



Review

# Past, Present, and Future of Genome Modification in *Escherichia coli*

Hirotada Mori <sup>1,\*</sup>, Masakazu Kataoka <sup>2</sup> and Xi Yang <sup>1</sup>

<sup>1</sup> Innovation Laboratory of Systems Microbiology and Synthetic Biology, Institute of Animal Sciences, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

<sup>2</sup> Department of Environmental Science and Technology, Faculty of Engineering, Shinshu University, Nagano 390-8621, Japan

\* Correspondence: hirotada.mori@gmail.com; Tel.: +86-20-3876-5025

**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* has 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  $\lambda$  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.

**Keywords:** *Escherichia coli* K-12; mutation; homologous recombination; HR; site-specific recombination; genome modification;  $\lambda$  Red; P1 transduction; recombineering



**Citation:** Mori, H.; Kataoka, M.; Yang, X. Past, Present, and Future of Genome Modification in *Escherichia coli*. *Microorganisms* **2022**, *10*, 1835. <https://doi.org/10.3390/microorganisms10091835>

Academic Editors: Tamás Fehér and Antonia P. Sagona

Received: 21 August 2022

Accepted: 5 September 2022

Published: 14 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The discoveries of bacterial conjugation [1] and of generalized transduction [2] have enabled genetic research in *Escherichia coli* K-12. Subsequent genetic investigations of *E. coli* K-12 and its bacteriophages have increased the knowledge of gene structure and function, and have led to the emergence of molecular biology. Although DNA transfer by transformation had been previously described [3] and was shown to occur naturally in *Pneumococcus* [3], *Haemophilus* [4], and *Bacillus subtilis* [5], *E. coli* proved to be recalcitrant. Treatment with  $\text{CaCl}_2$  allowed the transformation (transfection) of *E. coli* with bacteriophage [6] and plasmid DNA [7], 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 [8]. Expression of the  $\lambda$  Red recombinase (*exo*, *bet*, and *gam*) was shown to increase the efficiency of homologous recombination (HR) with linear DNA [9], leading to the use of  $\lambda$  Red recombinase in highly efficient systems for direct modification of chromosomal genes via HR [10,11].

Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas systems participate in acquired immunity in archaea and bacteria [12]. Although these unusual repetitive DNA sequences were first described in 1987 [13], molecular understanding of their function was first determined 25 years later [14], leading to dramatic advances in the technology of genome modification [15].

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 [16]. 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* [17,18], it was not completed by design. The development of whole-cell metabolic models has been steadily progressing [19–23], 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 [24]. Metabolic and whole-cell models have been constructed for *Mycoplasma* [25] and are progressing steadily for *E. coli* [26], 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 [1], which was thought to be sexless and to grow monogamously, and of transduction [2], 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 [27], the detailed molecular structure of which was visualized by cryo-electron microscopy [28]. Results showing that genetic transformation requires extracellular DNA, not protein or RNA [3], provided proof that genes are composed of DNA. Moreover, bacterial species such as *Pneumococcus* [3], *Haemophilus influenzae* [4], and *Bacillus subtilis* [5] were shown to have the ability to take up extracellular DNA.

The discovery of conjugation and transduction in *E. coli* made genetics possible. Table 1 shows genes associated with recombination in *E. coli*, including genes required for HR, site-specific recombination, and transposition. Table 2 summarizes methods used in *E. coli* genome-scale studies. This review focuses on the use of HR for genome modification. HR methods based on  $\lambda$  Red have been expanded for genome-scale functional analyses of *E. coli* and its phages [29,30]. Selected references are in Table 2, with additional references cited in a previous review [31].

Table 1. Recombination related genes.

ECK	Gene	Synonym	Left	Right	Ori	EcoCyc	UniProtKB	Description	Class	Origin
ECK2751	<i>casE</i>	<i>cas6e, cse3, ygcH</i>	2,873,696	2,874,295	C	G7426	Q46897	pre-CRISPR RNA endonuclease	nuclease	
ECK0232	<i>dinB</i>	<i>dinP</i>	247,385	248,440		G6115	Q47155	DNA polymerase IV	polymerase	
ECK0183	<i>dnaE</i>	<i>polC, sdgC</i>	201,613	205,095		EG10238	P10443	DNA polymerase III subunit $\alpha$	polymerase	
ECK0215	<i>dnaQ</i>	<i>mutD</i>	232,554	233,285		EG10243	P03007	DNA polymerase III subunit $\epsilon$	polymerase	
ECK0464	<i>dnaX</i>	<i>dnaZ</i>	488,097	490,028		EG10245	P06710-2	DNA polymerase III subunit $\gamma$	polymerase	
ECK0633	<i>holA</i>		666,579	667,610	C	EG11412	P28630	DNA polymerase III subunit $\delta$	polymerase	
ECK1085	<i>holB</i>		1,151,767	1,152,771		EG11500	P28631	DNA polymerase III subunit $\delta'$	polymerase	
ECK4252	<i>holC</i>		4,474,203	4,474,646	C	EG11413	P28905	DNA polymerase III subunit $\chi$	polymerase	
ECK4363	<i>holD</i>		4,598,169	4,598,582		EG11414	P28632	DNA polymerase III subunit $\psi$	polymerase	
ECK1843	<i>holE</i>		1,919,914	1,920,144		EG11505	P0ABS8	DNA polymerase III subunit $\theta$	polymerase	
ECK4339	<i>hsdM</i>	<i>hsm, hsp</i>	4,571,825	4,573,414	C	EG10458	P08957	Type I restriction enzyme EcoKI methylase subunit	others	
ECK4340	<i>hsdR</i>	<i>hsp, hsr</i>	4,573,615	4,577,127	C	EG10459	P08956	Type I restriction enzyme EcoKI endonuclease subunit	nuclease	
ECK4338	<i>hsdS</i>	<i>hsp, hss, rm</i>	4,570,434	4,571,828	C	EG10460	P05719	Type I restriction enzyme EcoKI specificity subunit	nuclease	
ECK1145	<i>mcrA</i>	<i>rgIA</i>	1,206,351	1,207,184		EG10573	P24200	e14 prophage; Type IV methyl-directed restriction enzyme EcoKMcrA	nuclease	e14 prophage
ECK4336	<i>mcrB</i>	<i>rgIB</i>	4,568,324	4,569,703	C	EG10574	P15005	McrB hexamer	nuclease	
ECK4335	<i>mcrC</i>		4,567,278	4,568,324	C	EG10575	P15006	Type IV methyl-directed restriction enzyme EcoKMcrBC subunit	nuclease	
ECK2152	<i>nfo</i>		2,244,868	2,245,725		EG10651	P0A6C1	endonuclease IV	nuclease	
ECK1144	<i>pinE</i>	<i>pin</i>	1,205,690	1,206,244		EG10737	P03014	Site-specific DNA recombinase of e14 prophage.	recombinase	e14 prophage
ECK1538	<i>pinQ</i>	<i>ydfl</i>	1,628,428	1,629,018		G6819	P77170	Predicted recombinase PinQ	recombinase	Qin prophage
ECK1369	<i>pinR</i>	<i>ynaD</i>	1,427,890	1,428,480	C	G6697	P0ADI0	Predicted site-specific recombinase	recombinase	Rac prophage
ECK3855	<i>polA</i>	<i>resA</i>	4,040,875	4,043,661		EG10746	P00582	DNA polymerase I	polymerase	
ECK0061	<i>polB</i>	<i>dinA</i>	63,429	65,780	C	EG10747	P21189	DNA polymerase II	polymerase	
ECK1348	<i>racC</i>	<i>sbcA</i>	1,412,294	1,412,569	C	EG10813	P15033	Rac prophage protein RacC		Rac prophage
ECK4381	<i>radA</i>	<i>sms</i>	4,616,278	4,617,660		EG11296	P24554	DNA recombination protein		
ECK1345	<i>ralA</i>	<i>ydaB, lar</i>	1,408,539	1,408,733	C	EG11900	P33229	endodeoxyribonuclease toxin RalR	nuclease	Rac prophage
ECK0883	<i>rarA</i>	<i>ycaJ, mgsA</i>	933,999	935,342		EG12690	P0AAZ4	Replication-associated recombination protein A.		
ECK2694	<i>recA</i>	<i>lexB, recH, rnmB, srf, tif, umuB, umuR, zab</i>	2,816,616	2,817,677	C	EG10823	P0A7G6	DNA recombination and repair protein; ssDNA-dependent ATPase; synaptase; ssDNA and dsDNA binding protein; ATP-dependent	recombinase	
ECK2816	<i>recB</i>	<i>ior, rorA</i>	2,946,369	2,949,911	C	EG10824	P08394	homologous DNA strand protein exodeoxyribonuclease V subunit RecB	nuclease	
ECK2818	<i>recC</i>		2,952,968	2,956,336	C	EG10825	P07648	exodeoxyribonuclease V subunit RecC	nuclease	
ECK2815	<i>recD</i>	<i>hopE</i>	2,944,543	2,946,369	C	EG10826	P04993	exodeoxyribonuclease V subunit RecD	nuclease	
ECK1347	<i>recE</i>	<i>rmuB, rac, sbcA</i>	1,409,592	1,412,192	C	EG10827	P15032	exonuclease VIII	nuclease	Rac prophage
ECK3692	<i>recF</i>	<i>uvrF</i>	3,874,057	3,875,130	C	EG10828	P0A7H0	Recombination mediator protein RecF		
ECK3642	<i>recG</i>	<i>radC, spoV</i>	3,819,119	3,821,200		EG10829	P24230	ATP-dependent DNA helicase	helicase	
ECK2887	<i>recJ</i>		3,030,281	3,032,014	C	EG10830	P21893	ssDNA-specific exonuclease	nuclease	
ECK3808	<i>recL</i>	<i>uvrD, uvrE, srjC, dar-2, dda, mutII, pdeB, rad</i>	3,991,892	3,994,054		EG11064	P03018	DNA helicase II	helicase	

Table 1. Cont.

ECK	Gene	Synonym	Left	Right	Ori	EcoCyc	UniProtKB	Description	Class	Origin
ECK2612	<i>recN</i>	<i>radB</i>	2,745,703	2,747,364		EG10831	P05824	DNA repair protein RecN		
ECK2563	<i>recO</i>		2,695,649	2,696,377	C	EG10832	P0A7H3	Recombination mediator protein RecO		
ECK3816	<i>recQ</i>		3,999,773	4,001,602		EG10833	P15043	ATP-dependent DNA helicase RecQ	helicase	
ECK0466	<i>recR</i>		490,410	491,015		EG10834	P0A7H6	Recombination mediator protein RecR		
ECK1346	<i>recT</i>		1,408,790	1,409,599	C	EG11899	P33228	recombinase RecT	recombinase	Rac prophage
ECK2693	<i>recX</i>	<i>oraA</i>	2,816,047	2,816,547	C	EG12080	P33596	RecA inhibitor RecX		
ECK3826	<i>rmuC</i>	<i>dinK, sosB, yigN</i>	4,011,242	4,012,669		EG11472	P0AG71	DNA recombination protein		
ECK3398	<i>rpnA</i>	<i>yhgA</i>	3,537,075	3,537,953		EG11750	P31667	Recombination-promoting nuclease RpnA	transposase	transposon
ECK2299	<i>rpnB</i>	<i>yfcI</i>	2,416,677	2,417,567	C	G7197	P77768	Recombination-promoting nuclease RpnB	transposase	transposon
ECK0131	<i>rpnC</i>	<i>yadD</i>	143,455	144,357		EG11748	P31665	Recombination-promoting nuclease RpnC	transposase	transposon
ECK4329	<i>rpnD</i>	<i>yjiP, yjiQ</i>	4,559,364	4,560,284		G7934	P0DP21, P0DP22	Recombination-promoting nuclease RpnD	transposase	transposon
ECK2236	<i>rpnE</i>	<i>yfaD</i>	2,350,932	2,351,831		EG12323	P37014	Inactive recombination-promoting nuclease-like protein RpnE	transposase	transposon
ECK1862	<i>ruvA</i>		1,940,171	1,940,782	C	RUVA	P0A809	Holliday junction branch migration complex subunit RuvA		
ECK1861	<i>ruvB</i>		1,939,152	1,940,162	C	RUVB	P0A812	Holliday junction branch migration complex subunit RuvB		
ECK1864	<i>ruvC</i>		1,941,661	1,942,182	C	EG10925	P0A814	Crossover junction endodeoxyribonuclease RuvC	nuclease	
ECK2005	<i>sbcB</i>	<i>exoI, cpeA, xonA</i>	2,076,786	2,078,213		EG10926	P04995	exodeoxyribonuclease I	nuclease	
ECK0391	<i>sbcC</i>	<i>rmuA</i>	408,612	411,758	C	EG10927	P13458	ATP-dependent structure-specific DNA nuclease—SbcC subunit	nuclease	
ECK0392	<i>sbcD</i>	<i>yajA</i>	411,755	412,957	C	EG11094	P0AG76	ATP-dependent structure-specific DNA nuclease—SbcD subunit	nuclease	
ECK4051	<i>ssb</i>	<i>exrB, lexC</i>	4,264,602	4,265,138		EG10976	P0AGE0	ssDNA-binding protein		
ECK3833	<i>tatD</i>	<i>yigX, yigW, mttC</i>	4,017,463	4,018,245		EG11481	P27859	3' → 5' ssDNA/RNA exonuclease TatD (exonuclease XI)	nuclease	
ECK1172	<i>umuC</i>	<i>uvm</i>	1,227,191	1,228,459		EG11056	P04152	DNA polymerase V catalytic protein	polymerase	
ECK1171	<i>umuD</i>		1,226,772	1,227,191		EG11057	P0AG11	DNA polymerase V protein UmuD	polymerase	
ECK3806	<i>xerC</i>		3,990,196	3,991,092		EG11069	P0A8P6	Site-specific tyrosine recombinase	recombinase	
ECK2889	<i>xerD</i>	<i>xprB</i>	3,032,755	3,033,651	C	EG11071	P0A8P8	Site-specific recombinase	recombinase	
ECK2505	<i>xseA</i>	<i>xse</i>	2,628,140	2,629,510		EG11072	P04994	exodeoxyribonuclease VII subunit XseA	nuclease	
ECK0416	<i>xseB</i>	<i>yajE</i>	437,106	437,348	C	EG11098	P0A8G9	exodeoxyribonuclease VII subunit XseB	nuclease	
ECK1747	<i>xthA</i>	<i>xth</i>	1,827,234	1,828,040		EG11073	P09030	exodeoxyribonuclease III	nuclease	
ECK0535	<i>ybcK</i>		564,907	566,433		G6300	P77698	Predicted recombinase YbcK	recombinase	DLP12 prophage
ECK2639	<i>yfjX</i>		2,769,827	2,770,285		G7378	P52139	Predicted antirestriction protein YfjX	nuclease	CP4-57 prophage
ECK2793	<i>ygdG</i>	<i>xni, exo</i>	2,924,963	2,925,718		EG12372	P38506	Ssb-binding protein; misidentified as ExoIX; synthetic lethal with polA; no nuclease activity detected; flap endonuclease family protein	nuclease	

