COMPARATIVE GENOMICS AND THE GENE COMPLEMENT OF A MINIMAL CELL

SARA ISLAS [1], ARTURO BECERRA [1], P. LUIGI LUISI [2] and ANTONIO LAZCANO [1]*

[1] *Facultad de Ciencias, UNAM, Apdo. Postal 70-407, Cd. Universitaria, 04510 Mexico D.f., Mexico;* [2] *ETH-Zentrum, Institut für Polymere, Universitätstrasse 6, CH-8092 Zürich, Switzerland*

(author for correspondence, e-mail: alar@correo.unam.mx)*

(Received 21 January 2002; accepted in revised form 12 May 2003)

Abstract. The concept of a minimal cell is discussed from the viewpoint of comparative genomics. Analysis of published DNA content values determined for 641 different archaeal and bacterial species by pulsed field gel electrophoresis has lead to a more precise definition of the genome size ranges of free-living and host-associated organisms. DNA content is not an indicator of phylogenetic position. However, the smallest genomes in our sample do not have a random distribution in rRNA-based evolutionary trees, and are found mostly in (a) the basal branches of the tree where thermophiles are located; and (b) in late clades, such as those of Gram positive bacteria. While the smallest-known genome size for an endosymbiont is only 450 kb, no free-living prokaryote has been described to have genomes $<$1450 kb. Estimates of the size of minimal gene complement can provide important insights in the primary biological functions required for a sustainable, reproducing cell nowadays and throughout evolutionary times, but definitions of the minimum cell is dependent on specific environments.

Keywords: minimum gene set, minimal cellular genomes, genetic redundancy, DNA content

1. Introduction

Definition of the properties of a minimal cell is a notoriously complex question which is related not only to the understanding of the essential properties of a living system, but is also germane to the issue of the origin of life and early stages of cellular evolution. Several different, complementary approaches to this problem are already feasible or may be available in the near future, including the development of experimental systems based on populations of replicating polymers such as RNA molecules (Joyce, 2002), the *in vitro* synthesis of artificial cells which can metabolize, multiply and adapt (Szostak *et al.*, 2001; Pohorille and Deamer, 2002), the empirical characterization of intracellular endosymbionts and obligate parasites with highly streamlined genomes (Morowitz, 1967; Morowitz and Wallace, 1973; Mira *et al.*, 2001; Gil *et al.*, 2002), the trimming of extant prokaryotic genomes by knock-out experiments and transposon mutagenesis (Itaya, 1995; Hutchinson *et al.*, 1999), and the recently advertised attempt to design a novel form of life with a completely artificial genome (Marshall, 2002).

Origins of Life and Evolution of the Biosphere **34:** 243–256, 2004.

The characteristics of a minimal cell may be inferred from the existence of the basic components required for reproduction and self-maintenance under given environmental conditions (Luisi *et al.*, 2002). From the viewpoint of comparative genomics, the characterization of a minimal cell is equivalent to the identification of the minimum number of genes required by an unicellular organism. Such estimates can provide important insights in the primary biological functions required for a sustainable, reproducing cell nowadays and throughout evolutionary times. However, the definition of minimal genome is determined to a considerable extent by the specific environment in which the presumed minimal cell is found (Space Science Board/National Research Council, 1999; Riley and Serres, 2000). Free-living, unicellular organisms may exist with genomes smaller than the 1.45 Mb lower-limit exhibited by extant prokaryotes (see below), but all the available evidence suggests that nowadays reduced, highly-streamlined genomes like those of *Buchnera* and the mycoplasma are viable only under the permissive, nutrient-rich, stable intracellular environment of their hosts (Mira *et al.*, 2001; Gil *et al.*, 2002). However, the situation must have been different during the earliest stages of biological evolution, when it is assumed that simpler, free-living cells with genomes even smaller than those of *Buchnera* and *Mycoplasma genitalium* must have proliferated.

