Perspective: SBE/AIChE 2012 James E. Bailey Award Winner

Exploring the Capabilities of the Geobiosphere's Microbial Genome

Eleftherios T. Papoutsakis
Dept. of Chemical and Biomolecular Engineering, Dept. of Biological Sciences, and the Delaware
Biotechnology Institute, University of Delaware, 15 Innovation Way, Newark, DE 19711

DOI 10.1002/aic.14032
Published online February 6, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Keywords: metabolic engineering, genome engineering, synthetic biology, metagenomics, genomic diversity, synthetic biochemistry, complex phenotypes

Introduction: Biological Systems: Ensembles of Complex Molecular Interactions

Most important properties of a cell or an ensemble of cells are the result of a complex integration of metabolic pathways, biophysical interactions, and regulatory/signal transduction events involving the products of "genes", sometimes a handful, but in most cases dozens, hundreds or more. The term "gene" here is used to represent a coding nucleic-acid polymer, largely DNA (but could also be RNA, like in retroviruses) that results eventually in the production of a functional RNA, which could be either translated into a protein or act as a regulatory, biophysical or catalytic RNA. In most cases, these genes and their interactions are not precisely known, and in fact, the complexity of their combinatorial interactions (discussed later), may prevent complete resolution by humans and their tools, in the foreseeable future at least. These genes and the resulting interactions give rise to complex biological or cellular phenotypes, i.e., "observable states". Virtually every observable biological phenotype is a complex phenotype.

Complex phenotypes span all autonomous or semiautonomous biological systems from the simplest ones (small phages or viruses) to the most complex, hierarchically, ones: physical or temporal ensembles of cells. A manifestation of a physical cell ensemble is an organ of an animal body (e.g., brain, heart or bone marrow), while a temporal ensemble would be a cellular differentiation process, such as the development of mature blood cells from hematopoietic stem cells in the bone marrow. Temporal and physical cell ensembles lead also to simple and complex ecosystems, including microbiota (the "totality" of microbes in a particular environment, such as the microbiota of the human gut; the corresponding set of genomes is referred to as the microbiome), and, to the extreme of complexity, the whole geobiosphere. There are also phenotypes that may not exist in nature, but which can be developed using the existing biological capabilities of the geobiopshere.

Understanding biological systems and their complexity and using this understanding to purposefully alter biological systems, from the simplest to the most complex, or to create new totally synthetic or semisynthetic biological or hybrid (biological/nonbiological) ensembles is the purview of biological engineering. Here, "biological engineering" is meant to encompass metabolic engineering, cell engineering, biomedical engineering, environmental engineering, and several other related "engineering" disciplines or fields dealing, in part at least, with biological systems. Tools or approaches such as genetic engineering, systems biology, synthetic biology, are engaged in the analysis as well the synthesis facets of biological engineering. What biological engineers deal with is the analysis and purposeful modification of the reaction space and the biophysical space of biological systems. This takes place at different scales from the macro to the molecular scales of biological systems. At almost all scales, fundamental principles of chemistry, biophysics, transport and interfacial phenomena, macromolecular science, and many others are at work here, and, thus, chemical and biological/biomolecular engineering, as integrative disciplines, have a major role to play in the development and growth of biological engineering. Historically, the scale of endeavor started at the macro scale (as for example in biochemical (or fermentation) engineering, a field developed by chemical engineers, two giants of chemical engineering, RH Wilhelm[1] and EL Gaden[2]) and moved in time, over the last 50+ years to the current molecular-level scale. The frequently used term biomolecular engineering is precisely used to stress the current molecular-level state of art in the field. In the last 10–15 years, the development of computational and experimental genetic and genomic tools have reached a level of sophistication that make it possible now to start contemplating the exploration of the biological capabilities of the geobiosphere for beneficial applications in biological engineering, and this

Correspondence concerning this article should be address to T. Papoutsakis at papoutsakis@dbi.udel.edu.

is the focus of this Perspective article. What are these capabilities of the geobiosphere, and what could be assessed and possibly used in the context of biological engineering, and, yes, synthetic biology.

Engineering complex phenotypes: from bioprocessing to medical engineering of the human microbiota

Engineered complex phenotypes are those we deem as potentially useful to humans or the health of a subsystem of the geobiosphere. In the rest of this work, the term complex phenotype will mean a desirable phenotype we desire to understand and/or develop in the context of *biological, cellular or metabolic engineering*. Here, these terms are used in their broadest, most fundamental sense of "engineering": to "alter purposefully" any or all components of a biological entity or ensemble.

Microbial Examples. Microbial examples of engineered phenotypes are abundant,[3–8] but in this Perspective the author would like to focus on some that underline challenging contemporary problems that have not yet had satisfactory solutions. One example is when one desires to endow a "platform" organism (e.g., *Escherichia coli* or *Saccharomyces cerevisiae*) with a desirable biosynthetic or catabolic pathway or a biophysical program that another organism may possess. Such a pathway may involve several enzyme-coding as well as regulatory genes, but what these genes are may not be known, or are only partially known. A more specific example would be to endow *E. coli* with a pathway for the degradation of a toxic organic compound (such, as, e.g., anthropogenic halogenated hydrocarbons[9]).

Another example is the ability to endow a platform organism with a "biophysical" program (e.g., a robust membrane system) that would enhance the organism's resistance to solvents or other toxic chemicals, and which some other organism may possess.[7] The latter could include a *Lactobacillus* species[10] or *Pseudomonas putida,*[11] both of which exhibit significantly higher resistance to solvents, or *Radiococcus radiodurans,* which exhibits remarkable resistance to ionizing radiation and toxic chemicals.[12] Tolerant strains would be important in bioprocessing and advanced bioremediation applications, where, in addition to maximizing the flux for a desirable product, the robustness and prolonged productivity of the cells, under realistic bioprocessing conditions, are equally important issues. Endowing cells with such desirable capabilities will likely require several or many genes or genetic loci from other organisms, most of which are likely not precisely known.

Mammalian Examples. Mammalian examples are also many, but one stands out for ingenuity and impact, and for which Shinya Yamanaka was awarded the 2012 Nobel prize in Medicine: to reprogram adult, differentiated human or mammalian cells using a set of transcription factors (TFs) to bring them "back" to a potent embryonic-stem (ES)-cell like state.[13] These are known as iPS (induced pluripotent stem) cells. Reprogramming is achieved at low frequency by introducing a set of TFs (originally four: Oct3/4, Sox2, c-Myc, and Klf4; other combinations have been since explored) under ES cell culture conditions. Once selected after reprogramming, iPS cells exhibit phenotypic characteristics of ES cells including the expression of canonical ES genes and markers. iPS have since been widely explored as a source of many differentiated cell types for applications in regenerative medicine and *in vitro* toxicology. ES cells engage a distinct set of TFs, each of which controls a distinct set of genes (known as the regulon of each TF), whose orchestrated action leads to the totipotency of ES cells in their ability to generate every cell type of an animal. The ingenuity of Yamanaka's approach derives from the bold hypothesis (which was viewed as simplistic at the time) that the forced expression of a small set of ES TFs would lead to the correct orchestration of events to reconstitute the ES phenotype and potency in adult cells. This potent example gives rise to many important questions. What does underlie the ability of the cell to engage this, rather crude, expression of few TFs to reconstitute a complex phenotype hitherto viewed as uniquely associated with the primordial ES state, and permanently lost once differentiation ensues? How do cells rewire their cellular program to achieve this remarkable feat? How does empowering simplicity overtake and stabilize the cellular complexity? One senses that these fundamental questions will occupy the minds and hands of generations of scientists, including engineers, exploring biological complexity and the rise of complex phenotypes.

