Evolutionary approaches for engineering industrial phenotypes in bacterial cell factories

Journal:	*Biotechnology Journal*
Manuscript ID	biot.201800439
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	03-Mar-2019
Complete List of Authors:	Fernández-Cabezón, Lorena; The Novo Nordisk Foundation Center for Biosustainability Cros, Antonin; The Novo Nordisk Foundation Center for Biosustainability Nikel, Pablo ; The Novo Nordisk Foundation Center for Biosustainability
Primary Keywords:	Metabolic engineering, Synthetic biology
Secondary Keywords:	Bacteria
Additional Keywords:	Industrial phenotypes, Combinatorial engineering

Review Article

Evolutionary approaches for engineering industrial phenotypes in bacterial cell factories

by

Lorena Fernández-Cabezón, Antonin Cros, and Pablo I. Nikel*

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs Lyngby, Denmark

Running title: Engineering industrial phenotypes in bacteria

Keywords: Metabolic engineering · Industrial phenotypes · Adaptive laboratory evolution · Synthetic Biology · Combinatorial engineering

Abbreviations: **ALE**, adaptive laboratory evolution; **CRISPRi**, CRISPR interference; **dCas9**, inactive Cas9; **dsDNA**, double-stranded DNA; **FACS**, fluorescence-activated cell sorting; **GRAS**, generally recognized as safe; **MCF**, microbial cell factory; **MMR**, mismatch repair system; **nCas9**, Cas9 nickase; **sgRNA**, single-guide RNA

* **Correspondence to:** *Pablo I. Nikel* (pabnik@biosustain.dtu.dk)
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
2800 Kgs Lyngby, Denmark
Tel: (+45 93) 51 19 18
Fax: (+45 45) 25 80 00

Abstract

The use of bacterial platforms for the bio-based production of added-value compounds (e.g. pharmaceuticals, biofuels, food ingredients, and building blocks, among other applications) has become a well-established industrial activity. The construction of microbial cell factories (MCFs) with robust and stable industrial phenotypes, however, remains one of the biggest challenges of contemporary biotechnology. In this review, we discuss methodologies for the optimization of bacterial MCFs for industrial bioprocesses, following the engineering principle of natural evolution based on genetic variation and selection. We present the state-of-the-art regarding techniques to manipulate and increase genetic variation in the bacterial population or to construct combinatorial libraries of strains, both globally (genome) and locally (individual genes or pathways, and entire sections and clusters of the bacterial genome). We also summarize different screening and selection technologies that can be applied to exploit this genetic variation with the ultimate goal of isolating mutant MCFs with (enhanced) phenotypes. Based on these principles, we close the article by presenting future trends in the design and construction of a new generation of MCFs that contribute to the long-sought-after transformation from a petrochemical industry to a veritable sustainable bio-based industry.

1. Introduction

The use of microorganisms as platforms for the production of added-value compounds has a long tradition and, over the years, has become a well-established industrial activity.[1,2] Biologically-based compounds are increasingly been used as pharmaceuticals, food ingredients, biofuels, and chemical precursors (building blocks), among other relevant applications.[3] The development of sustainable lifestyles, which meet current human needs without compromising the Earth's resources or its future, requires novel, efficient bio-manufacturing methods for the biosynthesis of such products.[4] Along the same line, streamlined workflows for the design, construction, and testing of *microbial cell factories* (MCFs) is key to support the long-sought-after transformation from a petrochemical industry to a more sustainable bio-based industry. An ideal MCF is expected to meet certain properties such as fast growth, non-pathogenicity (and, if possible, *generally recognized as safe*—GRAS—status), genetic stability, growth in low-cost defined media up to high cell densities, high product yield and efficient secretion, and limited formation of by-products.[5,6] In addition, MCFs should grow and produce the compound of interest under culture conditions that are optimal for the overall industrial process, including ease of media preparation and simple downstream steps. This last aspect is particularly challenging, since cell viability can be compromised by factors such as adverse industrial conditions, presence of growth inhibitors in raw materials, or accumulation of toxic compounds during fermentation. Thus, the construction of MCFs with robust phenotypes continues to be one of the grand challenges of industrial biotechnology.[7–11] Historically, this overarching task has been addressed through different methodologies that include (i) classical mutagenesis approaches, (ii) adaptive laboratory evolution (ALE), (iii) rational engineering, and (iv) combinatorial approaches. We start by summarizing the main aspects of each approach to set the stage for the discussion on novel, emerging methodologies for the optimization of bacterial cell factories.

In classical approaches for strain improvement, microorganisms capable of producing the compound(s) of interest are isolated and subjected to successive rounds of random mutagenesis followed by screening of enhanced properties. Although useful and simple, these approaches are usually laborious and time-consuming. Along with the emergence of Metabolic Engineering as a

field in the 1990s, new rational approaches began to be implemented for the construction of MCFs.[12,13] As a discipline, Metabolic Engineering seeks to improve cellular activities through the targeted manipulation of enzymatic, transport, and regulatory cell functions with the use of recombinant DNA technologies. In some cases, new enzymatic activities and biosynthetic routes are implanted in well-established hosts with the objective of producing non-natural products and expanding the chemical repertoire.[14,15] In time, innovative experimental and computational tools have been developed that allowed the emergence of *Systems Metabolic Engineering*, a more systematic and high-throughput discipline that integrates traditional Metabolic Engineering approaches with other fields such as Systems Biology and Synthetic Biology.[16–18] Thus, MCF design strategies are now taken onto a global scale, simultaneously considering upstream (i.e. strain development), midstream (i.e. fermentation), and downstream (i.e. separation and purification) stages of the bioprocesses. Several examples in the primary literature have demonstrated the efficacy of such approaches in improving phenotypes of industrial interest. Cost-efficient production of the amino acids L-lysine and L-arginine by *Corynebacterium glutamicum*, and the bulk chemical 1,4-butanediol and the drug precursor artemisinic acid by *Escherichia coli*, are just a few instances of bioprocesses that have found their way to industrial setups.[19–22]

Although Systems Metabolic Engineering opened up avenues for MCF design and construction, these approaches still require a substantial quantitative knowledge of the microbial host and the ability to reliably predict responses to multiple genetic or environmental manipulations (i.e. genotype-phenotype relationships). These issues remain to be considerable challenges, especially when using alternative microbial hosts other than *E. coli* or yeast. Low product yields, even after a comprehensive optimization of biosynthetic pathways, are often tied to a decrease in cell viability under industrial production conditions. In addition, stablishing a genetic toolbox for engineering non-model microbes is still a non-trivial and laborious task. All these aspects slow down the development of economically-viable MCFs and bioprocesses—and, indeed, a relatively small number of these strains can actually be scaled-up in an industrial context. Therefore, nowadays, random whole-genome engineering approaches, e.g. *via* ALE, have taken the field of industrial phenotype optimization into a wider space as compared to the use of rational approaches.

Evolutionary engineering, also known as *ALE*, follows natural evolution's "engineering principles" through genetic variation and selection.[23] Individual cells in a population randomly accumulate mutations, either as consequence of natural DNA replication errors or *via* externally induced mutagenesis mechanisms. By chance, some (beneficial) mutations encode phenotypic changes that allow cells to grow and divide faster in the culture conditions than other cells in the population, eventually taking progressive control of the population. As the selective pressure is increased, other beneficial mutations can be selected until the desired objective is achieved *via* gradual increases in fitness. ALE experiments have been applied successfully to improve a number of wild-type and engineered microorganisms and bioprocesses.[24–28] These approaches possess clear benefits over rational methods, including broad applicability to different microbial hosts, ease of practical implementation, discovery of new, non-intuitive regulatory mechanisms, and, most importantly, they guarantee at least *some* improvement of industrially-relevant phenotypes. On the other hand, the construction of combinatorial libraries of strains are also attracting increasing attention for optimizing cellular phenotypes.[7,29–31] These methodologies, if coupled with adequate screening and selection assays, can be applied to fine-tune the expression of metabolic pathway genes (i.e. combinatorial pathway engineering) and optimize complex phenotypes (whole-genome combinatorial techniques; e.g. whole-genome mutant libraries, genome shuffling, ribosome engineering). One way or the other, highly-efficient MCFs can only be obtained via increasing genetic variation and selecting towards the phenotype of interest.

From the examples above, it becomes clear that there is room for the development of novel, emerging approaches aimed at MCF optimization. In this review, we discuss methodologies for the improvement of bacterial MCFs for industrial bioprocesses, following the engineering principle of natural evolution based on genetic variation and selection (Fig. 1). Firstly, we present the most relevant techniques available to date to increase genetic variation in the bacterial population, both globally (genome) and locally (individual genes or sections of the genome). Secondly, different screening and selection technologies that can be applied to exploit genetic variation with the ultimate goal of isolating mutant strains with improved phenotypes are likewise discussed. Based on these principles, we close the article by presenting future trends in the design and construction of a new generation of MCFs that contribute to a more sustainable bio-based industry.

2. Strategies to increase the genetic variation in bacterial populations

Different techniques have been implemented in recent years to increase genetic variation in bacterial populations (Table 1). Several criteria must be considered when selecting the technique to be used, e.g. characteristics of the bacterial host, the type of phenotype to be improved, and the expected application/use of the resulting strain. For instance, the improvement of complex phenotypes often requires the implementation of whole-genome approaches, since the exact mechanisms (metabolic, regulatory, or both) underlying the phenotype of interest are not known. Targeted mutagenesis approaches, however, can be implemented to generate genetic variation (i.e. mutations) in specific genes or genomic regions. These targeted approaches have been shown to be especially useful to optimize phenotypes whose underlying molecular mechanisms or bottlenecks are fully or partially known (e.g. fine-tuning the expression of heterologous biosynthetic pathways or discovering new protein variants with improved properties). The most relevant whole-genome and targeted mutagenesis approaches described to date are discussed in the next two sections below.

