［招待論文：総説・レビュー論文］

Past, Present, and Future of Genome Modification in *Escherichia coli*
大腸菌ゲノム改変の過去・現在・未来

Hirotada Mori
Professor Emeritus, Nara Institute of Science and Technology (NAIST)
森 浩禎
奈良先端科学技術大学院大学名誉教授
Correspondence to: hmori@gtc.naist.jp

Abstract: *Escherichia coli* K-12 is one of the most well-studied species of bacteria. This species, however, is much more difficult to modify by homologous recombination (HR) than other model microorganisms. Research on HR in E. coli led to a better understanding of the molecular mechanisms of HR, resulting in technical improvements and rapid progress in genome research, and allowing whole-genome mutagenesis and large-scale genome modifications. Developments using λ Red (*exo*, *bet*, and *gam*) and CRISPR-Cas have made *E. coli* as amenable to genome modification as other model microorganisms, such as Saccharomyces cerevisiae and Bacillus subtilis. This review describes the history of recombination research in *E. coli,* as well as improvements in techniques for genome modification by HR. This review also describes the results of large-scale genome modification of *E. coli* using these technologies, including DNA synthesis and assembly. In addition, this article reviews recent advances in genome modification, considers future directions, and describes problems associated with the creation of cells by design.

大腸菌（*Escherichia coli*）K-12 株は、最もよく研究されている細菌種の 1 つである。しかし、この種は他のモデル微生物に比べ、相同組換え（HR）による改変が非常に困難であった。大腸菌の HR の研究は、HR の分子機構の理解を深め、ゲノム研究の技術的向上と急速な進展をもたらし、全ゲノム変異誘発と大規模なゲノム改変を可能にしてきた。λ Red（*exo*、*bet*、*gam*）や CRISPR-Cas を用いた開発により、大腸菌は酵母や枯草菌といった他のモデル微生物と同様にゲノム改変を行うことができるようになった。本総説では、大腸菌の組換え研究の歴史と、HR によるゲノム改変技術の向上について述べる。また、これらの技術を用いた大腸菌の大規模なゲノム改変の成果として、DNA 合成やアセンブリーなどについて解説する。さらに、最近のゲノム改変の進展を概観し、今後の方向性を考察するとともに、新規細胞の設計に伴う問題点についても述べる。

Keywords: *E. coli,* genome modification, synthetic biology
大腸菌、ゲノム改変、合成生物学

1 Introduction

The discoveries of bacterial conjugation (Lederberg and Tatum, 1946) and of generalized transduction (Lennox, 1955) have enabled genetics research in *Escherichia coli* K-12. Subsequent genetic investigations of *E. coli* K-12 and its bacteriophages have increased knowledge of gene structure and function, and have led to the emergence of molecular biology. Although DNA transfer by transformation had been described previously (Griffith, 1928) and was shown to occur naturally in *Pneumococcus* (Griffith, 1928), *Haemophilus* (Alexander and Leidy, 1950) and *Bacillus subtilis* (Spizizen, 1958), *E. coli* proved to be recalcitrant. Treatment with $CalCl_2$ allowed the transformation (transfection) of *E. coli* with bacteriophage (Mandel and Higa, 1970) and plasmid DNA (Cohen et al., 1972), but not transformation by (linear) chromosomal DNA. Based on a hypothesis that an endogenous exonuclease in *E. coli* degrades linear DNA, *E. coli recBCD* mutants, which lack the RecBCD exonuclease, were found to be transformable by chromosomal DNA, provided the strain carried a cryptic prophage encoding the SbcBC(D) recombinase (Cosloy and Oishi, 1973). Expression of the λ Red recombinase (*exo*, *bet*, and *gam*) was shown to increase the efficiency of homologous recombination (HR) with linear DNA (Murphy, 1998), leading to the use of λ Red recombinase in highly efficient systems for direct modification of chromosomal genes via HR (Datsenko and Wanner, 2000; Yu et al., 2000).

Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas systems participate in acquired immunity in Archea and Bacteria (Ishino et al., 2018). Although these unusual repetitive DNA sequences were first described in 1987 (Ishino et al., 1987), molecular understanding of their function was first determined 25 years later (Jinek et al., 2012), leading to dramatic advances in the technology of genome modification (Wanner et al., 2014).

Genome research has become increasingly important in the 21st century.

Technological innovations in the 1990s increased the efficiency and reduced the costs of genome modification and analysis, such as DNA sequencing and DNA synthesis. Construction of a minimal genome enabled improvements in the ability to synthesize antibiotics and produce other valuable materials.

To date, large-scale deletions of genes other than those that are clearly unnecessary, such as prophage, transposon regions, and insertion sequences, have been unsuccessful (Kolisnychenko et al., 2002). Knowledge of the principles of genome construction is still incomplete even in model organisms such as *E. coli.* Although a fully chemical synthetic bacterial genome has been constructed in *Mycoplasma* (Gibson et al., 2010; Hutchison et al., 2016), it was not completed by design. The development of whole-cell metabolic models has been steadily progressing (Edwards and Palsson, 2000; Feist et al., 2007; Monk et al., 2017; Orth et al., 2011; Reed et al., 2003), and these models have become platforms for genome design. These models, even those in *E. coli*, contain large numbers of genes with unknown or incomplete functions (Orth and Palsson, 2012). Metabolic and whole-cell models have been constructed for Mycoplasma (Karr et al., 2012) and are progressing steadily for *E. coli* (Macklin et al., 2020), although additional research is required to create an *E. coli* model cell for genome design. This review summarizes the historical background of technological improvements, shows examples of past and ongoing research, and considers the current status and future of this research.

2　Historical perspective of *E. coli* as a biological research tool

2.1　Before the molecular biology era

The discovery of conjugation in *E. coli* K-12 (Lederberg and Tatum, 1946), which was thought to be sexless and to grow monogamously, and of transduction (Lennox, 1955), showed that genetic traits could be transferred between bacteria, leading to the use of *E. coli* K-12 as a model cell. Studies of the molecular mechanism of conjugation showed that it required a fertility factor and that DNA is transferred by the Type IV secretion system (Lawley et al., 2003), the detailed molecular structure of which was visualized by cryo-electron microscopy (Hu et al.,

2019). Results showing that genetic transformation requires extracellular DNA, not protein or RNA (Griffith, 1928), provided proof that genes are composed of DNA. Moreover, bacterial species, such as *Pneumococcus* (Griffith, 1928), *Haemophilus influenzae* (Alexander and Leidy, 1950), and *Bacillus subtilis* (Spizizen, 1958), were shown to have the ability to take up extracellular DNA.

The discovery of conjugation and transduction in *E. coli* made genetics possible. HR methods based on λ Red have been expanded for genome-scale functional analyses of *E. coli* and its phages (Marinelli et al., 2012; Murphy, 2016).

