Genome Scale Engineering Techniques for Metabolic Engineering

Abstract

Metabolic engineering has expanded from a focus on designs requiring a small number of genetic modifications to increasingly complex designs driven by advances in genome-scale engineering technologies. Metabolic engineering has been generally defined by the use of iterative cycles of rational genome modifications, strain analysis and characterization, and a synthesis step that fuels additional hypothesis generation. This cycle mirrors the Design-Build-Test-Learn cycle followed throughout various engineering fields that has recently become a defining aspect of synthetic biology. This review will attempt to summarize recent genome-scale design, build, test, and learn technologies and relate their use to a range of metabolic engineering applications.

1. Introduction

Metabolic engineering is concerned with the engineering of biological systems for the purpose of manipulating flux towards desired products. A central goal of the field is to develop *forward* engineering approaches that are driven by predictive models and associated theory. Since such approaches require both sufficient understanding to develop models and genetic engineering tools to construct and test model predictions, the history of the field has focused primarily on the modification of a small number of genes with clear links to a targeted pathway. Typical modifications include overexpression of rate-limiting steps in the pathway, introduction of heterologous genes, and/or removal of competing pathways. Efforts along these lines have proven successful in increasing production titers from a broad range of platform strains, with applications ranging from bulk chemicals (Song et al., 2013; Yang et al., 2014), biofuels (Choi et al., 2012; Jang et al., 2012), to pharmaceuticals (Martin et al., 2003; Paddon and Keasling, 2014) and food derivatives (Kaur et al., 2014), among others

In the last decade, metabolic engineering has shifted from designs targeting a handful of genes with close metabolic network relationships to increasingly complex designs requiring the modification of dozens of genes spanning a broad range of metabolic functions (transporters, pathway enzymes, tolerance genes, etc.). To support this increased engineering complexity, metabolic engineering can now be generally defined by the use of iterative cycles of rational genome modification, systems level characterization, and sophisticated analysis. This approach

mirrors the Design-Build-Test-Learn (DBTL) cycle from the computational and engineering sciences (Figure 1). Here, we review applications and successes of genome scale engineering techniques for metabolic engineering based on the DBTL concept that link i) pathway design algorithms with active machine learning, ii) next-generation DNA synthesis and assembly with genome-engineering, and iii) laboratory automation with ultra-high throughput and sensitive genomics methods.

2. Pathway design algorithms with active machine learning

Conventional "design" typically involves a combination of literature searching, metabolic modeling, and heuristics. This design approach has limited throughput, where typically only a handful of designs are considered in depth. Recently, our understanding of microbial metabolism has greatly increased with accumulating bio-information on gene functions (Kan et al., 2012), genome structures (Lam et al., 2012), biological pathways (Peralta-Yahya et al., 2012), metabolic and regulatory networks (Gerosa and Sauer, 2011), and evolution of genomes (Blount et al., 2012). This knowledge makes it possible for the DBTL design to provide the complete set of build instructions for any target molecule, enabling rapid discovery of pathway configurations for reliable target molecule production.

Computational algorithms such as constraint-based flux balance analysis (FBA) are essential tools to predict phenotypic properties in genome scale modeling, which was widely used in different model strains. As an example, *E. coli*'s genome-scale metabolic network models have been updated over 20 years (McCloskey et al., 2013; Orth et al 2011). These databases are critical to improve the accuracy of the prediction of cellular phenotypes. More than 100 genome-scale metabolic network models were constructed for a wide range of different microorganisms, including *Saccharomyces cerevisiae* (Förster et al., 2003), *Corynebacterium glutamicum* (Shinfuku et al., 2009), *Mannheimia succiniciproducens* (Kim et al., 2007), *Bacillus subtilis* (Henry et al., 2009), *Clostridium acetobutylicum* (Lee et al., 2008), *Clostridium beijerinckii* (Milne et al., 2011), *Lactococcus lactis* (Flahaut et al., 2013), *Pichia pastoris* (Sohn et al., 2010), *Pseudomonas putida* (Puchałka et al., 2008), and so on. Recently, the ensemble modeling (EM) approach has shown promise in capturing kinetic and regulatory effects in the modeling of metabolic networks in comparison to FBA (Tran et al., 2008). It can simultaneously consider alternative model structures and parameter sets, such as identifying genetic/enzyme perturbations to minimize the number of models retained in the ensemble after each round of model screening (Zomorrodi et al., 2013). Ensemble Modeling for Robustness

Analysis (EMRA), which combines a continuation method with the Ensemble Modeling approach, can be used for investigating the robustness of non-native pathways. By comparing possible designs of two nonnative pathways (non-oxidative glycolysis and reverse glyoxylate cycle), EMRA resulted in the selection of targets for flux improvement by considering both performance and robustness (Lee et al., 2014).

A number of algorithms based on the above genome-scale models have been developed to identify network manipulation strategies while predicting their system-wide effects (Table 1). OptKnock (Choon et al., 2014) is one popular computational algorithm, capable of suggesting gene deletion strategies that lead to the overproduction of a target metabolite. A nested optimization framework identifies gene deletions targets considering both the production of the desired compounds and biomass formation. OptKnock was applied to develop strategies for the metabolic engineering of *E. coli* for the production of 1,4-butanediol (BDO), leading to a strain capable of producing 18 g/L BDO from renewable carbohydrate feedstocks. Beyond gene knockouts, the design of strains involving overexpression and down-regulation have also been shown to enhance biochemical production by computational algorithms. OptForce contrasts the metabolic flux patterns observed in a parent strain and a strain overproducing the chemical at the targeted yield (Ranganathan et al., 2010). By applying the OptForce algorithm, the effect of redirecting malonyl-CoA flux towards resveratrol production was evaluated, and shake flask experiments yielded 1.6 g/L of resveratrol without the need of using expensive inhibitors of fatty acid metabolism (Bhan et al., 2013).

Every predicted mutation should be associated with a specific design and measured effect on metabolism. However, to fully learn microbial metabolism and its responses to environmental factors, it is necessary to functionally characterize and accurately quantify all levels of gene products, mRNAs, proteins and metabolites, as well as their interaction. These requirements led to the generation of omics platform techniques, such as transcriptomics (Sorek and Cossart, 2010), proteomics (Otto et al., 2012), metabolomics (Hou et al., 2012) and interactomics (Janga et al., 2011). However, these techniques also generate a substantial amount of data that is hard to process and analyse for functional patterns. Several tools for Omics data analysis have been developed, such as GIMME(Becker and Palsson, 2008), E-Flux(Colijn et al., 2009), TIGER(Jensen et al., 2011), GIMMEp(Bordbar et al., 2012) (Table 1). GIMME produces a guaranteed functional metabolic model specific to transcriptomics data and quantifies the agreement between gene expression data and one or more metabolic objectives, which can be used for adaptive evolution of bacteria and rational design of metabolic

engineering strains. Furthermore, by integrating proteomics and metabolomics data, GIMMEp and GIM^{3}E methods were developed based on the GIMME.

Machine-learning methods, instead, seek to use intrinsic data structure, as well as the expert annotations of biologists to infer models that can be used to solve versatile data analysis tasks. (Domingos, 2012). The process of Machine learning can be seen as two phases. In the preparational phase, the data of model samples is used to build the computer system for learning the relationship between the data and the target phenotype. Then, this system can be used to predict the desired trace in larger scale screening. (Hastie et al., 2009; de Ridder et al., 2013). Machine learning has been successfully applied in different biological fields, such as high-content screening (Collinet et al., 2010; Mercer et al., 2012), drug development (Castoreno et al., 2010; Murphy, 2011), DNA sequence analysis (Ben-Hur et al., 2008) and proteomics (Datta and Pihur, 2010; Reiter et al., 2011). In a DBTL-cycle, Active learning refers to a class of semi-supervised machine learning techniques in which a data-driven algorithm iteratively selects new sets of pathway designs to evaluate in order to quickly learn a mathematical function that relates pathway design to pathway performance (King et al., 2009, 2004). This process can build accurate predictive models while minimizing the required number of expensive and often difficult-to-obtain data points. The genome design landscape is combinatorial in the number of possible pathway components, and thus far too large to ever be searched exhaustively. If machine learning can be iteratively used to ensure efficient modeling and optimization of engineered pathways, successful designs for a targeted molecule production can be easily and efficiently achieved.

3. Genome-Scale Building Techniques

Genome engineering has arisen primarily as a combination of advances in technologies for cost-efficient DNA synthesis and genome-editing. Compared to the traditional gene and pathway-level engineering, genome-scale approaches can increase the throughput at which we can search the sequence space for improved phenotypes. This is essential when dealing with complex traits involved in industrial environment robustness, like tolerance to osmotic stress, pH and product tolerance.

Although genome-scale modifications have been employed for decades in random mutagenesis and directed evolution studies, these tools were employed to search and understand the genetic basis of complex phenotypes (*reverse* engineering). The new toolsets allow *forward* engineering at the genome-scale, establishing the basis for an entirely new focal

point in metabolic engineering. In this section, we review important techniques that have enabled metabolic engineering at the genome scale.