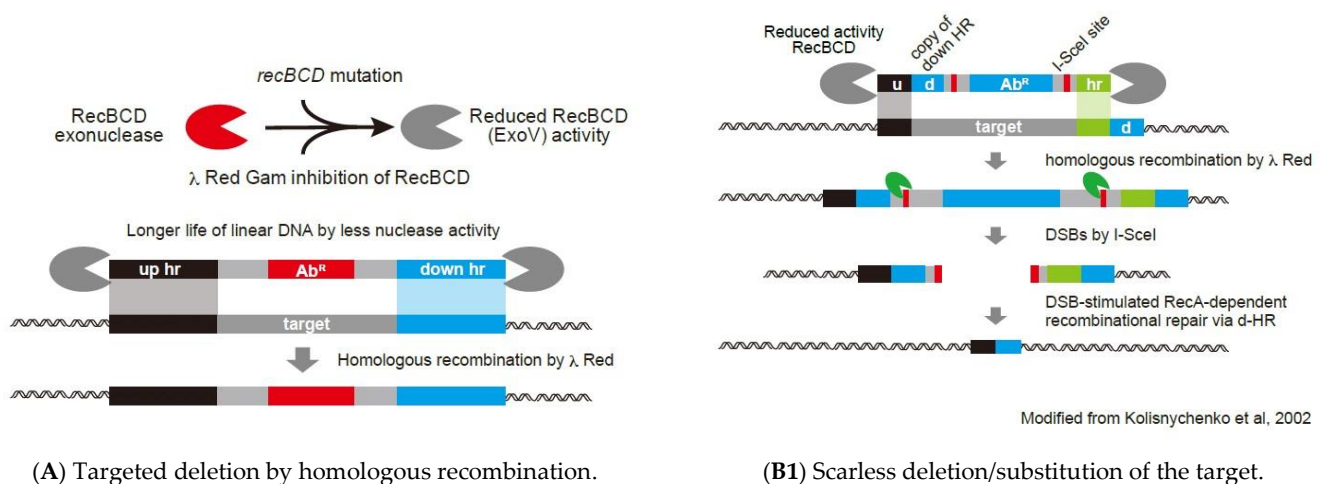
ECK: ECK ID defined by Riley et al., PubMed: 16397293; Gene: gene name; Synonym: alternate name; (Left, Right): coordinates on BW25113 genome; Ori: orientation on BW25113 genome (blank: clockwise, C: counterclockwise); EcoCyc: Entry ID PubMed: 30406744, EG IDs PMC3531124; UniProtKB: Entry ID PubMed ID: 18287689; Description: annotation of gene and gene product; Class: protein functional classification; Origin: ancestral origin.

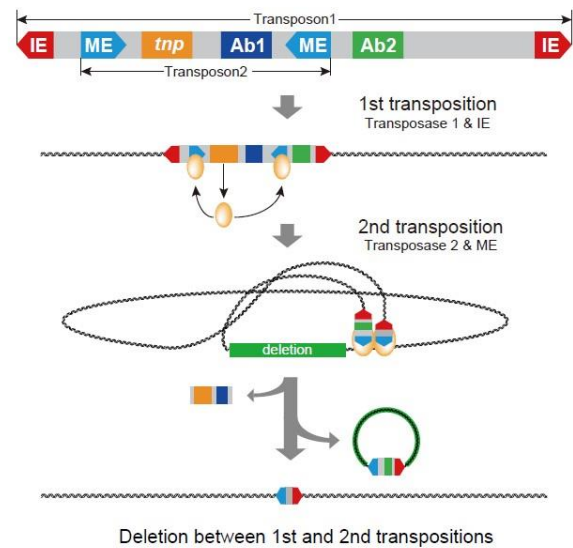
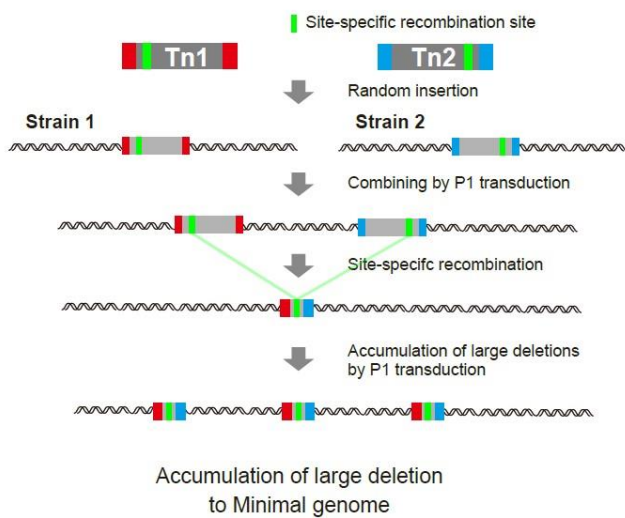
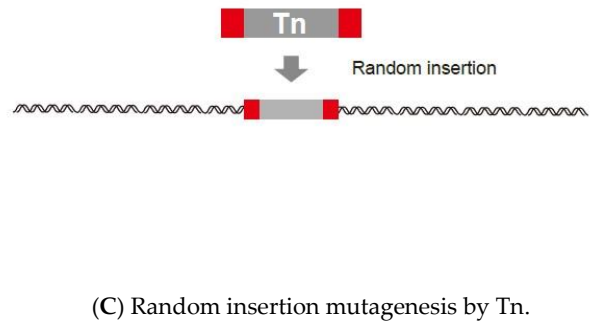
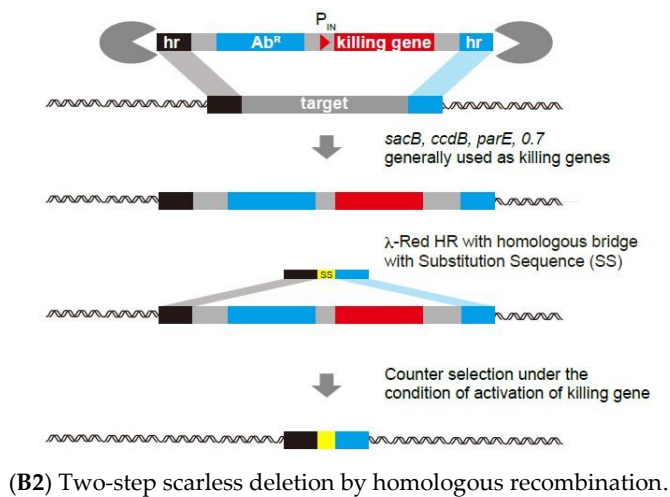
**Table 2.** Classification of genome modification methods in *E. coli*.

Recombination Category	System	Factors	Reference
Homologous	<i>recBCD</i>	Exonuclease VIII	[32]
	$\lambda$ Red recombinase	Red recombinase	[30]
Site-specific	<i>Int-att</i>	Integrase, <i>att</i> site	[33]
	<i>Cre-lox</i>	Cre site-specific recombinase	[34]
	FLP-FRT	Flippase, FRT-site	[35]
Mobile element	Mu phage		[36]
	Tn5		[37]
	Tn7		[38]
	Tn10		[39]
	Mariner Tn		[40]
	Group II intron		[41]

## 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 [42]. This method was modified during the development of an in vivo cloning (iVEC) protocol, which requires *xthA* (exonuclease III) and is independent of RecA and RecET [43]. Table 1 lists genes related to HR in *E. coli*. Mutations improving HR have been identified in the *recB*, *recC*, *recD*, *sbcA*, *sbcB*, *sbcC*, and *hsdR* genes. Moreover, findings showing that  $\lambda$  Red markedly increased the efficiency of HR in *E. coli* [9] and led to the development of tools for genetic modification and their use in *E. coli* and other microorganisms [10,11,44]. The method illustrated in Figure 1A requires only 35 bp of homology for efficient recombination and was adopted by the Japan genome project to construct the Keio collection of single-gene *E. coli* deletions [10,45]. Greater understanding of the molecular mechanism of  $\lambda$  Red HR led to many improvements for its use in *E. coli*, phages, and other bacteria [46–50]. Targeted replacement by HR requires suppression or inhibition of the RecBCD exonuclease. Historically, the mutants used had reduced RecBCD activity and a mutation in *sbcA* for activation of the *recET* pathway, which together leads to HR by the RecET recombinase. Major breakthroughs came from utilizing  $\lambda$  Red, which encodes the *gam*, *bet*, and *exo* genes, for recombination of PCR products with short homologies flanking the chromosomal target [10,11].

**Figure 1.** Cont.



Modified from Goryshin et al, 2003

(C1) Two random insertion mutageneses by Tn and site-specific recombination.

(C2) Random insertion and deletion by Tn-in-Tn.

