Emergent Properties of Reduced-Genome *Escherichia coli*

György Pósfai,[1,2*] Guy Plunkett III,[2,3,4] Tamás Fehér,[1] David Frisch,[2,4] Günther M. Keil,[5] Kinga Umenhoffer,[1] Vitaliy Kolisnychenko,[1†] Buffy Stahl,[2] Shamik S. Sharma,[6‡] Monika de Arruda,[2] Valerie Burland,[2,3] Sarah W. Harcum,[7] Frederick R. Blattner[2,3,4*]

With the use of synthetic biology, we reduced the *Escherichia coli* K-12 genome by making planned, precise deletions. The multiple-deletion series (MDS) strains, with genome reductions up to 15%, were designed by identifying nonessential genes and sequences for elimination, including recombinogenic or mobile DNA and cryptic virulence genes, while preserving good growth profiles and protein production. Genome reduction also led to unanticipated beneficial properties: high electroporation efficiency and accurate propagation of recombinant genes and plasmids that were unstable in other strains. Eradication of stress-induced transposition evidently stabilized the MDS genomes and provided some of the new properties.

Escherichia coli K-12 is one of the best understood and most thoroughly analyzed organisms and is the platform of choice for genetic, biochemical, and metabolic simulation research. Commercially, it is used for production of metabolites such as amino acids and proteins of therapeutic or commercial interest. K-12 is also gaining ground for production of DNA for gene therapy, DNA vaccines, and interference RNA. The genomes of two closely related K-12 strains, MG1655 and W3110, have been sequenced (*1*–*3*), and 87% of their genes have functional assignments (*4*). Because *E. coli* evolved in animal intestines and in the environment, parts of its genome are unnecessary for some applications, possibly even counterproductive. By eliminating as many of these gene segments as possible, we have constructed genetically stable "tabula rasa" strains with robust metabolic performance, to which genes for practical applications may be added.

Genome reductions may improve metabolic efficiency and decrease the redundancy among *E. coli* genes and regulatory circuits. Disseminated throughout the genome are mobile DNA elements that mediate recombination events such as transposition and horizontal gene transfer, including insertion sequence (IS) elements, transposases, defective phages, integrases, and site-specific recombinases (*5*). Multiple elements also provide DNA sequence repeats that mediate inversions, duplications, and deletions by homologous recombination even without transposase. To stabilize the genome and streamline metabolism, these elements must be deleted and unwanted functions removed, such as those specific for human hosts or particular environments. By means of a rational design strategy, we avoided loss of robustness that would result from more extensive deletions or an attempt to construct a minimal genome.

Predicting genes to be deleted without detrimental effect is not trivial. MG1655 reduced by 29.7% (*6*) had severely impaired growth and chromosomal segregation, whereas a strain reduced by 7% grew normally (*7*). We used a series of genomic sequence comparisons (Fig. 1) to identify segments present in K-12 but absent from five other *E. coli* (*8*). The analysis yielded nearly 100 proposed deletions (20% of the genome), encoding 900 genes. Initially we targeted large islands, IS-containing islands, and individual genes containing IS elements for removal. Deletion methods were based on recombination mediated by the phage lambda Red system. Beginning with prototype strain MDS12 (*9*), "scarless" deletions were made by removing the targeted segment and resealing the genome so that markers used in the construction were eliminated. Resulting strains were tested for robust growth on minimal medium, and deletions were serially accumulated into a single strain by P1 transduction. Deletion endpoints were verified by sequencing and by DNA microarray hybridization (Fig. 1) (*8*). Physical characteristics of the MDS strains are summarized in Table 1; deletion endpoints are in table S1, deleted genes in table S3, and strain request information in (*8*). Generation of double-strand breaks (DSBs) in each deletion process might have induced error-prone repair, but experiments designed to detect this showed that a single transient break would have no detectable effect on the accumulation of point mutations.

