Bacterial genome reductions: Tools, applications, and challenges

Nicole LeBlanc [1]* and Trevor C. Charles [1,2]

[1]Department of Biology, University of Waterloo, Waterloo, ON, Canada, [2]Metagenom Bio Life Science Inc., Waterloo, ON, Canada

Bacterial cells are widely used to produce value-added products due to their versatility, ease of manipulation, and the abundance of genome engineering tools. However, the efficiency of producing these desired biomolecules is often hindered by the cells' own metabolism, genetic instability, and the toxicity of the product. To overcome these challenges, genome reductions have been performed, making strains with the potential of serving as chassis for downstream applications. Here we review the current technologies that enable the design and construction of such reduced-genome bacteria as well as the challenges that limit their assembly and applicability. While genomic reductions have shown improvement of many cellular characteristics, a major challenge still exists in constructing these cells efficiently and rapidly. Computational tools have been created in attempts at minimizing the time needed to design these organisms, but gaps still exist in modelling these reductions *in silico*. Genomic reductions are a promising avenue for improving the production of value-added products, constructing chassis cells, and for uncovering cellular function but are currently limited by their time-consuming construction methods. With improvements to and the creation of novel genome editing tools and *in silico* models, these approaches could be combined to expedite this process and create more streamlined and efficient cell factories.

KEYWORDS

synthetic biology, bacteria, genome reduction, genome engineering, minimal genome

1 Introduction

From researching disease to producing various materials applicable in countless industries including food, pharmaceuticals, and textiles, bacteria have expanded what is possible and contributed to incredible advancements. This is highlighted by the industrial use of bacteria that naturally produce value added products such as antibiotics, amino acids, therapeutic products, biofuels, and materials for textiles and medical devices to name a few (Quillaguamán et al., 2005; Wendisch et al., 2006; Olano et al., 2008; De Eugenio et al., 2010; Choi and Lee, 2013; Shi et al., 2014; Jiang et al., 2018; Samrot et al., 2021). However, these processes rely, for the most part, on organisms that evolved in nature, and were not created for these industrial processes. Consequently, they have many

cellular functions that are irrelevant to the desired application which limits the efficiency of producing the target end-product (Choe et al., 2016; Weiser et al., 2019). Breakthroughs in DNA synthesis and sequencing technologies and the ever-increasing data on metabolic pathways have enabled the creation of novel methods to engineer optimized bacteria with more efficient and higher yielding production in a multitude of applications.

Some of the greatest challenges within a cell that limit its applicability are complexity and absence of predictability. All bacterial cells are composed of a range and diversity of molecules interacting in complex networks for countless cellular functions (Fehér et al., 2007). This is further complicated by random mutations, some of them caused by mobile elements, that lead to unpredictable cellular behavior. To combat these issues, genomic reductions have been performed to remove dispensable genes from the genome. Genome sequencing and functional assays have revealed essential genes like those responsible for core survival, those involved in industrially relevant processes, and non-essential genes that contribute to genome instability and superfluous or unknown cellular functions (Fehér et al., 2007; Park et al., 2014). When using a cell to produce a specific biomolecule in a defined environment, many genes whose functions do not contribute to the intended process could be candidates for removal. This would allow for the creation of tailored cell factories with improved physiological characteristics for the specific application. This can also be more generalized to create a chassis with only the genes required for cell survival and proliferent growth that can be further engineered for downstream applications.

2 Benefits of a reduced genome cell

Prior to the publication of the first full bacterial genome, genomic reductions in *E. coli* were suggested based on the presence of genes unnecessary for growth under defined conditions (Koob et al., 1994). Further, the use of *Mycoplasma* strains as a model for minimal genome construction was suggested due to their naturally minimized genomes (Morowitz, 1984). Now, with countless constructed reduced genome strains, the benefits of these smaller genomes are evident. First, decreasing the number of genes and functions within a cell reduces the complexity of the organism and makes modelling of its metabolism and functional predictions much simpler (Choe et al., 2016). Next, genomic stability has been greatly improved in genome reduced strains. This is highlighted by the improved growth characteristics including genomic stability following the deletion of biosynthetic clusters in *Streptomyces chattanoogensis* (Bu et al., 2019). Also in *E. coli*, the deletion of error-prone DNA polymerases, that are expressed during SOS response and implicated in induced mutagenesis, resulted in a 50% decrease in the spontaneous mutation rate and improved genetic stability (Csörgo et al., 2012). Another

advantage is the possibility of cells requiring less energy to replicate the smaller genome as well as lower transcriptional and translational costs. This correlation has not been fully investigated but many reduced genome strains display faster growth rates and higher cell density that could be attributed to these factors (Kolisnychenko et al., 2002; Mizoguchi et al., 2007; Zhu et al., 2017; Qiao et al., 2022). For example, the genome of *Lactococcus lactis* N8 was reduced by 6.9% by deleting prophages and genomic islands resulting in a shortened generation time by 17% (Qiao et al., 2022). Other observed benefits to genome-reduced strains include increased production of desired products, and improved transformation efficiency (Bu et al., 2019). Finally, the ease of genetic manipulation is one of the biggest advantages of a reduced genome strain. With improved growth characteristics, simpler metabolism, and fewer functions being performed within the cell, there is the potential to use it for many downstream applications such as expressing heterologous genes and producing biomolecules using tailored metabolic pathways.

3 Genome reductions occurring in nature

The idea of reducing bacterial genomes to improve physiological characteristics and create optimized hosts stems from this process occurring naturally through evolution. Cells evolve under strong selective pressure to maintain homeostasis against environmental changes. This forces cells to expand metabolic and signalling pathway redundancy to become more robust, which has resulted in increased genome sizes (Kurasawa et al., 2020). Thus, when cells are grown in laboratory and controlled conditions, this redundancy becomes unnecessary, making it possible to significantly reduce the size of the genome. This is displayed in obligately symbiotic bacteria that have undergone significant genome reduction through evolution (Choe et al., 2016). For example, *Buchnera* sp., an insect endosymbiont and a relative of *E. coli* (4.5 Mb genome), have genomes that are as small as 450 kb (Wernegreen, 2002; Fehér et al., 2007), approximately a 10th of the size of their *E. coli* relatives. Similar reductions have been observed in other obligate symbiont bacterial classes including member of the Gammaproteobacteria (*Buchnera aphidicola, Wigglesworthia, Blochmannia*) and Spirochaetes (*Borrelia burgdorferi*) (Fraser et al., 1995; Andersson et al., 1998). In these obligate symbionts, smaller genomes are possible because these bacteria have access to functions and metabolites from their host, obviating the need to encode them in their own genomes, compared to their free-living bacterial relatives. When a free-living bacterium becomes restricted to a host, the process of evolutionary genomic reductions begins with large and small deletions of genes no longer required, often accompanied by chromosomal rearrangements (McCutcheon and Moran, 2012; Bobay and

Ochman, 2017). This further develops with long-term obligate symbionts of pathogens which lose many pseudogenes and almost all mobile elements, resulting in a more stable chromosome (Bobay and Ochman, 2017). To reach those tiny-genome symbionts, there is an ongoing gene loss as the organism evolves to survive in the given conditions (Bobay and Ochman, 2017). Compared to the smallest free-living bacteria, *Mycoplasma genetalium* with a genome of 580 kb, genome-reduced symbiotic bacteria have genomes two to four times smaller such as *Candidatus Tremblaya princeps,* a betaproteobacteria with a genome of 138 kb, *Candidatus Sulcia muelleri*, a bacteroidetes with a genome of 245 kb, and *Candidatus Hodgkinia cicadicola*, an alphaproteobacteria with a genome of 143 kb (McCutcheon and Moran, 2012).

While genomic reductions were first observed in obligately symbiotic bacteria, studies comparing genomic data reveal that they are also prevalent in free-living bacterial genomes (Wolf and Koonin, 2013; Albalat and Canestro, 2016). Genome evolution occurs both through expansion by horizontal gene transfer and duplication events, and through genomic reduction from large-scale gene deletion (Thomas and Nielsen, 2005; Lynch, 2006). Interestingly, bacterial genomes have a bias towards deletion events over expansion events and these large-scale deletions can occur in a relatively short evolutionary time frame, as highlighted by experimental evolution studies (Mira et al., 2001; Kunin and Ouzounis, 2003; Nilsson et al., 2005; Koskiniemi et al., 2012; Lee and Marx, 2012). Another important aspect to consider is the deleterious effect large-scale genomic reductions can have on a cell and how the organization of the genome needs to protect against this (Fehér et al., 2007). By computationally analysing the metabolism of 55 bacterial species at the genome scale, it was elucidated that the bacterial genomes are organized in such a way as to increase robustness of metabolic genes against the deletion of contiguous genes (Hosseini and Wagner, 2018). This is the result of segregation of essential and non-essential metabolic gene clusters and the separation of synthetic lethal gene pairs. The adaptive forces that favor this organization despite genomic elements like transposons that cause random genome rearrangements have been identified by computationally modelling a reduced bacterial cell. (Hosseini and Wagner, 2018).

4 Designing a reduced-genome bacterium

4.1 Determining gene essentiality

Prior to reducing an organism's genome, studies are often performed to determine which genes are essential. Essentiality is entirely dependent on the environment the cells are grown in and their desired application. Genes that would be considered essential when grown on minimal media with limited nutrient supplementation, such as amino acid synthesis genes, would not
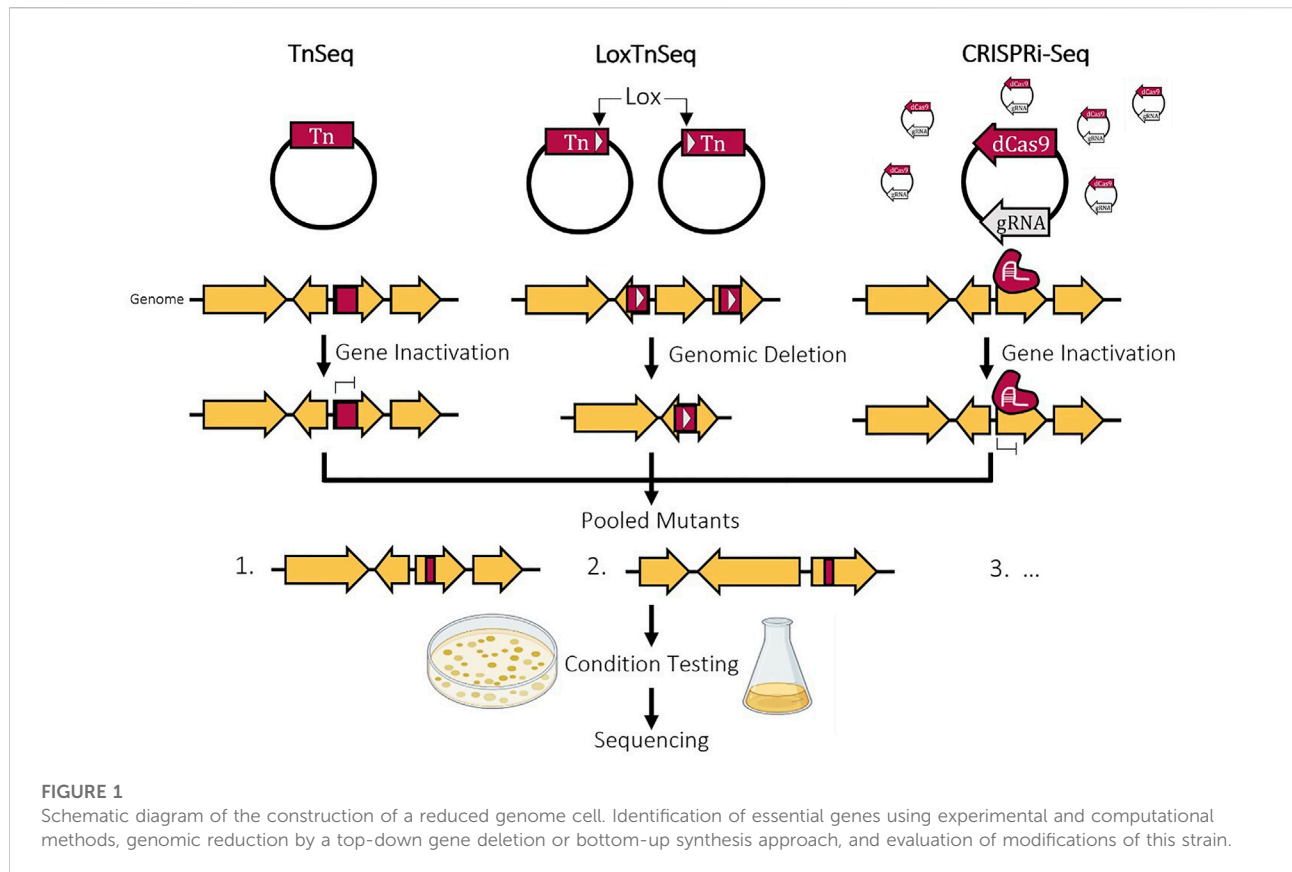
necessarily be essential when grown on nutrient-rich complex media (Baba et al., 2006; Patrick et al., 2007). Frequently, there are already target genes that can be identified for deletion, such as genes in competing metabolic pathways when attempting to produce a specific biomolecule. But when looking to make significant genomic reductions, it is important to evaluate what genes are needed for survival in the given conditions. While the focus of this review is the application of gene essentiality determination studies in constructing reduced-genome strains, there are other applications of this data that are of value to note. Identifying genes essential to an organism's survival will advance overall understanding of the fundamental principles of life (Moya et al., 2009; Jewett and Forster, 2010; Juhas et al., 2011; Commichau et al., 2013). Also, genes that are essential to a bacterium's survival could be targets for novel antimicrobial development (Bumann, 2008; Juhas et al., 2012a, 2012b). Three approaches are commonly taken for determining gene essentiality in bacteria: comparative genomics, large-scale gene inactivation studies, and *in silico* modelling. All three of these routes have various advantages and disadvantages to the genome design and determination of essential genes often involves combining data from each approach.

4.1.1 Essentiality determination by comparative genomics

First, the comparative genomics approach compares the genome of the target species to the genome of both closely and distantly related organisms to determine a core set of essential genes. Comparing the genomes of distantly related bacteria, 262 genes were found shared amongst them (Mushegian and Koonin, 1996). When more species were added into this comparison, the set of genes shrinks to 63 common core genes, mostly all related to basic components for gene expression and replication, identified across the Bacteria and Eukarya domains (Koonin, 2003). This can give a good idea of genes necessary for life across all species, but these alone would not be sufficient for survival. Accessory genes and metabolic genes are also required for survival in the given environment (Tarnopol et al., 2019). Additionally, proteins with similar functions may not necessarily have sequence similarities (Riley and Serres, 2000). Thus, the size of the minimal gene set may be underestimated as only those shown to be conserved across all species tested will be considered 'essential' and this also doesn't include environmental dependence of genes (Fehér et al., 2007).

4.1.2 Essentiality determination by gene inactivation

Large-scale gene inactivation studies, such as transposon mutagenesis, can be utilized. These types of analysis are able to score mutants on their ability to survive. With technological advancements, the set of known essential genes is always evolving. This is highlighted by *B. subtilis* initially having a

FIGURE 1
Schematic diagram of the construction of a reduced genome cell. Identification of essential genes using experimental and computational methods, genomic reduction by a top-down gene deletion or bottom-up synthesis approach, and evaluation of modifications of this strain.

set of 271 ORFs considered essential in 2003 and reduced even further to 253 in 2014 (Kobayashi et al., 2003; Juhas et al., 2014). Transposon sequencing (TraDIS, Inseq, TnSeq) is a very common tool to use to identify essential gene regions in the genome and has been applied in many different species and for many different growth conditions (Goodman et al., 2009; Langridge et al., 2009; Wong et al., 2016; Dejesus et al., 2017; Higgins et al., 2017; Baby et al., 2018; de Maat et al., 2020; Johnson et al., 2020; Matern et al., 2020). This involves the creation of a saturated transposon library where each genomic region is interrupted and theoretically inactivated by a transposon (Figure 1). By culturing this library in different growth conditions, the cells that contain a transposon insertion in an essential genomic region will not survive. The location of the remaining transposons is determined using sequencing, identifying the genes that are nonessential. This can then be taken one step further by assigning a fitness score based on the transposon insertion frequency in every gene. This helps elucidate non-essential genes and quasi-essential genes that contribute to cell growth but are not essential to survival. While this method can rapidly identify thousands of nonessential genes across the entire genome, it only shows the effect of single mutations and gene knockouts on viability. This fails to capture epistatic

interactions in the genome when multiple deletions are made such as synthetic lethal pairs or deletion combinations that could hinder the ability of the cell to grow rapidly (Yu et al., 2002, 2006). Also, the loss of individual genes can affect or control the essentiality of other genes. When studying *M. pneumoniae* and *Mycoplasma agalactiae*, the genes involved in the production of an essential metabolite in a linear metabolic pathway were found to be essential (Montero-Blay et al., 2020). However, with two pathways that produce the same essential metabolite, the genes from both were classified as fitness genes and not essential, showing that both can be deleted but in reality, one of the two is necessary (Montero-Blay et al., 2020). The opposite of this is also seen where essential genes can be rendered nonessential in response to the deletion of a different gene in the genome. The issues with epistatic interactions and redundancies within the genome can be addressed by performing multiple rounds of Tn-Seq after genomic deletions but this is time consuming and labour intensive (Hutchison et al., 2016). Also, this does not take into account genes that do not act as part of larger networks.

To complement the identification of individual nonessential genes, a method combining Tn-Seq and Cre-LoxP named LoxTnSeq was developed to highlight large

TABLE 1 Comparison between TnSeq and CRISPRiSeq methods to determine gene essentiality.

