Minimization of the *Escherichia coli* genome using a Tn*5*-targeted Cre/*loxP* excision system

Byung Jo Yu[1]*, Bong Hyun Sung[1]*, Michael D. Koob[2], Choong Hoon Lee[1], Jun Hyoung Lee[1], Won Sik Lee[1], Mi Sun Kim[3], and Sun Chang Kim[1†]

Published online: 16 September 2002, doi:10.1038/nbt740

An increasing number of microbial genomes have been completely sequenced, and functional analyses of these genomic sequences are under way. To facilitate these analyses, we have developed a genome-engineering tool for determining essential genes and minimizing bacterial genomes. We made two large pools of independent transposon mutants in *Escherichia coli* using modified Tn*5* transposons with two different selection markers and precisely mapped the chromosomal location of 800 of these transposons. By combining a mapped transposon mutation from each of the mutant pools into the same chromosome using phage P1 transduction and then excising the flanked genomic segment by Cre-mediated *loxP* recombination, we obtained *E. coli* strains in which large genomic fragments (59–117 kilobases) were deleted. Some of these individual deletions were then combined into a single "cumulative deletion strain" that lacked 287 open reading frames (313.1 kilobases) but that nevertheless exhibited normal growth under standard laboratory conditions.

The complete genomic sequences of >135 microorganisms represent an important resource in the efforts of biologists to achieve a comprehensive understanding of cellular life at the molecular level (see TIGR Microbial Database, The Institute for Genomic Research, Rockville, MD, February 8, 2002; http://www.tigr.org/ tdb/mdb/mdbcomplete. html/). Identifying the genes and other DNA sequences that are most essential for maintaining and replicating a free-living unicellular organism is a logical first step in this process, and a number of groups have begun to pursue this goal[1–6].

Escherichia coli has been the preeminent prokaryotic organism in research laboratories since the beginning of molecular biology and is arguably the most completely characterized single-cell life form[7]. Yet we still do not know the functions of >1,000 of its 4,392 predicted genes (Supplementary data are available at Summary of *E. coli*, Strain K-12, version released: http://www.ecocyc.org:1555/ECOLI/organism-summary?object=ECOLI.) Functional analyses have shown that *E. coli* cells grown under a given condition use only a fraction of their genes[8]. As one of us (M.D.K.) has proposed[9], deletion of genes that are nonessential under a given set of growth conditions should identify a minimized set of essential *E. coli* genes and DNA sequences.

The most widely used methods for deleting genes in the *E. coli* genome are the site-specific recombination systems, Flp/*FRT*, Cre/*loxP*, and λ-Red[10–14]. However, for each deletion experiment these methods require either the creation of targeting vectors or complex PCR experiments. To delete many nonessential genes scattered along the chromosome, it is advisable to carry out numerous deletion experiments to generate a custom-designed strain. The need for many deletion experiments led us to develop a rapid and efficient deletion procedure.

Here we describe a combinatorial deletion technique that combines Tn*5* transposon mutagenesis, the Cre/*loxP* excision system, and phage P1 transduction (Fig. 1). In this method, a modified Tn*5* transposon is inserted randomly into genes on the chromosome together with a *loxP* recombination site. Two *loxP* sites are brought in parallel into a single strain by phage P1 transduction, and the selected region between the two *loxP* sites is deleted by the action of Cre recombinase, an approach similar to that of ref. 15. Our Tn*5*-targeted Cre/*loxP* excision system is a rapid and efficient genome-engineering tool that does not require the construction of targeting vectors for each deletion experiment. To illustrate the utility of our approach, we generated numerous *E. coli* deletion strains in which specific chromosomal regions of 59–117 kilobases were deleted. Some of the individual deletion regions of these strains were then combined into a single "cumulative-deletion strain" in which a total of 108 open reading frames (ORFs) of known function and 179 ORFs of unknown function were deleted. Despite the deletion of ~313.1 kilobases from its genome, this strain still exhibited normal growth under standard laboratory conditions.

Results

Random transposition of transposons. After transposition with transposons TnKloxP and TnCloxP, 400 *E. coli* mutants with a TnKloxP on the chromosome were selected on Luria–Bertani (LB) plates containing kanamycin (Km). Similarly, 400 mutants, each containing a TnCloxP, were obtained on LB plates containing chloramphenicol (Cm). Southern blot analysis showed that most of the selected mutants (97%) had only one transposon insertion on the chromosome under our experimental conditions (data not shown). The DNA regions flanking the insertion transposon were sequenced to identify the position and direction of the inserted

[1]*Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Taejon, 305-701, Korea.* [2]*Institute of Human Genetics, University of Minnesota, Minneapolis, MN 55455, USA.* [3]*Biomass Team, Korea Institute of Energy Research, Taejon 305-343, Korea.* **These two authors contributed equally to this work.* [†]*Corresponding author (sckim@mail.kaist.ac.kr).*