2.2 Genetic and genomic engineering in the molecular biology era

Following studies showing that *recBCD* mutations suppress linear DNA degradation and *sbcA* mutations activate the *recET* pathway, HR was developed for cloning PCR products (Oliner et al., 1993). Findings showing that λ Red markedly increased the efficiency of HR in *E. coli* (Murphy, 1998) led to the development of tools for genetic modification and their use in *E. coli* and other microorganisms (Datsenko and Wanner, 2000; Fels et al., 2020; Yu et al., 2000). The method illustrated in Fig. 1 requires only 35 bp of homology for efficient recombination and

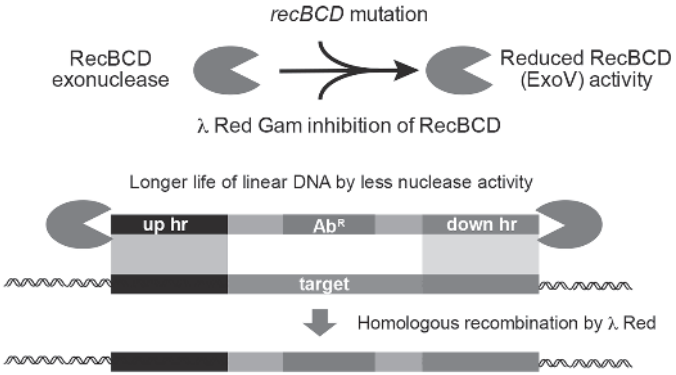

Fig. 1 Targeted deletion by homologous recombination
Two methods have been used to block RecD exonuclease: (1) using recBC mutations and (2) λ Red Gam synthesis. Cells are transformed with linear double-stranded (ds) DNA encoding an antibiotic resistance (Ab^R) cassette and ends homology regions (hr) of upstream (up) and downstream (down) regions of the target.

was adopted by the Japan genome project to construct the Keio collection of single-gene *E. coli* deletions (Baba et al., 2006). Greater understanding of the molecular mechanism of λ Red HR led to many improvements for its use in *E. coli*, phages, and other bacteria (Ellis et al., 2001; Li et al., 2013; Marinelli et al., 2008; Wang et al., 2006; Wannier et al., 2020). Targeted replacement by HR requires suppression or inhibition of the RecBCD exonuclease.

A strategy was subsequently developed for constructing "markerless" (scarless) gene replacements (deletions or substitutions) (Pósfai et al., 1999). This strategy involved the introduction of I-SceI sites into the *E. coli* chromosome, which had been absent from the genome. This scarless protocol was subsequently combined with λ Red HR to create scarless deletions (Fig. 2A).

Fig. 2B illustrates an alternative protocol for constructing scarless mutations. Unwanted sequences may be removed from genomes using a counter-selection method. A killing gene, also called a suicide gene, which can provide strong negative selection, can be introduced into cells under permissive conditions when the gene product is inactive. Treating cells with an inducer of its synthesis or a compound that inhibits cells harboring its product provides selection against cells bearing the inhibitor gene. This method has been used to target different regions of the *E. coli* chromosome, including *sacB* (Metcalf et al., 1996), *tetA* (Metcalf et al., 1996), and *tolC* (Gregg et al., 2014). Two-step protocols in Fig. 2B have provided powerful tools for recombineering *E. coli* and related bacteria with single-stranded (ss) and double-stranded (ds) DNA (Leatham-Jensen et al., 2012).

2.3　Genome-scale modification in the genome research era

2.3.1　Deletion of a large genomic region by random Tn insertion

Site-specific recombination at the *loxP* site resulted in two different types of Tn insertion mutations. Tn insertions located at both ends to be deleted were selected from each insertion mutation library and combined on one genome using P1 transduction. In the presence of overexpressed Cre protein, the fragment located between the two types of Tn was removed by Cre-*loxP* site-specific recombination

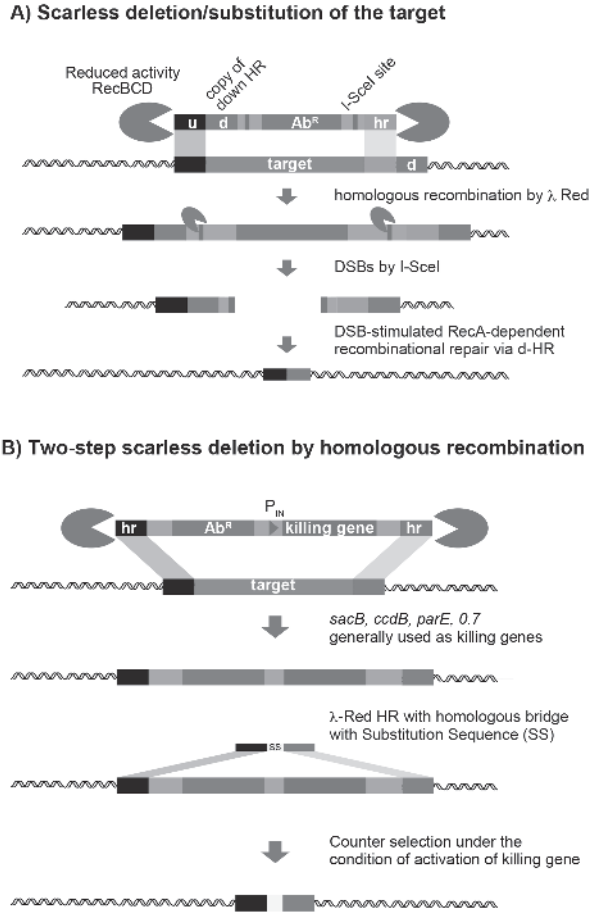

Fig. 2A) Scarless deletion/substitution of the target
Scarless deletion using I-SceI nuclease. The drug-resistant fragment flanked by I-SceI restriction enzyme sites is amplified with up and the next to the downside homology regions (u and hr) and introduced into the genome by λ Red homologous recombination (Kolisnychenko et al., 2002). The I-SceI-flanked segment is eliminated by expressing the meganuclease I-SceI, resulting in a double-strand break (DSB), DSB-stimulated DNA repair, and RecA-dependent recombination between the d direct repeats.
B) Two-step scarless deletion by homologous recombination
The Ab (antibiotic resistance gene) with the killing gene, such as *sacB*, *ccdB*, *parE*, or phage T7 *0.7* under the control of the tightly regulated promoter, such as *rhaBp* (Haldimann et al., 1998) is amplified with 36- to 40-nt homology region (hr) to target. The amplified fragment has then been transformed into a strain expressing λ Red to insert into the genome. Double-stranded "Substitution Sequence (SS)" with flanking hrs is transformed into the Ab resistant fragment integrated strain expressing λ Red. The transformants are selected in the presence of L-rhamnose, preferably with L-rhamnose as the sole carbon source.

(Fig. 3A) (Yu et al., 2002). These transposons were used to construct separate large-scale Tn insertion libraries, which were subsequently combined in the same strain by P1 transduction to yield a double Tn mutant. A site-specific recombinase was introduced to remove the chromosome segments between the Tn elements. This method has been used to construct large-scale genomic deletions by, for example, deleting 60 to 120 kb between pairs of Tn elements and choosing those mutants that did not impair cell growth. Cre was subsequently used to eliminate segments between *loxP* sites, followed by the introduction via P1 transduction of Tn elements for an additional large deletion and the repeat of the entire process (Fig. 3A).

Another approach for deleting large chromosomal segments randomly relies on a complex transposon with one Tn element inside another, Tn-in-Tn, carrying two types of Tn elements with opposite terminal repeat directions (Fig. 3B) (Goryshin et al., 2003). Following random transposition of Tn-in-Tn into the chromosome, synthesis of the transposase for the internal element is induced, leading to its transposition to a new site and elimination of DNA between the original and new sites. When this process was repeated 20 times, the average deletion length was found to be about 10 kb, with a total of about 200 kb being successfully deleted (Goryshin et al., 2003).