Characteristic	TnSeq	CRISPRiSeq
Mutant Library Generation	**Mutant library construction:** Conjugation of strain harboring suicide plasmid with transposon and transposase. Induce transposase activity, select, and pool resistant mutants	**gRNA library construction:** Random selection and synthesis of 10,000s of gRNAs, PCR amplification of synthesised guides, assembly into dCas9/guide expression plasmid, and transformation
		Mutant library construction
		Plasmid extraction from sgRNA library and electroporation into target *E. coli* strains
Library Screening	Growth of mutant library in desired conditions/media over multiple generations	Induction of dCas9 expression and growth of mutant library in desired conditions/media over multiple generations
Sequencing Preparation	Varies depending on the transposon used but for Mariner transposons; genome digestion to isolate transposon, blunting, and PCR with sequencing adaptors	PCR to amplify gRNAs
Duplicated Gene Screening	Only 1 transposon insertion per cell so cannot identify and screen those with duplicated genes	sgRNA can bind to multiple copies of the same region so can more effectively determine essentiality of those regions
Inactivation control	No control	Can modify the repression level by altering the sgRNA homology
Genomic target	Random integration[a]	Targeted integration by sgRNA sequence

[a]Depending on the transposon used (ie. mariner based transposons integrate into TA, dinucleotides but randomly across the genome).

genomic regions that are nonessential (Shaw et al., 2020). By modifying the transposon to also carry a lox site, two lox sites can be randomly inserted into the genome and a deletion of random size is created by activating the Cre recombinase, which causes recombination between two lox sites orientated in the same direction. These deletion mutants can then be screened in different conditions and sequenced, same as Tn-Seq. LoxTnSeq was applied in *Mycoplasma pneumoniae* yielding a pool of mutants with 285 unique deletions ranging from 50 bp to 28 kb (21% of the total genome) (Shaw et al., 2020). Following deletion, a pitfall in previous methods is that the selection of the deletion mutant relies on screening for the loss of antibiotic resistance (Tsuge et al., 2007; Leprince et al., 2012). LoxTnSeq is designed to address this by forming an inactive lox72 site after deletion which cannot be acted on by Cre (Shaw et al., 2020). The constitutively expressed Cre recombinase is lethal in *M. pneuomiae* in the presence of active loxP sites, killing cells that do not have the deletion (Shaw et al., 2020).

Another approach to determining gene essentiality is CRISPR interference sequencing (CRISPRi-Seq), which employs a catalytically dead mutant of Cas9 that silences gene expression rather than causing a double stranded break (Figure 1) (Rousset et al., 2018; Liu et al., 2021a). By using a pool of over 90,000 sgRNAs that target random genomic loci within *E. coli*, 21% of the previously annotated essential genes were found to be nonessential. Initially, there were toxicity issues when dCas9 was combined with some specific PAM-proximal sequences in sgRNAs (Cui et al., 2018). After analysing the guides within the constructed sgRNA library, the pool was filtered down to ~23,000 sgRNAs. Like Tn-Seq, after integration of the sgRNA library, silenced essential genes will be depleted and

those with repression in nonessential regions will survive. Thus, following sequencing, the number of reads from the remaining cells can be compared to the initial pool to determine gene essentiality. Another similar study was run using around 60,000 guides within *E. coli* and despite having almost 3 times the number of guides in the library, both studies achieved similar results (Wang et al., 2018). Improving sgRNA design methods and minimizing the library size would ultimately enable more experiments to run concurrently and decrease the cost of DNA sequencing and synthesis. A comparison between using CRISPRi-Seq and TnSeq to determine gene essentiality can be found in Table 1. One advantage of this method over Tn-Seq is the ability to study duplicated regions in the genome. The gRNAs will target a specific region in the genome based on homology surrounding a PAM site, so multiple copies of the same gene region given identical sequences can be repressed at the same time, providing a more accurate representation of that gene repression. On the other hand, with TnSeq, the transposon will randomly integrate into one genomic region and is not sequence specific so if there are duplicate copies of the same gene, even if its essential, it may read as nonessential due to the complementation from the other copies. Additionally, the repression level can be somewhat regulated by modifying the gRNA target or sequence homology where transposons will insert and interrupt a gene without any level of control on the effect of such insertion. Finally, CRISPRi-Seq can have wider applications by targeting specific locations of genes by simply altering the gRNA pool where transposons integrate randomly with no specificity. This method has since been applied to create a library of gRNAs to target the core essential genome of 18 *E. coli* strains to compare gene essentiality within

various genetic backgrounds (Rousset et al., 2021). Similar methods were also applied in *Streptococcus pneumoniae*, *Vibrio natriegens*, and *Synechocystis* sp. PCC 6803 (Lee et al., 2019; Yao et al., 2020; Liu et al., 2021a).

4.1.3 Essentiality determination by computational approaches

Computational programs and models have been developed to complement and fill in some gaps of time consuming and expensive experimental approaches of determining the essential genes. Using computational biology to assist with essential gene predictions is not something new. Comparative genomics, discussed above, has been used since 1996 and machine learning models to predict protein dispensability were developed in 2005 (Mushegian and Koonin, 1996; Chen and Xu, 2005). With technology advancements, these models and programs have recently become more sophisticated to address some challenges with computational approaches. One of the biggest challenges is that a minimal set of genes with a minimized metabolic network may not be viable in a cellular environment and/or not kinetically feasible. Thus, this is considered at all stages when developing a program to simulate essential gene predictions. The main features considered in computational models include expression level, sequence composition, evolutionary conservation, domain information, and network topology (Dong et al., 2018). These features have been reviewed in depth by Dong et al., 2018 but are briefly described below. First, the expression level of essential genes is typically higher than nonessential genes (Lloyd et al., 2015). This feature can be used to complement other essentiality prediction features, but not used as a method of assessment on its own due to the variability of this occurrence. Next, sequence composition can be analysed as there are differences between the amino acid sequences of essential and nonessential genes. This is a newer method of essentiality determination but is applicable in all sequenced genomes without the requirement of protein functional data (Zhang and Zhang, 1994; Sarangi et al., 2013; Ning et al., 2014). Evolutionary conservation is commonly used as a marker for essentiality and is based on the notion that genes that persist against negative selection through evolution are likely to be involved in essential functions within the cell (Jordan et al., 2002; Bergmiller et al., 2012). Like evolutionary conservation, domain information of proteins is the specific conservation of regions of proteins that perform similar functions. Also, essential genes encoding proteins will have various domains found infrequently in other proteins, making the substitution of these functions difficult (Chen et al., 2015; Peng et al., 2015). The first application of domain information is from 2011 and since then has been used in various studies, but not as commonly used as evolutionary conservation (Deng et al., 2011; Cheng et al., 2013; Lin et al., 2019). Finally, network topology refers to the way proteins interact within the cell with essential genes more frequently found in the center of complex protein-protein interaction networks (Yu et al., 2004; Wang et al., 2012). This also applies to metabolic, gene co-expression, and transcriptional networks (Acencio and Lemke, 2009; da Silva et al., 2008). For the most part, essentiality studies combine many of these features such as looking at the topologies of networks combined with gene expression profiles (Plaimas et al., 2008, 2010; Deng et al., 2011).

Based on both experimental data and computational models, online databases of the essential genes for various species have been created. First, the Database of Essential Genes (DEG) has lists of essential genes in 66 bacterial strains based on experimental data and this list is constantly being updated (Zhang et al., 2004; Luo et al., 2021). This database is the most widely used thanks to its practical tools such as homology searches using the embedded BLAST tool and since it contains essential genetic elements outside of just protein coding genes like non-coding RNAs, regulatory sequences, and essential promoters (Peng et al., 2017). An associated resource, database of predicted essential genes (pDEG), was constructed in 2011 in the same format as DEG to contain predicted essential genes for various *Mycoplasma* genomes but has not been updated since then to include other species or strains (Lin and Zhang, 2011). Two databases of essential genes for thousands of species have been more recently released, ePath and NetGenes (Kong et al., 2019; Senthamizhan et al., 2021). ePath, released in 2019, contains essential gene predictions for more than 4,000 bacteria (Kong et al., 2019). Its predictions are based on KEGG Ortholog (KO) annotations for biological functions, thus limiting this program to genes with KO numbers available. For example, a *Streptococcus* strain has 819 of its 2,270 genes annotated with KO so only those genes can be assessed for essentiality. This can be bypassed by using a built-in KEGG program, BlastKOALA to computationally assign KO numbers, but the obvious issue is that this is an estimation and not from experimental data (Kanehisa et al., 2016). The output for ePath includes essentiality scores based on experimental data (E-score) and the gene's involvement in critical cellular processes (P-score) with prediction accuracies of 75–91% (Kong et al., 2019). NetGenes, released in 2021, has predictions for over 2,700 bacterial species and includes information on the essential genes like essentiality scores and feature vectors (Senthamizhan et al., 2021). The essentiality predictions rely on protein-protein interaction network-based features from the STRING database described in depth in an initial publication from 2018 (Azhagesan et al., 2018). This database provides a strong model for essential and nonessential gene classification but fails to capture fitness genes that are not essential but required for robust growth. So, while this data is important, this model is not yet constructed to be used independently to design a strong growing minimized genome cell. Other databases including OGEE (Online GEne Essentiality database), EGGS (Essential Genes on Genome Scale), CEG (database of essential gene clusters) are briefly described in Table 2.

TABLE 2 Essential gene databases and computational programs to predict essential genes and design genomic deletions.

Program	Function	URL	References
DEG 15	Database of essential genomic regions based on experimental data with embedded BLAST tools for homology searches	www.essentialgene.org/	Luo et al. (2021)
pDEG	Database of predicted essential genes within *Mycoplasma* strains	http://tubic.org/pdeg/	Lin and Zhang, (2011)
NetGenes	Database of predicted essential genes for more than 2,700 organisms based on protein-protein interaction network features	https://rbc-dsai-iitm.github.io/NetGenes/	Senthamizhan et al. (2021)
ePath	Database of predicted essential genes for more than 4,000 organisms based on previous experimental data and functional KEGG orthologs	https://www.pubapps.vcu.edu/epath/	Kong et al. (2019)
OGEE	Database of essential genes from experimental data for 91 bacterial strains with additional gene features like expression profiles, conservation, evolutionary origins etc.	http://ogeedb.embl.de	Chen et al. (2017)
EGGS	Database of essential genes from experimental data for 11 different species with visualization and analysis on a subsystem diagram	https://pubseed.theseed.org/FIG/eggs.cgi	Gerdes et al. (2006) Overbeek et al. (2005)
CEG	Database of essential genes based on data from DEG that are clustered based on function	http://cefg.uestc.cn/ceg	Liu et al. (2020a) Ye et al. (2013)
CEG_Match	Within the CEG database, bases the prediction of essential genes on function	http://cefg.uestc.cn/ceg	Liu et al. (2020a) Ye et al. (2013)
ZCURVE 3.0	Predicts genes from an unannotated genome and will also predict essential genes from that	http://guolab.whu.edu.cn/zcurve/	Hua et al. (2015)
Geptop 2.0	Predicts gene essentiality based on sequence conservation and orthology with comparing a fully sequenced organism to 37 other species	http://guolab.whu.edu.cn/geptop/	Wen et al. (2019)
EGP	Machine learning-based method applying sequence composition features to predict essential genes with the input of nucleotide sequence	http://cefg.uestc.edu.cn:9999/egp	Ning et al. (2014)
Deeply Essential	Deep neural network to identify essential genes in bacteria based on sequence features only	https://github.com/ucrbioinfo/DeeplyEssential	Hasan and Lonardi, (2020)
DELEAT	Prediction of essential genes based on 6 features not dependent on experimental or functional data with design of large genomic deletions to minimize the organism's genome	https://github.com/jime-sg/deleat	Solana et al. (2021)
MinGenome	Minimal genome design with large deletion predictions using whole cell models using biological knowledge on gene locations and essentiality	https://github.com/maranasgroup/MinimalGenome	Wang and Maranas, (2018)
GAMA	Comprehensive and computationally intensive simulation of gene essentiality with gene deletions to construct a true minimal cell that cannot be run on a single computer	https://github.com/GriersonMarucciLab	Rees-Garbutt et al. (2020)
Minesweeper	Simultaneously assesses genomic regions that can be deleted, starting with larger deletion regions moving to individual genes	https://github.com/GriersonMarucciLab	Rees-Garbutt et al. (2020)

In addition to all the databases on essential genes, open-access programs have been created to run essential gene predictions, summarized in Table 2. Geptop predicts bacterial essential genes based on phylogeny, assessing evolutionary distance between species using composition vector method, and orthology, finding similar proteins across genomes using the reciprocal best hit method (Wei et al., 2013). This program, initially created in 2013, has recently been updated to include more essentiality data, increasing from 19 to 37 species, and to increase computation speed (Wen et al., 2019). Geptop 2.0 is simple to use with an interface to input DNA or protein sequences and receive the predicted essentiality with probabilities of genes or proteins but can only be used with fully sequenced organisms (Wen et al., 2019). CEG_Match, an extension on the CEG database, predicts essential genes based on function (Ye et al., 2013). More specifically, it matches the annotated gene names with the cluster gene names within the

CEG database, avoiding issues with BLAST searches by eliminating the misclassification of genes with different sequences but similar functions (Ye et al., 2013; Peng et al., 2017). So here, the obvious limitation is that this prediction method only works with genes with known functions and names.

Machine learning is also an increasingly popular route of determining gene essentiality and has been reviewed extensively (Liu et al., 2020b, 2021b; Aromolaran et al., 2021). Briefly, machine learning is the ability of a computer system to 'improve' and 'learn' using inputted data to make predictions despite not being programmed to do so accurately (Aromolaran et al., 2021). So, data from model organisms on essential and nonessential genes are used to train a classifier that is then applied to predict gene essentiality in the same or a different organism. Many machine learning models have been created to analyze protein and genomic features which have been now applied to essential gene determination (Peng et al., 2017).

These models are trained using sequence derived features and context-dependant features (Cheng et al., 2013; Ning et al., 2014). Sequence derived features include various factors like GC content, codon usage, protein length (more large and small proteins compared to medium sized proteins coded by essential genes), strand bias, and more (Lipman et al., 2002; Gong et al., 2008; Peng et al., 2017). Context-dependent features include those previously discussed like protein domain properties, protein-protein interaction networks, protein localization, and gene expression (Jansen et al., 2002; Seringhaus et al., 2006; Acencio and Lemke, 2009; Deng et al., 2011; Peng and Gao, 2014). Machine learning methods to predict essential genes are improved compared to homology mapping because they can use more features when constructing the prediction model (Deng et al., 2011; Lu et al., 2014). A machine learning-based method for essentiality predictions called Essential Gene Prediction (EGP) is freely accessible and only requires nucleotide sequence input (Ning et al., 2014). EGP uses amino acid, codon, and nucleotide usage as well as codon features independently to build the prediction model. This is done using training datasets from 16 genomes with known essential genes (Ning et al., 2014). It has been successfully used to identify essential genes in many organisms including *Salmonella typhimurium* and *E. coli* (Plaimas et al., 2010; Deng et al., 2012). But the selection of features and combinations may influence the performance of prediction and there is no clear method of selecting suitable features for differing organisms (Mobegi et al., 2017). Also, it has lower accuracy than some newer methods due to the limited reference species and parameters used for estimation (Peng et al., 2017). In general, a limitation of machine learning models is the inability to predict quasi-essential genes. Furthermore, there is a lack of complete and correct data from experimental and computational studies which impacts the accuracy of essential gene prediction in machine learning models.

