Chapter 2
Escherichia coli Genome Engineering and Minimization for the Construction of a Bioengine

Bong Hyun Sung, Jun Hyoung Lee, and Sun Chang Kim

Contents

Abstract A profusion of diverse genome-related information has been obtained by the sequencing of genomes from many microorganisms, functional analyses of these genomes, and the application of bioinformatics techniques to genomics, proteomics, and systems biology. The resulting barrage of data coupled with large-scale gene inactivation studies have allowed researchers to produce a genetic blueprint for a streamline, custom-designed microbe that carries the minimal gene set required for the organism to replicate in a given environment. On the basis of this minimal genome information, several research groups have generated minimal-genome *Escherichia coli* strains using sophisticated genome engineering techniques, such as the dual transposition, site-specific recombinations, and markerless genome recombination. These minimal genomes display various desirable traits for biological researches, such as improved genome stability, increased transformation efficacy, and higher production of biological materials. Therefore, the generation of

S.C. Kim (✉)
Department of Biological Sciences, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea
e-mail: sunkim@kaist.ac.kr

S.Y. Lee (ed.), *Systems Biology and Biotechnology of Escherichia coli*,
DOI 10.1007/978-1-4020-9394-4_2,

a large number of deletion mutants of the minimal *E. coli* genomes coupled with restructuring of regulatory circuits may lead to facilitate the construction of a variety of custom-designed bacterial strains (also called a "bioengine") with myriad practical and commercial applications.

2.1 Introduction

Nucleotide sequencing and comparative analysis of multiple diverse genomes is revolutionizing contemporary biology by providing a framework for interpreting and predicting the physiological properties of an organism. A variety of emerging postgenomic techniques, such as genome-wide gene expression profiling and monitoring of interactions among macromolecules, are helping to define the molecular compositions of cells. Scientists have developed, and continue to refine, sophisticated new computational approaches that allow one to explore the inherent organization of cellular networks as well as the mode and dynamics of interactions among cellular constituents (Hasty et al. 2001, Herring et al. 2006, Ishii et al. 2007, Kitano 2002, Koonin et al. 2002). These intricate tools and techniques have introduced a new paradigm in cell biology research: the construction of custom-designed, minimal-genome microbes (bioengines) that perform functions that raise the quality of life for human beings.

A minimal genome can be defined as a one that contains the smallest set of genes that allows the organism to replicate in a given environment (Mushegian 1999). The creation and study of minimal microbial genomes can help to increase our understanding of complex genetic material and provide a basis for the design of custom bacterial strains. Drawing on the complete genome sequences of more than 800 microorganisms (June 2008, Genomes OnLine Database http://www.genomesonline.org/) as well as extensive functional analyses of their gene products, researchers have proposed two different approaches for the construction of minimal genomes (Cho et al. 1999, Luisi 2002, Maniloff 1996). The first is a "top-down" approach, which involves trimming the genome of sequences that appear to be unnecessary on the basis of functional genomic studies of microorganisms. The second is a "bottom-up" approach, which entails synthesizing the proposed minimal genome from basic chemical building blocks and inserting it into an environment that allows metabolic activity and replication (Forster and Church 2006, 2007, Szostak et al. 2001). Although simple biological constructs can be synthesized artificially (Cello et al. 2002, Gibson et al. 2008, Itaya et al. 2008, Smith et al. 2003, Tumpey et al. 2005), the bottom-up approach is technically challenging and the actual synthesis of an artificial minimal genome from chemical building blocks is not possible if one lacks a complete functional analysis of the genes needed for life. However, the top-down approach, which starts with the intact genome of a well-characterized microorganism, is more technically feasible; this is because the top-down approach can be initiated in parallel with rapidly progressing functional genomics research in microorganisms.

Back in 1984, H. Morowitz first raised the idea of using mycoplasmas as models for the construction of a minimal genome in a living cell (Razin 1997a, b). The complete genomic sequence of the human pathogen *Mycoplasma genitalium* (Fraser et al. 1995) revealed that the genome is only 580 kb in length and contains only about 470 predicted protein coding genes, as compared with 1,703 and 4,288 in the *Haemophilus influenzae* (Fleischmann et al. 1995) and *Escherichia coli* (Blattner et al. 1997) genomes, respectively. However, mycoplasmas are parasites that evolved, by degenerative or reductive evolution, from Gram-positive bacteria and they have not been well characterized, because they are hard to grow and their genomes are difficult to manipulate. Therefore, scientists chose to attempt minimal genome construction in free-living microorganisms (Koob et al. 1994).

Among the various free-living microorganisms, *E. coli* is the most logical choice for minimal genome construction experiments, as it is more fully defined at the molecular level than any other microorganism. Furthermore, because of its nearly ubiquitous use as a research tool and its favorable growth characteristics, this bacterium has been the organism of choice in the development of tools for sophisticated genetic engineering. Finally, *E. coli* is one of the best commercially applicable hosts for the pharmaceutical and fermentation industries (Blattner et al. 1997, Riley et al. 2006).

Functional analysis of the *E. coli* genome has revealed that bacteria that are grown under a given condition use only a fraction of their genes for growth, replication, and production of important biological materials (Richmond et al. 1999, Tao et al. 1999). This is because the *E. coli* genome contains nonessential genes, transposable elements, bacteriophage DNA, cryptic prophages, pseudogenes, gene remnants, damaged operons, and virulence factors, some of which yield unnecessary or unwanted products that interfere with rational strain improvement and the production of desired biological substances (Blattner et al. 1997, Hayashi et al. 2006). Therefore, the identification and deletion of nonessential genes and other dispensable sequences in the microbial genome is necessary for the construction of a custom-designed bioengine, in which cellular metabolites and energy sources are efficiently optimized and directed toward the production of both essential and desired gene products (Edwards and Palsson 2000, Park et al. 2007). Concomitantly, the bioengine's metabolic waste and bio-pollution can be minimized, and the quality and stability of its products can be maximized (Cho et al. 1999, Kolisnychenko et al. 2002, Westers et al. 2003).

In this chapter, we describe approaches for minimizing the *E. coli* genome by eliminating unnecessary genes, to create a self-sustaining, self-replicating, artificial bioengine.

2.2 Estimating the Size and Gene Content of Minimal Genomes

Identification of the regulatory and protein-coding DNA sequences that are most essential for maintaining and replicating a free-living cellular organism is a logical first step in the construction of a custom-designed bioengine. Several approaches

have been used to identify essential and nonessential genes in microorganisms under given conditions, with the goal of defining the minimal set of genes necessary for cell survival and self-replication.

2.2.1 Identification of Minimal Essential Gene Sets by Sequence Comparison in silico

The sequenced genome of the parasitic bacterium *M. genitalium* contains only about 470 identified protein-coding genes, and these have been dubbed a minimal gene complement (Fraser et al. 1995). Although the complete gene complement of *M. genitalium* is nearly the smallest one among cellular life forms with sequenced genomes, there is no evidence that this collection of 470 genes represents a minimal self-sufficient gene set. To derive such a set, Mushegian and Koonin (1996) compared the 468 predicted *M. genitalium* protein sequences with the 1,703 predicted protein sequences encoded by the other completely sequenced microbial genome, that of *H. influenzae* (Fleischmann et al. 1995). Because these microorganisms belong to two distinct ancient bacterial lineages [that is, Gram-positive (*M. genitalium*) and Gram-negative (*H. influenzae*) bacteria], genes that are conserved in these two organisms are almost certainly essential for growth and replication. This genomic comparison suggested that the minimal number of genes necessary and sufficient to sustain the existence of a modern-type cell is closer to 256 (Koonin et al. 2002, Mushegian 1999). On the basis of these 256 genes, Mushegian and Koonin suggested the following key features that must be specified by a minimal gene set: rudimentary systems for gene transcription, protein translation, DNA replication, recombination, and repair; chaperone-like proteins and machinery for protein export and metabolite transport; and nucleotide salvage pathways.