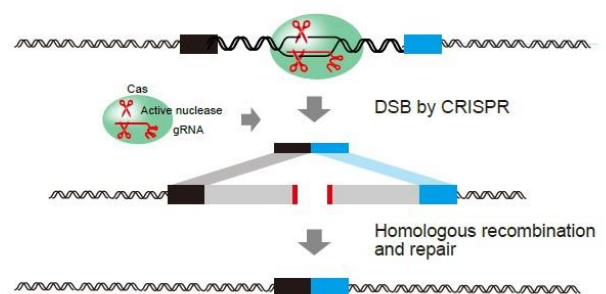
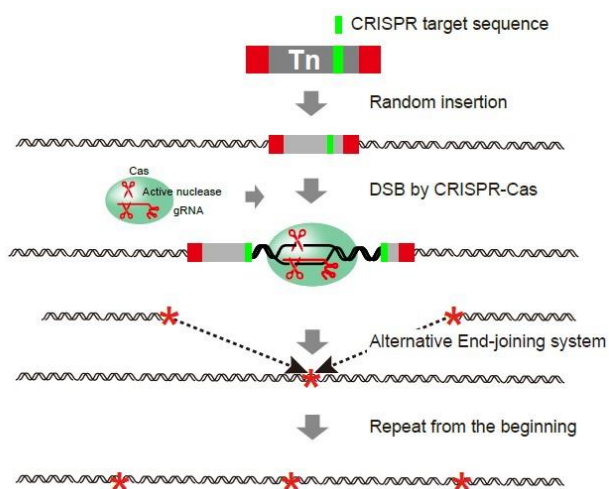
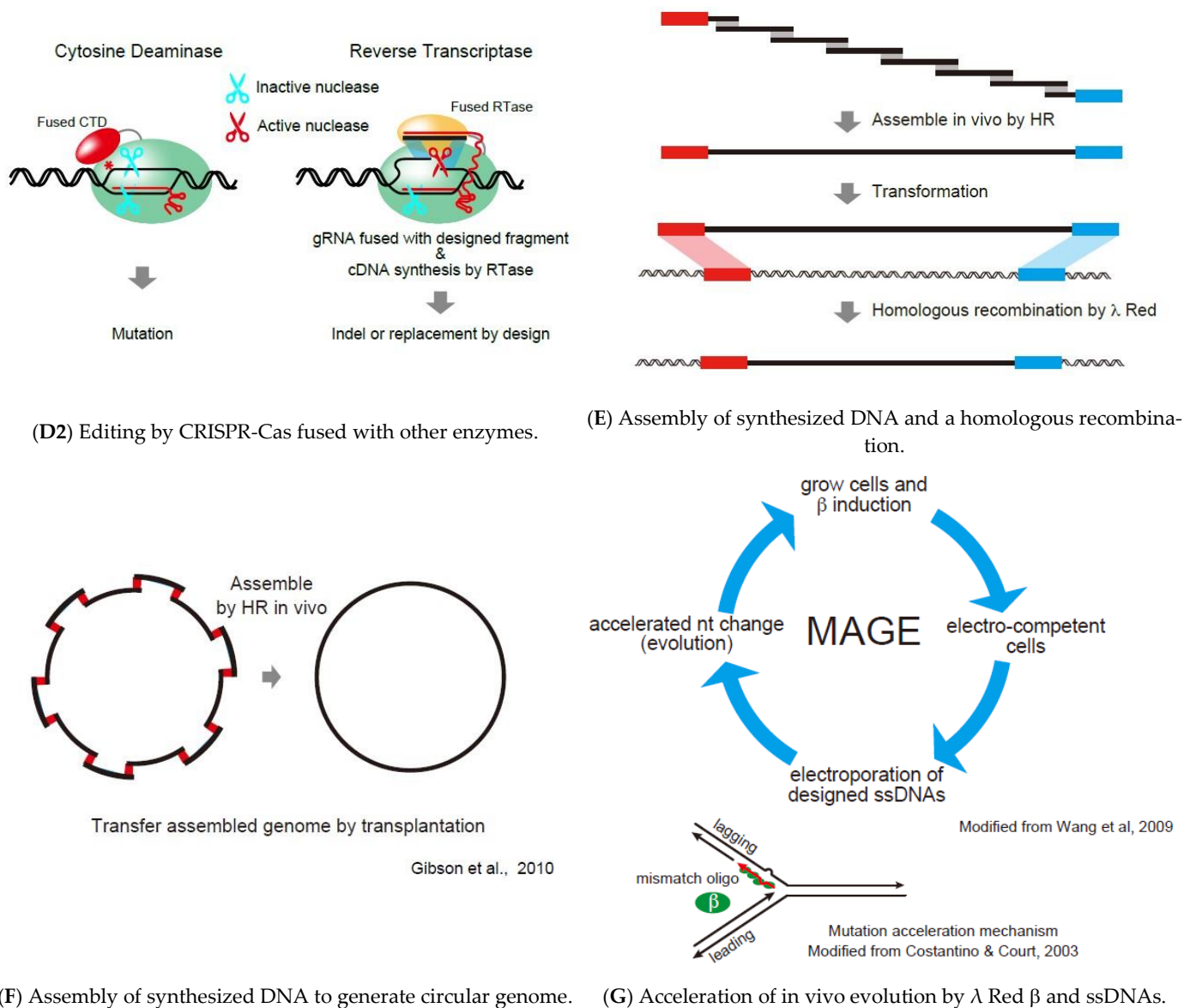


Figure 1. Cont.



**Figure 1.** Schematic illustration of genome-scale modification methods. (A) Two methods have been used to block RecD exonuclease: (1) using *recBC* mutations and (2)  $\lambda$  Red Gam synthesis. Cells are transformed with linear double-stranded (ds) DNA encoding an antibiotic resistance ( $Ab^R$ ) cassette and ends at homology regions (hr) of upstream (up) and downstream (down) regions of the target. (B1) Scarless deletion using I-SceI nuclease. The drug-resistant fragment flanked by I-SceI restriction enzyme sites is amplified with up and next to the downside homology regions (u and hr) and introduced into the genome by  $\lambda$  Red homologous recombination [16]. 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 direct repeats. This figure is modified from Kolisnychenko et al. [16] (B2) 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* [51], is amplified with 36 to 40 nt homology region (hr) to target. The amplified fragment is then transformed into a strain expressing  $\lambda$  Red to insert into the genome. Double-stranded “Substitution Sequence (SS)” with flanking hrs is transformed into the  $Ab$ -resistant fragment integrated strain expressing  $\lambda$  Red. The transformants are selected in the presence of L-rhamnose, preferably with L-rhamnose as the sole carbon source. (C) Random insertion mutagenesis by Tn. (C1) 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, with the location of insertion on the genome 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. (C2) A complex with Tnp that recognizes IE is introduced into the cell to obtain the first random insertion mutation. Synthetic induction of Tnp recognizing the internal ME is then performed to obtain a transition mutation; the direction of the second transition results in a deletion between two different Tn insertion sites. This figure is modified from Goryshin et al. [52] (C3) 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. (D1) DSB was induced by CRISPR-Cas, with DNA fragments transformed by bridging homologous regions at both ends of each double-strand break, resulting in genome repair and yielding to the circular genome. (D2) 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 [53]. 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 77. (E) Introduction of synthetic DNA fragments into cells, generally yeast cells, resulting in assembly of the fragments by in vivo homologous recombination. After the assembly, the assembled fragment is collected and transformed into *E. coli* cell and replacement of the target region in the  $\lambda$  Red-induced strain by homologous recombination. (F) Assembly of the synthetic DNA fragments in the cell, followed by circularization to reconstruct the genome. The synthetic genome was subsequently transferred to bacterial cells by cell fusion [17,18]. This figure is summarized from Gibson et al. [17] (G) Introduction of an ssDNA about 90 bp in length to be mutated in the cell via the induction of  $\lambda$  Red  $\beta$  protein, which promoted the introduction of mutations on the lagging strand during DNA replication and accelerated the introduction of mutations throughout the genome. This figure is modified from Wang et al. [54] and Costantino and Court [55].

The original method employs an antibiotic resistance fragment carrying a resistance gene flanked by FLP recognition target (FRT) sites, which allow elimination of the resistance cassette with an FLP expression plasmid, leaving behind “scar” sequences. A strategy was subsequently developed for constructing “markerless” (scarless) gene replacements (deletions or substitutions) [56]. 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  $\lambda$  Red HR to create scarless deletions (Figure 1(B1)).

Figure 1(B2) 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* [57–59], *tetA* [58–60], and *tolC* [61–63]. The leakiness of these genes, however, often prevents their use as general tools in genomics studies, requiring the implementation of both positive and negative selection with a dual *tetA-sacB* cassette [58] or an optimization protocol to interfere with escape [63]. Two-step protocols in Figure 1(B2) have provided powerful tools for recombineering *E. coli* and related bacteria with single-stranded (ss) and double-stranded (ds) DNA [31,64–73].

### 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 (Figure 1C) [74]. 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 repeating of the entire process (Figure 1(C,C1)).

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 (Figure 1(C2)) [52]. 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 [52]. In addition to Tn elements for making simple deletions, a Tn element carrying a conditional replication origin was constructed, allowing recovery of the deleted fragment as a plasmid. Therefore, this system may allow deletion of essential genes. To date, a comprehensive set of single-gene deletion mutants has been constructed [45,75], but it would be advantageous to simultaneously obtain clones of these fragments. Of the 15 cases examined, 11 were free of essential genes because growth was observed even when the plasmid was removed, whereas the other four were no longer viable after the plasmid was removed.

### 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 [76,77]. 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 (Figure 1(B1)) [16]. This method, which was based on the accumulation of scarless deletions by HR and DSB, allowed the 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 24 of 44 transposon regions, were deleted [16]. 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.

Another study first compiled a list of predicted essential genes, followed by the use of an HR method to delete regions between these genes [53]. Using  $\lambda$  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  $\lambda$  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. Figure 1(B1) shows how targeting the I-SceI site to a specific location generates a DSB that leads to the formation of the designed deletion. Figure 1(C3) shows how random Tn mutagenesis can be used to generate DSBs and nearly random deletions by CRISPR-Cas [78].

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) [79]. NHEJ, however, is usually not possible in bacteria due to a 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 [80,81] by LigA (Figure 1(C3)). Combining CRISPR-Cas with  $\lambda$  Red HR permits fashioning scarless deletions, insertions, or substitutions by design, limited only by occurrences of PAM sequences (Figure 1(D1)) [82–86].

The 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 [87–92], and various point mutations and small insertions/deletions by fusion of the reverse transcriptase to a single active Cas (nickase) subunit [93–96] (Figure 1(D2)). This subject has previously been reviewed [97].

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 [98–102].

CRISPRi was used to analyze a group of essential genes in *B. subtilis*, although this approach did not focus directly on genome modification [99]. 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 [100], thus identifying *E. coli* essential genes and genes essential for phage  $\lambda$  growth [101]. 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 [102]. Essentiality was tested with a pooled library, with the results evaluated by determining the relative change in read count by next-generation sequencing (NGS). The rules for effective gRNA design have been summarized [100,102].

### 2.4.2. Acceleration of Evolution under the Constraint of Mutation Direction by Oligo DNA

The 1990s marked a turning point in biological research, beginning with the automation of DNA sequencing and the development of technologies for the production of large amounts of data [103]. At the start of the 21st century, the pace of development of many technological innovations continued to increase. Genome modification using HR has been based on methods of design and recombination. However, 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. The relatively few analyses of genetic interactions have made it difficult to design methods that take these relationships into account.

Therefore, a method was devised to accelerate the evolution of mutagenesis by adding the constraint of viability and utilizing the principle of HR of  $\lambda$  Red, while restricting sequences using synthetic DNA. This Multiplex Automatable Genome Engineering (MAGE) method uses the  $\beta$  protein of  $\lambda$  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 (Figure 1G) [54]. Although MAGE was originally performed robotically [54], it can also be performed manually. MAGE has provided a powerful tool for genome-wide codon replacement [104], metabolic engineering [105], and other biological functions (Figure 1G) [106].  $\lambda$  Red HR is increased in mismatch repair mutants and shows strand bias [55,107]. MAGE technology has been further improved by using MAGE and  $\lambda$  Red HR to stimulate the evolution of host *E. coli* primase and helicase [108], which control the length of Okazaki lagging-strand fragments [109].

### 2.5. Impact of Plasmid Clones on Genome-Scale Analysis and Resource Construction

The construction of plasmid libraries encoding genes is important for understanding gene function. Resources required before starting a genome project include plasmid clones and mutants of the target gene, making their construction the first step. Once construction is underway and a clear blueprint of the target organism has been obtained, it is desirable to have systems to analyze the organism in its entirety. High-density DNA membrane filters and microarrays for *E. coli* have made global analysis possible [110–113]. Once all the gene clones and mutant libraries have been established, complete global comparative analysis of the entire gene set under the same conditions is feasible.

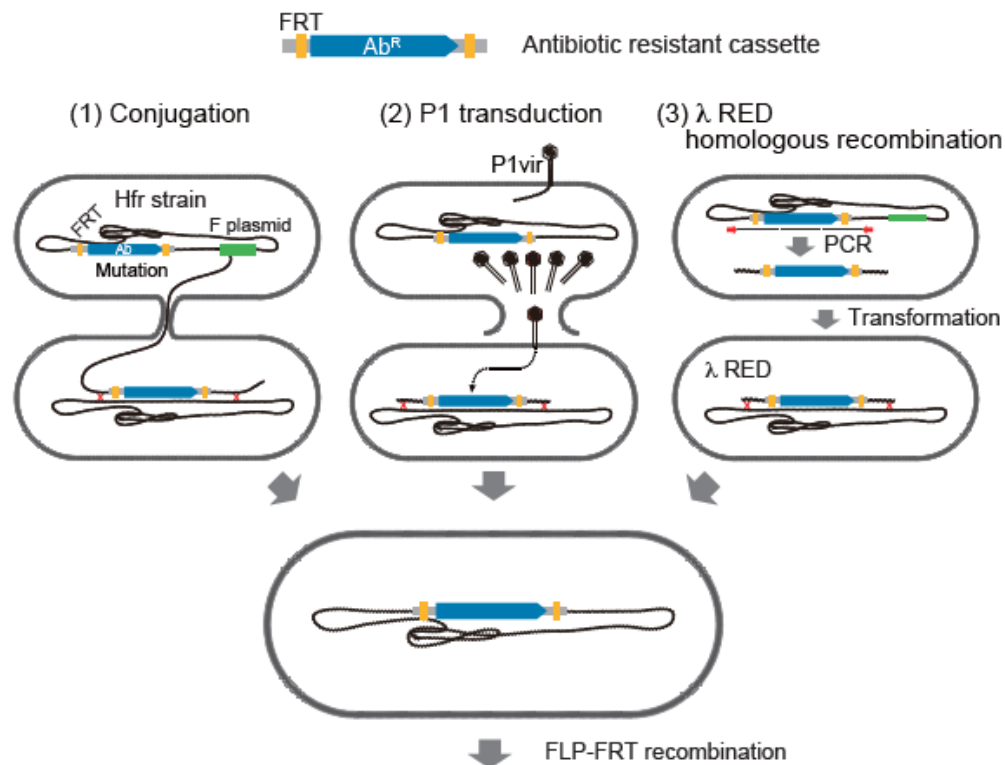
These requirements have given rise to global resource-building activities, and the construction of plasmid clone and mutant strain libraries [45,114–116], including promoter libraries [117]. The plasmid clone library was started using the restriction enzymes cloning method rather than recombination. Development of many of the methods described below allowed editing of the *E. coli* genome. Cloning of individual genes allowed the creation of synthetic metabolic pathways, fine-tuning of promoters, and the alteration of codons, after which the genes were recombined into the genome. These protocols not only optimize synthesis of useful products but modify *E. coli* biosynthetic and catabolic pathways. For example, *E. coli* genes for glycolysis are scattered throughout the genome. Modulating expression of these genes requires bringing them together [118] by using Ordered Gene Assembly in *Bacillus subtilis* (OGAB) [119] to create a series of plasmids with genes of the *E. coli* glycolytic pathway clustered in a variety of arrangements.