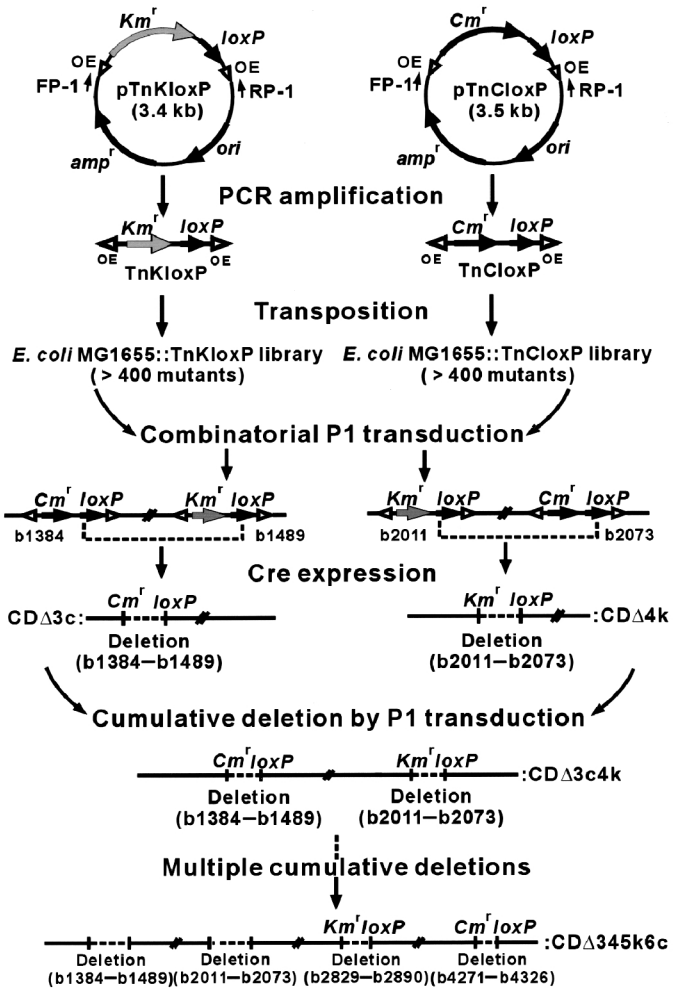

Figure 1. Tn*5*-mediated insertions of the Cre/*loxP* excision system and deletion of the target regions. Two modified Tn*5*-transposons, TnKloxP and TnCloxP, are generated from pTnKloxP and pTnCloxP, respectively, by PCR amplification using primers FP-1 and RP-1. These transposons are introduced into *E. coli* chromosome randomly, producing two groups of mutant libraries (*E. coli* MG1655::TnKloxP and MG1655::TnCloxP). 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 selected two *loxP* sites are brought in parallel into a single strain by phage P1 transduction. The selected region between the two *loxP* sites is deleted by the action of Cre. The individual deletion regions are then combined into a single "cumulative-deletion strain" (refer to Experimental Protocol and Table 1). OE denotes outer-end transposase recognition sequence.

loxP site in each of the mutants. The *loxP* sites were oriented in both directions on the chromosome, and 70% of the inserted transposons were located at hypothetical or unknown genes, whereas 30% were inserted in known genes. Our two *E. coli* mutant libraries totaling 800 strains provide a valuable pool of mutant strains for the combinatorial deletion of the *E. coli* genome. The location and direction of the *loxP* site inserted in the 800 mutant strains are shown on the physical maps of the *E. coli* genome in Figure 2.

Combinatorial deletions. Depending on the genes to be deleted, we chose a pair of mutant strains, one from the TnK*loxP* group and one from the TnC*loxP* group. Using phage P1 transduction, we constructed *Cm*^{r}-*Km*^{r} double-resistant strains containing two *loxP* sites in tandem (Fig. 3A). In choosing the mutant strains from the Tn*5*-mutant libraries for genomic deletion, we considered the position and direction of the two *loxP* sites and the possible presence of essential genes in the region flanked by the two *loxP* sites. Subsequent expression of Cre recombinase induces recombination between the *loxP* sites, resulting in excision of the flanked region.

We started with 13 pairs of insertion mutants selected from the two Tn*5*-mutant libraries for the deletion of *E. coli* genomic fragments in the range of 59–192 kilobases. However, we succeeded in generating only six chromosomal deletion (CD) mutants: CDΔ1k, CDΔ2c, CDΔ3c, CDΔ4k, CDΔ5k, and CDΔ6c (Table 1). The failed deletion regions are also listed in Table 1. The failure of the deletion in the remaining seven pairs suggests that one or more essential genes are present in the selected targeted region. We therefore segmented the failed deletion regions into smaller regions using our Tn*5*-mutant libraries. All deletions were verified by PCR using primers specific to the endpoints of each deletion (Fig. 3A, B). We generated six strains lacking a total of 504.7 kilobases (472 genes) that did not seem to be essential for cell survival in the laboratory.

Cumulative deletions. Next we constructed four cumulative-deletion mutants: CDΔ345k6c, CDΔ3c4k56, CDΔ34k56c, and CDΔ3c45k6. These strains have a total of 313.1 kilobases of deletion sequences in four separate deletion regions (DR) on the same chromosome and are identical except for the location of *Km*^{r} and *Cm*^{r} (Table 1). We verified the successful deletions of these four separate regions in the cumulative strains by multiplex PCRs using locus-specific primers. PCR bands, which indicate the deletion of the regions flanked by two *loxP* sites, are shown in Figure 3B. The PCR products observed on the gel for the cumulative-deletion strains reflect the location of *Cm*^{r} and *Km*^{r}. To confirm the absence of the deleted regions from the deletion strains, we conducted multiplex PCRs using two pairs of primers specific to the internal genes of each deletion region. The PCR products shown in *E. coli* MG1655 disappeared completely in the PCRs of deletion strains, indicating the absence of the deleted genomic segments in each deletion strain (Fig. 3C).