2.3.2 Large-scale deletion by HR

E. coli is thought to have acquired many genes to survive in diverse environments. Shrinking the *E. coli* genome is thought to improve the efficiency of metabolic functions and reduce redundancy in genomic and regulatory structures (Fehér et al., 2008; Pósfai et al., 2006). Using HR, the *E. coli* K-12 genome lacking K-islands, which were identified by comparative genomics as recent horizontal acquisitions to the genome, was reduced (Fig. 2A) (Kolisnychenko et al., 2002). This method, which was based on the accumulation of scarless deletions by HR and DSB, allowed elimination of 15% of the *E. coli* K-12 genome. Twelve K-islands, containing fragments of cryptic phage, transposons, disrupted pseudogenes, and genes of unknown function, were deleted. Ultimately, 9.3% of genes in the genome, including

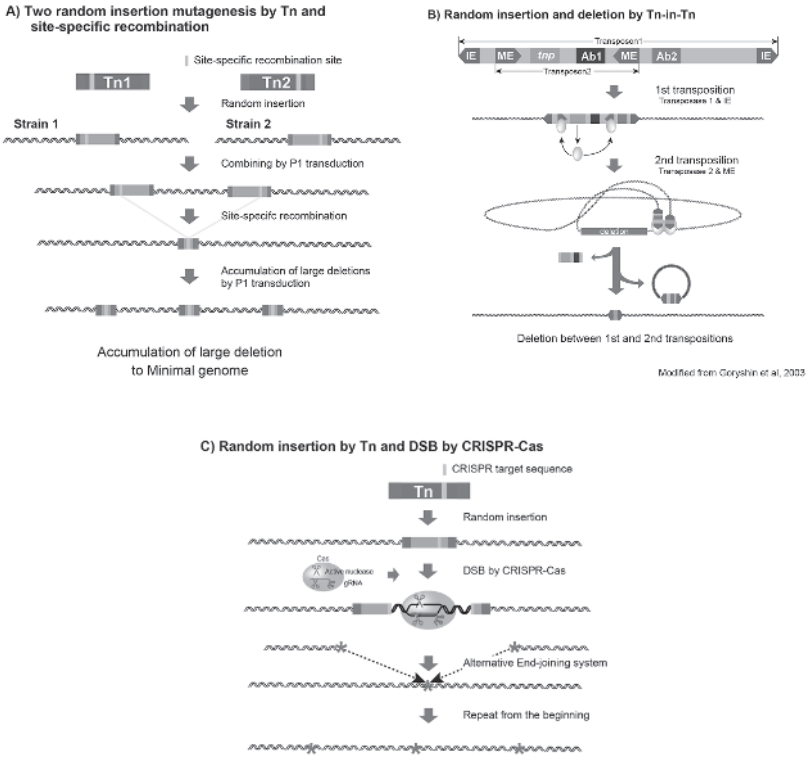

Fig. 3A) Two random insertion mutagenesis by Tn and site-specific recombination
Mutant Tns with less sequence specificity for insertion sites on chromosomes have been developed, with transposons such as Tn5, Tn10, and Mariner often used. Two different Tns of two different drug resistance genes were randomly mutated, and the location of insertion on the genome was determined by PCR and sequencing. E. coli strains with insertions at appropriate positions were selected and combined into a single *E. coli* strain using the P1 transduction method. This method used site-specific recombination at each Tn and induced recombination by increasing the production of site-specific recombinase and deleting the region between Tns.
B) Random insertion and deletion by Tn-in-Tn
A complex with Tnp that recognizes IE is introduced into the cell to obtain the first random insertion mutation. Then, synthetic induction of Tnp recognizing the internal ME is performed to obtain a transition mutation; the direction of the second transition results in a deletion between two different Tn insertion sites.
C) Random insertion by Tn and DSB by CRISPR-Cas
Insertion of a Tn fragment into the genome, followed by CRISPR-Cas cutting of the inside of the Tn fragment. This yielded a strain in which nuclease activity deleted the periphery.

24 of 44 transposon regions, were deleted (Kolisnychenko et al., 2002). Furthermore, strains with large deletions grew as well on minimal medium as wild-type strains, confirming that these K-islands did not contain essential genes.

These results suggest that mobile elements, such as IS, which may drive evolution but induce genomic instability, can be deleted, as can genes with unnecessary function, and groups of genes that adversely affect the bacterial growth environment, including in humans. However, it is not easy to predict which genes have those functions. A comparison of genomes of different *E. coli* strains enabled the selection of genes that were present in K-12 but absent in other *E. coli* strains. This resulted in the identification of a set of candidate genes, constituting about 20% of the genome, for deletion. This is an example of purposefully designed deletions that contain unstable factors and gene groups that are not necessary for bacterial growth. Moreover, the strains with large deletions, such as MDS42 and MDS43, grew almost as well as wild-type, with the stability of their genomes and their transformation efficiency being improved. An improved *E. coli* strain MDS69 with additional deletions is commercially available from Scarab Genomics.

Another study first compiled a list of predicted essential genes, followed by the use of an HR method to delete regions between these genes (Hashimoto et al., 2005). Using λ Red HR, alternate Ab resistance cassettes were inserted into intergenic regions between two essential genes, followed by combining them by P1 transduction and eliminating the inserted cassettes with λ Red HR. About 30% of the *E. coli* genome was deleted by combining the largest deletions between essential genes using P1 transduction, with the resulting phenotypes analyzed by determining their cell shape and nucleoid organization.

2.4　Genome-scale genetic modification in systems and synthetic biology

2.4.1　CRISPR-Cas application

Technological developments since the elucidation of the CRISPR-Cas mechanism have marked a major turning point in the biological sciences, just as the discovery of junctions, restriction enzymes, and vectors paved the way for molecular

biology. CRISPR-Cas provides a process to design cleavage sites, called double-strand breaks (DSBs), at will. Fig. 2A shows how targeting the I-SceI site to a specific location generates a DSB that leads to the formation of the designed deletion. Fig. 3C shows how random Tn mutagenesis can be used to generate DSBs and nearly random deletions by CRISPR-Cas (Ma et al., 2022).

CRISPR-Cas is most often used to create DSBs at locations governed by the sequence of guide RNA (gRNA). Accumulation of DSBs if unrepaired is lethal. In eucaryotes, most broken DNA ends are bridged by nonhomologous end-joining (NHEJ) (Wyman and Kanaar, 2006). NHEJ, however, is usually not possible in bacteria due to lack of the key NHEJ proteins Ku and Ligase-D. In *E. coli*, DSBs are most often repaired by HR and less frequently by alternative end-joining. An alternative end-joining (A-EJ) mechanism of repairing DSBs involves end-resection by RecBCD, end synapsis via microhomologies, and ligation of DNA ends (Bhattacharyya et al., 2018; Chayot et al., 2010) by LigA (Fig. 3C). Combining CRISPR-Cas with λ Red HR permits fashioning scarless deletions, insertions, or substitutions by design, limited only by occurrences of PAM sequences (Fig. 4A) (Adli, 2018).

Development of mutant Cas proteins lacking endonuclease activity has allowed precise base editing at very limited target sites in genomes by fusion of the enzyme cytosine deaminase to an inactive Cas subunit (Banno et al., 2018), and various point mutations and small insertions/deletions by fusion of the reverse transcriptase to a single active Cas (nickase) subunit (Park et al., 2021) (Fig. 4B). This subject has been reviewed previously (Dong et al., 2021).

The generation of catalytically inactive Cas by mutation has allowed repurposing CRISPR as an RNA-guided platform that can specifically interfere with transcription elongation, RNA polymerase binding, or transcription factor binding (CRISPR interference; CRISPRi) using a single guide RNA (sgRNA) chimera (Cui et al., 2018; Peters et al., 2016; Qi et al., 2013; Rousset et al., 2018; Wang et al., 2018).

CRISPRi was used to analyze a group of essential genes in *B. subtilis*, although this approach did not focus directly on the genome modification (Peters et al., 2016).