Examples of Heterogeneous Cellular Ensembles. Examples involving ensembles of different cell types beyond cells of the same origin as in tissues are also growing in scope and impact. A case of tremendous complexity, but also interest in understanding and treating human disease is that of understanding and "engineering" the human microbiota.[14,15] There are at least 10 times more microbial cells than human cells in the average adult human body, and this is enormous, in both numbers and variety, and population of microbes (frequently viewed now as another "organ") is indispensable for the development and health of the human body. Humans and animals are not alone with their own cells. They have developed, evolutionarily, to adapt to, coexist with, benefit from and, frequently, suffer from the microbes that live within (e.g., in the gut) and on them (skin). The best understood case is the physiology of the gut, which has evolved to depend on the presence of microbes for its function and health. The gut and other human microbiota modulate the development of innate and adaptive immunity, the biology (nutrition, cycling, apoptosis and transformation to neoplastic states) of the cells in the gut, intestinal barrier functions, immune tolerance and the susceptibility to infection. The dynamic state of the microbiota ensures a healthy homeostasis, but perturbations can lead to chronic or acute disease states, several of which (gastroenteritis, *clostridium difficile* associated colitis, irritable-bowel syndrome, among others)[16] are now viewed as requiring "engineering" of the microbiota employing established (antibiotics) and new tools (pre- or probiotics, cell therapies). While there seems to be a developing rational basis for such " engineering", a detailed molecular or mechanistic basis is not likely in the foreseeable future due to the complexity of the system, but mapping at least of the complexity will be an essential first step.

In most if not all these paradigms, and all other complex phenotypes, even when the "engineering" involves a "deterministic" set of genetic interventions, like the expression of a sequence of genes coding for the enzymes of a novel

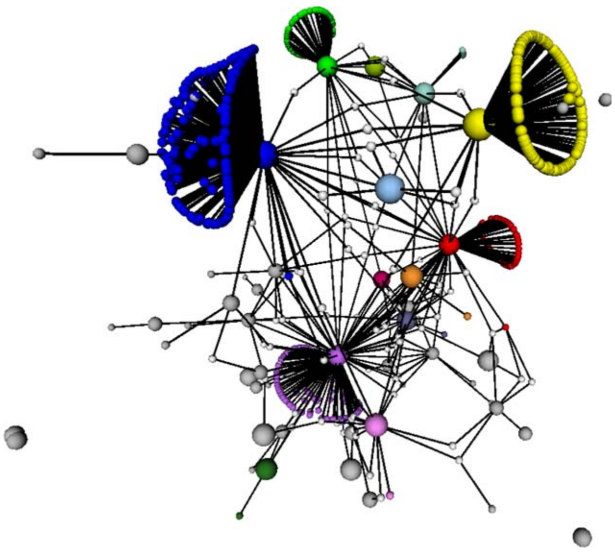


Figure 1. A partial and simplified depiction of the complex interactions among the major transcriptional regulators (transcription factors (TFs) and sigma factors (SFs)) of *C. acetobutylicum* based on the model organism of endospore formers, *B. subtilis*.

Each TF or SF (depicted as the spheres with multiple connections to other TF or SF spheres) controls alone (and sometimes in collaboration with other TFs and SFs) a set of genes (depicted as small colored spheres in symmetrical conical formation emanating from the larger spheres of the same color depicting a TF or SF) known as its regulon. The level of activity of each TF or SF activates a different subset (from 0 to 100%) of the genes of its regulon, and with different expression strength. Thus, a multitude of phenotypic states can be generated leading to many possible phenotypic outcomes, and, thus, increased stochasticity and population heterogeneity.

pathway, even from clonal cells, i.e., starting from single cells of the modified organism, there will be variation in the exhibited phenotype, and, thus, selection and adaptation will be necessary to arrive to a desirable phenotypic state. A well-known example is the aforementioned generation of the iPS cells,[13] which are rare, thus, requiring isolation by stringent selection. These phenomena derive from several sources that have as their foundation stochastic and nonstochastic events in a cell, whereby a small variation in the expression of a gene and/or the corresponding protein, like a transcriptional regulator, can lead to dramatically different phenotypic outcomes. Figure 1 depicts a very simplified version of the complex interactions among the major transcriptional regulators (TFs and sigma factors (SFs) of *Clostridium acetobutylicum,* based on the model organism of endospore-forming bacteria, *Bacillus subtilis*. Many TFs and SFs in these two and all organisms are expressed in a bimodal (early low, later high) or multimodal pattern and this allows them to alter the expression of the genes they control, i.e., their regulon. This derives from the fact that the promoters (the DNA sequence in front of a gene where a TF or SF binds to initiate transcription) of the genes in a single regulon vary enough on purpose so that the binding of the TF or SF will vary, thus, allowing to tune the expression of each gene according to the level of expression of the TF or SF. Thus, one can generate a multitude of phenotypic states, as can be readily visualized by the reader when assessing the possible states that can be generated in the system of Figure 1.

In a number of cases, stochastic and nonstochastic events along these principles can give rise to bistability or multistability, bet hedging and epigenetic inheritance,[17,18] to name some of the most explored phenomena in this fast-evolving field. This is a form of biological ”plasticity” that has been anecdotally known for many years, but which can be now interrogated at the molecular level with the precision that modern genetic and genomic tools can afford. When the phenotype engages heterogeneous cell populations, such as the aforementioned microbiota, phenotypic outcomes can vary enormously, spatially and temporally. So, in the development of these complex phenotypes, one follows a more demanding path toward its “optimization”, a path that has been visually abstracted in Figure 2. Here, engaging a variety of interventions (A, B, C…), the state of the population exhibiting a desirable complex phenotype (as represented, e.g., by its metabonome, typically defined as the set of biological elements of a biological system that provide a complete description of its metabolic state) moves up the “desirability” scale. Interventions may involve changes in environmental or culture conditions (e.g., use of nutrients or small molecules); genetic modifications (precise genetic changes, but also those resulting from accumulation of mutations); epigenetic selectivity; and, more generally, clonal selection for a selectable trait.

Where Do Complex Biological Traits Come from: Reaction and Nonreaction Based Capabilities of the Geobiosphere’s Plastic “Genome”

Complex biological traits are the result of a coordinated set of steps, many reaction based (the metabolic pathways of a cell and all cells), but many also are based on nonreactive

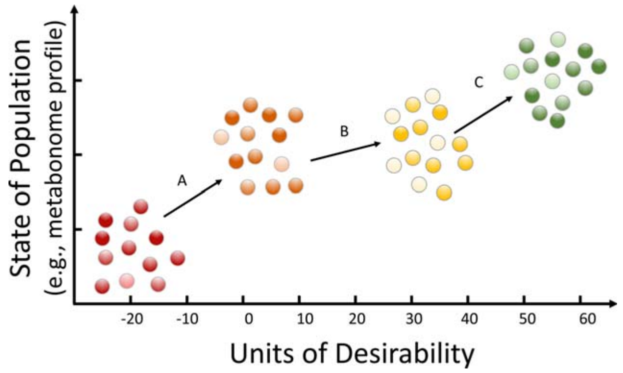


Figure 2. In engineering complex phenotypes, a population of cells whether originally homogeneous or even deriving from a single cell leads soon to increasing heterogeneity.