2.1 Whole-genome approaches to increment genetic variation

Whole-genome random mutagenesis: external mutagenic agents and mutator strains

Mutations in DNA are one of the most important sources of genetic variation and, therefore, are fundamental to natural evolution. Small and transient increases in mutation rate(s) have shown to significantly improve the probability of generating beneficial mutations and accelerate the improvement of MCFs, due to intrinsic mutation rates are usually low for most organisms (in the order of 10^{-9}–10^{-10} events per base pair per generation).[32,33] DNA mutagenesis can be externally induced by chemical or physical mutagens or internally mediated by (conditional) mutator phenotypes (Table 1; Fig. 2A).

Chemical and/or physical mutagenesis has traditionally been used for the optimization of industrial strains, due to its technical simplicity and its applicability to virtually any microorganism.[34,35] The

design of a successful mutagenesis protocol, however, is not a trivial task. The selection of an optimal dose of mutagen, which generates sufficient genetic variation without increasing too much the frequency of detrimental mutations in the genome, is a particularly critical step. In addition, since most mutagens preferentially introduce certain types of mutations (e.g. alkylating agents such as ethyl methane sulfonate and nitrosomehtyl guanidine induce predominately GC to AT transitions), mutagens should be switched periodically during the protocol. Most chemical mutagens used are base analogs, intercalating agents and base modifiers, while physical mutagens include ultraviolet light, electromagnetic radiation (e.g. γ radiation and X radiation) and particle radiation (e.g. β and α particles).[34,35] An alternative mutagenic method, termed *atmospheric and room temperature mutagenesis*, offers advantages over traditional mutagens (e.g. larger variation of mutants and safer operating conditions than the use of radiation sources), yet it requires specialized equipment that can limit its widespread use.[36]

Genome-wide random mutagenesis can also be genetically induced by using conditional mutator phenotypes. Mutator strains are bacterial strains displaying high mutation rates, since they have mutations in one or several genes encoding DNA repair or error-avoidance cellular systems.[37] For example, specific mutations in the components of the DNA mismatch repair system (MMR; e.g. *mutL*, *mutH*, and *mutS*) or in proofreading polymerases (e.g. *dnaQ*), as well as the overexpression of certain dominant mutator alleles of the same genes have shown to result in strong mutator phenotypes. Mutator strains have been investigated for decades, and appear to occur naturally in bacterial populations that are propagated for prolonged periods under identical conditions and also in pathogenic bacteria.[38,39] Mutator strains were initially used to introduce random mutations in extrachromosomal DNA (e.g. plasmids);[34,40] however, mutator phenotypes have subsequently been applied to the phenotypic optimization of several bacterial species (e.g. *E. coli*, *Lactobacillus casei*, *Synechococcus* sp., *Bacillus subtilis*, and *Clostridium acetobutylicum*).[41–45] Since mutation rates must be controlled to avoid extensive accumulation of deleterious mutations and prevent genomic instability, the overexpression of dominant mutator alleles is generally driven from tightly-regulated expression systems that can be subsequently removed from isolated cells displaying the phenotype of interest. Most mutator allelles preferentially introduce certain types of mutations because they act on different cell mechanisms that ensure DNA replication fidelity.[37] For example,

transition and frameshift mutations or tranversions can be randomly created in bacterial cells by interfering with the MMR system or the proofreading activity encoded by the *dnaQ* gene, respectively. A broader mutational spectrum can be achieved by simultaneously interfering with several of these DNA fidelity ensuring mechanisms, e.g., by overexpressing multiple mutator alleles.[46,47] Due to the difficulty of achieving an optimal mutation rate for each experiment, mutagenesis is sometimes coupled with selection and, at the same time, modulation of mutation rates during the run. For this purpose, Luan *et al*.[48] constructed a library of *E. coli* mutants that overexpress *dnaQ* variants displaying several mutation rates and mutational spectra. This strategy was successfully applied to improve the tolerance of *E. coli* towards *n*-butanol, acetate, and thermal challenges.[48,49] Alternatively, genetic circuits that enable dynamic changes in the mutation rate according to a particular phenotype can be used (e.g. based on biosensors). Chou and Keasling,[50] for instance, developed an adaptive control system, called *evolution of feedback-regulated phenotypes*, which allows to regulate the mutation rate by expression of a dominant mutator allele in a fashion inversely proportional to the concentration of target metabolite. This system was used to increase the production of tyrosine and isoprenoids in engineered *E. coli* strains. Similarly, Pham *et al*.[51] designed a riboswitch-based pH-sensing genetic device to control gene expression according to the environmental pH, and used this device for evolutionary engineering of *E. coli* for improved tolerance to a broad spectrum of organic acids.

Genome shuffling

Genome shuffling was first implemented in 2002 to improve the production of the antibiotic tylosin by *Streptomyces fradiae* and to increase acid tolerance of *Lactobacillus* strains.[52,53] This technology, based on the construction of libraries of mutant strains through the recursive fusion of protoplasts (i.e. forcing recombination events between genomes), enables the introduction of multiple genetic changes without any knowledge of genome sequences or genetic networks (Table 1; Fig. 2B). Thus, genome shuffling is especially useful for the rapid improvement of complex phenotypes in non-well-characterized bacterial hosts. In addition, the resulting (i.e. shuffled) strains are not considered to be genetically-modified microorganisms.[54–56]

Genome shuffling requires a parental population with good genetic variation and sufficient variation in traits relevant for the complex phenotype of interest. Therefore, this technique is typically initiated by subjecting the parental population to repeated cycles of random mutagenesis.[54–56] Next, several parental strains are selected and their protoplasts are prepared and fused in several shuffling rounds. Finally, the shuffled strains are screened and the most promising strains can be subjected to new genomic shuffling cycles if needed. The phenotypic improvements achieved after only a few cycles have been demonstrated to be comparable to those resulting from multi-year classical strain improvement.[52,53] In addition, genome shuffling is a quite cost-effective and easy-to-handle technology—although its application is largely limited to Gram-positive bacteria. Only a small number of examples of genome shuffling in Gram-negative bacteria are available (e.g. *E. coli*, *Pseudomonas parafulva*, and *Sphingobium chlorophenolicum*), probably due to the fact that protoplast fusion is less efficient in these bacteria.[57–59] However, mechanisms for horizontal gene transfer (e.g. conjugation, transduction, and transformation), many of which have higher efficiency in Gram-negative bacteria, might be exploited to foster recombination in these species.[55]

Construction of whole-genome mutant libraries

Gene function(s) can be investigated in a dynamic living system through systematic gene knock-down, knock-out, or overexpression. On the basis of this premise, whole-genome combinatorial libraries of mutants have been constructed and applied to identify relevant genetic targets for strain improvement in different bacterial species (Table 1).[7,8] The main advantage of this approach is that the resulting phenotype can be easily linked back to a single genetic perturbation. In addition, genes encoding regulatory proteins or proteins with unknown functions participating in the relevant phenotype (which could not be predicted by rational approaches), can be identified. Once the combinatorial libraries have been constructed, they can be used in multiple screenings of individual mutants or in competition fitness assays with pools of mutants.[60]

Random knock-out libraries have traditionally been constructed by transposon insertion mutagenesis that involves the random integration of an antibiotic resistance gene into the genome mediated by the activity of a transposase (Fig. 2C).[60] This technique has several advantages, such

as technical simplicity and broad applicability to multiple microbial species; however, an unbiased integration of transposons or possible pleiotropic effects may hinder the identification of relevant targets to unravel genotype-phenotype relationships. Some of these limitations can be overcome by constructing in-frame, single-gene knock-out mutant libraries, such as the KEIO collection created in the strain background *E. coli* K-12 BW25113.[61] The creation of such libraries has been possible thanks to the development of new efficient techniques of bacterial genome engineering (e.g. recombineering). Alternatively, knock-down libraries can be obtained using clustered regularly interspaced short palindromic repeats (CRISPR) interference (CRISPRi), which allows for transcriptional downregulation of almost any gene in bacterial genomes.[62] CRISPRi targets can be easily programmed *in silico* by substituting the first 20 nucleotides of the single-guide RNA (sgRNA) sequence to match the non-template strand of the target gene. As such, the design and construction of CRISPRi libraries (i.e. genome-wide libraries of sgRNAs) that target specific sets of genes or the entire genome becomes a straightforward task. Since CRISPRi knock-downs can be induced, titrated, and tuned, all genes—including essential ones, which cannot be accessed through knock-outs—can be systematically targeted. Despite these advantages, the construction of CRISPRi libraries has been used in only a few bacterial species (*E. coli*, *B. subtilis*, *Streptococcus pneumoniae*),[63–67] due to the fact that CRISPRi has been transferred using species-specific or narrow host range strategies and because components need to be optimized for proper function in different species. To overcome this barrier, Peters *et al*.[68] have recently developed '*mobile-CRISPRi*'—a suite of modular and transferable CRISPRi components that can stably integrate into the genomes of diverse bacteria.

