Systematic Genome Reductions: Theoretical and Experimental Approaches

Tamás Fehér,[†] Balázs Papp,[†,‡] Csaba Pál,[§] and György Pósfai*[,†]

Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary, Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom, and Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom

Received October 18, 2006

Contents

1. Introduction 3498
2. Minimal Gene-Set Concept 3500
3. Naturally Evolved Minimal Gene Sets 3500
4. Estimating the Size and Gene Content of Minimal Genomes 3501
   4.1. Comparative Genomics Approach 3501
   4.2. Large-Scale Gene Inactivation Studies 3501
   4.3. Computational Systems Biology Approaches 3502
5. Techniques of Experimental Genome Reductions 3503
   5.1. General Strategies 3503
   5.2. Basic Deletion Methods 3503
      5.2.1. Suicide Plasmid-Mediated Procedures 3503
      5.2.2. Linear DNA-Mediated Procedures 3504
      5.2.3. Additional Tools: Site-Specific Recombinases and Transposons 3504
6. Genome Reduction Projects 3505
   6.1. *E. coli* 3505
      6.1.1. Random Genome Reduction by Transposon Technology 3506
      6.1.2. Semirandom Genome Reduction 3506
      6.1.3. Genome Reductions via Targeted Strategies 3507
      6.1.4. Lessons of Parallel *E. coli* Genome Reductions 3508
   6.2. *B. subtilis* 3508
   6.3. *Corynebacterium glutamicum* 3509
7. Concluding Remarks 3510
8. Acknowledgements 3511
9. Note Added in Proof 3511
10. References 3511

1. Introduction

The living cell is the most complex structure, known to man, in the micrometer size range. On the road to understand its many and complex chemical reactions and exploit the enormous industrial potential it represents, a reductionist approach at the whole-cell level can clearly be beneficial. Molecular genetic engineering offers a way to create a simplified cell.

Even the simplest cell is made up of tens of millions of molecules of thousands of different kinds interacting in a complex cellular network.[1] However, since the cell is a self-organizing entity with most of the cell's hereditary information needed for structure and function coded in its genes, cell simplification can be reduced to the task of engineering the genome, the long chain of the DNA molecule.

Genome reduction projects are motivated by both academic and industrial interests with the most fundamental questions of life at heart. How far can the simplification go? What is the minimal gene set needed for sustaining life in a defined environment? Can we reduce the genome to a point where all components and reactions can be fully cataloged and characterized? Such whole-scale reductions as well as bottom-up approaches of synthesizing entire genomes with the ultimate goal of building a living cell might be far off. However, on the practical side, improvement of certain cellular functions by streamlining genomes is existing reality.

Currently, the subjects of such undertakings are prokaryotic genomes. The gene set of a prokaryotic cell can be relatively small (down to ∼400 genes),[2−5] and bacteria exhibit lower cellular complexities than eukaryotes.[6] Over 500 prokaryotic genomes have been fully sequenced to date (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). The enormous amount of data produced by the genome sequencing projects combined with the genetic/biochemical knowledge accumulated over the past decades provides an unprecedented degree of understanding of the prokaryotic cell. For the best-known model systems, like the *Escherichia coli* cell, systems biology approaches have been producing constantly improving metabolic and regulatory network models displaying increasing predictive power.[7,8] From the technical point of view, some well-studied prokaryotes are readily amenable to genome engineering. Their easy handling is due to their single-celled nature, the fast replication cycle resulting in large population size, the established selection schemes, and an array of sophisticated *in vivo* DNA-modifying techniques.[9]

As early as 1994, somewhat before the publication of the first bacterial genome sequence, systematic genome size reduction of *E. coli* was suggested by Koob et al.[10] The proposal was based on the notion that the full complement of the bacterial gene set is not needed under defined conditions in the test tube. Selective removal of genes needed only to meet the challenges of a changing natural environment would relieve the cells from running unused pathways and building up unnecessary products. A simplified, fully defined cell to which metabolic activities needed for a particular application are added could than be more efficiently grown under a controlled laboratory or bioreactor environment. Although this mechanistic assumption of engineer's remains largely to be proven, most current experimental genome reduction projects follow this line of

* To whom correspondence should be addressed. Phone: +36-62-599-778. Fax: +36-62-433-506. E-mail: posfaigy@brc.hu.
[†] Biological Research Center of the Hungarian Academy of Sciences.
[‡] The University of Manchester.
[§] University of Oxford.

10.1021/cr0683111 CCC: $65.00
Published on Web 07/18/2007

Tamás Fehér earned his M.D. at the University of Debrecen, Hungary, in 2000. His undergraduate research, conducted under the supervision of János Matkó, concentrated on mouse-thymocyte differentiation with special attention to purinoceptor-linked intracellular signal transduction pathways. After graduation he joined the group of György Pósfai to take an active part in his bacterial genome-engineering project. In 2006, he successfully defended his Ph.D. thesis written on "Construction and characterization of a multi-deletion *Escherichia coli* strain". He is presently focusing on the role of transposable elements in the adaptation of prokaryotic cells to environmental stress.

Balázs Papp earned his M.Sc. degree from the University of Debrecen in 2001 and his Ph.D. degree in Theoretical Biology from Eötvös Loránd University (Budapest, Hungary) in 2004. He was a postdoctoral fellow at the Faculty of Life Sciences of The University of Manchester. He is currently a Human Frontier Science Program Fellow at the Biological Research Center, Szeged, Hungary. His research interests focus on understanding the evolution of genetic systems and metabolic networks in microbes by combining systems biology and comparative genomics tools. In his recent works he has investigated the causes and evolution of gene essentiality and employed computational models to predict minimal gene sets.

Csaba Pál received his doctorate degree in evolutionary genetics in Hungary. After finishing his Ph.D. degree he worked as a postdoctoral fellow at the University of Bath (U.K.) and European Molecular Biology Laboratory (EMBL, Heidelberg). He is currently a research fellow at the University of Oxford. His main focus is on evolutionary systems biology using theoretical modeling, comparative genomics, and microbial selection experiments. His current research addresses the evolution of minimal genomes.

György Pósfai studied biology at the University of Szeged, Hungary, and was awarded his M.Sc. degree in 1981 and Ph.D. degree in 1989, studying genes of bacterial restriction-modification systems. In the late 1980s to early 1990s he worked as a postdoctoral fellow in the laboratory of Waclaw Szybalski at the University of Wisconsin and became involved in genome engineering and sequencing. Returning to Hungary, he held an International Research Fellowship of the Howard Hughes Medical Institute in 1996–2000 and was awarded a Széchenyi Professorship in 2000–2003. Currently he is Director of the Institute of Biochemistry at the Biological Research Center of the Hungary Academy of Sciences, Szeged, Hungary. His research focuses on microbial genomics with special interest in mobile genetic elements and synthetic/systems biology of minimal genomes.

thought. Thus, it is important to note that current genome minimization projects do not aim directly to create a fully minimal genome. Rather, the immediate goal is to reduce the gene content to a point where the streamlined cell still retains its favorable characteristics, including robust growth without the need for special nutritional requirements.

It is not a trivial task to define portions of the genome that are not needed under defined conditions. Obviously, there is a core set of cellular functions, mostly those involved in information processing, that must be kept intact.[11] The genomes of endosymbionts, products of intensive natural genome reductions, can give us clues as to what must be retained.[12] Extended genome comparisons as well as genome reduction simulations based on *in silico* deletions in metabolic network models help to predict the effects of deletions.[13] Gene essentiality studies,[14] based mostly on random[15−18] (e.g., high-throughput transposon mutagenesis) or systematic[19−22] gene inactivation and use of antisense RNA,[23,24] provide empirical data on individual genes. Genome comparisons of closely related bacterial strains are highly informative, revealing a core genome for the particular species and identifying a set of genomic islands specific for the particular strain.[25−27] These islands, presumably horizontally acquired gene packages with niche-specific functions, are obvious targets of systematic reduction efforts.[28−30] Deletion of parasitic DNA, phages, and transposons can reduce the mutagenic flexibility of the genome, needed in a stressful environment, but might be beneficial in the laboratorial/industrial setting by increasing genomic stability.[31]

All the information inferred from computational and empirical approaches might not be enough to predict the

effect of gene deletion on cell physiology. Even the best-known organisms contain a large number of genes (>10% of the total) with unknown functions.[32] Moreover, most of the interactions of the cellular constituents and, consequently, the concerted effects of individual, sequential deletions are largely unknown. However, this is by no means a fatal problem for experimental genome reduction approaches. If a particular deletion in a sequential series proves to be deleterious under the desired conditions, taking one step back and continuing with another reduction path can alleviate the problem.

In the past 5 years systematic experimental deletion works resulting in substantial genome reductions were reported for three prokaryotic organisms. *E. coli* and *Bacillus subtilis* are important model systems with biotechnological and industrial interests, while *Corynebacterium glutamicum* is a principal object of the amino acid fermentation industry. Improvements of certain strain characteristics, including increased genome stability and higher plasmid and protein yields in *E. coli*[31] or increased transformation efficiency and plasmid stability in *B. subtilis*,[30] have been reported in connection with the genome reductions. Clearly, the reductionist approach is starting to bear fruit, and the present paper aims to review recent theoretical and experimental progresses on genome minimizations.

2. Minimal Gene-Set Concept

The availability of hundreds of complete genome sequences of vastly different sizes reformulates one of the central questions in biology: What is the smallest set of genes sufficient and necessary to maintain cellular life? Clearly, the notion of minimal gene set is only meaningful when associated with well-defined sets of environmental conditions.[11] For instance, genes encoding enzymes for the biosynthesis of amino acids should only be present in the minimal genome if amino acids were not available in the environment. Thus, the absolute minimal gene set would correspond to the smallest group of genes that is sufficient to sustain a functioning cellular life form under the most favorable conditions, that is, in the presence of a full complement of essential nutrients and in the absence of environmental stress.[33]

There are currently two conceptually different ways to derive a minimal genome.[34] The bottom-up approach aims at constructing artificial chemical supersystems capable of replication and evolution with no macromolecules provided as nutrients.[35,36] This research program, traditionally pursued by investigators of prebiotic chemistry, is motivated not only by the prospect of building a simple model of complex natural cells but also by the possibility to gain better insight into the origin of life on Earth.[37,38] Although lab-designed cellular life is far from reach, there have been serious attempts to build subsystems of primitive cells (protocells) (for reviews, see refs 36 and 38). Moreover, plans have been put forward to construct modern cells by encapsulating *in vitro* synthesized DNA and cellular protein machineries within a lipid bilayer.[35,39] In addition to defining the list of genes to be encoded by such a synthetic genome, bottom-up approaches also need to identify those macromolecular components and small molecular metabolites that have to be provided to kick start the artificial cell.

In contrast to these bottom-up strategies, the top-down approach starts from existing organisms with the aim of simplifying their genome, arriving at minimal (or at least reduced) gene sets. Although genome minimization sounds less complicated than synthesis of a cell from scratch, efforts to streamline existing genomes face several theoretical and technical challenges: What is the most efficient experimental protocol to delete a large number of genes? How to decide which genes to delete and which ones to keep? How one can predict the effect of deletions on cell physiology? What is the impact of environmental circumstances on the composition and size of minimal gene sets? Is there a unique minimal gene repertoire under a given condition, or may many complementary minimal gene sets be available?

Table 1. Lifestyle and Number of Protein Coding Genes in Different Organisms[a]

organism	lifestyle	no. of protein coding genes
γ-Protebacteria		
E. coli K12	free living	4289
P. aeruginosa	opportunistic pathogen	5570
B. aphidicola str. APS	intracellular mutualist	564
B. aphidicola str. Cc	intracellular mutualist	357
W. glossinidia	intracellular mutualist	621
α-Protebacteria		
C. crescentus	free living	3737
A. tumefaciens	plant pathogen	5402
R. prowazekii	intracellular parasite	834
*W. pipientis w*Mel	intracellular parasite	1158
*W. pipientis w*Bm	intracellular mutualist	805
Firmicutes		
B. subtilis	free living	4099
M. genitalium	obligate host-associated parasite	482
U. urealyticum	obligate host-associated parasite	650
P. asteris	intracellular parasite	754
Archaebacteria		
P. furiosus	free living	2065
M. mazei	free living	3371
N. equitans	obligate host-associated parasite	536

[a] Gene numbers of bacteria with different lifestyles information on gene numbers were extracted from the Genomes OnLine Database[164] (GOLD, http://www.genomesonline.org/) except for *M. genitalium*, where an updated number was taken from ref 18.

3. Naturally Evolved Minimal Gene Sets

Organisms with nearly minimal number of genes not only are goals for biotechnological applications but also occur in nature. Although most bacteria possess ~2000−8000 genes, the smallest natural bacterial genomes contain less than 1000 genes, often in the range of 400−600 (Table 1). Organisms with an especially small number of genes are invariably obligate host-associated bacteria, suggesting that a small complement of genes is sufficient to maintain life under the nutrient-rich, constant environment provided by the host.[40] Obligate dependence on host cells could result from either parasitic or mutually beneficial relationships. One of the well-described examples on the latter comes from studies on endosymbiotic (intracellular) bacteria of insects, such as *Buchnera aphidicola* strains. These bacteria provide certain nutrients missing from the host diet (for a review, see ref 41). In contrast, several *Mycoplasma* species are human pathogens and harbor only a few hundred genes.