The cloning of synthetic DNA fragments into plasmids for growth in *E. coli*, other bacteria, and yeast is a key step in the chemical synthesis of genomes [120]. Methods using HR have been developed, including commercially available In-Fusion [121], SLiCE [122], and Gibson Assembly [123] methods. These technologies, which are often available as kits, have made genomic manipulations easier and faster, and have contributed to genome modification as the first step for cloning target genes.

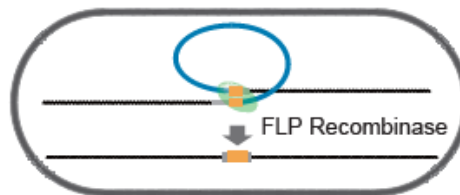
### 2.6. Accumulation of Gene Modifications on One Genome

Once each of the regional genomic modification has been created, it is often important to transfer mutations resulting in antibiotic resistance to other strains, thereby aiding the functional analysis or accumulation of mutations on one genome. Three convenient methods of mutation transfer are currently available to transfer mutations into the target host strain: (1) conjugation, (2) P1 transduction, and (3) PCR fragmentation (Figure 2). All of these methods create mutants by HR.

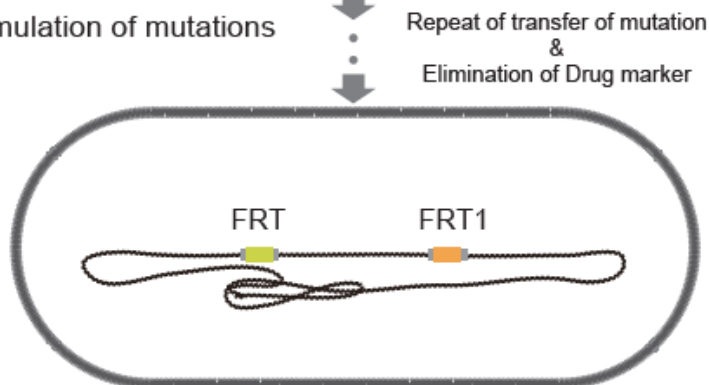
(A) Three standard methods to transfer mutation to another strains



(B) Elimination of drug resistance marker



(C) Accumulation of mutations



**Figure 2.** Common methods to transfer mutations to another strain using Keio collection mutants. (A) Methods of transferring mutations into the target host strain. (1) Conjugation, consisting of the recombination of an oriT onto the chromosome and use of an F plasmid to provide conjugative transfer factors in trans. (2) P1 transduction, using phage P1 lysate prepared on the mutant to infect new strains. (3) λ Red homologous recombination. (B) Elimination of drug-resistant marker by Flp-FRT site-specific recombination. After elimination, one copy of a 34 bp FRT scar remained. (C) Repetition of steps (A,B), resulting in the accumulation of mutations with FRT copies as a scar.

Historically, most mutations are detected by selection markers, with one example being the antibiotic-resistant template F<sub>1</sub>p-FRT (Figure 2) [10]. Other markers can also be used if they can be selected. In addition, site-specific recombination between FRTs can be significantly altered by changing the spacer sequences inside the FRTs [124]. By contrast, CRISPR enables markerless accumulation of mutations using PCR fragments (Figure 1(D1)).

### 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 [125,126], 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 [127].

##### 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 [74]. 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, with 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 markers 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 a second transposon transposition from within the transposon once inserted into the genome by adding another set of transposon recombination sites. Repeating these deletions resulted in a minimal genome [52]. 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. In addition to designing, a Tn carrying a condition-sensitive replication origin was designed to rescue as a plasmid the fragment to be deleted in the deletion strain. This system may allow the introduction of deletions into essential gene regions. A comprehensive single-gene deletion strain has been constructed, and the essential genes, as shown by single deletion, have been identified [45,75]. The essentiality of large-scale deletions can be analyzed by simultaneously obtaining clones of these fragments. Of the 15 cases analyzed, 11 lacked the essential genes because growth was observed even when the plasmid was removed. By contrast, the other four cases were no longer viable after removal of the plasmid. One case contained an essential gene, whereas the others contained deletions of short regions that did not contain any ORFs, making their situation unstable. The number of copies of the plasmid, however, may be a factor.

##### 3.1.2. Large Scarless Deletion by HR

The reduced genome lacking K-islands [16] has been further improved [76]. 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 [76]. 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, resulting in the selection of a set of candidate genes, comprising about 20% of the genome, for deletion.

The deletion method [16] 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 a practical, reduced-genome *E. coli*. MDS69, an improved *E. coli* strain with additional deletions, which is currently available commercially from Scarab Genomics (<https://www.scarabgenomics.com/products/clean-genome-e-coli/>), accessed on 4 September 2022).

A method of scarless deletion of a region of non-essential genes between essential genes involved the performance of two HR events (Figure 1(B2)) [53]. 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 the growth rate was inversely associated with the size of the deleted region [53]. Deletion also altered cell morphology, with changes in cell length and width and in nucleoid organization. Attempts to combine these large deletions showed that some could not be combined (Kato, J., personal communication), perhaps because 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, Machida, Japan.

Efforts have been made to develop bacteria with beneficial genomes for the production of materials, especially with industrial applications, without the inhibition of cell growth. One strain, MGF-01, was generated by deleting 1.03 Mb from 53 regions using P1 [128]. These deletions increased glucose consumption 1.44-fold and acetate accumulation 0.09-fold, confirming the efficacy of this method [128].

The strain MS56 has a genome reduced by 23% [129]. 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 [129]. Evaluation of the scarless HR deletion method [16,76] 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. Large-Scale Genome Modification by Synthesis and Recombineering

#### 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  $\lambda$  Red recombination has shown that ssDNA predominantly introduces mutations into the replicating lagging strand through the activity of the beta protein alone [46]. 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  $\beta$  proteins

(Figure 1G) [54]. A homemade automated facility was also developed to automate this cycle, resulting in the mutation in a single step of 24 genes in the 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway scattered throughout the genome in a single step. Evaluation of the resulting mutations for optimization of DXP synthesis showed that, for 20 of the 24 genes, 90-nucleotide long ssDNAs were designed to optimize the ribosome binding site, increasing their levels of expression. For the four other genes, ssDNAs were designed to introduce nonsense codons, making them non-functional in the MAGE method. The running of 5–35 MAGE cycles resulted in  $\sim 10^5$  mutant strains and increased the production of the target product, isoprenoid lycopene, up to 5-fold within 3 days [54].

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 [104]. Because these 314 genes are scattered throughout the genome, the genome was divided into 32 regions, with the MAGE cycle 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 [108].

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 [130]. These results showed that genome-wide codon replacement is feasible and that codons can be replaced using MAGE, with minimal or no effect on cell growth. Providing artificially modified organisms with a genetic code that does not exist in the natural world 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 (Figure 1E). 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 [131]. In the other 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 [132]. This method resulted in the construction of a partially 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  $\lambda$  Red recombinase. This step was repeated to construct the entire recoded genome [133]. 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.

Only two assembled fragments significantly affected cell growth [131]. The responsible genes were identified and individually modified to overcome this drawback. One gene was found to be insufficiently expressed in the fatty acid biosynthesis operon *rpmF-accC*, an insufficiency circumvented by improving the promoter in the duplicated region using the MAGE method [131].

### 3.3. Resources for Genome-Scale Functional Analysis towards Genome Design

The genomic sequences of the *E. coli* K-12 strains, MG1655 and W3110, were compared to more accurately determine their sequences [134] and annotation [135]. Experimental resources were also designed to analyze *E. coli* genes globally. The initial focus was on construction of an ORF plasmid clone library with PCR-amplified genes from predicted ORF regions within the genome (Figure 3) [115]. Full-length cDNA microarrays were developed and shared with the research community to launch OMICS research [110,111]. Efforts were also made to construct gene deletions using the Kohara  $\lambda$  phage ordered clone library of the *E. coli* K-12 genome [136]. However, HR with synthesizable base length was problematic at that time. Immediately after  $\lambda$  Red HR was first used to disrupt genes on the *E. coli* chromosome with PCR products [10], a comprehensive library of *E. coli* single-gene deletion mutants, the Keio collection [45], was constructed, with this library made freely available to the research community as an open resource.

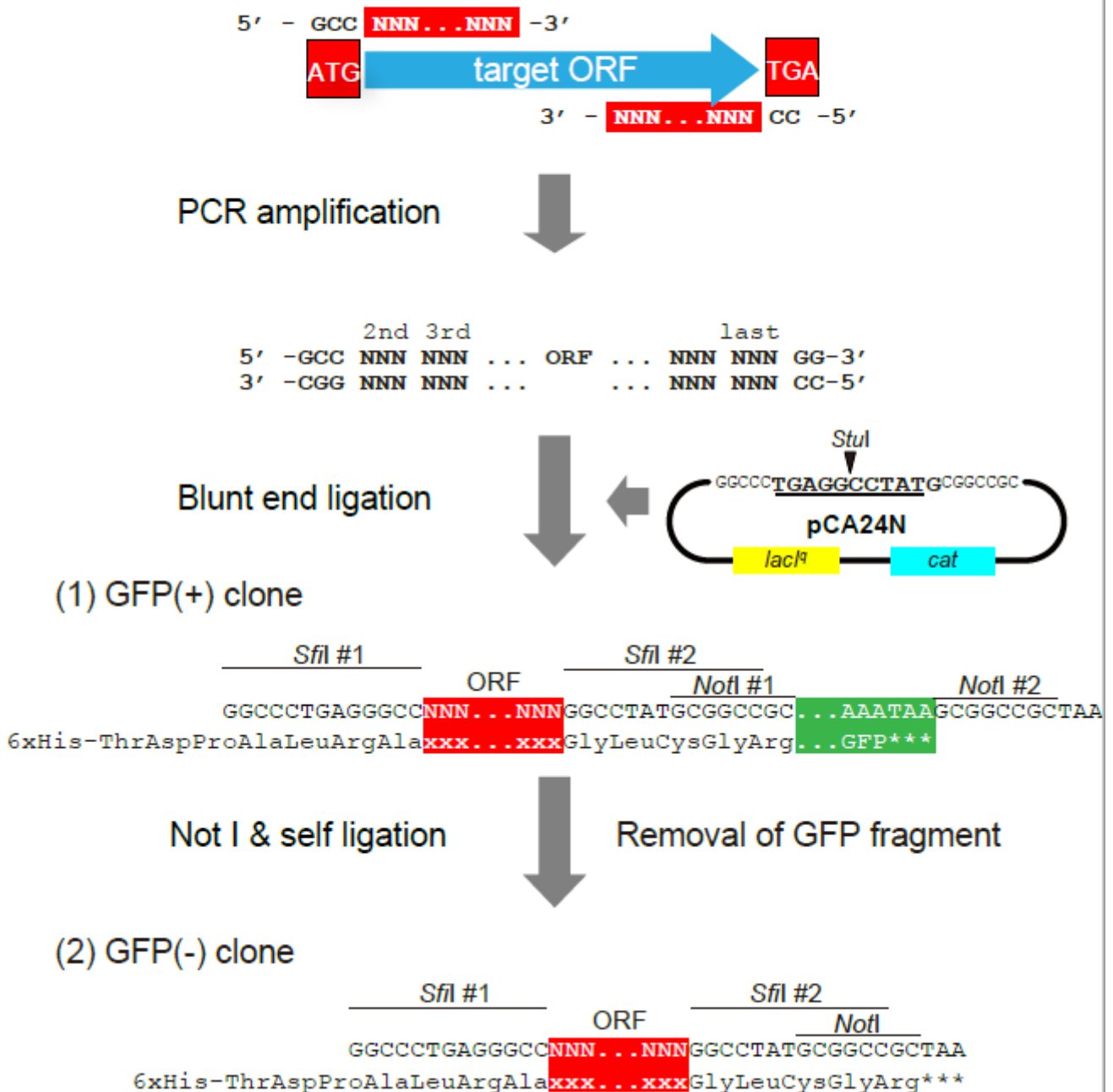
The Keio collection was used to examine the effects of central metabolic pathway gene deletions on transcription, translation, and intracellular metabolites levels [137]. Quantitative fluctuations of metabolites were small and remained stable compared with transcription alterations. Robustness was further addressed by analyzing comprehensive synthetic lethality through double gene knockout. This was accomplished by developing a second library of single-gene deletion mutants, the ASKA barcode deletion collection, which has not yet been completed. In addition to changing the antibiotic resistance gene, this library carried a 20 nt random sequence as a barcode (Figure 4) [138].