MDS39, the first in the series designed to be IS-free, was examined by genomic DNA hybridization to NimbleGen genome scanning microarrays, which included IS elements, phages, and plasmids absent from K-12 (*8*) as well as the K-12 genomic sequence in the form of 24-base oligonucleotides tiled about every 50 bases on both strands. Alarmingly, we found five unexpected copies of IS that had transposed to new locations (*8*) since the project began in 2002. Specific deletions later removed these IS and other segments, resulting in MDS41, 42, and 43 (*8*).

The reduced strains functioned comparably to the parent, MG1655. Growth rates were very similar (*8*). The slight changes in replichore lengths (Table 1) had no impact. Electroporation efficiencies of DH10B, MG1655, and MDS42 were compared (Table 2) for a small multicopy plasmid (pUC19) and pCC145, a bacterial artificial chromosome with a 145-kb insert. The efficiency for MDS42 was two orders of magnitude higher than that of MG1655 ($P = 0.002$), comparable to that of DH10B (regarded as best for electroporation). MDS42 efficiencies equaled or exceeded those of purchased competent cells, both under conditions optimal for MDS and according to the manufacturer's protocol

Table 1. Summary of the deletion strains and MG1655, based on the updated MG1655 sequence (U00096.2, June 2004); MDS12 values include the original MD1 deletion (see table S1). Gene counts are based on recently updated annotations (*3*).

	MG1655	MDS12	MDS41	MDS42	MDS43
Total number of genes	4,434	4,011	3,731	3,730	3,691
Genome size (bp)	4,639,675	4,263,492	3,977,067	3,976,359	3,931,408
Replichore imbalance (bp)	30,517	141,360	139,331	138,623	183,574
Total no. genes deleted	0	423	703	704	743
Total bp DNA deleted	0	376,183	662,608	663,316	708,267
Percent of genome deleted	0	8.11%	14.28%	14.30%	15.27%

[1]Institute of Biochemistry, Biological Research Center, H-6726 Szeged, Hungary. [2]Scarab Genomics LLC, Madison, WI 53713, USA. [3]Department of Genetics, University of Wisconsin, Madison, WI 53706, USA. [4]Genome Center of Wisconsin, University of Wisconsin, Madison, WI 53706, USA. [5]Federal Research Centre for Virus Diseases of Animals, Institute of Molecular Biology, Friedrich-Loeffler Institutes, D-17493 Greifswald-Insel Riems, Germany. [6]Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC 29634, USA. [7]Department of Bioengineering, Clemson University, Clemson, SC 29634, USA.

*To whom correspondence should be addressed. E-mail: fred@genome.wisc.edu (F.R.B.); posfaigy@nucleus.szbk.u-szeged.hu (G.P.)

†Present address: National Heart, Lung, and Blood Institute, 9000 Rockville Pike, Building 10, Room 7D05, Bethesda, MD 20852, USA.

‡Present address: New England Biolabs, 240 County Road, Ipswich, MA 01938, USA.