The heterogeneity can be narrowed or directed toward a more desirable phenotypic state (as assessed for example by the ensemble’s metabonome) by interventions, A, B, C,… Interventions may involve changes in: environmental or culture conditions (e.g., use of nutrients or small molecules), genetic modifications (precise genetic changes, but also those resulting from accumulation of mutations), epigenetic selectivity, and, more generally, clonal selection for a selectable trait.

interactions, broadly known as biophysical interactions: self- or assisted assembly of membranes, organelles, and cellular programs (e.g., the stress-response system) that engage biophysical interactions such as; protein-protein and protein-DNA interactions. How did cells evolve to engage this complexity, how do they manage this complexity with such remarkable coordination and efficiency, and what can we learn from that to allow us to engineer this complexity? These are fundamentally the core ideas and principles that drive modern biology and bioengineering, whereby we tackle parts of these big questions that are amenable to investigation with the state-of-the-art tools of the era. In evolutionary context, simple primordial, self-sustaining biological processes gave rise to increasing complexity by engaging a larger repertoire of large and small molecules and their interactions and driven by evolutionary adaptation. This biological complexity has now developed such an enormous and plastic, self-sustaining and ever evolving capability to generate new phenotypes that transcends for now the capabilities of the human mind to comprehend it.

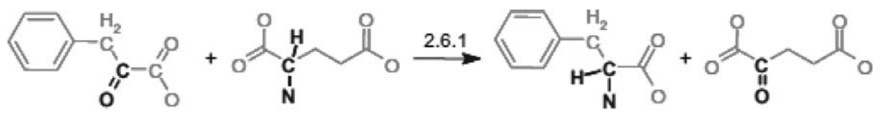
One perhaps can size these capabilities by considering the largest class of cells in the geobiosphere, the simplest microbes, namely the prokaryotes (bacteria), which, because of their relatively fast growth and diversity, constitute an enormous incubator of evolution. Each one of these simple microbes is a separate genetic-engineering laboratory, continuously evolving and responding to the changing geobiosphere, whether due to human or nonhuman activities (e.g., driven by geological events or by sun-driven weather changes), to give rise to new catalytic and biophysical capabilities and new phenotypes, including new cells and cellular ensembles, and in the long term new plant and animal species. This pool of capabilities, combined with those of other microbes (e.g., archaea or lower eukaryotes like fungi), and all eukaryotes (i.e., cells and their ensembles such as those in plants, and animals), constitute a dynamic and enormously potent machine for generating new capabilities. It is now accepted that microbial diversity has no limits in the context of human comprehension. The ongoing evolutionary process is faster, more plastic, and productive that we will ever be able to assess: trillions of organisms' continuously evolving, and giving rise to new complex traits and species. This could be viewed as the geobiosphere's "plastic and expanding genome", or at least this is the author's preferred way in describing it. It is not a "fixed" genome, but rather a plastic and expanding one. How big is this genome? The next section will address this question.

The Unimaginable Richness of the Metagenome. Part 1: What is Metagenomics and How Big is the Metagenomic Space?

The aim here is to provide a sense of the size the genomic space of the geobiosphere by briefly addressing some simple questions. How many organisms are there? How big is the diversity they encompass? These fundamental questions deserve and have received serious considerations,[19] but there is no sufficient space in this Perspective for an extensive analysis. Instead, some numbers will provide the essential part of the answers.

The number of prokaryotes on earth is estimate to be 4–6 × 10,[30] involving an estimated 10^6 to 10^8 separate genospecies.[20] Most of this microbial diversity cannot be readily captured as it is estimated that only less than 1% (0.001–0.01% from seawater, 0.25% from fresh water, 0.25% from sediments, 0.3% from soil) of them can be cultivated,[21] and this fraction is unlikely to change in the foreseeable future. So, this diversity can only be explored through metagenomics, which is both an evolving set of research methods and tools as well as a research field.[19] The methods and tools of metagenomics aim to overcome the two major problems of the unculturability and genomic diversity of the geobiosphere. Metagenomics, as a science, seeks to understand biology at the "aggregate level",[19] transcending the individual organism to focus on the genes in ensembles of cells, as they develop naturally or synthetically in various spaces and subspaces in natural or synthetic habitats, in animals, …in the geobiosphere in general. The tools are both experimental[22,23] and computational,[24,25] and for now at least, it is accepted *a priori* that the complexity of the systems, which metagenomics aims to study, is such that these systems "can only be sampled, never completely characterized".[19] The key experimental tool is the construction and screening of metagenomic libraries.[22,23,26–28] One samples a specific part of the geobiosphere (e.g., an ocean habitat, an area polluted by humans with xenobiotics, the human gut, the microbiota of an animal's skin; selection can be applied previously to sample more specific cell ensembles[29]), extracts the DNA, and cuts the DNA into appropriate size. One then inserts these DNA pieces into an expression vector (a plasmid or a fosmid) that would allow the expression of the genes on these DNA pieces into a screening host (typically *Escherichia coli*, but not only), and transforms this vector library into the screening host to construct the metagenomic library in the microbial host. This library is then screened by an appropriate assay for any screenable trait, such as an enzymatic activity, a cellular trait (e.g., tolerance to a toxic chemical) or a catabolic activity (e.g., the ability of the cell containing a library fragment to degrade a specific chemical). The key hypothesis is that the host organism has the ability to transcribe (produce the mRNA for) and translate (produce the protein from the mRNA) the foreign genes on the metagenomic DNA fragments. We discuss this crucial hypothesis later.

So, what could we sample metagenomically? It is best to start with a "small" example and calculation. It has been estimated that there are 10 trillion (10^{13}) genes in 1 g of soil, and because these are derived from at least 10^3 microbial species (different types of cells), one can estimate that there are at least 1 million *different* genes in 1 g of soil.[30] This number is actually about 3–4 million since a typical prokaryotic genome codes for about 4,000 genes. To estimate the number of genes in the geobiosphere, if we take the number (see earlier) of distinct bacterial genospecies to be 10^8, with an average number of 4×10^3 genes per genospecies, we could estimate a total of 4×10^{11} prokaryotic genes. If one includes also the large number of genes from archaea, lower (i.e., microbial, like fungi) and higher (plants and animals) eukaryotes, and viruses (their number is enormous; larger than those of bacteria; see earlier), the total number of distinct genes in the geobiosphere likely exceeds a trillion (10^{12}). Although many of these genes may code for similar functionalities (e.g., similar enzymes catalyzing the same or similar reactions), this number of genes constitutes



$$\underbrace{\begin{array}{c|ccccc|}
C & 0 & 2 & 0 & 0 & 0 \\
O & 2 & 4 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 1 & 1 \\
H & 0 & 0 & 1 & 0 & 0 \\
N & 0 & 0 & 1 & 0 & 2
\end{array}}_{\text{reactants}}
+
\underbrace{\begin{array}{c|ccccc|}
C & 0 & -2 & 0 & 1 & 1 \\
O & -2 & 4 & 2 & 0 & 0 \\
C & 0 & 2 & 0 & -1 & -1 \\
H & 1 & 0 & -1 & 0 & 0 \\
N & 1 & 0 & -1 & 0 & 2
\end{array}}_{\substack{\text{generalized reaction} \\ \text{operator for 2.6.1}}}
\longrightarrow
\underbrace{\begin{array}{c|ccccc|}
C & 0 & 0 & 0 & 1 & 1 \\
O & 0 & 4 & 2 & 0 & 0 \\
C & 0 & 2 & 0 & 0 & 0 \\
H & 1 & 0 & 0 & 0 & 0 \\
N & 1 & 0 & 0 & 0 & 2
\end{array}}_{\text{products}}$$

Figure 3. Example of bond-electron matrix (BEM) for the reacting portions of phenylpyruvate and glutamate (reactants) and phenylalanine and 2-oxoglutarate (products) along with the reaction operator matrix for the EC 2.6.1 generalized enzyme reaction (based on the enzyme commission (EC) classification system).

