Protein Cell 2010, 1(5): 427–434
DOI 10.1007/s13238-010-0064-4

Protein & Cell

MINI-REVIEW

How to make a minimal genome for synthetic minimal cell

Liu-yan Zhang, Su-hua Chang, Jing Wang ✉

Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
✉ Correspondence: wangjing@psych.ac.cn
Received May 6, 2010 Accepted May 12, 2010

ABSTRACT

As a key focus of synthetic biology, building a minimal artificial cell has given rise to many discussions. A synthetic minimal cell will provide an appropriate chassis to integrate functional synthetic parts, devices and systems with functions that cannot generally be found in nature. The design and construction of a functional minimal genome is a key step while building such a cell/chassis since all the cell functions can be traced back to the genome. Kinds of approaches, based on bioinformatics and molecular biology, have been developed and proceeded to derive essential genes and minimal gene sets for the synthetic minimal genome. Experiments about streamlining genomes of model bacteria revealed genome reduction led to unanticipated beneficial properties, such as high electroporation efficiency and accurate propagation of recombinant genes and plasmids that were unstable in other strains. Recent achievements in chemical synthesis technology for large DNA segments together with the rapid development of the whole-genome sequencing, have transferred synthesis of genes to assembly of the whole genomes based on oligonucleotides, and thus created strong preconditions for synthesis of artificial minimal genome. Here in this article, we review briefly the history and current state of research in this field and summarize the main methods for making a minimal genome. We also discuss the impacts of minimized genome on metabolism and regulation of artificial cell.

KEYWORDS synthetic biology, minimal genome, essential gene, minimal cell

INTRODUCTION

The title “synthetic biology” appeared in the literature in 1980 when it was used by Barbara Hobom (Hobom, 1980) to describe bacteria that had been genetically engineered using recombinant DNA technology, which was mostly synonymous with “bioengineering”. In 2000, the term “synthetic biology” was again introduced by Eric Kool and other speakers at the annual meeting of the American Chemical society in San Francisco (Rawls, 2000). Here, the term was used to describe the synthesis of unnatural organic molecules that function in living systems. It is close to the meaning “redesign life” (Benner and Sismour, 2005). Now, as described by Synthetic Biology Community (http://syntheticbiology.org/), synthetic biology is the design and construction of new biological parts, devices and systems, and the re-design of existing, natural biological systems for useful purposes. This is the simplest and perhaps the most widely accepted definition.

Synthetic biology, as a field, has developed over last decade, hanging on the advances in biology, genetics and genome sequencing coupled to the increase in the speed and storage capacity of computers and the internet. One of the key features of synthetic biology is the application of rigorous engineering principle to biological system design and development (Kitney, 2007; The Royal Academy of Engineering, 2009). A significant challenge to engineering in biology is the inherent complexity of the cells in which the modified DNA is embedded in order to produce the desired device and system. It is important that the synthetic device or system is either decoupled from the essential metabolic processes whose disordered can be lethal for the cell, or does not adversely affect these processes (The Royal Academy of Engineering, 2009). One approach to this problem is to simplify the chassis by reducing the genome and hence the complexity of the chassis. In accordance with this sense, minimizing the

genome of natural bacterial is one of the key branches of synthetic biology research (Diez et al., 2009).

Living organism contain, in their genome, genes coding for functions that are not essential for survival under the controlled and stable conditions. Determining and removing these unessential genes would help to decrease undesired interplay between device or system and chassis, and thus increase the predictive power. The practice for estimating the size and gene content of minimal genome have been implemented through comparative (Mushegian and Koonin, 1996) and experimental (Hutchison et al., 1999) approaches since the end of last century. Genome reduction projects (Yu et al., 2002; Fehér et al., 2007) in several model organisms were also reported in last decade. In particular, recent achievements in chemical synthesis technology for large DNA segments together with the rapid development of the whole-genome sequencing, have transferred synthesis of genes to assembly of the whole genomes based on oligonucleotides (Gibson et al., 2008a, b), and thus created strong preconditions for synthesis of artificial minimal genome based on derived minimal gene sets. In this review, we introduce the concept of minimal genome and review the strategies for determining minimal gene sets and essential genes. We also discuss the impacts of reduced genome on living cell and the practical significance of an artificial minimal cell.

