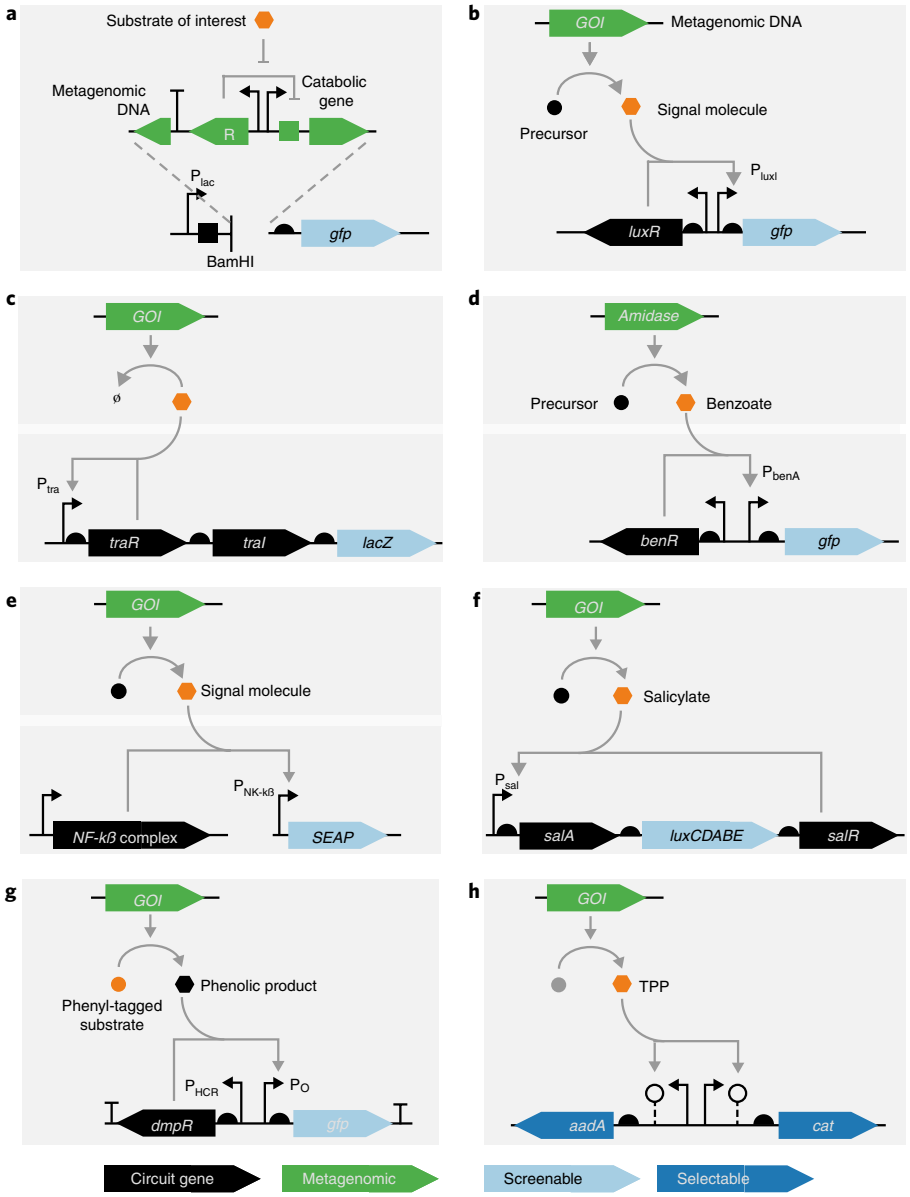

Fig. 2 | Genetic circuits used to mine metagenomic libraries. a, Schematic representation of the SIGEX procedure[32]. The platform plasmid is digested by BamHI and ligated with metagenomic DNA. Clones that self-ligate or contain a short insert are sorted out in a preround of FACS by inducing the P_{lac} 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[32], benzoate served as this substrate of interest. **b**, The quorum-sensing genetic circuit used in the METREX strategy[38] 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[41]. 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[42]. 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[43]; 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[50]. The activated SalR protein in turn initiates transcription of the *luxCDABE* reporter from the P_{sal} 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*[51]. The activated transcriptional regulator DmpR subsequently binds the P_o 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)[56]. 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.