Deep learning, a subset of machine learning, has networks that can use unlabeled or unstructured data to learn unsupervised. DeeplyEssential is a deep neural network that utilizes this learning model to predict essential genes by using only sequence information (Hasan and Lonardi, 2020). This model was able to achieve higher sensitivity and precision compared to clustered and down-sampled datasets used previously (Liu et al., 2017; Hasan and Lonardi, 2020). DeeplyEssential has countless applications since it only requires the genome sequence of the organism compared to other models that require topological or structural data which may not be available. Another deep learning model for essentiality predictions was developed taking a different approach using a framework that automatically learns biological features without the requirement of prior information (Zeng et al., 2021). This network uses information on gene expression, subcellular localization, and protein-protein interaction networks to learn topological features (Zeng et al., 2021). A major drawback to deep learning models is the high computational costs when training the network (Aromolaran et al., 2021). And when using them specifically for predicting essential genes, they require training with big data to outperform traditional machine learning algorithms and there is high complexity with tuning the parameters in deep learning models (Aromolaran et al., 2021). The biggest advantage of this approach is the ability to train algorithms with data from model organisms and use that to predict essential genes in poorly annotated organisms. But, as mentioned previously, it is unable to predict conditionally essential and quasi-essential genes. So in general, computational methods for essential gene prediction are great tools to supplement the experimental methods but require further improvements to be used as an independent tool.
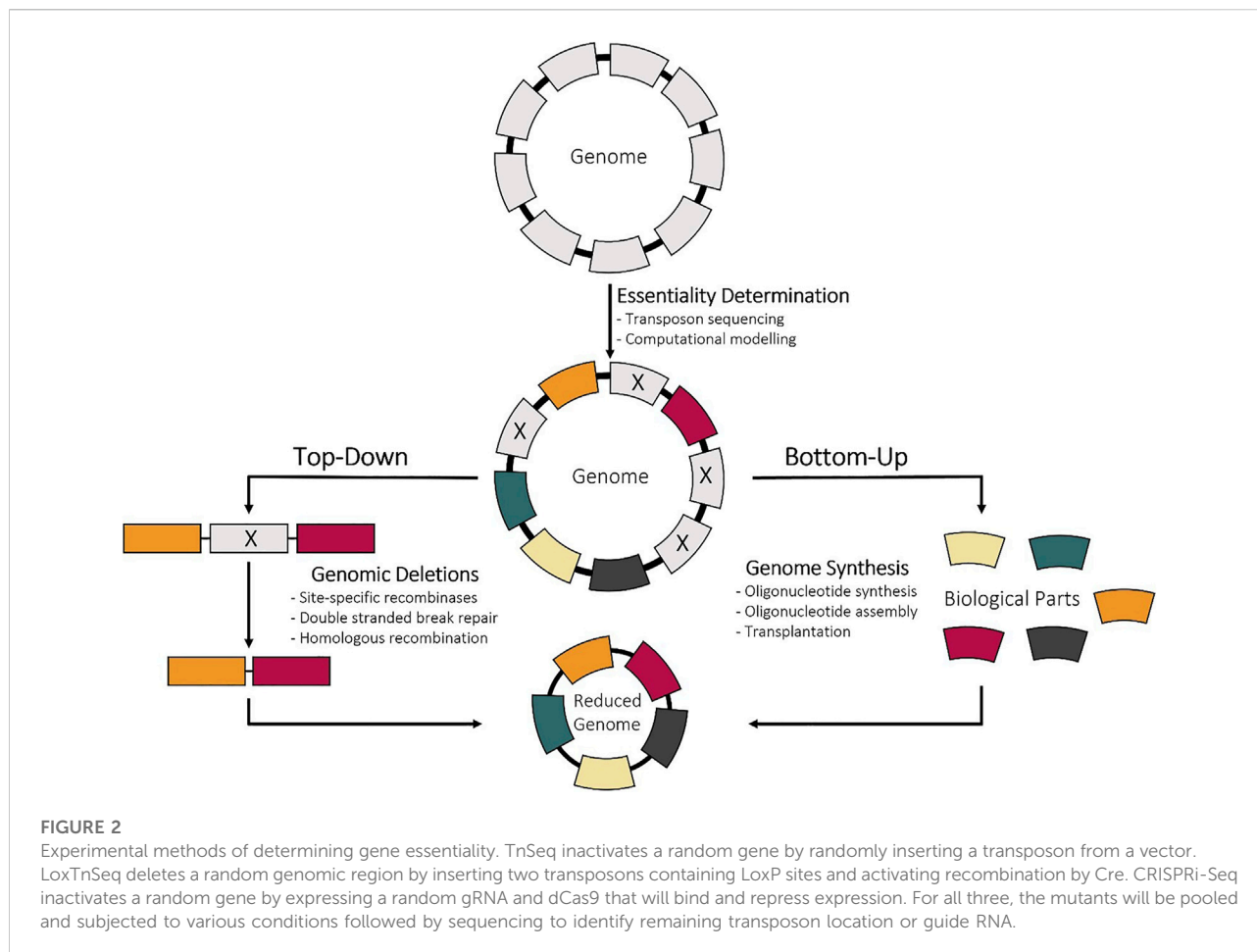
4.2 Strategies for constructing reduced-genome bacterial cells

Once the essential genes are identified by combining the computational and experimental methods outlined above, the next step is to reduce genome. This can be done either using bottom-up methods, chemical synthesis of the minimized genome, or top-down methods, deletion of nonessential genes from the genome (Figure 2). Bottom-up approaches have been enabled with advancements in sequencing, gene synthesis, and assembly technologies. This method also offers the unique opportunity to not only create a cell with a smaller genome, but to also restructure the genome for biotechnological applications. However, top-down construction approaches are more popular and have been utilized more since they do not require completed genetic information (Sung et al., 2016). Also, the wide variety of deletion methods, many of which tailored to specific species, enables the use of this approach in almost any bacterial strain. Since the focus of this review is the minimization of genomes, methods that are used to make large deletions or multiple sequential deletions will be discussed, and are summarized in Figure 3. Prior to making deletions, what deletions to make and the order to make them needs to be determined.

4.2.1 Computation design of genome reductions

As more is learned about bacterial genomes, the process of deciding which genes to remove and how to remove those genes becomes increasingly complex. Similar to using computational tools to predict gene essentiality, a few programs have been developed to assist in the deletion selection and genome design. The design of a tailored cell factory that produces key biomolecules or is a chassis for downstream applications is hindered by the segmented nature of our knowledge. Even with the knowledge and tools of how to build synthetic genomes bottom-up, very few have been constructed and

FIGURE 2
Experimental methods of determining gene essentiality. TnSeq inactivates a random gene by randomly inserting a transposon from a vector. LoxTnSeq deletes a random genomic region by inserting two transposons containing LoxP sites and activating recombination by Cre. CRISPRi–Seq inactivates a random gene by expressing a random gRNA and dCas9 that will bind and repress expression. For all three, the mutants will be pooled and subjected to various conditions followed by sequencing to identify remaining transposon location or guide RNA.
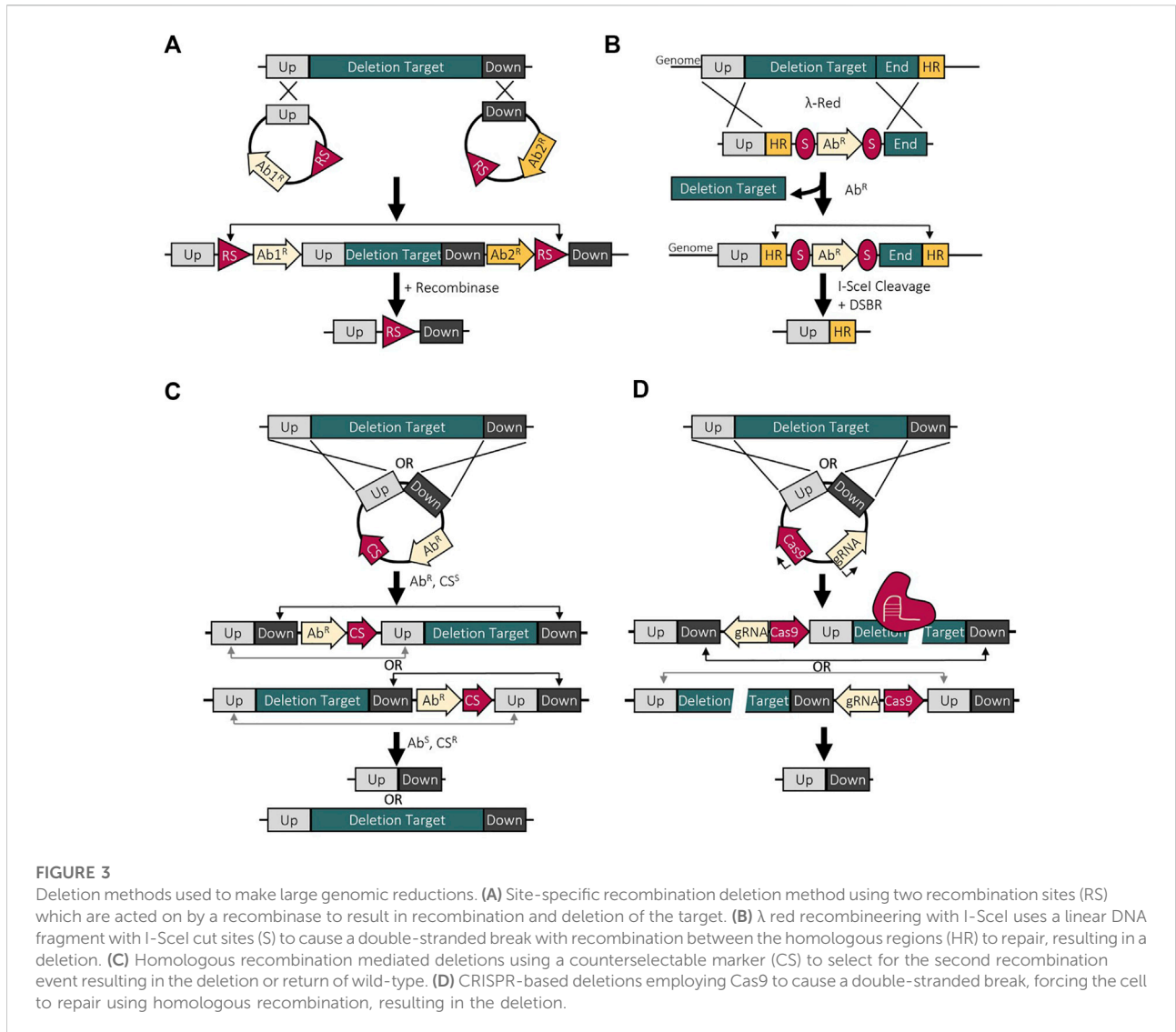
reported due to the difficulty of designing such genomes (Hutchison et al., 2016; Richardson et al., 2017; Fredens et al., 2019). This stems from little understanding of genome design principles due to the inordinate complexity of target organisms. There is also a lack of ability to analyze and evaluate genomic designs and an overwhelming number of possible genome configurations even for bacteria with small genomes (Matteau et al., 2020). Even taking the smallest organisms, like *M. genitalium* with a total of 525 genes, there are 2^{525} possible genome-scale designs, making it impossible to assess all these designs *in vivo*. Thus, computational whole-cell models (WCMs) have been developed to simulate the dynamics of a cell but are currently limited mainly to model organisms due to lack of genomic annotation (Karr et al., 2012; Chalkley et al., 2019; Münzner et al., 2019; Norsigian et al., 2020). These algorithms can model the effect of large genomic deletions on growth rate and metabolism prior to experimental testing and assist in designing genomes with a minimal set of genes in an optimal configuration.

Three algorithms currently exist to run these genome reduction simulations; MinGenome, Minesweeper, and GAMA

(Guess/Add/Mate Algorithm) (Table 2) (Wang and Maranas, 2018; Rees-Garbutt et al., 2020). MinGenome highlights long regions of nonessential genes by incorporating biological knowledge like gene location and essentiality with a genome metabolic model (Wang and Maranas, 2018). It also assesses large genomic regions that could be deleted even including regions that contain one or two essential genes that could be reintroduced after. This algorithm was applied to the *E. coli* MG1655 genome and showed similar results to experimental studies as well as alternative deletion combinations that have not been attempted in cells yet. Minesweeper takes a slightly different approach in that it assesses all genes that can be removed from the genome simultaneously, resulting in multiple different genome constructs since the order of deletions matters (Rees-Garbutt et al., 2020). It starts by deleting genes in groups then moves to deleting individual genes. This algorithm is based on gene knockout simulations to determine essential and nonessential genes. GAMA, developed at the same time as Minesweeper, first considers nonessential gene deletions in the guess/add stage before adding in essential genes in the final mate stage (Rees-Garbutt et al., 2020). In the guess phase, all the

FIGURE 3
Deletion methods used to make large genomic reductions. **(A)** Site-specific recombination deletion method using two recombination sites (RS) which are acted on by a recombinase to result in recombination and deletion of the target. **(B)** λ red recombineering with I-SceI uses a linear DNA fragment with I-SceI cut sites (S) to cause a double-stranded break with recombination between the homologous regions (HR) to repair, resulting in a deletion. **(C)** Homologous recombination mediated deletions using a counterselectable marker (CS) to select for the second recombination event resulting in the deletion or return of wild-type. **(D)** CRISPR-based deletions employing Cas9 to cause a double-stranded break, forcing the cell to repair using homologous recombination, resulting in the deletion.

nonessential genes from the input (which is a preprocessing stage to determine nonessential genes) are broken into four sets of genes which are used to make around 400 subsets where combinations of genes are deleted and determined if cell division can occur with those deletions. These viable sets are taken to the add phase, where deletion sets from previous groups are combined into a larger deletion set. About 3,000 of these combined deletion sets are tested and the ones that produce a viable cell are ranked, taking 50 of the smallest genomes to the mate phase. In this phase, two of the 50 minimized genomes are mated with random knockouts and knock-ins from a pool of the protein coding genes. 1,000 simulations per mating is performed and the updated strain is passed back into the pool for another round, which is continued for 100 generations. The simulations with GAMA are very computationally heavy, using between 400 and 3,000 CPUs, taking over 2 months to process on a standard supercomputer to generate minimal genome size reductions. Thus, a previously developed genome design suite was utilized to implement GAMA (Chalkley et al., 2019). Whole cell models were used to simulate 10 sets of minimal genes from literature of *M. genetalium in silico* and it was found that those cells could not divide and replicate based on this algorithm (Rees-Garbutt et al., 2021). Each of these gene sets had deleted essential genes, even those that were compiled from TnSeq knockout studies. After reintroducing up to 26 genes considered to be essential and 'low essential', these gene sets could divide again *in silico*. This highlights a disconnect between available data on minimized gene sets and strains developed with various *in vivo* studies. While the new strains developed from this modeling have not been tested experimentally, this shows the potential of such models to improve existing strains and create new reduced genome strains.

Finally, another program called DELEAT (DELetion design by Essentiality Analysis Tool) has been created which combines *in silico* gene essentiality predictions and automatic large-scale deletion design across all bacterial genomes (Solana et al., 2021). To estimate gene essentiality, genes are assigned an essentiality score from 0 to 1 based on 6 gene features that don't rely on functional annotation or experimental data. The deletion design is based on two parameters, the minimum desired deletion length (L) and the essentiality score (E). Following manual revision of the deletions, the program will provide a summary with various factors including deletion size, number of deletions, and deletion order. An additional feature will design primers to use to make the deletion constructs based on megapriming. This was used to highlight 35 deletions that could be made in *Bartonella quintana* to reduce the genome by 29% (Solana et al., 2021). While this program has yet to be fully tested, it provides promising insights into rapidly designing reduced cells. This is the first program that enables deletion design, from what to delete all the way to the deletion order and primer design.

4.2.2 Site-specific recombination deletion methods

Moving to making genomic deletions, site-specific recombinases are often used due to their high deletion efficiency and functionality across multiple strains. Two methods that are frequently used and function similarly are Flp/FRT and Cre/loxP which causing recombination between *FRT* or *loxP* sites by the Flp and Cre recombinases respectively. When these sites are oriented in the same direction, recombination will result in the excision of the region between the two sites where recombination between two sites facing opposite direction results in an inversion. So, two vectors containing *FRT* of *loxP* sites can be introduced into the genome via homologous recombination, one upstream and one downstream of the deletion target (Figure 3A). After introduction of the recombinase, the deletion target is excised from the genome leaving behind one *FRT* or *loxP* site. This method has been used to make very large deletions at high efficiency but with the left-over recombination site, sequential deletions cannot be made without removing the leftover site (Komatsu et al., 2010). Improvements to Cre/loxP have been made to bypass this issue by utilizing *loxP* mutations to render the left-over *loxP* site non-functional. This allows new functional *loxP* sites to be reintroduced into the strain for subsequent deletions. This has been applied in *Bacillus pumilus* by using mutated *lox71* and *lox66* sites which result in a double mutant *lox72* site after recombination which has very low affinity for Cre (Guan et al., 2017). This does allow for multiple deletions but there will still be many *lox72* sites throughout the genome which is not ideal when constructing a reduced-genome chassis strain. Another downfall with site-specific recombination deletion methods is the tedious deletion mutant selection process. Each vector inserted with the recombination sites would have a selectable antibiotic resistance gene marker to select for the double integration. The cells would then have to be screened for sensitivity to the selected antibiotics to distinguish between those with the deletion and those with an inversion. This can be avoided in some species since the constitutive activity of Cre on active *loxP* sites can be toxic, like in *M. pneumoniae*, but this is not true for all bacteria. There are also modifications to the Flp/FRT system to include a counter-selectable marker to enable selection of deletion mutants (Ishikawa and Hori, 2013).

4.2.3 Homologous recombination deletion methods

To combat some of the issues highlighted previously, alternative methods are used including markerless, homologous recombination-mediated, counter-selectable deletions. This involves the insertion of homologous regions up and downstream of the deletion target into a suicide vector that contains a selectable antibiotic resistance marker and a counter-selectable marker (Figure 3B). Common counter-selectable marker systems include sucrose (*sacB*), fusaric acid (*tetAR*), streptomycin (*rpsL*), and 5-fluorouracil (*upp*) (Reyrat et al., 1998; Fabret et al., 2002; Kang et al., 2002; Kristich et al., 2005; Goh et al., 2009; Keller et al., 2009). The first recombination event (insertion of the vector into the genome) can be selected for on the antibiotic plate and the second recombination event (excision from the genome) can be counter-selected on one of the above substances. If the vector is still present, the counter-selectable marker will result in cell death. This method is widely used as it is a 'scarless' deletion method and leaves no trace in the genome. Deficits of this method include the lower deletion efficiency compared to site-specific deletion methods. In the second recombinant deletion event, there is a 50% chance to excise only the plasmid from the genome, leaving the deletion target intact. In practice, this 50% efficiency of deletion is rarely obtained with a more realistic efficiency ranging from 10–40% (Graf and Altenbuchner, 2011). Also, this efficiency drops quickly as the size of deletion increases. For example, deletions of 26 and 64 kb achieved efficiencies of 40 and 20% respectively (Graf and Altenbuchner, 2011). It is possible to make very large deletions shown by the deletion of 1.4 Mbp in *Streptomyces*, but extensive screening was required to isolate cells with the correct deletion (Komatsu et al., 2010). This same deletion made with Cre-LoxP achieved 100% deletion efficiency (Komatsu et al., 2010). So, while this is a scarless deletion method with counterselection, screening is required to isolate wildtype and deletion mutants which is very time consuming when making multiple deletions.