Phylogenetic studies demonstrated that host-dependent bacteria with near-minimal gene sets are by no means remnants of ancient life forms but rather represent evolution-

ary derived conditions:[42−44] their tiny gene sets evolved from much larger genomes through massive loss of genes, including those no longer required in the intracellular environment.[40,45] In addition to relaxed selection pressure on specific gene functions, host-associated lifestyle also entails reduced population sizes relative to free-living relatives, resulting in mutational deterioration of weakly selected genes.[45] Drastic genome shrinkage can occur on a relatively short evolutionary timescale: the free-living ancestor of *Buchnera* has lost approximately 75% of its genome since it switched to an endosymbiotic lifestyle approximately 200 million years ago.[46] Moreover, the process of bacterial genome reduction has been observed in laboratory experiments[47] and chronological isolates of a human pathogen from the same patient,[48] suggesting that rates of DNA loss can be occasionally very high and possibly moulded by natural selection.

Although different host-associated bacteria have nearly the same number of genes (Table 1), comparative analyses of fully sequenced genomes revealed substantial differences in the composition of the gene sets retained by these organisms. For example, only approximately 50% of the genes conserved among intracellular mutualistic bacteria have been detected in obligate host-associated parasite species.[40] The observed diversity of evolved minimal gene sets may be the product of at least three fundamental processes. First, in many cases, genome reduction independently occurred in different lineages; hence, there must have been significant differences in the initial genetic makeup. Second, due to variation in selective forces imposed by the host, host-associated bacteria have evolved different lifestyles, reflected in variation in their genetic repertoires, and functional and metabolic capacities.[41] Although the biggest difference lies between parasitic and mutualistic lifestyles, more subtle variations in the nature of the host−endosymbiont relationships are also prevalent. For instance, even though the mutualistic bacteria of aphids and tsetse flies are phylogenetically closely related, they provide different sets of nutrients for their hosts (essential amino acids and vitamins, respectively[41]) and hence show considerable differences in their gene contents. Finally, diversity in reduced genomes could also arise from historical contingencies. Differences in the order of gene deletions can have an influence on the evolutionary outcome when essential cellular functions are encoded by redundant gene sets in the ancestral genome.[45] In this case, elimination of one such set would render the other one indispensable and vice versa.[13,46] The role of historical contingency on the outcome of reductive evolution could only be assessed if we could "replay life's tape",[49] that is, if evolution of minimal genomes could be repeated from the same ancestor and under identical selective conditions. In principle, both laboratory microbial evolution experiments[47] and computational tools[13] could be used to investigate repeated evolution of minimal gene sets under controlled conditions (see below).

The rapid accumulation of information on extremely reduced genomes and the fact that many of these organisms are phylogenetically related to well-studied free-living bacteria provide a unique opportunity not only to better understand the evolution of endosymbiosis but also to aid laboratory genome minimization efforts. First, investigation of the lifestyles and gene inventories of endosymbionts could give clues on the set of genes that are indispensable and hence should be retained in an engineered minimized genome.[12] Second, the selective regimes leading to genome shrinkage in host-associated bacteria could be mimicked in laboratory evolution experiments with the aim of evolving, rather than designing, reduced genomes.[47]

4. Estimating the Size and Gene Content of Minimal Genomes

Any engineering approach on genome minimization faces the problem of identifying those genes that can be removed from the genome without seriously compromising cell viability. Several strategies, both computational and experimental, have been suggested to predict the set of genes that should be retained in a minimal genome. We discuss these approaches and their potential limitations below.

4.1. Comparative Genomics Approach

Attempts to determine the minimal gene set by genome comparisons are based on the idea that genes that are shared between distantly related organisms are likely to be essential for cellular life.[12,50] Although a comparison of the first two completed bacterial genomes provided an estimate of 256 genes for the minimal set,[50] analysis of ∼100 genomes shows that only 63 genes are ubiquitous.[11] This discrepancy points to the main limitation of the method: numerous essential cellular functions can be performed by unrelated proteins which show no sequence similarities to each other in different organisms, resulting in the nonubiquitous presence of the corresponding genes across species. For example, the principal replicative DNA polymerase of eubacteria fails to show sequence similarity to its functional counterparts found in archaebacteria and eukaryotes.[33] Moreover, when each ubiquitous gene is required to be present in every genome, the size of the ubiquitous set will be artificially small due to sequencing and annotation errors or detection difficulties at low degrees of sequence conservation.[51] Hence, one cannot rely exclusively on the comparative approach as it may substantially underestimate the size of the minimal gene set. A further limitation of the method is that environmental dependence of the minimal gene repertoire cannot be directly addressed under this framework.

Despite these apparent difficulties to define the smallest set of genes by comparing extremely distant genomes, the comparative approach could offer invaluable insights on core gene sets conserved across closely related organisms by identifying strain-specific, horizontally transferred genomic islands,[27] which are most likely to be dispensable in lab. For example, genome comparison of different *E. coli* strains was employed to infer a reduced though not absolute minimal *E. coli* genome (a so-called core genome).[31]

4.2. Large-Scale Gene Inactivation Studies

An alternative approach to infer the set of genes essential for cellular life is to experimentally identify those genes whose individual inactivation causes inviability. Recent advances in functional genomics techniques enabled the genome-wide identification of such essential genes in various organisms, including bacteria,[15−20,22,24,52−55] yeast,[56] and worm.[57] Surprisingly, the fraction of essential genes proved to be low in almost all organisms investigated, typically in the range of 10−30%[11] (Table 2). The one current exception is the highly reduced genome of an obligate host-associated parasite *Mycoplasma genitalium* in which more than three-quarters of the genes are required for growth.[18]

Table 2. Percentage of Essential Genes and Number of Protein Coding Genes in Different Organisms[a]

organism	no. of protein coding genes	est % of essential genes	ref
M. genitalium	482	79	18
H. pylori	1590	17	52
H. influenzae Rd	1850	38	19
S. aureus N315	2594	25	24
M. tuberculosis H37Rv	4402	15	17
B. subtilis	4099	6.6	20
E. coli K12	4296	7.1	22
S. typhimurium LT2	4597	11	53
C. glutamicum R	3052	22.6	54
F. novicida U112	1719	23	55
S. cerevisiae	5794	19	56
C. elegans	19 099	7	57

[a] Note that estimates of gene essentiality are based on different gene inactivation methods and therefore these figures should be compared with caution. Information on gene numbers were extracted from the Genomes OnLine Database[164] (GOLD, http://www.genomesonline.org/) except for *M. genitalium*,[18] *E. coli*,[22] and *S. cerevisiae*,[165] where updated data were used from the literature.

Why do so many genes seem to be dispensable under laboratory conditions? First and foremost, most assays typically score mutants as being viable or not. This is the crudest of distinctions and fails to capture even relatively large effects on growth rates. Indeed, a nontrivial fraction of viable yeast mutants exhibits a measurable growth defect under normal conditions.[58−60] Second, laboratory conditions often fail to detect genes required under special environmental conditions and may therefore overestimate the fraction of nonessential genes.[61,62] It is tempting to speculate that the relatively low frequency of dispensable genes in the genome of *Mycoplasma* reflects its strict host-associated lifestyle that has enabled the evolutionary loss of genes needed only under specific environmental conditions.[61] Finally, loss of many dispensable genes can be compensated by other genes in the genome[63] due to either the presence of functionally redundant gene copies[64] or operation of alternative cellular pathways.[65]

The presence of such compensatory genetic interactions pose a serious difficulty for attempts to determine the minimal gene set based on single-gene inactivation experiments or by comparative genomics: genes that are individually dispensable may encode essential functions and hence may not be simultaneously dispensable.[12] Thus, the set of essential genes of any organism is expected to represent only a subset of the minimal genome. A better understanding of the cell's compensatory capacity would be needed to circumvent this difficulty. We need a catalog of gene sets that encode all essential cellular functions. Although promising approaches have been developed for high-throughput identification of pairwise genetic interactions in yeast,[66,67] an exhaustive mapping of interactions requires construction of an enormous number of combinations of multiple-gene deletant strains.

While the above considerations suggest that the complete set of nonessential genes is unlikely to be simultaneously dispensable, the reverse might also be true: many apparently essential genes can be deleted in combination with other genes. For example, although construction of an *E. coli* strain carrying a single-gene deletion in the *yefM* antitoxin gene was not successful,[22] *yefM* could be deleted as part of larger chromosomal segments containing its toxin-encoding locus, *yoeB*.[68,69] Moreover, transposon mutagenesis might overestimate the essential gene set by misclassification of nonessential genes that slow down cellular growth without arresting it (but might also miss those that can tolerate transposon insertion).[20]

4.3. Computational Systems Biology Approaches

A complete understanding of the relationship between genotype and phenotype would greatly facilitate the design of minimal genomes and render the above inference approaches unnecessary. Mathematical models relating gene content to cell physiology would inherently account for genetic interactions and enable the simulation of minimal gene sets under various environmental conditions. Although such a comprehensive mathematical representation of a whole cell is out of reach at present, models of various cellular subsystems (e.g., metabolism,[70,71] cell cycle,[72] signal transduction[73]) are becoming increasingly available.

Most computational efforts[13,74,75] to minimize biological systems have focused on metabolic networks, the best characterized cellular subsystem. Genomic information coupled with biochemical and physiological knowledge has enabled the reconstruction of genome-scale biochemical reaction networks for microorganisms.[70] Although traditional dynamic analysis of these large-scale networks is currently hindered by the lack of detailed kinetic information, a new modeling framework, the constraint-based approach, has been introduced to deduce the metabolic phenotype from the genotype.[76] The constraint-based approach attempts to narrow the range of possible phenotypes that a metabolic system can display based on the successive imposition of governing physicochemical, biological, and evolutionary constraints.[76,77] Despite its simplicity, applications of constraint-based modeling of the *E. coli* metabolic network have already yielded numerous key theoretical insights on the nature and evolution of minimal genomes. First, it has been quantitatively demonstrated that the minimal set of reactions capable of supporting growth is strongly dependent on the environment and growth efficiency requirements imposed on the network.[74] Second, theoretical support has been given to the notion that the catalog of essential genes is only a subset of the minimal genome: apparently, single-gene deletion studies underestimate the minimal metabolic gene set by about 45%.[13] Third, repeated simulations of successive gene loss events have revealed that various functionally equivalent minimal networks, differing in both gene content and number, can evolve even under identical conditions and starting gene sets.[13] Differences in the composition of the minimal networks can be attributed to the presence of alternative pathways in the ancestral network and differences in the order of gene deletion events during the repeated simulations. Thus, different sets of alternative metabolic routes can be lost in repeated rounds of evolutionary simulations, resulting in a distribution of minimal networks. Finally, the same simulation framework has been applied to model the gene loss of endosymbionts, which are closely related to *E. coli*. Starting from the present-day metabolic network of *E. coli* and mimicking the endosymbiotic lifestyle, the model predicts the outcome of genome reduction leading to endosymbiont *Buchnera* with 80% accuracy.[13] There is, however, a general caveat for application of constrained-based methods: some of the computationally derived minimal metabolic networks might not be kinetically feasible.[75]

An alternative approach is to design a minimal cell model that can perform well-defined sets of cellular functions and try to infer the required genetic instructions from the

biochemical blueprint.[78] This goal can be accomplished by building a minimal coarse-grained model of a bacterial cell, which is functionally complete, with growth rate, composition, division, and changes in cell morphology as natural outputs from dynamic simulations.[79] Such a coarse-grain model uses pseudochemical components (or modules) that are aggregates of distinct chemical species that share similar chemistry and metabolic dynamics. This bottom-up computational representation provides a framework in which the modules can be delumped into chemical and genetic descriptions while maintaining interconnections and dynamic features with all other components in the model. For example, a detailed description of nucleotide metabolism, including 12 genes, has been successfully integrated into such a whole-cell model.[78]

A key future goal would be to combine theoretical findings with experimental genome minimization efforts. First, it remains to be seen whether computationally predicted minimal gene sets would be able to maintain a living cell. Second, general theoretical conclusions on the nature of alternative minimal genomes desperately need experimental verifications: it is unclear, for example, whether functionally (nearly) equivalent alternative minimal gene sets may be achieved experimentally.

5. Techniques of Experimental Genome Reductions

5.1. General Strategies

Theoretically, culturing a population of cells by serial passage under conditions favoring loss of genetic material could lead to smaller genomes. Selection for faster replication or reduction of the energy burden on cells growing in poor nutrient conditions combined with periodical population bottlenecks allowing fixation of mutations have been suggested to favor deletional processes.[80,81] Defects of the methyl-directed mismatch repair system might also result in increased deletion rates.[82] However, there is no obvious relationship between genome size and doubling time. The primary determinant of growth rate seems to be translational efficiency.[83] Experimental evolution by serial passage, designed to test spontaneous genome reduction in *Salmonella enterica*, yielded a 0.05−2.5 bp per chromosome per division deletion rate.[47] This rate is too low for practical applications. A much higher rate of random deletion generation could be achieved by combination of engineered composite transposons and serial passage.[84] An attractive feature of this approach is that it allows us to explore different order of deletion events, possibly leading to alternative minimal gene sets. However, it also suffers from at least two major drawbacks. First, an adequate selective protocol for smaller genomes is still lacking. Second, by the very nature of this strategy, deletions are largely random with respect to genomic location and hence can only be identified by laborious genetic mapping or resequencing the genome.

In contrast, targeted approaches provide straightforward, controlled genome reduction schemes. In well-studied organisms molecular genetic engineering tools allow rapid, precise deletion construction, and the depth of the genomic and physiological knowledge of the target cell allows hypothesis-driven deletion path design. Therefore, genome reduction by rational design, i.e., by targeted removal of genes judged to be dispensable on the basis of available empirical knowledge, is currently the most powerful approach.