3.1. Assembly of DNA constructs

An important breakthrough for genome engineering was the development of methods for rapid assembly of DNA fragments, replacing the laborious cut and paste approaches and increasing the throughput at which we can design and test designs. This idea was pioneered by Stemmer in a work that relied on 40 bp oligonucleotides to assemble entire plasmids (Stemmer et al., 1995). Modern assembly methods can be divided in two broad categories: homology-based methods and restriction-based methods (Figure 2). They all support combinatorial DNA assembly, an important feature for library construction in genome-wide engineering studies.

Homology-based approaches offer the advantages of scarless multi-part DNA assembly that relies on short (20-40 bp) overlaps between the fragments. Since the structural features (homology overlap) are the same for all these approaches, the assembly can be performed using most of these methods without any redesign required. Further, the lack of requirement for restriction sites makes it highly sequence-independent.

Gibson assembly is a gold standard in this category, being successfully applied to assemble 133-166 kb fragments of a bacterial genome (Gibson et al., 2008) and the entire 16.3 kb mouse mitochondrial genome from 60-mers oligonucleotides (Gibson et al., 2010). The assembly is a one-pot isothermal reaction that involves three enzymatic reactions. T5 exonuclease catalyzes a 5’ strand resection to create 3’ overhangs. The exposed 3’ overhangs can anneal to other exposed ends of fragments with homologous sequences, priming each other so that Phusion polymerase can fill the gap. Finally, Taq ligase seals the nick to generate the assembled product (Gibson et al., 2009).

Circular Polymerase Extension Cloning (CPEC) consists of an alternative method in which fragments can be cloned into a vector by short PCR-like reactions. The method employs cycles of heating to denature the DNA pieces, cooling to anneal homologous sequences and polymerase extension to stitch the pieces together (Quan and Tian, 2011, 2009). The generated nicked product can then be sealed *in vivo*.

Homology-based assembly can also be accomplished *in vivo* by relying on native cell homologous recombination machinery. Transformation-associated recombination (TAR) in *Saccharomyces cerevisiae* is a key method in this category, with applications ranging from simple cloning of small inserts to the assembly of an entire eukaryotic chromosome (Annaluru et

al., 2014) or a bacterial genome (Gibson et al., 2008) from overlapping pieces. In this approach, yeast is transformed simultaneously with all pieces to be assembled, each containing proper homology overlaps (Kouprina and Larionov, 2008; Shao et al., 2009). The assembled product usually contains a selectable marker for isolation of cells that successfully assembled the construct. TAR has also been employed in *E. coli*, relying on the recombination functions of the prophage RecET (Zhang et al., 2000) or the Redαβ proteins of the lambda phage (Sharan et al., 2009). Although functionally similar, the RecET proteins seems to work mechanistically different from Redαβ, with the former performing better on linear-linear recombination. Therefore, direct cloning using the RecET system have been reported for fragments up to 52kb, comprising an important tool for bioprospecting (Fu et al., 2012). A similar approach relies on bacterial cell extracts rather than whole cells, termed seamless ligation cloning extract or SLiCE (Zhang et al., 2012).

Recently, a new approach termed Ligase Cycling Reaction (LCR) was characterized as being faster, cheaper and more convenient than other homology-based DNA assembly methods (Kok et al., 2014). Using optimized conditions, LCR enabled rapid assembly of up to 20 kb constructs from multiple parts, outperforming the efficiency of CPEC or Gibson Assembly. This method relies on bridging oligos that can anneal to complementary ends of two DNA parts to be assembled. After denaturation and annealing, the two DNA parts brought together by the bridging oligo are ligated using a thermostable DNA ligase. Multiple cycles of denaturation-annealing-ligation allow assembly of complex DNA constructs from multiple parts.

Despite the usefulness of homology-based approaches, some sequences can be challenging to assemble using these methods. Sequences that have stable secondary structure in the termini (such as hairpins) would preclude proper annealing/priming with the correct template. Additionally, sequences with repeated regions can result in inaccurate assembly due to incorrect annealing between the repeated homologous sequences. In these cases, restriction-based methods, which do not rely on homologous recombination, are often employed.

The first attempt to facilitate restriction-based approaches in order to increase the throughput of cloning was introduced by Biobricks (Shetty et al., 2008). The idea is to standardize biological parts by flanking them with the same set of restriction sites in the 5’ and 3’. By utilizing enzymes that recognize different sites but generate the same single strand overhangs, this approach allows modular assembly of parts in a way that the restriction sites are retained in the 5’ and 3’ termini of the final product, but lost in between the assembled parts. This allows recycling of the restriction sites and continued assembly to larger and larger products.

Biobricks allows assembly of up to three parts at a time, working in a hierarchical stepwise fashion, which can be time-consuming for large-size constructions. Golden Gate assembly alleviated this issue by employing Type IIS restriction endonucleases, which recognizes specific sequences but then cut at a defined distance away from the recognition site. Therefore, it is possible to generate user defined overhangs such that simultaneous assembly of multiple fragments in a defined order is possible (Engler et al., 2008).

With this broad range of DNA assembly tools available, genome engineers can choose the method that fits best to the DNA fragments that will be utilized. While homology-based approaches allow easy multi-part assembly, some sequences can be troublesome. On the other hand, restriction-based approaches can allow assembly of difficult templates but can result in scars and require the restriction site to beabsent in the assembled parts. A summary of the strengths and limitations of each discussed method is listed in Table 2. Ultimately, the method of choice will depend on the nature of the DNA fragments to be assembled.

3.2. Recombination-based genome-wide engineering strategies

Although *in vitro* assembly provides a powerful tool for metabolic engineering, realization of this advance requires an ability to insert such *in vitro* assembled DNA into the host chromosome and similarly to modify the host at additional locations as is required in almost all metabolic engineering designs. Several methods exist to recombine DNA parts into a host chromosome, but only a few of them present the efficiency and throughput required for genome-scale applications (Figure 3).

The first recombination machinery that was explored for genome-wide engineering purposes was the lambda red system in *E. coli*. This system comprises of three proteins that by themselves are able to integrate linear DNA fragments into the *E. coli* genome: gam (Redγ), exo (Redα) and bet (Redβ). Gam prevents the *E. coli* exonuclease RecBCD from degrading the DNA fragment, Exo is a 5’ to 3’ dsDNA exonuclease that generates ssDNA intermediates, and Bet promotes annealing of the ssDNA intermediate with exposed ssDNA in the host genome (Sharan et al., 2009). Since Bet cannot perform strand invasion, the current proposed mechanism suggests that the ssDNA intermediate is annealed to the lagging strand of the open replication fork, replacing an okasaki fragment. (Mosberg et al., 2010).

The introduction of the lambda red-assisted recombination facilitated engineering efforts in the *E. coli* chromosome, establishing the term “recombineering” (recently reviewed by Pines et al., 2015a). This allowed for the first time single-step gene integration and inactivation from

linear dsDNA PCR products (Datsenko and Wanner, 2000; Murphy, 1998). However, the low efficiency of gene-size modifications requires the integration of selectable markers, limiting the applicability to genome-wide endeavors. Small modifications, on the other hand, can be directly encoded in oligonucleotides and can be integrated with enough efficiency to completely eliminate the requirement of selection (Sharan et al., 2009). This idea was scaled up to the genome level in an approach termed Multiplex Automated Genome Engineering or MAGE (Wang et al., 2009).

MAGE leveraged array-based DNA synthesis with cyclic rounds of transformation, using a pool of oligos to continuously introduce diversity in a set of user-defined targets. This strategy was successfully employed to optimize lycopene production by targeting 24 genomic sites simultaneously (Wang et al., 2009). The large multiplex capability of MAGE was also demonstrated by the replacement of all 314 TAG stop codons with synonymous TAA codons in *E. coli* (Isaacs et al., 2011). This same strategy of oligo-mediated recombination was also employed in *Saccharomyces cerevisiae* in a similar approach termed YOGE (DiCarlo et al., 2013a). Despite the simplicity behind MAGE, this strategy requires careful oligo designs to optimize the recombineering efficiency of each variant in the library. Among the design considerations, secondary structure and the position of the mutation can influence recombineering efficiency. Optimal designs of MAGE oligos can be rapidly accomplished using a computational tool termed MAGE Oligo Design Tool or MODEST (Bonde et al., 2014).

MAGE is able to generate diversity at defined targets in a multiplex fashion, enabling metabolic engineers to address a broad range of applications such as metabolic flux optimization and protein engineering. Importantly, MAGE requires a priori knowledge of the engineered targets, which are not known for many complex traits such as tolerance and toxicity. Trackable Multiplex Recombineering (TRMR) addresses this need by efficiently mapping the genetic basis of targeted traits. TRMR similarly takes advantage of recombineering technology and multiplex oligo-synthesis to integrate a pool of barcoded dsDNA constructs with the end goal of up-regulating or down-regulating every gene in the *E. coli* genome (Warner et al., 2010). These libraries are subjected to growth selections, and promoter mutations that affect fitness are identified in parallel from the barcodes. This results in multiplex tracking of trait-defining genes, which was successfully employed to identify novel targets for cellulosic hydrolysate tolerance and growth in inhibitors such as methylglyoxal. Other approaches also contributed on the search for trait-related genes by assessing the impact of gene disruptions in a genome-wide scale. Bar-seq leveraged deep sequencing technologies to sequence barcodes in a yeast deletion collection, effectively counting barcodes as a proxy for the fitness effect of that gene

(Smith et al., 2009). Similarly, Tn-seq relied on a saturated transposon insertion library to disrupt genes and measure its fitness contribution by sequencing the transposon flanking regions (van Opijnen et al., 2009).