To avoid the time-consuming screening, two routes have been taken, making concurrent deletions and combining into one strain after and the development of novel deletion methods. A deletion method combining λ red recombineering and I-SceI mediated double-stranded break repair was applied in *Pseudomonas* to make sequential

genomic deletions (Chen et al., 2016). λ red recombineering employs three enzymes to catalyze homologous recombination of double stranded linear DNA with chromosomal DNA, avoiding the need to do any assembly steps outside of the cell (Yu et al., 2000). So, the linear substrate DNA contains an antibiotic resistance gene flanked by two I-SceI recognition sites (S) with 500 base pair homologous regions up- and down-stream of the deletion target (Figure 3C). After electroporation, I-SceI recombination removes the antibiotic resistance cassette, and the cell repairs the double stranded break using RecA-mediated homologous recombination (Chen et al., 2016). Similar methods have been applied in other species such as *Corynebacterium glutamicum* and *E. coli* (Vernyik et al., 2020; Wu et al., 2020). In *E. coli*, an approach using a Tn5 transposon containing I-SceI made successive, scarless deletions in the genome in random locations (Vernyik et al., 2020). This resulted in strains with up to a 2.5% genomic reduction with improved growth characteristics including biomass yield. Interestingly, of the 60 genomes sequenced, deletions were observed in the same 12 regions (Vernyik et al., 2020). The second method used to avoid time consuming sequential deletions employs the use of phage transduction to combine independent genomic deletions into one strain (Pósfai et al., 2006; Umenhoffer et al., 2017; Saragliadis et al., 2018). This allows for multiple deletions to be made at the same time, lowering the overall time required. This is also useful for transferring deletions made in one strain to a different strain and has been applied extensively in *E. coli* (Umenhoffer et al., 2017). This is limited by the host range of the phage but with the existence of multiple transducing phages, there are many strains that are compatible with this method of deletion combination (Kang et al., 2002; Lee et al., 2004; Saragliadis et al., 2018).

4.2.4 CRISPR/Cas deletion methods

One of the most popular genome editing tools in members of the Eukarya domain are CRISPR-Cas systems. Despite being successful in some bacterial model strains, CRISPR technologies are not widely used in other bacteria in contrast to the increased use in Eukarya organisms (Dicarlo et al., 2013; Jiang et al., 2013; Mali et al., 2013). CRISPR-Cas systems are divided into two classes; class 1 consisting of multi subunit complexes and class 2 which are large, multi-domain proteins composed of one single unit (Makarova et al., 2020). Class 2 systems have been most used in bacteria, including Cas9, and Cas12, and typically involve the creation of double stranded breaks in DNA which signal repair through homologous recombination or non-homologous end joining. The Cas nuclease can be targeted to a specific location by gRNAs and will cleave complementary DNA that is flanked by a protospacer adjacent motif (PAM) (Gasiunas et al., 2012; Jinek et al., 2012). Often in bacteria,

DNA cleavage and the overexpression of these large nucleases are lethal, limiting their use for making genomic deletions (Vento et al., 2019; Arroyo-Olarte et al., 2021).

With regards to making large-scale genomic reductions, CRISPR-Cas systems have been applied in two ways, to counterselect successful deletion mutants, and to make deletions. The lethality exhibited by the DNA breakage by a Cas nuclease can be used for counterselection. Cleavage is prevented when the PAM site is removed by a genomic deletion and those without the deletion will have a double stranded break, resulting in cell death (Oh and van Pijkeren, 2014; Banno et al., 2018; Penewit et al., 2018; Wirth et al., 2020). This is a useful method with high selection efficiencies with small deletions, but this efficiency drops significantly with larger deletions (Aparicio et al., 2018).

Additionally, there have been CRISPR-Cas systems developed to make genomic deletions successfully and rapidly in various bacterial species (Huang et al., 2015; So et al., 2017; Li K. et al., 2018; Zhang et al., 2019). These methods typically result in a deletion following DNA cleavage by a class 2 Cas nuclease and homologous recombination repair (Figure 3D). This employs either a heterologous recombinase or relies on endogenous homologous recombination machinery. Using native recombination machinery does simplify the overall process as it only needs one or two vectors harboring the CIRSPR-Cas elements and editing template, but this machinery may be lacking in some bacteria. For example, gene clusters in *Streptomyces coelicolor* of between 21 and 82 kb were deleted by using Cas9 with efficiencies between 38 and 100% (Huang et al., 2015). In addition, these gene clusters were deleted simultaneously with efficiencies between 29 and 54%, reducing the overall time to achieve a deletion by 3 times compared to other methods like Cre-LoxP. This involves the conjugation of *E. coli* harboring a plasmid with constitutive expression of Cas9 and gRNA as well as the up and downstream regions homologous to the deletion target. Following cleavage and two crossover events, only cells with the homologous recombination repair can survive. The plasmid can then be cured for another deletion round. A similar method using Cas9 was applied in *Bacillus subtilis* to achieve deletion efficiencies around 100% for a single gene and around 80% for a 38-kb region (So et al., 2017). Here, double stranded breaks were caused at the ends of the deletion region with homologous regions mediating repair in a two-plasmid system.

All these CRISR-Cas9-based editing methods that utilize recombineering either use linear DNA or circular DNA as the editing template, which each have their advantages and disadvantages (Jiang et al., 2013; Huang et al., 2015; Feng et al., 2018). A circular DNA has higher editing and recombination efficiencies since it can be copied along with plasmid replication and is not attacked by DNA exonucleases, but it is possible that the entire plasmid will be integrated into the genome, leading to high false positive rates (Huang et al., 2020).

Linear DNA on the other hand does not have the issue with genomic integration so it has a higher positive rate, but it can be degraded by exonucleases. These challenges were confronted by bringing elements of both circular and linear DNA targets into one model. By adding the deletion template into a plasmid flanked by the target sequence, Cas9 cleavage can release this fragment from the plasmid. This protects it from degradation during the transformation process and achieves high positive rates of deletions by utilizing linear DNA editing templates. This method was used to delete 187 kb DNA regions from the *E. coli* genome with much higher editing efficiency and positive rates than other CRISPR-based λ-Red recombineering methods (Huang et al., 2020). This was taken a step further to make 12 sequential deletions totalling 370 Kb in *E. coli* (Huang et al., 2020).

Other barriers to using CRISPR-Cas9 deletion methods is the cytotoxicity observed in bacteria from over-expression of Cas9, even in its catalytically dead form (Li L. et al., 2018; Cho et al., 2018). Even in strains that can tolerate Cas9, editing efficiencies are lowered since there is a lower number of cells that survive (Xu et al., 2015; Li Q. et al., 2016; Song et al., 2017). This was improved by controlling the expression of Cas9 with inducible promoters, but leaky expression could still result in toxicity (Reisch and Prather, 2015; Wasels et al., 2017; Vento et al., 2019). Other alternatives include light-inducible systems, not yet attempted in bacteria, and the growth of cells at high temperatures to prevent Cas9 function, which is inactive above 42°C, but this requires bacteria that can tolerate this temperature (Polstein and Gersbach, 2015; Mougiakos et al., 2017; Nihongaki et al., 2018; Zhou et al., 2018). Using alternative Cas enzymes or mutated Cas9 variants is an option to circumvent this cytotoxicity such as the use of Cas9n that has a mutation that prevents double-stranded breaks, only allowing the enzyme to nick one strand of DNA (Jinek et al., 2012; Standage-Beier et al., 2015). Large deletions in a few different species have been achieved by using this method (Standage-Beier et al., 2015; Li K. et al., 2018). However, the Cas9n mutant has less efficient editing, especially with low expression of the enzyme (Song et al., 2017; Malzahn et al., 2019). So, this is a viable route to explore for a potentially more stable and universal CRISPR-based deletion method but needs to be enhanced first.

Another promising route is to explore the benefits and limitations of other Cas nucleases. A study has shown that the Cas12a nuclease could achieve efficient editing and transformation into *Corynebacterium glutamicum* while Cas9 and Cas9n could not (Jiang et al., 2017). Cas12a has been applied in *Clostridium difficile* to make a large 49 kb deletion and multiplex deletions, in *Streptomyces* to make single and double gene deletions with efficiencies between 75 and 95%, and in other species including *E. coli* and *Mycobacterium smegmatis* (Yan et al., 2017; Li L. et al.,

2018; Hong et al., 2018). While these are a few successful examples, Cas12a is still a class 2 nuclease, meaning it is one large multi-subunit protein that can cause cellular toxicity when overexpressed.

More recent efforts have been using class 1 nucleases which are composed of multiple subunits (Xu et al., 2020). One use of these includes 'built in' genome editing using endogenously encoded CRISPR-Cas systems. This involves the identification of the CRISPR-Cas system that exists within the strain of interest and assembling a targeting plasmid that contains a minimal CRISPR array to target a specific genomic region. The deletion is mediated by cleavage followed by homology-directed DNA repair (Xu et al., 2020). Single deletions have been made successfully in various strains with high efficiencies including *C. saccharoperbutylacetonicum* (100%), *C. pasteurianum* (100%), *C. tyrobutyricum* (100%), *C. difficile* (30–100%), *L. crispatus* (100%), *P. aeruginosa* (50%), and *Z. mobilis* (100%) (Pyne et al., 2016; Zhang et al., 2018; Atmadjaja et al., 2019; Hidalgo-Cantabrana et al., 2019; Maikova et al., 2019; Xu et al., 2019; Zheng et al., 2019). Multiplex gene deletions have also been performed to simultaneously delete 2 genes in *C. tyrobutyricum* with 100% efficiency and 3 genes from *Z. mobilis* with 18.75% efficiency (Zhang et al., 2018; Zheng et al., 2019). Another example of a class 1 nuclease is the use of Cas3 to make non-specific deletions endogenously ranging from 7 to 424 kb in *Pseudomonas aeruginosa* with efficiencies reaching 100% as well as heterologously in *E. coli* and *Pseudomonas syringae* without a repair template (Csörgő et al., 2020). Though the targeting of Cas3 was not specific, it has great potential to make deletions much larger than Cas9 with higher efficiencies to make significant genomic reductions. To modify the functionality of Cas3, the nuclease was mutated to remove its helicase activity, converting the enzyme into a nickase (nCas3) that can nick single stranded DNA (Hao et al., 2022). Thus, two crRNAs were required to simultaneously target two genomic loci to get double nicking. This method was applied in *Z. mobilis* to generate a single deletion of 9 kb with an efficiency of 93.75% and to simultaneously delete two regions with an efficiency of 75% (Hao et al., 2022). Some other improvements include the creation of class 1 CRISPR-Cas systems that can be used in heterologous hosts. A transferable system employing a typeI-F cascade was created and used to make deletions up to 21 kb in *Pseudomonas* spp. with improved efficiencies compared to Cas9 systems (Xu et al., 2021). This was also further modified to incorporate λ-red for use in strains with poor homologous recombination and cells with anti-CRISPRs (Xu et al., 2021). While all these class 1 CRISPR-Cas tools were not used to construct a reduced genome cell, they still show promise for developing a system to make genomic reductions in strains with and without native CRISPR systems on a large scale.

Most studies making genomic deletions using CRISPR-based methods only highlighted its use to make a few deletions in a single strain. There are few studies that highlight a strategy that

can be used across multiple strains and few that compare different methods of making the same deletions. One example is a strategy employing a RecT recombinase and Cas12a in various *Corynebacterium glutamicum* strains but was not very applicable to large deletions due to low efficiencies (Jiang et al., 2017). Another is the use of the typeI-F cascade in various *Pseudomonas* strains but has also only been tested on smaller, single deletions (Xu et al., 2021). Thus, with CRISPR-based methods, there is not one prevailing method, deletion strategies need to be tailored to the strain, genome size, deletion location, and number of deletions to be made.

Overall, CRISPR-based deletion methods show promising results for both improving counterselection methods and deletions made using homologous recombination but there is plenty of research still lacking on improving the efficiency of this system in many bacterial genetic backgrounds. Without a deletion method applicable across multiple species and for multiple deletion targets, this technology is very limited compared to other more universal methods. Many class 2 nucleases including Cas9 and Cas 12a have also only been applied in various model organisms with high transformation efficiencies, so are not directly applicable to many species yet. Class 1 nucleases including Cas3 show significant promise in improving cell toxicity and the size of genomic deletions but have not been used extensively in reducing the size of genomes or in the making of multiple genomic deletions. Taken together, reducing cytotoxicity, increasing deletion efficiencies, continued exploration of alternative Cas nucleases, and testing larger genomic reductions need to be addressed to make these technologies more applicable to large scale deletions in a wide range of bacteria.

5 Applications of reduced-genome bacterial strains

5.1 Genomic deletions for investigative research

The earliest reports of genomic deletions were applied to investigate the effect of various genes within the cell by deleting them from the genome. This often involved deleting only one gene at a time from the genome. For example, numerous studies investigated the effects of the deletion of *recA* in various strains. In *Mycobacterium bovis* BCG, a *recA* deletion revealed that the cell was more susceptible to DNA-damaging agents (Sander et al., 2001). In *E. coli*, a *recA* deletion increased transformation efficiency and improved *in vivo* phage packaging (Kurnit, 1989). The deletion of genes can also make some unexpected discoveries. In *Streptococcus pneumoniae*, the deletion of a zinc uptake lipoprotein, *adcAII*, revealed an unpredicted relationship with capsule thickness (Durmort et al., 2020). The capsule is the main virulence factor, but mechanisms involved in the regulation

of its thickness is not well understood. Partial deletion of *adcAII* resulted in increased capsule thickness, making it hypervirulent and more resistant to neutrophil attack in mouse models (Durmort et al., 2020). Another discovery of novel functions to a previously annotated gene region was made in *Sinorhizobium meliloti* when removing almost half of its genes (diCenzo et al., 2016; Milunovic et al., 2014). The genome of *S. meliloti* is divided into three components, a circular chromosome of 3.7 Mb, a chromid of 1.7 Mb, and a megaplasmid of 1.3 Mb. In this reduction, the chromid and megaplasmid were removed, moving the essential genes the chromid carries to the chromosome which uncovered four unexpected essential toxin/antitoxin genes, showing the first report that two of them even function as a toxin/antitoxin system (Milunovic et al., 2014).

Furthermore, deletion studies are also performed to investigate the role of bacterial genes in human pathogenesis. Again in *S. pneuomiae*, the deletion of a few genes residing in the Entner-Doudoroff pathway resulted in increased virulence and mortality in chinchilla models of otitis media (ear infection) (Hu et al., 2019). Overall, this study identified the role of various metabolic genes in virulence and pathogenicity including glucose dehydrogenase, the Entner-Doudoroff pathway, and ketogluconate degradation genes (Hu et al., 2019). These few examples of how genomic deletions can be utilized to investigate the role of various genes in a vast array of processes highlights their importance in both genomic and functional discoveries.

5.2 Genomic reductions to improve biomolecule production

Since bacteria are widely used in industrial processes to produce a variety of biomolecules, strains that are robust, able to survive in strenuous conditions, and have the natural ability to produce specific products are shortlisted for use at the industrial scale. These include strains of *Streptomyces sp.*, *Pseudomonas sp.*, *B. subtilis*, and *E. coli* (Kolisnychenko et al., 2002; Commichau et al., 2013; Belda et al., 2016; Calero and Nikel, 2019). By further modifying these organisms to synthesize products more efficiently, their applicability expands, and production costs are lowered. Techniques to improve biomolecule production include increasing precursor supply, enhancing flux through specific biosynthetic pathways, and reducing formation of by-products from alternative pathways (Gao et al., 2010). Genomic reductions can be targeted to these three areas such as making large genomic deletions to remove nonessential energy and resource consuming functions or through targeted deletions of competing pathways. Some biomolecules of interest and strains with genomic reductions to improve their production are highlighted below and summarized in Table 3.

One area of interest is increasing the production of polyhydroxyalkanoates (PHAs). PHAs are biopolymers

TABLE 3 Previous genome reduced bacterial strains.