Physicochemical specificity of a selected genomic locus lies in its nucleotide sequence. This specificity provides the basis for targeting the locus by the homologous recombination machinery, a ubiquitous enzymatic system involved primarily in DNA maintenance.[85] Statistically, an 11-base segment represents a unique sequence in a 4-megabase bacterial genome. In practice, longer (40−1000 bp) segments are usually involved in the gene-targeting process depending on the mode of action of the particular recombination enzymes exploited in the engineering process. Deletion constructions usually require two consecutive recombination steps. First, an artificially assembled DNA construct, carrying one or two "homology arms" (DNA segments homologous to the genomic regions flanking the desired deletion) and a selectable marker gene, is inserted into the targeted locus of the genome by homologous recombination. In the second recombination step the marker gene is excised from the genome, leaving behind a clean deletion, preferably devoid of all exogenous sequences ("scarless deletion"). Otherwise, littering the genome with remnants of the constructions (e.g., target sites of site-specific recombinases or resistance markers) can result in polar effects (e.g., turning off or on neighboring genes) or genomic rearrangements and can prevent subsequent, serial manipulations.[86,87]

These basic recombination steps come in many variations, both species- and laboratory-specific, involving a plethora of enzymes. Circular plasmid DNA as well as linear DNA segments can be used to target the genome. Homologous (general) recombination is mediated by the cell's own enzymatic (RecA-centered) machinery or phage-borne recombinases (e.g., RecET[87] or lambda Red system[88,89]). Excision of the exogenous sequences can be selected for by application of counterselectable markers.[90] Site-specific (Cre/*lox*,[91,92] Flp/*frt*[93,94]) recombination is frequently used to facilitate the excision step. General, P1 phage-mediated transduction[95] can complement the engineering arsenal, allowing assembly of individually constructed deletions in a single strain.

5.2. Basic Deletion Methods

5.2.1. Suicide Plasmid-Mediated Procedures

Suicide plasmids are convenient delivery vehicles of DNA constructs destined for insertion into the genome.[96] The plasmids can be made nonreplicative in the target host by creating nonpermissive conditions for replication (e.g., by inactivating a heat-labile replication initiator protein at higher temperatures[97] or withdrawal of a replication protein supplied in trans[98,99]). Alternatively, the plasmid construct can be made in a permissive host and then electroporated in a target cell not supporting replication.[10,100,101] The method is outlined in Figure 1. A suicide plasmid, carrying a targeting DNA fragment composed of two homology arms (500−1000 bp each), assembled by recombinant PCR and homologous to the flanking regions of the planned deletion, is transformed in the cell. The plasmid can integrate into the chromosome via RecA-mediated single crossover involving one of the homology arms and the corresponding chromosomal region. The cointegrate is selected by its antibiotic resistance at the nonpermissive conditions for plasmid replication. It is noted that the flanking homology arms are duplicated in the genome at this point. Resolution of the cointegrate can occur spontaneously by intramolecular recombination involving either pair of homologies, resulting in either restoration of

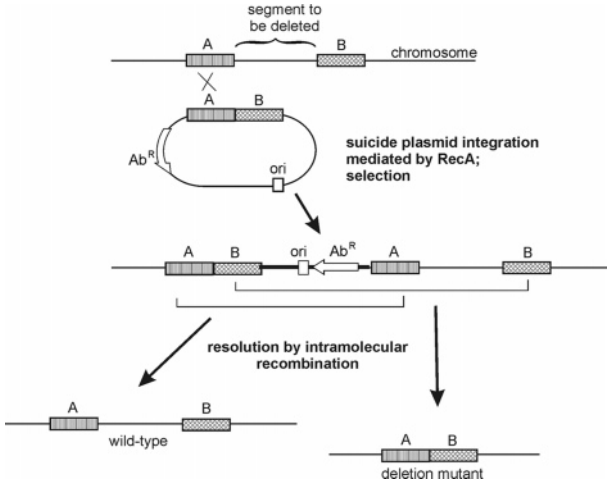

Figure 1. General scheme of the suicide plasmid-based deletion procedure. Boxes A and B represent >500-bp DNA segments flanking the genomic region to be deleted. Ab^R stands for an antibiotic resistance marker gene; ori indicates a replication origin functioning only under permissive conditions.

the wild-type sequence or the desired, scarless deletion. To identify the resolved products, screening or counterselection schemes, based on the integration of additional marker genes in the suicide plasmid (e.g., *lacZ* for screening,[102] *sacB*[103−105] or *rpsL*[68] in *E. coli* and *upp*[106] in *B. subtilis* for counterselection), have been developed.[90,107,108] This system can be improved even further by facilitating excision of the integrated plasmid. One example for such a mechanism is induction of replication of the integrated plasmid by providing permissive conditions. This activity not only enriches the culture for resolved cells by impairing chromosomal replication in unresolved ones[97] but also, in the case of rolling-circle plasmid replication, can increase the rate of cointegrant resolution.[109] Another efficient mode of enhancing the excision step is based on introduction of an 18-bp meganuclease cleavage site[110] in the suicide plasmid. Cleavage of the genome at this unique site creates a double-strand break, stimulating recombination and selecting for the cointegrate resolution simultaneously.[111] In any case, a final screen (e.g., phenotypic analysis, PCR test, Southern hybridizadion, or sequencing) is needed to distinguish the wild-type and the deletional products of resolution.

5.2.2. Linear DNA-Mediated Procedures

Linear DNA-based methods[87,88,112,113] are particularly suitable for large-scale genome modifications because no time-consuming cloning steps are required. (Variants of the method, worked out for *E. coli*, are depicted in Figure 2.) Targeting fragments, carrying a selectable marker gene and terminal homology arms, are assembled *in vitro* by PCR and transformed in the host cell. The construct is integrated into the genome by a double crossover via the terminal homologies, replacing the segment to be deleted. To evict the inserted exogenous sequences in a second recombination step, various alternatives have been worked out including (i) site-specific recombination between specific sequences flanking the marker gene,[86,87] (ii) replacement of the insertion with a second targeting fragment comprising only of the flanking homologies, combined with a selection scheme based on a counterselectable marker carried by the first fragment,[87,108] and (iii) excision by intramolecular recombination via duplicated segments created by the inserted fragment.[28,105,106]

The procedure developed for *B. subtilis* by Fabret et al.[106] is based on the ability of the bacterium to bind, take up, and integrate exogenous DNA in a naturally competent physiological state. During internalization the DNA is cleaved randomly and one of the strands is degraded. For efficient integration the internalized strand must carry relatively long (>400−500 bp) terminal homologies. Subsequent removal of the inserted selection marker is selected for on 5-fluorouracil-containing plates by the *upp* counterselection scheme, involving intramolecular recombination between short flanking direct repeats (30 bp) included in the design of the targeting fragment.

In *E. coli* internalization of double-stranded linear DNA can be achieved by electroporation.[114] However, the DNA is rapidly degraded in the cell by the exonuclease activity of the RecBCD complex,[115] precluding simple use of PCR-generated fragments. The stability problem is circumvented by use of bacteriophage-encoded recombination systems. In a popular scheme the λ phage *red* and *gam* genes are transiently expressed in the host cell prior to electroporation.[86,88,89,116] While Gam inhibits the exonucleolytic activity of RecBCD,[117,118] Red Exo (5′-3′ exonuclease[119,120]) and Red Bet (ssDNA binding protein[121,122]) promote recombination. Since efficient recombination by the Red system requires only short (40−60 bp) terminal homologies,[89,116] short homology extensions of PCR primers allow simple assembly of the targeting fragments. For efficient marker gene excision, enhancement techniques (e.g., introduction of a double stranded break [DSB] between duplicated segments[28] or use of counterselectable markers[106]) are usually needed.

5.2.3. Additional Tools: Site-Specific Recombinases and Transposons

Site-specific recombinases Cre[91,92] and FLP[93,94] are efficient tools for various genome manipulations. Randomly integrating transposons as well as targeting fragments inserted into the genome by either a suicide plasmid- or by a linear DNA-mediated method are often designed to carry recognition sites for site-specific recombinases. Recombination between these sites can then be used to remove marker genes or delete genomic segments between two insertions.[86,87,123−125] The frequently used Cre/*lox* and FLP/*frt* systems share many features. Both Cre and FLP recognize specific, 34-bp sequences (called *loxP* and *frt*, respectively) and catalyze cleavage and ligation of a pair of sites in a wide range of hosts.[87,102,123,124] If two recognition sites *in cis* are in the same directional orientation, the reaction results in deletion of the genomic segment flanked by them. The process is highly efficient; however, from the point of view of serial genome manipulations, it has the major drawback of leaving a single recognition site in the genome. Multiple use of the recombinase would result in accumulation of recognition sites, making recombination unpredictable. This problem can be circumvented using mutant recombinase target sites. The 34-bp recognition site consists of a 13-bp palindrom interrupted by an 8-bp asymmetric sequence. A pair of recognition sites carrying a single mutation each, on different halves of the palindrom, can be good substrates of the recombinase.[126,127] In turn, recombination between the single-mutant sites results in a double-mutant site, which is a poor substrate, and will not interfere with further manipulations involving the recombinase. This ingenious solution allows repeated use of

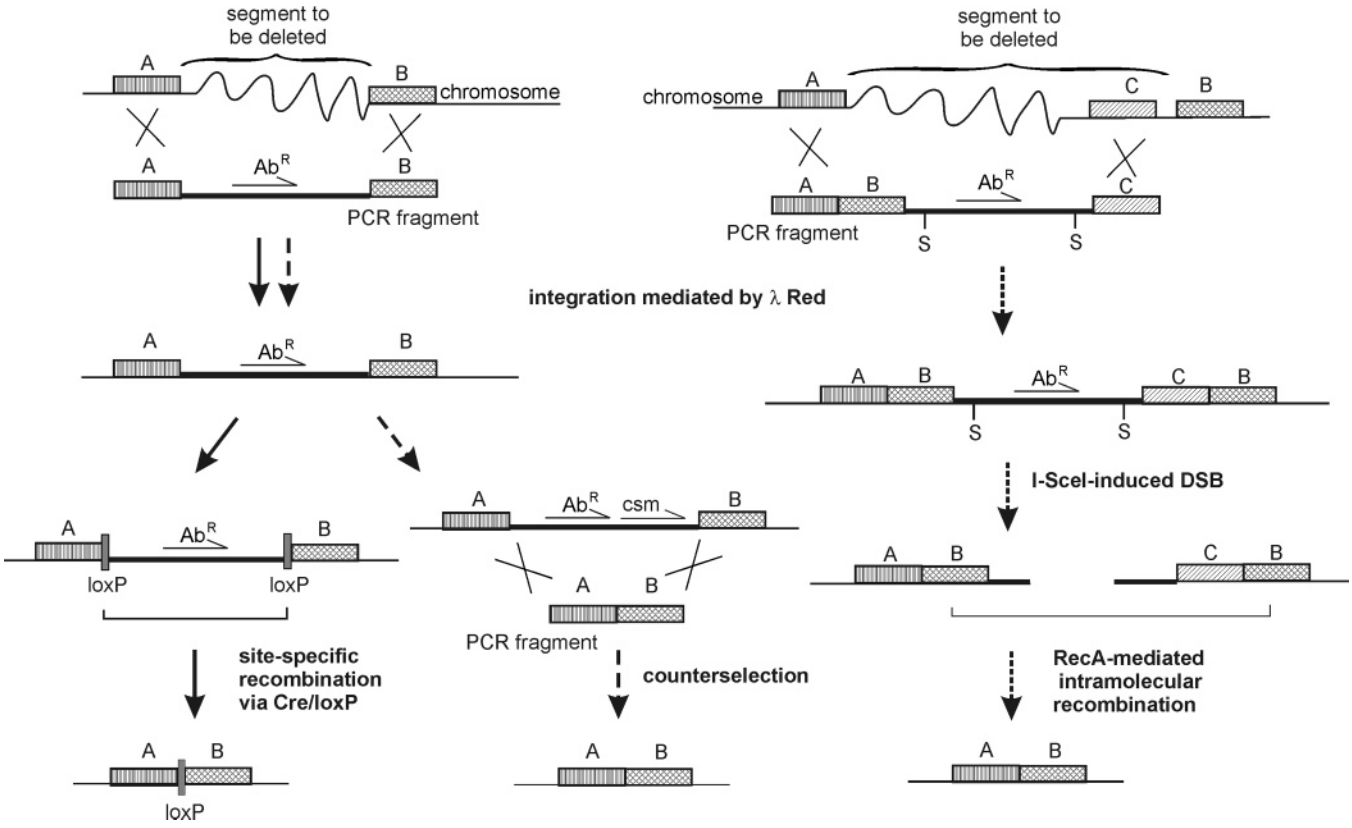

Figure 2. Overview of the λ Red-mediated, linear DNA-based deletion method. Arrows in different styles indicate three alternative routes for generating deletions. A, B, and C represent arbitrarily chosen 40−60-bp DNA segments (homology boxes). Ab^R and csm stand for an antibiotic resistance marker and a counterselectable gene, respectively. Target sites for Cre recombinase are indicated as loxP. S represents an I-SceI cleavage site.

the site-specific system;[128] however, the deletion created is not scarless, and problems of genomic instability and polar effects associated with introducing repeat sequences in the genome still persist.

