Genesis of Artificial Strains Based on Microbial Genomics

Sun Chang Kim, Bong Hyun Sung and Byung Jo Yu
Department of Biological Sciences, KAIST, Taejon 305-701, Korea

Abstract

Creating an artificial strain with a minimal gene set for a specific purpose is every biologist's dream. With the complete genome sequencing of more than 50 microorganisms and extensive functional analyses of their genes, it is possible to design a genetic blueprint for a simple custom-designed microbe with the minimal gene set. Two different approaches are being considered. The first "top-down" approach is trimming the genome to a minimal gene set by selectively removing genes of an organism thought to be unnecessary based on microbial genomics. The second "bottom-up" approach is to synthesize the proposed minimal genome from basic chemical building blocks. The "top-down" approach starting with the genome of a well known microorganism is more technically feasible, whereas the bottom-up approach may not be attainable in the nearest future because of the lack of the complete functional analysis of the genes needed for a life.

Here in this study, we used the top-down approach to minimize the *E. coli* genome to create an artificial organism with "core" elements for self-sustaining and self-replicating cells by eliminating unnecessary genes. Using several different kinds of sophisticated deletion techniques combined with a phage and transposons, we deleted about 19% of the *E. coli* genome without causing any damages to cellular growth. This smaller *E. coli* genome will be further reduced to a genome with a minimal gene set essential for cell life. This minimized *E. coli* genome can lead to the construction of many custom-designed strains with myriad practical and commercial applications.

Introduction

Perhaps no other area of research has been so energized by the application of genome technology than the microbial field. Since TIGR published the first complete genome sequence for a free-living organism, *Haemophilus influenzae*, 6 years ago, the sequences of almost 50 microbial genomes have been completed and more than 100 genomes should be sequenced completely in the next 1 to 2 years. Soon, completed microbial genome sequences will represent a collection of > 200,000 predicted coding sequences (1). Creating artificial bugs may sound like science fiction. However, we are now getting ever closer to designing a genetic blueprint of a minimal gene set for a simple man-made microbe due to the availability of the complete genome sequences of many microorganisms and to the advancement in functional genomics and bioinformatics. A minimal genome is generally defined as the smallest set of genes that allows for self-replication of the organism in a particular environment. The minimal gene set for cellular life can be estimated by comparison of complete bacterial genomes and functional analysis of their genes. In addition, to further estimate the minimum number of genes essential for life, some scientists are developing computer models of the minimal number of biochemical pathways needed for basic metabolic and reproductive functions and others are developing models by comparing the genes of *Mycoplasma genitalium*, which has the smallest known genome of any free-living organism, with the

genomes of other bacteria. The assumption is that common genes are important for survival. So far, by these approaches, researchers believe that the simplest living thing on earth would be a bacterium with between 265 to 350 genes (2, 3). To create a novel organism with a minimal genome, scientists must (i) determine which genes are the minimal set necessary for basic metabolism and replication, (ii) construct this minimal gene set, and (iii) provide or create the necessary nongenetic components for successful gene expression. From the derived minimal gene set, even further reduction may proceed in several directions (2): (i) examine pathways requiring complex cofactors and eliminate those of them that can be bypassed without the use of the cofactors, (ii) eliminate the remaining regulatory genes. (iii) delineate paralogs and replace at least the most highly conserved families with a single, presumably multifunctional "founder", and (iv) apply the parsimony principle (4).

Scientists have proposed, and are working on, two different ways of achieving a minimal gene set. The first, a "top-down" approach, entails removing or inactivating the entire set of genes of an microorganism thought to be unnecessary. The second, and more technically challenging, "bottom-up" approach, entails synthesizing the proposed minimal genome and inserting it into an environment that allows metabolic activity and replication. The bottom-up approach requires complete knowledge of what other cellular components (including proteins, lipids, and sugars) are necessary for metabolism and replication, and of how to assemble all these components along with DNA. Therefore, the bottom-up approach is not feasible in the nearest future. However, the top-down approach can be started now by selectively removing unnecessary genes from the genome of a well-studied microorganism based on microbial genomics and produces a subset of the minimal genome. So we started making artificial strains by minimizing *E. coli* genome based on its functional genomics data because a great deal is known from experimental data about *E. coli*. It uses standard metabolic pathways to derive energy and reducing power from any of a variety of organic compounds under a variety of cultural conditions (5). Many pathways and reactions of the central intermediary metabolism in *E. coli* are also present in both the Archaea and Eucaryota, suggesting that their evolution preceded separation of these major domains of life. All these information led us to choose *E. coli* as a starting microorganism for the creation of an artificial strain with a minimal gene set using the top-down approach.