THE MINIMAL GENOME CONCEPT

One of the basic purposes of the "minimal genome" research is to find out "how many genes can make a cell" (Koonin, 2000). It is closely related to the concept of "minimal gene set", which seeks to estimate the smallest number of genetic elements sufficient to build a modern-type free-living cellular organism. However, the phrase "minimal gene set" in itself makes no sense, except when associated with a defined set of conditions (Koonin, 2003), such as species, environmental conditions (e.g., culture medium, temperature, habitat) and the purpose of the research (e.g., main for theoretical development or for bioengineering application). It is not hard to understand that, functional completeness may be not so important for theoretical development, but it is really significant for bioengineering applications. Out of survival and developmental considerations, cells always keep some redundant genome elements for special cases, such as resource starvation and poor environment. 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 medium. So the minimal gene set should be defined under ideal conditions (Koonin, 2000), that is, in the presence of unlimited amounts of all essential nutrients and in the absence of any adverse factors, including competition. Whereas, Foley and Shuler (2010) think that defining such a minimal gene set required for life is necessary, but not sufficient for synthesizing a minimal cell, especially for a biotechnological platform cell. Functional gene fractions should be considered when deriving a minimal gene set.

There are currently two different ways to get a minimal genome, bottom-up approach and top-down approach (Szathmáry, 2005). The bottom-up approach constructs an artificial genome by using chemical synthesis method. After the repertoire for minimal gene set was figured out, the minimal genome could be synthesized based on this blueprint. In contrast, the top-down approach starts from existing organisms with the aim of simplifying their genome to a reduced genome. Although genome minimization seems easier than genome synthesis from scratch, when coming to the details, they have nearly the same challenges: how to decide which genes to delete and which ones to keep?

The naturally evolved minimal genomes may give us a reference (Fraser et al., 1995; Wann, 2000; Gil et al., 2002) . After entering 21th century, more and more complete genome sequences are becoming available for a large number of diverse bacterial species. According to statistics of GOLD (Genome OnLine Database V3.0, http://genomesonline.org/index2.htm) (Liolios et al., 2008), till March 18th, 2010, there are total 6937 genome projects, and 1224 complete genomes have been published including 1020 bacterial, 124 eukaryal and 80 archaeal genomes. Although most bacterial possess more than 2000 genes, the smallest bacterial genomes contain less than 600 genes (Table 1). Among these small genomes, *Mycoplasma genitalium* G-37 has the minimal genome sized 580 kb coding 564 open reading frames (ORFs) as a free-living bacterial. But mutagenesis experiment indicated that only 256–350 protein coding genes of *M. genitalium* are essential under laboratory growth condition (Hutchison et al., 1999). More theoretical and practical work tend to give a range of minimal gene set about 200–300 genes (Itaya, 1995; Mushegian and Koonin, 1996). Mushegian and Koonin used the comparative genomic approach to find ubiquitous genes existing in the first two completely sequenced genomes. These ubiquitous genes supplemented by some non-orthologous displacement genes composed a minimal gene set for cellular life (Mushegian and Koonin, 1996). Many groups have tried to derive essential genes through single gene inactivation experiments. Genes that were put into minimal gene set can be called essential genes. A number of studies have addressed this issue on *in silico* or *in vitro* level.