To date, two independent representatives of about 3000 genes have been successfully isolated, confirmed, and stored. Each of the independent representatives of the same gene deletion has a unique barcode. A method was also developed to efficiently produce double gene deletion strains by combining two types of deletion strains and measuring their growth, thereby enabling analyses of genetic interactions [139,140].

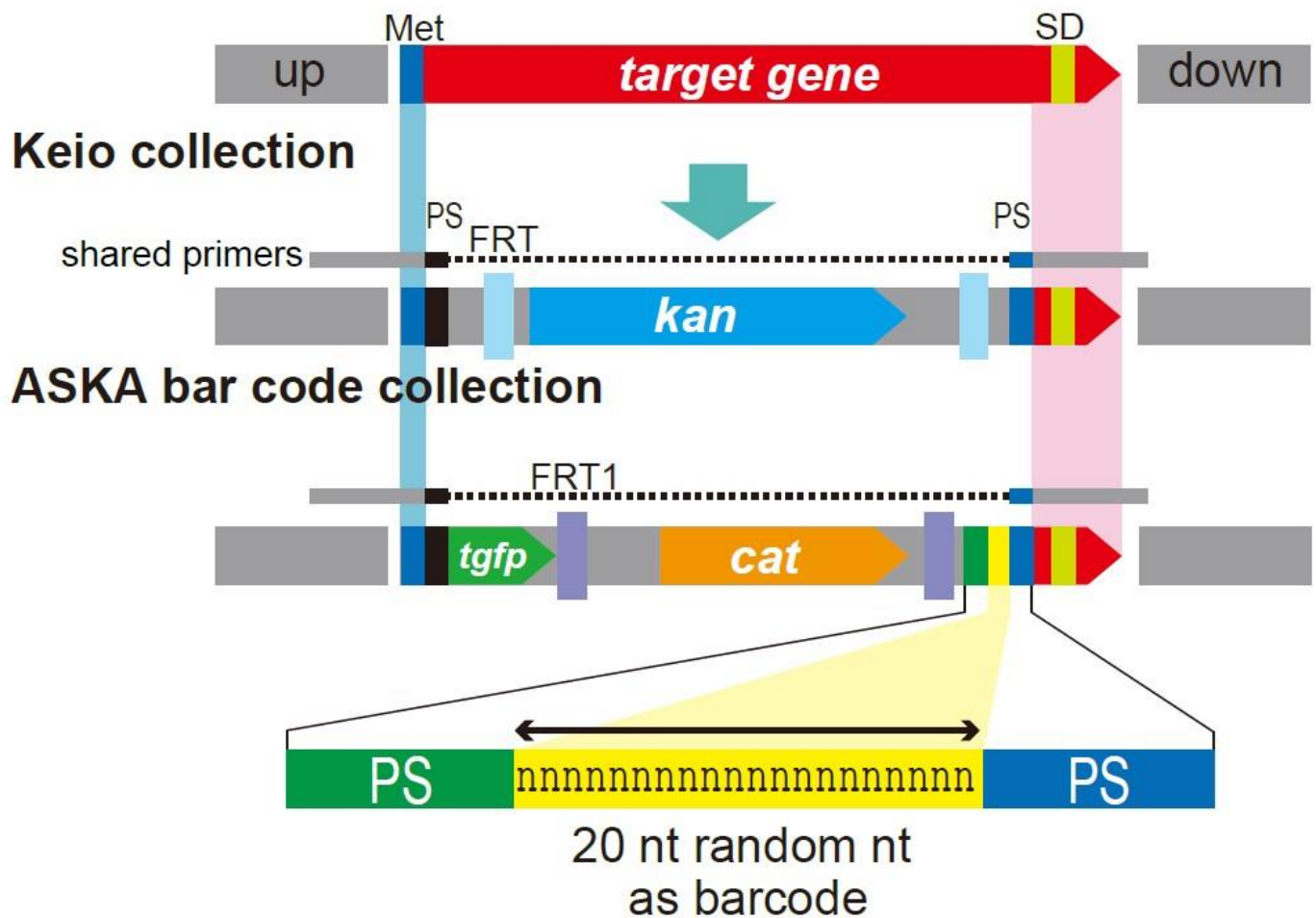
Catalytically inactive Cas resulting from gene mutation has allowed repurposing CRISPR as an RNA-guided platform that can specifically interfere with transcription elongation, RNA polymerase binding, or transcription factor binding (CRISPR interference; CRISPRi) by using a single guide RNA (sgRNA) chimera [98–102]. CRISPRi has been utilized to analyze a group of essential genes in *B. subtilis*, although this approach did not focus directly on genome modification but on knockdown of gene expression [99]. *E. coli* was subjected to CRISPRi screening by synthesizing a library of 92,000 sgRNA sequences covering the entire genome randomly, with PAM sequences as the only constraint [100]. This enabled identification of *E. coli* essential genes and genes essential for phage  $\lambda$  growth [101]. In a separate study, 60,000 sgRNAs were evaluated for testing essentiality while assessing the design of sgRNAs for all genes, including non-coding RNA genes [102]. Essentially was tested using a pooled library, with the results evaluated by determining the relative change in read count by NGS. The rules for effective gRNA design have been described [100,102].



# ASKA plasmid clone library



**Figure 3.** Construction of the ASKA plasmid clone library. The sequence corresponding to all of the amino acids in the coding region, except for the first Met codon, were PCR amplified, with additional GCC and CC nucleotides at the N- and C-termini, respectively. Translation from termination codon is shown by \*\*\*. The amplified fragments were subsequently cloned into the StuI site of pCA24N. Only clones with predicted orientation could generate fluorescence from the eGFP peptide. After the structures of the cloned plasmids were validated, the plasmids were cut with NotI and self-ligated to eliminate eGFP. The structures are of (1) a fusion type with eGFP and (2) a non-fusion type plasmid clone.



**Figure 4.** Structure of Keio single-gene deletion and ASKA barcode deletion collections. The coding regions of both deletion strains, except for the initiation codon and the codons encoding the six amino acids at the C-terminal, were replaced by drug resistance fragments. Site-directed recombination of FLP-FRT removed the drug resistance region. After removal from the Keio collection, the initiation codon and the codons encoding the C-terminal six-amino-acid region of the target gene were fused in frame with codons from the FRT site to suppress the polar effect of the downstream gene 35. The shared-primers contained the initiation codon (blue) and 50 bases upstream and downstream (gray), including the six C-terminal codons and the terminal codon (red) as chromosomal homologous regions. The black and blue sequences represent primers amplifying the template plasmid with the drug resistance gene and FRT sites both of Keio and ASKA barcode collections, respectively. For introducing 20 nt length random sequence as a barcode, PS for amplifying the resistant fragment (green), 20 nt random sequence (yellow), and PS of shared-primer (blue) was synthesized and amplified with PS (black) to prepare the template fragment. The resistant fragment with barcode was then amplified by shared-primers. The amplified drug resistance fragments for the Keio collection and ASKA barcode collection were used to transform  $\lambda$  Red-induced strains to generate deletion strains by homologous recombination. The barcode deletion strains are available for about 3000 genes.

### 3.4. 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 [141]. Genome-scale metabolic models of *E. coli* have been developed, such as iJE660 [19], iJR904 [20], iAF1260 [21], iJO1366 [22], and iML1515 [23], with others still being developed and im-

proved. Refinements of these models have shown the presence of as yet unidentified alternative pathways and isozymes and gaps in metabolic networks (orphan reactions) [24,142]. Knowledge of *E. coli* is also incomplete [22,24,142].

The minimal gene set concept [125] 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 [143]. These concepts have now expanded to include the concept of the chassis genome in synthetic biology [127,144].

Evaluation of *E. coli* identified 325 genes that could not be singly deleted [45,75]. However, even non-essential genes can be lethal in combination with other genes, making them synthetic lethal genes. Epistasis or interactions between genes and mutations are important for understanding gene function in *E. coli* and other cells [139,140,145–149].

Gene combinations of alternative pathways, compensatory pathways, and isozymes often exhibit synthetic lethality or synthetic sickness. Such genetic interactions can provide new insights into gene function [99,139,140,150,151]. Advances in genome design or genome deletions by design will require the systematization of knowledge from many sources, computer models for design, and genome editing technologies to enable experimental validation [152].

#### 4. Genomes by Synthesis

Although, to our knowledge, the *E. coli* genome has not been chemically synthesized, a genome has been reconstructed using completely synthetic DNA and *Mycoplasma* cells (Figure 1F) [17,18]. The genomic sequence of a living bacterium that serves as a template is required, although, in the near future, genomes may be reconstructed using fully synthetic DNA, based on the genome sequence design in *E. coli*. Determination of the rules for genome construction is absolutely required.

### 5. Discussion and Perspective

#### 5.1. 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 1000-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  $\lambda$  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 [153,154]. For example, participating institutions in Europe and Japan worked together as a community to build a *B. subtilis* deletion strain library [155,156].

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 from 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.

## 5.2. 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. Initially, the direction of the project was purely biological: to construct the minimum genome necessary for cell growth in nutrient media. Approaches included the removal of large areas of the genome that could be removed [74,157] and the removal by design of areas considered unnecessary for cell growth (see Section 2.5) [76]. 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, with quantitative analyses of the effects of single-gene deletions on intracellular conditions [137] suggesting 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 [129,143]. Moreover, the beginning of synthetic biology has resulted in expansion of the concept of the “chassis genome” [127,144].

Completely designing a genome is still not possible. This possibility may be enhanced by accumulating and analyzing genetic interaction information, including all genetic interactions and trans-omics analyses.

### 5.3. New Research Targets and Fields

The deletion of a single gene can affect the expression and function of a group of related genes. Although the individual effects may be small, the overall effect can be significant. Methods that look at the whole picture, such as RNA-seq, have made possible the analysis of how these initial disturbances affect the whole. The availability of comprehensive research resources has made possible the comparisons of the phenotypes associated with all gene deletions. In combination with other analytic methods, phenotypic analyses have made multidimensional analyses possible, as well as assessments of the position of individual gene groups in the overall activity of the cell. Thus, it has become experimentally possible to view cellular activity as a network [158].

Research on *E. coli* has followed this trend, with *E. coli* being one of the model organisms leading the new biology of the 21st century. This has led to the development and utilization of comprehensive research resources, including plasmid clones [115,116,159], deletion strains [43], and promoter fragment clones [117]. Several other resources remain under development.

### 5.4. 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 antimicrobial antibiotics and has developed research using comprehensive resources. As soon as resources become available, they are used for exhaustive screening [160], leading to the acceleration of research and development, and further expansion [161]. This group is therefore a good example of the successful use of both comprehensive and individual approaches in a single laboratory.

## 6. Epilog

The flow of genome modification in *E. coli* has increased expectations that researchers will have sufficient knowledge and technology to design organisms. To date, however, the knowledge to design genomes is insufficient. However, the steady accumulation of knowledge, evolving technologies, and progress in modeling has increased the understanding of genome organization. The ability to completely design artificial microbial cells may enable the construction of cells that can be used in medicine, antibiotic discovery, and engineering.