Imprecise excision of transposons can generate random deletions in the chromosome.[129,130] A Tn5-based deletion technology was described for *E. coli*; however, due to the promiscuity of Tn5, it should be applicable to a wide range of bacterial species (Figure 3). A composite Tn5 derivative, carrying an external and a different internal pair of transposon ends, is integrated in the genome by electroporating preformed transposome complexes in the cell. Integration occurs via the external ends. A mutant transposase, active on the internal ends, is then expressed from the transposon. Intramolecular transposition via the internal ends can produce either inversions or deletions of various sizes, extending from the site of the original insertion. Deletion formation results in loss of transposon DNA, with the exception of a linker sequence. The procedure can be repeatedly used in the same cell, resulting in a series of random, scarred deletions.[84]

6. Genome Reduction Projects

6.1. *E. coli*

E. coli K-12, despite its relatively large genome size (∼4500 genes[131]), is an ideal subject of genome reduction projects. Due to its long history as a favored model organism in the laboratory, a wealth of physiological and genomic knowledge has accumulated over the years. Simple culturing conditions, the availability of extensive molecular genetics tools, and its short generation time make the technically challenging genome minimization feasible. Thus, it is not surprising that most genome reduction efforts to date have focused on this organism. In addition to the knowledge the genome minimization generates, the practical aspects of the projects are also motivating. *E. coli* has been extensively used for production of DNA, recombinant proteins, and metabolites of pharmaceutical, industrial, and agricultural interests. Streamlining the genomes might result in cell factories with enhanced production capabilities.

Comparison of the pathogenic O157:H7 EDL933 and CFT073 and the nonpathogenic K-12 MG1655 genomes revealed a mosaic-like structure: the common backbone of gene strings was interrupted by genomic islands (GIs) specific for the particular strain.[25,26] The backbone regions generally encode basic core functions that are necessary regardless of the environmental niche. GIs are presumably horizontally acquired DNA fragments, which contain a disproportionate share of genes with unknown functions as well as genes for toxins, virulence factors, and metabolic capabilities that may be advantageous in the niche the strain is adapted to. Islands may also be loaded with parasitic or "junk" DNA: transposable elements, phages, pseudogenes, and gene remnants. Strain-specific islands (K-islands) represent ∼20% of the total K-12 MG1655 genome. It can be argued that since *E. coli* evolved in the intestinal tracts of animals, it has many genes, most of them coded on GIs, that are not relevant to practical laboratory or industrial applications and some that may be detrimental. This notion is supported by the findings that even under poor nutritional conditions, only 75−80% of the genes are expressed at detectable levels.[132,133]

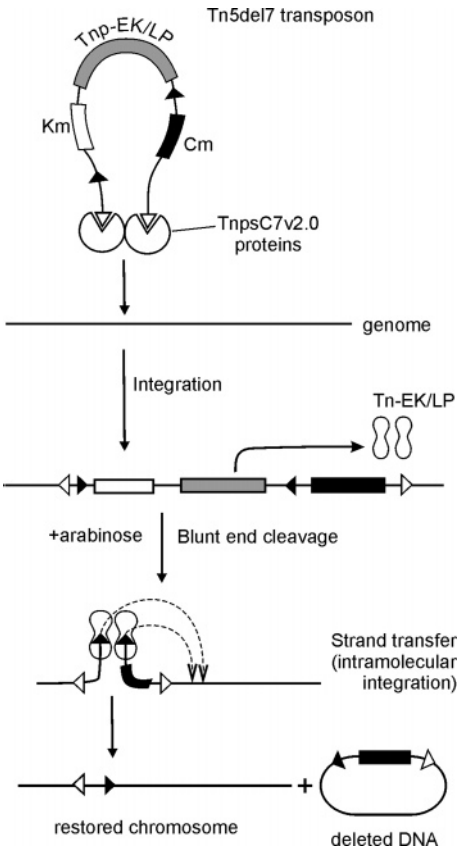

Figure 3. Details of the Tn5-based deletion technique.[84] Km and Cm denote resistance genes for Kanamycin and Chloramphenicol, respectively. White and black triangles represent the respective external and internal ends of the transposon. Tn-EK/LP is a mutant transposase, active only on internal ends. The deleted DNA can be rescued as a plasmid if an appropriate origin of replication is present next to the Cm gene.

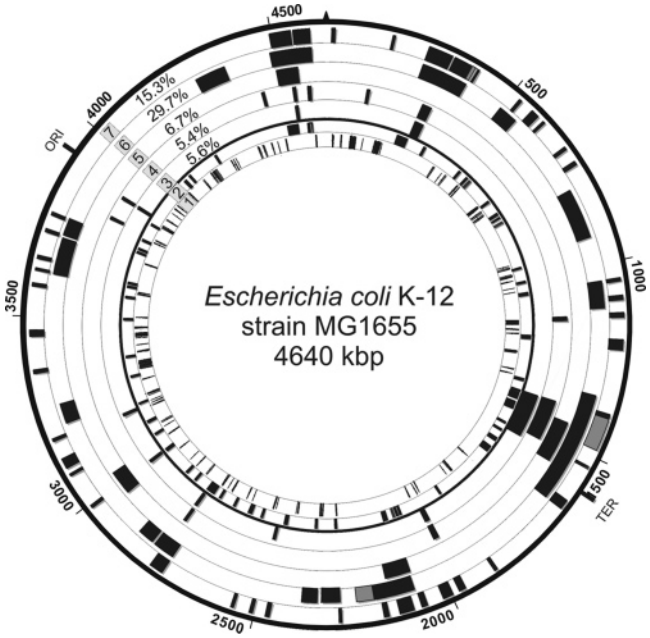

Figure 4. Deletion map of reduced-genome *E. coli* strains. Rings depict features mapped to the genome of *E. coli* K-12 MG1655, numbered on the perimeter in kbp. Outward from the center, (1) essential genes (www.shigen.nig.ac.jp/ecoli/pec/index.jsp), (2) strain-specific K-12 genomic islands longer than 4 kbp, (3) set of deletions constructed by Goryshin et al.,[84] (4) another set of deletions constructed by Goryshin et al.,[84] (5) set of deletions constructed by Yu et al.,[69] (6) set of deletions constructed by Hashimoto et al.,[68] and (7) set of deletions constructed by Pósfai et al.[31] The fraction of the genome deleted in each project is indicated on the respective ring. If two deletions were close to each other, one was colored gray for better visibility. ORI and TER indicate the origin and terminus of replication, respectively.

6.1.1. Random Genome Reduction by Transposon Technology

Specializing in transposon biology and utilization, Goryshin et al.[84] developed a powerful tool that can be applied to gene essentiality studies and minimal genome construction work.

The specialized Tn5 transposon-derived deletion system, used in the study, is described above (section 5.2.3). By repeating the random integration/deletion cycle 20 times, several different multideletion strains of *E. coli* MG1655 were constructed (Figure 3). Using pulse field electrophoresis, the extent of genomic reduction was estimated to lie between 100 and 262 kbp. DNA-microarray analysis was performed with the genomes of the two most extensively engineered strains for precise mapping of the deletions. Interestingly, only 9 and 11 chromosomal deletions were detected. This indicates that not all rounds resulted in loss of genomic material. Even more surprising was the fact that large segments of up to 146 kbp were deleted. Using a variation of this technique they managed to rescue the genomic segments neighboring different Tn insertions on a conditionally replicative plasmid and were thus able to investigate the essentiality of the genes carried by them. The advantages of the transposon-based deletion method are (i) no previous knowledge is needed concerning the sequence or dispensability of the targeted bacterial genes, (ii) the Tn5 transposase was shown to be active in all tested bacterial species, and (iii) the deleted segments can be saved as plasmids. These make the technique highly adequate for screening of different bacterial genomes for essential genes. Moreover, several different multideletional strains can arise in a single experiment, allowing directed evolution, provided that appropriate selective conditions can be imposed. On the other hand, the randomness of the integration site and deletion size renders this approach less useful for the precise, planned construction of a minimal genome. Also, a 64-bp linker is retained in the chromosome after each deletion event, and the probability of unwanted genomic rearrangements increases with the accumulation of sequence repeats.

6.1.2. Semirandom Genome Reduction

To demonstrate the feasibility of a combinatorial deletion technique in identification of essential genes and genome minimization, Yu et al.[69] performed a 6.7% reduction of the *E. coli* MG1655 genome.

Using modified Tn5 transposons, carrying either Kan or Cm resistance markers, two random Tn-insertion libraries, consisting of 400 mutants each, were constructed. The location and orientation of each insertion mutant was mapped by sequencing. Next, selected mutants were combined in a single cell using P1-transduction, and the chromosomal segment flanked by the insertions was excised by Cre-mediated recombination of the tandem *loxP* sites of the transposons. By repeating the procedure, four large deletions were accumulated in the genome, totaling 313 kbp (Figure 4). A total of 287 ORFs were removed, 179 of which encoded proteins of unknown function at the time of publishing. Not all deletions could be combined: "mutually

exclusive" deletions were thought to harbor genes coding for alternative essential functions. The growth rate of the deletion strain was measured in LB medium, and it did not significantly differ from that of the wild-type.

This semirandom deletion procedure has some limitations. In order to reach a saturated Tn5 mutant pool, where all nonessential genes are interrupted, an extremely large number of transposon mutants needs to be sequenced. Without such a saturated mutant pool not all deletions can be precisely engineered. Furthermore, use of this genome-modification system requires multiple steps, and the Cre/*lox* excision results in exogenous sequences remaining at the deletion sites. Cumulative piling of deletions in the genome thus requires removal of these exogenous sequences by λ Red-mediated manipulation. Nevertheless, the availability of mapped mutant pools allows the rapid construction of virtually any single genomic deletion, facilitating gene essentiality studies and specific strain constructions.

6.1.3. Genome Reductions via Targeted Strategies

In an attempt to probe the limits of genome reductions in *E. coli* MG1655, Hashimoto et al.[68] constructed a set of medium-sized and long deletions and combined some of them to maximize the total amount of DNA deleted (Figure 4).

Deletions were constructed using the linear DNA/λ Red-mediated method. In order to stabilize the targeting DNA fragment and facilitate its integration into the genome, deletions were constructed in a *rpsL* strain in which *recBC* was replaced with the λ *red* region. Targeting DNA fragments were constructed by a two-step PCR reaction and comprised of a *cat, rpsL, sacB* (CRS) cassette, flanked by short homology arms. While *cat* allows positive selection by chloramphenicol, *rpsL* and *sacB* are used for counterselection by streptomycin and saccharose, respectively. After electroporation of the fragment into the bacterial cell, genomic cointegrants were selected on Cm-containing plates. Such intermediate deletion constructs were transferred in a multideletion strain by P1 transduction. The CRS cassette was then eliminated by P1-transducing the cells with DNA fragments containing the "clean" deletion and applying counterselection. (Strains with individual "clean" deletions were separately constructed by transforming the cointegrant cells with the PCR-generated "clean" deletion joint and applying counterselection.)

End points for the 16 long deletions, ranging in size from 41 to 301 kbp, were planned by combining data from the literature concerning gene essentiality with previous experimental results of separately engineering 75 medium-sized deletions. By stepwise transduction of the 16 long-sized deletions into one cell a series of deletion strains was constructed with genome size reductions up to 29.7% (1 377 172 bp) of the parental chromosome. Deletion accumulation was confirmed by pulse-field gel electrophoresis and DNA microarray analysis. The DNA content of the deletion mutants was also measured by flow cytometry and found to be decreased compared to the parental strain.

Phenotypic analysis of the deletion mutants revealed a significant increase of the generation time of growing cells, more or less in parallel with the increasing number of deletions. In addition, a marked change of cell size and nucleoid organization was observed in the multideletion strains. Cells proved to be wider and those harboring 13 or more deletions were also found to be longer as compared to wild-type. Accumulation of deletions reduced cellular protein content by 23−25%. Also, instead of having a single nucleoid midcell or two nucleoids at 1/4 and 3/4 cell lengths, the multideletion cells had four or more randomly distributed nucleoids in the cellular periphery. The proportion of such cells increased in parallel with the combined deletion size. The cause of this abnormal nucleoid positioning remains unclear. The phenotypic traits listed above were not observed on the individual deletion strains, which brought about the conclusion that they arose as the result of synergistic defects in multiple genetic systems controlling related cellular processes.

In a mostly technical paper Fukiya et al. reported the individual deletions of a 117 and a 165 kbp region of *E. coli*.[125] Their method involves integration of *loxP*-containing DNA fragments by the Red system into the two ends of the target sequence. This is followed by introduction of the Cre recombinase into the cells on a plasmid, leading to efficient deletion of the genomic fragment flanked by the *loxP* sites. Retainment of one *loxP* site and an antibiotic marker within the genome, however, necessitates use of further engineering techniques if serial accumulation of deletions is desired. Nevertheless, the Red−Cre/*lox* technique can be a useful engineering tool, especially for construction of extremely large deletions.

The genome reduction work by Kolisnychenko et al.[28] targeted the standard laboratory strain *E. coli* K-12 MG1655. The objective of this research was to eliminate as many of the unnecessary genes as possible without interfering with beneficial *E. coli* characteristics, including the robust metabolic performance and rapid doublings in both rich and minimal media.