The reaction operator matrix is added to the BEM of the reactants to produce the BEM of the products. Negative number in the operator matrix represents bonds broken and positive number represents bonds created in the 2.6.1 reaction. (Reprinted with permission from Ref. 31).

an unimaginable richness, which when viewed from the reaction-space point of view, it is a staggering number to contemplate let alone understand and use. Given this enormous number of genes, one would readily accept as a lemma that *a protein that can catalyze whichever reaction is biologically possible already exists, but is most likely undiscovered.* As a result, the author dares to suggest that whatever is biologically "reaction possible", by whatever sequence of individual steps (as in a cellular pathway), it probably already exists, albeit perhaps not assembled in a single organism. What does "reaction possible" mean? This is a problem that has occupied the author's mind for many years since his graduate-school years, and although he has not had the expertise and the opportunity to work on it, he is happy that this problem has been fundamentally resolved by dear colleagues and friends of his, as discussed in the following section.

What is biologically "reaction possible": let us BNICE

Hatzimanikatis and coworkers[31,32] aimed to develop a reliable computational framework that would allow the identification of all possible biochemical reaction pathways for the production of any given biochemical compound from a given set of starting compounds. Earlier efforts by Mavrovouniotis and coworkers[33] had set the stage for exploring this important problem. The work of Hatzimanikatis and coworkers[31,32] constitutes a major extension of computational chemistry tools that make possible to identify every possible reaction for a given set of chemical reaction rules and starting compounds.[34] The computational framework, named biochemical network integrated computational explorer (BNICE), address this core problem in synthetic metabolic biochemistry. BNICE is based on chemical reaction rules, knowledge of enzyme-catalyzed reactions, and the outcomes are assessed further for thermodynamic feasibility and efficiency. BNICE employs the graph-theoretic matrix representation of biochemical compounds[35] and enzyme reaction rules. Molecules are coded as a bond-electron matrix (BEM),[35] whereby each atom in a molecule is represented by a row (or a column), while the other elements of the BEM represent the bond properties between atoms. Enzyme reactions are represented using a similar approach. Figure 3 depicts examples of this BEM representation. Following identification of all possible enzymatic steps that lead to the targeted chemical, thermodynamic analysis estimates the energetic efficiency of all these possible pathways. The authors have applied this approach to analyze both biosynthetic[31] and catabolic pathways.[36] For example, they examined the biosynthesis of aromatic amino acids (Figure 4) to identify almost 75,000 novel biochemical routes from chorismate (a core metabolic intermediate in amino acid biosynthesis) to phenylalanine, and more than 350,000 routes from chorismate to tyrosine.[36] The pathways that can be generated by BNICE involve compounds that exist in biological databases and chemical databases but also novel compounds. Thus, BNICE can be used to explore a broader (one would argue the broadest possible) biochemical reaction landscape and identify novel biochemical routes and compounds that remain to be discovered experimentally. To recap, computational tools like BNICE and extensions thereof can be used to explore and map the broad reaction space that is spanned by biological reactions (as we understand them on the geobiosphere), and can prioritize their thermodynamic efficiency, while adding constraints of pathway length but also additional constraints, such as those that come from metabolomics. Such constraints could include ranges of intracellular metabolite/intermediate concentrations, toxicity of the intermediate chemicals, segregation/compartmentalization into organelles, etc. Noting that biological systems can readily couple energy generation (e.g., from ATP hydrolysis) to energy needs aiming to carry out efficiently endergonic (energy requiring) reactions, the reaction landscape that can be explored is enormous. So, to come back to the basic assertion, in view of the trillion-like genes in the geobiosphere, a large fraction of which code enzyme-catalyzed reactions, it is quite likely that any predicted reaction and pathway by BNICE, or any other robust computational scheme in the future, likely already exists in the geobiosphere, or it can be "induced" to develop under evolutionary pressure. One can then extrapolate this logical conclusion to genes that code for biophysical traits and capabilities.

A related capability and field: can we expand the explorable genome-engineering space? Can we make microbial alloys?

Genome engineering aims to move the concept of metabolic or cell engineering to genome-scale modifications and

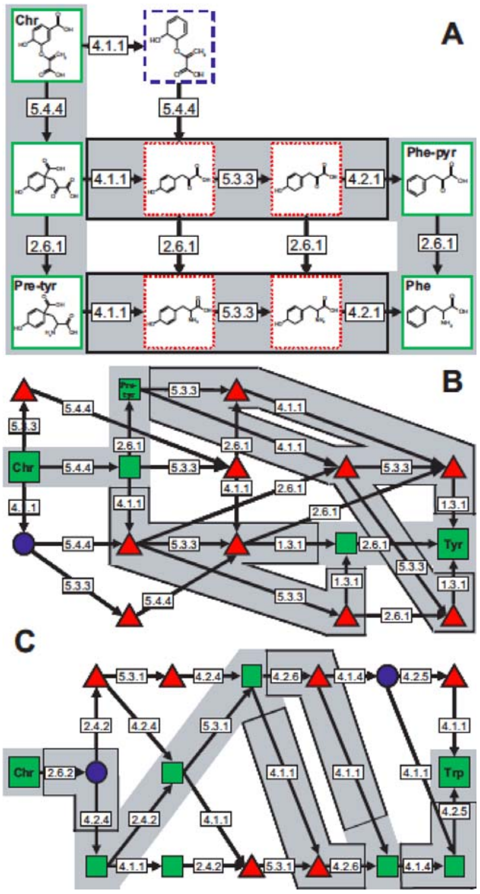


Figure 4. Alternative pathways for the biosynthesis of aromatic amino acids from chorismate: to phenylalanine (A), tyrosine (B) and tryptophan (C), with length equal to the length of the original pathway. Chorismate (Chr), phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp). Solid-line boxes (A) or squares (B and C) indicate compounds that exist in the KEGG database, dashed line boxes (A) or circles (B and C) indicate compounds that exist in the CAS database, and dotted-line boxes (A) or triangles (B and C) indicate novel compounds not found in KEGG or CAS databases. The structures of the compounds in the tyrosine (B) and tryptophan (C) pathways can be found in Ref. 31. Shaded pathways indicate the pathways that correspond to the native pathways; portions of the shaded native pathway outlined in black represent dissected reactions. (Reprinted with permission from Ref. 31).

complex traits derived from designed or evolved genetic changes, such as accumulation of beneficial mutations under selective pressure. Beyond classical mutagenesis, the first and still prototypical genome engineering method is whole genome shuffling (WGS). In WGS, mutant strains from a parent organism are generated by some mutagenic process and are then recombined (based on regions of homology) by repeated protoplast fusions. Thus, WGS aims to combine in a single-strain desirable gene mutations in order to generate improved bacterial phenotypes that can be screened for [37]. Genomic diversity can also be increased by techniques like global transcription machinery (gTME),[41] where a transcription factor is mutated to influence the entire transcriptome, multiplex automated genome engineering (MAGE),[42] whereby *in vivo* homologous recombination accelerates the rate of evolution, or trackable multiplex recombineering (TRMR)[43] that combines recombineering (genetic engineering mediated by DNA recombination processes) at specific locations with barcodes that can be traced using DNA-microarray analysis.