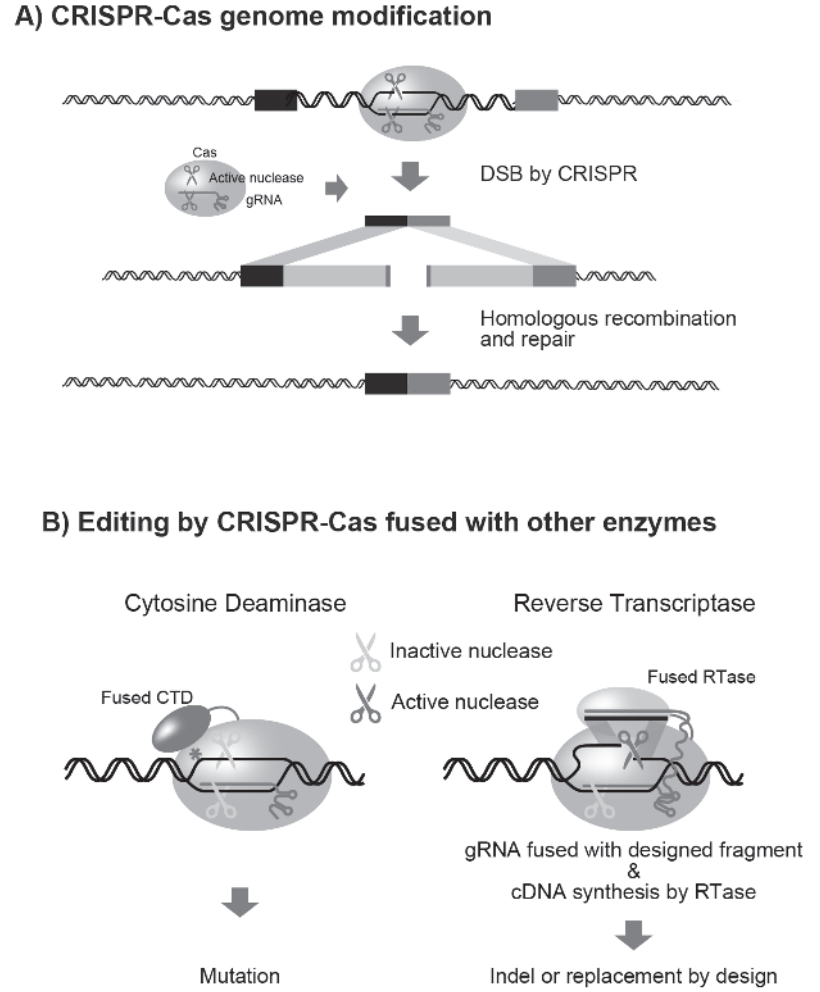

Fig. 4A) Scarless deletion/substitution using CRISPR-Cas
DSB was induced by CRISPR-Cas, and DNA fragments were transformed by bridging homologous regions at both ends of each double-strand break, resulting in genome repair and yielding to the circular genome.
B) Editing by CRISPR-Cas fused with other enzymes
Genome editing by fusion protein with a function different from that of Cas protein. Left panel: fusion of cytosine deaminase to a Cas protein with mutation-inactivated DNase activities (Banno et al., 2018). Right panel: fusion of reverse transcriptase to a Cas, which inactivates only the nick on the other strand, providing a template for repairing the nick site and introducing the mutation by a reverse-transcribed sequence (Tong et al., 2021).

CRISPRi screening of *E. coli* was performed by synthesizing a library of 92,000 sgRNA sequences covering the entire genome, with PAM sequences as the only constraint (Cui et al., 2018), thus identifying *E. coli* essential genes and genes essential for phage l growth (Rousset et al., 2018). In a separate study, 60,000 sgRNAs were evaluated to test essentiality while also assessing the design of sgRNAs for all genes, including those that did not encode RNA (Wang et al., 2018). Essentiality was tested with a pooled library, with the results evaluated by determining the relative change in read count by next-generation sequencing (NGS).

2.4.2 Acceleration of evolution under the constraint of mutation direction by oligo DNA

The mutation of many genes simultaneously can result in the synthetic lethality of genetic interactions, in which one mutation affects other genes, making the accumulation of mutations difficult. Therefore, a method was devised to accelerate the evolution of mutagenesis by adding the constraint of viability and utilizing the principle of HR of λ Red, while restricting sequences using synthetic DNA. This Multiplex Automatable Genome Engineering (MAGE) method uses the b protein of l Red and long (90 nt) synthetic ssDNAs, allowing acquisition of mutations on a genome-scale without lethal or severe growth-defect mutations because mutant selection is based on cell growth (Fig. 5) (Wang et al., 2009). Although MAGE was originally performed robotically, it can also be performed manually. MAGE has provided a powerful tool for genome-wide codon replacement (Isaacs et al., 2011), metabolic engineering (Wang and Church, 2011), and other biological functions (Fig. 5) (Lajoie et al., 2013). λ Red HR is increased in mismatch repair mutants and shows strand bias (Costantino and Court, 2003; Li et al., 2003). MAGE technology has been further improved by using MAGE and λ Red HR to stimulate the evolution of host *E. coli* primase and helicase (Lajoie et al., 2012), which control the length of Okazaki lagging-strand fragments (Balakrishnan and Bambara, 2013).

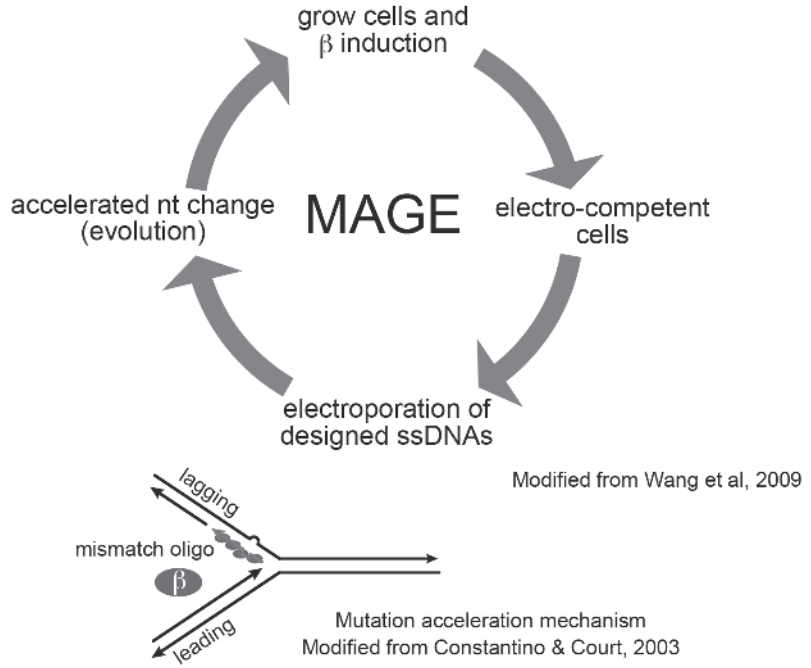

Fig. 5 Acceleration of in vivo evolution by λ Red β and ssDNAs
Introduction of an ssDNA about 90 bp in length to be mutated in the cell via the induction of λ Red λ protein, which promoted the introduction of mutations on the lagging strand during DNA replication and accelerated the introduction of mutations throughout the genome.

3 Genome-scale projects

3.1 Minimal genome projects

The minimal genome project, which began in 1997 based on a concept of minimal gene sets (Koonin, 1997; Mushegian and Koonin, 1996), has been ongoing for many years but has not yet been completed. The main reasons for non-completion are the existence of alternative pathways and compensatory circuits in the intracellular functional network, and the existence of orphan enzymes whose genes have not yet been identified. The minimal genome concept was later linked to the concept of chassis genomes in synthetic biology (Danchin, 2012).

