REVIEW ARTICLE

Toward minimal bacterial cells: evolution vs. design

Andrés Moya[1,2,3], Rosario Gil[1,2,3], Amparo Latorre[1,2,3], Juli Peretó[1,2,4], Maria Pilar Garcillán-Barcia[5] & Fernando de la Cruz[5]

[1]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, València, Spain; [2]CIBER de Epidemiología y Salud Pública, Madrid, Spain; [3]Departament de Genètica, Universitat de València, València, Spain; [4]Departament de Bioquímica i Biologia Molecular, Universitat de València, València, Spain; and [5]Departamento de Biología Molecular, Universidad de Cantabria, Santander, Spain

Abstract

Recent technical and conceptual advances in the biological sciences opened the possibility of the construction of newly designed cells. In this paper we review the state of the art of cell engineering in the context of genome research, paying particular attention to what we can learn on naturally reduced genomes from either symbiotic or free living bacteria. Different minimal hypothetically viable cells can be defined on the basis of several computational and experimental approaches. Projects aiming at simplifying living cells converge with efforts to make synthetic genomes for minimal cells. The panorama of this particular view of synthetic biology lead us to consider the use of defined minimal cells to be applied in biomedical, bioremediation, or bioenergy application by taking advantage of existing naturally minimized cells.

Introduction

Synthetic biology is neither a new science nor a clearly defined research program yet. Although recent years have witnessed great enthusiasm around the field, it is simply not true that 'synthetic biology' is a newly coined term. We must go to the beginning of the past century to find that the French biophysicist Stéphane Leduc used this term in 1912 as the title of his book on what he thought was the synthesis of artificial life. Meanwhile, the German-American biochemist Jacques Loeb in 1906 defined the synthesis of life as the goal of biology (Pereto & Catala, 2007). Moreover, synthetic biology can even be included in an older tradition of thinking that was introduced by Goethe when he pointed out on the intrinsic value of a better understanding of life from a holistic perspective rather than when approached by the more successful analytic one (Goethe, 1817–1824). Biology was dominated in the last century by the powerful analytic approach. The advent of genomics, with its avalanche of data, led us to consider again the possibility of a synthetic biology. The recent history of the discipline showed us that the notion of synthesis is not unique, and at least three different categories with different aims, meth-odologies and techniques can be distinguished: protocell creation, DNA-based device construction, and genome-driven cell engineering (see table 1 by O'Malley et al., 2008). The aims of the 'protocell creation' approach are to construct viable approximations to cells and to understand the fundamental biological principles in general and the origin of life in particular. The aim of the 'device construction' approach is to apply the engineering principles to biology and to construct standardized biological devices. The 'cell engineering' approach pursues the synthesis of minimal but complete genomes and their insertion in cells to redesign and control metabolic processes.

In the present review, we focus on the genome-driven cell engineering view of synthetic biology and emphasize the value of learning about naturally reduced genomes as a tool to engineer synthetically reduced ones, based on the hypothesis that simplified cells should be easier to study and predictably manipulate.

Natural minimal cells

In a classical attempt to define a minimal cell, Morowitz (1992) proposed two kinds of ecologically dependent

simplicity: (1) cells harboring a minimized metabolic network able to thrive heterotrophically in a chemically complex environment and (2) autotrophic cells that make the least demands on the environment (i.e. CO_2, N_2, and minerals) to self-construct and maintain all their biological components. Instances of cultivable bacteria of those two classes are found among *Mycoplasmatales* and *Cyanobacteria*, respectively. In terms of bacterial genome minimization in nature, all known cases to date are associated with very specific lifestyles linked to very stable environments: obligate symbiosis (either parasitic or mutualistic) and adaptation to narrow and unique ecological niches in free-living organisms. In both cases we find examples of heterotrophic or autotrophic metabolic modes establishing a sort of hierarchy of minimal life forms that expands Morowitz efforts (Table 1). The reasons for the reductive evolution of these genomes are absolutely dependent on the need for particular genetic adaptations to their specific ecological niches. The reductive genome syndrome (Moya *et al*., 2008) implies additional genomic changes such as a severe decrease in G+C content, the loss of mechanisms for DNA repair and recombination (but see the case of *Nanoarchaeum equitans*), and an increased evolutionary rate of the protein-coding genes. However, the changes undergone in cell size and shape do not follow a general rule. Some bacteria can be much bigger than their closest free-living relatives, as it is the case of mutualistic endosymbionts, or much smaller as it occurs in highly abundant marine bacteria or in parasites (Table 1).

Symbiosis as a genome reductive force

The smallest genome sizes have been detected in prokaryotic cells living in symbiosis with other organisms. Notable examples are the human parasite *Mycoplasma genitalium* (Fraser *et al*., 1995), the archaeal exosymbiont *N. equitans* (Waters *et al*., 2003), and the insect endosymbionts *Buchnera aphidicola* BCc (Perez-Brocal *et al*., 2006), *Candidatus* Carsonella ruddii and *Candidatus* Sulcia muelleri (McCutcheon & Moran, 2007). All these microorganisms have a heterotrophic lifestyle, i.e. they are dependent on the chemically complex environment represented by their respective host cells. To a lesser extent, minimization has also occurred in autotrophic deep-sea clam endosymbionts *Candidatus* Ruthia magnifica (Newton *et al*., 2007) and *Candidatus* Vesicomyosocius okutanii (Kuwahara *et al*., 2007).

Phylogenetic analyses (Baumann, 2005; Brochier *et al*., 2005; Kuwahara *et al*., 2007) have shown that host-dependent bacteria and archaea derived from free-living ancestors. As an adaptation to the symbiotic lifestyle, their genomes underwent a reductive process in which genes that were unnecessary in the new protected environment, or redundant because their functions were provided by the host, tended to be lost. However, each of these nearly minimal genomes is substantially different from the others due to the maintenance of different sets of genes depending on the host's needs.