**Author Contributions:** Conceptualization, H.M.; methodology H.M.; investigation, H.M., M.K. and X.Y.; writing—H.M., M.K. and X.Y.; supervision H.M. and M.K.; project administration, H.M.; funding acquisition, H.M., M.K. and X.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Grant-in-Aid for Scientific Research (A), grant number 16H02485, from the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Project of Collaborative Innovation Center of GDAAS, grant number XTXM202202-2021A1515012401; the Key Research and Development Program of Guangdong Province, grant number 2020B0202080004; and the Special Fund for Scientific Innovation Strategy-Construction of High-Level Academy of Agriculture Science, grant number R2020PY-JC001 and R2020YJ-LJ001, Interdisciplinary Cluster for Cutting Edge Research, Shinshu University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Lederberg, J.; Tatum, E.L. Gene Recombination in *Escherichia coli*. *Nature* **1946**, *158*, 558. [[CrossRef](#)]
- Lennox, E. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1955**, *1*, 190–206. [[CrossRef](#)]
- Griffith, F. The Significance of Pneumococcal Types. *J. Hyg.* **1928**, *27*, 113–159. [[CrossRef](#)]
- Alexander, E.H.; Leidy, G. Transformation of type specificity of *Hemophilus influenzae*. *A.M.A. Am. J. Dis. Child.* **1950**, *80*, 877–878.
- Spizizen, J. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 1072–1078. [[CrossRef](#)]
- Mandel, M.; Higa, A. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **1970**, *53*, 159–162. [[CrossRef](#)]
- Cohen, S.N.; Chang, A.C.Y.; Hsu, L. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 2110–2114. [[CrossRef](#)]
- Cosloy, S.D.; Oishi, M. Genetic Transformation in *Escherichia coli* K12. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 84–87. [[CrossRef](#)]
- Murphy, K.C. Use of Bacteriophage  $\lambda$  Recombination Functions To Promote Gene Replacement in *Escherichia coli*. *J. Bacteriol.* **1998**, *180*, 2063–2071. [[CrossRef](#)]
- Datsenko, K.A.; Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6640–6645. [[CrossRef](#)]
- Yu, D.; Ellis, H.M.; Lee, E.-C.; Jenkins, N.A.; Copeland, N.G.; Court, D.L. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5978–5983. [[CrossRef](#)] [[PubMed](#)]
- Ishino, Y.; Krupovic, M.; Forterre, P. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *J. Bacteriol.* **2018**, *200*, e00580-17. [[CrossRef](#)] [[PubMed](#)]
- Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **1987**, *169*, 5429–5433. [[CrossRef](#)]
- Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816–821. [[CrossRef](#)]
- Wanner, B.L.; Teramoto, J.; Mori, H. What hath DNA wrought? CRISPR-CAS gene silencing and engineering from bacteria to humans. *Phys. Life Rev.* **2014**, *11*, 144–145. [[CrossRef](#)]
- Kolisnychenko, V.; Plunkett, G.; Herring, C.D.; Fehér, T.; Pósfai, J.; Blattner, F.R.; Pósfai, G. Engineering a Reduced *Escherichia coli* Genome. *Genome Res.* **2002**, *12*, 640–647. [[CrossRef](#)] [[PubMed](#)]
- Gibson, D.G.; Glass, J.I.; Lartigue, C.; Noskov, V.N.; Chuang, R.-Y.; Algire, M.A.; Benders, G.A.; Montague, M.G.; Ma, L.; Moodie, M.M.; et al. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010**, *329*, 52–56. [[CrossRef](#)]
- Hutchison, C.A.; Chuang, R.-Y.; Noskov, V.N.; Assad-Garcia, N.; Deerinck, T.J.; Ellisman, M.H.; Gill, J.; Kannan, K.; Karas, B.J.; Ma, L.; et al. Design and synthesis of a minimal bacterial genome. *Science* **2016**, *351*, aad6253. [[CrossRef](#)]
- Edwards, J.S.; Palsson, B.O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5528–5533. [[CrossRef](#)]
- Reed, J.L.; Vo, T.D.; Schilling, C.H.; O Palsson, B. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **2003**, *4*, R54. [[CrossRef](#)]
- Feist, A.M.; Henry, C.S.; Reed, J.L.; Krummenacker, M.; Joyce, A.R.; Karp, P.D.; Broadbelt, L.J.; Hatzimanikatis, V.; Palsson, B. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **2007**, *3*, 121. [[CrossRef](#)] [[PubMed](#)]
- Orth, J.D.; Conrad, T.M.; Na, J.; A Lerman, J.; Nam, H.; Feist, A.M.; Palsson, B. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol. Syst. Biol.* **2011**, *7*, 535. [[CrossRef](#)] [[PubMed](#)]

23. Monk, J.M.; Lloyd, C.J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **2017**, *35*, 904–908. [[CrossRef](#)] [[PubMed](#)]
24. Orth, J.D.; Palsson, B. Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst. Biol.* **2012**, *6*, 30. [[CrossRef](#)] [[PubMed](#)]
25. Karr, J.R.; Sanghvi, J.C.; Macklin, D.N.; Gutschow, M.V.; Jacobs, J.M.; Bolival, B.; Assad-Garcia, N.; Glass, J.I.; Covert, M.W. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* **2012**, *150*, 389–401. [[CrossRef](#)]
26. Macklin, D.N.; Ahn-Horst, T.A.; Choi, H.; Ruggero, N.A.; Carrera, J.; Mason, J.C.; Sun, G.; Agmon, E.; DeFelice, M.M.; Maayan, I.; et al. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science* **2020**, *369*, eaav3751. [[CrossRef](#)]
27. Lawley, T.; Klimke, W.; Gubbins, M.; Frost, L. F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* **2003**, *224*, 1–15. [[CrossRef](#)]
28. Hu, B.; Khara, P.; Christie, P.J. Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 14222–14227. [[CrossRef](#)]
29. Marinelli, L.J.; Hatfull, G.F.; Piuri, M. Recombineering. *Bacteriophage* **2012**, *2*, 5–14. [[CrossRef](#)]
30. Murphy, K.C.  $\lambda$  Recombination and Recombineering. *EcoSal Plus* **2016**, *7*. [[CrossRef](#)]
31. Leatham-Jensen, M.P.; Fridomt-Møller, J.; Adediran, J.; Mokszycki, M.E.; Banner, M.E.; Caughron, J.E.; Krogfelt, K.A.; Conway, T.; Cohen, P.S. The Streptomycin-Treated Mouse Intestine Selects *Escherichia coli* *envZ* Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. *Infect. Immun.* **2012**, *80*, 1716–1727. [[CrossRef](#)] [[PubMed](#)]
32. Michel, B.; Leach, D. Homologous Recombination—Enzymes and Pathways. *EcoSal Plus* **2012**, *5*. [[CrossRef](#)] [[PubMed](#)]
33. Hartley, J.L.; Temple, G.F.; Brasch, M.A. DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res.* **2000**, *10*, 1788–1795. [[CrossRef](#)] [[PubMed](#)]
34. Herrmann, S.; Siegl, T.; Luzhetska, M.; Petzke, L.; Jilg, C.; Welle, E.; Erb, A.; Leadlay, P.F.; Bechthold, A.; Luzhetskyy, A. Site-Specific Recombination Strategies for Engineering Actinomycete Genomes. *Appl. Environ. Microbiol.* **2012**, *78*, 1804–1812. [[CrossRef](#)]
35. Cherepanov, P.P.; Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. *Gene* **1995**, *158*, 9–14. [[CrossRef](#)]
36. Casadaban, M.J. Fusion of the *Escherichia coli* lac genes to the ara promoter: A general technique using bacteriophage Mu-1 insertions. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 809–813. [[CrossRef](#)]
37. Reznikoff, W.S. Transposon Tn5. *Annu. Rev. Genet.* **2008**, *42*, 269–286. [[CrossRef](#)]
38. Craig, N.L. Tn7: A target site-specific transposon. *Mol. Microbiol.* **1991**, *5*, 2569–2573. [[CrossRef](#)]
39. Kleckner, N.; Bender, J.; Gottesman, S. Uses of transposons with emphasis on Tn10. *Methods Enzymol.* **1991**, *204*, 139–180. [[CrossRef](#)]
40. Rubin, E.J.; Akerley, B.J.; Novik, V.N.; Lampe, D.J.; Husson, R.N.; Mekalanos, J.J. In vivo transposition of *mariner*-based elements in enteric bacteria and mycobacteria. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1645–1650. [[CrossRef](#)]
41. Zhuang, F.; Karberg, M.; Perutka, J.; Lambowitz, A.M. Ecl5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* **2009**, *15*, 432–449. [[CrossRef](#)] [[PubMed](#)]
42. Oliner, J.D.; Kinzler, K.W.; Vogelstein, B. In vivo cloning of PCR products in *E. coli*. *Nucleic Acids Res.* **1993**, *21*, 5192–5197. [[CrossRef](#)] [[PubMed](#)]
43. Nozaki, S.; Niki, H. Exonuclease III (XthA) Enforces In Vivo DNA Cloning of *Escherichia coli* To Create Cohesive Ends. *J. Bacteriol.* **2019**, *201*, e00660-18. [[CrossRef](#)] [[PubMed](#)]
44. Fels, U.; Gevaert, K.; Van Damme, P. Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics. *Front. Microbiol.* **2020**, *11*, 548410. [[CrossRef](#)] [[PubMed](#)]
45. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K.A.; Tomita, M.; Wanner, B.L.; Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2006**, *2*, 50. [[CrossRef](#)] [[PubMed](#)]
46. Ellis, H.M.; Yu, D.; DiTizio, T.; Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6742–6746. [[CrossRef](#)]
47. Marinelli, L.J.; Piuri, M.; Swigoňová, Z.; Balachandran, A.; Oldfield, L.M.; van Kessel, J.C.; Hatfull, G.F. BRED: A Simple and Powerful Tool for Constructing Mutant and Recombinant Bacteriophage Genomes. *PLoS ONE* **2008**, *3*, e3957. [[CrossRef](#)]
48. Wannier, T.M.; Nyerges, A.; Kuchwara, H.M.; Czikkely, M.; Balogh, D.; Filsinger, G.T.; Borders, N.C.; Gregg, C.J.; Lajoie, M.J.; Rios, X.; et al. Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 13689–13698. [[CrossRef](#)]
49. Wang, J.; Sarov, M.; Rientjes, J.; Fu, J.; Hollak, H.; Kranz, H.; Xie, W.; Stewart, A.F.; Zhang, Y. An Improved Recombineering Approach by Adding RecA to  $\lambda$  Red Recombination. *Mol. Biotechnol.* **2006**, *32*, 43–54. [[CrossRef](#)]
50. Li, X.-T.; Thomason, L.C.; Sawitzke, J.A.; Costantino, N.; Court, D.L. Bacterial DNA polymerases participate in oligonucleotide recombination. *Mol. Microbiol.* **2013**, *88*, 906–920. [[CrossRef](#)]
51. Haldimann, A.; Wanner, B.L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **2001**, *183*, 6384–6393. [[CrossRef](#)]