Precise, scarless deletions were constructed mostly by a λ Red-mediated method. The PCR-generated targeting fragments were constructed to carry the actual postdeletion sequence joints. Thus, integration of a fragment by λ Red-type recombination creates a duplication of the segment flanking the joint. Cleavage of the inserted DNA by meganuclease I-SceI 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 genomic deletion (Figure 2). Individually constructed deletions were accumulated in a single strain in a cyclic fashion: deletion intermediates carrying the selection marker gene were sequentially transferred into the target cell by P1 transduction, and the deletion was made scarless by DSB-stimulated recombination.

Deletion target selection was primarily based on identification of strain-specific genomic islands by comparative genomics involving three *E. coli* strains. The largest K-islands were selected as the targets of the deletion work. Deletions were in some cases extended into neighboring genes judged to be dispensable based on the available experimental evidence. A total of 12 segments, ranging in size from 7 to 82 kbp, were sequentially deleted from the MG1655 chromosome, resulting in an 8.1% (or 376 kbp) reduction in genome size. The multiple deletion strain (MDS) with 12 deletions (MDS12) displayed growth rates similar to that of the parent strain in both rich and minimal/glucose media.

Following this pilot project, trimming of the *E. coli* genome was continued. Starting with MDS12, Pósfai et al. constructed a series of MDS cells up to MDS43, a strain carrying 43 deletions[31] (totaling 15.3% of the genome or 743

genes) (Figure 4). To identify deletion targets (primarily GIs and mobile genetic elements) genome comparisons were extended to six sequenced *E. coli* genomes.

One of the major goals of this round of deletions was to remove all mobile genetic elements from the genome. Prophages, insertion elements, transposases, and multiple sequence repeats (e.g., Rhs elements) disseminated throughout the genome, mediate genomic rearrangements, including inversions, deletions, transpositions, and horizontal gene transfers.[134] A major consequence of the lack of mobile elements was an increased genomic stability. It was shown that the overall mutation rate of the genome decreased by 20−25%, and this decrease was due to the lack of ISs. Mobile elements in the chromosome of the host present a constant source of contamination of DNA propagated in the cell. Plasmid DNA prepared from non-MDS hosts was always contaminated with IS-containing DNA. In an extreme case a plasmid carrying an apparently toxic chimeric gene could be recovered from non-MDS hosts only in IS-mutated forms. In contrast, plasmids prepared from IS-less MDS cells were free of IS contamination, and the unstable plasmid could be cloned and recovered in unaltered form. Increased stability of plasmids with inverted repeat sequences was also observed in MDS cells. The streamlined strain MDS42 (42 deletions) was tested for electroporation efficiency. Surprisingly, the deletion strain performed 2 orders of magnitude better than MG1655. The increased efficiency was probably the result of multiple, uncharacterized synergistic interactions resulting from removal of a number of genes (>180) which are known or predicted to be associated with the cell membrane. One could speculate that their cumulative influence on the membrane composition might result in better access for DNA to the depolarized membrane.

6.1.4. Lessons of Parallel E. coli Genome Reductions

It is important to formulate the goals of a particular genome reduction approach. Random and semirandom reduction schemes[69,84] can be useful for gene essentiality studies or directed evolution experiments. A targeted approach aiming at construction of improved biotechnological strains, in conjunction with an empirical data-driven procedure to carefully select deletion targets, has already produced reduced-genome cells (MDS) displaying emergent beneficial properties.[31] More extensive deletions aimed at maximizing the extent of reduction yielded the smallest *E. coli* genome to date (Δ16), but the somewhat indiscriminative removal of large segments resulted in aberrant cell morphology and increased doubling time.[68]

It is becoming clear that, in accordance with early predictions,[135] a large proportion of the genome is not needed under defined conditions in the laboratory. It is also clear that there are several different but partially convergent pathways leading to reduced *E. coli* genomes. This is illustrated by the differences and overlaps in the targets of the various approaches (Figure 4). Nineteen of the 43 deletions in MDS43[31] (15.3% reduction) represent a partial subset of the 16 large deletions (29.7% reduction) present in the strain maximally reduced to date.[68] The other 24 of 43 deletions are, however, unique to MDS43.

Surprisingly, a large number of deleted genes are listed as essential in a transposon mutagenesis-based gene-essentiality study.[16] For instance, MDS43, a strain with no apparent disabilities, lacks 121 genes marked as essential by Gerdes et al.[16] This underscores inherent uncertainties of the high-throughput transposon mutagenesis method: the transposon mutant pool might not be large enough to include hits of all nonessential genes, growth disadvantage of some mutants might mask their dispensability, and some individually lethal mutants might be viable in combination with other mutations. We note that a list of essential *E. coli* genes (PEC; http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp) identified by more stringent criteria, including the availability of conditionally lethal mutants or failure of obtaining deletion mutants, shows no discrepancy with the MDS deletions: all genes classified as essential are left intact in the genome of MDS43 (Figure 4).

All systematic deletion efforts view the genome as a set of genes and ignore the poorly characterized structural organization of the nucleoid or other potential higher-order spatial genome patterns.[136] The surprisingly large average size of deletions generated *in vivo* by the transposon-mediated technique employed by Goryshin et al.[84] contrasts their *in vitro* results and supports the notion that chromosomal DNA has a compact nucleoid body with supercoiled domains that brings distant points into close proximity. However, the remarkable tolerance of the multiple deletion *E. coli* cells indicates that the structural requirements of higher organization might be flexible enough to accommodate large-scale changes of the genome.

The two replichores (oppositely replicating halves) of the *E. coli* genome are nearly identical in size. Rearrangements affecting this size balance can have deleterious effects on replication.[137,138] The moderate change in replichore length difference in the MDS cells (up to 183 kbp) had no significant impact.[31] On the other hand, the Δ16 strain displayed a reduced growth rate that can be due to the larger difference of the replichores (256 kbp) and/or removal of the terminus region.[68]

The chromosomal architecture was also shown to have an impact on gene expression: the orientation and chromosomal position of certain highly expressed genes seem to be shaped by selective pressure. For instance, ribosomal RNA genes are co-oriented with DNA replication and clustered in the proximity of the replication initiation region.[139,140] This ensures efficient transcription and high gene dosage in fast-growing cells due to the simultaneous progression of multiple replication forks. Apparently, the serial deletions of MDS cells do not significantly perturb these architectural effects.

6.2. *B. subtilis*

The well-studied Gram-positive *B. subtilis* is a favored organism of the fermentation industry. *B. subtilis* is considered to be a GRAS organism (generally recognized as safe), contains no lipopolysaccharides (endotoxin) in its cellular boundary, and has a natural transforming ability and a high secretory capacity to export proteins into the surrounding medium.[141] These advantages underline the importance of genome modification techniques available in this species and explain the fairly large amount of genome engineering work that has been carried out in attempt to optimize *B. subtilis* for industrial use.

A prototype general system for generating unmarked gene replacements in *B. subtilis* was developed by Leenhouts et al.[98] This technique is a suicide plasmid-mediated, homologous recombination-based genome engineering method, adequate for use in cells which can be electroporated efficiently. Using this method several genes of *B. subtilis* and *L. lactis* were mutated. In these studies only a few

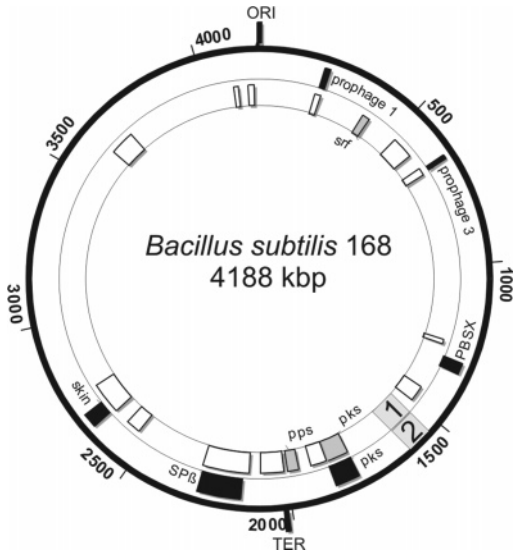

Figure 5. Genome reduction map of *B. subtilis* 168. Nucleotide numbering is shown on the perimeter in kbp. Outward from the center, rings depict (1) AT-rich islands[166] (>60% in sliding windows of 10 kbp with a step of 5 kbp; white boxes) and gene clusters involved in the synthesis of polyketide and peptide antibiotics[30,151] (shaded boxes) and (2) set of deletions constructed by Westers et al.[30] ORI and TER indicate the origin and terminus of replication, respectively.

hundred base pairs were deleted from the genome. However, serial deletion of large genomic segments can be achieved with the use of this technique or its variants.

Systematic genome reduction of *B. subtilis* targeted large, presumably dispensable regions, including prophages (SPβ, PBSX), prophage-like elements (prophage 1, prophage 3, *skin*), and the polyketide synthase (*pks*) operon, ranging in size from 13 to 134 kbp[30] (Figure 5). These regions are potentially important for survival in soil but seem to be expressed at low levels under laboratory conditions. Dorenbos et al.[142] constructed a *B. subtilis* genome deleted for prophage SPβ. This strain was the starting point of a large-scale reduction project, conducted by Westers and co-workers, leading to *B. subtilis*Δ*6*, a strain harboring six deletions.[30] Interestingly, several variations of plasmid-based deletion techniques were used to construct this strain. Deletion of the *skin* element was performed by cloning the flanking regions into a plasmid carrying a temperature-sensitive rolling circle origin of replication.[109] After transformation of the target cells and selection of genomic cointegrants the second recombination leading to plasmid excision was facilitated by activation of the rolling circle replication. The PBSX region was deleted in a similar manner. Next, prophage 1 was deleted using another suicide plasmid.[98] The *pks* operon region was deleted separately by use of an integration vector[143] with an antibiotic-resistance marker remaining in the chromosome, allowing transformation of the multideletional cell to yield a genome with five deletions. Finally, prophage 3 was deleted in a multistep procedure using additional suicide plasmids.[144] Correct deletion sizes were confirmed by PCR and/or Southern blotting.

Deletion of the six genomic segments reduced the *B. subtilis* genome by 320 kbp (7.7%). An extensive comparison of the multiple-deletion and parental strain concluded that the characteristics relating to growth, viability, carbon metabolism, protein secretion, competence, and sporulation were unaffected by the deletions. Compared to wild-type, the multiple-deletion strain was found to have a reduced motility in 0.25% agar but an increased motility in 0.5% agar. The secretion rate of a heterologous amylase protein (AmyQ), though, was not changed by the deletions, indicating that no large energy resources were redirected toward product formation or secretion. However, *B subtilis*Δ*6* has one major advantage over the wild-type: deletion of the BsuM restriction-modification system can increase the transformation efficiency and plasmid stability in this host.[30] In summary, the fact that the deletions left the physiology of *B. subtilis* virtually unaffected confirms the feasibility of large-scale genome reduction in this species. *B. subtilis*Δ*6* could be a convenient starting point for future cell-optimization projects, especially those aiming to minimize the indigenous contamination of industrially produced molecules.

6.3. *Corynebacterium glutamicum*

Recently, *C. glutamicum* also became the object of extensive genome engineering. *Corynebacterium* strains are widely used for industrial production of amino acids, DNA, and organic acids.[145,146] Suzuki et al.[102] identified strain-specific islands (SSI) of the *C. glutamicum* R genome by comparison to strain ATCC 13032 and developed several versions of the Cre/*lox* method to delete them. In all cases, plasmids with conditional replication origins were electroporated into *C. glutamicum* R and recombined into the genome. These plasmids are easily propagated and engineered in *E. coli* but do not replicate in *Corynebacterium*. Two plasmids, carrying different resistance markers and a *loxP* site each, were inserted into the two ends of the planned deletion via cloned homology arms. After selection the recombinants were transformed with a plasmid constitutively expressing the Cre recombinase. In every case this leads to loss of the genomic segment flanked by the inserted plasmids, leaving a *loxP* site and a resistance marker behind.

Using a second, markerless version of the method Suzuki et al.[29] deleted 11 genomic islands of *C. glutamicum* R ranging in size from 9.8 to 55.6 kbp harboring from 4 to 58 ORFs, respectively. However, the individual deletions were not accumulated in a single genome, most likely due to the undesirable recombination potential of the multiple *loxP* sites remaining in the genome. To circumvent this two further techniques suitable for accumulating deletions within one cell were developed.

In the first improved method[105] one of the homology arms is duplicated following integration of the construct into the genome, similarly to the technique developed by Kolisnychenko et al.[28] This provides adequate substrates for a second homologous recombination event resulting in scarless loss of the flanked genomic segment. This recombination step is facilitated by I-SceI-generated DSBs. A *sacB* and a *lacZ* gene, carried on the integrated plasmids, provide proper selection and screening for this step, respectively. To increase the ratio of recombinants losing the segment to be removed, the DSB induction is combined with the action of the Cre recombinase. This way, 25−50% of the resultant colonies lose the targeted segment. The cells are cured of the plasmid expressing Cre and I-SceI prior to the next deletion by culturing in complex medium without antibiotics. In their work, Suzuki et al.[105] described the consecutive deletions of three genomic islands of *C. glutamicum* R, removing 73 genes and reducing the genome size by 67 kbp (Figure 6).