Downloaded from http://science.sciencemag.org/ on April 19, 2017

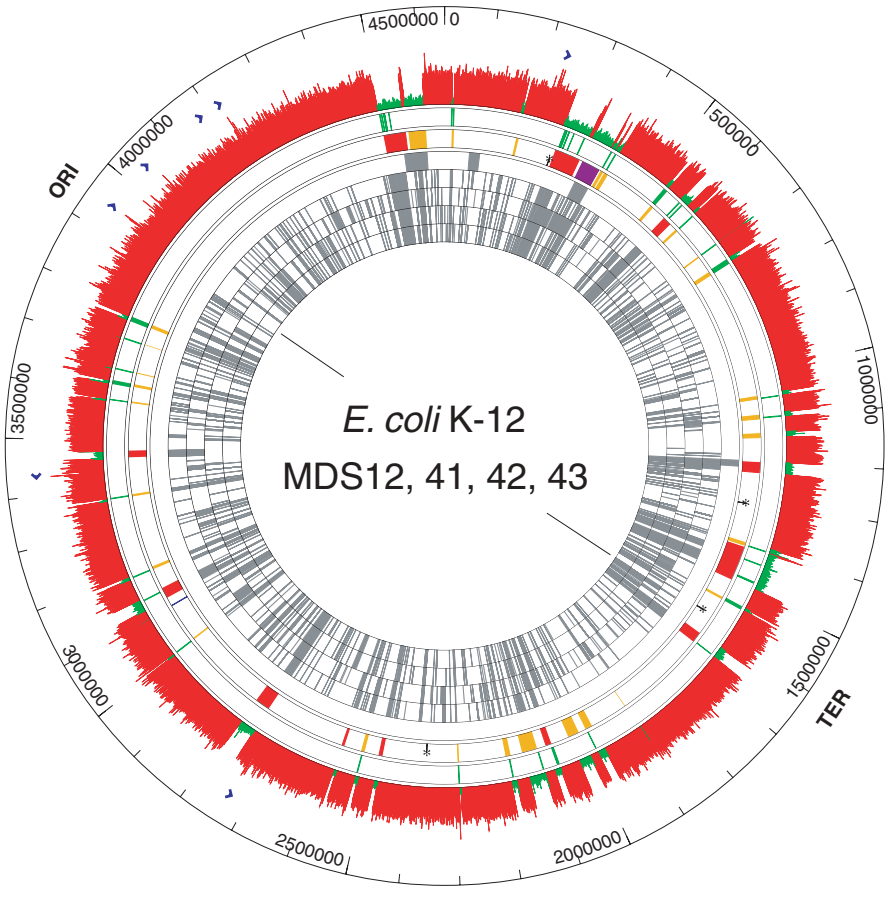


Fig. 1. Design and validation of MDS. Rings depict features mapped to the genome of *E. coli* K-12 strain MG1655, numbered on the outer ring. Outward from the center (rings 1 to 5; gray) are regions of K-12 that are absent in other *E. coli* genomes: RS218, CFT073, *S. flexneri* 2457T, O157:H7 EDL933, and DH10B. Ring 6: Regions targeted for deletion (red, MDS12; yellow, MDS41; blue, MDS42; purple, MDS43; asterisks, IS elements detected in MDS39 and later removed). Ring 7: Native IS and Rhs repeat elements (green). Ring 8: Experimental confirmation of the deletions in MDS43 by a genome-scanning DNA microarray (green, probes corresponding to deletions; red, other probes). Outer ring: ORI and TER, origin and terminus of replication, respectively; rRNA operons are in blue. See (*8*) for further details.

for DH10B. Chemically competent (*10*) MDS42 performed similarly to DH10B (*8*). In fermentations, MDS strains grew to high cell densities by a fed-batch protocol on minimal medium (Fig. 2A). Recombinant protein expression for the model protein chloramphenicol acetyltransferase (CAT) was similar for MDS41 and MG1655 grown to high cell densities (Fig. 2B). An exogenous DNA methyltransferase was also expressed efficiently in MDS42 but displayed low yields in an undeleted host (*11*).

To track IS transposition during experimental procedures, we examined plasmid DNAs isolated from different hosts (*8*). When purified from hosts with IS elements in their genomes, the plasmids were frequently contaminated with elements carried in (or that cotransform with) the plasmid DNA (fig. S1). Elements present in the host were detectable in transformants, even when purchased plasmids were used.

To verify that MDS strains are free from IS-mediated mutagenesis, we examined mutant bacteria that gain spontaneously the ability to use salicin as the sole carbon source. Metabolism of salicin by *E. coli* requires activation of the *bgl* operon, which occurs primarily by IS insertion into the promoter region (*12*). When MDS41 and MG1655 mutants were selected with salicin as the sole carbon source, the activation rate for MDS41 was less than 8% of that for MG1655 (Fig. 3A). The polymerase chain reaction (PCR) confirmed the absence of IS-generated mutations in MDS41, whereas numbers of IS-unrelated mutations were the same in both strains (fig. S4A).

We used classical fluctuation assays to analyze point mutations, deletions, and insertions in growing populations of bacteria by selection of mutants resistant to D-cycloserine, an antibiotic imported by CycA permease (*13*, *14*). Resistance arises almost

Table 2. Electroporation efficiency of MDS42 and undeleted *E. coli* strains with large single-copy and small multicopy plasmids (*8*).