Mycoplasma genitalium, with a of 580-kb genome encoding 477 proteins (Table 1), has the smallest genome among

Table 1. The smallest published natural genomes

Metabolic mode	Lifestyle	Organism	Environment	Genome size (kb)	ORFs	G+C content (%)	Cell size, diameter (mm)*
A. Heterotrophy	A.1. Symbionts	*Mycoplasma genitalium*[†]	Human cell	580	477	31.7	0.3
		Nanoarchaeum equitans[‡]	*Ignicoccus* sp.	490	536	31.6	0.4
		Buchnera aphidicola BCc[§¶]	*Cinara cedri*	420	362	20.1	3.0
		Carsonella ruddii[§‖]	*Pachypsylla venusta*	159	182	16.0	3.0
		Sulcia muelleri[**††]	*Homalodisca coagulate*	245	227	22.4	30.0
	A.2. Free living	*Pelagibacter ubique*[‡‡]	Sea water	1308	1354	29.7	0.3
B. Autotrophy	B.1. Symbionts	*Ruthia magnifica*[§]	*Calyptogena magnifica*	1200	976	34.0	1.0
		Vesicomyosocius okutanii[§]	*Calyptogena okutanii*	1000	937	31.6	1.0
	B.2. Free living	*Prochlorococcus marinus*[§§]	Sea water	1657	1717	30.8	0.6

All genomic dates were retrieved from the NCBI Entrez Genome page (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html)
*All bacteria are spherical in shape except *Sulcia muelleri*. In this case the size refers to length.
[†]Mollicutes.
[‡]Although the new phylum *Nanoarchaeota* was proposed (Huber *et al*., 2002), a reassessment of its phylogenetic position indicates that maybe represents a fast-evolving euryarchaeal lineage related to *Thermococcales* (Brochier *et al*., 2005).
[§]*Gammaproteobacteria*.
[¶]Coexists with *Serratia symbiotica*.
[‖]Although it has not been described, its metabolic impairment suggests the existence of another symbiont (Tamames *et al*., 2007).
[**]*Bacteroidetes*.
[††]Coexists with *Baumannia cicadellinicola*.
[‡‡]*Alphaproteobacteria*.
[§§]*Cyanobacteria*, ecotype MED4.

organisms that can be grown as pure cultures (Fraser *et al.*, 1995). The drastic economization in genetic information is associated with its parasitic way of life. Thus, the *M. genitalium* genome carries all essential genes for informational processes (replication, transcription, translation, and protein folding and processing) plus a severely limited metabolism, including very few genes for the synthesis of vitamins, nucleic acid precursors, energy supply, and those devoted to parasitism.

The hyperthermophilic archaeon *N. equitans* is the only known archaeon exhibiting a symbiotic lifestyle. It must attach to the surface of the submarine hot-vent crenarchaeon *Ignicoccus* for survival. With a genome of 490 kb (Waters *et al.*, 2003) and a high coding density, it contains 536 protein-coding genes. As described for bacteria with reduced genomes, this archaeon presents a complete information-processing system but a highly simplified metabolic apparatus. However, as a consequence of the dual adaptation of *N. equitans* to high temperature and to an obligate symbiotic lifestyle, it has also retained genes involved in DNA repair and recombination (Das *et al.*, 2006).

Three bacterial endosymbionts of insects possess the smallest microbial genomes sequenced to date. Two *Gammaproteobacteria*, *B. aphidicola* BCc (Perez-Brocal *et al.*, 2006) and '*C.* Carsonella ruddii' (Nakabachi *et al.*, 2006), are primary endosymbionts of phloem sap-feeding insects, aphids and psyllids, respectively. The third symbiont is the bacteroidetes '*C.* Sulcia muelleri', primary endosymbiont of sharpshooters, xylem sap-feeding insects. Their role in their respective symbiosis is to supply the nutrients that lack in the insect host diet, mainly essential amino acids and vitamins. All these bacteria depend on their hosts for survival. In fact, a characteristic of obligatory endosymbiotic bacteria is that, so far, they have not been cultured outside the insect hosts.

Four different genomes of the aphid endosymbiont *B. aphidicola* have been sequenced to date, providing an interesting model for the analysis of the genome-reductive evolutionary process. The cedar aphid symbiont, *B. aphidicola* BCc, represents a very advanced stage in the reductive process. With only 422 kb and just 362 protein-coding genes, its genome is about 200 kb smaller than the other three *B. aphidicola* sequenced strains (Perez-Brocal *et al.*, 2006). This dramatic genome reduction is due to the loss of most genes involved in the biosynthesis of nucleotides, cofactors, transporters, peptidoglycan, as well as all ATPase subunits. Nevertheless, it still retains complete machinery for informational processes, and a simplified metabolic network for energy production, an indication that its minimal gene set is still able to support cellular life. Remarkably, it is engaged on a symbiotic consortium with *Candidatus* Serratia symbiotica, a second symbiont massively present and with a long-term association in the cedar aphid

(Gomez-Valero *et al.*, 2004; Lamelas *et al.*, 2008), for the provision of some essential nutrients (Gosalbes *et al.*, 2008). Complementarity in metabolic capabilities also occurs in the bacteroidetes '*C.* Sulcia muelleri' (its genome being 250 kb in length and containing 227 protein-coding genes) (McCutcheon & Moran, 2007), which is a coresident, with the gammaproteobacterium *Candidatus* Baumannia cicadellinicola of the sharpshooter *Homalodisca vitripennis*. Thus, '*C.* Sulcia muelleri' provides essential amino acids whereas '*C.* Baumannia cicadellinicola' primarily contributes with vitamins and cofactors to the consortium (Wu *et al.*, 2006; McCutcheon & Moran, 2007).

The hackberry petiole gall psyllid *Pachypsylla venusta* symbiont, '*C.* Carsonella ruddii', with a genome size of 160 kb, has been proposed as the bacterial endosymbiont with the smallest known genome. This highly compact genome contains overlapping adjacent ORFs, and only 182 protein-coding genes have been predicted (Nakabachi *et al.*, 2006), a much lower number than previous proposals for minimal genomes, and almost half of the genes identified in *B. aphidicola* BCc. More than half of the '*C.* Carsonella ruddii' genes are devoted to translation and amino acid metabolism, but just a few tRNA genes could be identified, and some essential amino acid biosynthetic steps and pathways are missing. Furthermore, there is a total absence of genes for most bacterium-specific processes (Tamames *et al.*, 2007). It remains to be elucidated if many of the lost genes have been transferred from the genome of a *Carsonella* ancestor to the genome of a psyllid ancestor, now being expressed under the control of the host nucleus (Nakabachi *et al.*, 2006). This hypothesis would be in accordance with the endosymbiont theory for the origin of organelles. Maybe this strain of '*C.* Carsonella ruddii' is on its way to become a new subcellular entity, closer to an organelle than to a minimal living cell (Tamames *et al.*, 2007).