A minimal gene set can be estimated by the presence or absence of homologous genes based on whole-genome computational sequence comparisons (Mushegian and Koonin, 1996) and, similarly, by the determination of the set of sequences shared among fully sequenced proteomes, i.e., the universal protein families (Kyrpides *et al.*, 1999; Hutchinson *et al.*, 1999). Significant variations may exist between the lenghts of prokaryotic genes (Tekaia *et al.*, 2002). However, on a first approximation bacterial genes may be considered of similar size and tightly packed, i.e., the number of prokaryotic genes is proportional to genome size (Casjens, 1998). Hence, additional insights on the minimal amount of DNA required by extant cells may also be achieved by a statistical analysis of prokaryotic genome sizes (Herdman, 1985; Casjens, 1998; Shimkest, 1998). Previous attempts to analyze the distribution of bacterial DNA content were based on a sample of 603 prokaryotic genome sizes derived by different methodologies, such as renaturation kinetics and colorimetric techniques (Herdman, 1985), which have very different degrees of accuracy. With the development of pulse-field gel electrophoresis (PFGE), a technique that allows the separation and analysis of large DNA fragments and the direct study of the physical structure of genomes, however, the accuracy in the determination of genome sizes has been significantly improved. Here we report the results of an analysis of a database of 641 prokaryote genome sizes determined by PFGE that we have compiled from the published literature, and discuss its significance in providing insights in a minimal cellular genome. The approach developed here is very similar to that reported by Shimkets (1998), and may be considered complementary. We also discuss here how the high levels of genetic redundancy detected in all sequenced genomes can be used to obtain insights in simpler living systems without the large sets of enzymes and the soph-

isticated regulatory abilities of contemporary organisms, that are hypothesized to have existed prior to the divergence of the three major domains, which lacked.

2. Material and Methods

A genome size database has been constructed with the 641 prokaryotic DNA content values determined by PFGE reported in publications included in the NCBI/PubMed database (http://www.ncbi.nlm.nih.gov/PubMed/) as of November 2002. The organisms in this database have been divided into four major groups: (i) free-living Archaea and Bacteria (including pathogens and symbionts that remain separate and have free-living stages); (ii) thermophilic prokaryotes (optimal growth temperature $>45\,^{\circ}C$); (iii) obligate parasites; and (iv) endosymbionts, excluding mitochondria and chloroplasts. The information was completed with the phylogenetic position (not shown) and lifestyle of each organism, based both on the original reports and on data from the *Bergey's Manual of Bacterial Determination* (Holt *et al.*, 1994). The database is periodically updated and is available upon request. We have estimated the levels of genetic redundancy in the smallest genomes of endosymbionts and obligate parasites using the database of levels of paralogy (Total Proteins Hits) available from the Institute for Genomic Research (TIGR, http://www.tigr.org). To be considered redundant, all the ORFs in a given genome, whether annotated or not, were compared using BLAST and had to exhibit at least 60% sequence similarity ($P < 0.0001$). The result of this comparison is shown in Table II, where the sizes of some of the smallest known cellular genomes are indicated in kb, together with the number of ORFs, the number of redundants found in each genome, and the corresponding percentage per genome.

3. Results

The genome size distribution in our database is shown in Figure 1. The values of DNA content of free-living prokaryotes can vary over a tenfold range, from *Halomonas halmophila*, a moderately halophilic gamma proteobacteria endowed with a small 1450 kb genome (Mellado *et al.*, 1998), to the 9700 kb genome of *Azospirillium lipoferum* Sp59b (Martin-Didonet *et al.*, 2000). The widest range of genome sizes is exhibited by the proteobacteria, from the 450 kb *Buchnera* genome, to the largest ones in the sample, which correspond to aerobic organisms with complex life cycles which can include formation of spores and mycelia. There are no reports of archaeal genomes as large as those of *Azospirillum* and *Stigmatella*, perhaps due to incomplete sampling. All the archaeal genomes in our sample are small and fall within the 500 to 5100 kb range. These size ranges correspond in fact to those of thermophilic bacterial and archaeal genomes, were the lower and upper limits appear to correspond to extreme cases, i.e., the 500 kb chromosome of