However, all these techniques engage and manipulate only the genome of the host organism, and, thus, can only explore intrinsic genetic diversity. New, exogenous genetic programs or pathways are, thus, not accessible. Heterologous genes can be introduced via genomic libraries to enlarge the genomic space beyond a single species, but the genetic information on these libraries must be recognized by the host. This is then virtually identical to the process of screening metagenomic libraries. Can we develop novel capabilities based on "gene sets or programs" from two or more distinctly different organisms, or the metagenome? Can we develop genome engineering strategies aiming to explore the genetic diversity of unrelated genomes or the metagenome? Can we develop hybrid organisms combining programs and cellular machineries from different genomes? Such organisms would not merely combine the properties of the parents, but rather would have novel properties that derive from some combination of the two or more genomes, analogous to the properties of an alloy or amalgam when we combine metal and nonmetal elements. It would be then appropriate to call them microbial alloys. I will discuss in the following example.

The Unimaginable Richness of the Metagenome, Part 2: The Staggering Barriers in Exploring the Diversity of the Geosphere's Genome

To regroup, effective functional screening of metagenomic libraries can unlock the hidden potential of the genetic diversity in nature and lead to the identification of novel or potent enzymatic and biophysical activities, which could be used to engineer superior organisms or ensembles of organisms (as in microbiota) for biotechnological applications.[44] So, how does one explore and use this exceptional potential of the geobiosphere's genome? While metagenomics is clearly the only means (for now at least), the current limitations of the methods and the science are prohibitive. To make progress in this field, the author believes that the following three key problems must be resolved.

Problem one: the ability to express genes from metagenomic libraries in screening hosts. "Just as staggering as these potential riches are, so are the barriers to discovery of genes by functional screening. The approach is grossly limited by the ability of the organism that is hosting the metagenomic library to express genes from anonymous organisms represented in the library."[19] This sobering statement refers to but one of the major difficulties in exploring the genomic diversity of the geobiosphere: the ability to express genes from metagenomic libraries in screening hosts, the most widely used and most likely to be used in the near future being *E. coli*. The expression difficulty is largely attributed

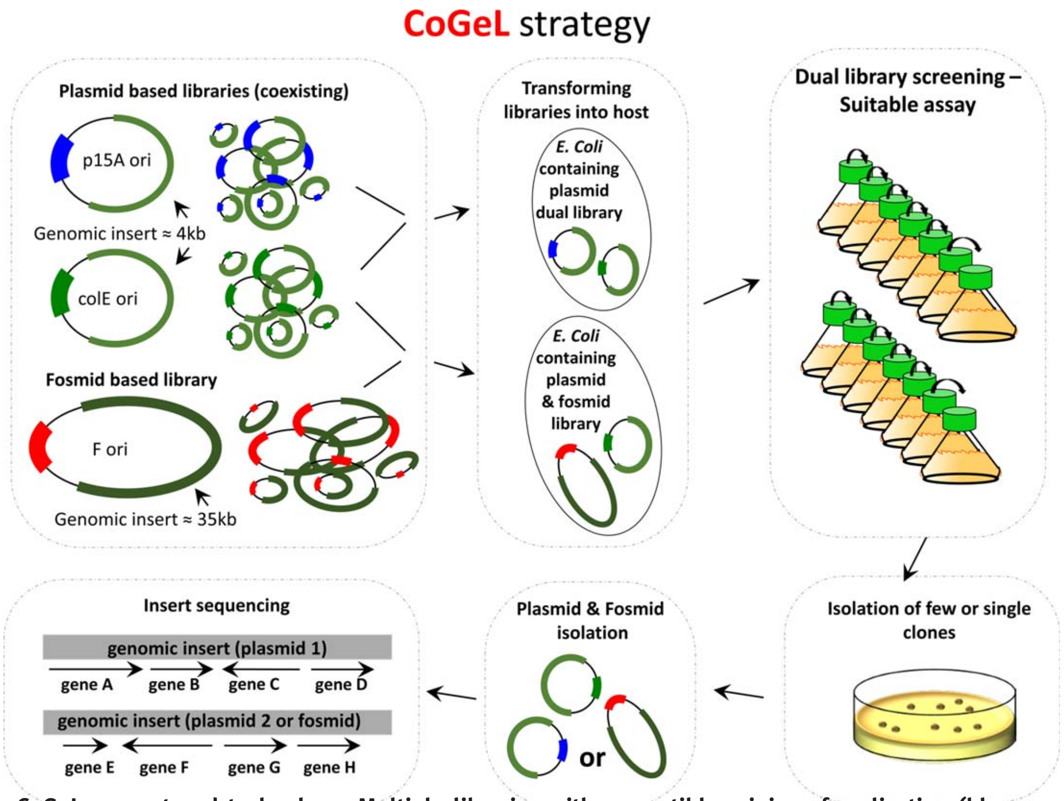

CoGeL strategy
Plasmid based libraries (coexisting)
p15A ori
Genomic insert ≈ 4kb
colE ori
Fosmid based library
F ori
Genomic insert ≈ 35kb
Transforming libraries into host
E. Coli containing plasmid dual library
E. Coli containing plasmid & fosmid library
Dual library screening – Suitable assay
Isolation of few or single clones
Plasmid & Fosmid isolation
or
Insert sequencing
genomic insert (plasmid 1)
gene A
gene B
gene C
gene D
genomic insert (plasmid 2 or fosmid)
gene E
gene F
gene G
gene H

Figure 5. The CoGeL concept and technology. Multiple libraries with compatible origins of replication (blue: p15A ori, green: colE1 ori, red: F ori) and different insert sizes (plasmid &3.5 kb, fosmid &35 kb) are constructed and transformed in a desired host (dual plasmid: blue and green, plasmid-fosmid combination: red and green). Cells containing two CoGeLs are screened for a specific phenotype, here shown as serial transfers under selective pressure, e.g. increasing concentrations of a toxic chemical (stressant) when selecting for a resistance phenotype. After enrichment, single clones are isolated by plating and nonchromosomal DNA is extracted. Genes on the selected clones can be identified by sequencing the identified library inserts. (Reprinted with permission from Ref. 45).

to the ability of the transcriptional machinery of the host organism to recognize the promoters from the metagenome and possibly also translate metagenomic transcripts.[19] However, it was concluded that "… it is essential to develop techniques that enable *E. coli* to express a greater array of genes (e.g., providing alternative sigma factors or tRNAs) and to screen libraries in bacteria from other divisions."[19]