In contrast with the above mentioned heterotrophic endosymbionts, the two *Gammaproteobacteria* '*C.* Ruthia magnifica' and '*C.* Vesicomyosocius okutanii' (Kuwahara *et al.*, 2007), primary endosymbionts of deep-sea clams of the genus *Calyptogena*, present a chemolithoautotrophic metabolism providing almost all nutrients to the host. These bacteria fix CO_2 using electrons from H_2S to drive an autotrophic metabolism based on a slightly modified version of the reductive pentose phosphate pathway (Calvin–Benson cycle). They have the potential for all essential biomolecule anabolism, except threonine (and isoleucine) and ubiquinone, and recycle nitrogen. The few missing steps in those pathways could be explained by the function of a yet unidentified enzyme, or by complementation with the host metabolism. Thus, these deep-sea clams with only a vestigial gut and greatly reduced filtering ability depend on their endosymbionts for most nutrients, including amino acids, vitamins, fatty acids, and nucleotides. On the bacterial side,

their position on specialized bacteriocytes in the outermost layer of the gill epithelial cells allows the bacteria to use sulfide as electron donor and oxygen (or nitrate) as electron acceptor, from the clam blood and environmental water, respectively, thus avoiding the spontaneous reaction of both components (Lane, 2007). These two autotrophic bacteria, endosymbionts of eukaryotic cells, could represent an intermediate state of metabolic minimization between the extreme reduction shown by heterotrophic endosymbionts and the minimization undergone by free-living marine bacteria.

Resources economization as a genome-reductive force

Two marine free-living bacteria, *Prochlorococcus marinus* (a cyanobacterium) and, *Pelagibacter ubique* (an alphaproteobacterium) have also streamlined genomes (< 2 Mb) but still live as independent cells. Both organisms are among the most abundant and successful organisms on Earth and also present a small cell volume (i.e. a high surface : volume ratio). Although, as above stated, the genome reduction syndrome shares many features with that of host-dependent prokaryotes, it appears to be a consequence of a selective process favouring bacterial adaptation to an environment with low concentrations of nutrients, allowing these organisms to make a substantial reduction of their cell machinery and still thrive in remarkable oligotrophic conditions, most likely as a consequence of the increase in their surface : volume ratio (Dufresne *et al.*, 2005).

In the marine microbial community, *Cyanobacteria* are among the most relevant organisms and, in particular, the genus *Prochlorococcus* (Partensky *et al.*, 1999). The smallest genome sizes are found in the strains of *P. marinus* (1.7–2.7 Mb), which also is the smallest known oxygen-evolving photoautotrophic organism in the ocean (Chisholm *et al.*, 1988). The high light-adapted ecotype MED4 has a genome of 1.66 Mb that codes for 1717 proteins (Rocap *et al.*, 2003). The comparative analysis with other *Prochlorococcus* genomes revealed the constant changes that take place in response to the different environmental challenges (Dufresne *et al.*, 2003; Rocap *et al.*, 2003; Hess, 2004). Yet, all of their genomes code for a complete metabolic network based on carbon fixation by the Calvin–Benson cycle, and a photoelectronic transport chain that allows the biosynthesis of all cellular constituents from CO_2, mineral salts, and visible light. DNA repair systems, chaperones, nitrogen metabolism, transport systems, and the motility machinery are simplified or absent in comparison to freshwater cyanobacteria. The reduction of about 30% in genome size, compared with other close relatives such as *Synechococcus*, is compatible with the assumption that a massive gene loss has occurred during the evolution from a *Prochlorococcus* ancestor (Dufresne *et al.*, 2005).

The smallest genome of any cell known to replicate independently in nature corresponds to the photoheterotrophic marine bacterium *P. ubique*, which belongs to the abundant clade SAR11 (Morris *et al.*, 2002). Its 1.308 Mb genome codes just for 1354 ORF, the smallest number for any free-living cell sequenced so far (Giovannoni *et al.*, 2005). This bacterium has the potential to biosynthesize all protein amino acids and most cofactors. Genome minimization is reflected by the nearly complete absence of nonfunctional or redundant DNA, with very short intergenic regions, and the lack of pseudogenes and phage genes. Its pattern of genome reduction is consistent with the hypothesis of genome streamlining driven by selection acting on a very large population occupying a very low nutrient and stable habitat. In this sense, *P. ubique* appears to employ an adaptive strategy that resembles the highly successful marine unicellular cyanobacteria in its simple metabolism and small genome size.

Defining minimal cells: experimental and computational approaches

The natural phenomenon of genome downsizing observed in the above mentioned specialized bacteria poses the question about how many genes are necessary to support cellular life, the first step to approach the challenge of making a minimal cell. A living cell is, at least, the sum of many essential functions that need to be performed for cellular survival and replication. Therefore, no mater how small, bacterial reduced genomes must still retain all essential genes involved in such housekeeping functions, as well as a limited amount of metabolic transactions needed to survive in a nutrient-rich and stable environment.

One way to approach the gene composition of a minimal genome is by comparing processes of natural minimization. The first attempt to define a minimal gene set for life was made by comparative genomics, soon after the first two bacterial genomes, from *Haemophilus influenzae* and *M. genitalium* (Mushegian & Koonin, 1996), were sequenced. The underlying hypothesis was that genes shared between distantly related species are likely to be essential. In addition, both bacteria are human parasites with reduced genomes. *Mycoplasma genitalium* genome, as stated above, still represents the smallest known genome of a bacterium that can be grown in the laboratory, and therefore must be close to a minimal autonomous genome. The comparative analysis led to the proposal of a minimal gene-set composed only by 256 genes, mostly involved in genetic information storage and processing, protein chaperoning and a limited metabolic capability. Although many additional comparative analyses have been performed since then (reviewed in Carbone, 2006), there is an intrinsic limitation of this method that needs to be empathized: as it was already

addressed since the very beginning of comparative genome analyses (Mushegian & Koonin, 1996), many essential cellular functions can be performed by several alternative and unrelated (nonorthologous) proteins, and will be mislead by comparative approaches. In fact, the comparative minimal core (common set of genes) only retrieves those genes involved in functions for which there is no alternative in nature (e.g. the complex translational machinery, including the ribosome).