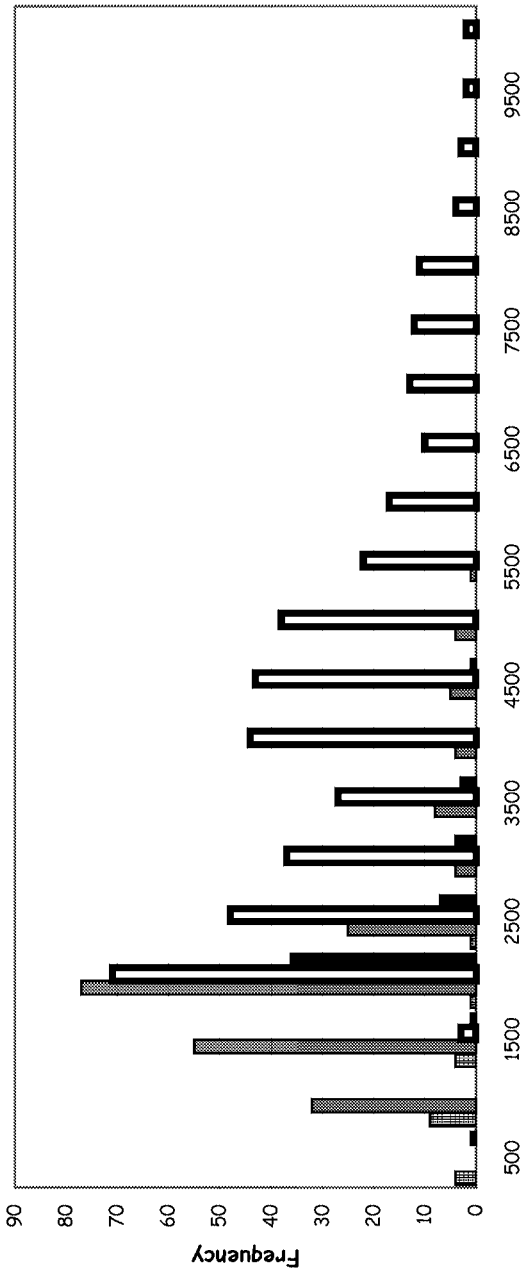

Figure 1. Prokaryotic genome size distribution ($N = 641$). Open boxes, free-living prokaryotes; grey boxes, obligate parasites; black boxes, thermophiles; boxes with horizontal lines, endosymbionts.

the thermophilic ectosymbiont *Nanoarchaeon equitans* (Hubert *et al.*, 2002), and the 5100 kb of the facultative thermophilic *Methanosarcina acetivorans* (Sowers *et al.*, 1988).

Classification of endosymbionts as a group by themselves shows that although their genome size distribution overlaps with that of obligate parasites (Figure 1), their DNA content can reach values significantly smaller that those of the smallest parasites, i.e., the mycoplasma. The smallest-known cellular genome is only 450 kb and corresponds to the obligate endosymbiont proteobacterium *Buchnera* spp. (Gil *et al.*, 2002), significantly smaller than the lower limit of 580 kb of the Mollicutes, which corresponds to the obligate parasite *Mycoplasma genitalium* (Fraser *et al.*, 1995). Other groups with reduced genome sizes are the rickettsia and several spirochaete. The DNA content values of other obligate parasites and organisms with stringent growth conditions, which we have grouped with the mycoplasma, however, can reach values as large as the 5016 kb of *Mycobacterium intracellulare* (Kim *et al.*, 1996).

4. Discussion

The data summarized in Figure 1 is clearly biased and does not reflect in an accurate way the actual levels of prokaryotic diversity. Because of their significance in medical and economical significance in human, animal, and crop plant life, pathogens and parasites are clearly overrepresented in our sample. Moreover, the overlap in the 2000 to 3000 kb region in Figure 1 of several of the categories used here to group the species in our sample shows that prokaryotes with similar genome sizes but different lifestyles can have very different complement of genes.

In spite of these limitations, the data summarized in Figure 1 provides useful insights into the evolution of prokaryotic DNA content and the size of a minimal cellular gene set. Considerable variations in DNA content may exist even within closely related bacterial species and strains (Bergthorsson and Ochman, 1995; Casjens, 1998), but as shown by the genomes of genera like *Helicobacter* and *Streptomyces*, this is not always the case (Shimkets, 1998). The size range of bacterial genome sizes are clearly less constrained than that of the archeal chromosomes. Our results also demonstrate the unsurpassed genome plasticity of the proteobacterial clade. While some members of the group like the myxobacteria have undergone major expansion of their encoding abilities adapting to oxygen-rich environments and developing complex life cycles, others like *Buchnera* have followed an opposite direction and lost considerable amounts of DNA as they adapted to an intracellular environment (Gil *et al.*, 2002).

The thermophilic bacterial and archaeal genomes tend to be relatively small, with the lowest limit represented by the 500 kb chromosome of the thermophilic ectosymbiont *Nanoarchaeon equitans* (Huber *et al.*, 2002). The 5100 kb genome of the facultative thermophilic *Methanosarcina acetivorans* is probably atypical.

However, the size range of thermophilic genomes does not necessarily reflect a correlation between DNA content, heat-loving microbial lifestyles and antiquity, since a wide variety of mesophilic bacterial groups, including leptospira, green-sulfur bacteria, cyanobacteria, spirochaetes, fusobacteria, and actinobacteria, can also exhibit small-sized genomes.