Experimental

Selection of target genes for deletion

As a first step toward minimizing the *E. coli* genome, we studied many data bases, including the EcoCyc (6, 7), to determine the genes not essential for cell growth. The selection of target genes for deletion is based on the strategy described in Fig. 1.

Genome engineering methods

Genesis of artificial microorganisms by the top-down approach requires simple and efficient genome manipulation methods because at least a thousand deletion experiments should be carried out. Therefore, several different approaches have been used for the gene deletion and the selection of the desired deletion mutants. At the early stage of this work, we used the most widely used method in which the mutant alleles were delivered into the genome on a suicide plasmid. Insertion of the circular molecule into the chromosome requires a single crossover between the mutant and the wilg-type (wt) alleles. This cointegrate was resolved by spontaneous recombination of the allele pair, resulting in cells

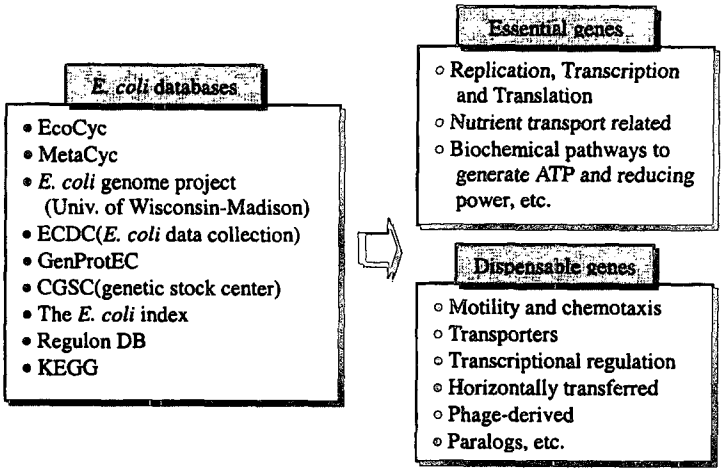

Fig. 1.

with either a wt or a mutant allele in the chromosome. Effective counter-selection was used to eliminate cells which retain the cointegrate structure and carry a counter selectable gene located on the inserted plasmid. The most widely used scheme was the *sac*B/sucrose counter-selection system (8). For the markerless deletion of the target genes, a double stranded breakage mediated homologous recombination method was been used (9). This method involves the integration of the mutant gene, carried on a circular plasmid, at a homologous locus into the chromosome, resulting in a direct duplication. Resolution of this cointegrate by intramolecular recombination was controlled by introducing a unique double-stranded break into the chromosome by the meganuclease I-*Sce*I (9).

The method of introducing the mutant allele into the chromosome on a linear DNA-fragment by a double crossover was also used. In addition, site-specific recombination systems such as Cre/*lox*P and Flp/FRT system, which delete the targeted genomic segment flanked by two recognition sites of a recombinase, was also used for the deletion of large genomic segments (10).

In parallel with the above methods of making precise chromosomal deletions, transposon-induced random deletions were also carried out. The pool of random deletion mutants was periodically screened for those cells that grow normally under the given conditions. The transposon-mediated deletion mutants were physically mapped and the deletion locations were confirmed by chromosomal sequencing. Although the mutant strains that were generated from this random deletion process were only partially characterized, the information obtained concerning the dispensable regions of the genome served as a reliable guide for deriving a fully characterized minimized strain of *E. coli.*

Results and Discussion

So far we deleted 19% of the *E. coli* genome (847 ORFs) without causing any significant damage to the growth of *E. coli*. Some of the mutants grow faster than wt *E. coli*, especially during the logarithmic and stationary phases.