The mutant strains with four deletions have growth rates comparable to that of the wild-type (wt) strain in LB broth (data not shown). From this result, we conclude that 313.1 kilobases of genomic sequence deleted in these cumulative-deletion strains are dispensable for *E. coli* MG1655 grown on LB broth and that these regions do not contain genes essential for cell survival.

As shown in Table 2, a total of 287 ORFs, including 179 unknown genes and 108 known genes, were deleted. These deletion regions contain some clustered genes. For instance, *hisLGDCBHAFI* (histidine biosynthesis), *rfcglfrfbXCADBgalF* (O-antigen biosynthesis), and *wcaMLKwzxCwcaJcpsBGwcaIHGgmdwcaFEDCBAJwzcwzbwza* (colanic acid biosynthesis) were removed by deletion of DR4(b2011–b2073) of CDΔ4k. In addition, *fecEDCBARI* (iron dicitrate ABC transporter) and *fimBEAICDFGH* (type 1 fimbriae biosynthesis), located in DR6(b4271–b4326) of CDΔ6c, were also eliminated. The function of the deleted *fecEDCBARI* might be complemented by its paralog or by clustered genes playing a similar role, such as *fepAECGDB-fes-entDFCEBA*, *feoB* ferrous transporter, *fhuACDB* ferrichrome, and the *ftnEBA* uptake system. However, the gene for type 1 fimbriae biosynthesis might be dispensable in *E. coli*, although it may play an important role in pathogenic *E. coli* strains. The dispensability of genes involved in the biosynthesis of fimbriae was also confirmed by Tn mutagenesis[16]. Some phage-derived genes scattered in each deletion region also appear to be nonessential.

In addition to the above cumulative deletions, we also attempted to combine other individual deletions into the same chromosome. However, we were not able to combine DR1(b0273–b0369) of CDΔ1k into the deletion strain CDΔ2c with

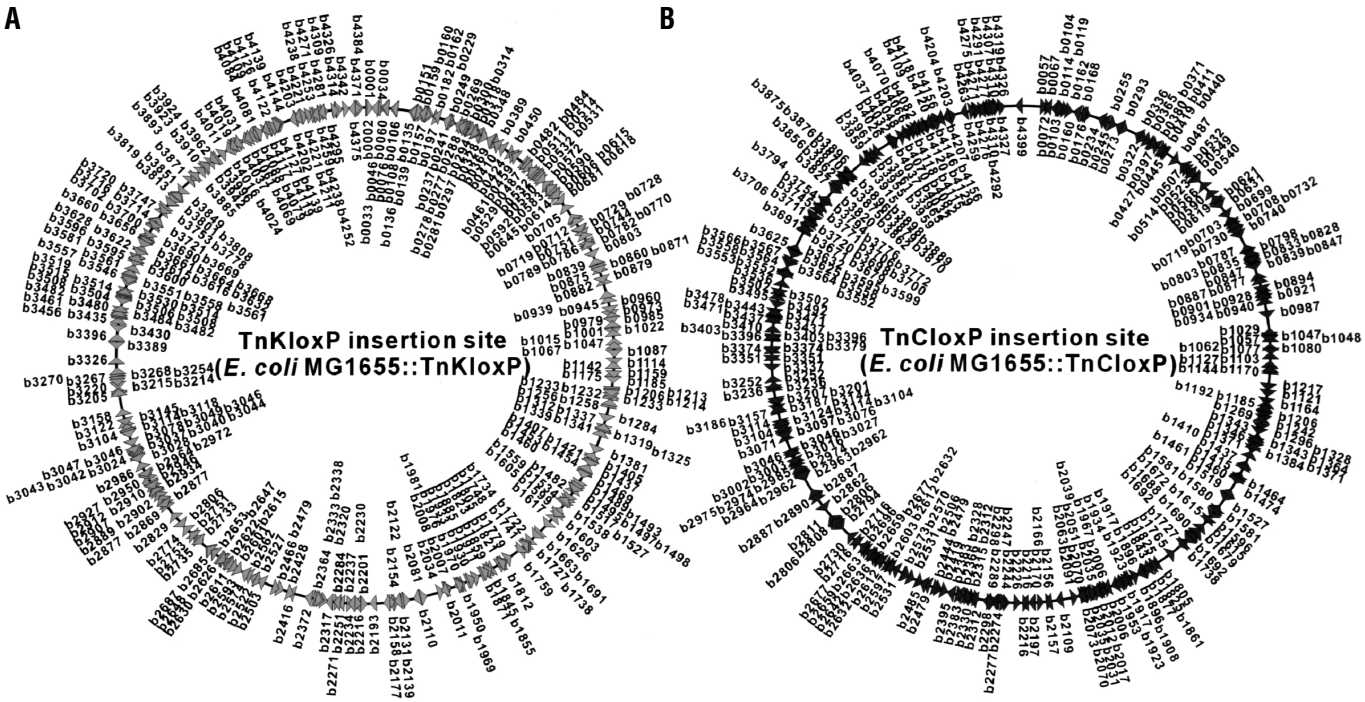

Figure 2. Physical maps of the *Cm*^{r} and *Km*^{r} inserts. Physical maps of *E. coli* MG1655 targeted with TnKloxP (A) and with TnCloxP (B) are drawn in a circular chromosome. The arrows in the *E. coli* MG1655::TnKloxP and MG1655::TnCloxP libraries indicate the insertion sites of TnKloxP and TnCloxP. The positions of the inserted transposons are indicated as Blattner numbers. The Blattner numbers located inside and outside the circle represent leftward and rightward orientations of the inserted *loxP*, respectively. We have constructed 800 *E. coli* insertion mutants, and the number will be increased to >4,000 (an average of one inserted *loxP* site every 1.15 kb on the chromosome). Information concerning the strains in these libraries is also available on our web page (http://bioneer.kaist.ac.kr/labs/mbt/), and requests for strains can be made through this site.