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

Type of biosensor	Examples	Advantage	Disadvantage
Transcription-based	Transcription factor	Can be used modularly	Slow response, relies on gene expression[95]
Translation-based	Riboswitch, ribozyme	Natural riboswitches can have high sensitivity and specificity[13]	Not broadly applicable as design of novel riboswitches is challenging
Post-translational-based	Fluorescence, FRET, protease	Fast response[96]	Absolute signals often low[12]

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.

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 synthases[46]. 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 effector of human–microbe interactions[45,46]. While much work in the microbiome field currently relies on sequence-based metagenomics[47–49], the use of such synthetic biology–powered functional metagenomics can yield new bioactive molecules that modulate the gut microbiome–human host system and would likely 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, [^{13}C]naphthalene was incubated with a contaminated groundwater sample, and the ^{13}C-incorporated DNA was subsequently isolated using ultracentrifugation[50]. 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 *nag2* 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[51,52]. As over 300 enzyme species have been reported to generate phenols or phenyl compounds as byproducts of their reactions[51], 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 genetic 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 *Sphingomonas* alkaline phosphatase. Further characterization of this alkaline phosphatase identified its thermolabile properties, making it a potentially useful tool in molecular biology[53].

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[54]; the need for refactoring has also been observed in the optimization process of other biosensors[55].

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[56]. Placing the thiamine pyrophosphate (TPP, the biologically active form of thiamine)-responsive ThiM19 riboswitch[57] 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)[56]. 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[58] 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[59,60], and natural allosteric transcription factors have been identified for 243 compound classes[59]. This broad palette of compounds includes vitamins, amino acids and antibiotics, and recent studies suggest that many more riboswitches are yet to be found[61]. 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.

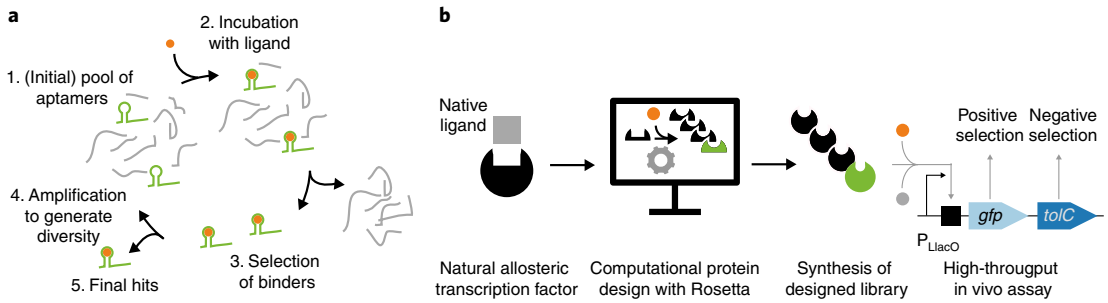

Fig. 3 | Expanding the range of compounds that can be detected using biosensors. a, Diagram of the SELEX process[94]. An initial pool of aptamers is incubated with a ligand of interest. Ligand-bound aptamers are separated from the nonbound aptamers. Using reverse transcription and PCR amplification a new pool of aptamers is generated using the ligand-bound aptamers as input. The cycle can be repeated to optimize the binding characteristics of the aptamers. **b**, Allosteric transcription factors binding a natural ligand are used as starting point to engineer binding of a ligand of interest (orange). Using Rosetta software the ligand-binding pocket is computationally redesigned[66]. A DNA library is synthesized containing the design variants. These protein-expressed designs are interrogated using a genetic circuit by both positive (using GFP) and negative (using TolC–colicin E1) selection.

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[62]. 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[63]. 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[64]. 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[65–67]. As an example, bacterial allosteric transcription factors were redesigned using the Rosetta framework (Fig. 3b)[66]. 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, gentiobiose, lactitol and sucralose. Another method to broaden the range of available biosensors relies on ligand-binding domains that contain engineered destabilizing mutations[67]. 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[68] 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[69]. 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, parathion, 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[69]. 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 phloretin[70]. 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)[11]. 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[71,72].