Gene overexpression libraries have been traditionally designed by using genomic expression libraries that allow the identification of positive genetic targets, i.e. individual genes or operons found in the cloned genomic fragments that improve the desired phenotype in the host strain (Fig. 2D). However, new technological advances and availability of genome sequence data enabled the creation of full-length open reading frame (ORF) libraries, such as the *ORFeome* collection of *E. coli* K-12 (ASKA collection),[69] which has considerably simplified the investigation of genotype-phenotype relationships and the optimization of gene expression. ORF libraries are constructed by PCR amplification or chemical synthesis of each individual ORF and its subsequent cloning in a

suitable expression vector. Therefore, overexpression libraries can now be constructed using genomic DNA or synthetic DNA based on the genomic information of the microbial host or other organisms with traits of interest, as well as with metagenomic DNA samples. For example, the heterologous expression of genomic libraries from *Piriformospora indica* and *Marinobacter aquaeolei* in *E. coli* allowed the identification of genes conferring salt and terpene tolerance, respectively.[70,71]

All these combinatorial libraries have been successfully adopted to improve growth-related phenotypes such as stress tolerance and substrate utilization. For example, they have been remarkably successful to improve phenotypes of industrial interest such as lycopene production in engineered *E. coli*[72] and acid tolerance in *Clostridium acetobutylicum*.[73] However, these approaches only allow for the identification of individual gene contributions in the phenotype of interest, so they often need to be combined with each other or with other methodologies to achieve phenotypic improvements. For example, more complex gene interactions can be investigated by using *co-expressing genomic libraries*, created with two compatible overexpression libraries.[74] Pooled competition fitness assays can also be coupled to next generation sequencing technologies, which allows monitoring the enrichment of certain mutants in the population during the experiment (e.g. transposon insertion sequencing, SCALE).[75–77] Emerging genome editing technologies, combined with barcoding technology, can also be implemented for the easy, time-efficient construction of whole-genome combinatorial libraries.[78,79]

Engineering transcriptional regulatory proteins

In recent years, several research groups have focused their attention on the direct or indirect manipulation of the transcriptional regulatory network in bacteria for the construction of efficient MCFs.[7,29,80,81] The engineering of transcriptional regulatory proteins (TRPs) allows to alter the expression of hundreds of genes at the same time, eliciting the appearance of complex traits (Table 1). In this approach, TRPs capable of reprogramming cell metabolism and regulation are firstly identified to improve complex phenotypes. The TRPs can be native, exogenous, or synthetic (i.e. artificial) regulators, modulating cellular functions at the specific, medium, or global level. After the

selection of one or several TRP(s), the workflow involves the creation of a library of variants of the proteins, the expression of the library in the host cell, and the subsequent selection of the desired phenotype, repeating the entire process as many times as necessary (Fig. 2E). The identification and characterization of TRPs, e.g. by randomized knock-out and overexpression libraries, is essential for the successful application of this methodology.

One of the most widely used approaches has been the modification of σ factors in bacteria (*global transcription machinery engineering*).[82,83] These regulatory proteins bind to the promoter regions with varying degrees of affinity and help to recruit the holoenzyme RNA polymerase to initiate transcription. Therefore, small variations in σ factors have the potential to affect greatly the subset of genes that are bound by RNA polymerase and are expressed in the MCF. The σ factor σ^{70} (RpoD) and, to a lesser extent, σ^{38} (RpoS), have been targets for global transcription machinery engineering in several bacteria (*E. coli*, *Lactobacillus plantarum*, *Klebsiella pneumoniae*, and *Zymomonas mobilis*).[29,84,85] Other endogenous global regulators, e.g. the global regulator H-NS (histone-like nucleoid structuring factor) and the cAMP receptor protein (CRP), as well as the RNA polymerase itself (RpoA), have been engineered in *E. coli* to improve several complex phenotypes such as acid tolerance.[29,80,86] Analogously, the pleiotropic transcriptional regulators CodY and CcpA of *B. subtilis* were engineered to increase recombinant protein production.[87] In addition to these native TRPs, the exogenous global regulator IrrE of the radiation-resistant *Deinococcus radiodurans* has the potential to act as a global regulator in *E. coli* and to enhance its resistance against ethanol, butanol, and acetate stress—specially after being engineered.[88] Similarly, the σ factor σ^{HrdB} of *Actinoplanes missouriensis*, *Micromonospora aurantiaca*, and *Salinispora arenicola,* were selected for random mutagenesis/DNA shuffling and introduced into *Actinoplanes teichomyceticus* to improve the synthesis of the antibiotic teicoplanin.[89] Libraries of artificial TRPs can also be designed to reprogram the bacterial transcriptome. For example, synthetic TRPs based on zinc finger motifs have been used to optimize complex phenotypes in *E. coli* such as resistance to heat shock, osmotic pressure and cold shock.[90–92] In this case, mutant libraries were constructed by randomly assembling 3-4 zinc-finger motifs from dozens of available motifs and were sometimes fused to endogenous transcriptional effector domains (e.g. the cAMP receptor protein, CRP, from *E. coli*).

All these approaches have been successfully applied to improve tolerance to stress, the production of metabolites, and the use of alternative substrates in several bacterial species. Some of these phenotypes, which were previously unattainable through classical strain improvement, were generated in a highly efficient manner after a few rounds of selection. The selected mutants may possess multiple altered phenotypes. For instance, the best oxidative stress-tolerant *E. coli* mutant OM3, obtained *via* CRP transcriptional engineering, also displayed enhanced thermotolerance at 48°C and organic solvent tolerance, suggesting an overlap in microbial stress responses.[93]

Ribosome and RNA polymerase engineering

Ribosome engineering has been widely used to optimize the production of antibiotics and activate silent gene clusters involved in the biosynthesis of secondary metabolites in Actinobacteria. This technique, characterized by its simplicity, consists of the isolation of spontaneous mutants resistant to drugs that target bacterial ribosomes (e.g. streptomycin, gentamicin, kanamycin, and chloramphenicol) (Fig. 2F).[94] Although the mechanism leading to the phenotypic variation of these mutants is not well understood, it is believed to be related to the activation of a 'stringent response' that leads to elevated levels of protein synthesis during the late growth phase and is beneficial for certain specialized phenotypes (e.g. secondary metabolism). Mutants resistant to these antibiotics often have a deletion or point mutation within genes encoding ribosomal components (e.g. ribosomal protein, ribosomal RNA, or translation factor), thus altering their translation machinery.[94] Similarly, it has been shown that the isolation of mutants resistant to the antibiotic rifampicin, which inhibits RNA polymerase, is a useful approach to improve certain phenotypes. These mutants, with a modified transcriptional machinery, often display RNA polymerase variants with enhanced affinity for the promoter regions of certain genes related to secondary metabolism.[94]

Both techniques described above (i.e. ribosome and RNA polymerase engineering), allow for the identification of high-performance strains between drug-resistant mutants with a high frequency of 5–40% (Table 1). They have enabled the overproduction of antibiotics in several *Streptomyces* species and α-amylase and protease in *Bacillus subtilis*, and the improvement of tolerance to

various aromatic compounds in the platform soil bacterium *Pseudomonas putida*.[94] These techniques do not require any genomic information or the use of mutagens, overcoming the accumulation of deleterious mutations in other genomic regions. In addition, different concentrations and/or combinations of antibiotics can be used to isolate a collection of strains with a different spectrum of mutations affecting their transcription and translation machinery and therefore, with different phenotypic characteristics. For example, a 180-fold increase in the production of actinorhodin in *Streptomyces coelicolor* 114735 has been achieved by the introduction of eight drug-resistance mutations.[95] To date, however, these approaches have been mainly used to improve the production of secondary metabolites and only a few reports have focused on the production of primary metabolites (e.g. vitamin B_{12} in *Propionibacterium shermanii*).[96] Furthermore, since this approach is based on the isolation of spontaneous mutants, its applicability may sometimes be limited due to the low frequency of mutation in bacteria.

2.2 Targeted mutagenesis approaches

Combinatorial pathway engineering

The construction of combinatorial libraries encoding metabolic routes is progressively becoming the tool of choice for heterologous pathway optimization (e.g. to eliminate metabolic flux imbalances).[30,31] The increasingly lower price of *de novo* DNA synthesis, the development of sophisticated DNA assembly technologies, and the standardization of genetic parts have notably accelerated the construction of libraries of variants.[97,98] The combinatorial assembly of critical pathway and regulatory elements (e.g. plasmid backbones, promoters, ribosome binding sites, coding sequences, terminators) facilitates the simultaneous optimization at different levels of regulation (e.g. number of gene copies, transcriptional and translational regulation, and coding sequence). For example, Pfleger *et al.*[99] achieved a 7-fold increase in mevalonate production in engineered *E. coli* by generating libraries of tunable intergenic regions, recombining various post-transcriptional control elements and screening for the desired relative expression levels in the library.

Alternatively, combinatorial libraries of metabolic pathways can be constructed by efficient homologous recombination techniques at the genome level (Fig. 3A). *In vivo* recombination-mediated genetic engineering (i.e. recombineering), based on homologous recombination events mediated by bacteriophage proteins such as RecET and λ Red recombinases, have been employed successfully for genome editing in several industrial bacterial species (e.g. *E. coli*, *C. glutamicum*, *Mycobacterium smegmatis*, and *P. putida*).[100–104] In these techniques, linear double-stranded or single-stranded DNA substrates, equipped with homology arms that hybridize with the complementary strand in the replication fork, are used for targeted gene replacements, deletions, insertions, inversions, or point mutations. An automated multiplex version of recombineering has been successfully implemented in *E. coli*, thus allowing for the construction of mutant libraries within a few days by simultaneously generating genetic variety in multiple target loci (automated multiplex genomic engineering, MAGE). For instance, the expression levels of 24 genes were optimized in parallel to improve lycopene production in *E. coli*, yielding better results than all previously reported efforts combined.[105] Improvements in this technique include automating the design of oligonucleotides and making multiplex recombineering easily traceable (e.g. trackable multiplex recombineering).[78] Since recombineering systems currently available are not equally active in all bacterial systems, efforts should be also made to identify novel recombinases. The efficient implementation of recombineering (at least when a multiplex approach is needed or when selectable markers are not used) traditionally requires the utilization of bacterial strains with a deficiency of MMR system or RecA, which might yield off target mutations.

A further step in these manipulations is the coupling of recombineering with CRISPR-Cas9 technologies (i.e. CRISPR-Cas9-assisted recombineering).[101,106–109] In these approaches, recombineering mediates the repair of double-stranded DNA (dsDNA) breaks created by CRISPR-Cas9 through the integration of transformed dsDNA fragments, which can have degenerated positions that enable the creation of a higher genetic variation. A new methodology named CRISPR-enabled trackable genome engineering (*CREATE*), which builds upon MAGE by incorporating CRISPR-Cas9 and barcoding technology, has been recently developed.[79] This technique combines the automated design and synthesis of target specific sgRNA cassettes and dsDNA repair fragments, making it possible to mutate tens of thousands of loci in parallel. The

substantial improvements in throughput compared with the previous methods are illustrated by the implementation of CREATE to optimize an isopropanol biosynthetic pathway in *E. coli* by constructing and selecting a combinatorial library with >1,000 mutants.[110] In addition, barcoding technology allows for simultaneous mapping of all mutations, thus facilitating the identification of beneficial mutations that confer phenotypes of interest in ALE. So far, *CREATE* has been only applied to *E. coli*; however, this approach is likely to work in other bacteria for which high-efficiency recombinases are available.