The author and his collaborators have recently developed a strategy to overcome this limitation by engineering hybrid strains that possess a transcriptional machinery capable of recognizing promoters from nonself (heterologous) DNA. Such strains can then be employed for screening heterologous DNA libraries to express foreign genes, thus, enlarging the sampling space that can be engaged in functional metagenomics and, as already discussed, also in genome engineering. Their hypothesis is that by expressing a heterologous sigma factor, the core RNA polymerase (RNAP) that transcribes genes into mRNAs of the host (here *E. coli*) can be recruited to initiate transcription from heterologous promoters. As a proof of concept, they demonstrated that by expressing sigma factors in *E. coli* from the phylogenetically distant *Lactobacillus plantarum*, they enabled *E. coli* to recognize a large fraction of promoters from *L. plantarum* genomic libraries. In order to assess in a quantitative and high-throughput way, the fraction of heterologous promoters, and notably of *L. plantarum* promoters, that could be recognized by the native or engineered *E. coli* RNAP to initiate transcription, they constructed two promoter GFP-trap libraries. These libraries allow an accurate quantitation of the number of library inserts that contain a promoter that can be recognized by the native or engineered RNAP of *E. coli*. They also showed that one can express several heterologous SFs in *E. coli*, thus, advancing the concept that it is possible to engineer *E. coli* strains capable of recognizing many if not most of the promoters and thus express a large fraction of genes from metagenomic libraries. The author calls such strains *transcriptional alloys* (*Tr* alloys) because the RNAP complex of these strains behaves neither as the original RNAP of the *E. coli* host nor like the RNAP complex of the SF- "donor" host. Instead it has properties from both host and donor, and such properties are analogous to the properties of metallic alloys: different from either metal or nonmetal components of the metallic alloy.

Problem two: how to capture the combinatorial complexity in metagenomic (or genomic) libraries. Complex phenotypes arise in cells from interactions among genes, pathways, signaling events and cellular programs under changing environmental conditions. In screening either metagenomic or genomic libraries, what is identified is a single genetic locus made up or one of a few genes (for

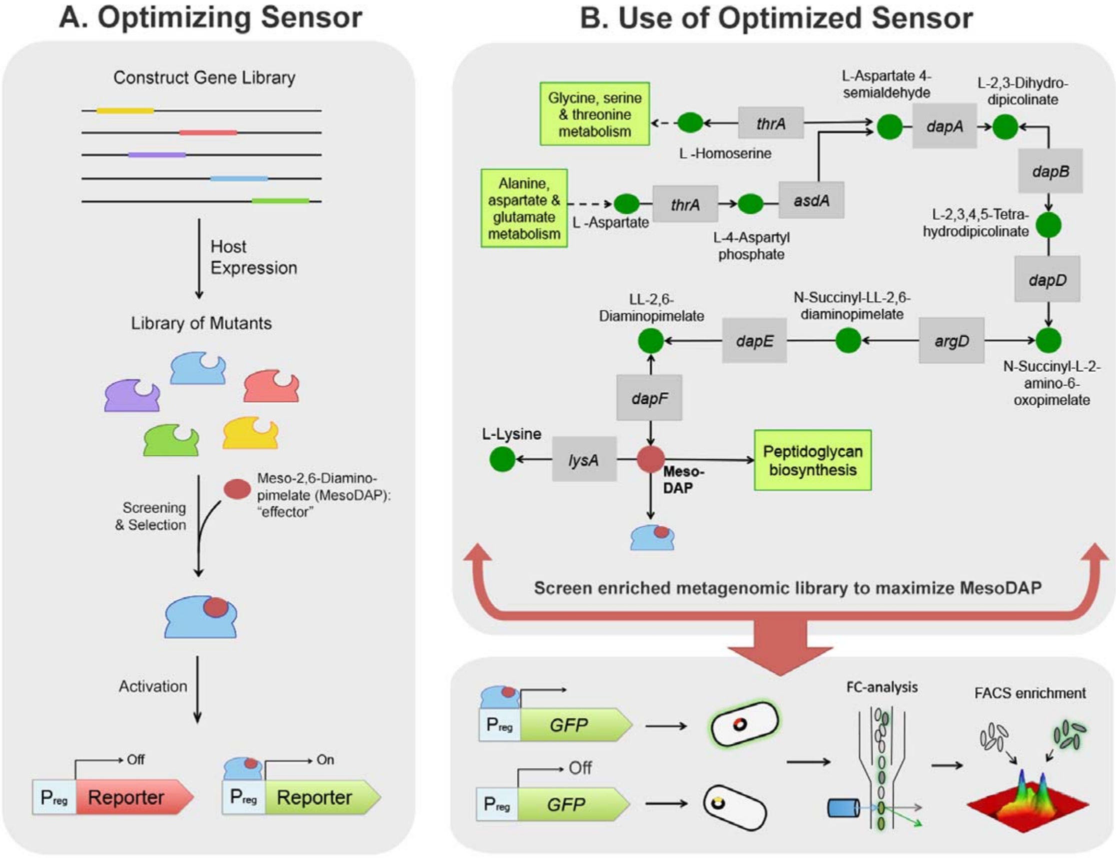

A. Optimizing Sensor
Construct Gene Library
Host Expression
Library of Mutants
Meso-2,6-Diamino-pimelate (MesoDAP): "effector"
Screening & Selection
Activation
Off
On
Preg
Reporter
Preg
Reporter
B. Use of Optimized Sensor
Glycine, serine & threonine metabolism
thrA
L-Aspartate 4-semialdehyde
dapA
L-2,3-Dihydro-dipicolinate
L -Homoserine
dapB
Alanine, aspartate & glutamate metabolism
L -Aspartate
thrA
asdA
L-4-Aspartyl phosphate
L-2,3,4,5-Tetra-hydrodipicolinate
dapD
LL-2,6-Diaminopimelate
dapE
N-Succinyl-LL-2,6-diaminopimelate
argD
N-Succinyl-L-2-amino-6-oxopimelate
dapF
L-Lysine
lysA
Meso-DAP
Peptidoglycan biosynthesis
Screen enriched metagenomic library to maximize MesoDAP
Preg
GFP
Off
Preg
GFP
FC-analysis
FACS enrichment

Figure 6. (A) Generating and optimizing intracellular sensors for small molecules by protein engineering (adapted from Ref. 53), and (B) using such a sensor to screen a targeted metagenomic library to identify genes that maximize the (intracellular) production of a small molecule (here: meso-2,6-Diamino-pimelate; mesoDAP) of the lysine biosynthesis pathway as a hypothetical goal.