A genetic approach can also be used for the experimental determination of essential genes. Large-scale inactivation studies have been performed to try to define which genes are essential for cell survival in several well-characterized bacterial models, such as *Escherichia coli*, *Bacillus subtilis*, *Staphylococcus aureus*, and *M. genitalium*, using several strategies such as massive transposon mutagenesis, the use of antisense RNA to inhibit gene expression, or the systematic inactivation of each individual gene present in a genome (reviewed in Gil *et al.*, 2004; Féher *et al.*, 2007; Reznikoff & Winterberg, 2008). But, still, all these methods have limitations, because they uncover several possibilities of misclassification of genes. Transposon mutagenesis might overestimate the minimal set by considering as essential some genes that slow down the growth without arresting it, but can also miss essential genes that tolerate transposon insertions. The use of antisense RNA is limited to genes for which an adequate expression of the inhibitory RNA can be obtained. Inactivation of single genes does not detect essential functions encoded by redundant genes, and some individually dispensable genes may not be simultaneously dispensable (a phenomenon called synthetic lethality).

A combined study of all published research lines using computational or experimental methods, including the comparison of reduced genomes from insect endosymbionts, has also been used to define the core of a minimal genome for a free-living bacterium thriving in a chemically rich environment (Gil *et al.*, 2004). For the first time, in this study the authors analyzed the functional completeness of the minimal metabolism coded by the proposed gene repertoire (involving 62 protein-coding genes out of a minimal set of 208 genes). This aspect was further explored, demonstrating the stoichiometric consistency and some architectural properties of the derived minimal metabolic network (Gabaldon *et al.*, 2007). However, it must be emphasized that metabolic complexity is ecologically dependent, i.e. it is a function of the chemical compounds and the primary energy source(s) available to the living system in a given environment (Morowitz, 1992). Therefore, the latter studies are just presenting one possible form of a minimal genome able to carry out one possible minimal metabolism.

Different metabolic networks derived from an experimental genome reductive evolution *in silico* have also been examined, in order to analyze the participation of contingency-dependent gene losses in the evolution of minimal genomes (Pal *et al.*, 2006). Using a computational representation of the metabolic network of *E. coli*, Pál and colleagues repeatedly simulated the successive loss of genes in a controlled environment. The different resulting minimal metabolic networks partially vary in both gene content and number, but a core metabolism, which is over-represented in bacterial insect endosymbionts, was always preserved. Moreover, by simulating their respective environmental conditions, the authors modeled the evolution of the gene content in *B. aphidicola* and *Wigglesworthia glossinidia* (primary endosymbiont of the tsetse fly) with over 80% accuracy. These results indicate that, at least in the examined cases, differences among ecologically dependent minimal networks were predictable, based on knowledge about its distant ancestors and its current lifestyle.

The simplification of modern cells has also been approached in a third way, by searching for the biochemical description of well-defined pathways that are needed to perform essential functions. Using this biochemical approach, Forster & Church (2006) described which are the main components needed to synthesize a minimal self-replicative system. Their proposed minimal genome contains just 151 genes, 38 RNA genes plus 113 protein-coding genes needed for informational processes, but does not include genes for intermediate metabolism (e.g. lipid metabolism and glycolysis). Self-reproductive vesicles encapsulating the biological system defined by their minimal genome, and allowing transmembranous small molecule transport, would be in principle spontaneously generated by the addition of amphiphilic molecules to the environment (Szostak *et al.*, 2001; Luisi, 2007; Mansy *et al.*, 2008), although much work has to be carried out on the biophysical properties of artificial vesicles with encapsulated materials.

This biochemically derived minimal gene-set is quite different from what has been proposed in the previously described approaches, but has the advantage of being completely composed by genes with well-defined functions, being designed by modules, and it allows a system-by-system debugging to attain self-replication. In any case, this seems a good start to approach the goal of synthesizing a useful, near-minimal, self-replicating system dependent on added energized substrates, that is, the closest approach ever made to an artificial minimal cell. Although some challenges still need to be faced before the goal can be accomplished (i.e. appropriate genes for tRNA modification, replication of monomer DNA circles over byproducts, ribosome assembly under physiological conditions, efficiency of translation, control and integration of subsystems, and cosegregation of genes with their products), the *in vitro* plan to synthesize it is on its way (Forster & Church, 2007), although on the basis of still unrealistic assumptions under the current technologies, like considering membranes as optional.

Synthetic genomes for minimal cells

One of the main goals of synthetic biology is the synthesis of a living cell, which will have a big impact both in biotechnology and in our basic understanding of modern living systems (Pereto & Catala, 2007). This goal can be approached in two complementary and alternative ways: top down or bottom up. The top-down approach, also called genome-driving cell engineering (O'Malley *et al.*, 2008) or minimal cell project (Luisi, 2007), starts from simple modern organisms and pretends to simplify them further through the removal of nonessential genetic elements. On the other side, the bottom-up approach attempts the synthesis of artificial minimal cells, component by component (Deamer, 2005; Forster & Church, 2006). While this second approach implies that, once the defined composition of a minimal genome has been envisaged, the next step would be to synthesize the genome that contains the required genetic information, the simplification of extant genomes can be performed both, by synthesizing whole near-minimal genomes, or by deletion of the nonessential parts of larger ones.