CRISPR-guided locus specific random mutagenesis

Novel and revolutionary tools to create locus-specific random mutations in bacterial genomes have been recently developed based on the CRISPR-Cas9 technology. These editing tools exploit the enzymatic activities of nuclease-deficient Cas9 proteins, avoiding the generation of dsDNA breaks that are lethal or severely toxic for many bacteria. Since no foreign DNA templates are required, these techniques may encounter less regulatory hurdles in industrial applications (Table 1).

A deaminase-mediated targeted mutagenesis tool was firstly developed in *E. coli* (Fig. 3B).[111] A catalytically-inactivated Cas9 protein (dCas9) fused to a cytosine deaminase activity, together with a 20-nucleotide sgRNA, allowed for the introduction of cytosine-to-thymine substitutions at the target locus without compromising cell growth. The substitutions were induced mainly within an approximately five-base window of target sequences on the protospacer adjacent motif-distal side, which can be shifted depending on the length of the sgRNA sequence. The use of a uracil DNA glycosylase inhibitor in combination with a degradation tag resulted in a more robust mutagenesis tool, which allowed for the simultaneous, multiplex editing of six different genes. As this approach does not rely on any host-dependent factor, it is extensible to other relevant bacterial species such as *C. glutamicum* and some *Pseudomonas* sp.[112,113]

Another locus-specific mutagenesis tool, termed EvolvR, has been recently developed in *E. coli* to diversify all nucleotides within a tunable window length (Fig. 3C).[114] EvolvR consists of a CRISPR-guided Cas9 nickase (nCas9) that nicks the target locus fused to a DNA polymerase that performs

error-prone nick translation. The construction of multiple nickase-polymerase variants allowed a wide range of targeted mutation rates (up to 7,770,000-fold greater than the mutagenesis rates observed in wild-type cells) and editing windows with lengths of up to 350 nucleotides.

3. Exploiting the genetic variation of the bacterial population

3.1 Screening and selection technologies

Irrespective of the methodology used to induce genetic variation within a bacterial population, efficient screening and selection methods are required to quickly fish out the MCF variants displaying the (enhanced) phenotype of interest (Fig. 1). To fulfill this purpose, different technologies have been developed over the years (Fig. 4).[23–26] Selection in solid culture media has been extensively used in the improvement of industrial MCFs, since it facilitates the screening of a large number of mutants by visual inspection of plates (Fig. 4). The detection of color changes around the colonies due to the formation of diffusing products or coupled enzymatic activities, and the differential growth in the presence of antibiotics, antimetabolites, or other toxic/stressor compounds, are commonly used criteria to select mutants with enhanced phenotype. These plate-based screening assays, however, may be inefficient for the improvement of complex phenotypes and have an inherent danger of selecting phenotypes that are not reproducible in liquid media, which is the ultimate destination of most production MCFs.

Serial batch liquid cultivations, performed in simple flasks/tubes or in automated sequential batch reactors, can be used to select for mutants with an increased maximum specific growth rate (Fig. 4). To ensure that the selection pressure is faster growth rate (instead of the myriad of process factors besides, e.g. nutrient starvation or oxygen limitation), it is important that mid-exponential phase cultures are used to seed the subsequent batch. Alternatively, continuous cultures in liquid media can be carried out to avoid variation in key culture parameters, e.g. growth rate, growth phase, pH, or cell density; although these long-term cultures may suffer from undesired microbial contamination. The most used continuous culture set-up is the chemostat, which operates under limiting nutrient conditions fixed with the culture dilution rate to select mutants with higher affinity

to the limiting substrate (e.g. mutants displaying high affinity transporters) (Fig. 4). When 'steady-state' conditions are attained in these prolonged cultivations, however, the ability of the microbial population to perform well under more dynamic or nutrient-excess conditions (typical of industrial setups) may become compromised. To overcome some of these limitations, other continuous culture setups and dynamic selection regimes have been developed, e.g. accelerostats, auxostats, and turbidostats.

An alternative culture technology for selecting slow-growing cells with a higher biomass yield is the serial propagation in emulsion.[115] By isolating individual cells in droplets, this strategy eliminates the competition for resources with neighboring cells, thus enabling for the progressive enrichment of mutants with an increased number of viable offspring after several rounds of propagation (Fig. 4). The viability of this concept was demonstrated by the individual compartmentalization of *L. lactis*.[116] A similar approach, based on cell compartmentalization in hollow fibers, enabled the progressive enrichment of *B. subtilis* mutants that secreted large amounts of enzyme into the selection medium.[117] Intrinsic fluorescence properties of certain bioproducts of interest or the use of biocompatible fluorescent substrates that can be modified by endogenous enzymatic activities, sometimes allow for straightforward selection of higher-producer MCFs by fluorescence-activated cell sorting (FACS) equipment (Fig. 4).[118] The development of small-molecule biosensors that enable the expression of fluorescent proteins in response to the intracellular accumulation of a target compound, has been also used for FACS-based sorting.[118–120]

All these culture technologies described above have proved useful in selecting certain phenotypes; however, their application is mostly limited to traits that are directly related to growth. For this reason, the individual screening of mutant collections, by means of small-scale fermentations coupled to the measurement of metabolites or products of interest in the culture broth, is still frequently performed in industrial applications (Fig. 4). A massive parallelization, automation, and miniaturization of these screening assays will be key to accelerate the characterization of mutant collections.

3.2 Coupling phenotypes of interest to bacterial growth

Linking any specific, relevant phenotype to an increase of a macroscopic parameter (e.g. growth rate, substrate affinity) and/or the survival of individual mutant cells in the culture conditions is still one of the biggest experimental challenges in industrial biotechnology, especially if the production of a target compound imposes an additional burden on the cells and reduces their growth. To overcome this issue, several rational experimental design and technological approaches are being developed.

Growth-coupling production strategies have become particularly relevant for Systems Metabolic Engineering over the past decade.[27,121–123] Their central goal is to render the synthesis of a desired metabolite essential for bacterial growth (Fig. 5). The formation of the target metabolite thus becomes an integral part of the MCF metabolic network, making growth the very driving force of production. Growth-coupled MCFs for efficient production can be designed in a such a way that the synthesis of the desired metabolite is necessary to attain maximal growth rate or to confer the ability to grow under a restrictive condition. Therefore, in most cases further improvement of the MCFs performance can be achieved through ALE merely by selecting for maximum growth. Examples of growth-coupling strategies include the targeted manipulation of native metabolism, the introduction of heterologous reactions to force coupling, the use of biosensors to tie cell survival to the target metabolite concentration, and by changing the growth conditions.[27,119] The challenge for the future is to generalize these growth-coupling strategies in order to be applied to more products, both native and heterologous, without the need to implement specialized selection systems for each product separately. In addition, the appearance of selection 'escapers' (i.e. non-producing mutants that are able to survive even in the presence of the selective pressure) and the difficulty to implement these rational approaches in all bacterial hosts (specially non-model bacteria), make the screening and testing of individual mutants still necessary.

3.3 Reverse engineering of relevant phenotypes

An essential part of the process of obtaining industrially-relevant phenotypes in bacterial cell factories is the identification of the genotype-phenotype relationships. In this regards, the best performing MCFs, isolated during the screening and selection tasks described in the previous sections, can either be used directly for production purposes and/or can be further optimized in new cycles of improvement as needed (Fig. 1). However, the investigation of genotype-phenotype relationships can be useful for new rational designs of strains with completely novel engineering principles (i.e. a reverse engineering approach). Evolved populations and/or selected mutant strains with enhanced performance can be analyzed using whole-genome sequencing and Systems Biology tools, with the ultimate goal of identifying causal beneficial mutations, uncovering selection escapers and unravelling new cell mechanisms leading to robust phenotypes. The identified beneficial mutations can be implemented into clean production background strains (i.e. naïve strains), which harbor other features of interest and lack other detrimental or neutral mutations accumulated during the process of improvement (Fig. 1). Again, the combination of different mutations (usually in a non-linear fashion) usually leads to the phenotypes of interest, which would be difficult to predict *a priori*.

4. Conclusion and the way ahead

The examples surveyed in this review clearly illustrate the complexity of dynamic interactions in cellular systems that often hinders industrial applications of Metabolic Engineering approaches. Although the final objective of Systems Metabolic Engineering is the optimization of industrial phenotypes on MCFs in an altogether rational fashion, the need of comprehensive molecular or functional knowledge and the availability of an efficient molecular biology toolbox for the microorganism of interest are still major hurdles in the widespread application of rational approaches in non-model bacteria. The implementation of methodologies based on the core engineering principle of natural evolution (i.e. genetic variation and selection) has proven to offer several advantages for MCFs optimization for industrial applications.

The development of new techniques for manipulating the nature and rate of DNA mutations, both globally (genome) and locally (individual genes), has simplified the construction of combinatorial libraries of mutants and the genetic diversification of bacterial populations. Exploring novel, higher mutational landscapes that are not available in nature, due partly to the conservative nature of the genetic code and the lower mutation rates found in organisms, can also be achieved thanks to some of these techniques. Multiple screening and selection technologies are also available to isolate higher performing mutants from genetically-diverse populations or collection of mutants. Emerging microfluidic technologies will also drive a transition to automation, parallelization and miniaturization of many of screening assays. Understanding genotype-phenotype relationships of higher performing mutants will be also progressively more comprehensive thanks to the rapid development of whole-genome sequencing and Systems Biology tools, which will become useful for implementing new rounds of MCFs improvement.