3.1.1 Large-scale deletion by random transposon insertion

Strains with large-scale deletions have been constructed using *loxP* site-specific recombination sites embedded in two types of Tn5 and a fragment of Tn5 randomly inserted into the *E. coli* genome (Yu et al., 2002). For example, two types of Tn5 incorporating separate Kan and Chl antibiotic resistance genes and *loxP* site-specific recombination sites were constructed and used to generate random insertion mutant libraries. Mutants were selected from each Kan- and Chl-resistant library that flanked the region to be deleted, and the two insertion mutations combined into a single genome by P1 transduction. Addition of the Cre protein allowed the large-scale deletion of the genomic region between the *loxP* site-specific recombination sites. This method was used to introduce large-scale genomic deletions in six strains by combining two types of transposons, deleting 60 to 120 kb between them, and selecting mutations that did not impair the growth of the deleted cells. The selection marker between *loxP* sites were subsequently removed from the strains by Cre, followed by the introduction of another large deletion region into a single genome by P1 transduction, resulting in the construction of a minimal genome.

By contrast, Tn5 derivatives have been used to create Tn-in-Tn system to generate a large genomic deletion (Goryshin et al., 2003). Because the average length of each deleted region was about 10 kb, repeating this process 20 times successfully introduced deletions into a ~200 kb region.

3.1.2 Large scarless deletion by HR

The reduced genome lacking K-islands (Kolisnychenko et al., 2002) has been further improved (Pósfai et al., 2006). Although *E. coli* was originally described as an intestinal bacterium, it has acquired a diverse set of genes, enabling it to survive in various environments. Shrinking of the genome is thought to improve the efficiency of metabolic functions and reduce redundancy in genomic and regulatory structures (Pósfai et al., 2006). Mobile elements, such as IS, which may drive evolution but induce genomic instability, genes with unnecessary function, and groups of genes that adversely affect the bacterial growth environment, including in humans, can be

deleted. However, predicting the genes that have these functions is difficult. A comparison of the genomes of different *E. coli* strains identified selected genes that were present in K-12 but absent from other *E. coli* strains, resulted in the selection of a set of candidate genes, comprising about 20% of the genome, for deletion.

The deletion method (Kolisnychenko et al., 2002) was based on the accumulation of scarless deletions by HR and DSB, resulting in the successful deletion of 15% of the *E. coli* K-12 genome. This was an example of purposefully designed deletions of sequences that are unstable factors and gene groups that are not necessary for the growth of *E. coli*.

The growth of the final strains with large deletions, MDS42 and MDS43, were almost the same as that of the wild type, although these deletions improved genome stability and transformation efficiency, making these strains practical reduced-genome *E. coli*. MDS69, an improved *E. coli* strain with additional deletions, is currently available commercially from Scarab Genomics.

Another approach of scarless deletion of a large region between essential genes involved the performance of two HR events (Fig. 2B) (Hashimoto et al., 2005). This method resulted in the deletion of the largest possible region from the essential intergenic region and the accumulation of deletions by P1 transduction to yield a minimal genome with large deletions.

Analyses of the phenotypes of *E. coli* from which about 30% of the genome had been deleted showed that growth rate was inversely associated with the size of the deleted region (Hashimoto et al., 2005). Deletion also altered cell morphology, with changes in cell length and width and in nucleoid organization. Attempts to combine these large deletions showed some could not be combined (Kato, J., personal communication), perhaps because the combination resulted in synthetic lethality. Therefore, it is still difficult to determine the associations between combinations of gene deletions and specific phenotypes. This deletion project has since become a joint project with Kyowa Hakko Co. Ltd.

Efforts have been made to develop bacteria with beneficial genomes for the production of materials, especially with industrial applications, without inhibition of

cell growth. One strain, MGF-01, was generated by deleting 1.03 Mb from 53 regions using P1 (Mizoguchi et al., 2008). These deletions increased glucose consumption 1.44-fold and acetate accumulation 0.09-fold, confirming the efficacy of this method (Mizoguchi et al., 2008).

The strain MS56 has a genome reduced by 23% (Park et al., 2014). It was generated by removing IS and other factors that may cause instability in plasmids containing foreign genes, and its stability and efficiency of expression of foreign gene products was analyzed (Park et al., 2014). Evaluation of the scarless HR deletion method (Kolisnychenko et al., 2002; Pósfai et al., 2006) with human tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) and bone morphogenetic protein-2 (BMP2) showed its superiority. This led to the development of genome-reduced strains useful for biosynthesis in industrial applications.

3.2 Recoding genome

3.2.1 Recoding genome by recombineering and ssDNA accelerated evolution

Despite the significant progress of the Minimal genome project in accumulating individual genes with specific functions, it is still almost impossible to fully design a genome with deletion of many genes. These drawbacks may be overcome by genetic-interaction analysis of systematic double deletion strains.

Evolutionary methods have also been explored. Mutants that are not viable or grow very slowly are eliminated during the selection process. Clarification of the molecular mechanism of RED recombination has shown that ssDNA predominantly introduces mutations into the replicating lagging strand through the activity of the beta-protein alone (Ellis et al., 2001). These findings led to the development of the Multiplex Automated Genome Engineering (MAGE) method using in vivo evolution with multiple types of designed ssDNA and b proteins (Fig. 5) (Wang et al., 2009). The running of 5-35 MAGE cycles resulted in the ~10^5 mutant strains and increased the production of the target product, isoplenoid lycopen, up to 5-fold within 3 days (Wang et al., 2009).

This development has enabled genome modification by deliberately limiting the

direction of mutation and accelerating evolution by recombination of many parts of the genome at once. This technology has since been further improved, allowing its use on a larger scale. For example, a recoding genome was constructed by replacing the TAG termination codons on all 314 *E. coli* genes bearing these codons with TAA termination codons (Isaacs et al., 2011). Because these 314 genes are scattered throughout the genome, the genome was divided into 32 regions, and the MAGE cycle was run for each region to obtain evolutionary mutant strains. The Conjugative Assembly Genome Engineering (CAGE) method was also developed to integrate the mutated chromosomal sites into a single *E. coli* strain using conjugation, resulting in an *E. coli* strain in which all terminal TAGs were replaced by TAAs. Although recoding was expected to eliminate the need for TAG codons, *prfA*, the gene encoding releasing factor (RF1), which recognizes TAG codons, was deleted, shows the ability to replace TAG codons with other codons. This technology was further improved by developing primase and helicase mutant strains, which contain mutations that control the lengths of Okazaki fragments synthesized by the lagging strand (Lajoie et al., 2012).

It may also be possible to replace codons for a specific amino acid, rather than termination codons. Forty-two highly expressed essential genes were selected, and rare codons in these genes were replaced by DNA synthesis; if this method was unsuccessful, these codons were replaced using MAGE. Ultimately, 405 codons on 42 highly expressed essential genes were replaced, resulting in reductions in cell growth. These results showed that genome-wide codon replacement is feasible and that codons can be replaced using MAGE, with no or minimal effect on cell growth. Providing artificially modified organisms with a genetic code that does not exist in the natural would thus ensure the safety of these organisms, even if they are released to the outside world.

3.2.2 Recoding genome by synthesis

A method has been developed to replace a target region of the genome with a fragment of synthetic DNA designed for recoding by assembly in vivo, including

recoding of the entire *E. coli* genome (Fig. 6). For example, replacement of the codons UAG (stop), AGG-AGA (Arg), AGC-AGU (Ser), and UUG-UUA (Leu) with other synonymous codons from the genome resulted in the construction of an *E. coli* genome with 57 codons. Similarly, replacing the codons TCG, TCA (Ser), and TAG (stop) resulted in the construction of an *E. coli* genome with 61 codons. Both methods used designed synthetic DNA, with the genome-reduced strain MDS42 used as the parent strain.