STRATEGIES ON DERIVING MINIMAL GENOME

In 1995, Mitsuhiro Itaya (Itaya, 1995) made the first try to estimate the minimal genome size. He investigated 79 randomly selected chromosomal loci of *Bacillus subtilis* by mutagenesis, only found 6 mutations rendered *B. subtilis* unable to form colonies. The statistical analyses for the

Table 1 Genome sizes of different bacteria

organism name	size (kb)	ORFs	Gram staining	biotic relationships	classification
Buchnera aphidicola Cc	420	431	Gram−	symbiotic	Proteobacteria-γ
Mycoplasma genitalium G-37	580	564	Gram+	free living	Tenericutes
Mycoplasma pneumoniae M129	816	736	Gram+	free living	Tenericutes
Mesoplasma florum L1	793	760	Gram+	free living	Firmicutes
Haemophilus influenzae Rd (KW20)	1830	1774	Gram−	free living	Proteobacteria-γ
Staphylococcus aureus aureus N315 (MRSA)	2813	2746	Gram+	free living	Firmicutes
Mycobacterium leprae Br4923	3268	2767	Gram+	free living	Actinobacteria
Mycobacterium tuberculosis H37Rv (lab strain)	4411	4060	Gram+	free living	Actinobacteria
Bacillus subtilis subtilis 168	4214	4408	Gram+	free living	Firmicutes
Escherichia coli K-12, MG1655	4639	4612	Gram−	free living	Proteobacteria-γ
Frankia sp CcI3	5433	4670	Gram+	symbiotic	Actinobacteria
Pseudomonas aeruginosa PAO1	6264	5742	Gram−	free living	Proteobacteria-γ
Mycobacterium smegmatis MC2 155	6988	6978	Gram+	free living	Actinobacteria
Burkholderia cenocepacia J2315 (ch-1)	8060	7023	Gram−	free living	Proteobacteria-β

frequency of indispensable loci (6 out of 79) shown that total indispensable genetic material would be included within about 318–562 kb, which is in accord with those currently determined smallest genomes for bacteria. However, he did not give a reasonable repertoire of genes in such a genome ever though a hypothetical minimal genome size required for life was proposed. Next year, Mushegian and Koonin (Mushegian and Koonin, 1996) derived such a repertoire for minimal self-sufficient gene set using comparative genomics approach based on the first two completely sequenced small bacterial genomes, *Mycoplasma genitalium* and *Haemophilus influenza*, which belong to two ancient bacterial lineages, i.e., Gram-positive and Gram-negative bacteria, respectively. This minimal gene set contains 234 orthologs between *M. genitalium* and *H. influenza* plus 22 non-orthologous displacements, total 256 genes. Later on, large-scale gene inactivation experiments were finished by different groups. Their work indicates that the present of essential genes in the whole genome was no more than 25%, except for the naturally small genomes. Advances in systems biology approach provide another way to define essential genes *in silico*. Scientists derive *in silico* essential genes based on the genome-scale metabolic network using modified Flux Balance Analysis (FBA) algorithm or other mathematical methods (Covert and Palsson, 2003; Price et al., 2004; Zhang et al., 2009). Systems approach gives a chance to understand and obtain biological knowledge on a system or cellular level. Each of these approaches did its bit in deriving minimal gene set and essential genes, but they have their own pros and cons, respectively.

Comparative genomics approach

Comparative genomics is based on the hypothesis that the genes, which are conserved in distant related species, are almost certainly essential for cellular function and are likely to approximate the minimal gene set (Mushegian and Koonin, 1996). But the same essential function may be performed by different even non-orthologous genes, so it is not enough for this approach to only search for genes that have homolog in the genomes of groups of organisms, the non-orthologous gene displacements also need to be included in to ensure the completeness of essential cellular functions. Using this approach, Mushegian and Koonin suggested that the 256 genes are close to the minimal gene set that is necessary and sufficient to sustain the existence of a modern-type cell. Since this minimal gene set was determined under ideal conditions, which means rich nutrition and no pressure, this minimal gene set also defined a hypothetical simple cell with following principle features: a nearly complete system of translation, a virtually complete DNA replication machinery, a rudimentary system of recombination and repair, a simple transcription apparatus with four RNA polymerase subunits, a single σ factor and three transcription factors, a large set of chaperone-like proteins, no amino acid biosynthesis, a limited repertoire of metabolite transport systems, etc (Mushegian and Koonin, 1996). No matter whether their minimal gene set is really functional, they give a rough figure of minimal genome and the corresponding minimal cell. Based on comparative results, and taking experimental gene inactivation data into consideration, Gil determined the core of a

minimal bacterial gene set with 206 genes, and reconstructed a minimal metabolic machinery necessary to sustain life (Gil et al., 2004). This proposed minimal genome had similar features with Mushegian and Koonin's.