(A) Protein engineering is used to develop a regulatory protein to recognize a small molecule ("effector"; here mesoDAP is used as a specific example). A diverse library of appropriate genes is constructed and expressed in a host strain to produce a large and diverse library of variants of the regulatory protein (RP). This library is then exposed to the small molecule (effector) of interest. If the effector molecule binds one of the variants of RP to "activate" it, the activated RP binds to P_{reg} to enable expression of the reporter gene (e.g., a fluorescent protein like green fluorescent protein, GFP). Thus, desirable variants can be isolated by a screening method such as fluorescence-activated cell sorting (FACS; see part B). Enhanced responses and optimization of the RP properties, such as affinity and selectivity for the effector molecule, are achieved through iterative cycles of this process. P_{reg}: promoter to which the activated RP can bind to enable expression of the reporter. (B) In this hypothetical example based on the lysine biosynthesis pathway in *E. coli* and similar organisms, the goal is the identify genes from an enriched metagenomic library that can enhance the flux to and intracellular accumulation of mesoDAP, the important precursor to lysine as well as peptidoglycan biosynthesis. A sensor system for mesoDAP is developed as in A, and is used to screen the metagenomic library using GFP as the reporter protein. Screening is carried out through the flow cytometry (FC)-based FACS, whereby clones with strong GFP fluorescence signal are isolated and sequenced to identify genes from the metagenomic library that lead to enhanced mesoDAP synthesis in the cells.

plasmid-borne libraries) or more genes on large DNA fragments (35 kb) of fosmid libraries. Standard, i.e., single genomic or metagenomic libraries cannot capture interactions among distantly located loci on a chromosome (and/or multiple chromosomes for metagenomic libraries) necessary to create or improve a complex phenotype, such as a novel pathway or biophysical trait like those we discussed previously. This derives from the fact that each library cell contains one type of library vector and thus one DNA fragment. For plasmid-borne libraries, the insert size can be up to 6–8 kb of DNA or in the case of fosmid libraries about 35 kb of contiguous DNA. This inability to have multiple DNA fragments (library inserts) in a single cell (clone) combined with the DNA-fragment size limitation constrains the combinatorial genomic space that can be sampled and hinders the identification of beneficial interactions among distantly-located genetic loci.

To overcome these limitations, the author and his collaborators developed and demonstrated a novel strategy and technology, that of the coexisting genomic libraries (CoGeLs)[45] (Figure 5). CoGeLs enable two (and more, but practically only a few) genomic (and/or metagenomic) libraries to coexist in one cell thus allowing to screen for necessary or cooperative gene interactions in the development or improvement of a screenable phenotype. Genomic fragments normally distantly located in a genome or multiple genomes can be expressed together in a single cell and screened for beneficial interactions. For small-insert plasmid-based libraries, the number of binary (let alone trinary) combinations of DNA-fragment inserts to be screened is very large because even single-insert libraries require a large number of individual clones to achieve a desired genome-coverage probability. This limitation can be overcome by utilizing a fosmid library (ca. 35-kb insert size) in combination with a coexisting plasmid library, and/or by using enriched libraries. Thus, the number of individual CoGeL clones necessary for a desirable genome-coverage probability is reduced by one order of magnitude. The use of this technology was demonstrated by developing desirable complex traits (reconstitution of an incapacitated pathway, tolerance to acid and solvent stress) that depend on interacting genomic loci.[45,46]

Problem three: the screening assays are core to the success in exploring the genomic or metagenomic diversity.[23] In principle, any assay for any enzyme or cellular trait that can be implemented in the context of library screening would advance the cause. Typical assays include screens for enzymatic activities of proteases, lipolytic enzymes, polysaccharide degrading enzymes, and generally hydrolytic activities. Others include assays that result in the change of color or fluorescence of a compound. Perhaps the easiest and cleanest assays are those based on survival and cell growth (null, low or high) of the host cell under some screening condition, such exposure to a stressant or toxin, like acid stress, oxidative stress, toxic solvents or xenobiotics.[7,45–50] Similarly robust assays are those that screen to identify one or more genes that can complement known defects in a pathway or biophysical trait (e.g., a missing component of the cell's stress response system) of the host cell.[45] However, a major need exists to develop general and easy to adapt assays to screen for more complex traits, such as novel pathways, biochemical intermediates or final metabolic products, and do so in a high-throughput fashion. An example would be the use of flow cytometry in combination with specific screens based on "sensor" chemicals or fluorescence assays. FC-based assays,[51] as well as microfluidic-based assays[23] hold great promise, but much work remains to be done to make such assays are easy to adapt, validate and use. A relatively new research activity that will have a large impact in resolving this problem is the design and implementation of synthetic-biology based *in vivo* assays that can be made to sense small molecules in the cell.[52,53] The concept of "optimizing" an intracellular sensor for a small molecule, and using such a sensor for targeted screening of a metagenomic library to identify genes that maximize the production of small molecule is illustrated in Figure 6.

The same set of issues prevents the expansion of the usable genomic space that can be explored in the context of genome engineering,[45,54,55] whereby, so far, as already discussed in the previous section, the genomic diversity that can be generated and screened for useful phenotypes is confined to mutational perturbations of single genomes.[41–43]

Epilogue: Just a Beginning

Metagenomics and genome engineering crossing the boundaries of a single genome are at an infant state of development. From basic scientific endeavors to engineering and technological applications, a virtually inexhaustible space of reaction and biophysical capabilities awaits exploration. As this exploration develops, many new, unimaginable for now questions will arise that will get to the core of issues of the design principles of biological processes and self-sustaining life. How did biology and life processes develop? How did biological complexity arise and how does it "stabilize" biological entities? How do phenotypes "prevail"? How do cells assemble *ab initio* novel pathways and biophysical entities, and many more? As noted earlier, chemical and biological/biomolecular engineering, as integrative disciplines, will be an integral part of this exploration, of this journey.

Acknowledgments

I would like to acknowledge support by the National Science Foundation grant CBET-1033926, and by the USA Office of Naval Research (ONR) Grant N000141010161. I thank Sergios Nicolaou, Stefan Gaida and Kyle Zingaro for working with me on issues discussed in this manuscript. I thank Dr. Carlos Paredes, a former postdoctoral fellow in my group for inspiration, for asking penetrating questions and also assistance for preparing Figure 1. I thank Alan Fast for preparing Figure 2, and Stephanie Luff for preparing Figure 6.

Literature Cited

1. Bartholomew WH, Karow EO, Sfat MR, Wilhelm RH. Oxygen transfer and agitation in submerged fermentations - mass transfer of oxygen in submerged fermentation of streptomyces-griseus. *Ind Eng Chem.* 1950;42(9):1801–1809.
2. Hixson AW, Gaden EL. Oxygen transfer in submerged fermentation. *Ind Eng Chem.* 1950;42(9):1792–1801.
3. Woodruff LBA, Gill RT. Engineering genomes in multiplex. *Curr Opin Biotechnol.* 2011;22(4):576–583.
4. Patnaik R. Engineering complex phenotypes in industrial strains. *Biotechnol Progr.* 2008;24(1):38–47.
5. Wisselink HW, Toirkens MJ, Wu Q, Pronk JT, van Maris AJA. Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered saccharomyces cerevisiae strains. *Appl Environ Microbiol.* 2009;75(4):907–914.
6. Borden JR, Jones SW, Indurthi D, Chen Y, Papoutsakis ET. A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in Clostridium acetobutylicum involving non-coding RNAs and ribosomal RNA processing. *Metab Eng.* 2010;12(3):268–281.
7. Nicolaou SA, Gaida SM, Papoutsakis ET. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metab Eng.* 2010;12(4):307–331.
8. Blaby IK, Lyons BJ, Wroclawska-Hughes E, Phillips GCF, Pyle TP, Chamberlin SG, Benner SA, Lyons TJ, de Crécy-Lagard V, de Crécy E. Experimental evolution of a facultative thermophile from a mesophilic ancestor. *Appl Environ Microbiol.* 2012;78(1):144–155.
9. Smidt H, de Vos WM. Anaerobic microbial dehalogenation. *Annu Rev Microbiol.* 2004;58:43–73.
10. Knoshaug EP, Zhang M. Butanol tolerance in a selection of microorganisms. *Applied Biochem Biotechnol.* 2009;153(1-2):13–20.
11. Ruhl J, Schmid A, Blank LM. Selected pseudomonas putida strains able to grow in the presence of high butanol concentrations. *Appl Environ Microbiol.* 2009; 75(13):4653–4656.