Although bioproduction strategies implemented thus far rely on the synthesis of a handful of relatively simple products, the future of biotechnological production relies on the production of useful compounds that cannot be easily accessed by other means. The inclusion of chemical elements that seldom (if at all) found in biological systems (e.g. boron, silicon, and halogens) is a particularly attractive feature to explore for biotechnological production of highly added-value compounds containing these atoms.[5,124] The tools and approaches discussed in this article will become essential to engineer phenotypes based on these novel biochemistries, since tying novel reactions involving new-to-nature metabolites requires a multi-level adaptation of MCFs. By taking full advantage of genetic variability, ALE, and screening, the next generation of MCFs will allow to explore the untapped biochemical landscape, expanding the current boundaries of bioproduction beyond the customary compounds currently in the industrial production pipeline.

Acknowledgements

The financial support from The Novo Nordisk Foundation (grants NNF10CC1016517 and *LiFe*, NNF18OC0034818), the Danish Council for Independent Research (*SWEET*, DFF-Research Project 8021-00039B), and the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 814418 (*SinFonia*) to P.I.N. is gratefully acknowledged. L.F.C. is supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 713683 (*COFUNDfellowsDTU*).

Conflict of interest

The authors declare no financial or commercial conflict of interest.

References

[1] A. L. Demain, *Science*. **1981**, *214*, 987.

[2] K. Buchholz, J. Collins, *Applied Microbiology and Biotechnology*. **2013**, *97*, 3747.

[3] K. Timmis, W. M. de Vos, J. L. Ramos, S. E. Vlaeminck, A. Prieto, A. Danchin, W. Verstraete, V. de Lorenzo, S. Y. Lee, H. Brüssow, J. K. Timmis, B. K. Singh, *Microbial Biotechnology*. **2017**, *10*, 984.

[4] K. Timmis, V. de Lorenzo, W. Verstraete, J. L. Ramos, A. Danchin, H. Brüssow, B. K. Singh, J. K. Timmis, *Microbial Biotechnology*. **2017**, *10*, 1137.

[5] P. I. Nikel, V. de Lorenzo, *Metabolic Engineering*. **2018**, *50*, 142.

[6] P. Calero, P. I. Nikel, *Microbial Biotechnology*. **2019**, *12*, 98.

[7] C. N. S. Santos, G. Stephanopoulos, *Current Opinion in Chemical Biology*. **2008**, *12*, 168.

[8] R. Patnaik, *Biotechnology Progress*. **2008**, *24*, 38.

[9] S. Van Dien, *Current Opinion in Biotechnology*. **2013**, *24*, 1061.

[10] M. Gustavsson, S. Y. Lee, *Microbial Biotechnology*. **2016**, *9*, 610.

[11] Z. Gong, J. Nielsen, Y. J. Zhou, *Biotechnology Journal*. **2017**, *12*, 1700014.

[12] J. Bailey, *Science*. **1991**, *252*, 1668.

[13] G. Stephanopoulos, J. Vallino, *Science*. **1991**, *252*, 1675.

[14] G. S. Hossain, S. P. Nadarajan, L. Zhang, T.-K. Ng, J. L. Foo, H. Ling, W. J. Choi, M. W. Chang, *Frontiers in Microbiology*. **2018**, *9*, 155.

[15] L. Martinelli, P. I. Nikel, *Microbial Biotechnology*. **2019**, *12*, 187.

[16] J. W. Lee, D. Na, J. M. Park, J. Lee, S. Choi, S. Y. Lee, *Nature Chemical Biology*. **2012**, *8*, 536.

[17] S. Y. Lee, H. U. Kim, *Nature Biotechnology*. **2015**, *33*, 1061.

[18] K. R. Choi, W. D. Jang, D. Yang, J. S. Cho, D. Park, S. Y. Lee, *Trends in Biotechnology*. **2019**. DOI: 10.1016/j.tibtech.2019.01.003.

[19] J. Becker, O. Zelder, S. Häfner, H. Schröder, C. Wittmann, *Metabolic Engineering*. **2011**, *13*, 159.

[20] H. Yim, R. Haselbeck, W. Niu, C. Pujol-Baxley, A. Burgard, J. Boldt, J. Khandurina, J. D. Trawick, R. E. Osterhout, R. Stephen, J. Estadilla, S. Teisan, H. B. Schreyer, S. Andrae, T. H. Yang, S. Y. Lee, M. J. Burk, S. Van Dien, *Nature Chemical Biology*. **2011**, *7*, 445.

[21] C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, M. Fleck, S. Bajad, G. Dang, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, H. Jiang, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, J. D. Newman, *Nature*. **2013**, *496*, 528.

[22] S. H. Park, H. U. Kim, T. Y. Kim, J. S. Park, S.-S. Kim, S. Y. Lee, *Nature Communications*. **2014**, *5*, 4618.

[23] U. Sauer, in *Advances in Biochemical Engineering/Biotechnology*, Vol. 73 **2001**.

[24] M. Dragosits, D. Mattanovich, *Microbial Cell Factories*. **2013**, *12*, 64.

[25] H. Bachmann, J. T. Pronk, M. Kleerebezem, B. Teusink, *Current Opinion in Biotechnology*. **2015**, *32*. DOI: 10.1016/j.copbio.2014.09.003.

[26] R. Mans, J.-M. G. Daran, J. T. Pronk, *Current Opinion in Biotechnology*. **2018**, *50*, 47.

[27] D. Shepelin, A. Hansen, R. Lennen, H. Luo, M. Herrgård, *Genes*. **2018**, *9*, 249.

[28] Z. Zhu, J. Zhang, X. Ji, Z. Fang, Z. Wu, J. Chen, G. Du, *Applied Microbiology and Biotechnology*. **2018**, *102*, 4615.

[29] W. Liu, R. Jiang, *Applied Microbiology and Biotechnology*. **2015**, *99*, 2093.

[30] M. Jeschek, D. Gerngross, S. Panke, *Current Opinion in Biotechnology*. **2017**, *47*, 142.

[31] H. Lu, J. C. Villada, P. K. H. Lee, *Trends in Biotechnology*. **2019**, *37*, 152.

[32] J. W. Drake, *Proceedings of the National Academy of Sciences*. **1991**, *88*, 7160.

[33] G. I. Lang, A. W. Murray, *Genetics*. **2008**, *178*, 67.

[34] P. L. Foster, in *Methods in Enzymology*, Vol. 204, Elsevier **1991**.

[35] A. Kodym, R. Afza, in *Plant Functional Genomics*, Vol. 236 **2003**.

[36] X. Zhang, X.-F. Zhang, H.-P. Li, L.-Y. Wang, C. Zhang, X.-H. Xing, C.-Y. Bao, *Applied Microbiology and Biotechnology*. **2014**, *98*, 5387.

[37] J. Horst, *Trends in Microbiology*. **1999**, *7*, 29.

[38] A. Oliver, F. Baquero, J. Blazquez, *Molecular Microbiology*. **2002**, *43*, 1641.

[39] T. Swings, B. Van den Bergh, S. Wuyts, E. Oeyen, K. Voordeckers, K. J. Verstrepen, M. Fauvart, N. Verstraeten, J. Michiels, *eLife*. **2017**, *6*, e22939.

[40] A. Greener, M. Callahan, B. Jerpseth, *Molecular Biotechnology*. **1997**, *7*, 189.

[41] O. Selifonova, F. Valle, V. Schellenberger, *Applied and Environmental Microbiology*. **2001**, *67*, 3645.

[42] D. Emlyn-Jones, G. D. Price, T. J. Andrews, *Applied and Environmental Microbiology*. **2003**, *69*, 6427.

[43] A. Endo, M. Sasaki, A. Maruyama, Y. Kurusu, *Bioscience, Biotechnology, and Biochemistry*. **2006**, *70*, 2357.

[44] G. Luan, Z. Cai, F. Gong, H. Dong, Z. Lin, Y. Zhang, Y. Li, *Protein & Cell*. **2013**, *4*, 854.

[45] T. J. Overbeck, D. L. Welker, J. E. Hughes, J. L. Steele, J. R. Broadbent, *Applied and environmental microbiology*. **2017**, *83*, e01120.

[46] A. H. Badran, D. R. Liu, *Nature Communications*. **2015**, *6*, 8425.

[47] L. Zhu, Y. Li, Z. Cai, *Biotechnology for biofuels*. **2015**, *8*, 93.

[48] G. Luan, Z. Cai, Y. Li, Y. Ma, *Biotechnology for Biofuels*. **2013**, *6*, 137.

[49] G. Luan, G. Bao, Z. Lin, Y. Li, Z. Chen, Y. Li, Z. Cai, *New Biotechnology*. **2015**, *32*, 732.

[50] H. H. Chou, J. D. Keasling, *Nature Communications*. **2013**, *4*, 2595.

[51] H. L. Pham, A. Wong, N. Chua, W. S. Teo, W. S. Yew, M. W. Chang, *Nature Communications*. **2017**, *8*, 411.

[52] R. Patnaik, S. Louie, V. Gavrilovic, K. Perry, W. P. C. Stemmer, C. M. Ryan, S. del Cardayré, *Nature Biotechnology*. **2002**, *20*, 707.

[53] Y.-X. Zhang, K. Perry, V. A. Vinci, K. Powell, W. P. C. Stemmer, S. B. del Cardayré, *Nature*. **2002**, *415*, 644.

[54] J. Gong, H. Zheng, Z. Wu, T. Chen, X. Zhao, *Biotechnology Advances*. **2009**, *27*, 996.

[55] D. Biot-Pelletier, V. J. J. Martin, *Applied Microbiology and Biotechnology*. **2014**, *98*, 3877.

[56] T. A. Magocha, H. Zabed, M. Yang, J. Yun, H. Zhang, X. Qi, *Bioresource Technology*. **2018**, *257*, 281.