With more and more whole genome sequence are available, comparative genomic approach was used to analyze extended species. It is foreseeable that the number of orthologous genes drop-off with the increase of the number of chromosomes from different organisms. Analysis of about 100 genomes shows that only 63 genes are ubiquitous (Koonin, 2003). That is because numerous essential cellular functions can be performed by unrelated proteins that show no sequence similarities to each other in different organisms, resulting in the non-ubiquitous presence of the corresponding genes across species (Fehér et al., 2007). 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 (Charlebois and Doolittle, 2004). COG database (http://www.ncbi.nlm.nih.gov/COG/) (Tatusov et al., 2003) shows that nearly 40% ubiquitous COGs (clusters of orthologous groups of proteins) are poorly characterized. It is thus expected that this approach is totally powerless for non-orthologous gene displacement, and a minimal genome based on this approach alone would be inviable and it would not be possible to identify the missing essential genes. But the comparative approach still could offer invaluable insights on core gene sets conserved across closely related organisms, for example by identifying strain-specific, horizontally transferred genomic islands (Medini et al., 2005; Posfai et al., 2006).

Experimental gene inactivation approach

To fill the gap of comparative genomics approach and to give a direct performance of gene essentiality, experimental gene inactivation approach identifies those genes whose individual inactivation causes inviability. Several genome-scale identifications of such genes have been performed in prokaryotic and eukaryotic organisms by different groups with different strategies. Generally, there are three main experimental approaches to identify essential genes under particular growth conditions: massive transposon mutagenesis strategies (Judson and Mekalanos, 2000; Akerley et al., 2002; Salama et al., 2004; French et al., 2008), the use of antisense RNA (Ji et al., 2001; Forsyth et al., 2002) to inhibit gene expression, and the systematic inactivation of each individual gene present in a genome (Herring et al., 2003; Kobayashi et al., 2003). However, these results showed that percentage of essential genes in the whole genome is really low (Table 2). In addition, results from the same method may vary a lot. Many reasons and limitations make this result. First, genes whose deletions sharply slow down the growth may not be included in the essential gene set. Second, genes that are individually dispensable may not be simultaneously dispensable, because of the presence of alternative cellular pathways, functionally redundant gene copies, and genes that can be compensated by other genes in the genome. Finally,

Table 2 Results of large-scale gene inactivation experiments for prokaryotes

organism	ORFs total	No_E	E (%)	reference
Mycoplasma genitalium G-37	518	381	73.55%	Glass et al., 2006
Mycoplasma pulmonis UAB	815	310	38.04%	French et al., 2008
Helicobacter pylori	1616	344	21.29%	Salama et al., 2004
Haemophilus influenzae KW20 Rd	1738	670	38.55%	Akerley et al., 2002
Francisella tularensis	1767	396	22.41%	Gallagher et al., 2006
Streptococcus pneumoniae TIGR4	2175	113	5.20%	Thanassi et al., 2002
	2175	133	6.11%	Song et al., 2005
Staphylococcus aureus N315	2662	168	6.31%	Ji et al., 2001
	2662	658	24.72%	Forsyth et al., 2002
Acinetobacter sp. ADP1	3307	499	15.09%	de Berardinis et al., 2008
Mycobacterium tuberculosis H37Rv	4038	614	15.21%	Sassetti et al., 2003
Escherichia coli K-12	4321	620	14.35%	Gerdes et al., 2003
	4321	303	7.01%	Baba et al., 2006
Bacillus subtilis 168	4354	271	6.22%	Kobayashi et al., 2003
Salmonella typhimurium LT2	4541	257	5.66%	Knuth et al., 2004
Pseudomonas aeruginosa PAO1	5664	678	11.97%	Jacobs et al., 2003
Pseudomonas aeruginosa PA14	5964	335	5.62%	Liberati et al., 2006

Note: ORF, open reading frame; No_E, the number of essential genes; E (%), the percentage of essential genes among the total number of ORFs in the genome.

laboratory conditions often fail to detect genes that are required under special environmental conditions. Moreover, transposon mutagenesis might miss genes that can tolerate transposon insertion.