Since then, comparisons of protein-coding regions in complete genome sequences from diverse organisms have revealed clusters of orthologous groups (COGs), (Tatusov et al. 1997); because orthologous proteins likely have similar functions, the COGs have been used to define a minimal gene set of life (Mushegian 1999). Using a similar approach, researchers compared genome sequences from uropathogenic *E. coli* CFT073, enterohemorrhagic DEL933, and the *E. coli* laboratory strain MG1655 and defined a combined set of nonredundant protein-encoding genes. Of these genes, only 39.2% (2,996 genes) are common to all three strains (Welch et al. 2002). However, when such an analysis was carried out with about 100 sequenced genomes, only 63 genes were found to be ubiquitous; most of these genes encode translational components, and a few specify basic components of the transcription machinery (Koonin 2003).

The challenge of this comparative genomic approach to the identification of a minimal gene set is that certain function-related complexities are not taken into account. For example, the minimal number of essential functions may be larger than that predicted by genome sequence comparison, because not all proteins that perform the same function share detectable sequence similarity (Riley and Serres 2000). Therefore, because it may substantially underestimate the size of the min-

imal gene set, one cannot rely exclusively on the comparative approach (Feher et al. 2007).

2.2.2 Identification of Essential Genes by Large–Scale Gene Inactivation

Global transposition mutagenesis has been used to identify nonessential genes in an effort to learn whether the naturally occurring gene complements behave as true minimal genomes under laboratory conditions.

The positions of 2,209 transposon insertions in the completely sequenced genomes of *M. genitalium* and its close relative *Mycoplasma peumoniae* were determined by sequencing across the junctions of the transposons and the genomic DNA (Hutchison et al. 1999). These junctions defined 1,354 distinct sites in which transposon insertion did not lead to lethality. This analysis suggests that 265–350 of the 480 protein-coding genes of *M. genitalium* are essential under laboratory conditions, including about 100 genes of unknown function (Hutchison et al. 1999).

In *Bacillus subtilis*, Itaya (1995) introduced mutations in a small set of randomly selected genetic loci, examined the growth properties of the mutants, and determined the percentage of genes that could be disrupted without loss of viability. This led to the hypothesis that the minimal *B. subtilis* genome may comprise about 318–562 kb which, given the average size of ~1 kb for a bacterial protein-coding gene, corresponds to 300–500 genes (Kunst et al. 1997). And with a systematic approach that employed single-gene disruptions that covered the complete *Bacillus* genome, Kobayashi et al. (2003) identified about 270 genes that are indispensable for growth of *B. subtilis* in a rich medium at 37 °C.

In *Pseudomonas aeruginosa*, *H. influenzae*, *Corynebacterium glutamicum*, and *E. coli*, global transposition mutagenesis has identified 678, 670, 658, and 620 genes, respectively, those essential for growth under laboratory conditions (Akerley et al. 2002, Gerdes et al. 2003, Jacobs et al. 2003, Suzuki et al. 2006). However, when Baba et al. (2006) used Red recombination to generate a set of precisely defined, single-gene deletions of all putative protein-coding genes in *E. coli* K-12, of the 4,288 genes targeted, only 303 genes, including 37 of unknown function, were unable to be disrupted in Lulia-Bertani (LB) medium. When these 303 essential genes were divided into functional groups, the major subsets contained members of COGs that play roles in protein translation, ribosomal structure, cell division, lipid metabolism, transcription, and cell envelope biogenesis. And only 67% (205 genes) of the 303 essential genes overlap with those in the essential gene set predicted by global transposition (Baba et al. 2006, Gerdes et al. 2003). These differences can be attributed to the use of different mutagenesis strategies and different growth conditions. However, because the global transposition system measures the effect of mutations on cell populations, a mutation that causes very slow growth can appear to be lethal and hence be falsely classified as essential. Furthermore, of the 3,988 single-gene deletion mutants made by Baba et al., 119 gave rise to mutant *E. coli* strains that were unable to grow on glycerol-supplemented M9 minimal medium

Table 2.1 Number of estimated essential genes in various microorganisms

Method	Strain	No. of essential genes/total ORFs (%)	Ref.
Comparative genomics	*M. genitalium and H. influenzae*	256	Mushegian and Koonin 1999
	M. genitalium	265–350/468 (79%)	Hutchison et al. 1999
	H. influenzae RD	670/1703 (38%)	Akerley et al. 2002
Global transposition	*E. coli* K12	620/4296 (14%)	Gerdes et al. 2003
	P. aeruginosa	678/5500 (12%)	Jacobs et al. 2003
	H. pyroli	255/1590 (16%)	Salama et al. 2004
	C. glutamicum R	658/2990 (22%)	Suzuki et al. 2006
Single-gene deletion or knockout	*S. cerevisiae*	1105/5916 (19%)	Giaever et al. 2002
	B. subtilis 168	271/4099 (6.8%)	Kobayashi et al. 2003
	S. typhimurium LT2	490/4597 (11%)	Knuth et al. 2004
	E. coli K12	303/4296 (7%)	Baba et al. 2006

(Joyce et al. 2006). Information about the variously defined minimal gene sets is summarized in Table 2.1.

2.3 Techniques for Experimental Genome Minimization

For some bacteria, strains that carry spontaneous deletion mutations can be obtained by long-term serial passage of the cells under laboratory culture conditions (Cooper et al. 2001). However, the spontaneous deleterious mutation rate for the *E. coli* genome is too low ($\sim 2 \times 10^{-4}$ per generation) to allow deletion mutants to be efficiently created using this technique (Elena and Lenski 2003, Kibota and Lynch 1996). Therefore, for the rapid construction of minimal genomes by removing protein-coding genes and other genomic sequences that previously were shown to be nonessential, researchers have developed a variety of deletion methods, including homologous recombination using suicide plasmids, linear DNA recombination using the phage Red system, site-specific recombination system, and random deletion by double transposition.

2.3.1 Suicide Plasmid–Mediated Genomic Deletion

Suicide plasmids are convenient vehicles for the delivery of DNA into the *E. coli* chromosome. Link et al. (1997) have described a method for gene replacement in *E. coli* that uses a homologous recombination between the bacterial chromosome and a suicide plasmid, carrying cloned chromosomal fragments (homologous arms), whose replication ability is temperature sensitive (Fig. 2.1A). At the non-permissive temperature (42 °C), cells maintain drug resistance only if the plasmid integrates into the chromosome by homologous recombination between the cloned fragment and the bacterial chromosome (Hamilton et al. 1989, Posfai et al. 1999). Alternative suicide plasmids that contain plasmid R6K-origin of replication (*ori*) can be made in an *E. coli* cell that supports synthesis of the replication initiation protein and