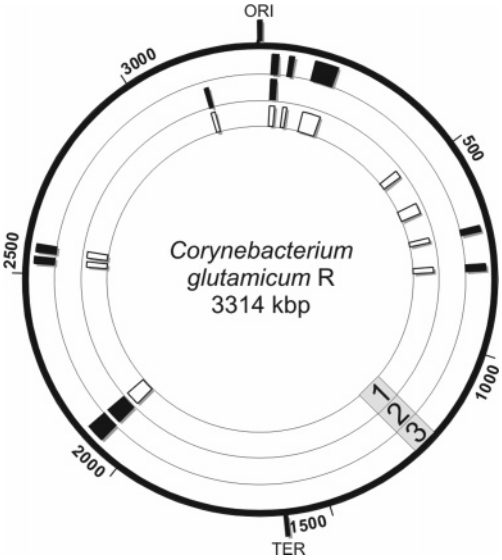

Figure 6. Genome reduction map of *C. glutamicum* R. Nucleotide numbering is shown on the perimeter in kbp. Outward from the center, rings depict (1) strain-specific islands longer than 10 kbp[29] (white boxes), (2) set of deletions constructed by Suzuki et al.,[105] and (3) another set of deletions constructed by Suzuki et al.[128] ORI and TER indicate the origin and terminus of replication, respectively.

The second method to serially introduce deletions into the *Corynebacterium* genome was employed to engineer an even more extensive genome reduction.[128] Using suicide plasmids different mutant *loxP* sites were introduced into the ends of the segment to be deleted. These single-point-mutant *loxP* sites are still good substrates of Cre. However, after their resolution, a double-mutated *loxP* remains in the genome, which, due to its low affinity to Cre, does not interfere with further genome manipulations involving the Cre/*lox* system. Eight SSIs were deleted this way, removing 188 genes from the *C. glutamicum* R genome, reducing its size from 3.31 to 3.12 Mb (Figure 6). Deleted genes included transposable elements, prophages, genes for phenylacetic acid degradation, and many genes of unknown function. Colony morphology and growth characteristics were investigated in minimal medium. In all cases, reduced-genome strains were found to be unchanged as compared to wild-type. These strains might represent a starting point for further streamlining and can provide a suitable host for merging beneficial mutations of production strains obtained by classical breeding techniques.

7. Concluding Remarks

Engineered biological systems have been used for countless human purposes. The objects of these engineering efforts were, basically, naturally evolved cells, adapted to promote their own survival in a particular environment. Recently, however, large-scale remodeling of cells to produce new, *tabula rasa* strains became possible, and rational genome reduction of bacterial cells can be an important avenue of deciphering the basic building plan of the cell and producing streamlined minimum genome cell factories[147] for industrial purposes.

Currently, most systematic deletion approaches are based on targeted removal of genomic segments. The segments are selected by a comprehensive strategy based on comparative genomics, gene essentiality analysis, and metabolic pathway studies. Despite the differences concerning the target cells and engineering techniques and of the *ad hoc* selection of reduction pathways some common themes emerge.

Several bacterial genomes display a dynamic, mosaic-like structure where the core genome is interrupted by horizontally acquired genomic islands.[25,26,29,148−153] These islands are loaded with phages, transposons, and genes with niche-specific metabolic and regulatory functions. The genomic islands can be relatively easily identified by comparing genomes of close evolutionary relationship, and removal of them could usually be accomplished without any deleterious effect on basic cell physiology. Many genome reduction projects targeted these genomic islands,[28,30,128] and the resulting cells represent various stages of core−genome reconstructions. It seems plausible that full core−genome constructs are technically and biologically achievable, and these "first-generation" minimal cells would perform comparably or even better than wild-type cells. This is well exemplified by the *E. coli* projects: removal of >40 genomic segments, carrying mobile genetic elements and genes with unknown or unnecessary functions, resulted in improved genomic stability and tolerance of some toxic clones.[31] A full core−genome *E. coli* is not out of reach: raising the number of deletions to ∼100 (or total genome reduction from the current 15% to ∼20%) would eliminate essentially all gene-sized genomic islands.

While comparative genomics can provide the guideline to first-generation minimal cells, construction of second-generation cells would require carving into the core functions of the cell. A key to this goal would be to successfully combine experimental approaches with theoretical knowledge of cellular networks, including computational predictions based on metabolic and regulatory network reconstructions.[154,155] On the other hand, experimental verification of theoretical predictions would greatly advance the model-building efforts.

As discussed above, a mere set of genes does not fully represent the information content of the genome. Beyond the genes, higher-order chromosome structures might also have a profound effect on cell physiology.[156] More extensive genome reduction plans will most likely benefit from a better understanding of the large-scale architectural organization of the genome.

It is anticipated that, starting with first-generation minimal cells, further reductions will take many specific directions. Construction of minimal cell factories for different products (e.g., DNA, recombinant proteins, or small molecules) would require different gene deletions. In addition to deletions, heterologous genes or gene circuits could be added to the genome to enhance certain characteristics.

An interesting avenue of minimal genome constructions can be exploitation of the evolving nature of living systems: combination of targeted modifications with evolutionary adaptation.[154,157,158] New ultra-low-cost sequencing technologies[159,160] will allow monitoring[161,162] compensating mutations at the whole-genome level in response to deletions. These data, integrated with transcriptome, proteome, and metabolic flux analysis,[163] will help mapping interactions in the cellular network.

The surprisingly extensive robustness and tolerance of bacterial cells to large-scale genome modification, predicted by systems biology approaches and confirmed by experimental reductions, indicates the enormous potential that lies in biological systems and awaits exploitation.

8. Acknowledgements

B.P. is a Fellow of the Human Frontier Science Program, while C.P. is supported by an EMBO postdoctoral fellowship and the Hungarian Research Fund (OTKA). G.P. is supported by the Hungarian Research Fund (OTKA).

9. Note Added in Proof

Three recent reviews report on further ongoing genome reduction projects in *E. coli*,[167] *B. subtilis*,[168] and the unicellular eukaryote *Schizosaccharomyces pombe*,[169] aiming at constructing minimum genome cells dedicated to protein production.

10. References

(1) Neidhardt, F. C.; Umberger, H. E. In *Escherichia coli and Salmonella thyphimurium, cellular and molecular biology*; Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umberger, H. E., Eds.; American Society of Microbiology: Washington, DC, 1996; Vol. 1.
(2) Fraser, C. M.; Gocayne, J. D.; White, O.; Adams, M. D.; Clayton, R. A.; Fleischmann, R. D.; Bult, C. J.; Kerlavage, A. R.; Sutton, G.; Kelley, J. M.; Fritchman, R. D.; Weidman, J. F.; Small, K. V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T. R.; Saudek, D. M.; Phillips, C. A.; Merrick, J. M.; Tomb, J. F.; Dougherty, B. A.; Bott, K. F.; Hu, P. C.; Lucier, T. S.; Peterson, S. N.; Smith, H. O.; Hutchison, C. A., III; Venter, J. C. *Science* **1995**, *270*, 397.
(3) Waters, E.; Hohn, M. J.; Ahel, I.; Graham, D. E.; Adams, M. D.; Barnstead, M.; Beeson, K. Y.; Bibbs, L.; Bolanos, R.; Keller, M.; Kretz, K.; Lin, X.; Mathur, E.; Ni, J.; Podar, M.; Richardson, T.; Sutton, G. G.; Simon, M.; Soll, D.; Stetter, K. O.; Short, J. M.; Noordewier, M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12984.
(4) Gil, R.; Sabater-Munoz, B.; Latorre, A.; Silva, F. J.; Moya, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4454.
(5) Perez-Brocal, V.; Gil, R.; Ramos, S.; Lamelas, A.; Postigo, M.; Michelena, J. M.; Silva, F. J.; Moya, A.; Latorre, A. *Science* **2006**, *314*, 312.
(6) Lodish, H.; Berk, A.; Matsudaira, P.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Zipursky, S. L.; Darnell, J. E. *Molecular Cell Biology*; W.H. Freeman & Company: New York, 2003.
(7) Reed, J. L.; Vo, T. D.; Schilling, C. H.; Palsson, B. O. *Genome Biol.* **2003**, *4*, R54.
(8) Covert, M. W.; Knight, E. M.; Reed, J. L.; Herrgard, M. J.; Palsson, B. O. *Nature* **2004**, *429*, 92.
(9) Watson, J. D.; Baker, T. A.; Bell, S. P.; Gann, A.; Levine, M.; Losick, R. *Molecular Biology of the Gene*; The Benjamin/Cummings Publishing Company, Inc.: Madrid, 2004.
(10) Koob, M. D.; Shaw, A. J.; Cameron, D. C. *Ann. N. Y. Acad. Sci.* **1994**, *745*, 1.
(11) Koonin, E. V. *Nat. Rev. Microbiol.* **2003**, *1*, 127.
(12) Gil, R.; Silva, F. J.; Pereto, J.; Moya, A. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 518.
(13) Pál, C.; Papp, B.; Lercher, M. J.; Csermely, P.; Oliver, S. G.; Hurst, L. D. *Nature* **2006**, *440*, 667.
(14) Judson, N.; Mekalanos, J. J. *Trends Microbiol.* **2000**, *8*, 521.
(15) Hutchison, C. A.; Peterson, S. N.; Gill, S. R.; Cline, R. T.; White, O.; Fraser, C. M.; Smith, H. O.; Venter, J. C. *Science* **1999**, *286*, 2165.
(16) Gerdes, S. Y.; Scholle, M. D.; Campbell, J. W.; Balazsi, G.; Ravasz, E.; Daugherty, M. D.; Somera, A. L.; Kyrpides, N. C.; Anderson, I.; Gelfand, M. S.; Bhattacharya, A.; Kapatral, V.; D'Souza, M.; Baev, M. V.; Grechkin, Y.; Mseeh, F.; Fonstein, M. Y.; Overbeek, R.; Barabasi, A. L.; Oltvai, Z. N.; Osterman, A. L. *J. Bacteriol.* **2003**, *185*, 5673.
(17) Sassetti, C. M.; Boyd, D. H.; Rubin, E. J. *Mol. Microbiol.* **2003**, *48*, 77.
(18) Glass, J. I.; Assad-Garcia, N.; Alperovich, N.; Yooseph, S.; Lewis, M. R.; Maruf, M.; Hutchison, C. A., III; Smith, H. O.; Venter, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 425.
(19) Akerley, B. J.; Rubin, E. J.; Novick, V. L.; Amaya, K.; Judson, N.; Mekalanos, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 966.
(20) Kobayashi, K.; Ehrlich, S. D.; Albertini, A.; Amati, G.; Andersen, K. K.; Arnaud, M.; Asai, K.; Ashikaga, S.; Aymerich, S.; Bessieres, P.; Boland, F.; Brignell, S. C.; Bron, S.; Bunai, K.; Chapuis, J.; Christiansen, L. C.; Danchin, A.; Debarbouille, M.; Dervyn, E.; Deuerling, E.; Devine, K.; Devine, S. K.; Dreesen, O.; Errington, J.; Fillinger, S.; Foster, S. J.; Fujita, Y.; Galizzi, A.; Gardan, R.; Eschevins, C.; Fukushima, T.; Haga, K.; Harwood, C. R.; Hecker, M.; Hosoya, D.; Hullo, M. F.; Kakeshita, H.; Karamata, D.; Kasahara, Y.; Kawamura, F.; Koga, K.; Koski, P.; Kuwana, R.; Imamura, D.; Ishimaru, M.; Ishikawa, S.; Ishio, I.; Le, Coq, D.; Masson, A.; Mauel, C.; Meima, R.; Mellado, R. P.; Moir, A.; Moriya, S.; Nagakawa, E.; Nanamiya, H.; Nakai, S.; Nygaard, P.; Ogura, M.; Ohanan, T.; O'Reilly, M.; O'Rourke, M.; Pragai, Z.; Pooley, H. M.; Rapoport, G.; Rawlins, J. P.; Rivas, L. A.; Rivolta, C.; Sadaie, A.; Sadaie, Y.; Sarvas, M.; Sato, T.; Saxild, H. H.; Scanlan, E.; Schumann, W.; Seegers, J. F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Simon, M.; Stragier, P.; Studer, R.; Takamatsu, H.; Tanaka, T.; Takeuchi, M.; Thomaides, H. B.; Vagner, V.; van Dijl, J. M.; Watabe, K.; Wipat, A.; Yamamoto, H.; Yamamoto, M.; Yamamoto, Y.; Yamane, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.; Zuber, U.; Ogasawara, N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4678.
(21) Kang, Y.; Durfee, T.; Glasner, J. D.; Qiu, Y.; Frisch, D.; Winterberg, K. M.; Blattner, F. R. *J. Bacteriol.* **2004**, *186*, 4921.
(22) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. *Mol. Syst. Biol.* **2006**, *2*, 0008.
(23) Ji, Y.; Zhang, B.; Van, S. F.; Horn Warren, P.; Woodnutt, G.; Burnham, M. K.; Rosenberg, M. *Science* **2001**, *293*, 2266.
(24) Forsyth, R. A.; Haselbeck, R. J.; Ohlsen, K. L.; Yamamoto, R. T.; Xu, H.; Trawick, J. D.; Wall, D.; Wang, L.; Brown-Driver, V.; Froelich, J. M.; C, K. G.; King, P.; McCarthy, M.; Malone, C.; Misiner, B.; Robbins, D.; Tan, Z.; Zhu, Zy, Z. Y.; Carr, G.; Mosca, D. A.; Zamudio, C.; Foulkes, J. G.; Zyskind, J. W. *Mol. Microbiol.* **2002**, *43*, 1387.
(25) Perna, N. T.; Plunkett, G., III; Burland, V.; Mau, B.; Glasner, J. D.; Rose, D. J.; Mayhew, G. F.; Evans, P. S.; Gregor, J.; Kirkpatrick, H. A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E. J.; Davis, N. W.; Lim, A.; Dimalanta, E. T.; Potamousis, K. D.; Apodaca, J.; Anantharaman, T. S.; Lin, J.; Yen, G.; Schwartz, D. C.; Welch, R. A.; Blattner, F. R. *Nature* **2001**, *409*, 529.
(26) Welch, R. A.; Burland, V.; Plunkett, G., III; Redford, P.; Roesch, P.; Rasko, D.; Buckles, E. L.; Liou, S. R.; Boutin, A.; Hackett, J.; Stroud, D.; Mayhew, G. F.; Rose, D. J.; Zhou, S.; Schwartz, D. C.; Perna, N. T.; Mobley, H. L.; Donnenberg, M. S.; Blattner, F. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 17020.
(27) Medini, D.; Donati, C.; Tettelin, H.; Masignani, V.; Rappuoli, R. *Curr. Opin. Genet. Dev.* **2005**, *15*, 589.
(28) Kolisnychenko, V.; Plunkett, G., III; Herring, C. D.; Feher, T.; Posfai, J.; Blattner, F. R.; Posfai, G. *Genome Res.* **2002**, *12*, 640.
(29) Suzuki, N.; Okayama, S.; Nonaka, H.; Tsuge, Y.; Inui, M.; Yukawa, H. *Appl. Environ. Microbiol.* **2005**, *71*, 3369.
(30) Westers, H.; Dorenbos, R.; van Dijl, J. M.; Kabel, J.; Flanagan, T.; Devine, K. M.; Jude, F.; Seror, S. J.; Beekman, A. C.; Darmon, E.; Eschevins, C.; de Jong, A.; Bron, S.; Kuipers, O. P.; Albertini, A. M.; Antelmann, H.; Hecker, M.; Zamboni, N.; Sauer, U.; Bruand, C.; Ehrlich, D. S.; Alonso, J. C.; Salas, M.; Quax, W. J. *Mol. Biol. Evol.* **2003**, *20*, 2076.
(31) Posfai, G.; Plunkett, G., III; Feher, T.; Frisch, D.; Keil, G. M.; Umenhoffer, K.; Kolisnychenko, V.; Stahl, B.; Sharma, S. S.; de Arruda, M.; Burland, V.; Harcum, S. W.; Blattner, F. R. *Science* **2006**, *312*, 1044.
(32) Riley, M.; Abe, T.; Arnaud, M. B.; Berlyn, M. K.; Blattner, F. R.; Chaudhuri, R. R.; Glasner, J. D.; Horiuchi, T.; Keseler, I. M.; Kosuge, T.; Mori, H.; Perna, N. T.; Plunkett, G., III; Rudd, K. E.; Serres, M. H.; Thomas, G. H.; Thomson, N. R.; Wishart, D.; Wanner, B. L. *Nucleic Acids Res.* **2006**, *34*, 1.
(33) Koonin, E. V. *Annu. Rev. Genomics Hum. Genet.* **2000**, *1*, 99.
(34) Szathmáry, E. *Nature* **2005**, *433*, 469.
(35) Forster, A. C.; Church, G. M. *Mol. Syst. Biol.* **2006**, *2*, 45.
(36) Luisi, P. L.; Ferri, F.; Stano, P. *Naturwissenschaften* **2006**, *93*, 1.
(37) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409*, 387.
(38) Szathmáry, E.; Santos, M.; Fernando, C. *Top. Curr. Chem.* **2005**, *259*, 167.
(39) Pohorille, A.; Deamer, D. *Trends Biotechnol.* **2002**, *20*, 123.
(40) Klasson, L.; Andersson, S. G. E. *Trends Microbiol.* **2004**, *12*, 37.
(41) Zientz, E.; Dandekar, T.; Gross, R. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 745.
(42) Lerat, E.; Daubin, V.; Moran, N. A. *PLoS Biol.* **2003**, *1*, E19.
(43) Sällström, B.; Andersson, S. G. E. *Curr. Opin. Microbiol.* **2005**, *8*, 579.
(44) Woese, C. R. *Microbiol. Rev.* **1987**, *51*, 221.
(45) Moran, N. A. *Cell* **2002**, *108*, 583.
(46) Moran, N. A.; Mira, A. *Genome Biol.* **2001**, *2*, RESEARCH0054.
(47) Nilsson, A. I.; Koskiniemi, S.; Eriksson, S.; Kugelberg, E.; Hinton, J. C.; Andersson, D. I. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12112.
(48) Teyssier, C.; Marchandin, H.; Simeon De Buochberg, M.; Ramuz, M.; Jumas-Bilak, E. *J. Bacteriol.* **2003**, *185*, 2901.