[57] M. Dai, S. D. Copley, *Applied and Environmental Microbiology*. **2004**, *70*, 2391.

[58] L. H. Reyes, M. P. Almario, J. Winkler, M. M. Orozco, K. C. Kao, *Metabolic Engineering*. **2012**, *14*, 579.

[59] L. Yi, Q. Peng, D. Liu, L. Zhou, C. Tang, Y. Zhou, L. Chai, *Environmental Technology*. **2018**. DOI: 10.1080/09593330.2018.1466918.

[60] J. Merritt, J. S. Edwards, *Metabolic Engineering*. **2004**, *6*, 212.

[61] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, *Molecular systems biology*. **2006**, *2*, 2006.0008.

[62] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, *Cell*. **2013**, *152*, 1173.

[63] J. M. Peters, A. Colavin, H. Shi, T. L. Czarny, M. H. Larson, S. Wong, J. S. Hawkins, C. H. S. Lu, B.-M. Koo, E. Marta, A. L. Shiver, E. H. Whitehead, J. S. Weissman, E. D. Brown, L. S. Qi, K. C. Huang, C. A. Gross, *Cell*. **2016**, *165*, 1493.

[64] X. Liu, C. Gallay, M. Kjos, A. Domenech, J. Slager, S. P. van Kessel, K. Knoops, R. A. Sorg, J. Zhang, J. Veening, *Molecular Systems Biology*. **2017**, *13*, 931.

[65] L. Cui, A. Vigouroux, F. Rousset, H. Varet, V. Khanna, D. Bikard, *Nature Communications*. **2018**, *9*, 1912.

[66] F. Rousset, L. Cui, E. Siouve, C. Becavin, F. Depardieu, D. Bikard, *PLoS genetics*. **2018**, *14*, e1007749.

[67] T. Wang, C. Guan, J. Guo, B. Liu, Y. Wu, Z. Xie, C. Zhang, X.-H. Xing, *Nature Communications*. **2018**, *9*, 2475.

[68] J. M. Peters, B.-M. Koo, R. Patino, G. E. Heussler, C. C. Hearne, J. Qu, Y. F. Inclan, J. S. Hawkins, C. H. S. Lu, M. R. Silvis, M. M. Harden, H. Osadnik, J. E. Peters, J. N. Engel, R. J. Dutton, A. D. Grossman, C. A. Gross, O. S. Rosenberg, *Nature Microbiology*. **2019**, *4*, 244.

[69] M. Kitagawa, T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga, H. Mori, *DNA Research*. **2006**, *12*, 291.

[70] S. Gahlot, A. Joshi, P. Singh, R. Tuteja, M. Dua, A. Jogawat, M. Kumar, S. Raj, V. Dayaman, A. K. Johri, N. Tuteja, *World Journal of Microbiology and Biotechnology*. **2015**, *31*, 1195.

[71] T. A. Tomko, M. J. Dunlop, *Biotechnology for Biofuels*. **2015**, *8*, 165.

[72] Y.-S. Jin, G. Stephanopoulos, *Metabolic Engineering*. **2007**, *9*, 337.

[73] J. R. Borden, S. W. Jones, D. Indurthi, Y. Chen, E. Terry Papoutsakis, *Metabolic Engineering*. **2010**, *12*, 268.

[74] S. A. Nicolaou, S. M. Gaida, E. T. Papoutsakis, *Nucleic Acids Research*. **2011**, *39*, e152.

[75] M. D. Lynch, T. Warnecke, R. T. Gill, *Nature Methods*. **2007**, *4*, 87.

[76] T. van Opijnen, A. Camilli, *Nature Reviews Microbiology*. **2013**, *11*, 435.

[77] V. K. Mutalik, P. S. Novichkov, M. N. Price, T. K. Owens, M. Callaghan, S. Carim, A. M. Deutschbauer, A. P. Arkin, *Nature Communications*. **2019**, *10*, 308.

[78] J. R. Warner, P. J. Reeder, A. Karimpour-Fard, L. B. A. Woodruff, R. T. Gill, *Nature Biotechnology*. **2010**, *28*, 856.

[79] A. D. Garst, M. C. Bassalo, G. Pines, S. A. Lynch, A. L. Halweg-Edwards, R. Liu, L. Liang, Z. Wang, R. Zeitoun, W. G. Alexander, R. T. Gill, *Nature Biotechnology*. **2016**, *35*, 48.

[80] Z. Lin, Y. Zhang, J. Wang, *Biotechnology Advances*. **2013**, *31*, 986.

[81] L. Tripathi, Y. Zhang, Z. Lin, *Frontiers in bioengineering and biotechnology*. **2014**, *2*, 33.

[82] H. Alper, G. Stephanopoulos, *Metabolic Engineering*. **2007**, *9*, 258.

[83] A. M. Lanza, H. S. Alper, in *Synthetic Gene Networks*, Vol. 813 (Eds: W. Weber, M. Fussenegger), Humana Press, Totowa, NJ **2012**.

[84] F. Tan, B. Wu, L. Dai, H. Qin, Z. Shui, J. Wang, Q. Zhu, G. Hu, M. He, *Microbial cell factories*. **2016**, *15*. DOI: 10.1186/s12934-015-0398-y.

[85] X.-W. Guo, Y. Zhang, L.-L. Li, X.-Y. Guan, J. Guo, D.-G. Wu, Y.-F. Chen, D.-G. Xiao, *Biotechnology for Biofuels*. **2018**, *11*, 307.

[86] X. Gao, X. Yang, J. Li, Y. Zhang, P. Chen, Z. Lin, *Microbial Cell Factories*. **2018**, *17*, 118.

[87] H. Cao, J. Villatoro-Hernandez, R. D. O. Weme, E. Frenzel, O. P. Kuipers, *Metabolic Engineering*. **2018**, *49*, 143.

[88] T. Chen, J. Wang, R. Yang, J. Li, M. Lin, Z. Lin, *PLoS ONE*. **2011**, *6*, e16228.

[89] H. Wang, L. Yang, K. Wu, G. Li, *Microbial Cell Factories*. **2014**, *13*, 10.

[90] K.-S. Park, Y.-S. Jang, H. Lee, J.-S. Kim, *Journal of Bacteriology*. **2005**, *187*, 5496.

[91] J. Y. Lee, B. H. Sung, B. J. Yu, J. H. Lee, S. H. Lee, M. S. Kim, M. D. Koob, S. C. Kim, *Nucleic Acids Research*. **2008**, *36*, e102.

[92] J. Y. Lee, K. S. Yang, S. A. Jang, B. H. Sung, S. C. Kim, *Biotechnology and Bioengineering*. **2011**, *108*, 742.

[93] S. Basak, R. Jiang, *PLoS ONE*. **2012**, *7*, e51179.

[94] K. Ochi, *The Journal of Antibiotics*. **2017**, *70*, 25.

[95] G. Wang, T. Hosaka, K. Ochi, *Applied and Environmental Microbiology*. **2008**, *74*, 2834.

[96] Y. Tanaka, K. Kasahara, M. Izawa, K. Ochi, *Bioscience, Biotechnology, and Biochemistry*. **2017**, *81*, 1636.

[97] S. Kosuri, G. M. Church, *Nature Methods*. **2014**, *11*, 499.

[98] A. Casini, M. Storch, G. S. Baldwin, T. Ellis, *Nature Reviews Molecular Cell Biology*. **2015**, *16*, 568.

[99] B. F. Pfleger, D. J. Pitera, C. D. Smolke, J. D. Keasling, *Nature Biotechnology*. **2006**, *24*, 1027.

[100] K. Murphy, in *Mycobacteria Protocols*, Vol. 1285 (Eds: T. Parish, D. M. Roberts), Third edition ed., Humana Press, New York **2015**.

[101] G. Pines, E. F. Freed, J. D. Winkler, R. T. Gill, *ACS Synthetic Biology*. **2015**, *4*, 1176.

[102] Y. Huang, L. Li, S. Xie, N. Zhao, S. Han, Y. Lin, S. Zheng, *Scientific Reports*. **2017**, *7*, 7916.

[103] J. A. Sawitzke, L. C. Thomason, N. Costantino, D. L. Court, in *Reference Module in Life Sciences*, Elsevier, **2017**. DOI: 10.1016/B978-0-12-809633-8.07023-0.

[104] D. E. Ricaurte, E. Martínez-García, Á. Nyerges, C. Pál, V. de Lorenzo, T. Aparicio, *Microbial Biotechnology*. **2018**, *11*, 176.

[105] H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest, G. M. Church, *Nature*. **2009**, *460*, 894.

[106] W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, *Nature Biotechnology*. **2013**, *31*, 233.

[107] Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang, S. Yang, *Applied and Environmental Microbiology*. **2015**, *81*, 2506.

[108] C. R. Reisch, K. L. J. Prather, *Scientific Reports*. **2015**, *5*, 15096.

[109] C. Ronda, L. E. Pedersen, M. O. A. Sommer, A. T. Nielsen, *Scientific Reports*. **2016**, *6*, 19452.

[110] L. Liang, R. Liu, A. D. Garst, T. Lee, V. S. i. Nogué, G. T. Beckham, R. T. Gill, *Metabolic Engineering*. **2017**, *41*, 1.

[111] S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, *Nature Microbiology*. **2018**, *3*, 423.

[112] W. Chen, Y. Zhang, Y. Zhang, Y. Pi, T. Gu, L. Song, Y. Wang, Q. Ji, *iScience*. **2018**, *6*, 222.

[113] Y. Wang, Y. Liu, J. Liu, Y. Guo, L. Fan, X. Ni, X. Zheng, M. Wang, P. Zheng, J. Sun, Y. Ma, *Metabolic Engineering*. **2018**, *47*, 200.

[114] S. O. Halperin, C. J. Tou, E. B. Wong, C. Modavi, D. V. Schaffer, J. E. Dueber, *Nature*. **2018**, *560*, 248.