Although these methods showed some differences in their details, both involved assembly of the genome by HR in yeast cells and its transfer to *E. coli* cells. In one method, the target region on the genome was removed, the assembled recoded genome was inserted into the target region using *attL-attP* site-specific recombination, and the vector region was deleted by CRISPR (Ostrov et al., 2016). In the other

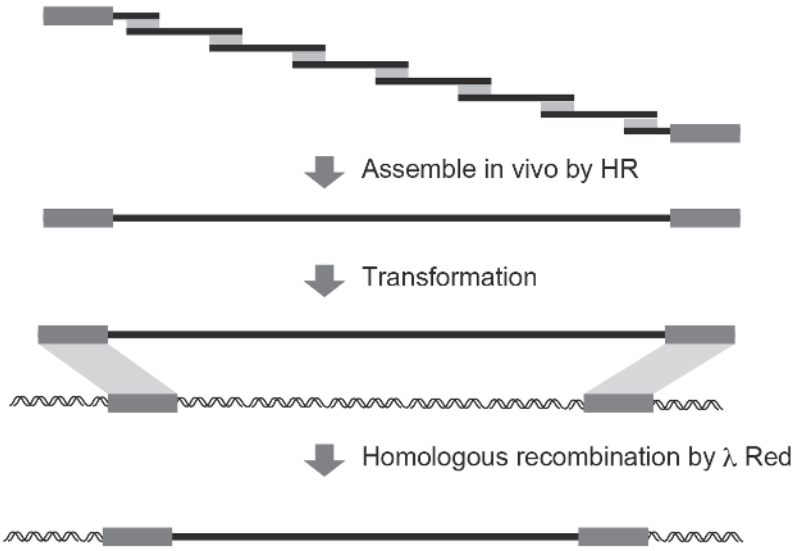

Fig. 6 Assembly of synthesized DNA and homologous recombination.
Introduction of synthetic DNA fragments into cells, generally yeast cells, assembly of the fragments by in vivo homologous recombination. After assembling, the assembled fragment is collected and transformed into *E. coli* cell and replacement of the target region in the λ Red-induced strain by homologous recombination.

method, fragments that accumulated on the BAC vector were integrated into the recipient genome by transferring them to the recipient using conjugation, although assembly in yeast cells was the same (Wang et al., 2016). This method resulted in the construction of a partial recoded genome by linearizing both ends of the fragments that had accumulated on the vector by CRISPR double-strand breaks and replacing them with the target regions of the genome by HR using λ RED recombinase. This step was repeated to construct the entire recoded genome (Fredens et al., 2019). The resulting *E. coli* strain Syn61 with a recoded genome was found to grow more slowly and have a longer cell length than the parental strain MDS42.

3.3 Genome-scale analysis towards genome design platform

In genomics, the construction of mutants is an important first step in analyzing the biological functions of genes and their products. As of September 2020, the *E. coli* genome, annotated as GenBank entry U00096.3, included 4609 genes, with 4285 of these genes encoding proteins, many of which have unknown functions. In addition, this genome was found to include genes encoding small proteins and non-coding RNAs (Hobbs et al., 2010). Genome-scale metabolic models of *E. coli* have been developed, such as iJE660 (Edwards and Palsson, 2000), iJR904 (Reed et al., 2003), iAF1260 (Feist et al., 2007), iJO1366 (Orth et al., 2011), and iML1515 (Monk et al., 2017), with others still being developed and improved. Refinements of these models have shown the present of as yet unidentified alternative pathways and isozymes and gaps in metabolic networks (orphan reactions) (Orth and Palsson, 2012; Orth and Palsson, 2010). Knowledge of *E. coli* is also incomplete (Orth et al., 2011; Orth and Palsson, 2012; Orth and Palsson, 2010).

The minimal gene set concept (Koonin, 1997) has become the minimal genome project, but it is still far from complete. By contrast, the minimal genome concept has expanded to the concept of a minimal genome factory that optimizes the genome to produce valuable products (Mizoguchi et al., 2007). These concepts have now expanded to include the concept of the chassis genome in synthetic biology (Calero and Nikel, 2019; Danchin, 2012).

4 Discussion and perspective

4.1 Transition of biological concepts

About 50 years have elapsed since recombinant *E. coli* gene modification technology was initially introduced, from the elucidation of its molecular mechanism to the development of methods based on phage recombination mechanisms and improvement of the technology. This has enabled almost any type of sequence modification, from genome-scale large modification to base-level modification. Genome research in the 1990s may therefore represent a turning point in biology both technologically and conceptually. This situation was similar in the 1970s, when genetic modification techniques were developed and molecular biology began to make significant progress.

Often, new concepts are not immediately accepted. One such example is the difference between “forward genetics” and “reverse genetics”, which was initially developed by physicists, but was not fully accepted by researchers in genetics (Yura, T., personal communication). Differences in acceptance were not likely due to differences in ways of thinking, but increased understanding was likely due to experimental efforts. The throughput of sequencing technology has expanded about 1,000-fold from the start of the project in 1990 to the completion of the first draft sequence of the human genome. With the technology available at that time, the genome of *E. coli* took 7 years to complete. However, the 21st century has seen the development of sequencing technologies based on novel concepts, and sequencing is now more than a billion times more efficient than it was in 1990. The driving force behind this development dates back to the 1990s. Researchers understood the importance of comparative analysis of individual human genomes and the need for further development of sequencing technology as the next step after the completion of the human genome.

Technological innovation has not only affected the speed of analysis, but has made possible more precise and diverse analyses, and at lower cost. For example, the range of applications of sequencing technology has rapidly expanded to include analyses of gene expression, protein-DNA interactions, protein-protein interactions,

nucleoids, and species distribution in populations. Moreover, genome editing in the 21st century has been revolutionized by determining the molecular mechanism underlying CRISPR.

Although the recombination efficiency of *E. coli* is lower than that of other model microorganisms, such as *B. subtilis* and yeast, the tooling of λ Red recombinase made possible the modification of *E. coli* by genome-scale recombination, increasing its recombination efficiency. The availability of genome modification techniques by recombination in Gram-positive, Gram-negative, and eukaryotic unicellular organisms has created a favorable research environment for comparative analysis.

Progress in genome analysis led to the development of research resources in yeast, the creation of databases, the construction of gene clone and deletion strain libraries, and their sharing as community assets (Botstein and Fink, 2011; Giaever and Nislow, 2014). For example, participating institutions in Europe and Japan worked together as a community to build a *B. subtilis* deletion strain library (Kobayashi et al., 2003; Vagner et al., 1998).

Although a similar comprehensive experimental resource community has been proposed for *E. coli*, few laboratories agreed to participate. Groups at Keio University, the Nara Institute of Science and Technology, and Purdue University therefore agreed to develop this resource. Although the time to completion was undetermined, the development of many innovations resulted in the completion of the entire project in 3 years. These developed resources were subsequently shared with the research community as open resources. *E. coli* could therefore be positioned as a model microorganism alongside *B. subtilis* and yeast.

The importance of putting all the pieces together is enormous. The importance of analyzing the structure and function of target genes and proteins in detail through individually targeted analyses will likely remain unchanged. However, looking at the entire picture revealed aspects that could not be determined on individual studies of a narrow range of targets. Molecular biology in the 20th century has been described as a very precise “science of parts”, whereas genome research, starting in the 1990s and

extending into the 21st century, can be described as “science as a system”. Alignment of the two will likely greatly advance our understanding of the whole picture of life.

4.2 Concept of minimal genome

The minimal genome project, which started with the minimal gene set, has been a long-term effort to realize the minimal genome concept. In particular, the large-scale deletion efforts showed that, although deletion of individual genes in a region did not significantly affect the growth of an organism, simultaneous deletion of many gene clusters significantly affects growth and may even be lethal. These findings emphasized the importance of determining genetic interactions under conditions of synthetic lethality and sickness, and may have a major impact on network biology. Indeed, findings showing that cell lethality was due to multiple gene deletions and quantitative analyses of the effects of single-gene deletions on intracellular conditions (Ishii et al., 2007) suggested that many genes that are considered non-essential may repeatedly interact with each other to maintain intracellular stability, at least at the metabolite level. Efforts are underway to analyze genetic interactions through the systematic construction of double gene deletion strains. The saying in Japan, “If the wind blows, the bucket makers prosper”, is equivalent to the “Butterfly Effect” in chaos theory. Cells dynamically regulate transcription, translation, and enzyme activities while optimizing the balance between the genetic elements of the cell and the environmental factors affecting growth.