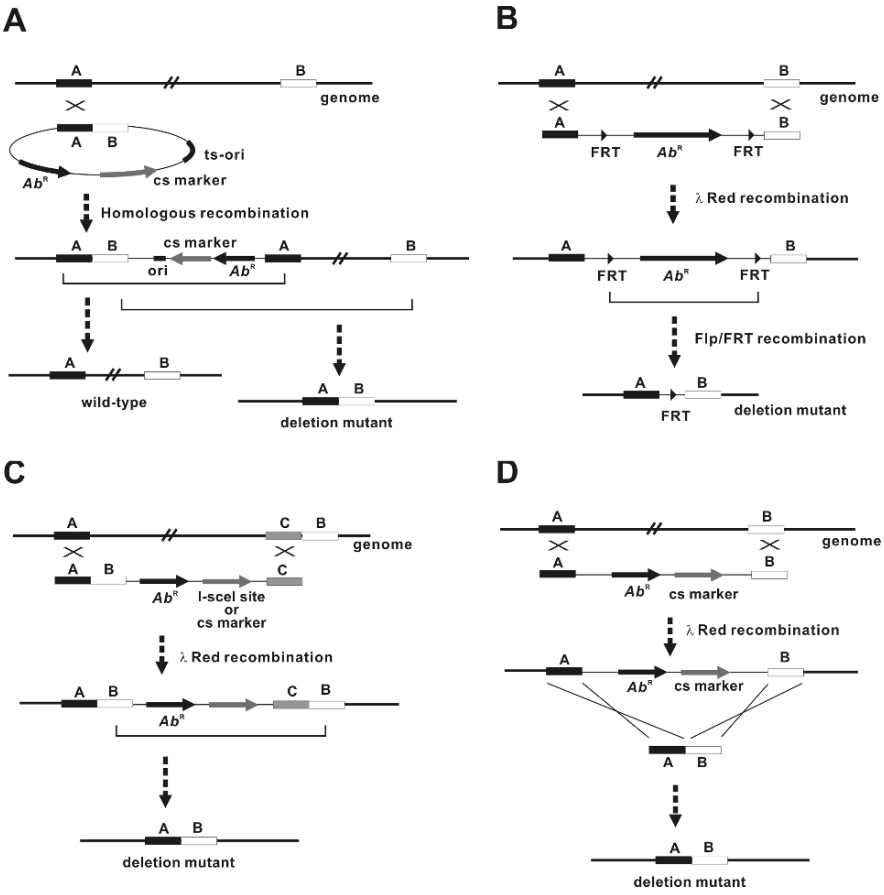

Fig. 2.1 General schemes of the deletion procedures. In all schemes, the chromosomal DNA is shown at the top of the figure, and the targeted plasmid or DNA fragment is shown below the chromosomal DNA schematic. The sequences of the **A** and **B** regions in the chromosome and the targeted plasmid or DNA fragment are the same thus can undergo homologous recombination. (**A**) Protocol for use of the suicide plasmid-mediated deletion system. The plasmid, which carries two PCR-amplified DNA fragments (**A** and **B**, >500 bp) that flank a target genomic region to be deleted, is transformed into *E. coli,* and cells are plated at the nonpermissive temperature (42 °C) of the plasmid replicon. When shifted to the permissive temperature (30 °C), the plasmid is excised from the chromosome by recombination at either A or B. The resistance colonies against the counterselection marker are screened for deletion of the target region (Link et al. 1997, Posfai et al. 1999). (**B**) Overview of the Red-mediated linear DNA deletion method. An *FRT*-flanked antibiotic resistance gene is amplified by PCR and transformed into *E. coli* expressing the Red recombinase. After selection of the antibiotic-resistant transformants, the inserted antibiotic resistance gene is eliminated by the Flp/*FRT* recombination system (Datsenko and Wanner 2000). (**C**) The scarless deletion system. The targeted DNA fragment, which carries the actual post-deletion sequence joints, was amplified by PCR and integrated into the genome by Red recombination. Integration creates a duplication of the segment that flanks the deletion target region. Cleavage of the inserted DNA by meganuclease I-*Sce*I introduces a DSB between the duplicated segments and stimulates their intramolecular recombination, resulting in a scarless genome deletion (Kolisnychenko et al. 2002). (**D**) Deletion of a genomic region by two serial linear DNA recombination events. A DNA

then transferred to a target cell that lacks the replication protein, which permits the plasmid to integrate into the *E. coli* genome (Koob et al. 1994, Posfai et al. 1994, Yoon et al. 1998).

After the integration of a suicide plasmid that contains a homologous arm into the chromosome, the suicide plasmid can be excised out by the second homologous recombination. Depending on the position of the second recombination event that excises the plasmid, the chromosome either retains the wild-type sequence or has a deletion between the target sites (Fig. 2.1A). For easy identification of the resolved products, a counterselection marker, such as *sacB* (sucrose sensitivity) (Dedonder 1966, Gay et al. 1985, Link et al. 1997) or *rpsL* (streptomycin sensitivity) (Hashimoto et al. 2005, Russell et al. 1989, Wang et al. 1993), is integrated in the suicide plasmids. The resolution of the plasmids is thus screened on media supplemented with sucrose or streptomycin, respectively. Another efficient mode of enhancing the plasmid excision step is the introduction of an 18-base pair (bp) meganuclease I-*Sce*I cleavage site in the suicide plasmid (Posfai et al. 1999). Cleavage of the genome at this unique site creates a double-strand break, which simultaneously stimulates recombination and selects for resolution of the integrated plasmid.

Although methods that employ suicide plasmids can be used to delete genomic segments, for each deletion experiment, these procedures require the creation of specific targeting vectors before recombining them into the chromosome.

2.3.2 Linear DNA–Mediated Genomic Deletion

To avoid the inconvenience of constructing the targeting vectors in the suicide plasmid-mediated genomic deletion, linear DNA-mediated genomic deletion approach has been introduced. In *Saccharomyces cerevisiae* and a few naturally competent bacteria, genes or genomic regions can be directly disrupted by transformation with double-stranded DNA (dsDNA) fragments, created with the polymerase chain reaction (PCR), that encode a selectable marker and have only about 50 bp of flanking DNA (called homology arms) that are homologous to the chromosome region of interest. Through homology arm-directed homologous recombination between the DNA fragments and the chromosome, this procedure facilitates the generation of specific chromosomal mutations and thus functional analysis of the genome (Baudin et al. 1993, Oliver et al. 1998, Wilson et al. 1999). In *E. coli*, however, intracellular exonucleases, such as RecBCD, degrade linear DNA (Lorenz and Wackernagel

Fig. 2.1 (continued) fragment that contains an antibiotic resistance gene and a counterselection marker cassette flanked by DNA segments that are homologous to the chromosomal target region was amplified by PCR and inserted into the chromosomal DNA by Red–mediated recombination. The inserted markers are replaced with a linear DNA fragment consisting of only the chromosomal fragment by a second Red recombination (Hashimoto et al. 2005). Ab^R stands for antibiotic resistance marker gene; ori indicates an origin of replication that functions only under permissive conditions; and cs marker indicates a counterselection marker

1994) and inhibit recombination with the PCR products. Linear DNA recombination systems in *E. coli* were developed by finding ways to inhibit the intracellular exonucleases. For example, one such system uses the RecET recombinase to disrupt plasmid-borne genes with linear DNA fragments (Zhang et al. 1998). Other methods make use of the fact that the phage lambda Red (*gam*, *bet*, *exo*) function promotes a greatly enhanced rate of recombination when using linear DNA (Datsenko and Wanner 2000, Murphy 1998, Yu et al. 2000). In the Red system, the Gam protein inhibits the RecBCD nuclease and prevents it from attacking the linear DNA fragments, and Exo (a 5′-to-3′ exonuclease) and Bet (a single-strand DNA binding protein) generate recombination activity for the linear DNA. Because Bet binds to linear DNA strands that are 36 bases in length or longer, it is recommended that, for efficient recombination, the DNA have homology arms of more than 40 bp (Yu et al. 2000).

One method that uses Red-mediated recombination for efficient deletion of specific genomic segments in *E. coli* is that of Datsenko and Wanner (2000). Their basic strategy (Fig. 2.1B) is to replace a chromosomal target with a linear DNA fragment that carries a selectable antibiotic resistance gene flanked by 50-nucleotide (nt) extensions that are homologous to selected sequences in the bacterial chromosome (Fig. 2.1B, segments labeled A and B). This is accomplished by Red-mediated recombination between the *E. coli* genome and these flanking homologies. After selection, the resistance gene can be eliminated by using a helper plasmid that expresses the Flp recombinase, a *S. cerevisiae* enzyme that acts on the directly repeated *FRT* sites that flank the resistance gene (Broach and Hicks 1980). However, because this method is dependent on yeast Flp/*FRT* site-specific recombination to eliminate the selection marker gene, each deletion event leaves behind one *FRT* site in the bacterial chromosome, which interferes with subsequent rounds of chromosomal deletions.