The genes deleted to date are as followings:

- Functionally clustered genes: 182 genes.
  - Anaerobic respiration related (*hya*ABCDEF, *tor*ACDRST, *nar*LXKGHJI, *hyc*HGFEDCBA, *hyp*ABCDE, *hyb*GFEDCBA, *fdh*ED*fdo*IHG*fdh*D, *nrf*ABCDEFG, *frd*DCBA)

- Core-polysaccharide of LPS(*rfa*DFCLKZYJIBSPGQ)
- Flagellar(*flg*NMABCDEFGHIJKL, *fli*YZACDST, *flh*ACDcheYBRWA*mot*BA*tap*, *fli*EFGHIJKLMNOPQR)
- Enterochellin related (*ent*DFCEBA*fep*AECGDB*fes*)
- Betaine, choline(*bet*ABIT), cyanate catabolism(*cyn*RTSX), lacotose(*lac*AYZI), nitrate reductase (*nar*VWYZU), formate dehydrogenase (*fdn*GHI), multi-antibiotic resistance (*mar*RAB), maltose PTS system (*mal*IXY), Iron-dicitrate ABC transporter (*fec*EDCBARI), Fimbriae (*fim*BEAICDFGH), D-glucuronate catabolism (*uxu*ABR), 4-aminobutyrate Transferase (*gab*DTP), glycine proline transport (*pro*VWX), multi-drug secretion (*emr*RAB) related genes.

• Phage derived genes: 41 genes.
• Scattered genes: 304 genes.
• Functional unknown genes: 320 genes.

This smaller *E. coli* genome will be further reduced to a genome with a minimal gene set essential for cellular life. Creating a custom-designed strain with a minimal genome would represent an important step forward in genetic engineering as it would permit the creation of organisms (new and existing) simply from knowing the sequence of their genomes. This research may provide insight into the origins of life, bacterial evolution, or the control of bacterial metabolism. In addition, definition of a minimal genome could lead to a better understanding of the genomes of more complex modern organisms. The minimized *E. coli* genome can lead to the construction of many custom-designed strains with myriad practical and commercial applications. The first practical benefits might be in microbial engineering. Bacteria are now commonly engineered to produce useful products, ranging from industrial chemicals to pharmaceutical proteins. A minimal organism might require less energy or produce fewer waste products that could contaminate the desired product. A minimal organism can also be used as the basis for novel "designer" bacteria that are created to perform specific tasks, such as the breakdown of environmental toxins, to clean up toxic messes, or to create renewable energy by splitting water into hydrogen and oxygen.

Acknowledgments

This work was supported by National Research Laboratory Program of the Ministry of Science and Technology (2000-N-NL-01-C-160) in Korea.

References

1. TIGR microbial database: http://www.tigr.org/tdb/mdb/mdb.html
2. Mushegian, A. and Koonin, E. V. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* **93:**10268-10273.
3. Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. and Venter, J. C. 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286:** 2165-2169.
4. Benner, S. A., Ellington, A. D. and Tauer, A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. USA* **86:**70547058.
5. Riley, M. and Serres, H. 2000. Interim report on genomics on *Escherichia coli. Annu. Rev. Micobiol.* **54:** 341-411.

6. Karp, P. D., Riley, M., Pasley, S. M., Pellegrini-Toole, A. and Krummenacker, M. 1999. EcoCyc: encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* 27: 55-58.
7. Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M. and Pellegrini-Toole, A. 2000. The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* 28: 56-59.
8. Link, A. J., Phillips, D. and Church, G. M. 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: 6628-6637.
9. Posfai, G., Kolisnychenko, V., Berexzki, Z. and Blattner, F. R. 1999. Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Res.* 27: 4409-4415.
10. Zhang, Y., Buchholz, F., Muyrers, J. P. P. and Stewart A. F. 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet.* 20: 123-128.