Oligo-mediated recombination, although remarkably useful for genome engineering as demonstrated by MAGE and TRMR, does not support larger size modifications. Common metabolic engineering endeavors require importing heterologous functions encoded across a number of genes. Such larger dsDNA based modifications cannot be inserted at the same efficiency and throughput as smaller ssDNA-oligo based modifications. Integrating larger size cassettes is even more troublesome for non-recombinogenic organisms like *E. coli*, since the gold standard lambda red system does not appear to support by itself efficient recombination for fragments larger than 2500 bp (Kuhlman and Cox, 2010).

Alternative approaches and recombination machineries have been explored to overcome this limitation in *E. coli*. A method introduced by Kuhlman and Cox in 2010 combined the recombination functions of lambda red system with precise double strand breaks introduced by I-SceI endonuclease in the chromosome and donor plasmids. The generated double strand breaks stimulated repair mechanisms that, associated with lambda red, resulted in efficient integration of a 7 kb fragment. Recombinase-assisted genome engineering (RAGE) relied on a different recombination machinery (Cre recombinase) to introduce a 34 kb alginate metabolism pathway into the *E. coli* genome (Santos et al., 2013). Importantly, RAGE allowed testing of the integrated pathway in multiple hosts, chromosomal position and varying copy number, developing a robust *E. coli* strain that produces ethanol directly from brown macroalgae.

Despite the achievements of these alternative approaches in integrating larger size DNA parts, the requirement of multiple steps still limits their applicability to genome-wide engineering. The field is in need of technologies that would allow rapid integration of multiple genes and pathways in a multiplex or recursive fashion in order to quickly test DNA fragments for desired functions. Development of new technologies based on the CRISPR-Cas system might contribute significantly toward this goal, as described on the next section.

3.3. CRISPR-assisted tools for rational genome engineering

Homologous recombination founded the basis for the rational genome engineering field, however, the efficiency of most systems remains too low for supporting forward engineering at the genome scale. Moreover, repair and segregation further reduce the recombination efficiency (Reynolds and Gill, 2015). Small mutations can be integrated with reasonable efficiencies

(Wang et al., 2009), which can be improved to efficiencies as high as 70% by optimizing the oligo designs to avoid the mismatch repair system (Sawitzke et al., 2011). However, larger gene-size integrations occur at efficiencies on the order of 10^{-4} per viable cell in *E. coli* (Murphy, 1998). The realization that introducing a double strand break in the genome significantly increases homologous recombination has directly addressed this issue.

Repair of double strand breaks is vital for genome stability, a process that is accomplished by two main pathways in most organisms. Non-homologous end joining (NHEJ) combines the break ends to repair the lesions, usually introducing indels of various lengths. Homology-directed repair (HDR) is a mechanistically distinct process that uses a template DNA with proper homology regions to repair the lesion. As a result, precise genomic modifications can be engineered by generating a double strand break at the genomic site of choice and providing a template DNA containing the desired modifications (point mutations or whole genes) flanked by homology arms encompassing the breakpoint.

Initial studies in mouse using the yeast endonuclease I-SceI determined that the efficiency of homologous recombination upon a double strand break increased by 2 orders of magnitude (Choulika et al., 1995; Rouet et al., 1994). This finding stimulated interest to develop tools that allows generation of double strand breaks at user-defined sites. The first generation of programmable endonucleases focused primarily on zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs).

ZFNs are engineered fusions of zinc-finger domains with the type IIS restriction enzyme FokI. Zinc-finger are the most abundant DNA-binding motif in eukaryotic cells, requiring approximately 30 amino acids to recognize 3 base pairs of DNA (Pavletich and Pabo, 1991). Therefore, assembling arrays of zinc-finger domains can recognize extended sequences with custom specificity. Similarly, TALENs are engineered fusions of TALE domains with the FokI endonuclease. However, TALE domains offer the advantage that the DNA recognition code is completely known (Boch et al., 2009; Joung and Sander, 2013).

ZFNs and TALENs have been employed for genome engineering purposes in several organisms (Gaj et al., 2013), but the requirement of reengineering the entire protein for every targeted site limit its applicability. This limitation was overcome by the newest generation of customized endonucleases, based on the bacterial adaptive immune system CRISPR (Sorek et al., 2013). CRISPR (clustered regularly interspaced short palindromic repeats) immune response is based on endonucleases whose specificity is determined by short RNA molecules. Therefore, these sets of endonucleases can be reprogrammed to specific sites simply by providing a synthetic guide RNA (gRNA), a significant improvement over ZFNs or TALENs.

Since the first demonstrations of these RNA-guided nucleases for genome engineering purposes (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013), their use expanded to multiple applications. Cas9, the nuclease from the type II system found in *Streptococcus pyogenes*, is the most extensively used to date. Cas9 cleavage requires the presence of a 5'-NGG-3' consensus sequence immediately downstream of the target site, which is called PAM ("protospacer adjacent motif"; (Mojica et al., 2009). The recent crystal structure of Cas9 identified the structural features of PAM recognition, establishing the basis for engineering this enzyme with distinct PAM requirements (Anders et al., 2014).

Successful CRISPR-based genome engineering applications are dependent on proper design of gRNAs. Identifying spacers adjacent to the PAM sequence in the correct orientation, abrogating further CRISPR cleavage after repair and minimizing off-target gRNA binding are among important design considerations. A number of computational tools are now available to aid users in correctly designing gRNAs for different targeted genomes (Aach et al., 2014; Bae et al., 2014; Heigwer et al., 2014).

Cas9 have been used for engineering the genome of a broad range of organisms, including bacteria (Jiang et al., 2013), yeast (DiCarlo et al., 2013b), plants (Gao and Zhao, 2014; Shan et al., 2013), mouse (Cong et al., 2013) and human cell lines (Mali et al., 2013). In *E. coli*, CRISPR-based genome editing also offers the advantage of working as a selection mechanism by itself, on account of this organism being inefficient in NHEJ (Dillingham and Kowalczykowski, 2008). As a result, cells that do not integrate a provided repair template for HDR should not survive, selecting for cells that correctly edited the genome. Importantly, the provided template for HDR must contain modifications that either eliminates the spacer or PAM sequence, preventing further Cas9 cleavage (Figure 3). A recent study in *E. coli* employed PCR products as HDR templates for codon saturation mutagenesis, containing PAM mutations outside of the coding region and non-synonymous mutations inside the coding region. They reported 99.7% PAM mutation efficiency and 81% efficiency for the codon mutations (Pines et al., 2015b).

In organisms proficient in NHEJ, engineering precise modifications rely on the balance between the NHEJ and HDR repair pathways. As a result, most applications so far relied on the NHEJ pathway for indels introduction at the breakpoint, efficiently knocking out affected genes. In these organisms, there is no selection for HDR template integration, since NHEJ repair will allow cell survival. Precise HDR modifications thus require selectable markers or strategies to favor HDR over NHEJ. Recent methods increased the efficiency of HDR in such organisms by inhibiting the NHEJ pathway (Chu et al., 2015; Maruyama et al., 2015).

The ability of CRISPR-induced HDR have also been explored for gene-size modifications. In *E. coli,* it was recently reported multigene editing of three targets, including deletions and insertions up to 4.5 kb (Jiang et al., 2015). In yeast, a recent method named CasEMBLR leveraged *in vivo* assembly of DNA parts (as described in section 3.1) with Cas9-assisted genome editing to introduce entire pathways from DNA parts (Jakočiu□nas et al., 2015). The authors successfully implemented a carotenoid pathway and engineered a tyrosine production strain in a single-step. This represents a substantial step-forward for genome-scale pathway engineering. In mice, fluorescent reporters were also integrated using CRISPR (Yang et al., 2013).

Shortly after the use of Cas9 as a programmable endonuclease was published, a variant harboring two point mutations in the RuvC1 and HNH nuclease domains was used to provide an easily programmable and broadly applicable gene regulation platform (Qi et al., 2013). This catalytically inactive or “dead” Cas9 (dCas9) was still able to bind to its target site, yet no cleavage occurred. By designing gRNAs that would direct dCas9 to promoter regions, strong gene repression was achieved. This allows easy repression of multiple genes when desired, which was called CRISPR interference (CRISPRi). This idea was recently applied in *E. coli* to adjust the carbon flux in a polyhydroxyalkanoate (PHA) producing strain and manipulate the content of 4-hydroxybutyrate (4HB) in the polymer (Lv et al., 2015). In addition to CRISPRi, dCas9 was fused to different domains and proved to work in versatile applications, such as gene activation (Bikard et al., 2013) and epigenome editing (Hilton et al., 2015).