To delete genomic segments without leaving behind remnants of the selection marker, researchers developed a scarless deletion method that combines Red–mediated recombination and double strand break (DSB)–stimulated recombination (Kolisnychenko et al. 2002). For this method, the PCR-generated, linear DNA fragments must be constructed so that they carry the actual post-deletion sequence joints (Fig. 2.1C). Thus, integration of such a DNA fragment by Red recombination creates a duplication of the segment that flanks the deletion target region. Cleavage of the inserted DNA by meganuclease I-*Sce*I introduces a DSB between the duplicated segments and stimulates their intramolecular recombination. Eventually, repair of the DSB by this recombination event results in a scarless genome deletion. In another system for markerless deletion, a counterselection marker, *sacB*, and an I-*Sce*I recognition site were used simultaneously to increase the efficiency of resolution of the inserted markers (Sung et al. 2006).

Hashimoto et al. (2005) developed yet another deletion method that goes through two serial Red–mediated recombination events (Fig. 2.1D). In the first recombination, the targeted genomic region is replaced with the Cm^{R}-*rpsL*-*sacB* (CRS) cassette flanked by DNA fragments that are homologous to the chromosomal target. In this step, the chloramphenicol-resistance gene (Cm^{R}) is used as a positive

selection marker for the deletion mutants. In the second step, the inserted CRS cassette is replaced with a linear DNA fragment that consists of only the chromosomal sequences, which produces a markerless deletion. For this second round of recombination, *rpsL* and *sacB* are used as the two counterselection markers, and sensitivity to chloramphenicol is assessed in the selected transformants.

In addition to the protocols described above, an improved method for the rapid markerless deletion with linear DNA was developed recently by Yu et al. (2008). In this method, the deletion process is mediated by a single helper plasmid that carries genes that encode the Red recombination proteins and the I-*Sce*I nuclease under the control of inducible arabinose and rhamnose promoters, respectively. Genomic deletions are performed by first growing the bacteria in a medium that contains arabinose as the carbon source (to spur synthesis of the Red recombination proteins, which introduces linear DNAs into the bacterial chromosome) and then changing the carbon source in the growth medium from arabinose to rhamnose (to stimulate production of the I-*Sce*I nuclease, which introduces a DSB that stimulates intramolecular recombination). Only two days are required for the deletion of a genomic segment without remaining a selection marker.

Finally, a DNA recombination system mediated by single-strand DNAs (ssDNAs) also has been developed. The ssDNA-binding protein Bet of phage lambda stimulates recombination in chromosomal DNA by using synthetic ssDNAs as short as 30 bases in length. This ssDNA recombination can be used to mutagenize or repair the chromosome with efficiencies that generate up to 6% recombinants among treated cells (Ellis et al. 2001).

2.3.3 Site-Specific Recombination–Mediated Genomic Deletion

Site-specific recombination is a useful genetic tool for deleting undesired DNA sequences and modifying chromosome architecture. The frequently used Cre/*loxP* and Flp/*FRT* site-specific recombination systems share many features. Cre (Flp) is a site-specific recombinase that mediates the recombination of a DNA sequence flanked by two 34-bp *loxP* (*FRT*) sites. A *loxP* (*FRT*) site consists of two 13-bp inverted repeats that flank an 8-bp core region. Intramolecular recombination between the two uni-directionally oriented *loxP* (*FRT*) sites that flank a genomic region of interest results in deletion of the intervening DNA fragment (Broach and Hicks 1980, Sternberg and Hamilton 1981).

To integrate a *loxP* (*FRT*) site and an antibiotic resistance marker into a predetermined chromosomal site, PCR-amplified segments of the selected chromosomal region are cloned into a suicide plasmid that contains the R6K-*ori* and an antibiotic-resistance marker. Independently, another PCR-amplified genomic segment is cloned into another suicide plasmid that carries a different antibiotic-resistance marker. The suicide plasmids are inserted into the genome by homologous recombination at predetermined sites. One of the inserted chromosomal sites is then transferred by P1 transduction into a cell that carries the other insertion (Miller 1992). The predetermined genomic segments, which are flanked by the *loxP* (*FRT*)

sequences, are then excised by the Cre (Flp) recombinase (Posfai et al. 1994, Yoon et al. 1998). For the rapid integration of *loxP* sites into the genome, Fukiya et al. (2004) also reported a method that involves the integration of *loxP*-containing DNA fragments into the two ends of the target genomic sequence using the Red system.

Yu et al. (2002) established a transposon-coupled, site-specific excision system (Fig. 2.2A). Using modified Tn*5* transposons, the authors constructed two large pools of independent transposon-generated mutants and then mapped precisely the chromosomal locations of 800 of these transposons, which carry a *loxP* site and either the chloramphenicol- or kanamycin-resistance markers (Cm^R or Km^R). Yu et al. then combined selected mutants in a single cell by P1 transduction, and large genomic target regions (57–117 kb) flanked by *loxP* sites were excised by Cre/*loxP* recombination. The advantage of this method is the rapid and easy deletion of almost any desired segment of the *E. coli* genome using the transposon-generated mutant libraries without having to go through the time-consuming process of generating

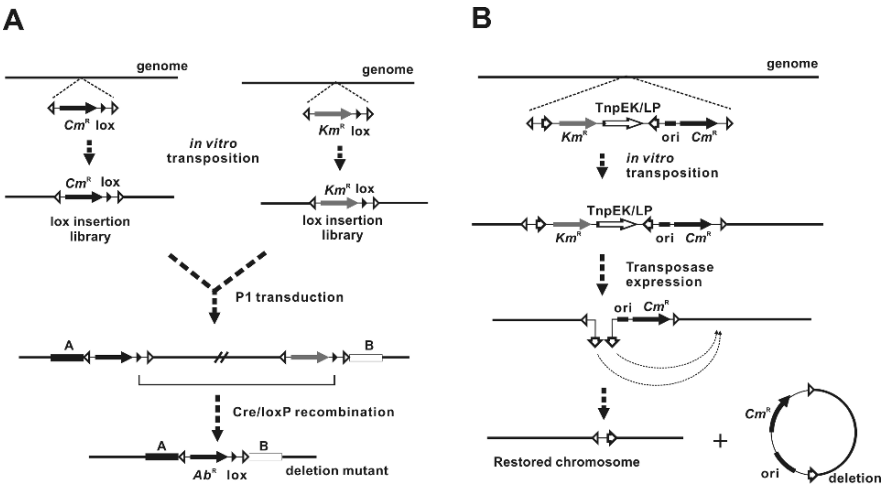

Fig. 2.2 General scheme of the Tn-coupled Cre/*loxP* excision system and random deletion by double transposition. (**A**) Tn*5*-mediated insertions and deletion of the target regions using the Cre/*loxP* excision system. Two modified Tn*5*-transposons are introduced into the *E. coli* genome randomly by *in vitro* transposition. Two mutant strains with a *loxP* site in the same orientation are selected, one from each mutant library, depending on the target region to be deleted. The two selected *loxP* sites are brought, in parallel, into a single strain by P1 transduction. The target region between the two *loxP* sites is deleted by the action of Cre recombinase (Yu et al. 2002). (**B**) The recursive deletion system. This strategy for deletion can be employed after integration of the transposon into the host genome using external transposon ends (*white triangles*). The internal transposon ends (*white arrows*) are used in the second transposition event. The intramolecular transposition leads to the removal of the internal part of the transposon and the deletion of a portion of the chromosome. The addition of a conditional origin of replication allows for capture of the deleted chromosomal material into a complementary self-replicating plasmid (Goryshin et al. 2003). Ab^R stands for antibiotic resistance marker gene; ori indicates an origin of replication; Cm^R *and* Km^R indicates a chloramphenicol resistance and a kanmycin resistance gene, respectively

either customized deletion constructs or PCR products. However, one *loxP* site and an antibiotic resistance marker are retained within the genome after a segment is deleted, which necessitates the use of further engineering techniques to remove the remnants if serial accumulation of deletions is desired. Nevertheless the Cre/*loxP* recombination system can be a useful engineering tool, especially for high-throughput construction of extremely large deletions.