DR2(b0532–b0619), even though the DR1(b0273–b0369) was successfully combined into the deletion strain CDΔ6c with DR6(b4271–b4326). Another failure of the cumulative deletions was with DR6(b4271–b4326) of CDΔ6c. We failed to combine the DR6(b4271–b4326) into the deletion strain CDΔ2c with DR2(b0532–b0619), even though the DR6(b4271–b4326) was successfully combined into other deletion strains CDΔ1k, CDΔ4k, and CDΔ5k. We are currently segmenting the large deletion regions that cannot be combined into a cumulative-deletion strain into smaller regions using our Tn*5*-mutant libraries to identify the essential regions.

Discussion

In this study, we describe an efficient method for functionally analyzing the *E. coli* genome. The advantage of our method is that we have generated two large *E. coli* transposon libraries with *loxP* sites distributed almost evenly on the chromosome. We can therefore select pairs of mutants from these libraries to delete quickly and easily almost any desired segment of the *E. coli* genome, thereby avoiding the time-consuming process of generating either customized deletion constructs or PCR products. Deletion experiments using these transposon libraries can be directly applied to minimization of the *E. coli* genome and will be especially useful for identifying genomic sequences that are important for cell survival. Our Tn*5*-mutant libraries and deletion strains could also be applied to improving industrial *E. coli* strains by restructuring metabolic and cellular pathways and by systematically eliminating genes that complicate the production of valuable industrial materials in *E. coli*.

Table 1. Deletion strains constructed by the Tn*5*-targeted Cre/*loxP* excision system

Combinatorial deletions		Cumulative deletions	
Deletion mutants[a,b]	Target regions/size	Deletion mutants	DR/size[c]
CDΔ1k	b0273 (TnC)–b0369 (TnK)/98.4 kb	CDΔ1k6c	DR1 + DR6 /157.4 kb
CDΔ2c	b0532 (TnK)–b0619 (TnC)/93.2 kb	CDΔ3c4k	DR3 + DR4 /187.4 kb
ND	b0703 (TnC)–b0882 (TnK)/192.0 kb	CDΔ3c5k	DR3 + DR5 /183.7 kb
CDΔ3c	b1384 (TnC)–b1489 (TnK)/117.0 kb	CDΔ4k6c	DR4 + DR6 /129.4 kb
ND	b1911 (TnK)–b2006 (TnC)/86.4 kb	CDΔ5k6c	DR5 + DR6 /125.7 kb
CDΔ4k	b2011 (TnK)–b2073 (TnC)/70.4 kb	CDΔ3c4k5	DR3 + DR4 + DR5 /254.1 kb
ND	b2034 (TnK)–b2156 (TnC)/142.5 kb	CDΔ3c45k	
ND	b2237 (TnK)–b2315 (TnC)/82.2 kb	CDΔ4k56c	DR4 + DR5 + DR6 /202.4 kb
ND	b2615 (TnK)–b2696 (TnC)/71.2 kb	CDΔ45k6c	
CDΔ5k	b2829 (TnK)–b2890 (TnC)/66.7 kb	CDΔ345k6c	DR3 + DR4 + DR5 + DR6 /313.1 kb
ND	b3396 (TnC)–b3461 (TnK)/77.9 kb	CDΔ3c4k56	
ND	b4016 (TnK)–b4146 (TnC)/157.1 kb	CDΔ34k56c	
CDΔ6c	b4271 (TnC)–b4326 (TnK)/59.0 kb	CDΔ3c45k6	

[a]CD stands for chromosomal deletion; k and c represent the remaining selection markers Km^r and Cm^r, respectively.
[b]ND, not deleted.
[c]DR, deletion region. DR1, b0273–b0369; DR2, b0532–b0619; DR3, b1384–b1489; DR4, b2011–b2073; DR5, b2829–b2890; DR6, b4271–b4326.
These deletion strains showed growth patterns comparable to that of the wild-type strain in LB broth (data not shown).