Fang et al. analyzed gene essentiality in bacteria *in silico* through four main parameters: persistence index, orthologs leading-strand index, sequence divergence, and codon adaptation index (Fang et al., 2005). They found that persistent nonessential (PNE) genes show an over-representation of genes related to maintenance and stress response. This outlines the limits of current experimental techniques to define gene essentiality and highlights the essential role of genes implicated in maintenance that, although dispensable for growth, are not dispensable from an evolutionary point of view. They also showed that some persistent genes have many characters in common with experimentally identified essential genes. They suggested to consider them as truly essential genes. So when deriving a minimal genome, the results from experimental gene inactivation also should be treated carefully.

Systems biology approach: genome-scale network model

A complete understanding of the relationship between genotype and phenotype would greatly facilitate the design of minimal genome. 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 (Fehér et al., 2007). Although such a comprehensive mathematical representation of a whole cell is out of reach at present, models of various cellular subsystems, especially genome-scale metabolic network models are becoming increasingly available (Reed et al., 2003; Duarte et al., 2004; Becker and Palsson, 2005; Puchalka et al., 2008; Christian et al., 2009; Zhang et al., 2009).

Genomic information coupled with biochemical and physiological knowledge has enabled the reconstruction of genome-scale biochemical reaction network for microorganism (Price et al., 2004). The constraint-based approach was introduced to deduce the metabolic phenotype from the genotype (Covert and Palsson, 2003). From genome-scale metabolic network to minimal gene set *in silico*, common practice is kind of like experimental gene inactivation approach. That is the network reaction(s) associated with each gene was individually "deleted" by setting the flux to 0 and optimizing for the biomass function (Price et al., 2004). This methods have yielded numerous key theoretical insights on the nature and evolution of minimal genome, but there is a general caveat for application of constrained-based methods: some of the computationally derived minimal metabolic networks might not be kinetically feasible (Holzhütter and Holzhütter, 2004).

Recently, Ying Zhang et al. introduced another algorithm to the minimal gene set calculation of metabolic network (Zhang et al., 2009). Their analysis also were carried out using the optimization and single gene deletion function of the COBRA toolbox, but their algorithm picks one gene each time to do "deletion", then uses judgment and iteration procedure to put genes into three categories: core-essential, synthetic lethal and non-essential genes. This *in silico* method may have its own limits, but it gives reliability to essentiality of each gene.

CHEMICAL SYNTHESIS OF DNA

Once the content of a genome has been determined, the next step may be to "re-write", or synthesis all or part of the genome. Technology for chemical synthesis of DNA is becoming mature. Synthesis of small DNA segments has been not a difficult task since the 21th century. Furthermore, there are a number of cases where the genome of an organism has been entirely synthesized. In 2002, Cello and coworkers at State University of New York, Stony Brook, synthesized the poliovirus genome (7741 bp) from its published sequence, producing the first synthetic organism (Cello et al., 2002). In 2003, the genome of the bacteriophage ΦX-174 (5386 bp) was assembled in just two weeks by a team at the J. Craig Venter Institute (Smith et al., 2003). In 2008, Hamilton Smith and co-workers again pushed forward the boundaries of synthesis with their reconstruction of an entire 489 kb synthetic genome of the bacterium, *Mycoplasma genitalium* (Gibson et al., 2008a). The synthesis capacity has shown a steady increase in the last decade. There has been a concurrent fall in cost to less than $0.55 per base pair, depending on sequence length and composition (The Royal Academy of Engineering, 2009). The technological barrier to progress in chemical synthesis of minimal genome will be mitigated with the development of new technology (The Royal Academy of Engineering, 2009).