2.3.4 Transposon–Mediated Random Deletion

Goryshin et al. (2003) have described a Tn*5*-based deletion technology (Fig. 2.2B) that uses a composite, linear Tn*5* derivative that carries a replication origin, Km^R and Cm^R selectable markers, Tn*5* transposon end sequences at its 5′ and 3′ termini (external ends), and an internal pair of end sequences from a different transposon (internal ends). In this method, the engineered transposon is introduced into the bacterial host genome by the electroporation of preformed transposome complexes into the cell (Goryshin et al. 2000). The external ends drive integration of this transposon into the host chromosome. A mutant transposase (TnpEK/LP) is expressed from the integrated transposon and then binds to and carries out blunt-end cleavage at the internal ends, which results in (i) the loss of a fragment of the integrated transposon that houses the Km^R gene and (ii) the facilitation of intramolecular transposition. The intramolecular transposition event can create host genome inversions or deletions that begin at the internal ends and extend for varying distances along the host chromosome. Deletions result in loss of the transposon DNA, with the exception of a linker sequence.

Repeated use of this procedure in the same cell, yields a series of random deletions. The average deletion size per round is about 11 kb. The addition of a conditionally active *ori* (one that is induced by IPTG) in the transposon allows for capture of the deleted chromosomal material into a self-replicating plasmid that is complementary to the host chromosome (Fig. 2.2B). Because the transposon integration sites and genomic deletions are random, screening of mutants obtained by this strategy requires significant time and effort. However, this approach allows the deletion of genomic segments in the absence of complete genome sequence information and without prior knowledge of which genes are dispensable for viability.

2.4 Genome Minimization of Microorganisms

E. coli minimal genomes have been constructed by the diverse genomic deletion methods described above, generating bacterial strains that house genomes that are 5–30% smaller than that of a wild-type *E. coli* strain. The genomes of other microorganisms, such as *B. subtilis*, *C. glutamicum*, and the yeasts also have been reduced for the construction of minimal-genome factories (Fujio 2007). The deletion sizes and characteristics of these minimal genomes are summarized in Table 2.2.

Table 2.2 Deleted functions and characteristics of minimal genomes

Strain	Deletion Size	Deleted Functions (D) and Characteristics (C)	Ref.
E. coli			
△20-4	218.7 kb (5.6%)	(D) Random genomic regions	Goryshin et al. 2003
CD△3456	313.1 kb (6.7%)	(D) Nonspecific target regions (C) Presentation of mutually exclusive regions	Yu et al. 2002
MDS12	376.1 kb (8.1%)	(D) K-strain–specific islands (C) No significant difference to the parent cell	Kolisnychenko et al. 2002
MDS43	708.3 kb (15.3%)	(D) K-islands, mobile elements (C) Increased genome stability and electroporation efficiency	Posfai et al. 2006
MGF-01	1.03 Mb (22%)	(D) Nonessential regions without growth deficiency (C) Increased threonine production (2-fold)	Mizoguchi et al. 2007
△16	1.38 Mb (29.7%)	(D) Nonessential genes in the literature (C) Growth deficiency, abnormal nucleoid location	Hashimoto et al. 2005
B. subtilis			
△6	320 kb (7.7%)	(D) Prophage, polyketide synthesis	Westers et al. 2003
MG1M	991 kb (24%)	(D) Prophage, polyketide synthesis, secondary metabolites (C) Growth deficiency	Ara et al. 2007
MGB874	873.5 kb (20.7%)	(C) Increased productivity of extracellular cellulase (1.7-fold) and protease (2.5-fold)	Morimoto et al. 2008
C. glutamicum	190 kb (5.7%)	(D) R-strain–specific regions	Suzuki et al. 2005b
S. pombe	∼500 kb (4%)	(C) Growth at a lower rate	Giga-Hama et al. 2007
S. cerevisiae	531.5 kb (5%)	(C) Increased production of ethanol (1.8-fold) and glycerol (2-fold)	Murakami et al. 2007

2.4.1 E. coli

2.4.1.1 Random Genome Minimization by Transposition

Using the composite Tn*5* derivative (Fig. 2.2B), Goryshin et al. (2003) developed a unique method for random and recursive deletion of genomic segments that can be applied to gene essentiality studies and minimal genome construction. The authors repeated the random integration/deletion process 20 times per cell to reduce the size of the *E. coli* MG1655 genome, generating several different multi-deletion strains (Fig. 2.3). For 4 of these minimized strains, pulsed-field gel electrophoresis (PFGE)

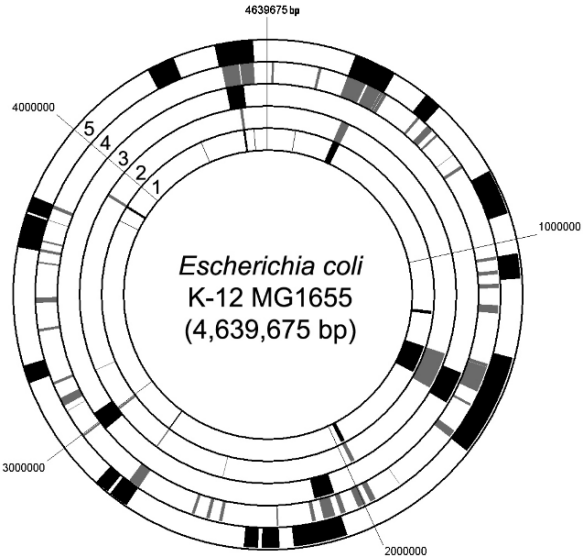

Fig. 2.3 Deletion map of minimal genome *E. coli* strains. Outward from center: 1, set of deletions (209 kb, 5.4% of the wild-type genome) constructed by Goryshin et al. (2003); 2, another set of deletions (219 kb, 5.6% of the wild-type genome) by Goryshin et al. (2003); 3, deletions (313 kb, 6.7% of the wild-type genome) by Yu et al. (2002); 4, deletions (708 kb, 15.3% of the wild-type genome) by Posfai et al. (2006); 5, deletions (1,377 kb, 29.7% of the wild-type genome) by Hashimoto et al. (2005). This figure was produced using the Genome Paint v. 3.0.1 software provided by the National Institute of Genetics, Japan (http://www.shigen.nig.ac.jp/tools/GenomePaint/v3.0)

estimated the total amount of deleted DNA to be between 100 and 262 kb (∼5.6% of the total genome). The locations of the deletions were mapped by microarray hybridization and revealed that, in the two minimized strains with the smallest genomes, only 9 and 11 chromosomal deletions were detected. These findings indicate that some deletions may be too small to detect or occurred within the transposons.