[115] H. Bachmann, F. J. Bruggeman, D. Molenaar, F. Branco dos Santos, B. Teusink, *Current Opinion in Microbiology*. **2016**, *31*, 109.

[116] H. Bachmann, M. Fischlechner, I. Rabbers, N. Barfa, F. Branco dos Santos, D. Molenaar, B. Teusink, *Proceedings of the National Academy of Sciences*. **2013**, *110*, 14302.

[117] D. Naki, C. Paech, G. Ganshaw, V. Schellenberger, *Applied Microbiology and Biotechnology*. **1998**, *49*, 290.

[118] J.-L. Lin, J. M. Wagner, H. S. Alper, *Biotechnology Advances*. **2017**, *35*, 950.

[119] T. C. Williams, I. S. Pretorius, I. T. Paulsen, *Trends in Biotechnology*. **2016**, *34*, 371.

[120] A. Carpenter, I. Paulsen, T. Williams, *Genes*. **2018**, *9*, 375.

[121] Y.-S. Tai, M. Xiong, P. Jambunathan, J. Wang, J. Wang, C. Stapleton, K. Zhang, *Nature Chemical Biology*. **2016**, *12*, 247.

[122] S. Aslan, E. Noor, A. Bar-Even, *Biochemical Journal*. **2017**, *474*, 3935.

[123] A. von Kamp, S. Klamt, *Nature communications*. **2017**, *8*, 15956.

[124] F. H. Arnold, *Angewandte Chemie International Edition*. **2018**, *57*, 4143.

Table 1. Overview of the main techniques implemented to increase genetic variation in bacterial populations.

Technique	Description	Advantages	Disadvantages	Main applications
Whole-genome random mutagenesis	Increase of intrinsic mutation rate with mutagen agents (*physical/chemical mutagenesis*) or dominant mutator alleles (*gene-encoded mutagenesis*), followed by screening and selection	- Technically simple - Applicable to almost any bacteria - Non-GMOs [a] - Mutagenesis can be coupled to selection - *A priori* genome knowledge not needed	- Low frequency of beneficial mutations - Accumulation of detrimental mutations - Narrow mutational spectra - Dangerous mutagens needed - Difficult to identify mutations responsible for observed phenotypes	Improvement of simple phenotypes in non-well characterized bacteria
Construction of whole-genome mutant libraries	Construction of mutant libraries (*knock-out*, *knock-down* and *overexpression libraries*), followed by screening and selection	- Applicable to many bacteria - Direct link between genotype and phenotype - Reutilization of libraries for several projects	- Laborious and time-consuming - Only individual contributions of genes can be identified - Variation followed by selection	Identification of molecular mechanisms underlying complex phenotypes of interest
Genome shuffling	Libraries of shuffled strains *via* recursive fusion of protoplasts of multiple parental strains, followed by	- Improvement of complex phenotypes - Non-GMOs - High genetic variation generated	- Variation followed by selection - Low efficiency in Gram-negative bacteria	Improvement of complex phenotypes in non-well characterized Gram-positive bacteria

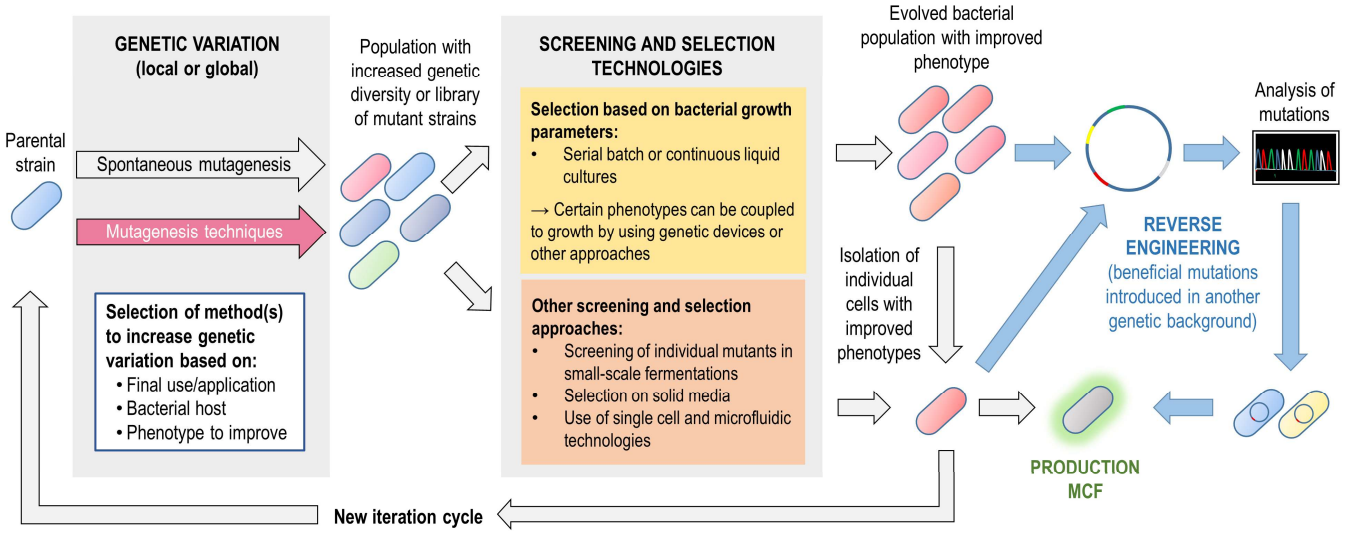
	screening and selection	- *A priori* genome knowledge not needed	- Difficult to identify mutations responsible for observed phenotypes	
Engineering transcriptional regulatory proteins (TRPs)	Construction of libraries of endogenous, exogenous, or artificial TRPs variants, followed by screening and selection	- Applicable to many bacteria - Rapid improvement of complex phenotypes - *A priori* genome knowledge not needed	- Variation followed by selection - Laborious and time-consuming - Only access to latent cellular potential - Genetic tools and protocols required	Optimization of stress tolerance phenotypes
Ribosome and RNA polymerase engineering	Isolation of spontaneous mutants with drug-resistant ribosomes and RNA polymerases, followed by screening and selection	- Technically simple - Applicable to almost any bacteria - Non-GMOs - *A priori* genome knowledge not needed	- Variation followed by selection - Low frequency of drug-resistant mutants - Global pleotropic effects - Only access to latent cellular potential	Overproduction of secondary metabolites and activation of silent gene clusters in Actinobacteria
Combinatorial pathway engineering	Construction of combinatorial libraries for the optimization of metabolic routes (encoded in plasmids or genome), followed by screening and selection	- Optimization at different levels of regulation (e.g. transcriptional and translational, coding sequences, number of gene copies)	- Variation followed by selection - Off-target mutations (CRISPR and recombineering) - Applicable to a number reduced of bacterial species	Optimization of heterologous biosynthetic pathways

CRISPR-guided local random mutagenesis	Creation of locus-specific random mutations *via* CRISPR-guided nuclease-deficient Cas9 proteins fused to deaminases or engineered DNA polymerases	- Donor DNA template not used (potentially non-GMOs) - Mutagenesis can be coupled to selection - High mutation rates - Broad mutational spectra	- Off-target mutations - Applicable to a reduced number of bacterial species	Phenotype optimization by coupling random mutagenesis of one or a few genes with selection

[a] GMO, genetically-modified organism.

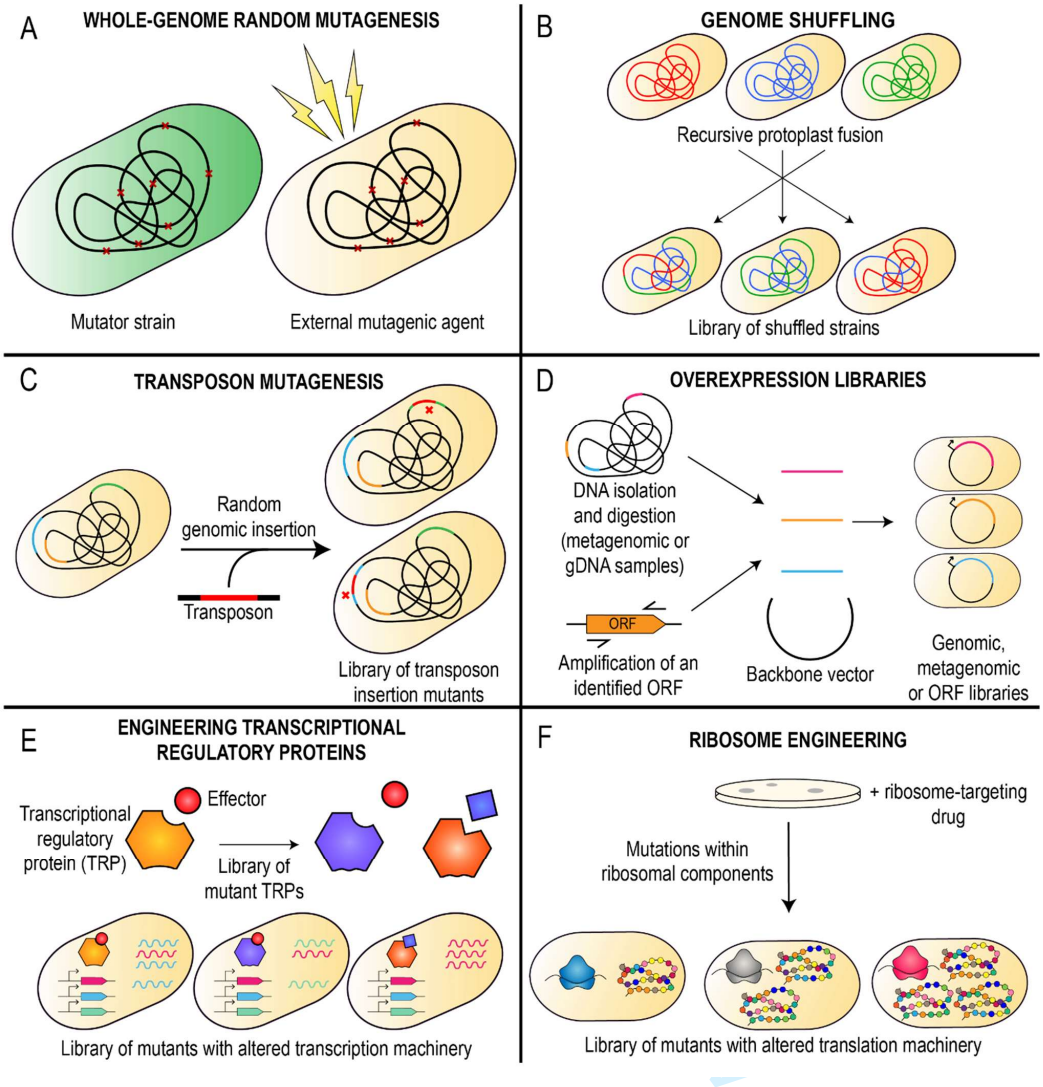
FIGURE LEGENDS

Figure 1. Overview of evolutionary approaches to engineer bacterial cell factories.