Strain	Plasmid	Transformants/μg
MG1655	pUC19†	0.7×10^8
DH10B	pUC19	35.0×10^8
DH10B*	pUC19	35.4×10^8
MDS42	pUC19	130.0×10^8
MG1655	pCC145‡	0
DH10B	pCC145	1.9×10^6
DH10B*	pCC145	6.5×10^6
MDS42	pCC145	10.0×10^6

*Purchased competent cells. †2.686 kb. ‡153 kb.

exclusively from loss-of-function mutations in *cycA*, which do not affect growth in minimal medium (*15*). The total mutation rates of *cycA* in MG1655 and MDS41 were 6.56×10^{-8} and 5.27×10^{-8}, respectively; this difference (21.2%) was significant ($P \leq 0.0001$; table S2) (*8*). In MG1655, PCR revealed that IS transpositions accounted for 24.2% of the mutations, deletions accounted for 1.5%, and point mutations and small indels accounted for 74.3% (Fig. 3B). In MDS41, no IS-related mutations were found, and frequencies of other mutation types were similar to those in MG1655; thus, the differences are explained by the absence of IS elements in MDS41.

Recombinant ectopic genes are not always tolerated by *E. coli*, and IS mutagenesis provides a defense against expression of products that are deleterious. A chimeric gene composed of VP60 of rabbit hemorrhagic disease virus (*16*) fused to the B subunit of cholera toxin (CTX) was very unstable in *E. coli*. Individually, both genes carried by a low copy number plasmid were stable in *E. coli* HB101, C600, and DH10B, but pCTXVP60 carrying the fusion gene in the same hosts did not produce fusion protein and was recovered in low yields. All recovered plasmids contained mutations in the CTXVP60 open reading frame, virtually all resulting from IS insertions (fig. S2). In contrast, the recombinant plasmid was completely stable in MDS; normal yields of plasmid DNA were obtained.

Plasmids based on adeno-associated virus are used as delivery vehicles in vaccine and gene therapy research (*17*). They are unstable when propagated in standard *E. coli* hosts. The plasmid pT-ITR contains both inverted terminal repeat sequences (ITRs) of the virus. The ITRs fold in perfectly paired, stable secondary structures with double arms (fig. S3A) that frequently delete in *E. coli* (*18*), necessitating extensive screening for intact plasmids before use in gene therapy. We tested pT-ITR stability in MG1655 and MDS42 over serial subcultures (fig. S3B). When grown in MG1655, plasmid restriction

Downloaded from http://science.sciencemag.org/ on April 19, 2017

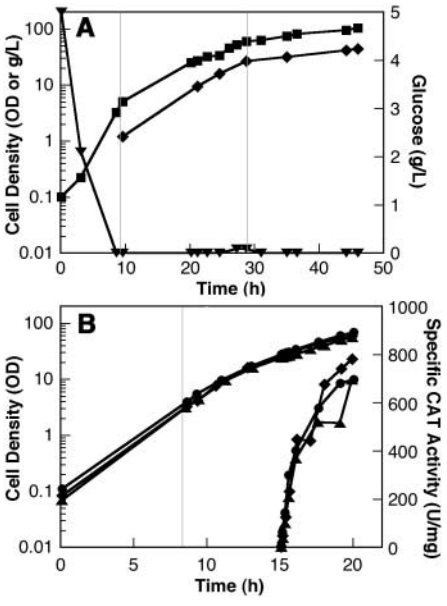


Fig. 2. Growth and protein production. (**A**) MDS41 in minimal medium. Three growth phases (phase 1, batch phase; phase 2, fed-batch, controlled growth rate 0.15 hour^{-1}; phase 3, fed-batch, controlled growth rate 0.03 hour^{-1} to avoid oxygen limitation), marked by vertical lines, were used to reach a dry cell weight (DCW) of 44 g/liter, optical density at 600 nm (OD_{600}) > 100. ■, optical density (left scale); ♦, DCW (left scale); ▼, glucose concentration (right scale). (**B**) Cell density and CAT expression from pProEX HT-CAT in MG1655 and MDS41 in minimal medium, a single fed-batch phase controlled to 0.25 hour^{-1}. Isopropyl-β-D-thiogalactopyranoside (IPTG; 5 mM) was added at 15 hours to induce CAT expression. ●, MG1655; ■ and ▲, MDS41 duplicates.