In addition to Cas9, genome engineering has also explored the use of the endonuclease from the CRISPR type I system. This system involves a multisubunit complex termed CRISPR-associated complex for antiviral defense (CASCADE), which is guided by CRISPR RNAs (crRNAs) to a complementary DNA sequence. Upon CASCADE recognition of target dsDNA, the endonuclease Cas3 is recruited for degradation of the target (Gong et al., 2014; Sorek et al., 2013). CASCADE was used in *E. coli* as a gene regulation tool in a strain not expressing the endonuclease Cas3 (Rath et al., 2015). The authors were able to efficiently silence gene expression using crRNAs that guide the CASCADE complex to promoter regions, in the same way as shown for dCas9.

The combination of tools described above sets the stage for forward genome-scale approaches to engineering metabolism. Precise genome edits can be easily accomplished using a variety of tools, summarized in Table 3. Flux can now be easily manipulated by employing arrays of gRNAs with dCas9/CASCADE as well as *in vitro* assembled constructs

containing various heterologous functions. Such increasing capabilities should expand the range of applications that can be addressed by metabolic engineering.

4. High-throughput characterization of built strains

Studying and learning from populations generated by genome-wide engineering strategies requires approaches to connect genotypes to phenotypes. To do this, the first step is to test the performance of different designs enabled by build techniques. This testing can be accomplished using screening and selection strategies which subsequently narrow library diversity to a handful of winners. The testing strategies are dependent on the engineering objective function. Ideally, an engineering objective which can be tied to survival will enable simultaneous testing of over 10^8 designs. When selections are unavailable, screening techniques using different readouts can be performed at lower throughputs. Libraries generated using genome-wide recursive multiplexed engineering approaches have been screened for lycopene production using colorimetric quantification, selected for tolerance to small molecules (hydrolysate and acetate) using inhibitory assays and selected for production of small molecules (naringenin and glucaric acid) via a sensor-reporter riboswitch system (Raman et al., 2014; Sandoval et al., 2012; Wang et al., 2009). Many options exist to test libraries and thorough reviews on the topic are available (Dietrich et al., 2010; Zhang and Keasling, 2011).

Tying the results of a screen or selection to a design (genotype) of interest can be performed by either serial or parallel approaches. When the results of testing reduce library diversity to a few winners, serial sequencing approaches (Sanger sequencing) are effective at analyzing targeted modifications in single genotypes with one sequencing reaction per modified site per genotype. Since the number of sequencing reactions scales proportionally with the number of genotypes assessed and number of sites assessed, this approach is limited towards analyzing 10^2 genotypes (Sandoval et al., 2012; Wang et al., 2009). Digital sequencing approaches like multiplexed allele specific PCR have also been used to improve the throughput of genotyping approaches (Ruano and Kidd, 1989; Wang and Church, 2011; Wang et al., 2012). With allele specific PCR, primers are specifically designed to amplify (or not amplify) sequences containing mutations. Although this approach can assess the presence of a mutation, it is limited towards binary solutions where a mutation either exists or does not exist. Other genotyping approaches such as full-genome sequencing using both shotgun sequencing or optical mapping approaches can be used to sequence genomes(Alkan et al., 2011; Metzker, 2010). These approaches can provide an evaluation of the full genotype of organisms in series

(or several organisms in parallel) and have been used with genome editing approaches to evaluate the combinations of mutations that occur in engineered or evolved populations (Minty et al., 2011; Raman et al., 2014).

Parallelized genotyping efforts enable the quantitative genotyping of a library. The ability to quantify mutations that exist in a population enable both simultaneous detection of the effect and composition of mutations. For example, genotypes can be quantitatively identified in parallel using a competitive hybridization microarray assays (Lynch et al., 2007; Warner et al., 2010). High-throughput deep-sequencing approaches can be used to assess mutations in populations by parallelizing sequencing by reading on the order of 10^7 sequences. When coupled to a screen or a selection, high-throughput sequencing has been used to engineer proteins, study the connection between codon usage and gene expression and also map the mutation landscape of an entire gene (Acevedo et al., 2014; Firnberg et al., 2014; Fowler and Fields, 2014; Goodman et al., 2013). Since most sequencing technologies are limited to short and continuous read lengths (~10^2-10^3 bp) accommodations are needed to assess distal mutations that can occur over the entire genome (10^6-10^9 bp) as is inherent to emerging genome engineering approaches. A technique, TRACE, was introduced which enables the high-throughput tracking of combinations of mutations in a population. With TRACE, sites of interest are assembled into a single construct using an emulsion multiplexed PCR approach (Zeitoun et al., 2015). Each assembled construct covalently links and condenses targeted genomic sites, subsequently enabling a single sequencing read to contain combinatorial mutation information that occur on a genome with single-cell resolution. This is compatible with both high- and low-throughput sequencing technologies allowing for thousands of combinatorial mutants to be identified in a single high-throughput sequencing experiment. In the original report, up to 10-sites were assembled in one TRACE reaction, thereby improving the throughput of analyzing mutants by Sanger sequencing by an order of magnitude. With high-throughput sequencing, up to 6 sites were assembled in a single TRACE reaction to assess the combinatorial mutation landscape of a MAGE library targeting the ribosome binding site of several membrane genes previously implicated in alcohol tolerance. Other approaches for overcoming short read sequencing technologies can include multiplexed single cell barcoding approaches (Craig et al., 2008; Smith et al., 2010). In this case, unique barcodes are ligated to amplicons or DNA at a single genotype level. This allows for the full genome, or targeted sequencing of many sites at once for hundreds to thousands of cells in a population.

In all of these discussed sequencing approaches, the sequenced site must be known prior to sequencing. In addition, noise occurring in the system through off-target random

mutations are always present and although transient expression of the mismatch repair system can reduce off-target mutations, they still influence the final strain, particularly in cases where many rounds of targeted mutagenesis are necessary towards engineering organisms (Nyerges et al. 2014). Regardless of these limitations, the depth and redundancy of high-throughput sequencing approaches can be directly used with multiplex genome editing approaches (Sims et al., 2014). As third generation sequencing technologies become ubiquitous and full-genome sequencing becomes less expensive, many of the problems of off-target mutations can be alleviated.

5. Conclusion

Recent metabolic engineering applications have demonstrated an increased emphasis on more complex genome-scale designs, which require methods to build and characterize strains at higher throughputs. In this review, we summarized recent tools that have facilitated metabolic engineering at the genome-scale. We related the metabolic engineering framework to the DBTL engineering cycle and used this to summarize technologies for i) Design. Computational tools are now available for genome-scale reconstruction, pathway prediction and optimization, and omics data analysis, being successfully employed for the metabolic engineering of several organisms to produce a range of industrially-relevant compounds. ii) Build. DNA assembly methods coupled to genome-scale engineering strategies enable generation of diversity in multiple levels across the genome, expanding the sequence space to search for improved phenotypes. iii) Test. Genome-scale tracking approaches enables parallel characterization of individuals generated at the population-level, efficiently tying improved phenotypes to the designed genotypes. iv) Learn. Machine learning algorithms allow quick identification of pathway designs that correlates with improved performance, fueling more DBTL cycles.

We expect to see a broad range of future applications from metabolic engineers that are enabled by such technologies. However, the field is still in need of technologies to fuel more comprehensive and reliable design cycling, including: i) Methods to quickly integrate and test multiple pathway libraries, ii) Multi-component genetic circuits with predictable and reliable performance and iii) Pathway construction for complicated chemistries, not normally encompassed by microbial metabolism. Realization of these and related advances should push forward creative new applications of metabolic engineering principles in a range of areas.

7. References

Aach, J., Mali, P., Church, G.M., 2014. CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. bioRxiv 005074.

Acevedo, A., Brodsky, L., Andino, R., 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505, 686–690.

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., Nielsen, J., 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput. Biol. 9, e1002980.

Alkan, C., Coe, B.P., Eichler, E.E., 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12, 363–376.

Anders, C., Niewoehner, O., Duerst, A., Jinek, M., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569–573.

Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., et al., 2014. Total synthesis of a functional designer eukaryotic chromosome. Science 344, 55–58.

Bae, S., Park, J., Kim, J.-S., 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473–1475.

Bates, J.T., Chivian, D., Arkin, A.P., 2011. GLAMM: Genome-Linked Application for Metabolic Maps. Nucleic Acids Res. 39, W400–5.

Becker, S.A., Palsson, B.O., 2008. Context-specific metabolic networks are consistent with experiments. PLoS Comput. Biol. 4, e1000082.

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., Rätsch, G., 2008. Support vector machines and kernels for computational biology. PLoS Comput. Biol. 4, e1000173.

Bhan, N., Xu, P., Khalidi, O., Koffas, M.A.G., 2013. Redirecting carbon flux into malonyl-CoA to improve resveratrol titers: Proof of concept for genetic interventions predicted by OptForce computational framework. Chem. Eng. Sci. 103, 109–114.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., Marraffini, L.A., 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 41, 7429–7437.