The smallest, highly-streamlined genomes in our sample do not have a random phylogenetic distribution. The phylogenetic mapping of genome sizes on the 16/18S rRNA tree (not shown) demonstrates that the reduction of prokaryotic genome size has occurred independently multiple times in separate lineages, and persists as an end-state character with the organisms deriving essential nutrients from a host. Although endosymbionts and intracellular parasites have many features in common, including massive gene losses as they adapted into the nutrient-rich environment provided by their hosts, grouping them into two different categories allows some insights into the differences that exist between these two lifestyles (Figure 1). For instance, it is likely that the larger size of intracellular parasite genomes, as compared to those of endosymbionts, is due to the presence of genetically encoded specifically related to parasitic lifestyles, such as sequences involved in host-parasite recognition and infection mechanisms.

Figure 1 provides no support for the hypothesis that the size distribution of extant prokaryotic chromosomes is the outcome of a series of whole genome duplications that begun with an ancestral 800 kb minigenome as suggested by Wallace and Morowitz (1973) and Herdman (1985). Since there are no known free-living prokaryotes with genomes smaller than the 1450 kb, 1500 kb, and 1530 kb of *Halomonas halmophila* (Mellado *et al.*, 1998), *Aquifex pyrophilus* (Shao *et al.*, 1994) and *Fervidobacterium islandicum*, respectively, the extrapolation of a normal distribution curve beyond this cut-off value does not seem justified. However, as argued by Shimkets (1998) on the basis of a smaller sample of 141 chromosomes of prokaryotes grouped as generalists and specialists, the minimum genome size for a living organism is approximately 600 kb, a figure that fits nicely with the small genomes of *Mycoplasma genitalium* and the different *Buchnera* species (Fraser *et al.*, 1995; Gil *et al.*, 2002). The independent, massive gene losses that these two types of bacteria have undergone suggest that their limited encoding capacities are feasible only because of their adaptation to the highly permissive intracellular environments provided by their hosts.

5. How Small Can Viable Cells Be?

One of the earliest attempts to describe both in functional and evolutionary terms the minimal set of characteristics that a cell must fulfill to be considered alive was undertaken by Morowitz (1967). Based on the enzymatic components of primary metabolism whose presence he assumed was required for DNA-based cell repro-

duction, Morowitz estimated the size of a minimal cell that turned out to be about one-tenth smaller than mycoplasma.

As reviewed elsewhere (Luisi *et al.*, 2002), the defining characteristics of a minimal cell now and throughout the past has been discussed by Varela *et al.* (1974), Woese (1983), Oro and Lazcano (1984), Dyson (1985), Jay and Gilbert (1987), Morowitz (1992), Walde *et al.* (1994), Oberholzer *et al.* (1995), Ganti (1997), and Szostak *et al.* (2001). Perhaps not surprisingly, the rapid pace at which more and more completely sequenced cellular genomes become available has shifted the emphasis towards deducing the minimum number of protein-encoding genes required for cellular life outside a host cell and under laboratory conditions.

Following the publication of the complete genomes of *Haemophilus influenza* and *M. genitalium*, Mushegian and Koonin (1996) published the results of a detailed comparison of these two species in conjuction with the fragmentary data from other organisms then available. Once parasite-specific sequences were discarded, the final outcome was an inventory of 256 genes that according to Mushegian and Koonin resembles not only the genetic complement of the ancestor of the Gram-negative and Gram-positive lineages to which *H. influenza* and *M. genitalium*, respectively, belong, but also the amount of DNA required to sustain a modern type minimal cell under permissible conditions. Since most of the 256 sequences shared by these two organisms have eukaryotic and/or archaeal homologs, Mushegian and Koonin also discussed how this figure could be reduced to describe the genome of the last common ancestor of the Bacteria, Archaea and Eukarya, and suggested that their results could provide insights into the earliest stages of biological evolution.

As underlined by Koonin (2000), the estimated 256 minimal gene set complement derived from the comparison of the *H. influenzae* and *M. genitalium* genomes is quite similar to the values of viable minimal genome sizes inferred by site-directed gene disruptions in *B. subtilis* (Itaya, 1995) and transposon-mediated mutagenesis knock-outs in *M. genitalium* and *M. pneumoniae* (Hutchinson *et al.*, 1999). These figures are also consistent with the estimate that the universal family of proteins shared among fully sequenced cellular genomes comprises 324 sequences (Kyrpides *et al.*, 1999) and, as summarized in Table I, with the sizes of the *Buchnera* genomes (Gil *et al.*, 2002), and the 551 kb vestigial nucleus or nucleomorph found in cryptomonads, and which is the outcome of a secondary endosymbiotic event in which a protist engulfed an already existing unicellular eukaryotic alga which was then reduced to a secondary plastid (Douglas *et al.*, 2001).