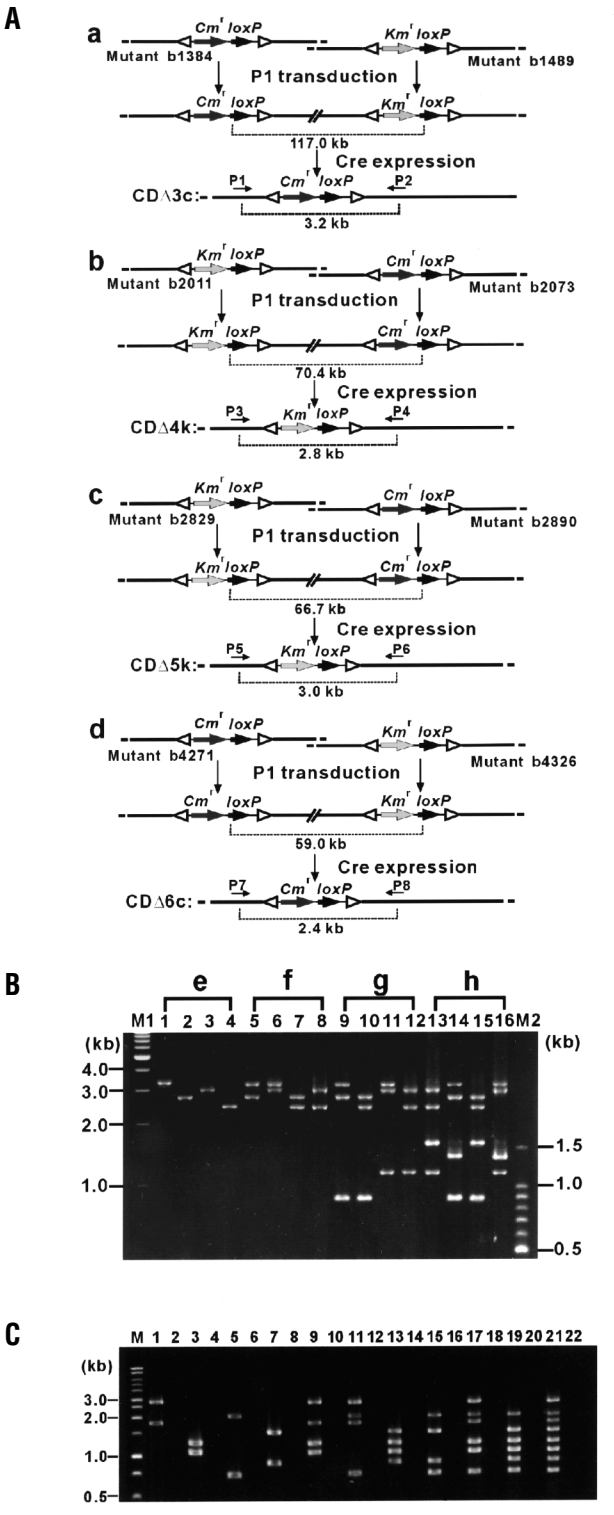

Figure 3. Deletion of genomic segments and confirmation by PCR. (A) A strategy for generating combinatorial deletions. Two *loxP* sites are chosen, one from each Tn*5*-mutant library, and recombined in tandem on the same chromosome by phage P1 transduction. After Cre expression, the selected regions flanked by the two *loxP* sites are deleted, producing deletion strains CDΔ3c (a), CDΔ4k (b), CDΔ5k (c), and CDΔ6c (d). The deletions of CDΔ3c, CDΔ4k, CDΔ5k, and CDΔ6c are confirmed by PCRs using locus-specific primers (P1–P8). (B) Confirmation of the combinatorial and cumulative deletions. The PCRs were conducted using a pair of primers specific to the endpoints of each deletion (P1–P2 for CDΔ3c, P3–P4 for CDΔ4k, P5–P6 for CDΔ5k, and P7–P8 for CDΔ6c, respectively). The successful deletion of CDΔ3c, CDΔ4k, CDΔ5k, and CDΔ6c produced the PCR products of 3.2, 2.8, 3.0, and 2.4 kb, respectively. Panels e, f, g, and h represent the PCR results of the combinatorial-deletion mutant strains containing one, two, three, and four deletion regions on the same chromosome, respectively. Lanes 1–4 indicate the PCR results of the combinatorial-deletion strains CDΔ3c, CDΔ4k, CDΔ5k, and CDΔ6c, respectively. Lanes 5–8 show the multiplex PCR results of the cumulative-deletion strains CDΔ3c4k, CDΔ3c5k, CDΔ4k6c, and CDΔ5k6c, respectively. Lanes 9–12 display the multiplex PCR results of the cumulative-deletion strains CDΔ3c4k5, CDΔ4k56c, CDΔ3c45k, and CDΔ45k6c, respectively. Lanes 13–16 show the multiplex PCR results of the cumulative-deletion strains CDΔ345k6c, CDΔ3c4k56, CDΔ34k56c, and CDΔ3c45k6, respectively. The different PCR products shown in the same deletion region were caused by the presence or absence of the selection markers. M1 and M2 are size markers. (C) Verification of the absence of the deleted region (DR) from each deletion strain. Absence of each DR in deletion strains was confirmed by multiplex PCRs using two pairs of primers specific to the internal sites (genes) of each deletion region (internal genes selected for primers: b1407 and b1468 for DR3, b2035 and b2060 for DR4, b2836 and b2870 for DR5, and b4291 and b4317 for DR6, respectively). Lanes 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21 show the multiplex PCR results of MGl655. Lanes 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, and 22 indicated the multiplex PCR results of deletion strains CDΔ3c, CDΔ4k, CDΔ5k, CDΔ6c, CDΔ3c4k, CDΔ3c5k, CDΔ4k6c, CDΔ5k6c, CDΔ3c4k5, CDΔ4k56c, and CDΔ345k6c, respectively. Lane M, a size marker.

Using our approach, we obtained several deletion mutants in which genomic fragments of ~59–117 kilobases were deleted. These deletion strains had growth patterns comparable to that of the wild-type strain in LB broth (data not shown), indicating that the deleted regions are nonessential for cell growth under these conditions.