Strain	Deletion	Deletion size	Deletion method	Characteristics (relative to parental strain)	References
Bacillus amyloliquefaciens					
GR167	Genomic islands, extracellular polysaccharide biosynthesis genes, prophages	167 Kb (4.18%)	HR with _upp_ CS	Faster growth, higher transformation efficiency, increased heterologous gene expression	Zhang et al. (2020)
Bacillus subtilis					
Δ6	Prophages, _pks_ operon	323 Kb (7.7%)	HR with no CS	Comparable growth rate	Westers et al. (2003)
MG1M	Prophages, antibiotic production genes	991 Kb (24%)	HR	Reduced growth rate, unstable recombinant protein production	Ara et al. (2007)
MGB874	74 regions including prophages, secondary metabolite producing genes, etc	873.5 Kb (20.7%)	HR with _upp_ CS	Increase in cellulase (1.7-fold) and protease (2.5-fold) production	Morimoto et al. (2008)
BSK814G2	Prophages, antibiotic production operons and other nonessential regions	814 Kb (20%)	HR with _upp_ CS	Decreased growth characteristics but 4.4-fold higher guanosine production	Li et al. (2016b)
BSK756T3	Prophages, antibiotic production operons and other nonessential regions	756 Kb (18.6%)	HR with _upp_ CS	Decreased growth characteristics but 5.2-fold higher thymidine production	Li et al. (2016b)
PG10	Many genes including those for sporulation, motility, secondary metabolism, prophages, secreted proteases, etc.	1.46 Mb (36%)	HR with _manP_ CS	Decreased growth rate, lower resource utilization for information processing, improved production of 'difficult proteins' that cannot be produced in other _Bacillus subtilis_ strains	Reuß et al. (2017) Suárez et al. (2019)
Corynebacterium glutamicum					
MB001	3 Prophages	204.7 Kb (6%)	HR with _SacB_ CS	Improved growth under stress conditions, increased transformation efficiency, 30% increase in heterologous protein production	Baumgart et al. (2013)
C1*	Non-essential genes including prophages, unknown genesetc.	440 Kb (13.4%)	HR with _sacB_ CS	Robust against stresses, improved growth stability, similar growth rates	Baumgart et al. (2018)
CR101	All prophages and IS elements	249.4 Kb (7.6%)	HR with _sacB_ CS	Similar growth rate and transformation efficiency to MB001	Linder et al. (2021)
Escherichia coli					
MDS42	Insertion sequences	663.3 Kb (14.3%)	λ-Red HR with I-SceI + P1 transduction	Improved electroporation efficiency, similar growth rates	Pósfai et al. (2006)
Δ16	Various deletions across the _E. coli_ genome	1.38 Mb (29.7%)	λ-Red HR with _sacB_ and _rpsL_ CS + P1 transduction	Slower growth and abnormal cell morphology	Hashimoto et al. (2005)
MGF-01	Various nonessential gene regions	1.03 Mb (22%)	λ-Red HR + P1 transduction	1.5-fold higher cell density and 2x threonine production from an introduced gene cassette	Mizoguchi et al. (2007)
MS56	IS Elements, K-islands, flagella genes, LPS synthesis genes	1.1 Mb (23%)	λ-Red HR with I-SceI + _sacB_ CS	1.6-fold faster growth and improved genomic stability	Park et al. (2014)
Lactococcus lactis					
9K-4	Prophages, integrases, and transposases	71 Kb (2.83%)	Cre-LoxP	Faster growth rate, increased biomass yield, improved heterologous gene expression 3-4-fold	Zhu et al. (2017)
N8-8	Prophages and genomic islands	176 Kb (6.86%)	Cre-LoxP	Shortened generation time by 17%, similar nisin yield	Qiao et al. (2022)
Magnetospirillum gryphiswaldense					
ΔTZ-17	Prophages, transposases, nitrogen fixation genes, _pks_ operon	227 Kb (5.5%)	HR with _galK_ CS	Comparable growth rate and magnetosome biosynthesis with improved genomic stability	Zwiener et al. (2021)
Mycoplasma mycoides					
JCVI-syn3A	All nonessential or quasi essential genes	669 Kb (55.2%)	Chemical synthesis	Improved growth rates compared to JCVI-syn3.0	Breuer et al. (2019)
Pseudomonas alloputida					
KTU-13	Genomic islands	254.5 Kb (4.1%)	HR with _sacB_ CS	45-fold increase in transformation efficiency, 9.4-fold increase in heterologous protein expression, 39% increase in PHA production	Liang et al. (2020)

TABLE 3 (*Continued*) Previous genome reduced bacterial strains.

Strain	Deletion	Deletion size	Deletion method	Characteristics (relative to parental strain)	References
EM383	Flagellar biosynthesis genes, prophages, transposases, recombinases	265.8 Kb (4.3%)	HR with ISce-I	Improved growth rate, heterologous protein expression, plasmid stability, stress resistance, and more	Lieder et al. (2015) Martínez-García et al. (2014)
Pseudomonas mendocina					
NKU421	Genomic island, prophages, hypothetical protein clusters	418 Kb (7.7%)	HR with *upp* CS	Increased ATP/ADP ratio by 11x, Improved mcl-PHA and alginate oligosaccharide production by 114.8 and 27.8% respectively	Fan et al. (2020)
Pseudomonas taiwanesis					
VBL120	Megaplasmid, prophages, flagellar biosynthesis genes, and biofilm genes	640 Kb (10.7%)	I-SceI HR with CS	Increased growth rates and biomass yield, improved production of chemicals including phenol	Wynands et al. (2019)
Sinorhizobium meliloti					
Rm1021	2 megaplasmids containing nonessential genes including toxin/ antitoxin systems	3.1 Mb (46%)	Flp/FRT	Identification of 4 toxin/antitoxin pairs that are essential	Milunovic et al. (2014)
Streptomyces albus					
J1074 (Del14)	15 biosynthetic secondary metabolite gene clusters	500 Kb (7.3%)	HR of mutant BAC library and *phi*C31 integrase	Comparable growth rates and improved heterologous gene expression of 7 products by 2–2.4 fold	Myronovski et al. (2018)
Streptomyces avermitilis					
SUKA17	Biosynthetic genes, prophages, transposases	1.67 Mb (18.5%)	Cre/LoxP	Increased streptomycin (4-fold) and cephamycin C (2-fold) production	Komatsu et al. (2010)
Streptomyces chattanoogensis					
L321	Biosynthetic clusters including the natamycin biosynthetic cluster	700 Kb (7.7%)	Cre/LoxP	Increased ATP and NADPH availability, higher transformation efficiency, improved heterologous gene expression, and increased genetic stability	Bu et al. (2019)

naturally synthesized by a variety of bacteria as a stored carbon source in limited nutrient environments with high availability of carbon sources (Khanna and Srivastava, 2005). They are a viable option for the replacement of environmentally harmful, petroleum-based plastics but adoption is hindered by the high production costs (Khanna and Srivastava, 2005; Mozejko-Ciesielska et al., 2019). Thus, genome reduction has been investigated as a strategy to improve PHA production. One strain of high interest due to its metabolic versatility and robustness, *Pseudomonas alloputida* KT2440, has been a target of such reductions (Mozejko-Ciesielska et al., 2019). Intuitively, the deletion of PHA depolymerase, encoded by *phaZ*, results in an increase of PHA yield by 38% compared to its parental strain (Poblete-Castro et al., 2014). Expanding on this, another study constructed a strain with the *phaZ* deletion and a deletion in two enzymes in a competing metabolic pathway, *fadB* and *fadA*, which resulted in a PHA yield increase of 13% when using an alternative carbon source, lignocellulosic biomass (Salvachúa et al., 2020). Next, the deletion of genomic islands accounting for approximately 4% of the genome also improved cell dry weight by 26.4% and PHA yield by 39.32% (Liang et al., 2020).

Genomic reductions in other PHA producing strains have also shown strong improvements in production. By reducing the *Pseudomonas mendocina* genome by 7.7% with 14 sequential deletions, creating NKU421, the ATP/ADP ratio was improved by a factor of 11 and PHA production was improved by 114.8% compared to the parental strain (Fan et al., 2020). Another key organism that is of high interest to engineer improved biomolecule production is *Bacillus subtilis*. *B. subtilis* is widely used to produce enzymes and other chemicals at the industrial scale (Li Y. et al., 2016). Deletions totalling 814 kb in *B. subtilis* 168 found that transformation efficiency and growth rates were slightly decreased but when this strain was engineered to produce guanosine and thymidine by overexpressing some genes, accumulation increased 4.4- and 5.2-fold respectively (Li Y. et al., 2016). So, although some growth characteristics were negatively impacted, biomolecule production was improved significantly. This was observed in another minimized *B. subtilis* strain PG10 with a 36% genome reduction (Reuß et al., 2017). Despite a decrease in growth rate, this strain was able to produce 'difficult proteins' that could not be produced by other *B. subtilis* strains like staphylococcal antigens (Reuß et al.,

2017; Suárez et al., 2019). Genome minimization has also been performed in non-model organisms to produce various biomolecules. For example, *Magnetospirillum gryphiswaldense* is a key organism in magnetosome biosynthesis and production (Zwiener et al., 2021). Magnetosomes have biological functions as magnetic sensors and have strong biomedical and biotechnological applications as magnetic nanoparticles in magnetic imaging, carriers for magnetic drug targeting and hyperthermia applications. Since large scale production is limiting its applicability, the genome of *M. gryphiswaldense* has been reduced by 5.5% to simplify its metabolism and work towards a chassis for magnetosome production (Zwiener et al., 2021). This strain displayed similar growth rates and magnetosome biosynthesis as the parental strain but had increased genetic stability and resilience (Zwiener et al., 2021). Overall, by reducing the genome size, producing biomolecules can be improved by freeing up more resources. This is often as a result of improved growth characteristics but there are cases where biomolecule production is increased, and other factors are worsened.

5.3 Genomic reductions to improve growth characteristics

Many of the deletions made above were made to improve growth characteristics and in turn increased biomolecule production. These factors include genomic stability, plasmid maintenance, growth rate, heterologous gene expression, and more. Research is also contributing to the creation of optimized chassis strains for any downstream application. These chassis are engineered for more rapid growth, improved stability, and improved heterologous gene expression. These strains can then be further engineered for uses such as producing a specific product otherwise unable to be produced by that organism or to utilize alternative carbon sources.

Looking at the prime example of *E. coli*, numerous studies have been conducted to improve its physiological characteristics. The genome of *E. coli* K12 MG1655, a common lab strain, was reduced by 14.3% by deleting insertion sequences resulting in improved electroporation efficiency and growth rates similar to the parental strain (Pósfai et al., 2006). In the same strain, a later study reduced the genome by 23% through the deletion of insertion sequences (ISs), K-islands, flagella genes, and some LPS synthesis genes (Park et al., 2014). These deletions resulted in 1.6-fold faster growth in minimal media and improved genome stability from the elimination of IS transposition (Park et al., 2014).

Moving away from *E. coli*, *Bacillus subtilis* 168 was reduced, creating MBG74, by deleting 874 kb of nonessential genomic regions, or 20% of the genome (Morimoto et al., 2008). This strain showed significant improvements in heterologous gene expression with increased yields of cellulase (1.7-fold) and

protease (2.5-fold) (Morimoto et al., 2008). In *Bacillus amyloquefaciens*, 4.18% of the genome was deleted, making strain GR167, improving transformation efficiency, growth rates, and heterologous gene expression (Zhang et al., 2020). These growth characteristics highlight this strain as a suitable chassis for further genetic modification and industrial applications. As a proof of concept, GR167 was engineered to produce surfactin with two deletions and introduction of a stronger promoter in front of the native surfactant producing gene. These modifications, making strain GR167IDS, resulted in a 10.4-fold increase in surfactin compared to GR167 (Zhang et al., 2020). This highlights both the ease of manipulation of the reduced genome strain as well as the benefit of combining these reductions with other modifications to improve the expression of either native or foreign genes.

Strains that show promising growth characteristic improvements can be utilized as chassis for further downstream engineering. By combining these more stable and faster growing strains with either the integration of various biomolecule producing cassettes or by introducing more targeted deletions, they can be used as molecular production factories. For example, in *B. amyloliquefaciens*, the deletion of three peptidoglycan hydrolase genes resulted in an increased production of alpha-amylase by 48% and increased cell viability because of decreased cell lysis (Zhang J. et al., 2021). A future step to assess further improvements on alpha-amylase production could be to combine these three deletions into the minimized *M. amyloliquefaciens* GR167 strain. Or similarly, in a reduced genome *E. coli* strain, the deletion of *fadR*, *fabR*, and *iclR* has previously been shown to increase L-threonine production and could be introduced into a genome reduced strain for further optimization (Yang et al., 2019).

5.4 The consequences of minimal genomes

When reducing a genome, there is always the possibility of removing genes that may not be essential to survival or biomolecule production but essential to robust cell growth. This is highlighted by the minimal genome strains which contain only the genes needed for cell survival. The initial focus when constructing minimal genomes was in *E. coli* with the first strains constructed in 2002. The genome of *E. coli* K-12 MG1655 was reduced by 6.8% with 313.1 Kb deleted containing 287 open reading frames and 179 unknown genes, yielding strain *E. coli* CDΔ3456 with no improvement to growth characteristics (Yu et al., 2002). The construction of this strain uncovered that the deletion of certain pairs of genes that are individually nonessential resulted in cell death, termed synthetic lethal pairs. Another *E. coli* minimal genome strain published in 2005 had a deletion of 29.7% of the MG1655 genome resulting in strain *E. coli* Δ16 (Hashimoto et al., 2005). The

deletion of these regions resulted in slower growth compared to the parental strain with almost two times the doubling time. The cells were also observed to have abnormal cell morphology and increased chromosome number per cell (Hashimoto et al., 2005). This was the first suggestion that minimal genomes may not be the end goal for having a chassis strain for downstream applications and that other genes unrelated to cell survival should be included in the final strain.

One of the most notable examples of a minimal genome is the *Mycoplasma mycoides* strain JCVI-syn3A (Breuer et al., 2019). After using whole genome chemical synthesis combined with assembly and cloning in yeast, *M. mycoides* synthetic genome was transplanted into *Mycoplasma capricolum*, making it the first cell that is controlled by a synthetic genome, JCVI-syn1.0 (Gibson et al., 2010; Sleator, 2010). The genome of this strain was further reduced to a total of 531 Kb and 473 genes, just under 50% the size of the parental strain (Hutchison et al., 2016). Named JCVI-syn3.0, this strain was an autonomously replicating cell with the smallest genome recorded (Hutchison et al., 2016). However, this strain displayed much slower growth rates, 2 to 3 times less than that of the parental *M. mycoides* strain and had some altered morphological traits. While this cell can survive and replicate, the elimination of all non-essential genes prevented it from being much use so the final revision to this strain occurred with the reintroduction of 20 genes to bring JCVI-syn3A to a total of 543 kb and 493 genes (Breuer et al., 2019). This included quasi-essential genes that are not required for survival but are required for robust growth with 149 of the protein coding genes having no known function (51). This emphasizes the fact that despite having the tools and technology to construct a synthetic minimal genome, there is still so much that is unknown about what set of functionalities are essential in enabling life and robust growth. A few studies following the strain construction looked at associating function to the unannotated proteins. By taking various approaches including sequence-based annotations, secondary structure matching, and multi-pipeline approaches, 66 proteins of unknown function in JCVI-syn3.0 were assigned function (Danchin and Fang, 2016; Hutchison et al., 2016; Yang and Tsui, 2018; Breuer et al., 2019). A more recent study annotated 50% of the proteins with unknown functions, 9 times more than existing UniProt annotations, by applying a novel pipeline that computationally predicted protein structure using map-based simulations followed by structural-based function annotation and protein-protein interaction predictions (Zhang C. et al., 2021).

There was initially a strong focus on creating minimal genome strains for use as chassis for downstream applications, but after seeing the effects on a cell, such as decreased growth rates and altered cell morphology, there was a shift towards only reducing a genome to the point where the organism can still have robust growth. Regardless, these minimal genomes are of immense value for research purposes and discovering the core set of genes required for bacterial life. The construction of minimal genomes has also elucidated many aspects of the genome such as the function and importance of specific genes or genomic interactions such as synthetic lethal pairs.

6 Discussion

Overall, genomic reductions are a promising avenue for uncovering gene functions, improving the production of valuable biomolecules and for creating chassis strains that can serve as model organisms on which to build applications. A combination of experimental and computational methods is likely to be most powerful for determining gene essentiality. For making defined deletions, many methods exist, and many are tailored and designed for specific species and strains With CRISPR becoming a popular tool in Eukaryal organisms, it is interesting to see it lagging in terms of applicability in bacteria, likely because powerful methods for genome manipulation have existed in bacteria for many years. Despite this, more recent research has been designing and implementing CRISPR-based methods for making deletions and with continued progress towards using alternative Cas enzymes, minimizing the toxicity of the class 2 nucleases, and improving deletion efficiencies, these efforts could result in improved approaches for large-scale deletion.

Since the ideation of minimal genomes, the goal with genomic reductions has shifted from making the smallest genome possible in a surviving cell to reducing the genome to make a cell that functions well and is not crippled. This is especially highlighted by the many studies that showed reduced growth characteristics in cells with extreme reductions. So, while minimal cells are valuable from a perspective of gaining knowledge on cellular functions and the essentials for survival, a reduced genome cell that has improved growth characteristics can be used for many applications and is the goal. One of the biggest driving forces behind the construction of genome-reduced hosts is the increasing demand for improved production of economically important bio-metabolites from stable, robust, and reliable strains. With an increasing number of reduced genome strains and more studies focused on improving methods of identifying essential genes and making deletions both experimentally and computationally, the area of large-scale genomic reductions is overcoming many challenges. Within the next few years, we can expect to see these reduced-genome alternatives serve as chassis for a broad range of applications.

Author contributions

NL conceived, wrote and edited the manuscript. TCC reviewed and edited the manuscript.