Blount, Z.D., Barrick, J.E., Davidson, C.J., Lenski, R.E., 2012. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489, 513–518.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509–1512.

Bonde, M.T., Klausen, M.S., Anderson, M.V., Wallin, A.I.N., Wang, H.H., Sommer, M.O.A., 2014. MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering. Nucleic Acids Res. 42, W408–15.

Bordbar, A., Mo, M.L., Nakayasu, E.S., Schrimpe-Rutledge, A.C., Kim, Y.-M., Metz, T.O., Jones, M.B., Frank, B.C., Smith, R.D., Peterson, S.N., Hyduke, D.R., Adkins, J.N., Palsson, B.O., 2012. Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. Mol. Syst. Biol. 8, 558.

Broeckling, C.D., Reddy, I.R., Duran, A.L., Zhao, X., Sumner, L.W., 2006. MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. Anal. Chem. 78, 4334–4341.

Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., Mittag, F., Schubert, M., Glont, M., Golebiewski, M., van Iersel, M., Keating, S., Rall, M., Wybrow, M., Hermjakob, H., Hucka, M., Kell, D.B., Müller, W., Mendes, P., Zell, A., Chaouiya, C., Saez-Rodriguez, J., Schreiber, F., Laibe, C., Dräger, A., Le Novère, N., 2013. Path2Models: large-scale generation of computational models from biochemical pathway maps. BMC Syst. Biol. 7, 116.

Campodonico, M.A., Andrews, B.A., Asenjo, J.A., Palsson, B.O., Feist, A.M., 2014. Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path. Metab. Eng. 25, 140–158.
Carbonell, P., Parutto, P., Baudier, C., Junot, C., Faulon, J.-L., 2014a. Retropath: automated pipeline for embedded metabolic circuits. ACS Synth. Biol. 3, 565–577.
Carbonell, P., Parutto, P., Herisson, J., Pandit, S.B., Faulon, J.-L., 2014b. XTMS: pathway design in an eXTended metabolic space. Nucleic Acids Res. 42, W389–94.
Castoreno, A.B., Smurnyy, Y., Torres, A.D., Vokes, M.S., Jones, T.R., Carpenter, A.E., Eggert, U.S., 2010. Small molecules discovered in a pathway screen target the Rho pathway in cytokinesis. Nat. Chem. Biol. 6, 457–463.
Chandrasekaran, S., Price, N.D., 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 107, 17845–17850.
Choi, Y.J., Park, J.H., Kim, T.Y., Lee, S.Y., 2012. Metabolic engineering of Escherichia coli for the production of 1-propanol. Metab. Eng. 14, 477–486.
Choulika, A., Perrin, A., Dujon, B., Nicolas, J.F., 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol. Cell. Biol. 15, 1968–1973.
Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., Kühn, R., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. doi:10.1038/nbt.3198
Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.-Y., Moody, D.B., Murray, M., Galagan, J.E., 2009. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. PLoS Comput. Biol. 5, e1000489.
Collinet, C., Stöter, M., Bradshaw, C.R., Samusik, N., Rink, J.C., Kenski, D., Habermann, B., Buchholz, F., Henschel, R., Mueller, M.S., Nagel, W.E., Fava, E., Kalaidzidis, Y., Zerial, M., 2010. Systems survey of endocytosis by multiparametric image analysis. Nature 464, 243–249.
Collins, S.B., Reznik, E., Segrè, D., 2012. Temporal expression-based analysis of metabolism. PLoS Comput. Biol. 8, e1002781.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.
Cotten, C., Reed, J.L., 2013. Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. Biotechnol. J. 8, 595–604.
Craig, D.W., Pearson, J.V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J.J., Pawlowski, T.L., Laub, T., Nunn, G., Stephan, D.A., Homer, N., Huentelman, M.J., 2008. Identification of genetic variants using bar-coded multiplexed sequencing. Nat. Methods 5, 887–893.
Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. 97, 6640–6645.
Datta, S., Pihur, V., 2010. Feature selection and machine learning with mass spectrometry data. Methods Mol. Biol. 593, 205–229.
De Ridder, D., de Ridder, J., Reinders, M.J.T., 2013. Pattern recognition in bioinformatics. Brief. Bioinform. 14, 633–647.
DiCarlo, J.E., Conley, A.J., Penttilä, M., Jäntti, J., Wang, H.H., Church, G.M., 2013a. Yeast oligo-mediated genome engineering (YOGE). ACS Synth. Biol. 2, 741–749.
DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., Church, G.M., 2013b. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. 41, 4336–4343.

Dietrich, J.A., McKee, A.E., Keasling, J.D., 2010. High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu. Rev. Biochem. 79, 563–590.
Dillingham, M.S., Kowalczykowski, S.C., 2008. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. Microbiol. Mol. Biol. Rev. 72, 642–671.
Engler, C., Kandzia, R., Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One 3, e3647.
Engler, C., Marillonnet, S., 2014. Golden Gate cloning. Methods Mol. Biol. 1116, 119–131.
Firnberg, E., Labonte, J.W., Gray, J.J., Ostermeier, M., 2014. A comprehensive, high-resolution map of a gene's fitness landscape. Mol. Biol. Evol. 31, 1581–1592.
Flahaut, N.A.L., Wiersma, A., van de Bunt, B., Martens, D.E., Schaap, P.J., Sijtsma, L., Dos Santos, V.A.M., de Vos, W.M., 2013. Genome-scale metabolic model for Lactococcus lactis MG1363 and its application to the analysis of flavor formation. Appl. Microbiol. Biotechnol. 97, 8729–8739.
Fowler, D.M., Fields, S., 2014. Deep mutational scanning: a new style of protein science. Nat. Methods 11, 801–807.
Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F., Zhang, Y., 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat. Biotechnol. 30, 440–446.
Förster, J., Famili, I., Fu, P., Palsson, B.Ø., Nielsen, J., 2003. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Res. 13, 244–253.
Gaj, T., Gersbach, C.A., Barbas, C.F., 3rd, 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405.
Gao, Y., Zhao, Y., 2014. Specific and heritable gene editing in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 111, 4357–4358.
Gawand, P., Hyland, P., Ekins, A., Martin, V.J.J., Mahadevan, R., 2013. Novel approach to engineer strains for simultaneous sugar utilization. Metab. Eng. 20, 63–72.
Gerosa, L., Sauer, U., 2011. Regulation and control of metabolic fluxes in microbes. Curr. Opin. Biotechnol. 22, 566–575.
Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., Merryman, C., Young, L., Noskov, V.N., Glass, J.I., Venter, J.C., Hutchison, C.A., 3rd, Smith, H.O., 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319, 1215–1220.
Gibson, D.G., Smith, H.O., Hutchison, C.A., III, Craig Venter, J., Merryman, C., 2010. Chemical synthesis of the mouse mitochondrial genome. Nat. Methods 7, 901–903.
Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345.
Gong, B., Shin, M., Sun, J., Jung, C.-H., Bolt, E.L., van der Oost, J., Kim, J.-S., 2014. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. Proc. Natl. Acad. Sci. U. S. A. 111, 16359–16364.
Goodman, D.B., Church, G.M., Kosuri, S., 2013. Causes and effects of N-terminal codon bias in bacterial genes. Science 342, 475–479.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Science & Business Media.
Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., Broadbelt, L.J., 2005. Exploring the diversity of complex metabolic networks. Bioinformatics 21, 1603–1609.
Heigwer, F., Kerr, G., Boutros, M., 2014. E-CRISP: fast CRISPR target site identification. Nat. Methods 11, 122–123.
Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat.

Biotechnol. 28, 977–982.
Henry, C.S., Zinner, J.F., Cohoon, M.P., Stevens, R.L., 2009. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. Genome Biol. 10, R69.
Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., Gersbach, C.A., 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. doi:10.1038/nbt.3199
Hou, Y., Braun, D.R., Michel, C.R., Klassen, J.L., Adnani, N., Wyche, T.P., Bugni, T.S., 2012. Microbial strain prioritization using metabolomics tools for the discovery of natural products. Anal. Chem. 84, 4277–4283.
Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T.A., Goodman, D.B., Reppas, N.B., Emig, C.J., Bang, D., Hwang, S.J., Jewett, M.C., Jacobson, J.M., Church, G.M., 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333, 348–353.
Jakočiu□nas, T., Rajkumar, A.S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C.B., Skjødt, M.L., Nielsen, A.T., Borodina, I., Jensen, M.K., Keasling, J.D., 2015. CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae. ACS Synth. Biol. doi:10.1021/acssynbio.5b00007
Jang, Y.-S., Lee, J.Y., Lee, J., Park, J.H., Im, J.A., Eom, M.-H., Lee, J., Lee, S.-H., Song, H., Cho, J.-H., Seung, D.Y., Lee, S.Y., 2012. Enhanced Butanol Production Obtained by Reinforcing the Direct Butanol-Forming Route in Clostridium acetobutylicum. MBio 3, e00314–12–e00314–12.
Janga, S.C., Díaz-Mejía, J.J., Moreno-Hagelsieb, G., 2011. Network-based function prediction and interactomics: the case for metabolic enzymes. Metab. Eng. 13, 1–10.
Jensen, P.A., Lutz, K.A., Papin, J.A., 2011. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst. Biol. 5, 147.
Jensen, P.A., Papin, J.A., 2011. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 27, 541–547.
Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239.
Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., Yang, S., 2015. Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. Appl. Environ. Microbiol. 81, 2506–2514.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816–821.
Joung, J.K., Sander, J.D., 2013. TALENs: a widely applicable technology for targeted genome editing. Nat. Rev. Mol. Cell Biol. 14, 49–55.
Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R., 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief. Bioinform. 11, 40–79.
Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr, Assad-Garcia, N., Glass, J.I., Covert, M.W., 2012. A whole-cell computational model predicts phenotype from genotype. Cell 150, 389–401.
Katajamaa, M., Miettinen, J., Oresic, M., 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 22, 634–636.
Kaur, B., Chakraborty, D., Kumar, B., 2014. Metabolic engineering of Pediococcus acidilactici BD16 for production of vanillin through ferulic acid catabolic pathway and process optimization using response surface methodology. Appl. Microbiol. Biotechnol. 98, 8539–8551.