2.4.1.2 Minimization of the *E. coli* Genome Using the Cre/*loxP* Excision System

To demonstrate the feasibility of a combinatorial deletion technique in the identification of essential genes and genome minimization, Yu et al. (2002) performed a 6.7% reduction of the *E. coli* MG1655 genome using the transposon-coupled Cre/*loxP* excision system (Fig. 2.2A) described above. From 13 pairs of genomic deletion targets, six mutant *E. coli* strains that lacked a total of 504.7 kb of DNA (472 genes) were generated. Yu et al. then combined each individual deletion into a single genome by P1 transduction. Repeating the Cre/*loxP* excision procedure on this combined deletion mutant strain, Yu et al. produced an *E. coli* mutant with an additional four large deletions (totaling 313 kb) in the genome (Fig. 2.3). Although

a total of 287 open reading frames (ORFs) had been removed from this ultimate strain, its growth rate in LB medium did not differ significantly from that of the wild-type strain.

Yu et al. also showed that, although many deletions could be successfully combined into a single strain, some deletions that are viable individually are not viable when combined with other deletions. This observation suggests that some mutations are 'mutually exclusive' (Smalley et al. 2003) [also referred to as 'synthetic lethal' (Hartman et al. 2001), or 'mutually essential' (Yu et al. 2006)].

As Yu et al. mentioned in their paper (Yu et al. 2002), in order to generate fully minimized strains and thereby define true minimal essential gene sets, transposon libraries must be expanded to include a total of about 4,000 mapped insertion mutants (saturation of the library). Also, if one wishes to construct a cumulative deletion strain that contains only bacterial DNA, a second recombination strategy, such as with the Red system, is required to eliminate both the selectable marker and the remaining *loxP* site from the deletion mutants. Nevertheless, the availability of mapped mutant pools allows the rapid construction of bacterial strains with virtually any single genomic deletion, which facilitates gene essentiality studies.

2.4.1.3 Genome Minimization Using Scarless Deletion Techniques

Reduced *E. coli* genomes have also been constructed through the generation of sequential large deletions using a combination of Red–mediated and DSB-stimulated recombination as described above (Fig. 2.1C). To stabilize the *E. coli* genome and streamline the bacterial metabolism, researchers have deleted from the genome troublesome DNA sequences and genes that encode proteins that perform unnecessary or unwanted functions, such as K-islands (genomic segments present in K-12, but absent from other *E. coli* strains) (Perna et al. 2001), mobile DNA elements [including insertion sequences (IS)], prophases, transposases, integrases, and site-specific recombinases. These deletions were serially introduced into a single strain by P1 transduction, which generated *E. coli* reduced strain MDS12 (Kolisnychenko et al. 2002), MDS42, and MDS43 (Posfai et al. 2006), which lack 8.1%, 14.3%, and 15.3% of the *E. coli* genome, respectively (Fig. 2.3).

Relative to its parent strain MG1655, MDS42 displays a comparable growth rate and a mutation rate that is reduced by ∼21%. Plasmids prepared from MDS42 cells are free of IS-contamination, and unstable plasmids, even those that carry a toxic chimeric gene, can be recovered in an unaltered form, which illustrates the increased genome stability of the deletion mutant. Also, MDS42 displays a transformation efficiency that is two orders of magnitude higher than that of the wild type. This unexpected increase in the electrocompetence of MDS42 is presumed to stem from uncharacterized synergistic effects (such as improved intracellular access for DNA through depolarized membranes) resulting from the removal of more than 180 genes that encode known or predicted membrane-associated proteins.

Another *E. coli* strain with a highly reduced genome, Delta16, has been characterized by Hashimoto et al. (2005). The size of the Delta16 genome (3.26 Mb) is

29.7% smaller than that of the parent strain (MG1655, 4.64 Mb). For the construction of Delta16, each deletion was generated through two serial lambda Red–mediated recombination events (Fig. 2.1D). The deletions were then accumulated, one at a time, in a single strain, by serial P1 transductions using the original single-deletion strains (Fig. 2.3). Phenotypic analysis revealed that the various large deletion-containing strains that gave rise to Delta16 grew more slowly than did the parental cell. In addition, mutant cells that harbored 13 or more deletions were longer in size and wider than the wild type. Hashimoto et al. (2005) also noticed that, while the parental cells contained one or two nucleoids localized at the midcell, or at the $^{1}/_{4}$ and $^{3}/_{4}$ positions, most of the mutant cells had four or more nucleoids that were localized at the periphery of the cell, near the envelope.

Recently, Kato and Hashimoto (2007) used consecutive genome deletion to discover that the origin of DNA replication is the only unique *cis*-acting DNA sequence in the *E. coli* genome that is necessary for survival.

To construct the minimal *E. coli* genome for industrial applications, Mizoguchi et al. (2007a) selected for deletion regions that were not expected to affect the growth or basic metabolism of the bacteria. Ninety-five regions of the *E. coli* genome, with a total size of 1.8 Mb, were deleted independently using markerless deletion mediated by the *sacB-cat* cassette (Mizoguchi et al. 2007b), and the individual deletions were transferred to a single chromosome by P1 transduction. Throughout the genome-size reduction process, Mizoguchi et al. assessed the growth properties (in minimal medium) of each intermediate strain and selected only those strains that displayed no growth-deficiency for subsequent deletions. In the final mutant strain, called MGF-01, the genome size was reduced by 22% (1.03 Mb). During the exponential growth phase, MGF-01 displayed doubling times in M9 minimal medium that were as fast as that of the wild-type strain, and the final cell density reached by MGF-01 was 1.5 times higher than that of the wild-type strain. When the genetic circuit for threonine production ($\triangle met$::*thrABC-Cm*^{R}) was integrated into the chromosomes of MGF-01 and its wild-type parental strain, the resulting wild-type- and MGF-01-based strains produced 5 and 10 g/l of threonine in 48 h, respectively.

2.4.2 Other Microorganisms

B. subtilis *B. subtilis*, one of the most extensively studied model microorganisms, displays a superior ability to produce various secretory enzymes. Many industrial scientists have exploited this capability in the production of a variety of useful enzymes (Westers et al. 2004). The 4.2-Mb *B. subtilis* genome contains 10 horizontally acquired prophage (SPβ and PBSX) and prophage-like (pro1-7 and skin) sequences (Kunst et al. 1997). In addition, 2.8% of the genome encompasses two large operons that produce secondary metabolites (pks and pps).

Using a suicide plasmid-based chromosomal integration-excision system (Leenhouts et al. 1996), Westers et al. (2003) have produced a *B. subtilis* Delta6 mutant strain, with a 7.7% reduction in the genome (0.53 Mb), by deleting two prophage (SPβ, PBSX) and three prophage-like sequences (pro1, pro6, skin) as

well as one of the secondary metabolite operon (pks) from wild-type *B. subtilis* 168. However, phenotypic characterization of the Delta6 cells disclosed no unique properties, relative to wild-type 168 cells. Recently, Ara et al. (2007) deleted, from the *B. subtilis* genome, all prophage and prophage-like sequences, with the exception of pro7, as well as the pks and pps operons, which resulted in a *B. subtilis* strain (MG1M) that lacked 0.99 Mb of the wild-type genome. However, MG1M strain displays unstable phenotypes with respect to growth rate, cell morphology, and recombinant protein production after successive culture, making it inappropriate for further study.

Another *B. subtilis* minimal genome strain, MGB874, was created recently by introducing deletions, step-by-step, into 28 regions in which single deletions do not affect cell growth (Morimoto et al. 2008). A total of 873.5 kb of DNA (20.7% of the genome) was deleted from wild-type *B. subtilis*. In order to assess the ability of MGB874 to synthesize and secrete proteins, wild-type *B. subtilis* and MGB874 were transformed with plasmids that encoded extracellular cellulase and protease enzymes, and protein production was measured. Cellulase and protease enzyme production was enhanced 1.7- and 2.5-fold, respectively, in MGB874, relative to the wild-type strain.