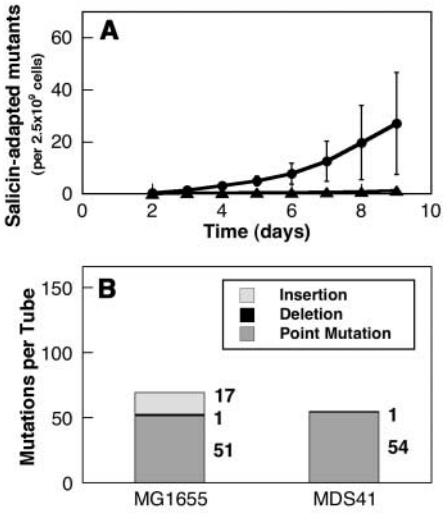


Fig. 3. Mutation rates and spectrum. (**A**) Adaptation of MG1655 (●) and MDS41 (▲) cells to salicin/minimal medium. (**B**) *cycA* mutations causing D-cycloserine resistance in MG1655 and MDS41. Total cell numbers and SD values are in table S2 (*8*).

digests produced multiple new DNA fragments, whereas propagation in MDS42 produced uniform digest patterns over several subcultures. The digest patterns indicated loss of both ITRs in MG1655, which we confirmed by DNA sequencing (*8*).

We showed that pCTXVP60 is not a specific target for IS transposition by performing bystander mutation assays (*8*) on MG1655 with pCTX or pCTXVP60, for mutations to salicin adaptation and D-cycloserine resistance (CS^R). Bacteria containing pCTXVP60 showed a rate of *bgl* mutations 4 times that seen in bacteria containing pCTX (fig S4B). Fluctuation assays also showed higher rates of CS^R when pCTXVP60 was present (fig. S4C, table S2B), and one-third of *cycA* insertion mutants also had IS in the plasmid. No transposition was detected in the plasmid-encoded CAT gene when expression was induced (table S2C), but CS^R insertion mutants were more than twice as frequent as they were without induction (fig. S5).

In this work, deletions totaling up to 15.27% of the *E. coli* genome produced stable strains without physiological compromise. Elimination of transposition artifacts was expected, but neither the increased electroporation efficiency nor the stability of plasmids unstable in K-12 was anticipated. The unexpectedly efficient electrocompetence of MDS42 counters the suggestion (*19*) that a *deoR* mutation in DH10B is the critical determinant for high electroporation efficiency; both MDS and DH10B (*20*) are *deoR*$^+$.

Removal of external structures such as fimbriae could allow better access for DNA to the depolarized membrane, but removal of an unknown deoxyribonuclease or restriction system or activation of an unknown DNA uptake factor could also affect the recovery of transformants. More than 180 of the genes deleted from MDS encode known or predicted membrane-associated proteins (e.g., fimbrial and flagellar structures, transport systems), membrane synthesis enzymes, or regulatory factors, all of which could influence membrane composition cumulatively to bring about altered sensitivity to depolarization. Because much of the K-12 protein interactome remains obscure, unexpected results of multiple deletions are likely. The synthetic biology of genome reduction could have produced synergistic interactions ("synthetic beneficials") between deletions that resulted in an altered phenotype such as high electrocompetence, whereas other combinations of deletions would result in less surprising synthetic lethals.

All strains tested, except MDS, were affected by contamination of isolated plasmids with IS-containing DNA from carryover of IS-containing genomic DNA, IS mini-circles (*21*), and plasmids carrying integrated IS. Our results show that any DNA propagated on *E. coli* containing IS elements is likely to be contaminated with IS elements, and that transposition can be frequent. IS transposition mutates both plasmid and chromosomal bystander genes in a manner consistent with stress-induced activity; powerful selection must also operate. Removing the IS eliminated the main source of instability.