Simplifying a modern-cell genome

Experimental genome reduction strategies have been extensively used in the past few years with several model and biotechnologically useful microorganisms, in order to learn about genome architecture and improve their characteristics (reviewed by Féher *et al.*, 2007). There are different ways to force a cellular population to go through a genome-reductive process (e.g. selection for faster replication, damaging some DNA repair systems, or forcing the cells to grow in nutrient-poor conditions and pass through population bottlenecks), but all of them are slow and highly unpredictable. Furthermore, there are no adequate protocols to select for smaller genomes and, due to the randomness of the process, it is quite difficult to identify the genomic location of the losses. It is, therefore, more convenient to use targeted approaches toward genome reduction. The minimization process can be accelerated using different experimental strategies, which include plasmid or linear DNA-mediated procedures, as well as the use of site-specific recombinases and transposons. Processes of minimization in the laboratory have been pursued with *E. coli* (Kolisnychenko *et al.*, 2002; Yu *et al.*, 2002; Fukiya *et al.*, 2004; Hashimoto *et al.*, 2005; Posfai *et al.*, 2006; Mizoguchi *et al.*, 2007), *B. subtilis* (Westers *et al.*, 2003; Ara *et al.*, 2007; Morimoto *et al.*, 2008), *Corynebacterium glutamicum* (Suzuki *et al.*, 2005a, b, c, d), and *M. genitalium* (Glass *et al.*, 2006). Some of these targeted approaches, aiming at the careful deletion of selected presumably dispensable genome regions, produced reduced-genome cells displaying virtually unaffected physiological features, and some emergent beneficial properties (Westers *et al.*, 2003; Posfai *et al.*, 2006; Morimoto *et al.*,

2008). Surprisingly, many of the removed genes were considered essential in a massive transposon mutagenesis experimental study (Gerdes *et al.*, 2003). However, when more extensive deletions were generated to maximize the extent of genome reduction, the indiscriminative removal of large DNA fragments resulted in aberrant cell morphology and reduced growth rate (Hashimoto *et al.*, 2005). In fact, it can be considered that an artificially reduced genome accepts further reductions with growing pains, making difficult the task of determining if there is a minimal deletion core. Three issues are critical to understand why. First, redundancy might be only apparent. Second, each part of the genome is loosely connected with the rest, so that each part functions worse when contiguous parts are eliminated, and deleterious effects are normally incremental. Third, the mechanism of cell division is not yet completely understood. It seems that a bacterium uses the bulk of the chromosome to take some measurements as to where and when to initiate cell division (Yu & Margolin, 1999). When increasing portions of the genome are deleted, the measuring system starts to malfunction (G. Pósfai, pers. commun.).

Thus, when we investigate the minimal deletion core, we face what we can call 'the three-leg chair paradox': if we have a four-leg chair and ask how many legs are needed for the chair to stand, the experimental answer will be three, but the chair will be always unbalanced (only apparent redundancy). However, the common set of legs (minimal deletion core) is zero, because no particular leg is essential if we eliminate them one by one. Proper function requires proper disposition of the four legs, and if it has only three then the chair has to readapt to the three legs' new disposition. Similarly, a simplified genome might require rebalancing due to the loose network connections that involve the whole genome. And we can go further: if we think on equilibrists or bicycles, they can also work on two-leg and even one-leg chairs (using wheels in movement we can limit the number of legs to just one). Similarly, under specific carefully selected conditions, genomes can be greatly simplified. We are convinced that a minimal cell is feasible in the long run, following an iterative procedure with deletion/adjustment/ deletion cycles. In this way, compensatory mutations may be necessary to rebalance the genome, allowing for the genome to spontaneously adjust to each substantial deletion before attempting a further reduction in size.

Synthesizing a modern-cell genome

At the turn of the millennium, the increasing genomics technologies generated a demand for *de novo* synthesis of long DNA sequences containing complex gene compositions. But current technologies were still too slow, expensive, and inaccurate to allow synthesizing whole small genomes, as a gateway to fabricate artificial cells. Multiple breakthrough

Table 2. Landmarks in the race toward the synthesis of artificial minimal genomes

References	Title	Remarks
Sekiya *et al.* (1979)	Total synthesis of a tyrosine suppressor tRNA gene. XVI. Enzymatic joinings to form the total 207-base pair-long DNA	Khorana's research group presents the last in a series of 16 papers, published along 5 years, describing the chemical synthesis of a 207-bp tRNA gene
Nambiar *et al.* (1984)	Total synthesis and cloning of a gene coding for the ribonuclease S protein	The first protein-coding gene is synthesized in the Sidney Brenner's group. Two people, working full-time during a year and a half, were needed to synthesize this 330-bp DNA fragment (S.A. Brenner, pers. commun.)
Stemmer *et al.* (1995)	Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides	The first report of the synthesis of extended stretches of DNA, using a PCR-based method for the synthesis of a 2.7-kb plasmid from a pool of short, overlapping synthetic oligonucleotides
Cello *et al.* (2002)	Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template	The first chemically synthesized genome. A functional 7740-bp poliovirus genome, with all expected biochemical and pathogenic properties was obtained
Yount *et al.* (2003)	Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus	Viral genome assembled from cDNAs
Smith *et al.* (2003)	Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides	Venter's group synthesizes a 5385-bp bacteriophage genome in 2 weeks by serialized oligonucleotides assembly and amplification
Kodumal *et al.* (2004)	Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster	A couple of months were sufficient to synthesize a genome fragment of 32 kb, the biggest piece of man-made DNA at that time, incorporating several genes needed for the synthesis of a pharmaceutical compound
Tian *et al.* (2004)	Accurate multiplex gene synthesis from programmable DNA microchips	Presents an oligonucleotide synthesis method miniaturized on photo-programmable microfluidic chips, allowing both reduced error frequencies and lower costs. The procedure allowed the synthesis of all 21 genes that encode proteins of the *E. coli* 30S ribosomal subunit and to improve translation efficiency *in vitro* through alteration of codon bias.
Gibson *et al.* (2008)	Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome	The complete synthesis of the 582 970-bp genome for the pathogenic bacterium *M. genitalium* is reported. Venter's group used assembled chemically synthesized oligonucleotides, joined by *in vitro* recombination to produce intermediate assemblies, which were all cloned as bacterial artificial chromosomes in *E. coli*, and transferred into yeast, to assemble the full-length genome

advances in DNA synthesis made this a reachable goal by late 2004, opening the door to what has been considered a new era for biotechnology (Andrianantoandro *et al.*, 2006; Heinemann & Panke, 2006; Serrano, 2007). Table 2 shows a summary of the major achievements in DNA synthesis that led us to fast, inexpensive, and accurate fabrication of long stretches of DNA.