Kim, H.U., Kim, W.J., Lee, S.Y., 2013. Flux-coupled genes and their use in metabolic flux analysis. Biotechnol. J. 8, 1035–1042.
Kim, J., Reed, J.L., 2012. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. Genome Biol. 13, R78.
Kim, T.Y., Kim, H.U., Park, J.M., Song, H., Kim, J.S., Lee, S.Y., 2007. Genome-scale analysis of Mannheimia succiniciproducens metabolism. Biotechnol. Bioeng. 97, 657–671.
Kok, S. de, Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G., Platt, D., Shapland, E.B., Serber, Z., Dean, J., Newman, J.D., Chandran, S.S., 2014. Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. ACS Synth. Biol. 3, 97–106.
Kouprina, N., Larionov, V., 2008. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat. Protoc. 3, 371–377.
Kuhlman, T.E., Cox, E.C., 2010. Site-specific chromosomal integration of large synthetic constructs. Nucleic Acids Res. 38, e92.
Kuscu, C., Arslan, S., Singh, R., Thorpe, J., Adli, M., 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat. Biotechnol. 32, 677–683.
Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., Kwok, P.-Y., 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol. 30, 771–776.
Lee, J., Yun, H., Feist, A.M., Palsson, B.Ø., Lee, S.Y., 2008. Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. Appl. Microbiol. Biotechnol. 80, 849–862.
Lee, Y., Lafontaine Rivera, J.G., Liao, J.C., 2014. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. Metab. Eng. 25, 63–71.
Lv, L., Ren, Y.-L., Chen, J.-C., Wu, Q., Chen, G.-Q., 2015. Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metab. Eng. 29, 160–168.
Lynch, M.D., Warnecke, T., Gill, R.T., 2007. SCALEs: multiscale analysis of library enrichment. Nat. Methods 4, 87–93.
Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-guided human genome engineering via Cas9. Science 339, 823–826.
Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., Keasling, J.D., 2003. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat. Biotechnol. 21, 796–802.
Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., Ploegh, H.L., 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. doi:10.1038/nbt.3190
McCloskey, D., Palsson, B.Ø., Feist, A.M., 2013. Basic and applied uses of genome□scale metabolic network reconstructions of Escherichia coli. Mol. Syst. Biol. 9, 661.
Mercer, J., Snijder, B., Sacher, R., Burkard, C., Bleck, C.K.E., Stahlberg, H., Pelkmans, L., Helenius, A., 2012. RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. Cell Rep. 2, 1036–1047.
Metzker, M.L., 2010. Sequencing technologies—the next generation. Nat. Rev. Genet. 11, 31–46.
Milne, C.B., Eddy, J.A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R.S., Jin, Y.-S., Blaschek, H.P., Price, N.D., 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain Clostridium beijerinckii NCIMB 8052. BMC Syst. Biol. 5, 130.
Minty, J.J., Lesnefsky, A.A., Lin, F., Chen, Y., Zaroff, T.A., Veloso, A.B., Xie, B., McConnell, C.A., Ward, R.J., Schwartz, D.R., Rouillard, J.-M., Gao, Y., Gulari, E., Lin, X.N., 2011. Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in

Escherichia coli. Microb. Cell Fact. 10, 18.
Miskovic, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an ORACLE. Trends Biotechnol. 28, 391–397.
Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740.
Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., Kanehisa, M., 2010. PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res. 38, W138–43.
Mosberg, J.A., Lajoie, M.J., Church, G.M., 2010. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics 186, 791–799.
Murphy, K.C., 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. J. Bacteriol. 180, 2063–2071.
Murphy, R.F., 2011. An active role for machine learning in drug development. Nat. Chem. Biol. 7, 327–330.
Navid, A., Almaas, E., 2012. Genome-level transcription data of Yersinia pestis analyzed with a new metabolic constraint-based approach. BMC Syst. Biol. 6, 150.
Nocon, J., Steiger, M.G., Pfeffer, M., Sohn, S.B., Kim, T.Y., Maurer, M., Rußmayer, H., Pflügl, S., Ask, M., Haberhauer-Troyer, C., Ortmayr, K., Hann, S., Koellensperger, G., Gasser, B., Lee, S.Y., Mattanovich, D., 2014. Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab. Eng. 24, 129–138.
Nyerges, A., Csorgo, B., Nagy, I., Latinovics, D., Szamecz, B., Posfai, G., Pal, C., 2014. Conditional DNA repair mutants enable highly precise genome engineering. Nucleic Acids Res. 42, e62.
Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.Ø., 2011. A comprehensive genome ‐ scale reconstruction of Escherichia coli metabolism—2011. Mol. Syst. Biol. 7, 535.
Otto, A., Bernhardt, J., Hecker, M., Becher, D., 2012. Global relative and absolute quantitation in microbial proteomics. Curr. Opin. Microbiol. 15, 364–372.
Paddon, C.J., Keasling, J.D., 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat. Rev. Microbiol. 12, 355–367.
Paley, S.M., Karp, P.D., 2006. The Pathway Tools cellular overview diagram and Omics Viewer. Nucleic Acids Res. 34, 3771–3778.
Pavletich, N.P., Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252, 809–817.
Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B., Keasling, J.D., 2012. Microbial engineering for the production of advanced biofuels. Nature 488, 320–328.
Pines, G., Freed, E.F., Winkler, J.D., Gill, R.T., 2015a. Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. ACS Synth. Biol. doi:10.1021/acssynbio.5b00009
Pines, G., Pines, A., Garst, A.D., Zeitoun, R.I., Lynch, S.A., Gill, R.T., 2015b. Codon Compression Algorithms for Saturation Mutagenesis. ACS Synth. Biol. 0, null.
Puchałka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., Papin, J.A., Martins dos Santos, V.A.P., 2008. Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput. Biol. 4, e1000210.
Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183.
Quan, J., Tian, J., 2011. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat. Protoc. 6, 242–251.

Quan, J., Tian, J., 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4, e6441.
Raman, S., Rogers, J.K., Taylor, N.D., Church, G.M., 2014. Evolution-guided optimization of biosynthetic pathways. Proc. Natl. Acad. Sci. U. S. A. 111, 17803–17808.
Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P.R., Lundgren, M., 2015. Efficient programmable gene silencing by Cascade. Nucleic Acids Res. 43, 237–246.
Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M.O., Aebersold, R., 2011. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat. Methods 8, 430–435.
Reynolds, T.S., Gill, R.T., 2015. Quantifying Impact of Chromosome Copy Number on Recombination in Escherichia coli. ACS Synth. Biol. doi:10.1021/sb500338g
Rodrigo, G., Carrera, J., Prather, K.J., Jaramillo, A., 2008. DESHARKY: automatic design of metabolic pathways for optimal cell growth. Bioinformatics 24, 2554–2556.
Rossell, S., Huynen, M.A., Notebaart, R.A., 2013. Inferring metabolic states in uncharacterized environments using gene-expression measurements. PLoS Comput. Biol. 9, e1002988.
Rouet, P., Smih, F., Jasin, M., 1994. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. 91, 6064–6068.
Ruano, G., Kidd, K.K., 1989. Direct haplotyping of chromosomal segments from multiple heterozygotes via allele-specific PCR amplification. Nucleic Acids Res. 17, 8392.
Sandoval, N.R., Kim, J.Y.H., Glebes, T.Y., Reeder, P.J., Aucoin, H.R., Warner, J.R., Gill, R.T., 2012. Strategy for directing combinatorial genome engineering in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 109, 10540–10545.
Santos, C.N.S., Regitsky, D.D., Yoshikuni, Y., 2013. Implementation of stable and complex biological systems through recombinase-assisted genome engineering. Nat. Commun. 4, 2503.
Sawitzke, J.A., Costantino, N., Li, X.-T., Thomason, L.C., Bubunenko, M., Court, C., Court, D.L., 2011. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. J. Mol. Biol. 407, 45–59.
Schmidt, B.J., Ebrahim, A., Metz, T.O., Adkins, J.N., Palsson, B.Ø., Hyduke, D.R., 2013. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics 29, 2900–2908.
Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences 99, 15112–15117.
Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., Gao, C., 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. 31, 686–688.
Shao, Z., Zhao, Hua, Zhao, Huimin, 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37, e16.
Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., Court, D.L., 2009. Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. 4, 206–223.
Shetty, R.P., Endy, D., Knight, T.F., Jr, 2008. Engineering BioBrick vectors from BioBrick parts. J. Biol. Eng. 2, 5.
Shinfuku, Y., Sorpitiporn, N., Sono, M., Furusawa, C., Hirasawa, T., Shimizu, H., 2009. Development and experimental verification of a genome-scale metabolic model for Corynebacterium glutamicum. Microb. Cell Fact. 8, 43.
Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15, 121–132.
Smith, A.M., Heisler, L.E., Mellor, J., Kaper, F., Thompson, M.J., Chee, M., Roth, F.P., Giaever, G., Nislow, C., 2009. Quantitative phenotyping via deep barcode sequencing. Genome Res. 19, 1836–1842.