References and Notes

1. F. R. Blattner *et al.*, *Science* **277**, 1453 (1997).
2. K. Hayashi *et al.*, *Mol. Syst. Biol.*, published online 21 February 2006 (10.1038/msb4100049).
3. M. Riley *et al.*, *Nucleic Acids Res.* **34**, 1 (2006).
4. M. H. Serres, S. Goswami, M. Riley, *Nucleic Acids Res.* **32**, 300D (2004).
5. L. S. Frost, R. Leplae, A. O. Summers, A. Toussaint, *Nat. Rev. Microbiol.* **3**, 722 (2005).
6. M. Hashimoto *et al.*, *Mol. Microbiol.* **55**, 137 (2005).
7. B. J. Yu *et al.*, *Nat. Biotechnol.* **20**, 1018 (2002).
8. See supporting material on *Science* Online.
9. V. Kolisnychenko *et al.*, *Genome Res.* **12**, 640 (2002).
10. H. Inoue, H. Nojima, H. Okayama, *Gene* **96**, 23 (1990).
11. A. Kiss, personal communication.
12. B. G. Hall, *Mol. Biol. Evol.* **15**, 1 (1998).
13. R. J. Wargel, C. A. Hadur, F. C. Neuhaus, *J. Bacteriol.* **105**, 1028 (1971).
14. R. R. Russell, *J. Bacteriol.* **111**, 622 (1972).
15. T. Fehér, B. Cseh, K. Umenhoffer, I. Karcagi, G. Pósfai, *Mutat. Res.* **595**, 184 (2006).
16. G. Meyers, C. Wirblich, H. J. Thiel, *Virology* **184**, 664 (1991).
17. L. Santat *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11053 (2005).
18. S. Chatterjee, personal communication.
19. S. G. Grant, J. Jessee, F. R. Bloom, D. Hanahan, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4645 (1990).
20. Our analysis of preliminary genome sequence data for DH10B obtained from the Human Genome Sequencing Center (www.hgsc.bcm.tmc.edu/projects/microbial/EcoliDH10B).
21. Y. Sekine, N. Eisaki, K. Kobayashi, E. Ohtsubo, *Gene* **191**, 183 (1997).

22. We thank K. Giesow, A. Szalkanovics, N. Zimmerman, N. Hermersmann, S. Phillips, E. Steffen, and J. Altmann for excellent technical contributions; M. Chandler for sequences of IS elements; I. Karcagi for help with DNA constructs; J. Luecke for help with microarrays; and T. Albert, S. C. Kim, and J. Campbell for critical reading of the manuscript. Supported by a grant from the Hungarian Scientific Research Fund (OTKA) (G.P.), NIH grant GM35682 (F.R.B.), and the Federal Ministry of Education and Research (G.M.K.). F.R.B. is president and CEO of and has a financial interest in Scarab Genomics, LLC, is president of and has a financial interest in DNASTAR, Inc., and is on the scientific advisory board of and has a financial interest in Nimblegen Systems, Inc.

Supporting Online Material

www.sciencemag.org.cgi/content/full/1126439/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 to S4
References and Notes

21 February 2006; accepted 19 April 2006
Published online 27 April 2006;
10.1126/science.1126439
Include this information when citing this paper.

Downloaded from http://science.sciencemag.org/ on April 19, 2017



Emergent Properties of Reduced-Genome *Escherichia coli*
György Pósfai, Guy Plunkett III, Tamás Fehér, David Frisch, Günther M. Keil, Kinga Umenhoffer, Vitaliy Kolisnychenko, Buffy Stahl, Shamik S. Sharma, Monika de Arruda, Valerie Burland, Sarah W. Harcum and Frederick R. Blattner (April 27, 2006)
Science **312** (5776), 1044-1046. [doi: 10.1126/science.1126439]
originally published online April 27, 2006

Editor's Summary

This copy is for your personal, non-commercial use only.

Article Tools Visit the online version of this article to access the personalization and article tools:
http://science.sciencemag.org/content/312/5776/1044

Permissions Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Downloaded from http://science.sciencemag.org/ on April 19, 2017