Genetic variation in a bacterial population can be generated through spontaneous mutagenesis events and/or artificially increased through mutagenesis techniques. The characteristics of the bacterial host, the type of phenotype to be improved, and the expected application/use of the resulting strain define the approach to be used. Regardless of the methodology used to increase genetic variation within a bacterial population and/or to generate a combinatorial library of strains, efficient screening and selection methods are required to quickly fish out the mutants displaying the (enhanced) phenotype of interest. The best-performing mutants can be used directly for production purposes and/or can be further optimized in new cycles of improvement. Evolved populations and/or selected mutant strains with (enhanced) phenotype can also be analyzed using whole-genome sequencing and Systems Biology tools, with the ultimate goal of identifying causal beneficial mutations that can be introduced into clean production background strains (i.e. reverse engineering approaches). MCF, microbial cell factory.

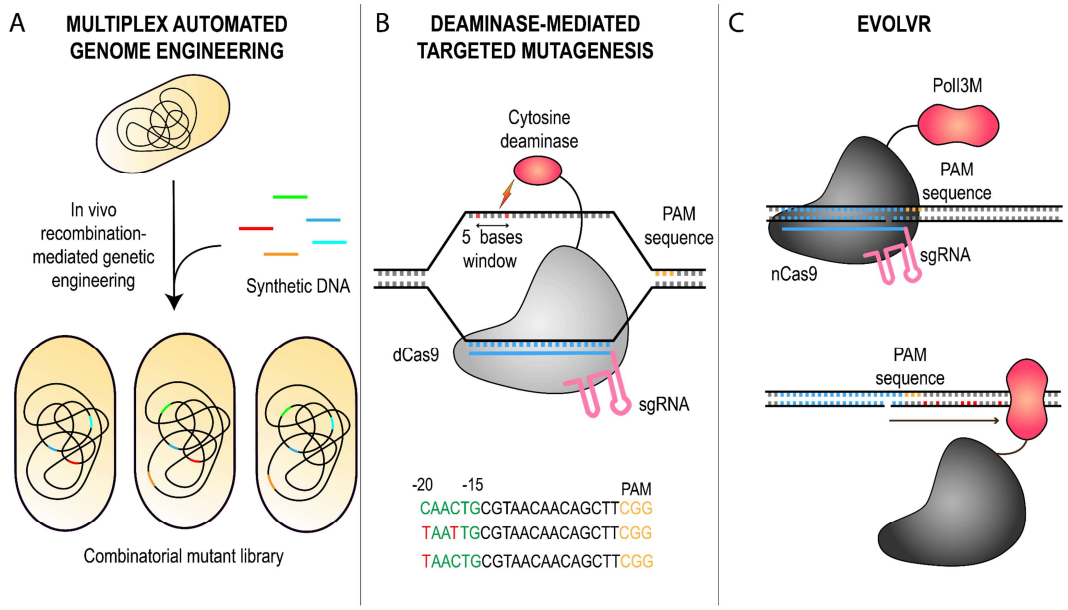
Figure 2. Whole-genome approaches to increase and exploit genetic variation.



(A) Whole-genome random mutagenesis induced by chemical or physical mutagens or mediated by mutator alleles. **(B)** Construction of a library of shuffled strains through the recursive fusion of protoplasts from several parental strains. **(C)** Creation of a library of transposon insertional mutants through the random genomic integration of an antibiotic resistance gene mediated by a transposase activity. **(D)** Construction of overexpression libraries using genomic DNA (gDNA), either from the bacterial host or from other organisms displaying phenotypic traits of interest, or using synthetic DNA or metagenomic samples. ORF, open reading frame. **(E)** Construction of a collection of mutants, which overexpress variants of transcriptional regulatory proteins (TRPs), with altered transcriptional responses that elicit the appearance of phenotypes of interest. **(F)** Isolation

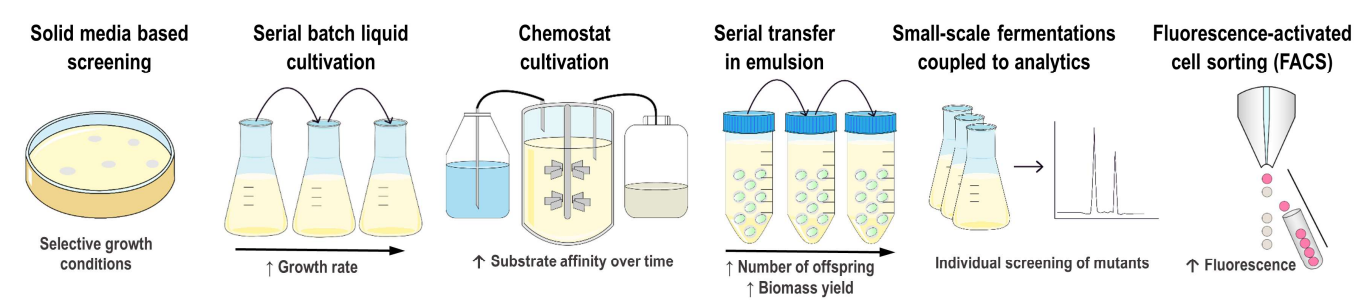
of spontaneous mutants resistant to ribosome-targeting drugs that have altered translational responses leading to specialized phenotypes.

Figure 3. Targeted mutagenesis techniques to increase and exploit genetic variation.



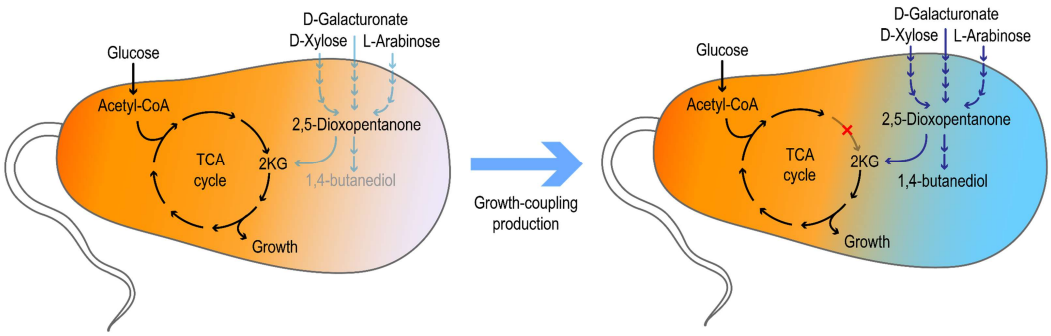
(A) Construction of combinatorial libraries of mutants by *in vivo* homologous recombination techniques (recombineering) using linear DNA substrates equipped with homology arms. **(B-C)** CRISPR-guided locus specific random mutagenesis approaches. The catalytically-inactivated Cas9 protein (dCas9) fused to a cytosine deaminase activity, together with a 20-nucleotide single-guide RNA (sgRNA), allows the introduction of cytosine-to-thymine substitutions in a window located approximately 15–20 bases upstream of the PAM (protospacer adjacent motif) sequence.[111] The EvolvR tool has been developed to diversify all nucleotides within a tunable window length.[114] EvolvR consists of a CRISPR-guided Cas9 nickase (nCas9) that nicks the target locus fused to a DNA polymerase that performs error-prone nick translation (PolI3M).

Figure 4. Comparison of screening and selection technologies applied to exploit the genetic variation of bacterial populations or mutant libraries.



Screening and selection technologies applied to exploit the (natural or induced) genetic variation of bacterial populations or libraries of mutant strains. Regardless of the methodology used to increase genetic variation, efficient screening and selection methods are required to quickly fish out mutants displaying the (enhanced) phenotype of interest. Different liquid culture setups (serial batch liquid cultivation, chemostat or serial transfer in emulsion) can be used to enrich bacterial populations with best performing mutants if the phenotype to be optimized is coupled to growth parameters (e.g. growth rate, substrate affinity, or biomass yield). Selective, solid culture media or fluorescence-activated cell sorting (FACS) based-screening can be applied to isolate individual mutants with enhanced performance for certain phenotypes. Additionally, small-scale fermentations of individual mutants coupled to analytic measurements can be performed to isolate superior mutants for almost any phenotype.

Figure 5. Growth-coupled production strategies to optimize the synthesis of a target product.



A non-phosphorylative metabolism coupled to the biosynthesis of the tricarboxylic acid (TCA)-cycle intermediate 2-ketoglutarate (2KG) from lignocellulosic sugars was engineered in *E. coli*.[121] This growth-based selection platform uses a 2KG auxotroph (identified with a red cross), which allowed to identify several gene clusters mediating the assimilation of these sugars into the TCA cycle via the non-native intermediate 2,5-dioxopentanoate. The application of this platform was demonstrated by creating an artificial pathway for the production of 1,4-butanediol from D-xylose, L-arabinose or D-galacturonate. Acetyl-CoA, acetyl-coenzyme A.