C. glutamicum The bacterium *C. glutamicum* is used widely for the industrial production of amino acid and organic acids. Therefore, Suzuki et al. (2005a, b, c, and d) used modified Cre/*loxP* recombination to generate a minimal-genome *C. glutamicum* strain that lacked 190 kb of the wild-type genome that encodes a total of 188 ORFs. This deletion mutant exhibits normal growth under standard laboratory conditions. In addition, Tsuge et al. (2007) have generated 42 *C. glutamicum* mutants (with deletions of 0.2–186 kb) using a deletion method similar to that described by Yu et al. (2002), which combines an transposon and the Cre/*loxP* excision system. Tsuge et al. showed that a total of 393.6 kb (11.9%) of the *C. glutamicum* R genome is nonessential for growth under standard laboratory conditions.

Schizosaccharomyces pombe In the fission yeast *S. pombe*, Hirashima et al. (2006) reported a method for the deletion of a large genomic region using homologous recombination between the chromosome and a fragment of linear DNA. Giga-Hama et al. (2007) used this method to create an *S. pombe* mutant dedicated to heterologous protein production. The authors deleted a total of more than 500 kb from a wild-type *S. pombe* strain by repeating the deletion procedure multiple times. Although the authors succeeded in developing a viable strain of *S. pombe* with a minimal genome, the phenotypic characteristics of this organism have not yet been reported.

S. cerevisiae To decipher the number of genes required for growth and the genome organization responsible for ethanol production, various segments of the *S. cerevisiae* genome were deleted, and a viable mutant was created that had lost about 5% (531.5 kb) of the wild-type genome (Murakami et al. 2007). This mutant displays an increase in ethanol (1.8-fold) and glycerol (2-fold) production, relative to the wild type, while also exhibiting levels of resistance to various stresses (heat-shock, acidic or alkaline growth conditions, the presence of 7.5% ethanol, 1 M NaCl, or 1.5 M sorbitol in the growth medium) that are nearly equivalent to those of the parental strain.

2.5 Conclusions and Perspectives

In order to realize our dream of creating ideal, robust host organisms for novel uses that benefit humankind, scientists must first unravel microbial genomes to determine the minimal components that are sufficient for life in specific controlled environments. Researchers are tackling this task by using a comprehensive approach based on computational, experimental, and literature-based studies. For some organisms, the removal of all genomic regions save those that are part of the defined core element has been accomplished, without any deleterious effect on basic cell physiology, by adapting well-characterized recombination systems to the deletion of large genomic segments. Indeed, some of the minimal *E. coli* genomes constructed to date even show improved genome stability and/or increased production of industrial products, relative to wild-type strains.

With the aid of systems biology and synthetic biology approaches, the minimal *E. coli* genomes now in existence may be reduced further to produce a genome that houses the absolute minimal number of genes essential for life. This would represent an important step toward acquiring the ability to genetically engineer organisms (novel and existing) knowing only the sequences of their genomes. Minimal genome research also may provide insights into the origins of life; bacterial evolution; regulation of microbial metabolism; and the genomes of more complex modern organisms. Finally, minimized *E. coli* genomes can lead to the construction of elite, custom-designed bioengines with a plethora of practical and commercial applications.

Bacteria are now commonly engineered to produce useful products, ranging from industrial chemicals to pharmaceutical proteins. Therefore, the first benefits of minimal-genome research likely will be in microbial engineering. A minimal-genome organism might require less energy to produce the same amount of a given protein or produce fewer waste products that contaminate the desired product. A minimal-genome organism also may be used as the basis for novel bioengines created to perform specific tasks, such as the breakdown of environmental toxins, the neutralization of toxic spills, or the creation of renewable energy.

References

Akerley BJ, Rubin EJ, Novick VL et al. (2002) A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proc Natl Acad Sci U S A 99(2): 966–71

Ara K, Ozaki K, Nakamura K et al. (2007) *Bacillus* minimum genome factory: effective utilization of microbial genome information. Biotechnol Appl Biochem 46(Pt 3):169–78

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

Baudin A, Ozier-Kalogeropoulos O, Denouel A et al. (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. Nucleic Acids Res 21(14):3329–30

Blattner FR, Plunkett G, 3rd, Bloch CA et al. (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277(5331):1453–74

Broach JR, Hicks JB (1980) Replication and recombination functions associated with the yeast plasmid, 2 mu circle. Cell 21(2):501–8
Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science 297(5583):1016–8
Cho MK, Magnus D, Caplan AL et al. (1999) Policy forum: genetics. Ethical considerations in synthesizing a minimal genome. Science 286(5447):2087, 2089–90
Cooper VS, Schneider D, Blot M et al. (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183(9): 2834–41
Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97(12):6640–5
Dedonder (1966) Levansucrase from *Bacillus subtilis*. Methods Enzymol 8:500–5
Edwards JS, Palsson BO (2000) Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. BMC Bioinformatics 1:1
Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4(6):457–69
Ellis HM, Yu D, DiTizio T et al. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98(12):6742–6
Feher T, Papp B, Pal C et al. (2007) Systematic genome reductions: theoretical and experimental approaches. Chem Rev 107(8):3498–513
Fleischmann RD, Adams MD, White O et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269(5223):496–512
Forster AC, Church GM (2006) Towards synthesis of a minimal cell. Mol Syst Biol 2:45
Forster AC, Church GM (2007) Synthetic biology projects in vitro. Genome Res 17(1):1–6
Fraser CM, Gocayne JD, White O et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270(5235):397–403
Fujio T (2007) Minimum genome factory: innovation in bioprocesses through genome science. Biotechnol Appl Biochem 46(Pt 3):145–6
Fukiya S, Mizoguchi H, Mori H (2004) An improved method for deleting large regions of *Escherichia coli* K-12 chromosome using a combination of Cre/loxP and lambda Red. FEMS Microbiol Lett 234(2):325–31
Gay P, Le Coq D, Steinmetz M et al. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J Bacteriol 164(2):918–21
Gerdes SY, Scholle MD, Campbell JW et al. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J Bacteriol 185(19):5673–84
Giaever G, Chu AM, Ni L et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418(6896):387–91
Gibson DG, Benders GA, Andrews-Pfannkoch C et al. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319(5867): 1215–20
Giga-Hama Y, Tohda H, Takegawa K et al. (2007) *Schizosaccharomyces pombe* minimum genome factory. Biotechnol Appl Biochem 46(Pt 3):147–55
Goryshin IY, Jendrisak J, Hoffman LM et al. (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol 18(1):97–100
Goryshin IY, Naumann TA, Apodaca J et al. (2003) Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis. Genome Res 13(4):644–53
Hamilton CM, Aldea M, Washburn BK et al. (1989) New method for generating deletions and gene replacements in *Escherichia coli*. J Bacteriol 171(9):4617–22
Hartman JL, 4th Garvik B, Hartwell L (2001) Principles for the buffering of genetic variation. Science 291(5506):1001–4
Hashimoto M, Ichimura T, Mizoguchi H et al. (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. Mol Microbiol 55(1):137–49

Hasty J, McMillen D, Isaacs F et al. (2001) Computational studies of gene regulatory networks: in numero molecular biology. Nat Rev Genet 2(4):268–79

Hayashi K, Morooka N, Yamamoto Y et al. (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol Syst Biol 2:7

Herring CD, Raghunathan A, Honisch C et al. (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38(12):1406–12

Hirashima K, Iwaki T, Takegawa K et al. (2006) A simple and effective chromosome modification method for large-scale deletion of genome sequences and identification of essential genes in fission yeast. Nucleic Acid Res 34(2)11

Hutchison CA, Peterson SN, Gill SR et al. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286(5447):2165–9

Ishii N, Nakahigashi K, Baba T et al. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 316(5824):593–7

Itaya M (1995) An estimation of minimal genome size required for life. FEBS Lett 362(3): 257–60

Itaya M, Fujita K, Kuroki A et al. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. Nat Methods 5(1):41–3