Most of the previously discussed examples are screening approaches, not for selection (except for the thiamine and xanthine transport mining circuits[56] shown in Fig. 2h). Several sources[73–76] 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[56] from ~10^4 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)[77]. 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^{-8} 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

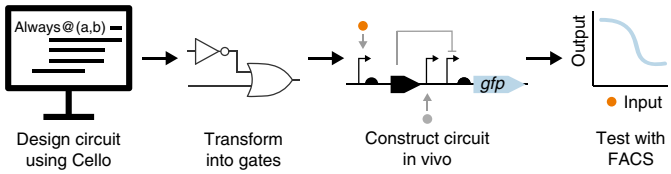

Fig. 4 | Genetic circuit design using Cello.[11] The user inputs the design requirements in Cello in Verilog code. The Verilog code is transformed into a circuit diagram, which is subsequently converted in a DNA sequence representing the genetic circuit. Eventually, the circuit can be tested in vivo using fluorescence-assisted cell sorting (FACS).

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[78,79]. These modes include point mutations, deletions and insertion sequence (IS) elements causing the circuit to fail[80]. However as a general rule these can be mitigated by introducing redundancy into the genetic circuit[78,79], for example by introducing two copies of a selection marker[56,77]. 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, cosmids 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. [24]) 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. [81]) 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[83]. Over the past decade, more expression hosts that enable a broader use of metagenomic DNA have been introduced. To study the differential expression, six proteobacteria—*A. tumefaciens* (Alphaproteobacteria), *Burkholderia graminis* (Betaproteobacteria), *Caulobacter vibrioides* (Alphaproteobacteria), *E. coli* (Gammaproteobacteria), *P. putida* (Gammaproteobacteria), and *Ralstonia metallidurans* (Betaproteobacteria)—were transformed with the same metagenomic library[84]. The transformed cells were then assayed for antibacterial activity, altered pigmentation, and altered colony morphology using the classic phenotypic readout method (Fig. 1). The behavior of the different species overlapped minimally, suggestive of differences in the ability of these hosts to functionally express metagenomics libraries. However, in practice, using a variety of proteobacteria is often an obstacle because broad-range shuttle vectors[85] need to be used and the screening assay or selection must be compatible with all of the hosts. Additionally, some host organisms might not be suitable for use with large metagenomic libraries because of their low transformation efficiency. Metagenomic DNA with a eukaryotic origin is unlikely to be heterologously expressed in the aforementioned organisms because of the lack of prokaryotic intron splicing and conservation of promoter sequences. To access eukaryotic DNA, a cDNA library can be constructed (by reverse transcribing environmental RNA) and cloned into a heterologous eukaryotic microbial host such as *Saccharomyces cerevisiae*. Using this method a histidine auxotrophic yeast mutant was rescued by cDNA, possibly originating from a eukaryotic basidiomycete fungal species and an ascomycete[86].

There was a minimum overlap in expression when six different hosts were transformed with the same metagenomic library[84]. This indicates, combined with the calculations that only 40% of heterologous proteins from a randomly cloned metagenomic library are expressed in *E. coli*[83], 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[87]. 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[88]. 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[89]. 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[90]. This 'enforced 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[91].

Challenges due to codon bias, hampering translation, can be solved by using high-throughput DNA synthesis technology[92], for example, by matching codon frequencies as near as possible to the native expression system[85].

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[93] 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