Methods for *de novo* synthesis of long DNA molecules were primarily based on the use of PCR technology to assemble pools of overlapping short-length oligonucleotides. Although these techniques allowed for the complete reconstruction of long functional sequences (from phages to complete genetic pathways) (Cello *et al.*, 2002; Kodumal *et al.*, 2004), they present some relevant inconveniences: oligonucleotide synthesis was expensive, and mutations could arise during their synthesis as well as in the PCR steps, leading to a too high error rate for its use on a genomic scale. Several correction procedures were designed to reduce the error rate, based on the use of photo-programmable micro-

fluidic chips for oligonucleotide assembly (which also exploit the advantage of miniaturization to reduce material costs) (Tian *et al.*, 2004), by adding natural DNA mismatch repair proteins to detect DNA mismatches and discriminate between correct and incorrect sequences (Carr *et al.*, 2004) or using a recursive method to generate error-free long DNA molecules form imperfect oligonucleotides (Linshiz *et al.*, 2008). Now that limiting factors such as DNA length and synthesis fidelity are in their way to be solved, constructing large and complex circuits by commercial *de novo* DNA synthesis will soon become cheaper than the use of standard gene manipulation and cloning technologies. This is why these new technologies are rapidly being adopted by DNA synthesis companies, which are developing new technological platforms that combine advanced informatics, sophisticated automation, and a battery of biochemical protocols to allow the design and massive scale production of biological devices, from full-length genes, operons, and pathways to, ultimately, genomes (Serrano, 2007).

At the beginning of 2008, Venter and colleagues announced the synthesis of the first complete genome of an organism, the small genome of *M. genitalium* (Gibson *et al.*, 2008), a milestone that has been considered 'the dawn of synthetic genomics' (Galperin, 2008). The authors used synthetic oligonucleotides to assemble 5–7-kb pieces that were joined by *in vitro* recombination to produce intermediate fragments, which were then cloned in *E. coli* as bacterial artificial chromosomes (BACs), and transferred into yeast to finally get the full-length genome assembled. The resulting genome is almost identical to the one from the original reference strain, except for one gene, which was disrupted by an antibiotic marker to block pathogenicity and to allow for selection. This monumental and challenging work has proven that DNA synthesis of a small genome is achievable with present-day technology.

An additional challenge would be to put a synthetic genome to work inside a real cell, and Venter's group is also trying to achieve this goal. As a step toward propagation of synthetic genomes, Lartigue *et al.* (2007) proved that it is possible to replace the genome of one organism by another. Intact genomic DNA from *Mycoplasma mycoides* LC, was transplanted into *Mycoplasma capricolum* cells, and the genome-transplanted cells were phenotypically identical to the donor strain. In this case, genome transfer was facilitated because mycoplasmas lack bacterial cell wall, and the two bacteria were closely related. Yet, it is not clear if it would work in other organisms and/or conditions (Pennisi, 2007). An additional challenge is still awaiting, because it is not known if naked synthetic DNA would work when introduced into a cellular compartment without a genome.

Concluding remarks: synthetic engineering of minimal cells

Synthetic biology is a re-emerging multifaceted research program. Its protocell, DNA-based device construction, and genome-driven cell engineering views have gained increased interest because, probably more than any other biological discipline, synthetic biology recreates nature, allowing the design and fabrication of an artificial minimal cell. But does this task require the complete understanding of cell functions? According to Haldane (1940), producing an artificial cell under laboratory conditions would occur even before we fully understand the processes going on inside cells. On the other hand, in direct confrontation with a very naïve reductionist approach made by engineers, some others think that it is necessary to gain first a better understanding on how the cell systems functionally integrate (Drubin *et al.*, 2007; Serrano, 2007; Isalan *et al.*, 2008). At any rate, biological organisms are highly complex, and we currently lack enough knowledge about how living systems work. In fact, even if we consider only the genetic informa-

tion represented by protein-coding genes, about a third of genes from any given natural genome have an unknown function, making it difficult to predict any organism capabilities. If we want to understand the functions that are performed by a synthetic cell, we need to get as close as possible to a minimal genome in which any given essential role is performed by the products of a well-characterized set of genes, which will make it easier to study and predictably engineer, if desired.

Once a minimal genetic complement is established for a system, a whole chromosome with this minimal genome could be obtained by DNA synthesis and introduced into a cellular compartment, or generated by a small number of sequential replacement steps, to fabricate minimal cells. These artificial cells would be ideal containers for inserting new genetic modules or modifying existing ones, molding new genomes to customize microorganisms for different purposes (Andrianantoandro *et al.*, 2006). Even more, with the ability to generate a synthetic specific genome, a modified genetic code could be defined so that no alternative codons were used, thus leaving additional codons for the incorporation of non-natural amino acids in cellular proteins. Provided that an organism with such a designed genome would function, the incorporation of a variety of non-natural amino acids will allow the fabrication of novel proteins with novel functions (Drubin *et al.*, 2007). Even though minimal cells can be less robust than natural ones (Gabaldon *et al.*, 2007), a detailed knowledge on their cellular functions would be a good starting point to add new components that increase redundancy and allow ulterior engineering manipulations.

We also would like to emphasize the usefulness of plasmids in synthetic biology. Plasmids are tools for genomic engineering, naturally exploited by bacteria to promote genetic exchange. We can use plasmids to tinker on genomes and even to try to produce minimal but functional genomes. In fact, BACs are nothing but F-derived plasmids, so that we can design BACs or BAC-like plasmids containing an *oriT* sequence and transfer them by conjugation to a cell devoid of its chromosome (a so-called maxicell) (Sancar *et al.*, 1979; Belogurov *et al.*, 1992; Heinemann & Ankenbauer, 1993). Complete genomes, genome segments, or functional modules can be synthesized on a BAC platform and tested for function. Conjugation is a very promiscuous mechanism and it works perfectly with minimal cells. For instance, *Prochlorococcus* has been found to be a good recipient that can accept and stably inherit plasmid RSF1010 from *Proteobacteria* (Tolonen *et al.*, 2006). In addition, we have found that all genes essential on the recipient cell side for conjugation to take place are included on the minimal genome, as judged from results on plasmid conjugation using the 3906 mutants of the Keio collection (Baba *et al.*, 2006) as recipients (D. Pérez-Mendoza, M.V. Mendiola & F. de la Cruz, submitted).