At present, a related technological challenge lies in the ability to successfully transfer large DNA segments, even whole genomes, into populations of cells (Foley and Shuler, 2010). In prokaryotic cells, circular plasmids of tens of kilo base pairs may be routinely transfected into cells. Whole chromosomes present a much greater challenge. And the other major challenge in genome transplantation is to understand the role of DNA methylation and histone modifications. These modifications directly affect gene activity and such modifications may play key roles in activating transplanted genomes (The Royal Academy of Engineering, 2009).

GENOME REDUCTION RROJECTS

An top-down approach to get a minimal genome is genome reduction, which cuts off non-essential DNA segment from the whole genome but still keeps the cell alive. This strategy starts from existing organisms with the aim of simplifying their

genome till a minimal, or at least reduced genome. *Escherichia coli*, as the favorite model Gram-positive bacterium, its reduced genome or minimal genome have been constructed by the diverse genomic deletion methods, such as suicide plasmid-mediated genomic deletion, linear DNA-mediated genomic deletion, site-specific recombination-mediated genomic deletion, and transposon-mediated random deletion. These methods were summarized in Fehér et al's review (Fehér et al., 2007) and the book named "Systems Biology and Biotechnology of *Escherichia coli*" by S.C. Kim et al. (Sung et al., 2009). The reduced-genome bacterial strains have genomes that are 5%–30% smaller than that of a wild-type *E. coli* strain. The genomes of other microorganisms, such as *B. subtilis* (Ara et al., 2007), *C. glutamicum* (Suzuki et al., 2005), and the yeasts (Giga-Hama et al., 2007; Murakami et al., 2007) also have been reduced for the construction of minimal-genome factories.

Consistent with the hypothesis that most of bacterial genomes contain redundant or non-essential segments, most new strains with reduced-genomes performed better in some aspect, and showed potential for industrial application.

CONCLUSION AND PERSPECTIVES

The development of biotechnology and bioengineering, coupled with the accumulating of knowledge about cellular life, have given us a chance to realize our dream of creating ideal and robust host organisms for novel usages that benefit humankind. Genome, as the most important basis for a living cell, is favored by synthetic biology researchers. Creating an artificial minimal genome, on one hand, is a basis for synthetic biology, on the other hand, provides a new way to understand living cell. Researchers are tackling this task by using a comprehensive approach based on computational, experimental and literature-based studies. Based on these approaches, many essential genes have been identified, and organisms with streamlined genomes also have been constructed. However, any result from single approach has not so much persuasion, so information from different strategies should be selected and integrated logically and appropriately to get a minimal genome.

It is apparently that after a suitable and robust minimal genome is constructed, it would provide synthetic minimal cell a good chassis to assemble kinds of functional modules. The applications of such a minimal cell will have a broad applications ranging from industrial chemicals to pharmaceutical proteins.

ACKNOWLEDGEMENTS

This work was sponsored by the National Natural Science Foundation of China (NSFC, Grant No. 30700441); and the Beijing New Star Project, Beijing Municipal Science & Technology Commission Foundation (2007A082).

REFERENCES

Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N., and Mekalanos, J.J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proc Natl Acad Sci U S A 99, 966–971.

Ara, K., Ozaki, K., Nakamura, K., Yamane, K., Sekiguchi, J., and Ogasawara, N. (2007). Bacillus minimum genome factory: effective utilization of microbial genome information. Biotechnol Appl Biochem 46, 169–178.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2, 2006.0008.

Becker, S.A., and Palsson, B.O. (2005). Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. BMC Microbiol 5, 12.

Benner, S.A., and Sismour, A.M. (2005). Synthetic biology. Nat Rev Genet 438, 533–543.

Cello, J., Paul, A.V., and Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: generation of infections virus in the absence of natural template. Science 297, 1016–1018.

Charlebois, R.L., and Doolittle, W.F. (2004). Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res 14, 2469–2477.

Christian, N., May, P., Kempa, S., Handorf, T., and Ebenhoh, O. (2009). An integrative approach towards completing genome-scale metabolic networks. Mol BioSyst 5, 1889–1903.

Covert, M.W., and Palsson, B.O. (2003). Constraints-based models: Regulation of gene expression reduces the steady-state solution space. J Theor Biol 221, 309–325.