The minimal genome concept has also expanded to optimize the practical use of cells for industrial purposes (Mizoguchi et al., 2007; Park et al., 2014). Moreover, the beginning of synthetic biology has resulted in expansion of the concept of the “chassis genome” (Calero and Nikel, 2019; Danchin, 2012).

4.3 Dramatic changes in quality and quantity of data require a variety of analyses

The existence of comprehensive research resources and the development of analytic methods have led to a rapid increase in data. Effective use of these data

requires mathematical analyses and information processing technologies such as statistical analysis and modeling. Although analyzing the data generated from a single comprehensive study provides many hints on physiological functions and molecular mechanisms, these hints may not be successfully verified by experimental methods. Although data registered in databases and accessible through Web systems have been used for informatics analysis, the contribution of these enormous amounts of data to the progress of individual research is unclear. The situation may be dependent on the presence of a framework for sharing information that may provide direct material for experimental validation, such as proposals for individual molecular mechanisms obtained from informatics analysis. Another factor may result from the same research group performing comprehensive analyses and individual targeted research. The accumulation of experience is important in deepening research in individual studies, making it difficult to complete both comprehensive and individual analyses in a single laboratory. Establishment of a community-oriented framework is likely necessary to share comprehensive data, along with interpretations and/or suggestions. For example, a group at McMaster University in Canada has specialized in discovering anti-microbial antibiotics and has developed research using comprehensive resources. As soon as resources become available, they are used for exhaustive screening (Pathania et al., 2009), leading to acceleration of research and development, and further expansion (Brown and Wright, 2016). This group is therefore a good example of the successful use of both comprehensive and individual approaches in a single laboratory.

References

Adli, M. (2018) “The CRISPR tool kit for genome editing and beyond”, *Nature Communications.* 9(1):1911.

Alexander, H.E. and Leidy, G. (1950) “Transformation of type specificity of *Hemophilus influenzae*”, *AMA Am J Dis Child* . 80(5), pp. 877-878.

Baba, T., et al. (2006) “Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection”, *Mol Syst Biol* . 2(1):2006 0008.

Balakrishnan, L. and Bambara, R.A. (2013) “Okazaki fragment metabolism”, *Cold Spring Harb Perspect Biol.* 5(2).

Banno, S., et al. (2018) “Deaminase-mediated multiplex genome editing in *Escherichia coli*”, *Nature Microbiology.* 3(4), pp. 423-429.

Bhattacharyya, S., et al. (2018) “Phage Mu Gam protein promotes NHEJ in concert with *Escherichia coli* ligase”, *Proc Natl Acad Sci U S A.* 115(50):E11614-E11622.

Botstein, D. and Fink, G.R. (2011) “Yeast: an experimental organism for 21st Century biology”, *Genetics*. 189(3), pp. 695-704.

Brown, E.D. and Wright, G.D. (2016) “Antibacterial drug discovery in the resistance era”, *Nature*. 529(7586), pp. 336-343.

Calero, P. and Nikel, P.I. (2019) “Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms”, *Microbial Biotechnology.* 12(1), pp. 98-124.

Chayot, R., et al. (2010) “An end-joining repair mechanism in *Escherichia coli*”, *Proc Natl Acad Sci U S A.* 107(5), pp. 2141-2146.

Cohen, S.N., Chang, A.C. and Hsu, L. (1972) “Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA”, *Proc Natl Acad Sci U S A* . 69(8), pp. 2110-2114.

Cosloy, S.D. and Oishi, M. (1973) “Genetic transformation in *Escherichia coli* K12”, *Proc Natl Acad Sci U S A*. 70(1), pp. 84-87.

Costantino, N. and Court, D.L. (2003) “Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants”, *Proc Natl Acad Sci U S A.* 100(26), pp. 15748-15753.

Cui, L., et al. (2018) “A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9”, *Nature Communications.* 9(1):1912.

Danchin, A. (2012) “Scaling up synthetic biology: Do not forget the chassis”, *FEBS Lett.* 586(15), pp. 2129-2137.

Datsenko, K.A. and Wanner, B.L. (2000) “One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products”, *Proc Natl Acad Sci U S A.* 97(12), pp. 6640-6645.

Dong, H., Cui, Y. and Zhang, D. (2021) “CRISPR/Cas Technologies and Their Applications in *Escherichia coli*”, *Front Bioeng Biotechnol* . 9:762676.

Edwards, J.S. and Palsson, B.O. (2000) “The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities”, *Proc Natl Acad Sci U S A.* 97(10), pp. 5528-5533.

Ellis, H.M., et al. (2001) “High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides”, *Proc Natl Acad Sci U S A.* 98(12), pp. 6742-6746.

Fehér, T., et al. (2008) “Scarless engineering of the *Escherichia coli* genome”, *Methods Mol Biol.* 416, pp. 251-259.

Feist, A.M., et al. (2007) “A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information”, *Mol Syst Biol.* 3(1), 121.

Fels, U., Gevaert, K. and Van Damme, P. (2020) “Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics”, *Frontiers in Microbiology.* 11:548410.

Fredens, J., et al. (2019) “Total synthesis of *Escherichia coli* with a recoded genome”, *Nature.* 569(7757), pp. 514-518.

Giaever, G. and Nislow, C. (2014) “The yeast deletion collection: a decade of functional genomics”, *Genetics.* 197(2), pp. 451-465.

Gibson, D.G., et al. (2010) “Creation of a bacterial cell controlled by a chemically synthesized

genome", *Science.* 329(5987), pp. 52-56.
Goryshin, I.Y., et al. (2003) "Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis", *Genome Res.* 13(4), pp. 644-653.
Gregg, C.J., et al. (2014) "Rational optimization of *tolC* as a powerful dual selectable marker for genome engineering", *Nucleic Acids Res.* 42(7), pp. 4779-4790.
Griffith, F. (1928) "The Significance of Pneumococcal Types", *J Hyg (Lond)* . 27(2), pp. 113-159.
Haldimann, A., Daniels, L.L. and Wanner, B.L. (1998) "Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon", *J Bacteriol.* 180(5), pp. 1277-1286.
Hashimoto, M., et al. (2005) "Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome", *Mol Microbiol.* 55(1), pp. 137-149.
Hobbs, E.C., Astarita, J.L. and Storz, G. (2010) "Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection", *J Bacteriol* . 192(1), pp. 59-67.
Hu, B., Khara, P. and Christie, P.J. (2019) "Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*", *Proc Natl Acad Sci U S A.* 116(28), pp.14222-14227.
Hutchison, C.A., 3rd, et al. (2016) "Design and synthesis of a minimal bacterial genome", *Science.* 351(6280):aad6253.
Isaacs, F.J., et al. (2011) "Precise manipulation of chromosomes in vivo enables genome-wide codon replacement", *Science.* 333(6040), pp. 348-353.
Ishii, N., et al. (2007) "Multiple high-throughput analyses monitor the response of *E. coli* to perturbations", *Science.* 316(5824), pp. 593-597.
Ishino, Y., Krupovic, M. and Forterre, P. (2018) "History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology", *J Bacteriol* . 200(7).
Ishino, Y., et al. (1987) "Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product", *J Bacteriol.* 169(12), pp. 5429-5433.
Jinek, M., et al. (2012) "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science.* 337(6096), pp. 816-821.
Karr, J.R., et al. (2012) "A whole-cell computational model predicts phenotype from genotype", *Cell*. 150(2), pp. 389-401.
Kobayashi, K., et al. (2003) "Essential *Bacillus subtilis* genes", *Proc Natl Acad Sci U S A.* 100(8), pp. 4678-4683.
Kolisnychenko, V., et al. (2002) "Engineering a reduced *Escherichia coli* genome", *Genome Res.* 12(4), pp. 640-647.
Koonin, E.V. (1997) "Big time for small genomes", *Genome Res*. 7(5), pp. 418-421.
Lajoie, M.J., et al. (2012) "Manipulating replisome dynamics to enhance λ Red-mediated multiplex genome engineering", *Nucleic Acids Res.* 40(22):e170.
Lajoie, M.J., et al. (2013) "Genomically recoded organisms expand biological functions", *Science*. 342(6156), pp. 357-360.
Lawley, T.D., et al. (2003) "F factor conjugation is a true type IV secretion system", *FEMS Microbiol Lett.* 224(1), pp. 1-15.
Leatham-Jensen, M.P., et al. (2012) "The streptomycin-treated mouse intestine selects *Escherichia coli envZ* missense mutants that interact with dense and diverse intestinal microbiota", *Infection and Immunity.* 80(5), pp. 1716-1727.