Funding

Conflict of interest

TCC is a major shareholder of Metagenom Bio Life Science Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acencio, M. L., and Lemke, N. (2009). Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information. *BMC Bioinforma.* 10, 290. doi:10.1186/1471-2105-10-290

Albalat, R., and Canestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* 12, 379–391. doi:10.1038/nrg.2016.39

Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C. M., Podowski, R. M., et al. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133–140. doi:10.1038/24094

Aparicio, T., de Lorenzo, V., and Martínez-García, E. (2018). CRISPR/Cas9-based counterselection boosts recombineering efficiency in Pseudomonas putida. *Biotechnol. J.* 13, e1700161. doi:10.1002/biot.201700161

Ara, K., Ozaki, K., Nakamura, K., Yamane, K., Sekiguchi, J., and Ogasawara, N. (2007). Bacillus minimum genome factory: Effective utilization of microbial genome information. *Biotechnol. Appl. Biochem.* 46, 169–178. doi:10.1042/ba20060111

Aromolaran, O., Aromolaran, D., Isewon, I., and Oyelade, J. (2021). Machine learning approach to gene essentiality prediction: A review. *Brief. Bioinform.* 22, bbab128. doi:10.1093/bib/bbab128

Arroyo-Olarte, R. D., Bravo Rodríguez, R., and Morales-Ríos, E. (2021). Genome editing in bacteria: CRISPR-cas and beyond. *Microorganisms* 9, 844. doi:10.3390/microorganisms9040844

Atmadjaja, A. N., Holby, V., Harding, A. J., Krabben, P., Smith, H. K., and Jenkinson, E. R. (2019). CRISPR-Cas, a highly effective tool for genome editing in Clostridium saccharoperbutylacetonicum N1-4(HMT). *FEMS Microbiol. Lett.* 366, fnz059. doi:10.1093/femsle/fnz059

Azhagesan, K., Ravindran, B., and Raman, K. (2018). Network-based features enable prediction of essential genes across diverse organisms. *PLoS ONE* 13, e0208722. doi:10.1371/journal.pone.0208722

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The keio collection. *Mol. Syst. Biol.* 2, 2006.0008. doi:10.1038/msb4100050

Baby, V., Lachance, J.-C., Gagnon, J., Lucier, J.-F., Matteau, D., Knight, T., et al. (2018). Inferring the minimal genome of Mesoplasma florum by comparative genomics and transposon mutagenesis. *MSystems* 3, e00198–17. doi:10.1128/mSystems.00198-17

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* 3, 423–429. doi:10.1038/s41564-017-0102-6

Baumgart, M., Unthan, S., Kloß, R., Radek, A., Polen, T., Tenhaef, N., et al. (2018). Corynebacterium glutamicum chassis C1*: Building and testing a novel platform host for synthetic biology and industrial biotechnology. *ACS Synth. Biol.* 7, 132–144. doi:10.1021/acssynbio.7b00261

Baumgart, M., Unthan, S., Rückert, C., Sivalingam, J., Grünberger, A., Kalinowski, J., et al. (2013). Construction of a prophage-free variant of Corynebacterium glutamicum ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl. Environ. Microbiol.* 79, 6006–6015. doi:10.1128/AEM.01634-13

Belda, E., van Heck, R. G. A., José Lopez-Sanchez, M., Cruveiller, S., Barbe, V., Fraser, C., et al. (2016). The revisited genome of Pseudomonas putida KT2440 enlightens its value as a robust metabolic chassis. *Environ. Microbiol.* 18, 3403–3424. doi:10.1111/1462-2920.13230

Bergmiller, T., Ackermann, M., and Silander, O. K. (2012). Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLoS Genet.* 8, e1002803. doi:10.1371/journal.pgen.1002803

Bobay, L. M., and Ochman, H. (2017). The evolution of bacterial genome architecture. *Front. Genet.* 8, 72–76. doi:10.3389/fgene.2017.00072

Breuer, M., Earnest, T. M., Merryman, C., Wise, K. S., Sun, L., Lynott, M. R., et al. (2019). Essential metabolism for a minimal cell. *Elife* 8, e36842–75. doi:10.7554/ELIFE.36842

Bu, Q. T., Yu, P., Wang, J., Li, Z. Y., Chen, X. A., Mao, X. M., et al. (2019). Rational construction of genome-reduced and high-efficient industrial Streptomyces chassis based on multiple comparative genomic approaches. *Microb. Cell Fact.* 18, 16–18. doi:10.1186/s12934-019-1055-7

Bumann, D. (2008). Has nature already identified all useful antibacterial targets? *Curr. Opin. Microbiol.* 11, 387–392. doi:10.1016/j.mib.2008.08.002

Calero, P., and Nikel, P. I. (2019). Chasing bacterial chassis for metabolic engineering: A perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* 12, 98–124. doi:10.1111/1751-7915.13292

Chalkley, O., Purcell, O., Grierson, C., and Marucci, L. (2019). The genome design suite: Enabling massive in-silico experiments to design genomes. *bioRxiv* [Preprint]. doi:10.10.1101/681270

Chen, L., Cheng, Y., Li, M., and Wang, J. (2015). Proteins involved in more domain types tend to be more essential. *Int. J. Bioinform. Res. Appl.* 11, 91–110. doi:10.1504/IJBRA.2015.068086

Chen, W. H., Lu, G., Chen, X., Zhao, X. M., and Bork, P. (2017). OGEE v2: An update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res.* 45, D940–D944. doi:10.1093/nar/gkw1013

Chen, Y., and Xu, D. (2005). Understanding protein dispensability through machine-learning analysis of high-throughput data. *Bioinformatics* 21, 575–581. doi:10.1093/bioinformatics/bti058

Chen, Z., Ling, W., and Shang, G. (2016). Recombineering and I-SceI-mediated Pseudomonas putida KT2440 scarless gene deletion. *FEMS Microbiol. Lett.* 363, fnw231–7. doi:10.1093/femsle/fnw231

Cheng, J., Wu, W., Zhang, Y., Li, X., Jiang, X., Wei, G., et al. (2013). A new computational strategy for predicting essential genes. *BMC Genomics* 14, 910. doi:10.1186/1471-2164-14-910

Cho, S., Choe, D., Lee, E., Kim, S. C., Palsson, B., and Cho, B. K. (2018). High-level dCas9 expression induces abnormal cell morphology in *Escherichia coli*. *ACS Synth. Biol.* 7, 1085–1094. doi:10.1021/acssynbio.7b00462

Choe, D., Cho, S., Kim, S. C., and Cho, B. K. (2016). Minimal genome: Worthwhile or worthless efforts toward being smaller? *Biotechnol. J.* 11, 199–211. doi:10.1002/biot.201400838

Choi, Y. J., and Lee, S. Y. (2013). Microbial production of short-chain alkanes. *Nature* 502, 571–574. doi:10.1038/nature12536

Commichau, F., Pietack, N., and Stulke, J. (2013). Essential genes in Bacillus subtilis: A re-evaluation after ten years. *Mol. Biosyst.* 9, 1068–1075. doi:10.1039/C3MB25595F

Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R., and Pósfai, G. (2012). Low-mutation-rate, reduced-genome *Escherichia coli*: An improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* 11, 11–13. doi:10.1186/1475-2859-11-11

Csörgő, B., León, L. M., Chau-Ly, I. J., Vasquez-Rifo, A., Berry, J. D., Mahendra, C., et al. (2020). A compact cascade–Cas3 system for targeted genome engineering. *Nat. Methods* 17, 1183–1190. doi:10.1038/s41592-020-00980-w

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.* 9, 1912. doi:10.1038/s41467-018-04209-5

da Silva, J. P. M., Acencio, M. L., Mombach, J. C. M., Vieira, R., da Silva, J. C., Lemke, N., et al. (2008). *In silico* network topology-based prediction of gene essentiality. *Phys. A Stat. Mech. Its Appl.* 387 (4), 1049–1055. doi:10.1016/j.physa.2007.10.044

Danchin, A., and Fang, G. (2016). Unknown unknowns: Essential genes in quest for function. *Microb. Biotechnol.* 9, 530–540. doi:10.1111/1751-7915.12384

De Eugenio, L. I., Escapa, I. F., Morales, V., Dinjaski, N., Galán, B., García, J. L., et al. (2010). The turnover of medium-chain-length polyhydroxyalkanoates in Pseudomonas putida KT2442 and the fundamental role of PhaZ depolymerase for the metabolic balance. *Environ. Microbiol.* 12, 207–221. doi:10.1111/j.1462-2920.2009.02061.x

de Maat, V., Arredondo-Alonso, S., Willems, R. J. L., and van Schaik, W. (2020). Conditionally essential genes for survival during starvation in Enterococcus faecium E745. *BMC Genomics* 21, 568. doi:10.1186/s12864-020-06984-2

Dejesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., et al. (2017). *Comprehensive essentiality analysis of the* Mycobacterium tuberculosis *genome via saturating transposon mutagenesis*, 8. doi:10.1128/mBio.02133-16MBio

Deng, J., Deng, L., Su, S., Zhang, M., Lin, X., Wei, L., et al. (2011). Investigating the predictability of essential genes across distantly related organisms using an integrative approach. *Nucleic Acids Res.* 39, 795–807. doi:10.1093/nar/gkq784

Deng, J., Tan, L., Lin, X., Lu, Y., and Lu, L. J. (2012). Exploring the optimal strategy to predict essential genes in microbes. *Biomolecules* 2, 1–22. doi:10.3390/biom2010001

Dicarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., and Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336–4343. doi:10.1093/nar/gkt135

diCenzo, G. C., Zamani, M., Milunovic, B., and Finan, T. M. (2016). Genomic resources for identification of the minimal N2-fixing symbiotic genome. *Environ. Microbiol.* 18, 2534–2547. doi:10.1111/1462-2920.13221

Dong, C., Jin, Y. T., Hua, H. L., Wen, Q. F., Luo, S., Zheng, W. X., et al. (2018). Comprehensive review of the identification of essential genes using computational methods: Focusing on feature implementation and assessment. *Brief. Bioinform.* 21, 171–181. doi:10.1093/bib/bby116

Durmort, C., Ercoli, G., Ramos-Sevillano, E., Chimalapati, S., Haigh, R., De Ste Croix, M., et al. (2020). Deletion of the zinc transporter lipoprotein AdcAII causes hyperencapsulation of Streptococcus pneumoniae associated with distinct alleles of the type I restriction modification system. *MBio* 11, e00445–20. doi:10.1128/mBio.00445-20

Fabret, C., Ehrlich, S. D., and Noirot, P. (2002). A new mutation delivery system for genome-scale approaches in Bacillus subtilis. *Mol. Microbiol.* 46, 25–36. doi:10.1046/j.1365-2958.2002.03140.x

Fan, X., Zhang, Y., Zhao, F., Liu, Y., Zhao, Y., Wang, S., et al. (2020). Genome reduction enhances production of polyhydroxyalkanoate and alginate oligosaccharide in Pseudomonas mendocina. *Int. J. Biol. Macromol.* 163, 2023–2031. doi:10.1016/j.ijbiomac.2020.09.067

Fehér, T., Papp, B., Pál, C., and Pósfai, G. (2007). Systematic genome reductions: Theoretical and experimental approaches. *Chem. Rev.* 107, 3498–3513. doi:10.1021/cr0683111

Feng, X., Zhao, D., Zhang, X., Ding, X., and Bi, C. (2018). CRISPR/Cas9 assisted multiplex genome editing technique in *Escherichia coli*. *Biotechnol. J.* 13, e1700604. doi:10.1002/biot.201700604

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., et al. (1995). The minimal gene complement of Mycoplasma genitalium. *Science* 270 (5235), 397–403. doi:10.1126/science.270.5235.397

Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., et al. (2019). Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569, 514–518. doi:10.1038/s41586-019-1192-5

Gao, H., Zhuo, Y., Ashforth, E., and Zhang, L. (2010). Engineering of a genome-reduced host: Practical application of synthetic biology in the overproduction of desired secondary metabolites. *Protein Cell* 1, 621–626. doi:10.1007/s13238-010-0073-3

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2579–E2586. doi:10.1073/pnas.1208507109

Gerdes, S., Edwards, R., Kubal, M., Fonstein, M., Stevens, R., and Osterman, A. (2006). Essential genes on metabolic maps. *Curr. Opin. Biotechnol.* 17, 448–456. doi:10.1016/j.copbio.2006.08.006

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi:10.1126/science.1190719

Goh, Y. J., Andrea Azcárate-Peril, M., O'Flaherty, S., Durmaz, E., Valence, F., Jardin, J., et al. (2009). Development and application of a upp-based counterselective gene replacement system for the study of the S-layer protein SlpX of Lactobacillus acidophilus NCFM. *Appl. Environ. Microbiol.* 75, 3093–3105. doi:10.1128/AEM.02502-08

Gong, X., Fan, S., Bilderbeck, A., Li, M., Pang, H., and Tao, S. (2008). Comparative analysis of essential genes and nonessential genes in *Escherichia coli* K12. *Mol. Genet. Genomics.* 279, 87–94. doi:10.1007/s00438-007-0298-x

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., et al. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6, 279–289. doi:10.1016/j.chom.2009.08.003

Graf, N., and Altenbuchner, J. (2011). Development of a method for markerless gene deletion in Pseudomonas putida. *Appl. Environ. Microbiol.* 77, 5549–5552. doi:10.1128/AEM.05055-11

Guan, Z. B., Wang, K. Q., Shui, Y., and Liao, X. R. (2017). Establishment of a markerless multiple-gene deletion method based on Cre/loxP mutant system for Bacillus pumilus. *J. Basic Microbiol.* 57, 1065–1068. doi:10.1002/jobm.201700370

Hao, Y., Wang, Q., Li, J., Yang, S., Zheng, Y., and Peng, W. (2022). Double nicking by RNA-directed Cascade-nCas3 for high-efficiency large-scale genome engineering. *Open Biol.* 1, 210241. doi:10.1098/rsob.210241

Hasan, M. A., and Lonardi, S. (2020). DeeplyEssential: A deep neural network for predicting essential genes in microbes. *BMC Bioinforma.* 21, 367. doi:10.1186/s12859-020-03688-y

Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., et al. (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* 55, 137–149. doi:10.1111/j.1365-2958.2004.04386.x

Hidalgo-Cantabrana, C., Goh, Y. J., Pan, M., Sanozky-Dawes, R., and Barrangou, R. (2019). Genome editing using the endogenous type I CRISPR-Cas system in Lactobacillus crispatus. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15774–15783. doi:10.1073/pnas.1905421116

Higgins, S., Sanchez-Contreras, M., Gualdi, S., Pinto-Carbó, M., Carlier, A., and Eberl, L. (2017). The essential genome of Burkholderia cenocepacia H111. *J. Bacteriol.* 199, e00260-17. doi:10.1128/JB.00260-17

Hong, W., Zhang, J., Cui, G., Wang, L., and Wang, Y. (2018). Multiplexed CRISPR-Cpf1-mediated genome editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. *ACS Synth. Biol.* 7, 1588–1600. doi:10.1021/acssynbio.8b00087

Hosseini, S. R., and Wagner, A. (2018). Genomic organization underlying deletional robustness in bacterial metabolic systems. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7075–7080. doi:10.1073/pnas.1717243115

Hu, F. Z., Król, J. E., Tsai, C. H. S., Eutsey, R. A., Hiller, L. N., Sen, B., et al. (2019). Deletion of genes involved in the ketogluconate metabolism, Entner-Doudoroff pathway, and glucose dehydrogenase increase local and invasive virulence phenotypes in Streptococcus pneumoniae. *PLoS ONE* 14, e0209688–25. doi:10.1371/journal.pone.0209688

Hua, Z. G., Lin, Y., Yuan, Y. Z., Yang, D. C., Wei, W., and Guo, F. B. (2015). Zcurve 3.0: Identify prokaryotic genes with higher accuracy as well as automatically and accurately select essential genes. *Nucleic Acids Res.* 43, W85–W90. doi:10.1093/nar/gkv491

Huang, C., Guo, L., Wang, J., Wang, N., and Huo, Y. X. (2020). Efficient long fragment editing technique enables large-scale and scarless bacterial genome engineering. *Appl. Microbiol. Biotechnol.* 104, 7943–7956. doi:10.1007/s00253-020-10819-1

Huang, H., Zheng, G., Jiang, W., Hu, H., and Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim. Biophys. Sin.* 47, 231–243. doi:10.1093/abbs/gmv007

Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351, aad6253. doi:10.1126/science.aad6253

Ishikawa, M., and Hori, K. (2013). A new simple method for introducing an unmarked mutation into a large gene of non-competent Gram-negative bacteria by FLP/FRT recombination. *BMC Microbiol.* 13, 86. doi:10.1186/1471-2180-13-86

Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Res.* 12, 37–46. doi:10.1101/gr.205602

Jewett, M. C., and Forster, A. C. (2010). Update on designing and building minimal cells. *Curr. Opin. Biotechnol.* 21, 697–703. doi:10.1016/j.copbio.2010.06.008

Jiang, G., Johnston, B., Townrow, D. E., Radecka, I., Koller, M., Chaber, P., et al. (2018). Biomass extraction using non-chlorinated solvents for biocompatibility improvement of polyhydroxyalkanoates. *Polym. (Basel)* 10, E731. doi:10.3390/polym10070731

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41, e188. doi:10.1093/nar/gkt780

Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., et al. (2017). CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum. *Nat. Commun.* 8, 15179. doi:10.1038/ncomms15179

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Johnson, A. O., Forsyth, V., Smith, S. N., Learman, B. S., Brauer, A. L., White, A. N., et al. (2020). Transposon insertion site sequencing of providencia stuartii: Essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements. *MSphere* 5, e00412–20. doi:10.1128/mSphere.00412-20

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* 12, 962–968. doi:10.1101/gr.87702

Juhas, M., Eberl, L., and Church, G. M. (2012a). Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol.* 30, 601–607. doi:10.1016/j.tibtech.2012.08.002

Juhas, M., Eberl, L., and Glass, J. I. (2011). Essence of life: Essential genes of minimal genomes. *Trends Cell Biol.* 21, 562–568. doi:10.1016/j.tcb.2011.07.005

Juhas, M., Reuß, D. R., Zhu, B., and Commichau, F. M. (2014). Bacillus subtilis and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiol. (United Kingdom)* 160, 2341–2351. doi:10.1099/mic.0.079376-0

Juhas, M., Stark, M., von Mering, C., Lumjiaktase, P., Crook, D. W., Valvano, M. A., et al. (2012b). High confidence prediction of essential genes in Burkholderia Cenocepacia. *PLoS ONE* 7, e40064. doi:10.1371/journal.pone.0040064

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi:10.1016/j.jmb.2015.11.006

Kang, H. Y., Dozois, C. M., Tinge, S. A., Lee, T. H., and Curtiss, R. (2002). Transduction-mediated transfer of unmarked deletion and point mutations through use of counterselectable suicide vectors. *J. Bacteriol.* 184, 307–312. doi:10.1128/JB.184.1.307-312.2002

Karr, J. R., Sanghvi, J. C., MacKlin, D. N., Gutschow, M. v., Jacobs, J. M., Bolival, B., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401. doi:10.1016/j.cell.2012.05.044