Jacobs MA, Alwood A, Thaipisuttikul I et al. (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 100(24):14339–44

Joyce AR, Reed JL, White A et al. (2006) Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. J Bacteriol 188(23):8259–71

Kato J, Hashimoto M (2007) Construction of consecutive deletions of the *Escherichia coli* chromosome. Mol Syst Biol 3:132

Kibota TT, Lynch M (1996) Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature 381(6584):694–6

Kitano H (2002) Computational systems biology. Nature 420(6912):206–10

Knuth K, Niesalla H, Hueck CJ et al. (2004) Large-scale identification of essential Salmonella genes by trapping lethal insertions. Mol Microbiol 51(6):1729–44

Kobayashi K, Ehrlich SD, Albertini A et al. (2003) Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A 100(8):4678–83

Kolisnychenko V, Plunkett G, 3rd, Herring CD et al. (2002) Engineering a reduced *Escherichia coli* genome. Genome Res 12(4):640–7

Koob MD, Shaw AJ, Cameron DC (1994) Minimizing the genome of *Escherichia coli*. Motivation and strategy. Ann NY Acad Sci 745:1–3

Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1(2):127–36

Koonin EV, Wolf YI, Karev GP (2002) The structure of the protein universe and genome evolution. Nature 420(6912):218–23

Kunst F, Ogasawara N, Moszer I et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390(6657):249–56

Leenhouts K, Buist G, Bolhuis A et al. (1996) A general system for generating unlabelled gene replacements in bacterial chromosomes. Mol Gen Genet 253(1–2):217–24

Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. J Bacteriol 179(20):6228–37

Lorenz MG, Wackernagel W (1994) Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev 58(3):563–602

Luisi PL (2002) Toward the engineering of minimal living cells. Anat Rec 268(3):208–14

Maniloff J (1996) The minimal cell genome: “on being the right size”. Proc Natl Acad Sci U S A 93(19):10004–6

Miller JH (1992) A short course in bacterial genetics: A laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, New York

Mizoguchi H, Mori H, Fujio T (2007a) *Escherichia coli* minimum genome factory. Biotechnol Appl Biochem 46(Pt 3):157–67

Mizoguchi H, Tanaka-Masuda K, Mori H (2007b) A simple method for multiple modification of the *Escherichia coli* K-12 chromosome. Biosci Biotechnol Biochem 71(12): 2905–11

Morimoto T, Kadoya R, Endo K et al. (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. DNA Res 15(2):73–81

Murakami K, Tao E, Ito Y et al. (2007) Large scale deletions in the Saccharomyces cerevisiae genome create strains with altered regulation of carbon metabolism. Appl Microbiol Biotechnol 75(3):589–97

Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J Bacteriol 180(8):2063–71

Mushegian A (1999) The minimal genome concept. Curr Opin Genet Dev 9(6):709–14

Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A 93(14):10268–73

Oliver SG, Winson MK, Kell DB et al. (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16(9):373–8

Park JH, Lee KH, Kim TY et al. (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. Proc Natl Acad Sci U S A 104(19):7797–802

Perna NT, Plunkett G, 3rd, Burland V et al. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409(6819):529–33

Posfai G, Kolisnychenko V, Bereczki Z et al. (1999) Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. Nucleic Acids Res 27(22): 4409–15

Posfai G, Koob M, Hradecna Z et al. (1994) In vivo excision and amplification of large segments of the *Escherichia coli* genome. Nucleic Acids Res 22(12):2392–8

Posfai G, Plunkett G, 3rd, Feher T et al. (2006) Emergent properties of reduced-genome *Escherichia coli*. Science 312(5776):1044–6

Razin S (1997a) Comparative genomics of mycoplasmas. Wien Klin Wochenschr 109(14–15): 551–6

Razin S (1997b) The minimal cellular genome of mycoplasma. Indian J Biochem Biophys 34(1–2):124–30

Richmond CS, Glasner JD, Mau R et al. (1999) Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res 27(19):3821–35

Riley M, Abe T, Arnaud MB et al. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot – 2005. Nucleic Acids Res 34(1):1–9

Riley M, Serres MH (2000) Interim report on genomics of *Escherichia coli*. Annu Rev Microbiol 54:341–411

Russell CB, Dahlquist FW (1989) Exchange of chromosomal and plasmid alleles in *Escherichia coli* by selection for loss of a dominant antibiotic sensitivity marker. J Bacteriol 171(5):2614–8

Salama NR, Shepherd B, Falkow S (2004) Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. J Bacteriol 186(23):7926–35

Smalley DJ, Whiteley M, Conway T (2003) In search of the minimal *Escherichia coli* genome. Trends Microbiol 11(1):6–8

Smith HO, Hutchison CA, 3rd, Pfannkoch C et al. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A 100(26):15440–5

Sternberg N, Hamilton D (1981) Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. J Mol Biol 150(4):467–86

Sung BH, Lee CH, Yu BJ et al. (2006) Development of a biofilm production-deficient *Escherichia coli* strain as a host for biotechnological applications. Appl Environ Microbiol 72(5): 3336–42

Suzuki N, Nonaka H, Tsuge Y et al. (2005a) Multiple large segment deletion method for *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 69(2):151–161

Suzuki N, Nonaka H, Tsuge Y et al. (2005b) New multiple-deletion method for the *Corynebacterium glutamicum* genome, using a mutant lox sequence. Appl Environ Microbiol 71(12):8472–80

Suzuki N, Okayama S, Nonaka H et al. (2005c) Large-scale engineering of the *Corynebacterium glutamicum* genome. Appl Environ Microbiol 71(6):3369–72

Suzuki N, Tsuge Y, Inui M et al. (2005d) Cre/loxP-mediated deletion system for large genome rearrangements in *Corinebacterium glutamicum*. Appl Microbiol Biotechnol 67(2):225–33

Suzuki N, Okai N, Nonaka H et al. (2006) High-throughput transposon mutagenesis of *Corynebacterium glutamicum* and construction of a single-gene disruptant mutant library. Appl Environ Microbiol 72(5):3750–5

Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409(6818):387–90

Tao H, Bausch C, Richmond C et al. (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J Bacteriol 181(20):6425–40

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278(5338):631–7

Tsuge Y, Suzuki N, Inui M et al. (2007) Random segment deletion based on IS31831 and Cre/loxP excision system in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 74(6):1333–41

Tumpey TM, Basler CF, Aguilar PV et al. (2005) Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310(5745):77–80

Wang G, Blakesley RW, Berg DE et al. (1993) pDUAL: a transposon-based cosmid cloning vector for generating nested deletions and DNA sequencing templates in vivo. Proc Natl Acad Sci U S A 90(16):7874–8

Welch RA, Burland V, Plunkett G, 3rd et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A 99(26):17020–4

Westers H, Dorenbos R, van Dijl JM et al. (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol Biol Evol 20(12):2076–90

Westers L, Westers H, Quax WJ (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. Biochim Biophys Acta 1694(1–3):299–310

Wilson RB, Davis D, Mitchell AP (1999) Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. J Bacteriol 181(6):1868–74

Yoon YG, Cho JH, Kim SC (1998) Cre/loxP-mediated excision and amplification of large segments of the *Escherichia coli* genome. Genet Anal 14(3):89–95

Yu BJ, KK, Lee JH (2008) Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. Nucleic Acids Research 36:84

Yu BJ, Sung BH, Koob MD et al. (2002) Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat Biotechnol 20(10):1018–23

Yu BJ, Sung BH, Lee JY et al. (2006) sucAB and sucCD are mutually essential genes in *Escherichia coli*. FEMS Microbiol Lett 254(2):245–50

Yu D, Ellis HM, Lee EC et al. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci U S A 97(11):5978–83

Zhang Y, Buchholz F, Muyrers JP et al. (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet 20(2):123–8