1. Turnbaugh, P. J. & Gordon, J. I. An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**, 708–713 (2008).
2. Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput. Biol.* **2**, e92 (2006).
3. Amann, R. I., Ludwig, W. & Schleifer, K.-H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
4. Torsvik, V., Goksøyr, J. & Daae, F. L. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**, 782–787 (1990).
5. Staley, J. T. & Konopka, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* **39**, 321–346 (1985).
6. Hess, M. et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467 (2011).
7. Burstein, D. et al. New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
8. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
9. Escobar-Zepeda, A., Vera-Ponce de León, A. & Sanchez-Flores, A. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Front. Genet.* **6**, 348 (2015).
10. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998).
11. Nielsen, A. A. et al. Genetic circuit design automation. *Science* **352**, aac7341 (2016).
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.
12. Eggeling, L., Bott, M. & Marienhagen, J. Novel screening methods–biosensors. *Curr. Opin. Biotechnol.* **35**, 30–36 (2015).
13. Michener, J. K., Thodey, K., Liang, J. C. & Smolke, C. D. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab. Eng.* **14**, 212–222 (2012).
14. Rondon, M. R. et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).
15. Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G. & Daniel, R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 3113–3116 (2000).
16. Brady, S. F. & Clardy, J. Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J. Am. Chem. Soc.* **122**, 12903–12904 (2000).
17. MacNeil, I. A. et al. Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* **3**, 301–308 (2001).
18. Gillespie, D. E. et al. Isolation of antibiotics turbomycin A and turbomycin B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* **68**, 4301–4306 (2002).
19. Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W. R. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl. Environ. Microbiol.* **67**, 89–99 (2001).
20. Simon, C., Herath, J., Rockstroh, S. & Daniel, R. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl. Environ. Microbiol.* **75**, 2964–2968 (2009).
21. Charlop-Powers, Z., Banik, J. J., Owen, J. G., Craig, J. W. & Brady, S. F. Selective enrichment of environmental DNA libraries for genes encoding nonribosomal peptides and polyketides by phosphopantetheine transferase-dependent complementation of siderophore biosynthesis. *ACS Chem. Biol.* **8**, 138–143 (2013).
22. Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* **6**, 981–989 (2004).
23. Wichmann, F., Udikovic-Kolic, N., Andrew, S. & Handelsman, J. Diverse antibiotic resistance genes in dairy cow manure. *MBio* **5**, e01017–13 (2014).
24. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
25. Culligan, E. P., Sleator, R. D., Marchesi, J. R. & Hill, C. Metagenomic identification of a novel salt tolerance gene from the human gut microbiome which encodes a membrane protein with homology to a brp/blh-family β -carotene 15,15′-monooxygenase. *PLoS One* **9**, e103318 (2014).
26. Culligan, E. P., Sleator, R. D., Marchesi, J. R. & Hill, C. Functional environmental screening of a metagenomic library identifies stlA; a unique salt tolerance locus from the human gut microbiome. *PLoS One* **8**, e82985 (2013).
27. Guazzaroni, M. E., Morgante, V., Mirete, S. & González-Pastor, J. E. Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environ. Microbiol.* **15**, 1088–1102 (2013).
28. Varaljay, V. A. et al. Functional metagenomic selection of ribulose 1,5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environ. Microbiol.* **18**, 1187–1199 (2016).
29. Henning, H. et al. Identification of novel benzoylformate decarboxylases by growth selection. *Appl. Environ. Microbiol.* **72**, 7510–7517 (2006).
30. Sommer, M. O., Church, G. M. & Dantas, G. A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. *Mol. Syst. Biol.* **6**, 360 (2010).
31. Forsberg, K. J. et al. Identification of genes conferring tolerance to lignocellulose-derived inhibitors by functional selections in soil metagenomes. *Appl. Environ. Microbiol.* **82**, 528–537 (2015).
32. Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* **23**, 88–93 (2005).
33. Uchiyama, T. & Miyazaki, K. Metagenomic screening for aromatic compound-responsive transcriptional regulators. *PLoS One* **8**, e75795 (2013).
34. Han, S. S., Lee, J. Y., Kim, W. H., Shin, H. J. & Kim, G. J. Screening of promoters from metagenomic DNA and their use for the construction of expression vectors. *J. Microbiol. Biotechnol.* **18**, 1634–1640 (2008).
35. Lee, S. H., Kim, J. M., Lee, H. J. & Jeon, C. O. Screening of promoters from rhizosphere metagenomic DNA using a promoter-trap vector and flow cytometric cell sorting. *J. Basic Microbiol.* **51**, 52–60 (2011).
36. Meier, M. J., Paterson, E. S. & Lambert, I. B. Use of substrate-induced gene expression in metagenomic analysis of an aromatic hydrocarbon-contaminated soil. *Appl. Environ. Microbiol.* **82**, 897–909 (2015).
37. de Lorenzo, V. Problems with metagenomic screening. *Nat. Biotechnol.* **23**, 1045–1046 (2005). author reply 1045–1046.
38. Williamson, L. L. et al. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl. Environ. Microbiol.* **71**, 6335–6344 (2005).
39. Guan, C. et al. Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. *Appl. Environ. Microbiol.* **73**, 3669–3676 (2007).
40. Nasuno, E. et al. Phylogenetically novel LuxI/LuxR-type quorum sensing systems isolated using a metagenomic approach. *Appl. Environ. Microbiol.* **78**, 8067–8074 (2012).
41. Schipper, C. et al. Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* **75**, 224–233 (2009).
42. Uchiyama, T. & Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl. Environ. Microbiol.* **76**, 7029–7035 (2010).
43. Lakhdari, O. et al. Functional metagenomics: a high throughput screening method to decipher microbiota-driven NF-κB modulation in the human gut. *PLoS One* **5**, e13092 (2010).
44. de Wouters, T. et al. A robust and adaptable high throughput screening method to study host-microbiota interactions in the human intestine. *PLoS One* **9**, e105598 (2014).
45. Cohen, L. J. et al. Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. *Proc. Natl Acad. Sci. USA* **112**, E4825–E4834 (2015).
46. Cohen, L. J. et al. Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* **549**, 48–53 (2017).
47. Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature* **12**, 174–180 (2011).
48. Schloissnig, S. et al. Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50 (2013).
49. Guo, C. J. et al. Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell* **168**, 517–526.e18 (2017).
50. Wang, Y. et al. A culture-independent approach to unravel uncultured bacteria and functional genes in a complex microbial community. *PLoS One* **7**, e47530 (2012).