52. Goryshin, I.Y.; Naumann, T.A.; Apodaca, J.; Reznikoff, W.S. Chromosomal Deletion Formation System Based on Tn5 Double Transposition: Use For Making Minimal Genomes and Essential Gene Analysis. *Genome Res.* **2003**, *13*, 644–653. [[CrossRef](#)] [[PubMed](#)]
53. Hashimoto, M.; Ichimura, T.; Mizoguchi, H.; Tanaka, K.; Fujimitsu, K.; Keyamura, K.; Ote, T.; Yamakawa, T.; Yamazaki, Y.; Mori, H.; et al. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* **2004**, *55*, 137–149. [[CrossRef](#)] [[PubMed](#)]
54. Wang, H.H.; Isaacs, F.J.; Carr, P.A.; Sun, Z.Z.; Xu, G.; Forest, C.R.; Church, G. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **2009**, *460*, 894–898. [[CrossRef](#)] [[PubMed](#)]
55. Costantino, N.; Court, D.L. Enhanced levels of  $\lambda$  Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15748–15753. [[CrossRef](#)]
56. Posfai, G.; Kolisnychenko, V.; Berezcki, Z.; Blattner, F.R. Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Res.* **1999**, *27*, 4409–4415. [[CrossRef](#)]
57. Gay, P.; Le Coq, D.; Steinmetz, M.; Berkelman, T.; I Kado, C. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* **1985**, *164*, 918–921. [[CrossRef](#)]
58. Li, X.-T.; Thomason, L.C.; Sawitzke, J.A.; Costantino, N.; Court, D.L. Positive and negative selection using the tetA-sacB cassette: Recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res.* **2013**, *41*, e204. [[CrossRef](#)]
59. Metcalf, W.; Jiang, W.; Daniels, L.L.; Kim, S.K.; Haldimann, A.; Wanner, B.L. Conditionally Replicative and Conjugative Plasmids Carrying lacZ $\alpha$  for Cloning, Mutagenesis, and Allele Replacement in Bacteria. *Plasmid* **1996**, *35*, 1–13. [[CrossRef](#)]
60. Bochner, B.R.; Huang, H.C.; Schieven, G.L.; Ames, B.N. Positive selection for loss of tetracycline resistance. *J. Bacteriol.* **1980**, *143*, 926–933. [[CrossRef](#)]
61. de Zwaig, R.N.; Luria, S.E. Genetics and Physiology of Colicin-tolerant Mutants of *Escherichia coli*. *J. Bacteriol.* **1967**, *94*, 1112–1123. [[CrossRef](#)] [[PubMed](#)]
62. Morona, R.; Reeves, P. The tolC locus of *Escherichia coli* affects the expression of three major outer membrane proteins. *J. Bacteriol.* **1982**, *150*, 1016–1023. [[CrossRef](#)]
63. Gregg, C.J.; Lajoie, M.J.; Napolitano, M.G.; Mosberg, J.A.; Goodman, D.B.; Aach, J.; Isaacs, F.J.; Church, G.M. Rational optimization of tolC as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res.* **2014**, *42*, 4779–4790. [[CrossRef](#)]
64. Cirino, P.C.; Sun, L. Advancing Biocatalysis through Enzyme, Cellular, and Platform Engineering. *Biotechnol. Prog.* **2008**, *24*, 515–519. [[CrossRef](#)]
65. Vaknin, A.; Berg, H.C. Physical Responses of Bacterial Chemoreceptors. *J. Mol. Biol.* **2007**, *366*, 1416–1423. [[CrossRef](#)]
66. Lobato-Márquez, D.; Molina-García, L.; Moreno-Córdoba, I.; Portillo, F.G.-D.; Díaz-Orejas, R. Stabilization of the Virulence Plasmid pSLT of *Salmonella Typhimurium* by Three Maintenance Systems and Its Evaluation by Using a New Stability Test. *Front. Mol. Biosci.* **2016**, *3*, 66. [[CrossRef](#)] [[PubMed](#)]
67. Lobato-Márquez, D.; Molina-García, L. Evaluation of Plasmid Stability by Negative Selection in Gram-negative Bacteria. *Bioprotocol* **2017**, *7*, 2261. [[CrossRef](#)] [[PubMed](#)]
68. Gerdes, K.; Maisonneuve, E. Bacterial Persistence and Toxin-Antitoxin Loci. *Annu. Rev. Microbiol.* **2012**, *66*, 103–123. [[CrossRef](#)]
69. Adedirán, J.; Leatham-Jensen, M.P.; Mokszycki, M.E.; Frimodt-Møller, J.; Kroghfelt, K.A.; Kazmierczak, K.; Kenney, L.J.; Conway, T.; Cohen, P.S. An *Escherichia coli* Nissle 1917 Missense Mutant Colonizes the Streptomycin-Treated Mouse Intestine Better than the Wild Type but Is Not a Better Probiotic. *Infect. Immun.* **2014**, *82*, 670–682. [[CrossRef](#)]
70. Viola, M.G.; LaBreck, C.J.; Conti, J.; Camberg, J.L. Proteolysis-Dependent Remodeling of the Tubulin Homolog FtsZ at the Division Septum in *Escherichia coli*. *PLoS ONE* **2017**, *12*, e0170505. [[CrossRef](#)]
71. DiBiasio, E.C.; Dickinson, R.A.; Trebino, C.E.; Ferreira, C.N.; Morrison, J.J.; Camberg, J.L. The Stress-Active Cell Division Protein ZapE Alters FtsZ Filament Architecture to Facilitate Division in *Escherichia coli*. *Front. Microbiol.* **2021**, *12*, 733085. [[CrossRef](#)] [[PubMed](#)]
72. LaBreck, C.J.; Trebino, C.E.; Ferreira, C.N.; Morrison, J.J.; DiBiasio, E.C.; Conti, J.; Camberg, J.L. Degradation of MinD oscillator complexes by *Escherichia coli* ClpXP. *J. Biol. Chem.* **2021**, *296*, 100162. [[CrossRef](#)] [[PubMed](#)]
73. Vos, M.R.; Piraino, B.; LaBreck, C.J.; Rahmani, N.; Trebino, C.E.; Schoenle, M.; Peti, W.; Camberg, J.L.; Page, R. Degradation of the *E. coli* antitoxin MqsA by the proteolytic complex ClpXP is regulated by zinc occupancy and oxidation. *J. Biol. Chem.* **2021**, *298*, 101557. [[CrossRef](#)] [[PubMed](#)]
74. Yu, B.J.; Sung, B.H.; Koob, M.D.; Lee, C.H.; Lee, J.H.; Lee, W.S.; Kim, M.S.; Kim, S.C. Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. *Nat. Biotechnol.* **2002**, *20*, 1018–1023. [[CrossRef](#)]
75. Yamamoto, N.; Nakahigashi, K.; Nakamichi, T.; Yoshino, M.; Takai, Y.; Touda, Y.; Furubayashi, A.; Kinjo, S.; Dose, H.; Hasegawa, M.; et al. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* **2009**, *5*, 335. [[CrossRef](#)]
76. Pósfai, G.; Plunkett, G.; Fehér, T.; Frisch, D.; Keil, G.M.; Umenhoffer, K.; Kolisnychenko, V.; Stahl, B.; Sharma, S.S.; de Arruda, M.; et al. Emergent Properties of Reduced-Genome *Escherichia coli*. *Science* **2006**, *312*, 1044–1046. [[CrossRef](#)]
77. Fehér, T.; Karcagi, I.; Györfy, Z.; Umenhoffer, K.; Csörgő, B.; Pósfai, G. Scarless Engineering of the *Escherichia coli* Genome. In *Microbial Gene Essentiality: Protocols and Bioinformatics*; Humana Press: Totowa, NJ, USA, 2008; Volume 416, pp. 251–259. [[CrossRef](#)]
78. Ma, S.; Su, T.; Liu, J.; Lu, X.; Qi, Q. Reduction of the Bacterial Genome by Transposon-Mediated Random Deletion. *ACS Synth. Biol.* **2022**, *11*, 668–677. [[CrossRef](#)]



79. Wyman, C.; Kanaar, R. DNA Double-Strand Break Repair: All's Well that Ends Well. *Annu. Rev. Genet.* **2006**, *40*, 363–383. [[CrossRef](#)]
80. Chayot, R.; Montagne, B.; Mazel, D.; Ricchetti, M. An end-joining repair mechanism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2141–2146. [[CrossRef](#)]
81. Bhattacharyya, S.; Soniat, M.M.; Walker, D.; Jang, S.; Finkelstein, I.J.; Harshey, R.M. Phage Mu Gam protein promotes NHEJ in concert with *Escherichia coli* ligase. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E11614–E11622. [[CrossRef](#)]
82. Li, Y.; Lin, Z.; Huang, C.; Zhang, Y.; Wang, Z.; Tang, Y.-J.; Chen, T.; Zhao, X. Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 mediated genome editing. *Metab. Eng.* **2015**, *31*, 13. [[CrossRef](#)] [[PubMed](#)]
83. Zhao, D.; Yuan, S.; Xiong, B.; Sun, H.; Ye, L.; Li, J.; Zhang, X.; Bi, C. Development of a fast and easy method for *Escherichia coli* genome editing with CRISPR/Cas9. *Microb Cell Factories* **2016**, *15*, 1–9. [[CrossRef](#)]
84. Tian, P.; Wang, J.; Shen, X.; Rey, J.F.; Yuan, Q.; Yan, Y. Fundamental CRISPR-Cas9 tools and current applications in microbial systems. *Synth. Syst. Biotechnol.* **2017**, *2*, 219–225. [[CrossRef](#)] [[PubMed](#)]
85. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **2018**, *9*, 1–13. [[CrossRef](#)] [[PubMed](#)]
86. Reisch, C.R.; Prather, K.L.J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **2015**, *5*, 15096. [[CrossRef](#)]
87. Banno, S.; Nishida, K.; Arazoe, T.; Mitsunobu, H.; Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **2018**, *3*, 423–429. [[CrossRef](#)]
88. Chen, W.; Zhang, Y.; Zhang, Y.; Pi, Y.; Gu, T.; Song, L.; Wang, Y.; Ji, Q. CRISPR/Cas9-based Genome Editing in *Pseudomonas aeruginosa* and Cytidine Deaminase-Mediated Base Editing in *Pseudomonas* Species. *iScience* **2018**, *6*, 222–231. [[CrossRef](#)]
89. Kantor, A.; McClements, M.E.; MacLaren, R.E. CRISPR-Cas9 DNA Base-Editing and Prime-Editing. *Int. J. Mol. Sci.* **2020**, *21*, 6240. [[CrossRef](#)]
90. Tong, Y.; Jørgensen, T.S.; Whitford, C.M.; Weber, T.; Lee, S.Y. A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.* **2021**, *12*, 1–11. [[CrossRef](#)]
91. Yuan, T.; Yan, N.; Fei, T.; Zheng, J.; Meng, J.; Li, N.; Liu, J.; Zhang, H.; Xie, L.; Ying, W.; et al. Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods. *Nat. Commun.* **2021**, *12*, 1–11. [[CrossRef](#)]
92. Zheng, K.; Wang, Y.; Li, N.; Jiang, F.-F.; Wu, C.-X.; Liu, F.; Chen, H.-C.; Liu, Z.-F. Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. *Commun. Biol.* **2018**, *1*, 1–6. [[CrossRef](#)] [[PubMed](#)]
93. Anzalone, A.V.; Randolph, P.B.; Davis, J.R.; Sousa, A.A.; Koblan, L.W.; Levy, J.M.; Chen, P.J.; Wilson, C.; Newby, G.A.; Raguram, A.; et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019**, *576*, 149–157. [[CrossRef](#)] [[PubMed](#)]
94. Yan, J.; Cirincione, A.; Adamson, B. Prime Editing: Precision Genome Editing by Reverse Transcription. *Mol. Cell* **2020**, *77*, 210–212. [[CrossRef](#)]
95. Park, J.; Yoon, J.; Kwon, D.; Han, M.-J.; Choi, S.; Park, S.; Lee, J.; Lee, K.; Lee, J.; Lee, S.; et al. Enhanced genome editing efficiency of CRISPR PLUS: Cas9 chimeric fusion proteins. *Sci. Rep.* **2021**, *11*, 1–9. [[CrossRef](#)]
96. Zhang, G.; Liu, Y.; Huang, S.; Qu, S.; Cheng, D.; Yao, Y.; Ji, Q.; Wang, X.; Huang, X.; Liu, J. Enhancement of prime editing via xrRNA motif-joined pegRNA. *Nat. Commun.* **2022**, *13*, 1–12. [[CrossRef](#)]
97. Dong, H.; Cui, Y.; Zhang, D. CRISPR/Cas Technologies and Their Applications in *Escherichia coli*. *Front. Bioeng. Biotechnol.* **2021**, *9*, 762676. [[CrossRef](#)] [[PubMed](#)]
98. Qi, L.S.; Larson, M.H.; Gilbert, L.A.; Doudna, J.A.; Weissman, J.S.; Arkin, A.P.; Lim, W.A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152*, 1173–1183. [[CrossRef](#)]
99. Peters, J.M.; Colavin, A.; Shi, H.; Czarny, T.L.; Larson, M.H.; Wong, S.; Hawkins, J.S.; Lu, C.H.; Koo, B.-M.; Marta, E.; et al. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **2016**, *165*, 1493–1506. [[CrossRef](#)]
100. Cui, L.; Vigouroux, A.; Rousset, F.; Varet, H.; Khanna, V.; Bikard, D. A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.* **2018**, *9*, 1–10. [[CrossRef](#)]
101. Rousset, F.; Cui, L.; Siouve, E.; Becavin, C.; Depardieu, F.; Bikard, D. Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLoS Genet.* **2018**, *14*, e1007749. [[CrossRef](#)]
102. Wang, T.; Guan, C.; Guo, J.; Liu, B.; Wu, Y.; Xie, Z.; Zhang, C.; Xing, X.-H. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat. Commun.* **2018**, *9*, 1–15. [[CrossRef](#)] [[PubMed](#)]
103. Hood, L.E.; Hunkapiller, M.W.; Smith, L.M. Automated DNA sequencing and analysis of the human genome. *Genomics* **1987**, *1*, 201–212. [[CrossRef](#)]
104. Isaacs, F.J.; Carr, P.A.; Wang, H.H.; Lajoie, M.J.; Sterling, B.; Kraal, L.; Tolonen, A.C.; Gianoulis, T.A.; Goodman, D.B.; Reppas, N.B.; et al. Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **2011**, *333*, 348–353. [[CrossRef](#)]
105. Wang, H.H.; Church, G.M. Multiplexed Genome Engineering and Genotyping Methods. *Methods Enzymol.* **2011**, *498*, 409–426. [[CrossRef](#)]
106. Lajoie, M.J.; Rovner, A.J.; Goodman, D.B.; Aerni, H.-R.; Haimovich, A.D.; Kuznetsov, G.; Mercer, J.A.; Wang, H.H.; Carr, P.A.; Mosberg, J.A.; et al. Genomically Recoded Organisms Expand Biological Functions. *Science* **2013**, *342*, 357–360. [[CrossRef](#)] [[PubMed](#)]