Smith, A.M., Heisler, L.E., St Onge, R.P., Farias-Hesson, E., Wallace, I.M., Bodeau, J., Harris, A.N., Perry, K.M., Giaever, G., Pourmand, N., Nislow, C., 2010. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res. 38, e142.
Sohn, S.B., Graf, A.B., Kim, T.Y., Gasser, B., Maurer, M., Ferrer, P., Mattanovich, D., Lee, S.Y., 2010. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnol. J. 5, 705–715.
Song, C.W., Kim, D.I., Choi, S., Jang, J.W., Lee, S.Y., 2013. Metabolic engineering of Escherichia coli for the production of fumaric acid. Biotechnol. Bioeng. 110, 2025–2034.
Sorek, R., Cossart, P., 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat. Rev. Genet. 11, 9–16.
Sorek, R., Lawrence, C.M., Wiedenheft, B., 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. 82, 237–266.
Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L., 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 164, 49–53.
Swainston, N., Smallbone, K., Mendes, P., Kell, D., Paton, N., 2011. The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. J. Integr. Bioinform. 8, 186.
Töpfer, N., Jozefczuk, S., Nikoloski, Z., 2012. Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in Escherichia coli. BMC Syst. Biol. 6, 148.
Tran, L.M., Rizk, M.L., Liao, J.C., 2008. Ensemble modeling of metabolic networks. Biophys. J. 95, 5606–5617.
Van Opijnen, T., Bodi, K.L., Camilli, A., 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat. Methods 6, 767–772.
Wang, H.H., Church, G.M., 2011. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. Methods Enzymol. 498, 409–426.
Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., Church, G.M., 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894–898.
Wang, H.H., Kim, H., Cong, L., Jeong, J., Bang, D., Church, G.M., 2012. Genome-scale promoter engineering by coselection MAGE. Nat. Methods 9, 591–593.
Wang, Y., Eddy, J.A., Price, N.D., 2012. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst. Biol. 6, 153.
Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B.A., Gill, R.T., 2010. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat. Biotechnol. 28, 856–862.
Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., Jaenisch, R., 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154, 1370–1379.
Yang, J., Wang, Z., Zhu, N., Wang, B., Chen, T., Zhao, X., 2014. Metabolic engineering of Escherichia coli and in silico comparing of carboxylation pathways for high succinate productivity under aerobic conditions. Microbiol. Res. 169, 432–440.
Zhang, F., Keasling, J., 2011. Biosensors and their applications in microbial metabolic engineering. Trends Microbiol. 19, 323–329.
Zhang, Y., Muyrers, J.P., Testa, G., Stewart, A.F., 2000. DNA cloning by homologous recombination in Escherichia coli. Nat. Biotechnol. 18, 1314–1317.
Zhang, Y., Werling, U., Edelmann, W., 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. 40, e55.
Zomorrodi, A.R., Lafontaine Rivera, J.G., Liao, J.C., Maranas, C.D., 2013. Optimization-driven

identification of genetic perturbations accelerates the convergence of model parameters in ensemble modeling of metabolic networks. Biotechnol. J. 8, 1090–1104.

Figure

Figures

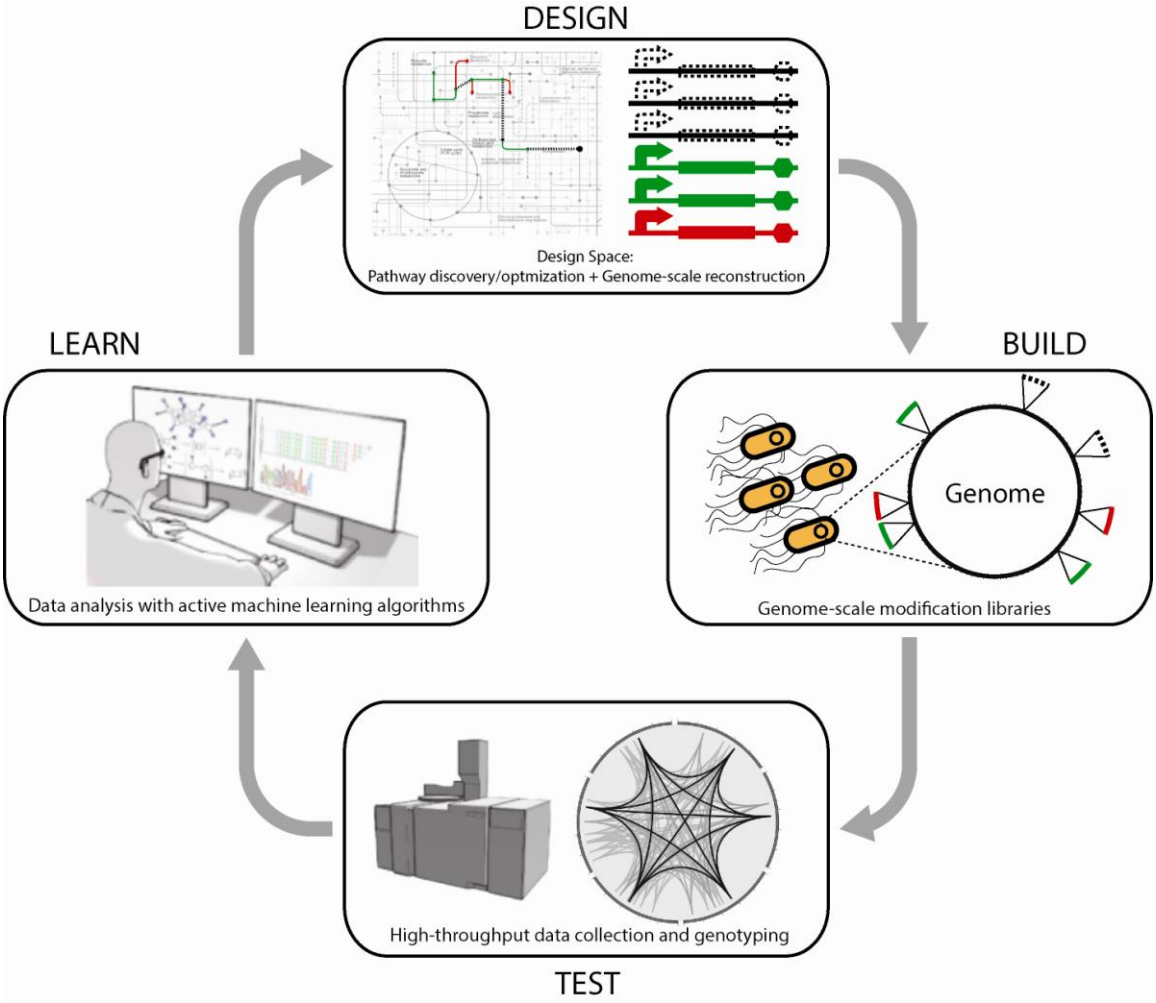

Fig. 1 The DBTL cycle applied to synthetic biology.