51. Choi, S.L. Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits. *ACS Synth. Biol.* **3**, 163–171 (2014). **This study is the first to couple a genetic circuit to mine metagenomic library to high-throughput FACS screening**.
52. Kim, H., Kwam, K. K., Rha, E. & Lee, S.-G. in *Hydrocarbon and Lipid Microbiology Protocols*. (eds. McGenity T., Timmis K., Nogales B.) 3–12 (Springer Protocols Handbooks, Springer, Berlin, Heidelberg, 2015).
53. Lee, D.-H. et al. A novel psychrophilic alkaline phosphatase from the metagenome of tidal flat sediments. *BMC Biotechnol.* **15**, 1–13 (2015).
54. Jeong, Y. S. et al. High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes. *Protein Eng. Des. Sel.* **25**, 725–731 (2012).
55. Siedler, S. et al. Development of a bacterial biosensor for rapid screening of yeast p-coumaric acid production. *ACS Synth. Biol.* **6**, 1860–1869 (2017).
56. Genee, H. J. et al. Functional mining of transporters using synthetic selections. *Nat. Chem. Biol.* **12**, 1015–1022 (2016). **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**.
57. Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in *Escherichia coli*. *Nucleic Acids Res.* **37**, e39 (2009).
58. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).
59. Novichkov, P. S. et al. RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genom.* **14**, 745 (2013).
60. Mellin, J. R. & Cossart, P. Unexpected versatility in bacterial riboswitches. *Trends Genet.* **31**, 150–156 (2015).
61. Dar, D. et al. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**, aad9822 (2016).
62. Wittmann, A. & Suess, B. Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* **586**, 2076–2083 (2012).
63. Espah Borujeni, A., Mishler, D. M., Wang, J., Huso, W. & Salis, H. M. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Res.* **44**, 1–13 (2016).
64. Domin, G. et al. Applicability of a computational design approach for synthetic riboswitches. *Nucleic Acids Res.* **45**, 4108–4119 (2017).
65. Jha, R. K., Chakraborti, S., Kern, T. L., Fox, D. T. & Strauss, C. E. M. Rosetta comparative modeling for library design: Engineering alternative inducer specificity in a transcription factor. *Proteins* **83**, 1327–1340 (2015).
66. Taylor, N. D. et al. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2016). **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**.
67. Feng, J. et al. A general strategy to construct small molecule biosensors in eukaryotes. *eLife* **4**, 7250–7257 (2015).
68. Tinberg, C. E. et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
69. Libis, V., Delépine, B. & Faulon, J.-L. Expanding biosensing abilities through computer-aided design of metabolic pathways. *ACS Synth. Biol.* **5**, 1076–1085 (2016).
70. Ausländer, S., Ausländer, D., Müller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* **487**, 123–127 (2012).
71. Chen, Y. J. et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).
72. Stanton, B. C. et al. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10**, 99–105 (2014).
73. Lorenz, P. & Eck, J. Metagenomics and industrial applications. *Nat. Rev. Microbiol.* **3**, 510–516 (2005).
74. Taupp, M., Mewis, K. & Hallam, S. J. The art and design of functional metagenomic screens. *Curr. Opin. Biotechnol.* **22**, 465–472 (2011).
75. Milshteyn, A., Schneider, J. S. & Brady, S. F. Mining the metabiome: identifying novel natural products from microbial communities. *Chem. Biol.* **21**, 1211–1223 (2014).