(49) Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History*; W. W. Norton: New York, 1989.
(50) Mushegian, A. R.; Koonin, E. V. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10268.
(51) Charlebois, R. L.; Doolittle, W. F. *Genome Res.* **2004**, *14*, 2469.
(52) Salama, N. R.; Shepherd, B.; Falkow, S. *J. Bacteriol.* **2004**, *186*, 7926.
(53) Knuth, K.; Niesalla, H.; Hueck, C. J.; Fuchs, T. M. *Mol. Microbiol.* **2004**, *51*, 1729.
(54) Suzuki, N.; Okai, N.; Nonaka, H.; Tsuge, Y.; Inui, M.; Yukawa, H. *Appl. Environ. Microbiol.* **2006**, *72*, 3750.
(55) Gallagher, L. A.; Ramage, E.; Jacobs, M. A.; Kaul, R.; Brittnacher, M.; Manoil, C. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1009.
(56) Giaever, G.; Chu, A. M.; Ni, L.; Connelly, C.; Riles, L.; Veronneau, S.; Dow, S.; Lucau-Danila, A.; Anderson, K.; Andre, B.; Arkin, A. P.; Astromoff, A.; El-Bakkoury, M.; Bangham, R.; Benito, R.; Brachat, S.; Campanaro, S.; Curtiss, M.; Davis, K.; Deutschbauer, A.; Entian, K. D.; Flaherty, P.; Foury, F.; Garfinkel, D. J.; Gerstein, M.; Gotte, D.; Guldener, U.; Hegemann, J. H.; Hempel, S.; Herman, Z.; Jaramillo, D. F.; Kelly, D. E.; Kelly, S. L.; Kotter, P.; LaBonte, D.; Lamb, D. C.; Lan, N.; Liang, H.; Liao, H.; Liu, L.; Luo, C.; Lussier, M.; Mao, R.; Menard, P.; Ooi, S. L.; Revuelta, J. L.; Roberts, C. J.; Rose, M.; Ross-Macdonald, P.; Scherens, B.; Schimmack, G.; Shafer, B.; Shoemaker, D. D.; Sookhai-Mahadeo, S.; Storms, R. K.; Strathern, J. N.; Valle, G.; Voet, M.; Volckaert, G.; Wang, C. Y.; Ward, T. R.; Wilhelmy, J.; Winzeler, E. A.; Yang, Y.; Yen, G.; Youngman, E.; Yu, K.; Bussey, H.; Boeke, J. D.; Snyder, M.; Philippsen, P.; Davis, R. W.; Johnston, M. *Nature* **2002**, *418*, 387.
(57) Kamath, R. S.; Fraser, A. G.; Dong, Y.; Poulin, G.; Durbin, R.; Gotta, M.; Kanapin, A.; Le Bot, N.; Moreno, S.; Sohrmann, M.; Welchman, D. P.; Zipperlen, P.; Ahringer, J. *Nature* **2003**, *421*, 231.
(58) Thatcher, J. W.; Shaw, J. M.; Dickinson, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 253.
(59) Warringer, J.; Ericson, E.; Fernandez, L.; Nerman, O.; Blomberg, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15724.
(60) Deutschbauer, A. M.; Jaramillo, D. F.; Proctor, M.; Kumm, J.; Hillenmeyer, M. E.; Davis, R. W.; Nislow, C.; Giaever, G. *Genetics* **2005**, *169*, 1915.
(61) Papp, B.; Pál, C.; Hurst, L. D. *Nature* **2004**, *429*, 661.
(62) Harrison, R.; Papp, B.; Pal, C.; Oliver, S. G.; Delneri, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2307.
(63) Wagner, A. *Bioessays* **2005**, *27*, 176.
(64) Gu, Z.; Steinmetz, L. M.; Gu, X.; Scharfe, C.; Davis, R. W.; Li, W. H. *Nature* **2003**, *421*, 63.
(65) Fischer, E.; Sauer, U. *Eur. J. Biochem.* **2003**, *270*, 880.
(66) Tong, A. H.; Evangelista, M.; Parsons, A. B.; Xu, H.; Bader, G. D.; Page, N.; Robinson, M.; Raghibizadeh, S.; Hogue, C. W.; Bussey, H.; Andrews, B.; Tyers, M.; Boone, C. *Science* **2001**, *294*, 2364.
(67) Schuldiner, M.; Collins, S. R.; Thompson, N. J.; Denic, V.; Bhamidipati, A.; Punna, T.; Ihmels, J.; Andrews, B.; Boone, C.; Greenblatt, J. F.; Weissman, J. S.; Krogan, N. J. *Cell* **2005**, *123*, 507.
(68) Hashimoto, M.; Ichimura, T.; Mizoguchi, H.; Tanaka, K.; Fujimitsu, K.; Keyamura, K.; Ote, T.; Yamakawa, T.; Yamazaki, Y.; Mori, H.; Katayama, T.; Kato, J. *Mol. Microbiol.* **2005**, *55*, 137.
(69) Yu, B. J.; Sung, B. H.; Koob, M. D.; Lee, C. H.; Lee, J. H.; Lee, W. S.; Kim, M. S.; Kim, S. C. *Nat. Biotechnol.* **2002**, *20*, 1018.
(70) Price, N. D.; Reed, J. L.; Palsson, B. O. *Nat. Rev. Microbiol.* **2004**, *2*, 886.
(71) Mulquiney, P. J.; Kuchel, P. W. *Biochem. J.* **1999**, *342 Pt 3*, 597.
(72) Chen, K. C.; Calzone, L.; Csikasz-Nagy, A.; Cross, F. R.; Novak, B.; Tyson, J. J. *Mol. Biol. Cell.* **2004**, *15*, 3841.
(73) Nelson, D. E.; Ihekwaba, A. E.; Elliott, M.; Johnson, J. R.; Gibney, C. A.; Foreman, B. E.; Nelson, G.; See, V.; Horton, C. A.; Spiller, D. G.; Edwards, S. W.; McDowell, H. P.; Unitt, J. F.; Sullivan, E.; Grimley, R.; Benson, N.; Broomhead, D.; Kell, D. B.; White, M. R. *Science* **2004**, *306*, 704.
(74) Burgard, A. P.; Vaidyaraman, S.; Maranas, C. D. *Biotechnol. Prog.* **2001**, *17*, 791.
(75) Holzhütter, S.; Holzhütter, H. G. *Chembiochem.* **2004**, *5*, 1401.
(76) Covert, M. W.; Famili, I.; Palsson, B. O. *Biotechnol. Bioeng.* **2003**, *84*, 763.
(77) Edwards, J. S.; Covert, M.; Palsson, B. *Environ. Microbiol.* **2002**, *4*, 133.
(78) Castellanos, M.; Wilson, D. B.; Shuler, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6681.
(79) Domach, M. M.; Leung, S. K.; Cahn, R. E.; Cocks, G. G.; Shuler, M. L. *Biotechnol. Bioeng.* **2000**, *67*, 827.
(80) Andersson, S. G.; Kurland, C. G. *Trends Microbiol.* **1998**, *6*, 263.
(81) Maniloff, J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10004.
(82) Lovett, S. T.; Feschenko, V. V. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7120.
(83) Mira, A.; Ochman, H.; Moran, N. A. *Trends Genet.* **2001**, *17*, 589.
(84) Goryshin, I. Y.; Naumann, T. A.; Apodaca, J.; Reznikoff, W. S. *Genome Res.* **2003**, *13*, 644.
(85) Rupp, W. D. In *Escherichia coli and Salmonella thyphimurium, cellular and molecular biology*; Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umberger, H. E., Eds.; American Society of Microbiology: Washington, DC, 1996; Vol. 2.
(86) Datsenko, K. A.; Wanner, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6640.
(87) Zhang, Y.; Buchholz, F.; Muyrers, J. P.; Stewart, A. F. *Nat. Genet.* **1998**, *20*, 123.
(88) Murphy, K. C. *J. Bacteriol.* **1998**, *180*, 2063.
(89) Muyrers, J. P.; Zhang, Y.; Buchholz, F.; Stewart, A. F. *Genes Dev.* **2000**, *14*, 1971.
(90) Reyrat, J. M.; Pelicic, V.; Gicquel, B.; Rappuoli, R. *Infect. Immun.* **1998**, *66*, 4011.
(91) Hoess, R. H.; Ziese, M.; Sternberg, N. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3398.
(92) Sternberg, N.; Hamilton, D. *J. Mol. Biol.* **1981**, *150*, 467.
(93) Beggs, J. D. *Nature* **1978**, *275*, 104.
(94) Broach, J. R.; Hicks, J. B. *Cell* **1980**, *21*, 501.
(95) Miller, J. H. *A Short Course in Bacterial Genetics: A Laboratory Manual for Escherichia coli and Related Bacteria*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1992.
(96) Ruvkun, G. B.; Ausubel, F. M. *Nature* **1981**, *289*, 85.
(97) Hamilton, C. M.; Aldea, M.; Washburn, B. K.; Babitzke, P.; Kushner, S. R. *J. Bacteriol.* **1989**, *171*, 4617.
(98) Leenhouts, K.; Buist, G.; Bolhuis, A.; ten Berge, A.; Kiel, J.; Mierau, I.; Dabrowska, M.; Venema, G.; Kok, J. *Mol. Gen. Genet.* **1996**, *253*, 217.
(99) Miller, V. L.; Mekalanos, J. J. *J. Bacteriol.* **1988**, *170*, 2575.
(100) Gutterson, N. I.; Koshland, D. E., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4894.
(101) Slater, S.; Maurer, R. *J. Bacteriol.* **1993**, *175*, 4260.
(102) Suzuki, N.; Tsuge, Y.; Inui, M.; Yukawa, H. *Appl. Microbiol. Biotechnol.* **2005**, *67*, 225.
(103) Blomfield, I. C.; Vaughn, V.; Rest, R. F.; Eisenstein, B. I. *Mol. Microbiol.* **1991**, *5*, 1447.
(104) Parish, T.; Stoker, N. G. *Microbiol.* **2000**, *146 (Pt 8)*, 1969.
(105) Suzuki, N.; Nonaka, H.; Tsuge, Y.; Okayama, S.; Inui, M.; Yukawa, H. *Appl. Microbiol. Biotechnol.* **2005**, *69*, 151.
(106) Fabret, C.; Ehrlich, S. D.; Noirot, P. *Mol. Microbiol.* **2002**, *46*, 25.
(107) Lederberg, J.; Zinder, N. *J. Am. Chem. Soc.* **1948**, *70*, 4267.
(108) Murphy, K. C.; Campellone, K. G.; Poteete, A. R. *Gene* **2000**, *246*, 321.
(109) Biswas, I.; Gruss, A.; Ehrlich, S. D.; Maguin, E. *J. Bacteriol.* **1993**, *175*, 3628.
(110) Monteilhet, C.; Perrin, A.; Thierry, A.; Colleaux, L.; Dujon, B. *Nucleic Acids Res.* **1990**, *18*, 1407.
(111) Posfai, G.; Kolisnychenko, V.; Bereczki, Z.; Blattner, F. R. *Nucleic Acids Res.* **1999**, *27*, 4409.
(112) Jasin, M.; Schimmel, P. *J. Bacteriol.* **1984**, *159*, 783.
(113) Russell, C. B.; Thaler, D. S.; Dahlquist, F. W. *J. Bacteriol.* **1989**, *171*, 2609.
(114) Sambrook, J.; Fritch, E. F.; Maniatis, T. *Molecular Cloning. A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1987.
(115) Benzinger, R.; Enquist, L. W.; Skalka, A. *J. Virol.* **1975**, *15*, 861.
(116) Yu, D.; Ellis, H. M.; Lee, E. C.; Jenkins, N. A.; Copeland, N. G.; Court, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5978.
(117) Karu, A. E.; Sakaki, Y.; Echols, H.; Linn, S. *J. Biol. Chem.* **1975**, *250*, 7377.
(118) Murphy, K. C. *J. Bacteriol.* **1991**, *173*, 5808.
(119) Carter, D. M.; Radding, C. M. *J. Biol. Chem.* **1971**, *246*, 2502.
(120) Little, J. W. *J. Biol. Chem.* **1967**, *242*, 679.
(121) Kmiec, E.; Holloman, W. K. *J. Biol. Chem.* **1981**, *256*, 12636.
(122) Muniyappa, K.; Radding, C. M. *J. Biol. Chem.* **1986**, *261*, 7472.
(123) Pomerantsev, A. P.; Sitaraman, R.; Galloway, C. R.; Kivovich, V.; Leppla, S. H. *Infect. Immun.* **2006**, *74*, 682.
(124) Stephan, J.; Stemmer, V.; Niederweis, M. *Gene* **2004**, *343*, 181.
(125) Fukiya, S.; Mizoguchi, H.; Mori, H. *FEMS Microbiol. Lett.* **2004**, *234*, 325.
(126) Albert, H.; Dale, E. C.; Lee, E.; Ow, D. W. *Plant J.* **1995**, *7*, 649.
(127) Araki, K.; Araki, M.; Yamamura, K. *Nucleic Acids Res.* **1997**, *25*, 868.
(128) Suzuki, N.; Nonaka, H.; Tsuge, Y.; Inui, M.; Yukawa, H. *Appl. Environ. Microbiol.* **2005**, *71*, 8472.
(129) Kleckner, N.; Roth, J.; Botstein, D. *J. Mol. Biol.* **1977**, *116*, 125.
(130) McHattie, L. A.; Jackowski, J. B. In *DNA, Insertion Elements, Plasmids and Episomes*; Bukhari, A. L., Shapiro, J. A., Adhya, S. L., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1977.