Synthetic biology has one of its main focuses on technological applications. The research program on minimal cells should be considered, although not exclusively, with the same applied orientation. We would like to point out that better than the use of a generalist minimal cell, the success of the applied program relies on the use of particular minimal cells as vessels for specific applications. That is, we can consider the use of different types of minimal cells when pretending any biomedical, bioremediation, or bioenergy applications, and we can take advantage of existing cells minimized by nature. Let us consider a few examples. Engineered mycoplasma with further-reduced genomes might be appropriate living vectors for the introduction of new genetic circuits designed to fight against human diseases; selected soil bacteria may be appropriate vessels where to implement and/or to modify genetic circuits for bioremediation, while particular cyanobacteria could be useful for H_2 production or CO_2 sequestration.

Finally, we would like to end with a reference to ethical and safety issues. There is no maturity in a field if it is not accompanied by ethical, environmental, and safety considerations (Bhutkar, 2005; Church, 2005). Even though, due to their lack of robustness, artificially designed cells are unlikely to outcompete natural organisms, it is imperative to think on the necessary safeguards to keep synthetic cells from spreading in the environment from the very beginning of any synthetic biology project. At any rate, we agree with Parens *et al.* (2008) that the issues raised by synthetic biology should be integrated within the general ethical enquiries of bioethics.

Acknowledgements

References

Andrianantoandro E, Basu S, Karig DK & Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* **2**. DOI: 10.1038/msb4100073.

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

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL & Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**. DOI: 10.1038/msb4100050.

Baumann P (2005) Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu Rev Microbiol* **59**: 155–189.

Belogurov AA, Delver EP & Rodzevich OV (1992) IncN plasmid pKM101 and IncI1 plasmid ColIb-P9 encode homologous antirestriction proteins in their leading regions. *J Bacteriol* **174**: 5079–5085.

Bhutkar A (2005) Synthetic biology: navigating the challenges ahead. *J Biolaw Bus* **8**: 19–29.

Brochier C, Gribaldo S, Zivanovic Y, Confalonieri F & Forterre P (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* **6**: R42.

Carbone A (2006) Computational prediction of genomic functional cores specific to different microbes. *J Mol Evol* **63**: 733–746.

Carr PA, Park JS, Lee Y-J, Yu T, Zhang S & Jacobson JM (2004) Protein-mediated error correction for de novo DNA synthesis. *Nucleic Acids Res* **32**: e162.

Cello J, Paul AV & Wimmer E (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**: 1016–1018.

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB & Welschmeyer NA (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340–343.

Church G (2005) Let us go forth and safely multiply. *Nature* **438**: 423.

Das S, Paul S, Bag SK & Dutta C (2006) Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* **7**: 186.

Deamer D (2005) A giant step towards artificial life? *Trends Biotechnol* **23**: 336–338.

Drubin DA, Way JC & Silver PA (2007) Designing biological systems. *Genes Dev* **21**: 242–254.

Dufresne A, Salanoubat M, Partensky F *et al.* (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *P Natl Acad Sci USA* **100**: 10020–10025.

Dufresne A, Garczarek L & Partensky F (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* **6**: R14.

Féher T, Papp B, Pal C & Posfai G (2007) Systematic genome reductions: theoretical and experimental approaches. *Chem Rev* **107**: 3498–3513.

Forster AC & Church GM (2006) Towards synthesis of a minimal cell. *Mol Syst Biol* **2**: 45.

Forster AC & Church GM (2007) Synthetic biology projects *in vitro*. *Genome Res* **17**: 1–6.

Fraser CM, Gocayne JD, White O *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397–403.

Fukiya S, Mizoguchi H & Mori H (2004) An improved method for deleting large regions *of Escherichia coli* K-12 chromosome using a combination of Cre/loxP and lambda red. *FEMS Microbiol Lett* **234**: 325–331.

Gabaldon T, Pereto J, Montero F, Gil R, Latorre A & Moya A (2007) Structural analyses of a hypothetical minimal metabolism. *Philos Trans Roy Soc B* **362**: 1751–1762.

Galperin MY (2008) The dawn of synthetic genomics. *Environ Microbiol* **10**: 821–825.

Gerdes SY, Scholle MD, Campbell JW *et al.* (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185**: 5673–5684.

Gibson DG, Benders GA, Andrews-Pfannkoch C *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215–1220.

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

Giovannoni SJ, Tripp HJ, Givan S *et al.* (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA III, Smith HO & Venter JC (2006) Essential genes of a minimal bacterium. *P Natl Acad Sci USA* **103**: 425–430.

Goethe (1817-1824) *Zur Naturwissenschaft überhaupt, besonders zur Morphologie*. Cotta, Stuttgart.

Gomez-Valero L, Soriano-Navarro M, Perez-Brocal V, Heddi A, Moya A, Garcia-Verdugo JM & Latorre A (2004) Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *J Bacteriol* **186**: 6626–6633.

Gosalbes MJ, Lamelas A, Moya A & Latorre A (2008) The striking case of tryptophan provision in the cedar aphid. *J Bacteriol* **190**: 6026–6029.

Haldane JBS (1940) Unsolved problems of science. IV: can we make life? *Keeping Cool and Other Essays* (Haldane JBS, ed), pp. 23–30. Chatto and Windus, London.

Hashimoto M, Ichimura T, Mizoguchi H *et al.* (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol Microbiol* **55**: 137–149.

Heinemann JA & Ankenbauer RG (1993) Retrotransfer of IncP plasmid R751 from *Escherichia coli* maxicells: evidence for the genetic sufficiency of self-transferable plasmids for bacterial conjugation. *Mol Microbiol* **10**: 57–62.

Heinemann M & Panke S (2006) Synthetic biology – putting engineering into biology. *Bioinformatics* **22**: 2790–2799.

Hess WR (2004) Genome analysis of marine photosynthetic microbes and their global role. *Curr Opin Biotech* **15**: 191–198.

Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC & Stetter KO (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 27–28.

Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, Raineri E, Garriga-Canut M & Serrano L (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**: 840–845.

Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M & Santi DV (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *P Natl Acad Sci USA* **101**: 15573–15578.

Kolisnychenko V, Plunkett G III, Herring CD, Feher T, Posfai J, Blattner FR & Posfai G (2002) Engineering a reduced *Escherichia coli* genome. *Genome Res* **12**: 640–647.

Kuwahara H, Yoshida T, Takaki Y *et al.* (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* **17**: 881–886.