76. Huang, W. E., Song, Y. & Xu, J. Single cell biotechnology to shed a light on biological 'dark matter' in nature. *Microb. Biotechnol.* **8**, 15–16 (2015).
77. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl Acad. Sci. USA* **111**, 17803–17808 (2014).
78. Gallagher, R. R., Patel, J. R., Interiano, A. L., Rovner, A. J. & Isaacs, F. J. Multilayered genetic safeguards limit growth of microorganisms to defined environments. *Nucleic Acids Res.* **43**, 1945–1954 (2015).
79. Chan, C. T. Y., Lee, J. W., Cameron, D. E., Bashor, C. J. & Collins, J. J. 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nat. Chem. Biol.* **12**, 82–86 (2016).
80. Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K. & Sommer, M. O. A. Diverse genetic error modes constrain large-scale bio-based production. *Nat. Commun.* **9**, 787 (2018).
81. Iftime, D. et al. Streptocollin, a type IV lanthipeptide produced by *Streptomyces collinus* Tü 365. *ChemBioChem* **16**, 2615–2623 (2015).
82. Laureti, L. et al. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl Acad. Sci. USA* **108**, 6258–6263 (2011).
83. Gabor, E. M., Alkema, W. B. L. & Janssen, D. B. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* **6**, 879–886 (2004).
84. Craig, J. W., Chang, F.-Y., Kim, J. H., Obiajulu, S. C. & Brady, S. F. 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 (2010).
85. Angov, E., Hillier, C. J., Kincaid, R. L. & Lyon, J. A. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One* **3**, e2189 (2008).
86. Bailly, J. et al. Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J.* **1**, 632–642 (2007).
87. Ferrer, M., Beloqui, A., Timmis, K. N. & Golyshin, P. N. Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* **16**, 109–123 (2009).
88. Stevens, D. C. et al. Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in *Escherichia coli*. *PLoS One* **8**, e64858 (2013).
89. Gaida, S. M. et al. Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nat. Commun.* **6**, 7045 (2015). **This paper shows how addition of heterologous sigma factors to a metagenomic expression host increases the fraction of transcribed DNA**.
90. Kim, Y. J. et al. Improved metagenome screening efficiency by random insertion of T7 promoters. *J. Biotechnol.* **230**, 47–53 (2016).
91. Kushwaha, M. & Salis, H. M. A portable expression resource for engineering cross-species genetic circuits and pathways. *Nat. Commun.* **6**, 7832 (2015).
92. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
93. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: managing the interplay between natural and engineered biology. *FEBS Lett.* **586**, 2199–2206 (2012).
94. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
95. Gredell, J. A., Frei, C. S. & Cirino, P. C. Protein and RNA engineering to customize microbial molecular reporting. *Biotechnol. J.* **7**, 477–499 (2012).
96. Mishra, D., Rivera, P. M., Lin, A., Del Vecchio, D. & Weiss, R. A load driver device for engineering modularity in biological networks. *Nat. Biotechnol.* **32**, 1268–1275 (2014).

Acknowledgements

This research was funded by the Novo Nordisk Foundation and the European Union Seventh Framework Programme (FP7-KBBE-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.J.G. and M.O.A.S. are co-founders of Biosyntia, 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.