Lederberg, J. and Tatum, E.L. (1946) “Gene recombination in *Escherichia coli*”, *Nature*. 158(4016), 558.

Lennox, E.S. (1955) “Transduction of linked genetic characters of the host by bacteriophage P1”, *Virology.* 1(2), pp. 190-206.

Li, X.T., et al. (2003) “Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*”, *Nucleic Acids Res.* 31(22), pp. 6674-6687.

Li, X.T., et al. (2013) “Bacterial DNA polymerases participate in oligonucleotide recombination”, *Mol Microbiol* . 88(5), pp. 906-920.

Ma, S., et al. (2022) “Reduction of the Bacterial Genome by Transposon-Mediated Random Deletion”, *ACS Synthetic Biology.* 11(2), pp. 668-677.

Macklin, D.N., et al. (2020) “Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation”, *Science.* 369(6502).

Mandel, M. and Higa, A. (1970) “Calcium-dependent bacteriophage DNA infection”, *J Mol Biol.* 53(1), pp. 159-162.

Marinelli, L.J., Hatfull, G.F. and Piuri, M. (2012) “Recombineering: A powerful tool for modification of bacteriophage genomes”, *Bacteriophage.* 2(1), pp. 5-14.

Marinelli, L.J., et al. (2008) “BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes”, *PLoS One.* 3(12):e3957.

Metcalf, W.W., et al. (1996) “Conditionally replicative and conjugative plasmids carrying *lacZα* for cloning, mutagenesis, and allele replacement in bacteria”, *Plasmid.* 35(1), pp. 1-13.

Mizoguchi, H., Mori, H. and Fujio, T. (2007) “*Escherichia coli* minimum genome factory”, *Biotechnol Appl Biochem.* 46(Pt 3), pp. 157-167.

Mizoguchi, H., et al. (2008) “Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome”, *DNA Res.* 15(5), pp. 277-284.

Monk, J.M., et al. (2017) “iML1515, a knowledgebase that computes *Escherichia coli* traits”, *Nat Biotechnol* . 35(10), pp. 904-908.

Murphy, K.C. (1998) “Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*”, *J Bacteriol.* 180(8), pp. 2063-2071.

Murphy, K.C. (2016) “λ Recombination and Recombineering”, *EcoSal Plus.* 7(1).

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(19), pp. 10268-10273.

Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1993) “In vivo cloning of PCR products in *E. coli*”, *Nucleic Acids Res.* 21(22), pp. 5192-5197.

Orth, J.D., et al. (2011) “A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism--2011”, *Mol Syst Biol.* 7(1), p. 535.

Orth, J.D. and Palsson, B.Ø. (2010) “Systematizing the generation of missing metabolic knowledge”, *Biotechnol Bioeng.* 107(3), pp. 403-412.

Orth, J.D. and Palsson, B.Ø (2012) “Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions”, *BMC Systems Biology.* 6(1), 30.

Ostrov, N., et al. (2016) “Design, synthesis, and testing toward a 57-codon genome”, *Science.* 353(6301), pp. 819-822.

Park, J., et al. (2021) “Enhanced genome editing efficiency of CRISPR PLUS: Cas9 chimeric fusion proteins”, *Scientific Reports.* 11(1):16199.

Park, M.K., et al. (2014) “Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences”, *Appl Microbiol Biotechnol.* 98(15), pp. 6701-6713.

Pathania, R., et al. (2009) “Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting”, *Nat Chem Biol.* 5(11), pp. 849-856.
Peters, J.M., et al. (2016) “A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria”, *Cell.* 165(6), pp. 1493-1506.
Pósfai, G., et al. (1999) “Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome”, *Nucleic Acids Res*. 27(22), pp. 4409-4415.
Pósfai, G., et al. (2006) “Emergent properties of reduced-genome *Escherichia coli*”, *Science.* 312(5776), pp. 1044-1046.
Qi, L.S., et al. (2013) “Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression”, *Cell.* 152(5), pp. 1173-1183.
Reed, J.L., et al. (2003) “An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)”, *Genome Biol.* 4(9):R54.
Rousset, F., et al. (2018) “Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors”, *PLoS Genet*. 14(11):e1007749.
Spizizen, J. (1958) “Transformation of Biochemically Deficient Strains of Bacillus Subtilis by Deoxyribonucleate”, *Proc Natl Acad Sci U S A.* 44(10), pp. 1072-1078.
Tong, Y., et al. (2021) “A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing”, *Nature Communications.* 12(1):5206.
Vagner, V., Dervyn, E. and Ehrlich, S.D. (1998) “A vector for systematic gene inactivation in *Bacillus subtilis*”, *Microbiology (Reading).* 144 (Pt 11), pp. 3097-3104.
Wang, H.H. and Church, G.M. (2011) “Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering”, *Methods Enzymol.* 498, pp. 409-426.
Wang, H.H., et al. (2009) “Programming cells by multiplex genome engineering and accelerated evolution”, *Nature.* 460(7257), pp. 894-898.
Wang, J., et al. (2006) “An improved recombineering approach by adding RecA to λ Red recombination”, *Mol Biotechnol.* 32(1), pp. 43-53.
Wang, K., et al. (2016) “Defining synonymous codon compression schemes by genome recoding”, *Nature.* 539(7627), pp. 59-64.
Wang, T., et al. (2018) “Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance”, *Nature Communications*. 9(1):2475.
Wanner, B.L., Teramoto, J. and Mori, H. (2014) “What hath DNA wrought? CRISPR-CAS gene silencing and engineering from bacteria to humans: comment on “Diversity, evolution, and therapeutic applications of small RNAs in prokaryotic and eukaryotic immune systems” by Edwin L. Cooper and Nicola Overstreet”, *Physics of Life Reviews.* 11(1), pp. 144-145; discussion pp. 149-151.
Wannier, T.M., et al. (2020) “Improved bacterial recombineering by parallelized protein discovery”, *Proc Natl Acad Sci U S A.* 117(24), pp. 13689-13698.
Wyman, C. and Kanaar, R. (2006) “DNA double-strand break repair: all's well that ends well”, *Annu Rev Genet.* 40, pp. 363-383.
Yu, B.J., et al. (2002) “Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/*loxP* excision system”, *Nat Biotechnol.* 20(10), pp. 1018-1023.
Yu, D., et al. (2000) “An efficient recombination system for chromosome engineering in Escherichia coli”, *Proc Natl Acad Sci U S A.* 97(11), pp. 5978-5983.

〔受付日　2022. 9. 7〕