Lamelas A, Perez-Brocal V, Gomez-Valero L, Gosalbes MJ, Moya A & Latorre A (2008) Evolution of the secondary symbiont *Candidatus* Serratia symbiotica in aphid species of the subfamily Lachninae. *Appl Environ Microbiol* **74**: 4236–4240.

Lane CE (2007) Bacterial endosymbionts: genome reduction in a hot spot. *Curr Biol* **17**: R508–R510.

Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA III, Smith HO & Venter JC (2007) Genome transplantation in bacteria: changing one species to another. *Science* **317**: 632–638.

Linshiz G, Yehezkel TB, Kaplan S, Gronau I, Ravid S, Adar R & Shapiro E (2008) Recursive construction of perfect DNA molecules from imperfect oligonucleotides. *Mol Syst Biol* **4**: 191.

Luisi PL (2007) Chemical aspects of synthetic biology. *Chem Biodivers* **4**: 603–621.

Mansy SS, Schrum JP, Krishnamurthy M, Tobe S, Treco DA & Szostak JW (2008) Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**: 122–125.

McCutcheon JP & Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *P Natl Acad Sci USA* **104**: 19392–19397.

Mizoguchi H, Mori H & Fujio T (2007) *Escherichia coli* minimum genome factory. *Biotechnol Appl Bioc* **46**: 157–167.

Morimoto T, Kadoya R, Endo K *et al.* (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res* **15**: 73–81.

Morowitz HJ (1992) *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis*. Yale University Press, Binghamton, NY.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA & Giovannoni SJ (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Moya A, Pereto J, Gil R & Latorre A (2008) Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nat Rev Genet* **9**: 218–229.

Mushegian AR & Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *P Natl Acad Sci USA* **93**: 10268–10273.

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA & Hattori M (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**: 267.

Nambiar KP, Stackhouse J, Stauffer DM, Kennedy WP, Eldredge JK & Benner SA (1984) Total synthesis and cloning of a gene coding for the ribonuclease S protein. *Science* **223**: 1299–1301.

Newton IL, Woyke T, Auchtung TA *et al.* (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315**: 998–1000.

O'Malley MA, Powell A, Davies JF & Calvert J (2008) Knowledge-making distinctions in synthetic biology. *Bioessays* **30**: 57–65.

Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG & Hurst LD (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440**: 667–670.

Parens E, Johnston J & Moses J (2008) Ethics. Do we need "synthetic bioethics"? *Science* **321**: 1449.

Partensky F, Hess WR & Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol R* **63**: 106–127.

Pennisi E (2007) Genetics. Replacement genome gives microbe new identity. *Science* **316**: 1827.

Pereto J & Catala J (2007) The renaissance of synthetic biology. *Biol Theor* **2**: 128–130.

Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A & Latorre A (2006) A small microbial genome: the end of a long symbiotic relationship? *Science* **314**: 312–313.

Posfai G, Plunkett G III, Feher T *et al.* (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**: 1044–1046.

Reznikoff WS & Winterberg KM (2008) Transposon-based strategies for the identification of essential bacterial genes. *Microbial Gene Essentiality: Protocols and Bioinformatics*, *Vol. 416* (Osterman AL & Gerdes SY, eds), pp. 13–26. Humana Press, Totowa, NJ.

Rocap G, Larimer FW, Lamerdin J *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Sancar A, Hack AM & Rupp WD (1979) Simple method for identification of plasmid-coded proteins. *J Bacteriol* **137**: 692–693.

Sekiya T, Takeya T, Brown EL *et al.* (1979) Total synthesis of a tyrosine suppressor tRNA gene. XVI. Enzymatic joinings to form the total 207-base pair-long DNA. *J Biol Chem* **254**: 5787–5801.

Serrano L (2007) Synthetic biology: promises and challenges. *Mol Syst Biol* **3**: 158.

Smith HO, Hutchison CA III, Pfannkoch C & Venter JC (2003) Generating a synthetic genome by whole genome assembly: {phi}X174 bacteriophage from synthetic oligonucleotides. *P Natl Acad Sci USA* **100**: 15440–15445.

Stemmer WPC, Crameri A, Ha KD, Brennan TM & Heyneker HL (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**: 49–53.

Suzuki N, Nonaka H, Tsuge Y, Inui M & Yukawa H (2005a) New multiple-deletion method for the *Corynebacterium glutamicum* genome, using a mutant lox sequence. *Appl Environ Microbiol* **71**: 8472–8480.

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

Suzuki N, Okayama S, Nonaka H, Tsuge Y, Inui M & Yukawa H (2005c) Large-scale engineering of the *Corynebacterium glutamicum* genome. *Appl Environ Microbiol* **71**: 3369–3372.

Suzuki N, Tsuge Y, Inui M & Yukawa H (2005d) Cre/loxP-mediated deletion system for large genome rearrangements in *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **67**: 225–233.

Szostak JW, Bartel DP & Luisi PL (2001) Synthesizing life. *Nature* **409**: 387–390.

Tamames J, Gil R, Latorre A, Pereto J, Silva FJ & Moya A (2007) The frontier between cell and organelle: genome analysis of *Candidatus* Carsonella ruddii. *BMC Evol Biol* **7**: 181.

Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X & Church G (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**: 1050–1054.

Tolonen AC, Liszt GB & Hess WR (2006) Genetic manipulation of *Prochlorococcus* strain MIT9313: green fluorescent protein expression from an RSF1010 plasmid and Tn5 transposition. *Appl Environ Microbiol* **72**: 7607–7613.

Waters E, Hohn MJ, Ahel I *et al.* (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *P Natl Acad Sci USA* **100**: 12984–12988.

Westers H, Dorenbos R, van Dijl JM *et al.* (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Mol Biol Evo* **20**: 2076–2090.

Wu D, Daugherty SC, Van Aken SE *et al.* (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* **4**: e188.

Yount B, Curtis KM, Fritz EA *et al.* (2003) Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *P Natl Acad Sci USA* **100**: 12995–13000.

Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, Lee WS, Kim MS & Kim SC (2002) Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. *Nat Biotechnol* **20**: 1018–1023.

Yu X-C & Margolin W (1999) FtsZ ring clusters in min and partition mutants: role of both the Min system and the nucleoid in regulating FtsZ ring localization. *Mol Microbiol* **32**: 315–326.