After deleting each genomic segment, we were able to readily combine individual deletions into a single chromosome by using phage P1 transduction and then screening on selective media. We used a "markerless" deletion method[11] to eliminate both selectable markers and the remaining *loxP* sites in a cumulative-deletion strain, so any number of deletions can be further combined into a single strain using this process.

Although many of the deletions generated could be successfully combined into a single strain, some deletions that were viable individually were not viable when combined with other deletions. This result clearly indicates that some genes are not dispensable simultaneously, even though they are dispensable individually. The genes belonging to this group may be those involved in alternative metabolic pathways. This observation also suggests that the number of essential genes is greater than that estimated by simple transposon mutagenesis experiments and further illustrates the utility of our combinatorial-deletion approach for the functional study of the *E. coli* genome.

Comparative genomic analysis of *E. coli* suggests that <50% of its genes might be needed for cell growth, with the rest of the genes either having evolved as accessory genes or having been horizontally transferred from other organisms to support growth under specific conditions[17]. This observation suggests that iterative and cumulative deletion experiments such as those described here will eventually generate "minimized" *E. coli* genomes consisting of only those genomic sequences and genes essential for survival under particular conditions. Our long-term goal is to generate these minimized strains and thereby conclusively define practical minimal gene sets that will focus characterization efforts on the most functionally significant sequences. To achieve this, we will expand our transposon libraries to ~4,000 mapped insertion mutants, construct a library of robust *E. coli* strains containing individual deletions, and combine these deletions into strains with various minimized gene sets through a large number of combinatorial and cumulative deletions.

Table 2. Genes and their functions in deletion regions

DR	Genes and functions
DR1 b0273–b0369 (97 ORFs)	*argF*: Ornithine carbamoyltransferase chain F *insB*: Transposase *intF*: Putative prophage cp4-6 integrase *betABIT*: Choline transport system *prpRBCDE*: Propionate catabolism operon *codBA*: Cytosine permease, deaminase *cynBARTS*: Cynate metabolism *lacAYZI*: lac operon *mhpRABCDFET*: 3-phenylpropionic acid metabolism *tauABCD*: Sulfate starvation inducers *hemB*: δ-Aminolevulinic acid dehydratase Unknown genes
DR2 b0532–b0619 (88 ORFs)	*aphCF*: Hydroperoxide detoxification *fimZ*: Fimbriae Z protein *intD*: Prophage dlp12 integrase *emrE*: Ethidium resistance protein *rus*: Crossover junction endodeoxyribonuclease rus *nmpC*: Outer-membrane porin protein nmpC precursor *envY*: Porin thermoregulatory protein env *nfrAB*: Bacteriophage N4 adsorption protein *entDFCEBA, fepAECGDB*: Enterochellin genes *pheP*: Phenylalanine-specific permease *nfnB*: Dihydropteridine reductase *cstA*: Carbon starvation protein A *ahpCF*: Alkyl hydroperoxide reductase *rnk*: Regulator of nucleoside diphosphate kinase *rna*: Nuclease I precursor Unknown genes
DR3 b1384–b1489 (106 ORFs)	*feaRB*: Phenylacetaldehyde dehydrogenase *tynA*: Copper amine oxidase precursor *mac, maoC*: maoC protein *acpD*: Acyl carrier protein phosphodiesterase *hrpA*: ATP-dependent helicase hrpA *aldA*: Aldehyde dehydrogenase A *gapC*: Glyceraldehyde 3-phosphate dehydrogenase C *cytB*: Cytochrome b561 *trg*: Methyl-accepting chemotaxis protein III *rimL*: Ribosomal-protein-serine acetyltransferase *tehAEB*: Tellurite resistance protein *ansP*: l-Asparagine permease *rhsE*: rhsE protein *narVYZU*: Respiratory nitrate reductase 2 *fdnGHI*: Formate dehydrogenase, nitrate-inducible *adhP*: Alcohol dehydrogenase, propanol-preferring *sfcA*: NAD-dependent malic enzyme *sra*: Ribosomal protein s22 *osmC*: Osmotically inducible protein c Unknown genes
DR4 b2011–b2073 (63 ORFs)	*sbcB*: Exodeoxyribonuclease I *hisLGDCBHAFI*: Histidine biosynthesis *wzzB*: Chain-length determinant protein *ugd*: UDP-glucose 6-dehydrogenase *gnd*: 6-Phosphogluconate dehydrogenase *rfcgIfrfbXCADBgalF*: O-antigen biosynthesis *wcaMLKwzxCwcaJcpsBGwcaIHGgmdwcaFEDCBAJw* *zcwzbwza*: Colanic acid biosynthesis *dcd*: Deoxycytidine triphosphate deaminase *udk*: Uridine kinase *alkA*: DNA-3-methyladenine glycosylase II Unknown genes
DR5 b2829–b2890 (62 ORFs)	*PtsP:* PEP-protein phosphotransferase *mutH*: DNA mismatch repair protein *aas*: Acyl-ACP synthetase *galR*: Galactose operon repressor *lysAR*: DAP decarboxylase *araE*: Arabinose-proton symporter *kduDI*: 2-Deoxyl-*O*-gluconate 3-dehydrogenase *lysS*: Lysine-tRNA ligase Unknown genes
DR6 b4271–b4326 (56 ORFs)	*fecEDCBARI*: Iron dicitrate ABC transporter *fimBEAICDFGH*: Type 1 fimbriae biosynthesis *gntP*: Gluconate transporter *uxuABR*: D-Glucuronate catabolism Unknown genes