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., *et al*. (2008). A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. Mol Syst Biol 4, 174.

Diez, M.S., Lam, C.M., Leprince, A., and dos Santos, V.A. (2009). (Re-)construction, characterization and modeling of systems for synthetic biology. Biotechnol J 4, 1382–1391.

Duarte, N.C., Herrgard, M.J., and Palsson, B.O. (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14, 1298–1309.

Fang, G., Rocha, E., and Danchin, A. (2005). How essential are nonessential genes? Mol Biol Evol 22, 2147–2156.

Fehér, T., Papp, B., Pál, C., and Pósfai, G. (2007). Systematic genome reductions: theoretical and experimental approaches. Chem Rev 107, 3498–3513.

Foley, P.L., and Shuler, M.L. (2010). Considerations for the design and construction of a synthetic platform cell for biotechnological applications. Biotechnol Bioeng 105, 26–36.

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., *et al*. (2002). A genome-wide strategy for the identification of essential genes in Staphylococcus aureus. Mol Microbiol 43, 1387–1400.

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., *et al*. (1995). The minimal gene complement of *Mycoplasma*

genitalium. Science 270, 397–403.

French, C.T., Lao, P., Loraine, A.E., Matthews, B.T., Yu, H., and Dybvig, K. (2008). Large-scale transposon mutagenesis of Mycoplasma pulmonis. Mol Microbiol 69, 67–76.

Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaul, R., Brittnacher, M., and Manoil, C. (2006). A comprehensive transposon mutant library of *Francisella novicida*, a bioweapon surrogate. Proc Natl Acad Sci U S A 104, 1009–1014.

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., *et al*. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J Bacteriol 185, 5673–5684.

Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., *et al*. (2008a). Complete chemical synthesis, assembly, and cloning of a mycoplasma genitalium genome. Science 319, 1215–1220.

Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., and lll, C.A. H. (2008b). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci U S A 105, 20404–20409.

Giga-Hama, Y., Tohda, H., Takegawa, K., and Kumagai, H. (2007). *Schizosaccharomyces pombe* minimum genome factory. Biotechnol Appl Biochem 46, 147–155.

Gil, R., Sabater-Munoz, B., Latorre, A., Silva, F.J., and Moya, A. (2002). Extreme genome reduction in *Buchnera spp*.: Toward the minimal genome needed for symbiotic life. Proc Natl Acad Sci U S A 99, 4454–4458.

Gil, R., Silva, F.J., Pereto, J., and Moya, A. (2004). Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68, 518–537.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., III, C.A.H., Smith, H.O., and Venter, J.C. (2006). Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A 103, 425–430.

Herring, C.D., Glasner, J.D., and Blattner, F.R. (2003). Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. Gene 311, 153–163.

Hobom, B. (1980). Surgery of genes — at the doorstep of synthetic biology. Med Klin 75, 14–21.

Holzhütter, S., and Holzhütter, H.-G. (2004). Computational design of reduced metabolic networks. ChemBioChem 5, 1401–1422.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., and Venter, J.C. (1999). Global transposon mutagenesis and a minimal mycoplasma genome. Science 286, 2165–2169.

Itaya, M. (1995). An estimation of minimal genome size required for life. FEBS Letters 362, 257–260.

Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., *et al*. (2003). Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A 100, 14339–14344.

Ji, Y., Zhang, B., Van, S.F., Horn, Warren, P., Woodnutt, G., Burnham, M.K.R., and Rosenberg, M. (2001). Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science 293, 2266–2269.

Judson, N., and Mekalanos, J.J. (2000). TnAraOut, A transposon-based approach to identify and characterize essential bacterial genes. Nat Biotechnol 18, 740–745.

Kitney, R.I. (2007). Synthetic biology — engineering biologically-based devices and systems. In 11th Mediterranean Conference on Medical and Biological Engineering and Computing 2007, Vols 1 and 2, T. Jarm, P. Kramar, and A. Zupanic, eds. (Berlin, Springer-Verlag Berlin), pp. 1138–1139.