Keller, K. L., Bender, K. S., and Wall, J. D. (2009). Development of a markerless genetic exchange system for Desulfovibrio vulgaris Hildenborough and its use in generating a strain with increased transformation efficiency. *Appl. Environ. Microbiol.* 75, 7682–7691. doi:10.1128/AEM.01839-09

Khanna, S., and Srivastava, A. K. (2005). Recent advances in microbial polyhydroxyalkanoates. *Process Biochem.* 40, 607–619. doi:10.1016/j.procbio.2004.01.053

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., et al. (2003). Essential Bacillus subtilis genes. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4678–4683. doi:10.1073/pnas.0730515100

Kolisnychenko, V., Plunkett, G., Herring, C. D., Fehér, T., Pósfai, J., Blattner, F. R., et al. (2002). Engineering a reduced *Escherichia coli* genome. *Genome Res.* 12, 640–647. doi:10.1101/gr.217202

Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E., and Ikeda, H. (2010). Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2646–2651. doi:10.1073/pnas.0914833107

Kong, X., Zhu, B., Stone, V. N., Ge, X., El-Rami, F. E., Donghai, H., et al. (2019). ePath: an online database towards comprehensive essential gene annotation for prokaryotes. *Sci. Rep.* 9, 12949. doi:10.1038/s41598-019-49098-w

Koob, M. D., Shaw, A. J., and Cameron, D. C. (1994). Minimizing the genome of *Escherichia coli*. Motivation and strategy. *Ann. N. Y. Acad. Sci.* 745, 1–3. doi:10.1111/j.1749-6632.1994.tb44359.x

Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136. doi:10.1038/nrmicro751

Koskiniemi, S., Sun, S., Berg, O. G., and Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet.* 8, e1002787–7. doi:10.1371/journal.pgen.1002787

Kristich, C. J., Manias, D. A., and Dunny, G. M. (2005). Development of a method for markerless genetic exchange in *Enterococcus faecalis* and its use in construction of a srtA mutant. *Appl. Environ. Microbiol.* 71, 5837–5849. doi:10.1128/AEM.71.10.5837-5849.2005

Kunin, V., and Ouzounis, C. A. (2003). The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* 13, 1589–1594. doi:10.1101/gr.1092603

Kurasawa, H., Ohno, T., Arai, R., and Aizawa, Y. (2020). A guideline and challenges toward the minimization of bacterial and eukaryotic genomes. *Curr. Opin. Syst. Biol.* 24, 127–134. doi:10.1016/j.coisb.2020.10.012

Kurnit, D. M. (1989). *Escherichia coli* recA deletion strains that are highly competent for transformation and for *in vivo* phage packaging. *Gene* 82, 313–315. doi:10.1016/0378-1119(89)90056-5

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., et al. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res.* 19, 2308–2316. doi:10.1101/gr.097097.109

Lee, H. H., Ostrov, N., Wong, B. G., Gold, M. A., Khalil, A. S., and Church, G. M. (2019). Functional genomics of the rapidly replicating bacterium Vibrio natriegens by CRISPRi. *Nat. Microbiol.* 4, 1105–1113. doi:10.1038/s41564-019-0423-8

Lee, M. C., and Marx, C. J. (2012). Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet.* 8, e1002651–9. doi:10.1371/journal.pgen.1002651

Lee, S., Kriakov, J., Vilcheze, C., Dai, Z., Hatfull, G. F., and Jacobs, W. R. (2004). Bxz1, a new generalized transducing phage for mycobacteria. *FEMS Microbiol. Lett.* 241, 271–276. doi:10.1016/j.femsle.2004.10.032

Leprince, A., de Lorenzo, V., Völler, P., van Passel, M. W. J., and Martins dos Santos, V. A. P. (2012). Random and cyclical deletion of large DNA segments in the genome of Pseudomonas putida. *Environ. Microbiol.* 14, 1444–1453. doi:10.1111/j.1462-2920.2012.02730.x

Li, K., Cai, D., Wang, Z., He, Z., and Chen, S. (2018a). Development of an efficient genome editing tool in Bacillus licheniformis using CRISPR-cas9 nickase. *Appl. Environ. Microbiol.* 84, e02608–17. doi:10.1128/AEM.02608-17

Li, L., Wei, K., Zheng, G., Liu, X., Chen, S., Jiang, W., et al. (2018b). CRISPR-Cpf1-Assisted multiplex genome editing and transcriptional repression in Streptomyces. *Appl. Environ. Microbiol.* 84 (18), e00827–18. doi:10.1128/AEM.00827-18

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., et al. (2016a). CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnol. J.* 11, 961–972. doi:10.1002/biot.201600053

Li, Y., Zhu, X., Zhang, X., Fu, J., Wang, Z., Chen, T., et al. (2016b). Characterization of genome-reduced Bacillus subtilis strains and their application for the production of guanosine and thymidine. *Microb. Cell Fact.* 15, 94–15. doi:10.1186/s12934-016-0494-7

Liang, P., Zhang, Y., Xu, B., Zhao, Y., Liu, X., Gao, W., et al. (2020). Deletion of genomic islands in the Pseudomonas putida KT2440 genome can create an optimal chassis for synthetic biology applications. *Microb. Cell Fact.* 19, 70–12. doi:10.1186/s12934-020-01329-w

Lieder, S., Nikel, P. I., de Lorenzo, V., and Takors, R. (2015). Genome reduction boosts heterologous gene expression in Pseudomonas putida. *Microb. Cell Fact.* 14, 23–14. doi:10.1186/s12934-015-0207-7

Lin, Y., Zhang, F. Z., Xue, K., Gao, Y. Z., and Guo, F. B. (2019). Identifying bacterial essential genes based on a feature-integrated method. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16, 1274–1279. doi:10.1109/TCBB.2017.2669968

Lin, Y., and Zhang, R. R. (2011). Putative essential and core-essential genes in Mycoplasma genomes. *Sci. Rep.* 1, 53. doi:10.1038/srep00053

Linder, M., Haak, M., Botes, A., Kalinowski, Jorn., and Rückert, C. (2021). Construction of an IS-free Corynebacterium glutamicum ATCC 13 032 chassis strain and random mutagenesis using the endogenous ISCg1 transposase. *Front. Bioeng. Biotechnol.* 9, 751334. doi:10.3389/fbioe.2021.751334

Lipman, D. J., Souvorov, A., Koonin, E. v., Panchenko, A. R., and Tatusova, T. A. (2002). The relationship of protein conservation and sequence length. *BMC Evol. Biol.* 2, 20–10. doi:10.1186/1471-2148-2-20

Liu, S., Wang, S. X., Liu, W., Wang, C., Zhang, F. Z., Ye, Y. N., et al. (2020a). Ceg 2.0: An updated database of clusters of essential genes including eukaryotic organisms. *Database* 2020, baaa112. doi:10.1093/database/baaa112

Liu, X., He, T., Guo, Z., Ren, M., and Luo, Y. (2020b). Predicting essential genes of 41 prokaryotes by a semi-supervised method. *Anal. Biochem.* 609, 113919. doi:10.1016/j.ab.2020.113919

Liu, X., Kimmey, J. M., Matarazzo, L., de Bakker, V., van Maele, L., Sirard, J. C., et al. (2021a). Exploration of bacterial bottlenecks and Streptococcus pneumoniae pathogenesis by CRISPRi-Seq. *Cell Host Microbe* 29, 107–120. doi:10.1016/j.chom.2020.10.001

Liu, X., Luo, Y., He, T., Ren, M., and Xu, Y. (2021b). Predicting essential genes of 37 prokaryotes by combining information-theoretic features. *J. Microbiol. Methods* 188, 106297. doi:10.1016/j.mimet.2021.106297

Liu, X., Wang, B. J., Xu, L., Tang, H. L., and Xu, G. Q. (2017). Selection of key sequence-based features for prediction of essential genes in 31 diverse bacterial species. *PLoS ONE* 12, e0174638. doi:10.1371/journal.pone.0174638

Lloyd, J. P., Seddon, A. E., Moghe, G. D., Simenc, M. C., and Shiua, S. H. (2015). Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. *Plant Cell* 27, 2133–2147. doi:10.1105/tpc.15.00051

Lu, Y., Deng, J., Carson, M., Lu, H., and Lu, L. (2014). Computational methods for the prediction of microbial essential genes. *Curr. Bioinform.* 9 (2), 89–101. doi:10.2174/1574893608999140109113434

Luo, H., Lin, Y., Liu, T., Lai, F. L., Zhang, C. T., Gao, F., et al. (2021). DEG 15, an update of the database of essential genes that includes built-in analysis tools. *Nucleic Acids Res.* 49, D677–D686. doi:10.1093/nar/gkaa917

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60, 327–349. doi:10.1146/annurev.micro.60.080805.142300

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., and Soutourina, O. (2019). Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.* 85, e01416–e01419. doi:10.1128/AEM.01416-19

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., et al. (2020). Evolutionary classification of CRISPR-cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi:10.1038/s41579-019-0299-x

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi:10.1126/science.1232033

Malzahn, A. A., Tang, X., Lee, K., Ren, Q., Sretenovic, S., Zhang, Y., et al. (2019). Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis. *BMC Biol.* 17, 9. doi:10.1186/s12915-019-0629-5

Martínez-García, E., Nikel, P. I., Aparicio, T., and de Lorenzo, V. (2014). Pseudomonas 2.0: Genetic upgrading of P. Putida KT2440 as an enhanced host for heterologous gene expression. *Microb. Cell Fact.* 13, 159. doi:10.1186/s12934-014-0159-3

Matern, W. M., Jenquin, R. L., Bader, J. S., and Karakousis, P. C. (2020). Identifying the essential genes of *Mycobacterium avium* subsp. hominissuis with Tn-Seq using a rank-based filter procedure. *Sci. Rep.* 10, 1095. doi:10.1038/s41598-020-57845-7

Matteau, D., Lachance, J., Grenier, F., Gauthier, S., Daubenspeck, J. M., Dybvig, K., et al. (2020). Integrative characterization of the near-minimal bacterium Mesoplasma florum. *Mol. Syst. Biol.* 16, e9844. doi:10.15252/msb.20209844

McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi:10.1038/nrmicro2670

Milunovic, B., diCenzo, G. C., Morton, R. A., and Finan, T. M. (2014). Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of Sinorhizobium meliloti. *J. Bacteriol.* 196, 811–824. doi:10.1128/JB.01104-13

Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596. doi:10.1016/S0168-9525(01)02447-7

Mizoguchi, H., Mori, H., and Fujio, T. (2007). *Escherichia coli* minimum genome factory. *Biotechnol. Appl. Biochem.* 46, 157–167. doi:10.1042/ba20060107

Mobegi, F. M., Zomer, A., de Jonge, M. I., and van Hijum, S. A. F. T. (2017). Advances and perspectives in computational prediction of microbial gene essentiality. *Brief. Funct. Genomics* 16, 70–79. doi:10.1093/bfgp/elv063

Montero-Blay, A., Piñero-Lambea, C., Miravet-Verde, S., Lluch-Senar, M., and Serrano, L. (2020). Inferring active metabolic pathways from proteomics and essentiality data. *Cell Rep.* 31, 107722. doi:10.1016/j.celrep.2020.107722

Morimoto, T., Kadoya, R., Keiji, E., Tohata, M., Sawada, K., Liu, S., et al. (2008). Enhanced recombinant protein productivity by genome reduction in Bacillus subtilis. *DNA Res.* 15, 73–81. doi:10.1093/dnares/dsn002

Morowitz, H. J. (1984). The completeness of molecular biology. *Isr. J. Med. Sci.* 20, 750–753.

Mougiakos, I., Bosma, E. F., Weenink, K., Vossen, E., Goijvaerts, K., van der Oost, J., et al. (2017). Efficient genome editing of a facultative thermophile using mesophilic spCas9. *ACS Synth. Biol.* 6, 849–861. doi:10.1021/acssynbio.6b00339

Moya, A., Gil, R., Latorre, A., Peretó, J., Pilar Garcillán-Barcia, M., and de La Cruz, F. (2009). Toward minimal bacterial cells: Evolution vs. design. *FEMS Microbiol. Rev.* 33, 225–235. doi:10.1111/j.1574-6976.2008.00151.x

Mozejko-Ciesielska, J., Szacherska, K., and Marciniak, P. (2019). Pseudomonas species as producers of eco-friendly polyhydroxyalkanoates. *J. Polym. Environ.* 27, 1151–1166. doi:10.1007/s10924-019-01422-1

Münzner, U., Klipp, E., and Krantz, M. (2019). A comprehensive, mechanistically detailed, and executable model of the cell division cycle in *Saccharomyces cerevisiae*. *Nat. Commun.* 10, 1308. doi:10.1038/s41467-019-08903-w

Mushegian, A. R., and Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273. doi:10.1073/pnas.93.19.10268

Myronovski, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., and Luzhetskyy, A. (2018). Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* 49, 316–324. doi:10.1016/j.ymben.2018.09.004

Nihongaki, Y., Otabe, T., and Sato, M. (2018). Emerging approaches for spatiotemporal control of targeted genome with inducible CRISPR-Cas9. *Anal. Chem.* 90, 429–439. doi:10.1021/acs.analchem.7b04757

Nilsson, A. I., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C. D., and Andersson, D. I. (2005). Bacterial genome size reduction by experimental evolution. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12112–12116. doi:10.1073/pnas.0503654102

Ning, L. W., Lin, H., Ding, H., Huang, J., Rao, N., and Guo, F. B. (2014). Predicting bacterial essential genes using only sequence composition information. *Genet. Mol. Res.* 13, 4564–4572. doi:10.4238/2014.June.17.8

Norsigian, C. J., Pusarla, N., McConn, J. L., Yurkovich, J. T., Dräger, A., Palsson, B. O., et al. (2020). BiGG models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* 48, D402–D406. doi:10.1093/nar/gkz1054

Oh, J. H., and van Pijkeren, J. P. (2014). CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic Acids Res.* 42, e131. doi:10.1093/nar/gku623

Olano, C., Lombó, F., Méndez, C., and Salas, J. A. (2008). Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab. Eng.* 10, 281–292. doi:10.1016/j.ymben.2008.07.001

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi:10.1093/nar/gki866

Park, M. K., Lee, S. H., Yang, K. S., Jung, S.-C., Lee, J. H., and Kim, S. C. (2014). Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Appl. Microbiol. Biotechnol.* 98, 6701–6713. doi:10.1007/s00253-014-5739-y

Patrick, W. M., Quandt, E. M., Swartzlander, D. B., and Matsumura, I. (2007). Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* 24, 2716–2722. doi:10.1093/molbev/msm204

Penewit, K., Holmes, E. A., McLean, K., Ren, M., Waalkes, A., and Salipante, S. J. (2018). Efficient and scalable precision genome editing in *Staphylococcus aureus* through conditional recombineering and CRISPR/Cas9-mediated counterselection. *MBio* 9, e00067–18. doi:10.1128/mBio.00067-18

Peng, C., and Gao, F. (2014). Protein localization analysis of essential genes in prokaryotes. *Sci. Rep.* 4, 6001. doi:10.1038/srep06001

Peng, C., Lin, Y., Luo, H., and Gao, F. (2017). A comprehensive overview of online resources to identify and predict bacterial essential genes. *Front. Microbiol.* 8, 2331. doi:10.3389/fmicb.2017.02331

Peng, W., Wang, J., Cheng, Y., Lu, Y., Wu, F., and Pan, Y. (2015). UDoNC: An algorithm for identifying essential proteins based on protein domains and protein-protein interaction networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 12, 276–288. doi:10.1109/TCBB.2014.2338317

Plaimas, K., Eils, R., and König, R. (2010). Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Syst. Biol.* 4, 56. doi:10.1186/1752-0509-4-56

Plaimas, K., Mallm, J. P., Oswald, M., Svara, F., Sourjik, V., Eils, R., et al. (2008). Machine learning based analyses on metabolic networks supports high-throughput knockout screens. *BMC Syst. Biol.* 2, 67. doi:10.1186/1752-0509-2-67

Poblete-Castro, I., Binger, D., Oehlert, R., and Rohde, M. (2014). Comparison of mcl-poly(3-hydroxyalkanoates) synthesis by different Pseudomonas putida strains from crude glycerol: Citrate accumulates at high titer under PHA-producing conditions. *BMC Biotechnol.* 14, 962. doi:10.1186/s12896-014-0110-z

Polstein, L. R., and Gersbach, C. A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* 11, 198–200. doi:10.1038/nchembio.1753

Pósfai, G., Plunkett, G., Feher, T., Frisch, D., Keil, G., Umenhoffer, K., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312 (5776), 1044–1046. doi:10.1126/science.1126439

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Sci. Rep.* 6, 25666. doi:10.1038/srep25666

Qiao, W., Liu, F., Wan, X., Qiao, Y., Li, R., Wu, Z., et al. (2022). Genomic features and construction of streamlined genome chassis of nisin Z producer Lactococcus lactis N8. *Microorganisms* 10 (1), 47. doi:10.3390/microorganisms10010047

Quillaguamán, J., Hashim, S., Bento, F., Mattiasson, B., and Hatti-Kaul, R. (2005). Poly(beta-hydroxybutyrate) production by a moderate halophile, Halomonas boliviensis LC1 using starch hydrolysate as substrate. *J. Appl. Microbiol.* 99, 151–157. doi:10.1111/j.1365-2672.2005.02589.x

Rees-Garbutt, J., Chalkley, O., Landon, S., Purcell, O., Marucci, L., and Grierson, C. (2020). Designing minimal genomes using whole-cell models. *Nat. Commun.* 11, 836. doi:10.1038/s41467-020-14545-0