Experimental protocol

Bacterial strains, plasmids, and enzymes. *E. coli* MG1655 was used for the chromosomal deletion. pMOD<MCS> was purchased from Epicentre Technologies (Madison, WI). pKKlox and pKClox were obtained from Y. Yoon[10] (KAIST, Taejon, Korea), and pKD4, pKD46, and pCP20 from B. Wanner[11] (Purdue University, West Lafayette, IN). pELCre was constructed by cloning a DNA fragment containing a *tetR-*P_{tet}*-cre*, which was PCR-generated from pTATCπ (ref. 10), into the *Bam*HI site of pEL3 (ref. 18). pST is a derivative of pKD4 that contains an *FRT*-flanked *sacB* and a Tc^r that were PCR-generated by using *Bacillus subtilis* and Tn*10* as a template, respectively. All the plasmids were electroporated by Gene Pulser system (Bio-Rad, Hercules, CA) and propagated in *E. coli*. All primers were synthesized from GenoTech (Taejon, Korea).

Construction of transposons. The linear transposons TnKloxP and TnCloxP were constructed for the random insertion of a *loxP* site into the *E. coli* genome (Fig. 1). TnKloxP consists of a Km^r and a *loxP* site, and TnCloxP consists of a Cm^r and a *loxP* site. Both transposons are flanked by the hyperactive 19 bp outer-end transposase recognition sequence (OE). The transposons TnKloxP and TnCloxP were generated from pTnKloxP and pTnCloxP, respectively, by PCR amplification using primers FP-1 (5′-ATTCAGGCTGCGCAACTGT-3′) and RP-1 (5′-TCAGTGAGCGAGGAAGCGGAAG-3′) (Fig. 1). For the construction of pTnKloxP, the 1.1 kb fragment containing a Km^r and a *loxP* site was isolated by digesting pKKlox with *Not*I + *Xba*I and cloned into the *Bam*HI site of pMOD<MCS> by blunt-end ligation. pTnCloxP was also constructed by inserting the 1.2 kb fragment containing a Cm^r and a *loxP* site, which was obtained by digesting pKClox with *Not*I + *Bam*HI, into the *Bam*HI site of pMOD<MCS> by blunt-end ligation.

Random transposition of TnKloxP and TnCloxP. TnKloxP (500 ng) or TnCloxP (500 ng) was incubated with 10 units (1 unit/μl) Tn*5* transposase (Epicentre Technologies) for 30 min at room temperature to form a transposome. The transposome (1 μl) was electroporated into *E. coli* MG1655 using standard procedures[19–21] in the presence of Mg^{2+}. Transposon insertion mutants were selected by using the selectable marker (Km^r or Cm^r) of the transposons. Random transposition of TnKloxP or TnCloxP was confirmed by Southern blot analysis[22] after digestion of the genomic DNA isolated from the transposon insertion mutants with *Cla*I.

The locations of the modified Tn*5* insertion sites were identified by direct chromosomal DNA sequencing using the Tn5Int primer (5′-TCGACCTGCAGGCATGCAAGCTTCA-3′), and the sequences were compared with the GenBank DNA sequence database using the BLAST program[23].

Combinatorial and cumulative deletion. The schematic diagram for the combinatorial deletion is shown in Figure 1. Two *loxP* sites chosen, one from each Tn*5* mutant library (TnKloxP mutants or TnCloxP mutants) depending on the targets to be deleted (Table 1), were recombined in tandem on the same chromosome by phage P1 transduction[24]. Lysates of phage P1 were prepared on the mutant carrying the TnKloxP and used for the transduction of the mutant carrying the TnCloxP or vice versa. The Cm^r-Km^r double-resistant strains containing two *loxP* sites in parallel were selected on LB plates supplemented with Cm (17 μg/ml) and Km (25 μg/ml), and transformed with pELCre. For the deletion of the regions flanked by two *loxP* sites, the *cre* of pELCre was induced by heat-inactivated chlortetracycline (15 μg/ml), and the pELCre was subsequently cured upon shifting to 42°C. The deletion mutants after Cre expression were selected by replica plates, first on an LB-Cm plate and then on an LB-Km plate or vice versa. Six chromosomal deletion mutants CDΔ1k, CDΔ2c, CDΔ3c, CDΔ4k, CDΔ5k, and CDΔ6c (Table 1), were obtained. The deletion of targeted regions was confirmed by PCRs using a pair of primers specific to the endpoints of deletions and two pairs of primers specific to the internal genes of each deletion region with *E. coli* MG1655 (wt) and the deletion strain as templates (Fig. 3).