107. Li, X.; Costantino, N.; Lu, L.; Liu, D.; Watt, R.M.; Cheah, K.S.E.; Court, D.L.; Huang, J. Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res.* **2003**, *31*, 6674–6687. [[CrossRef](#)] [[PubMed](#)]
108. Lajoie, M.J.; Gregg, C.J.; Mosberg, J.A.; Washington, G.C.; Church, G.M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res.* **2012**, *40*, e170. [[CrossRef](#)]
109. Balakrishnan, L.; Bambara, R.A. Okazaki Fragment Metabolism. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a010173. [[CrossRef](#)]
110. Oshima, T.; Aiba, H.; Masuda, Y.; Kanaya, S.; Sugiura, M.; Wanner, B.L.; Mori, H.; Mizuno, T. Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. *Mol. Microbiol.* **2002**, *46*, 281–291. [[CrossRef](#)]
111. Oshima, T.; Wada, C.; Kawagoe, Y.; Ara, T.; Maeda, M.; Masuda, Y.; Hiraga, S.; Mori, H. Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. *Mol. Microbiol.* **2002**, *45*, 673–695. [[CrossRef](#)]
112. Tao, H.; Bausch, C.; Richmond, C.; Blattner, F.R.; Conway, T. Functional Genomics: Expression Analysis of *Escherichia coli* Growing on Minimal and Rich Media. *J. Bacteriol.* **1999**, *181*, 6425–6440. [[CrossRef](#)] [[PubMed](#)]
113. Richmond, C.S.; Glasner, J.D.; Mau, R.; Jin, H.; Blattner, F.R. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **1999**, *27*, 3821–3835. [[CrossRef](#)] [[PubMed](#)]
114. Van Dyk, T.K.; Wei, Y.; Hanafey, M.K.; Dolan, M.; Reeve, M.J.G.; Rafalski, J.A.; Rothman-Denes, L.B.; LaRossa, R.A. A genomic approach to gene fusion technology. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2555–2560. [[CrossRef](#)] [[PubMed](#)]
115. Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res.* **2006**, *12*, 291–299. [[CrossRef](#)] [[PubMed](#)]
116. Saka, K.; Tadenuma, M.; Nakade, S.; Tanaka, N.; Sugawara, H.; Nishikawa, K.; Ichiyoshi, N.; Kitagawa, M.; Mori, H.; Ogasawara, N.; et al. A Complete Set of *Escherichia coli* Open Reading Frames in Mobile Plasmids Facilitating Genetic Studies. *DNA Res.* **2005**, *12*, 63–68. [[CrossRef](#)] [[PubMed](#)]
117. Zaslaver, A.; Bren, A.; Ronen, M.; Itzkovitz, S.; Kikoin, I.; Shavit, S.; Liebermeister, W.; Surette, M.G.; Alon, U. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **2006**, *3*, 623–628. [[CrossRef](#)] [[PubMed](#)]
118. Tsuge, K.; Nakahigashi, K.; Togashi, T.; Hasebe, M.; Takai, Y.; Hasegawa, M.; Tomita, M.; Itaya, M. An artificial glycolysis operon library toward elucidation of possible “operon rule”. *Genes Genet. Syst.* **2010**, *85*, 454.
119. Tsuge, K.; Matsui, K.; Itaya, M. One step assembly of multiple DNA fragments with a designed order and orientation in *Bacillus subtilis* plasmid. *Nucleic Acids Res.* **2003**, *31*, e133. [[CrossRef](#)]
120. Sánchez-Pascuala, A.; De Lorenzo, V.; Nikel, P.I. Refactoring the Embden–Meyerhof–Parnas Pathway as a Whole of Portable GlucoBricks for Implantation of Glycolytic Modules in Gram-Negative Bacteria. *ACS Synth. Biol.* **2017**, *6*, 793–805. [[CrossRef](#)]
121. Zhu, B.; Cai, G.; Hall, E.O.; Freeman, G.J. In-Fusion™ assembly: Seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *BioTechniques* **2007**, *43*, 354–359. [[CrossRef](#)]
122. Zhang, Y.; Werling, U.; Edelmann, W. Seamless Ligation Cloning Extract (SLiCE) Cloning Method. *Methods Mol. Biol.* **2013**, *1116*, 235–244. [[CrossRef](#)]
123. Gibson, D.G.; Young, L.; Chuang, R.-Y.; Venter, J.C.; Hutchison, C.A., III; Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345. [[CrossRef](#)] [[PubMed](#)]
124. Umlauf, S.W.; Cox, M.M. The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *EMBO J.* **1988**, *7*, 1845–1852. [[CrossRef](#)] [[PubMed](#)]
125. Koonin, E.V. Big Time for Small Genomes: Table 1. *Genome Res.* **1997**, *7*, 418–421. [[CrossRef](#)] [[PubMed](#)]
126. Mushegian, A.R.; Koonin, E.V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10268–10273. [[CrossRef](#)]
127. Danchin, A. Scaling up synthetic biology: Do not forget the chassis. *FEBS Lett.* **2012**, *586*, 2129–2137. [[CrossRef](#)]
128. Mizoguchi, H.; Sawano, Y.; Kato, J.-I.; Mori, H. Superpositioning of Deletions Promotes Growth of *Escherichia coli* with a Reduced Genome. *DNA Res.* **2008**, *15*, 277–284. [[CrossRef](#)]
129. Park, M.K.; Lee, S.H.; Yang, K.S.; Jung, S.-C.; Lee, J.H.; Kim, S.C. Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 6701–6713. [[CrossRef](#)]
130. Lajoie, M.J.; Kosuri, S.; Mosberg, J.A.; Gregg, C.J.; Zhang, D.; Church, G.M. Probing the Limits of Genetic Recoding in Essential Genes. *Science* **2013**, *342*, 361–363. [[CrossRef](#)]
131. Ostrov, N.; Landon, M.; Guell, M.; Kuznetsov, G.; Teramoto, J.; Cervantes, N.; Zhou, M.; Singh, K.; Napolitano, M.G.; Moosburner, M.; et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **2016**, *353*, 819–822. [[CrossRef](#)]
132. Wang, K.; Fredens, J.; Brunner, S.F.; Kim, S.H.; Chia, T.; Chin, J.W. Defining synonymous codon compression schemes by genome recoding. *Nature* **2016**, *539*, 59–64. [[CrossRef](#)] [[PubMed](#)]
133. Fredens, J.; Wang, K.; de la Torre, D.; Funke, L.F.H.; Robertson, W.E.; Christova, Y.; Chia, T.; Schmied, W.; Dunkelmann, D.L.; Beránek, V.; et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **2019**, *569*, 514–518. [[CrossRef](#)] [[PubMed](#)]
134. Hayashi, K.; Morooka, N.; Yamamoto, Y.; Fujita, K.; Isono, K.; Choi, S.; Ohtsubo, E.; Baba, T.; Wanner, B.L.; Mori, H.; et al. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* **2006**, *2*, 49. [[CrossRef](#)]
135. Riley, M. *Escherichia coli* K-12: A cooperatively developed annotation snapshot–2005. *Nucleic Acids Res.* **2006**, *34*, 1–9. [[CrossRef](#)]
136. Miki, T.; Yamamoto, Y.; Matsuda, H. A Novel, Simple, High-Throughput Method for Isolation of Genome-Wide Transposon Insertion Mutants of *Escherichia coli* K-12. *Methods Mol. Biol.* **2008**, *416*, 195–204. [[CrossRef](#)]

137. Ishii, N.; Nakahigashi, K.; Baba, T.; Robert, M.; Soga, T.; Kanai, A.; Hirasawa, T.; Naba, M.; Hirai, K.; Hoque, A.; et al. Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations. *Science* **2007**, *316*, 593–597. [[CrossRef](#)]
138. Otsuka, Y.; Muto, A.; Takeuchi, R.; Okada, C.; Ishikawa, M.; Nakamura, K.; Yamamoto, N.; Dose, H.; Nakahigashi, K.; Tanishima, S.; et al. GenoBase: Comprehensive resource database of *Escherichia coli* K-12. *Nucleic Acids Res.* **2014**, *43*, D606–D617. [[CrossRef](#)]
139. Typas, A.; Nichols, R.J.; Siegele, D.; Shales, M.; Collins, S.; Lim, B.; Braberg, H.; Yamamoto, N.; Takeuchi, R.; Wanner, B.L.; et al. High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods* **2008**, *5*, 781–787. [[CrossRef](#)]
140. Butland, G.; Babu, M.; Díaz-Mejía, J.J.; Bohdana, F.; Phanse, S.; Gold, B.; Yang, W.; Li, J.; Gagarinova, A.G.; Pogoutse, O.; et al. eSGA: *E. coli* synthetic genetic array analysis. *Nat. Chem. Biol.* **2008**, *5*, 789–795. [[CrossRef](#)]
141. Hobbs, E.C.; Astarita, J.L.; Storz, G. 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.* **2010**, *192*, 59–67. [[CrossRef](#)]
142. Orth, J.D.; Palsson, B. Systematizing the generation of missing metabolic knowledge. *Biotechnol. Bioeng.* **2010**, *107*, 403–412. [[CrossRef](#)] [[PubMed](#)]
143. Mori, H.; Mizoguchi, H.; Fujio, T. *Escherichia coli* minimum genome factory. *Biotechnol. Appl. Biochem.* **2007**, *46*, 157–167. [[CrossRef](#)] [[PubMed](#)]
144. Calero, P.; Nikel, P.I. Chasing bacterial chassis for metabolic engineering: A perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **2018**, *12*, 98–124. [[CrossRef](#)] [[PubMed](#)]
145. Elena, S.F.; Lenski, R.E. Epistasis between new mutations and genetic background and a test of genetic canalization. *Evolution* **2001**, *55*, 1746–1752. [[CrossRef](#)]
146. He, X.; Qian, W.; Wang, Z.; Li, Y.; Zhang, J. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat. Genet.* **2010**, *42*, 272–276. [[CrossRef](#)]
147. Phillips, P.C. Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **2008**, *9*, 855–867. [[CrossRef](#)]
148. Khan, A.I.; Dinh, D.M.; Schneider, D.; Lenski, R.E.; Cooper, T.F. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science* **2011**, *332*, 1193–1196. [[CrossRef](#)]
149. Klimova, A.; Sandler, S.J. An Epistasis Analysis of *recA* and *recN* in *Escherichia coli* K-12. *Genetics* **2020**, *216*, 381–393. [[CrossRef](#)]
150. Gagarinova, A.; Stewart, G.; Samanfar, B.; Phanse, S.; White, C.A.; Aoki, H.; Deineko, V.; Beloglazova, N.; Yakunin, A.F.; Golshani, A.; et al. Systematic Genetic Screens Reveal the Dynamic Global Functional Organization of the Bacterial Translation Machinery. *Cell Rep.* **2016**, *17*, 904–916. [[CrossRef](#)]
151. Côté, J.-P.; French, S.; Gehrke, S.S.; MacNair, C.R.; Mangat, C.S.; Bharat, A.; Brown, E.D. The Genome-Wide Interaction Network of Nutrient Stress Genes in *Escherichia coli*. *mBio* **2016**, *7*, e01714-16. [[CrossRef](#)]
152. Landon, S.; Rees-Garbutt, J.; Marucci, L.; Grierson, C. Genome-driven cell engineering review: In vivo and in silico metabolic and genome engineering. *Essays Biochem.* **2019**, *63*, 267–284. [[CrossRef](#)] [[PubMed](#)]
153. Botstein, D.; Fink, G.R. Yeast: An Experimental Organism for 21st Century Biology. *Genetics* **2011**, *189*, 695–704. [[CrossRef](#)] [[PubMed](#)]
154. Giaever, G.; Nislow, C. The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics* **2014**, *197*, 451–465. [[CrossRef](#)]
155. Vagner, V.; Dervyn, E.; Ehrlich, S.D. A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology* **1998**, *144*, 3097–3104. [[CrossRef](#)]
156. Kobayashi, K.; Ehrlich, S.D.; Albertini, A.; Amati, G.; Andersen, K.K.; Arnaud, M.; Asai, K.; Ashikaga, S.; Aymerich, S.; Bessieres, P.; et al. Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4678–4683. [[CrossRef](#)] [[PubMed](#)]
157. Kato, J.; Hashimoto, M. Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* **2007**, *3*, 132. [[CrossRef](#)]
158. Barabási, A.-L.; Oltvai, Z.N. Network biology: Understanding the cell’s functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [[CrossRef](#)]
159. Rajagopala, S.V.; Yamamoto, N.; E Zweifel, A.; Nakamichi, T.; Huang, H.-K.; Mendez-Rios, J.D.; Franca-Koh, J.; Boorgula, M.P.; Fujita, K.; Suzuki, K.-I.; et al. The *Escherichia coli* K-12 ORFeome: A resource for comparative molecular microbiology. *BMC Genom.* **2010**, *11*, 470. [[CrossRef](#)]
160. Pathania, R.; Zlitni, S.; Barker, C.; Das, R.; A Gerritsma, D.; Lebert, J.; Awuah, E.; Melacini, G.; A Capretta, F.; Brown, E.D. Chemical genomics in *Escherichia coli* identifies an inhibitor of bacterial lipoprotein targeting. *Nat. Chem. Biol.* **2009**, *5*, 849–856. [[CrossRef](#)]
161. Brown, E.D.; Wright, G.D. Antibacterial drug discovery in the resistance era. *Nature* **2016**, *529*, 336–343. [[CrossRef](#)] [[PubMed](#)]