12. Cox MM, Battista JR. Deinococcus radiodurans - the consummate survivor. *Nat Rev Microbiol.* 2005;3(11): 882–892.

13. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006;126(4):663–676.

14. Preidis GA, Versalovic J. Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era. *Gastroenterol.* 2009;136(6):2015–2031.

15. Sonnenburg JL, Fischbach MA. Community health care: therapeutic opportunities in the human microbiome. *Sci Trans Med.* 2011;3(78): 78ps12. doi: 10.1126/scitranslmed.3001626.

16. Zhu YM, Luo TM, Jobin C, Young HA. Gut microbiota and probiotics in colon tumorigenesis. *Cancer Lett.* 2011;309(2):119–127.

17. Veening JW, Smits WK, Kuipers OP. Bistability, epigenetics and bet-hedging in bacteria. *Ann Rev Microbiol.* 2008:62;193–210.

18. Veening JW, Stewart EJ, Berngruber TW, Taddei F, Kuipers OP, Hamoen LW. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Nat Acad Sci USA.* 2008;105(11):4393–4398.

19. USA National Research Council. Committee on Metagenomics: Challenges and Functional Applications. USA National Academies Press. In: The new science of metagenomics : revealing the secrets of our microbial planet. Washington, DC: National Academies Press; 2007.

20. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. *Proc Nat Acad Sci USA.* 1998;95(12):6578–6583.

21. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in-situ detection of individual microbial--cells without cultivation. *Microbiol Rev.* 1995;59(1): 143–169.

22. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Ann Rev Genet.* 2004;38:525–552.

23. Taupp M, Mewis K, Hallam SJ. The art and design of functional metagenomic screens. *Curr Opin Biotechnol.* 2011;22(3):465–472.

24. Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol.* 2005;71(3):1501–1506.

25. Schloss PD, Handelsman J. A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *Bmc Bioinform.* 2008;9:34.

26. Schmeisser C, Steele H, Streit WR. Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol.* 2007;75(5):955–962.

27. Warren RL, Freeman JD, Levesque RC, Smailus DE, Flibotte S, Holt RA. Transcription of foreign DNA in Escherichia coli. *Genome Res.* 2008;18(11):1798–1805.

28. Gabor EM, Alkema WB, Janssen DB. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol.* 2004;6(9): 879–886.

29. Schloss PD, Handelsman J. Biotechnological prospects from metagenomics. *Current Opin Biotechnol.* 2003; 14(3):303–310.

30. Schloss PD, Handelsman J. Toward a census of bacteria in soil. *Plos Comput Biol.* 2006;2(7):786–793.

31. Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ. Exploring the diversity of complex metabolic networks. *Bioinform.* 2005;21(8): 1603–1609.

32. Li CH, Henry CS, Jankowski MD, Ionita JA, Hatzimanikatis V, Broadbelt LJ. Computational discovery of biochemical routes to specialty chemicals. *Chem Eng Sci.* 2004;59(22-23):5051–5060.

33. Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G. Computer-aided synthesis of biochemical pathways. *Biotechnol Bioeng.* 1990;36(11):1119–1132.

34. Broadbelt LJ, Stark SM, Klein MT. Computer-generated pyrolysis modeling - on-the-fly generation of species, reactions, and rates. *Ind Eng Chem Res.* 1994;33(4):790–799.

35. Ugi I, Bauer J, Brandt J, Friedrich J, Gasteiger J, Jochum C, Schuber W. New applications of computers in chemistry. *Angew Chemie Int Ed.* 1979;18(2):111–123.

36. Finley SD, Broadbelt LJ, Hatzimanikatis V. Computational framework for predictive biodegradation. *Biotechnol Bioeng.* 2009;104(6):1086–1097.

37. Dai MH, Copley SD. Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by Sphingobium chlorophenolicum ATCC 39723. *App Environ Microbiol.* 2004;70(4):2391–2397.

38. Patnaik R, Louie S, Gavrilovic V, Perry K, Stemmer WPC, Ryan CM, del Cardayré S. Genome shuffling of Lactobacillus for improved acid tolerance. *Nature Biotechnol.* 2002;20(7):707–712.

39. Wang YH, Li Y, Pei XL, Yu L, Feng Y. Genome-shuffling improved acid tolerance and L-lactic acid volumetric productivity in Lactobacillus rhamnosus. *J Biotechnol.* 2007;129(3):510–515.

40. Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WPC, del Cardayre SB. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature.* 2002;415(6872):644–646.

41. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science.* 2006;314(5805):1565–1568.

42. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Chruch GM. Programming cells by multiplex genome engineering and accelerated evolution. *Nature.* 2009;460(7257):894–898.

43. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nature Biotechnol.* 2010;28(8):856–862.

44. Fernandez-Arrojo L, Guazzaroni ME, Lopez-Cortes N, Beloqui A, Ferrer M. Metagenomic era for biocatalyst identification. *Curr Opin Biotechnol.* 2010;21(6): 725–733.

45. Nicolaou SA, Gaida SM, Papoutsakis ET. Coexisting/coexpressing genomic libraries (CoGeL) identify interactions among distantly located genetic loci for developing

complex microbial phenotypes. *Nucleic Acids Res.* 2011;39(22):e152.

46. Nicolaou SA, Gaida SM, Papoutsakis ET. Exploring the combinatorial genomic space in Escherichia coli for ethanol tolerance. *Biotechnol J*. 2012;7(11):1337–1345.
47. Borden JR, Jones SW, Indurthi D, Chen YL, Papoutsakis ET. A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in Clostridium acetobutylicum involving non-coding RNAs and ribosomal RNA processing. *Metab Eng*. 2010;12(3):268–281.
48. Borden JR, Papoutsakis ET. Dynamics of genomic-library enrichment and identification of solvent tolerance genes for Clostridium acetobutylicum. *Appl Environ Microbiol*. 2007;73(9):3061–3068.
49. Zingaro KA, Papoutsakis ET. Toward a semisynthetic stress response system to engineer microbial solvent tolerance. *Mbio*. 2012;3(5):e00308-12. doi:10.1128/mBio.00308–12.
50. Zingaro KA, Terry Papoutsakis E. GroESL overexpression imparts Escherichia coli tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable patterns. *Metab Eng*. 2013;15(1):196–205.
51. Tracy BP, Gaida SM, Papoutsakis ET. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Curr Opin Biotechnol*. 2010;21(1):85–99.
52. Zhang FZ, Carothers JM, Keasling JD. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnol*. 2012;30(4):354–359.
53. Gredell JA, Frei CS, Cirino PC. Protein and RNA engineering to customize microbial molecular reporting. *Biotechnol J*. 2012;7(4):477–499.
54. Lynch MD, Warnecke T, Gill RT. SCALEs: multiscale analysis of library enrichment. *Nature Methods*. 2007;4(1):87–93.
55. Reyes LH, Almario MP, Winkler J, Orozco MM, Kao KC. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in Escherichia coli. *Metab Eng*. 2012;14(5):579–590.