Rees-Garbutt, J., Rightmyer, J., Chalkley, O., Marucci, L., and Grierson, C. (2021). Testing theoretical minimal genomes using whole-cell models. *ACS Synth. Biol.* 10, 1598–1604. doi:10.1021/acssynbio.0c00515

Reisch, C. R., and Prather, K. L. J. (2015). The no-SCAR (scarless Cas9 assisted recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* 5, 15096. doi:10.1038/srep15096

Reuß, D. R., Altenbuchner, J., Mäder, U., Rath, H., Ischebeck, T., Sappa, P. K., et al. (2017). Large-scale reduction of the Bacillus subtilis genome: Consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* 27, 289–299. doi:10.1101/gr.215293.116

Reyrat, J., Pelicic, V., Gicquel, B., and Rappuoli, R. (1998). Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* 66, 4011–4017. doi:10.1128/IAI.66.9.4011-4017.1998

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., et al. (2017). Design of a synthetic yeast genome. *Science* 355, 1040–1044. doi:10.1126/science.aaf4557

Riley, M., and Serres, M. H. (2000). Interim report on genomics of *Escherichia coli*. *Annu. Rev. Microbiol.* 54, 341–411. doi:10.1146/annurev.micro.54.1.341

Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., et al. (2021). The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat. Microbiol.* 6, 301–312. doi:10.1038/s41564-020-00839-y

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLoS Genet.* 14, e1007749. doi:10.1371/journal.pgen.1007749

Salvachúa, D., Rydzak, T., Auwae, R., De Capite, A., Black, B. A., Bouvier, J. T., et al. (2020). Metabolic engineering of Pseudomonas putida for increased polyhydroxyalkanoate production from lignin. *Microb. Biotechnol.* 13, 290–298. doi:10.1111/1751-7915.13481

Samrot, A. V., Samanvitha, S. K., Shobana, N., Renitta, E. R., Kumar, P. S., Kumar, S. S., et al. (2021). The synthesis, characterization and applications of polyhydroxyalkanoates (PHAs) and PHA-based nanoparticles. *Polymers* 13 (9), 3302. doi:10.3390/polym13193302

Sander, P., Papavinasasundaram, K. G., Dick, T., Stavropoulos, E., Ellrott, K., Springer, B., et al. (2001). Mycobacterium bovis BCG recA deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model. *Infect. Immun.* 69, 3562–3568. doi:10.1128/IAI.69.6.3562-3568.2001

Saragliadis, A., Trunk, T., and Leo, J. C. (2018). Producing gene deletions in *Escherichia coli* by P1 transduction with excisable antibiotic resistance cassettes. *J. Vis. Exp.* 139. doi:10.3791/58267

Sarangi, A. N., Lohani, M., and Aggarwal, R. (2013). Prediction of essential proteins in prokaryotes by incorporating various physico-chemical features into the general form of Chou's pseudo amino acid composition. *Protein Pept. Lett.* 20, 781–795. doi:10.2174/0929866511320070008

Senthamizhan, V., Ravindran, B., and Raman, K. (2021). NetGenes: A database of essential genes predicted using features from interaction networks. *Front. Genet.* 12, 722198. doi:10.3389/fgene.2021.722198

Seringhaus, M., Paccanaro, A., Borneman, A., Snyder, M., and Gerstein, M. (2006). Predicting essential genes in fungal genomes. *Genome Res.* 16, 1126–1135. doi:10.1101/gr.5144106

Shaw, D., Miravet-Verde, S., Piñero-Lambea, C., Serrano, L., and Lluch-Senar, M. (2020). LoxTnSeq: Random transposon insertions combined with cre/lox recombination and counterselection to generate large random genome reductions. *Microb. Biotechnol.* 14 (6), 2403–2419. doi:10.1111/1751-7915.13714

Shi, T., Wang, Y., Wang, Z., Wang, G., Liu, D., Fu, J., et al. (2014). Deregulation of purine pathway in Bacillus subtilis and its use in riboflavin biosynthesis. *Microb. Cell Fact.* 13, 101. doi:10.1186/s12934-014-0101-8

Sleator, R. D. (2010). The story of Mycoplasma mycoides JCVI-syn1.0: The forty million dollar microbe. *Bioeng. Bugs* 1, 229–230. doi:10.4161/bbug.1.4.12465

So, Y., Park, S. Y., Park, E. H., Park, S. H., Kim, E. J., Pan, J. G., et al. (2017). A highly efficient CRISPR-Cas9-mediated large genomic deletion in bacillus subtilis. *Front. Microbiol.* 8, 1167. doi:10.3389/fmicb.2017.01167

Solana, J., Garrote-Sánchez, E., and Gil, R. (2021). Deleat: Gene essentiality prediction and deletion design for bacterial genome reduction. *BMC Bioinforma.* 22, 444. doi:10.1186/s12859-021-04348-5

Song, X., Huang, H., Xiong, Z., Ai, L., and Yang, S. (2017). CRISPR-Cas9D10A nickase-assisted genome editing in Lactobacillus casei. *Appl. Environ. Microbiol.* 83, e01259–17. doi:10.1128/AEM.01259-17

Standage-Beier, K., Zhang, Q., and Wang, X. (2015). Targeted large-scale deletion of bacterial genomes using CRISPR-nickases. *ACS Synth. Biol.* 4, 1217–1225. doi:10.1021/acssynbio.5b00132

Suárez, R. A., Stülke, J., and van Dijl, J. M. (2019). Less is more: Toward a genome-reduced Bacillus cell factory for "difficult proteins. *ACS Synth. Biol.* 8, 99–108. doi:10.1021/acssynbio.8b00342

Sung, B. H., Choe, D., Kim, S. C., and Cho, B. K. (2016). Construction of a minimal genome as a chassis for synthetic biology. *Essays Biochem.* 60, 337–346. doi:10.1042/EBC20160024

Tarnopol, R. L., Bowden, S., Hinkle, K., Balakrishnan, K., Nishii, A., Kaczmarek, C. J., et al. (2019). Lessons from a minimal genome: What are the essential organizing principles of a cell built from scratch? *Chembiochem.* 20, 2535–2545. doi:10.1002/cbic.201900249

Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi:10.1038/nrmicro1234

Tsuge, Y., Suzuki, N., Inui, M., and Yukawa, H. (2007). Random segment deletion based on IS31831 and cre/loxP excision system in Corynebacterium glutamicum. *Appl. Microbiol. Biotechnol.* 74, 1333–1341. doi:10.1007/s00253-006-0788-5

Umenhoffer, K., Draskovits, G., Nyerges, Á., Karcagi, I., Bogos, B., Tímár, E., et al. (2017). Genome-wide abolishment of mobile genetic elements using genome shuffling and CRISPR/Cas-assisted MAGE allows the efficient stabilization of a bacterial chassis. *ACS Synth. Biol.* 6, 1471–1483. doi:10.1021/acssynbio.6b00378

Vento, J. M., Crook, N., and Beisel, C. L. (2019). Barriers to genome editing with CRISPR in bacteria. *J. Ind. Microbiol. Biotechnol.* 46, 1327–1341. doi:10.1007/s10295-019-02195-1

Vernyik, V., Karcagi, I., Tímár, E., Nagy, I., Gyorkei, A., Papp, B., et al. (2020). Exploring the fitness benefits of genome reduction in *Escherichia coli* by a selection-driven approach. *Sci. Rep.* 10, 7345. doi:10.1038/s41598-020-64074-5

Wang, J., Li, M., Wang, H., and Pan, Y. (2012). Identification of essential proteins based on edge clustering coefficient. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 9, 1070–1080. doi:10.1109/TCBB.2011.147

Wang, L., and Maranas, C. D. (2018). MinGenome: An *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synth. Biol.* 7, 462–473. doi:10.1021/acssynbio.7b00296

Wang, T., Guan, C., Guo, J., Liu, B., Wu, Y., Xie, Z., et al. (2018). Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat. Commun.* 9, 2475. doi:10.1038/s41467-018-04899-x

Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A. M., and Lopes Ferreira, N. (2017). A two-plasmid inducible CRISPR/Cas9 genome editing tool for Clostridium acetobutylicum. *J. Microbiol. Methods* 140, 5–11. doi:10.1016/j.mimet.2017.06.010

Wei, W., Ning, L. W., Ye, Y. N., and Guo, F. B. (2013). Geptop: A gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PLoS ONE* 8, e72343. doi:10.1371/journal.pone.0072343

Weiser, R., Green, A. E., Bull, M. J., Cunningham-Oakes, E., Jolley, K. A., Maiden, M. C. J., et al. (2019). Not all pseudomonas aeruginosa are equal: Strains from industrial sources possess uniquely large multireplicon genomes. *Microb. Genom.* 5. doi:10.1099/mgen.0.000276

Wen, Q. F., Liu, S., Dong, C., Guo, H. X., Gao, Y. Z., and Guo, F. B. (2019). Geptop 2.0: An updated, more precise, and faster geptop server for identification of prokaryotic essential genes. *Front. Microbiol.* 10, 1236. doi:10.3389/fmicb.2019.01236

Wendisch, V. F., Bott, M., and Eikmanns, B. J. (2006). Metabolic engineering of *Escherichia coli* and Corynebacterium glutamicum for biotechnological production of organic acids and amino acids. *Curr. Opin. Microbiol.* 9, 268–274. doi:10.1016/j.mib.2006.03.001

Wernegreen, J. J. (2002). Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* 3, 850–861. doi:10.1038/nrg931

Westers, H., Dorenbos, R., Van Dijl, J. M., Kabel, J., Flanagan, T., Devine, K. M., et al. (2003). Genome engineering reveals large dispensable regions in Bacillus subtilis. *Mol. Biol. Evol.* 20, 2076–2090. doi:10.1093/molbev/msg219

Wirth, N. T., Kozaeva, E., and Nikel, P. I. (2020). Accelerated genome engineering of Pseudomonas putida by I-SceI—mediated recombination and CRISPR-Cas9 counterselection. *Microb. Biotechnol.* 13, 233–249. doi:10.1111/1751-7915.13396

Wolf, Y. I., and Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *Bioessays.* 35, 829–837. doi:10.1002/bies.201300037

Wong, Y. C., el Ghany, M. A., Naeem, R., Lee, K. W., Tan, Y. C., Pain, A., et al. (2016). Candidate essential genes in Burkholderia cenocepacia J2315 identified by genome-wide TraDIS. *Front. Microbiol.* 7, 1288. doi:10.3389/fmicb.2016.01288

Wu, M., Xu, Y., Yang, J., and Shang, G. (2020). Homing endonuclease I-SceI-mediated Corynebacterium glutamicum ATCC 13032 genome engineering. *Appl. Microbiol. Biotechnol.* 104, 3597–3609. doi:10.1007/s00253-020-10517-y

Wynands, B., Otto, M., Runge, N., Preckel, S., Polen, T., Blank, L. M., et al. (2019). Streamlined Pseudomonas taiwanensis VLB120 chassis strains with improved bioprocess features. *ACS Synth. Biol.* 8, 2036–2050. doi:10.1021/acssynbio.9b00108

Xu, F., Li, Y., Li, M., Ziang, H., and Yan, A. (2020). Harnessing the type I CRISPR-Cas systems for genome editing in prokaryotes. *Environ. Microbiol.* 23, 542–558. doi:10.1111/1462-2920.15116

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., et al. (2015). Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase. *Appl. Environ. Microbiol.* 81, 4423–4431. doi:10.1128/AEM.00873-15

Xu, Z., Li, M., Li, Y., Cao, H., Miao, L., Xu, Z., et al. (2019). Native CRISPR-Cas-mediated genome editing enables dissecting and sensitizing clinical multidrug-resistant *P. aeruginosa*. *Cell Rep.* 29, 1707–1717.e3. doi:10.1016/j.celrep.2019.10.006

Xu, Z., Li, Y., Cao, H., Si, M., Zhang, G., Woo, P. C. Y., et al. (2021). A transferrable and integrative type I-F Cascade for heterologous genome editing and transcription modulation. *Nucleic Acids Res.* 49 (16), e94. doi:10.1093/nar/gkab521

Yan, M.-Y., Yan, H.-Q., Ren, G.-X., Zhao, J.-P., Guo, X.-P., and Sun, Y.-C. (2017). CRISPR-Cas12a-assisted recombineering in bacteria. *Appl. Environ. Microbiol.* 83, e00947–17. doi:10.1128/AEM.00947-17

Yang, J., Fang, Y., Wang, J., Wang, C., Zhao, L., and Wang, X. (2019). Deletion of regulator-encoding genes fadR, fabR and iclR to increase L-threonine production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 103, 4549–4564. doi:10.1007/s00253-019-09818-8

Yang, Z., and Tsui, S. K. W. (2018). Functional annotation of proteins encoded by the minimal bacterial genome based on secondary structure element alignment. *J. Proteome Res.* 17, 2511–2520. doi:10.1021/acs.jproteome.8b00262

Yao, L., Shabestary, K., Björk, S. M., Asplund-Samuelsson, J., Joensson, H. N., Jahn, M., et al. (2020). Pooled CRISPRi screening of the cyanobacterium Synechocystis sp PCC 6803 for enhanced industrial phenotypes. *Nat. Commun.* 11, 1666. doi:10.1038/s41467-020-15491-7

Ye, Y. N., Hua, Z. G., Huang, J., Rao, N., and Guo, F. B. (2013). Ceg: A database of essential gene clusters. *BMC Genomics* 14, 769. doi:10.1186/1471-2164-14-769

Yu, B. J., Sung, B. H., Koob, M. D., Lee, C. H., Lee, J. H., Lee, W. S., et al. (2002). Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. *Nat. Biotechnol.* 20, 1018–1023. doi:10.1038/nbt740

Yu, B. J., Sung, B. H., Lee, J. Y., Son, S. H., Kim, M. S., and Kim, S. C. (2006). sucAB and sucCD are mutually essential genes in *Escherichia coli*. *FEMS Microbiol. Lett.* 254, 245–250. doi:10.1111/j.1574-6968.2005.00026.x

Yu, D., Ellis, H. M., Lee, E.-C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5978–5983. doi:10.1073/pnas.100127597

Yu, H., Greenbaum, D., Lu, H. X., Zhu, X., and Gerstein, M. (2004). Genomic analysis of essentiality within protein networks. *Trends Genet.* 20, 227–231. doi:10.1016/j.tig.2004.04.008

Zeng, M., Li, M., Fei, Z., Wu, F. X., Li, Y., Pan, Y., et al. (2021). A deep learning framework for identifying essential proteins by integrating multiple types of biological information. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 296–305. doi:10.1109/TCBB.2019.2897679

Zhang, C., Zheng, W., Cheng, M., Omenn, G. S., Freddolino, P. L., and Zhang, Y. (2021a). Functions of essential genes and a scale-free protein interaction network revealed by structure-based function and interaction prediction for a minimal genome. *J. Proteome Res.* 20, 1178–1189. doi:10.1021/acs.jproteome.0c00359

Zhang, F., Huo, K., Song, X., Quan, Y., Wang, S., Zhang, Z., et al. (2020). Engineering of a genome-reduced strain Bacillus amyloliquefaciens for enhancing surfactin production. *Microb. Cell Fact.* 19 (1), 223. doi:10.1186/s12934-020-01485-z

Zhang, J., Xu, X., Li, X., Chen, X., Zhou, C., Liu, Y., et al. (2021b). Reducing the cell lysis to enhance yield of acid-stable alpha amylase by deletion of multiple peptidoglycan hydrolase-related genes in Bacillus amyloliquefaciens. *Int. J. Biol. Macromol.* 167, 777–786. doi:10.1016/j.ijbiomac.2020.11.193

Zhang, J., Yang, F., Yang, Y., Jiang, Y., and Huo, Y. X. (2019). Optimizing a CRISPR-Cpf1-based genome engineering system for Corynebacterium glutamicum. *Microb. Cell Fact.* 18, 60. doi:10.1186/s12934-019-1109-x

Zhang, J., Zong, W., Hong, W., Zhang, Z. T., and Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metab. Eng.* 47, 49–59. doi:10.1016/j.ymben.2018.03.007

Zhang, R., Ou, H. Y., and Zhang, C. T. (2004). Deg: A database of essential genes. *Nucleic Acids Res.* 32, 271–272. doi:10.1093/nar/gkh024

Zhang, R., and Zhang, C. (1994). Z curves, an intutive tool for visualizing and analyzing the DNA sequences. *J. Biomol. Struct. Dyn.* 11, 767–782. doi:10.1080/07391102.1994.10508031

Zheng, Y., Han, J., Liang, W., Li, R., Hu, X., Wang, B., et al. (2019). Characterization and repurposing of the endogenous Type I-F CRISPR-Cas system of Zymomonas mobilis for genome engineering. *Nucleic Acids Res.* 47, 11461–11475. doi:10.1093/nar/gkz940

Zhou, X. X., Zou, X., Chung, H. K., Gao, Y., Liu, Y., Qi, L. S., et al. (2018). A single-chain photoswitchable CRISPR-Cas9 architecture for light-inducible gene editing and transcription. *ACS Chem. Biol.* 13, 443–448. doi:10.1021/acschembio.7b00603

Zhu, D., Fu, Y., Liu, F., Xu, H., Saris, P. E. J., and Qiao, M. (2017). Enhanced heterologous protein productivity by genome reduction in Lactococcus lactis NZ9000. *Microb. Cell Fact.* 16, 1–13. doi:10.1186/s12934-016-0616-2

Zwiener, T., Dziuba, M., Mickoleit, F., Rückert, C., Busche, T., Kalinowski, J., et al. (2021). Towards a "chassis" for bacterial magnetosome biosynthesis: Genome streamlining of Magnetospirillum gryphiswaldense by multiple deletions. *Microb. Cell Fact.* 20, 35. doi:10.1186/s12934-021-01517-2