The cumulative deletion to bring each individual deletion into the same chromosome was conducted with CDΔ3c, CDΔ4k, CDΔ5k, and CDΔ6c. One of the deletion mutants, either CDΔ3c or CDΔ6c, was used as a phage P1 recipient; another mutant, CDΔ4k or CDΔ5k, was then introduced into the selected recipient strain (CDΔ3c or CDΔ6c) by phage P1 transduction, generating four new deletion strains, CDΔ3c4k, CDΔ3c5k, CDΔ4k6c, and CDΔ5k6c, each of which carries two deletion regions on the same chromosome. These deletion strains were selected on LB-Km-Cm plates, and the deletion regions were confirmed by using multiplex PCRs. Before the further introduction of the third deletion region with a Km^r into the above

four deletion strains, the *loxP* site and Km^r present in the strains were exchanged with the *FRT*-flanked *sacB-Tc*^{r} cassette that was generated from pST by PCR using primers with 50 nt homology extensions[11], and the inserted cassette was eliminated by using a helper plasmid pCP20 expressing Flp recombinase, generating four Km^r-free mutant strains, CDΔ3c4, CDΔ3c5, CDΔ46c, and CDΔ56c. For the construction of deletion strains containing three deletions, DR4(b2011–b2073) of CDΔ4k was combined into CDΔ3c5 and CDΔ56c by phage P1 transduction, producing CDΔ3c4k5 and CDΔ4k56c, whereas DR5(b2829–b2890) of CDΔ5k was combined into CDΔ3c4 and CDΔ46c, producing CDΔ3c45k and CDΔ45k6c. For the introduction and selection of the fourth deletion region with a Cm^r into the above four mutants containing three deletions, the *loxP* site and Cm^r present in the DR3(b1384–b1489) of CDΔ3c and DR6(b4271–b4326) of CDΔ6c of the above four mutants with three deletion regions were also eliminated as described above, producing four Cm^r-free strains, CDΔ34k5, CDΔ4k56, CDΔ345k, and CDΔ45k6. Finally, DR3(b1384–b1489) of CDΔ3c was combined into CDΔ4k56 and CDΔ45k6, generating CDΔ3c4k56 and CDΔ3c45k6, and DR6(b4271–b4326) of CDΔ6c was combined into CDΔ34k5 and CDΔ345k, producing CDΔ34k56c and CDΔ345k6c. These four mutant strains contain the same four deletion regions and the same markers on the same chromosome, but the markers are present at different sites. During the construction of cumulative-deletion strains, the complete deletion of the targeted regions from the deletion strains was also verified by multiplex PCRs using locus-specific primers.

Acknowledgments

This work was supported in part by grants from the National Research Laboratory Program (2000-N-NL-01-C-160) and the Molecular Medicine Research Program (00-J-MM-01-B-01) from the Korean Ministry of Science and Technology.

Competing interests statement

The authors declare that they have no competing interests.

Received 11 February 2002; accepted 15 July 2002

1. Schmid, M.B., Kapur, N., Isaacson, D.R., Lindroos, P. & Sharpe, C. Genetic analysis of temperature-sensitive lethal mutants of *Salmonella typhimurium*. *Genetics* **123**, 625–663 (1989).
2. Arigoni, F. *et al*. A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol*. **16**, 851–856 (1998).
3. 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* **93**, 10268–10273 (1996).
4. Winzeler, E.A. *et al*. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
5. Itaya, M. An estimation of minimal genome size required for life. *FEBS Lett.* **362**, 257–260 (1995).
6. Akerley, B.J. *et al*. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **99**, 966–971 (2002).
7. Blattner, F.R. *et al*. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
8. 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.* **181**, 6425–6440 (1999).
9. Koob, M.D., Shaw, A.J. & Cameron, D.C. Minimizing the genome of *Escherichia coli*. *Ann. N. Y. Acad. Sci.* **745**, 1–3 (1994).
10. Yoon, Y.G., Cho, J.H. & Kim, S.C. Cre/*loxP*-mediated excision and amplification of the *Escherichia coli* genome. *Genet. Anal.* **14**, 89–95 (1998).
11. 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* **97**, 6640–6645 (2000).
12. Posfai, G. *et al*. *In vivo* excision and amplification of large segments of the *Escherichia coli* genome. *Nucleic Acids Res.* **22**, 2392–2398 (1994).
13. Hasan, N., Koob, M. & Szybalski, W. *Escherichia coli* genome targeting, I. Cre-*lox*-mediated *in vitro* generation of *ori*- plasmids and their *in vivo* chromosomal integration and retrieval. *Gene* **150**, 51–56 (1994).
14. Kolisnychenko, V. *et al*. Engineering a reduced *Escherichia coli* genome. *Genome Res*. **12**, 640–647 (2002).
15. Szybalski, W. From the double-helix to novel approaches to the sequencing of large genomes. *Gene* **135**, 279–290 (1993).
16. Pratt, L.A. & Kolter, R. Genetic analysis of *Esherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol. Microbiol.* **30**, 285–293 (1998).
17. Lawrence, J.G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95**, 9413–9417 (1998).
18. Armstrong, K.A. A 37×10^3 molecular weight plasmid-encoded protein is required for replication and copy number control in the plasmid pSC101 and its temperature-sensitive derivatives pHS1. *J. Mol. Biol*. **175**, 331–348 (1984).
19. Thompson, J.R., Register, E., Curotto, J., Kurtz, M. & Kelly, R. An improved protocol for the preparation of yeast cells for transformation by electroporation. *Yeast* **14**, 565–571 (1998).
20. Grant, S.G., Jessee, J., Bloom, F.R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA* **87**, 4645–4649 (1990).
21. Goryshin, I.Y., Jendrisak, J., Hoffman, L.M., Meis, R. & Reznikoff, W.S. Insertional transposon mutagenesis by electroporation of released Tn*5* transposition complexes. *Nat. Biotechnol.* **18**, 97–100 (2000).
22. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual*, Edn. 3 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
23. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Miller, J.H. *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992).