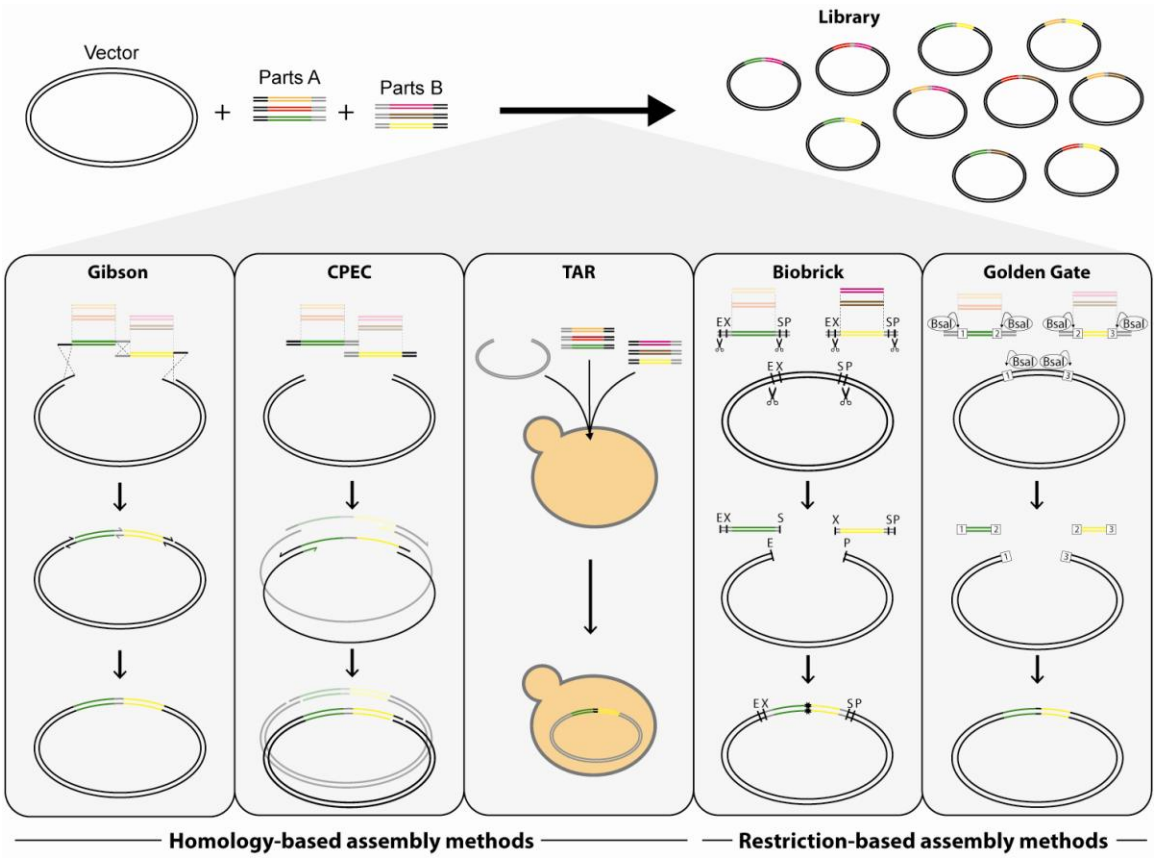

Fig. 2 DNA assembly technologies for libraries construction

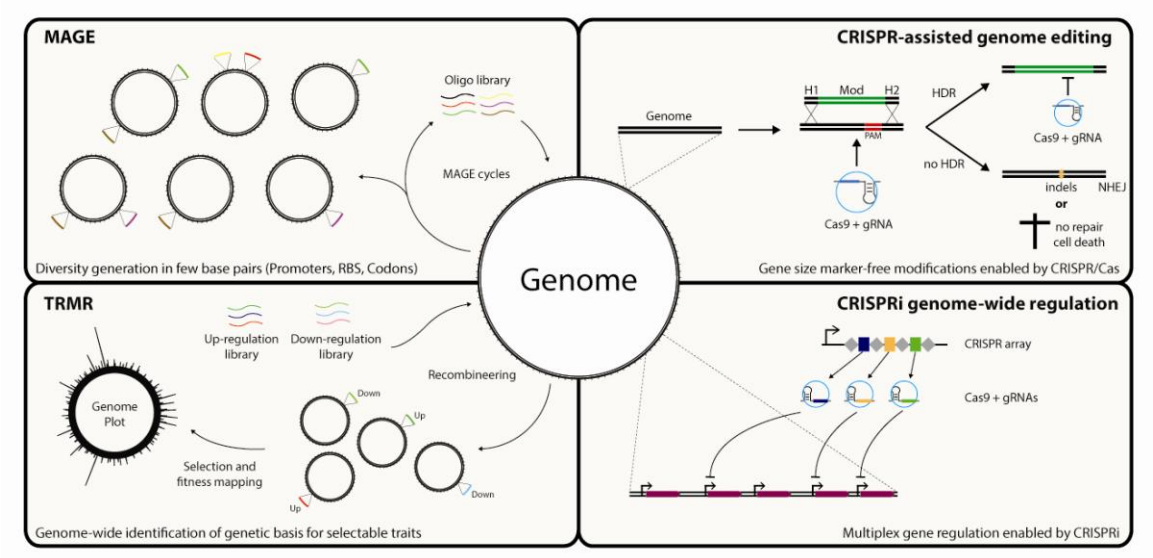

Fig. 3 Genome-scale engineering strategies

Tables

Table 1 Computational tools for genome-scale modeling

Tools	Description	Reference
Pathway Tools	Genome-scale reconstruction	Paley et al., 2006; Karp et al., 2010
Model SEED		Henry et al., 2010
GLAMM		Bates et al., 2011
SuBliMinaL toolbox		Swainston et al., 2011
RAVEN toolbox		Agren et al., 2013
Path2Models		Büchel et al., 2013
EMRA		Lee et al., 2014
BNICE	Pathway prediction	Hatzimanikatis et al., 2005
PathPred		Moriya et al., 2010
RetroPath		Carbonell et al., 2014a
DESHARKY		Rodrigo et al., 2008
PathPred		Moriya et al., 2010
RELATCH		Kim and Reed 2012
XTMS		Carbonell et al., 2014b
PROM		Chandrasekaran and Price 2010
ORACLE		Miskovic and Hatzimanikatis 2010
GEM-Path		Campodonico et al., 2014
OptKnock	Gene Knockout	Choon et al., 2014
OptGene		Patil et al., 2005
MOMA		Segrè et al., 2002
EMILiO		Yang et al., 2011
SIMUP		Gawand et al., 2013
GDLS		Lun et al., 2009
OptForce	Pathway amplification	Ranganathan et al., 2010
OptORF		Kim and Reed 2010

CosMos		Cotten and Reed 2013
FSEOF		Nocon et al., 2014
GIMME	transcriptome	Becker et al., 2008
E-Flux		Colijn et al., 2009
MADE		Jensen and Papin 2011
TIGER		Jensen et al., 2011
iMAT		Bordbar et al., 2012
TEAM		Collins et al., 2012
AdaM		Töpfer et al., 2012
GX-FBA		Navid and Almaas 2012
mCADRE		Wang et al., 2012
FCGs		Kim et al., 2013
EXAMO		Rossell et al., 2013
GIMMEp	Proteome	Bordbar et al., 2012
GIM^3E	metabolome	Schmidt et al., 2013
MET-IDEA		Broeckling et al., 2006
MZmine		Katajamaa et al., 2006

Table 2: DNA assembly methods summary

	DNA Assembly Method	Advantages	Limitations	Thoughput (number of parts/reaction for efficient assembly)	Assembly Time	References
Homology-based	Gibson	Fast cloning reaction	- Efficiency drops with increasing number of assembly parts - Homology-based cloning limitations*	~ 4	1 hour	Gibson et al., 2009 Kok et al., 2014
	CPEC	Relatively cheap (only DNA polymerase is required)	- Long assembly time compared to other methods - Homology-based cloning limitations*	~ 4	1-3 hours	Quan et al., 2011 Kok et al., 2014
	Yeast TAR	Efficient for complex and long assemblies	- Long assembly time compared to other methods - Homology-based cloning limitations*	at least 12	2-3 days	Shao et al., 2009 Kok et al., 2014
	SLiCE	Relatively cheap (only bacterial cell extract required)	Homology-based cloning limitations*	~ 7	1 hour	Zhang et al., 2012
	LCR	Fast and highly efficient reaction	Homology-based cloning limitations*	at least 12	1.5-2 hours	Kok et al., 2014
Restriction-based	Biocricks	A lot of parts are already available in the biobrick format	- Only 3 fragments can be assembled at a time - Restriction site must be	3	1.5-2 hours	Shetty et al., 2008

		absent in the assembled parts			
Golden Gate	Alow simultaneous assemblies of multiple parts compared to Biobricks	Restriction site must be absent in the assembled parts	at least 9	5 - 30 min	Engler et al., 2008 Engler et al., 2014

Table 3: Summary of different methods for targeted genome editing

Genome Editing Method	Advantages	Limitations	References
MAGE	- Markerless modifications - Highly multiplexible	Requires mismatch-repair deficient background	Wang et al., 2009 Bonde et al., 2014
TRMR / Bar-Seq / Tn-Seq	Allows identification of unknown genes involved in complex traits	Combination of mutations and epistatic interactions are not taken into account	Warner et al., 2010 Smith et al., 2009 van Opijnen et al., 2009
RAGE	Allows integration of large-size DNA constructs	Multiple steps required	Santos et al., 2013
TALENs	Allows targeted genome editing	- Requires engineering whole enzyme for every new targeted site - Off-target effects	Joung et al., 2013 Gaj et al., 2013
ZFNs	Allows targeted genome editing	- Requires engineering whole enzyme for every new targeted site - Off-target effects	Gaj et al., 2013
CRISPR	Allows targeted genome editing with easy target reprogramming capabilities	Off-target effects	Mali et al., 2013 Cong et al., 2013 Jinek et al., 2012 Kuscu et al., 2014