Knuth, K., Niesalla, H., Hueck, C.J., and Fuchs, T.M. (2004). Large-scale identification of essential Salmonella genes by trapping lethal insertions. Mol Microbiol 51, 1729–1744.

Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., *et al*. (2003). Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A 100, 4678–4683.

Koonin, E.V. (2000). How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet 01, 99–116.

Koonin, E.V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1, 127–136.

Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T., and Ausubel, F.M. (2006). An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A 103, 2833–2838.

Liolios, K., Mavromatis, K., Tavernarakis, N., and Kyrpides, N.C. (2008). The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. Nucl Acids Res 36, D475–479.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. Curr Opin Genet Dev 15, 589–594.

Murakami, K., Tao, E., Ito, Y., Sugiyama, M., Kaneko, Y., Harashima, S., Sumiya, T., Nakamura, A., and Nishizawa, M. (2007). Large scale deletions in the Saccharomyces cerevisiae genome create strains with altered regulation of carbon metabolism. Appl Microbiol Biotechnol 75, 589–597.

Mushegian, A.R., and Koonin, E.V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A 93, 10268–10273.

Posfai, G., III, G.P., Feher, T., Frisch, D., Keit, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., Arruda, M.d., *et al*. (2006). Emergent properties of reduced-genome *Escherichia coli*. Science 312, 1044–1046.

Price, N.D., Reed, J.L., and Palsson, B.O. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Micro 2, 886–897.

Puchalka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., Papin, J.A., and Martins dos Santos, V.A. (2008). Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol 4, e1000210.

Rawls, R.L. (2000). 'Synthetic biology' makes its debut. Chem Eng News 78, 49–53.

Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol 4, 12.

Salama, N.R., Shepherd, B., and Falkow, S. (2004). Global transposon mutagenesis and essential gene analysis of

Helicobacter pylori. J Bacteriol 186, 7926–7935.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48, 77–84.

Smith, H.O., Hutchison, C.A., Pfannkoch, C., and Venter, J.C. (2003). Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A 100, 15440–15445.

Song, J.-H., Ko, K.S., Lee, J.-Y., Baek, J.Y., Oh, W.S., Yoon, H.S., Jeong, J.-Y., and Chun, J. (2005). Identification of essential genes in Streptococcus pneumoniae by allelic replacement mutagenesis. Mol Cells 19, 365–374.

Sung, B.H., Lee, J.H., and Kim, S.C. (2009). *Escherichia coli* genome engineering and minimization for the construction of a bioengine. In Systems Biology and Biotechnology of *Escherichia coli*, S.Y. Lee, ed. (Daejeon, Springer), pp. 19–40.

Suzuki, N., Nonaka, H., Tsuge, Y., Okayama, S., Inui, M., and Yukawa, H. (2005). Multiple large segment deletion method for Corynebacterium glutamicum. Appl Microbiol Biotechnol 69, 151–161.

Szathmáry, E. (2005). Life: In search of the simplest cell. Nature 433, 469–470.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., *et al*. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 14.

Thanassi, J.A., Hartman-Neumann, S.L., Dougherty, T.J., Dougherty, B.A., and Pucci, M.J. (2002). Identification of 113 conserved essential genes using a high-throughput gene disruption system in Streptococcus pneumoniae. Nucleic Acids Res 30, 3152–3162.

The Royal Academy of Engineering (2009). Synthetic Biology: scope, applications and implications (London, The Royal Academy of Engineering).

Wann, E.R. (2000). Unlocking the secrets of another minimal genome. Trends Microbiol 8, 544–545.

Yu, B.J., Sung, B.H., Koob, M.D., Lee, C.H., Lee, J.H., Lee, W.S., Kim, M.S., and Kim, S.C. (2002). Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat Biotech 20, 1018–1023.

Zhang, Y., Thiele, I., Weekes, D., Li, Z., Jaroszewski, L., Ginalski, K., Deacon, A.M., Wooley, J., Lesley, S.A., Wilson, I.A., *et al*. (2009). Three-dimensional structural view of the central metabolic network of Thermotoga maritima. Science 325, 1544–1549.