(131) Blattner, F. R.; Plunkett, G., III; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. *Science* **1997**, *277*, 1453.
(132) Tao, H.; Bausch, C.; Richmond, C.; Blattner, F. R.; Conway, T. *J. Bacteriol.* **1999**, *181*, 6425.
(133) Wei, Y.; Lee, J. M.; Richmond, C.; Blattner, F. R.; Rafalski, J. A.; LaRossa, R. A. *J. Bacteriol.* **2001**, *183*, 545.
(134) Hacker, J.; Hentschel, U.; Dobrindt, U. *Science* **2003**, *301*, 790.
(135) Koob, M. D.; Shaw, A. J.; Cameron, D. C. *Ann. N. Y. Acad. Sci.* **1994**, *745*, 1.
(136) Allen, T. E.; Price, N. D.; Joyce, A. R.; Palsson, B. O. *PLoS Comput. Biol.* **2006**, *2*, e2.
(137) Hill, T. M. In *Escherichia coli and Salmonella thyphimurium, cellular and molecular biology*; Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umberger, H. E., Eds.; American Society for Microbiology,: Washington, DC, 1996; Vol. 2.
(138) Rothstein, R.; Michel, B.; Gangloff, S. *Genes Dev.* **2000**, *14*, 1.
(139) Couturier, E.; Rocha, E. P. *Mol. Microbiol.* **2006**, *59*, 1506.
(140) Guy, L.; Roten, C. A. *Gene* **2004**, *340*, 45.
(141) Westers, L.; Westers, H.; Quax, W. J. *Biochim. Biophys. Acta* **2004**, *1694*, 299.
(142) Dorenbos, R.; Stein, T.; Kabel, J.; Bruand, C.; Bolhuis, A.; Bron, S.; Quax, W. J.; Van Dijl, J. M. *J. Biol. Chem.* **2002**, *277*, 16682.
(143) Perego, M. In *Bacillus subtilis and other Grampositive bacteria*; Sonenshein, A. L., Hoch, J. A., Losick, R., Eds.; American Society of Microbiology: Washington, DC, 1993.
(144) Swinfield, T. J.; Oultram, J. D.; Thompson, D. E.; Brehm, J. K.; Minton, N. P. *Gene* **1990**, *87*, 79.
(145) Ikeda, M.; Nakagawa, S. *Appl. Microbiol. Biotechnol.* **2003**, *62*, 99.
(146) Kalinowski, J.; Bathe, B.; Bartels, D.; Bischoff, N.; Bott, M.; Burkovski, A.; Dusch, N.; Eggeling, L.; Eikmanns, B. J.; Gaigalat, L.; Goesmann, A.; Hartmann, M.; Huthmacher, K.; Kramer, R.; Linke, B.; McHardy, A. C.; Meyer, F.; Mockel, B.; Pfefferle, W.; Puhler, A.; Rey, D. A.; Ruckert, C.; Rupp, O.; Sahm, H.; Wendisch, V. F.; Wiegrabe, I.; Tauch, A. *J. Biotechnol.* **2003**, *104*, 5.
(147) Inui, M.; Tsuge, Y.; Suzuki, N.; Vertes, A. A.; Yukawa, H. *Appl. Environ. Microbiol.* **2005**, *71*, 407.
(148) Hacker, J.; Blum-Oehler, G.; Muhldorfer, I.; Tschape, H. *Mol. Microbiol.* **1997**, *23*, 1089.
(149) Lapidus, A.; Galleron, N.; Andersen, J. T.; Jorgensen, P. L.; Ehrlich, S. D.; Sorokin, A. *FEMS Microbiol. Lett.* **2002**, *209*, 23.
(150) Thomson, N. R.; Parkhill, J. *Trends Microbiol.* **2003**, *11*, 66.
(151) Koumoutsi, A.; Chen, X. H.; Henne, A.; Liesegang, H.; Hitzeroth, G.; Franke, P.; Vater, J.; Borriss, R. *J. Bacteriol.* **2004**, *186*, 1084.
(152) Tettelin, H.; Masignani, V.; Cieslewicz, M. J.; Donati, C.; Medini, D.; Ward, N. L.; Angiuoli, S. V.; Crabtree, J.; Jones, A. L.; Durkin, A. S.; Deboy, R. T.; Davidsen, T. M.; Mora, M.; Scarselli, M.; Margarit y Ros, I.; Peterson, J. D.; Hauser, C. R.; Sundaram, J. P.; Nelson, W. C.; Madupu, R.; Brinkac, L. M.; Dodson, R. J.; Rosovitz, M. J.; Sullivan, S. A.; Daugherty, S. C.; Haft, D. H.; Selengut, J.; Gwinn, M. L.; Zhou, L.; Zafar, N.; Khouri, H.; Radune, D.; Dimitrov, G.; Watkins, K.; O'Connor, K. J.; Smith, S.; Utterback, T. R.; White, O.; Rubens, C. E.; Grandi, G.; Madoff, L. C.; Kasper, D. L.; Telford, J. L.; Wessels, M. R.; Rappuoli, R.; Fraser, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13950.
(153) Young, J. P.; Crossman, L. C.; Johnston, A. W.; Thomson, N. R.; Ghazoui, Z. F.; Hull, K. H.; Wexler, M.; Curson, A. R.; Todd, J. D.; Poole, P. S.; Mauchline, T. H.; East, A. K.; Quail, M. A.; Churcher, C.; Arrowsmith, C.; Cherevach, I.; Chillingworth, T.; Clarke, K.; Cronin, A.; Davis, P.; Fraser, A.; Hance, Z.; Hauser, H.; Jagels, K.; Moule, S.; Mungall, K.; Norbertczak, H.; Rabbinowitsch, E.; Sanders, M.; Simmonds, M.; Whitehead, S.; Parkhill, J. *Genome Biol.* **2006**, *7*, R34.
(154) Fong, S. S.; Palsson, B. O. *Nat. Genet.* **2004**, *36*, 1056.
(155) Joyce, A. R.; Reed, J. L.; White, A.; Edwards, R.; Osterman, A.; Baba, T.; Mori, H.; Lesely, S. A.; Palsson, B. O.; Agarwalla, S. *J. Bacteriol.* **2006**.
(156) Peter, B. J.; Arsuaga, J.; Breier, A. M.; Khodursky, A. B.; Brown, P. O.; Cozzarelli, N. R. *Genome Biol.* **2004**, *5*, R87.
(157) Cooper, T. F.; Rozen, D. E.; Lenski, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1072.
(158) Sauer, U. *Adv. Biochem. Eng. Biotechnol.* **2001**, *73*, 129.
(159) Shendure, J.; Mitra, R. D.; Varma, C.; Church, G. M. *Nat. Rev. Genet.* **2004**, *5*, 335.
(160) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L.; Jarvie, T. P.; Jirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. *Nature* **2005**, *437*, 376.
(161) Honisch, C.; Raghunathan, A.; Cantor, C. R.; Palsson, B. O.; van den Boom, D. *Genome Res.* **2004**, *14*, 2495.
(162) Herring, C. D.; Raghunathan, A.; Honisch, C.; Patel, T.; Applebee, M. K.; Joyce, A. R.; Albert, T. J.; Blattner, F. R.; van den Boom, D.; Cantor, C. R.; Palsson, B. O. *Nat. Genet.* **2006**, *38*, 1406.
(163) Fong, S. S.; Nanchen, A.; Palsson, B. O.; Sauer, U. *J. Biol. Chem.* **2006**, *281*, 8024.
(164) Liolios, K.; Tavernarakis, N.; Hugenholtz, P.; Kyrpides, N. C. *Nucleic Acids Res.* **2006**, *34*, D332.
(165) Christie, K. R.; Weng, S.; Balakrishnan, R.; Costanzo, M. C.; Dolinski, K.; Dwight, S. S.; Engel, S. R.; Feierbach, B.; Fisk, D. G.; Hirschman, J. E.; Hong, E. L.; Issel-Tarver, L.; Nash, R.; Sethuraman, A.; Starr, B.; Theesfeld, C. L.; Andrada, R.; Binkley, G.; Dong, Q.; Lane, C.; Schroeder, M.; Botstein, D.; Cherry, J. M. *Nucleic Acids Res.* **2004**, *32*, D311.
(166) Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.; Danchin, A.; et al. *Nature* **1997**, *390*, 249.
(167) Mizoguchi, H.; Mori, H.; Fujio, T. *Biotechnol. Appl. Biochem.* **2007**, *46*, 157.
(168) Ara, K.; Ozaki, K.; Nakamura, K.; Yamane, K.; Sekiguchi, J.; Ogasawara, N. *Biotechnol. Appl. Biochem.* **2007**, *46*, 169.
(169) Giga-Hama, Y.; Tohda, H.; Takegawa, K.; Kumagai, H. *Biotechnol. Appl. Biochem.* **2007**, *46*, 147.CR0683111

CR0683111