However, considerable caution is required to avoid an overinterpretation of these different estimates. Although the backtrack methodology proposed by Mushegian and Koonin (1996) is quite straightforward, their estimates do not consider proteins that perform the same function but have different sequences (Riley and Serres, 2000), either because they have diverged beyond recognition or because they are in fact analogous. Equally important, they failed to consider polyphyletic

TABLE I
Some miniature cellular genomes

Species	Genome size (kb)	Lifestyle	Reference
Mycoplasma genitalium	580	obligate parasite	Fraser *et al.*, 1995
Buchnera spp.	450	endosymbiont	Gil *et al.*, 2002
crytomonad nucleomorph	551	secondary endosymbiont	Douglas *et al.*, 2001

gene losses which have been involved in the size reduction of the *M. genitalium* and *H. influenzae* genomes, and which led to the loss of purine- and pyrimidine nucleotide biosynthetic pathways, among others (Becerra *et al.*, 1997).

As the number of fully sequenced genomes has increased, their comparison has led to smaller sets of minimum gene complements, which are now reduced to approximately 80 orthologous sequences common to all life forms (Koonin, 2000). Quite surprisingly, some of the most likely *a priori* candidates for strict universality, such as those sequences involved in DNA replication, have also turned out to be not only poorly preserved but also, in some cases, of polyphyletic origin (Edgell and Doolittle, 1997; Olsen and Woese, 1996; Böhlke *et al.*, 2002). If the term 'universal distribution' is restricted to its most obvious sense, i.e., that of traits found in all completely sequenced genomes now available, then quite unexpectedly the resulting repertoire is formed by relatively few features and by incompletely represented biochemical processes (Tatusov *et al.*, 1997; Tekaia *et al.*, 1999; Brown *et al.*, 2001; Delaye *et al.*, 2002). As argued elsewhere (Islas *et al.*, submitted), such inventories include sequences that originated in different epochs, including some which may have arisen in the RNA/protein world (Tekaia *et al.*, 1999; Delaye and Lazcano, 2000; Lazcano, 2001; Anantharaman *et al.*, 2002). Hence, the figures reported by Mushegian and Koonin (1996) and Koonin (2000) represent, at the best, lower limits of the actual size of minimal gene-encoded functions required by a cell living under highly permissive environmental conditions. Thus, such estimates do not provide accurate models for the properties of ancestral Archean genomes.

6. The Search for a Minimal Cell: Beyond Genetic and Functional Redundancy

Recognition that the biochemical complexity of extant organisms is the outcome of process of biological evolution that started perhaps 4×10^9 years ago can lead to some inferences on smaller ancestral cells endowed with less complex genome replication apparatus and simpler gene expression mechanisms. In spite of the structural and functional similarities between the template-directed en-

zymatic synthesis of RNA and DNA, double-stranded DNA cellular genomes replicate via a large, complex array of molecular components in which proof-reading DNA polymerases play a central role. However, a number of experimental results and sequence comparisons suggest that replication of a DNA genome can be achieved with a simplified set of catalysts (Delaye *et al.*, 2002). For instance, the RNA-primer formation is catalyzed in mitochondria not by a primase but by the organellar DNA-dependent monomeric RNA polymerase (Frick and Richardson, 2001). This suggests that a smaller set of less-specific polymerases could be functional and, in fact, may have existed during the early stages of cell evolution. Thus, a working model of a simpler DNA-cell may be envisioned in which a single ancestral polymerase, whose evolutionary vestiges appear to be present in the catalytic palm domain of the DNA pol I and its homologs such as the T7 phage RNA polymerase (Delaye *et al.*, 2001), could play multiple roles as a DNA polymerase, a transcriptase and a primase.

Similar arguments can be advocated for a simplified version of protein synthesis requiring less components. For instance, the fact that RNA molecules are capable of perfoming by themselves all the reactions involved in peptide-bond formation suggests that protein biosynthesis evolved in an RNA world (Zhang and Cech, 1998), i.e., that the first ribosome lacked proteins and was formed only by RNA. This possibility is supported by the crystallographic data that has shown that ribosome catalytic site where peptide bond formation takes place is composed solely of RNA (Nissen *et al.*, 2000).

Additional clues to the genetic organization of primitive forms of translation involving less components are provided by paralogous genes, which are sequences that diverge not through speciation but after a duplication event. Such genetic redundancies are a common feature of all known cellular genomes, including those of the smallest described lifeforms (Table II). Accordingly, the presence in all known cells of pairs of homologous genes encoding two elongation factors, which are GTP-dependent enzymes that assist in protein biosynthesis, provide evidence of the existence of a more primitive, less-regulated version of protein synthesis took place with only one elongation factor. In fact, the experimental evidence of *in vitro* translation systems with modified cationic concentrations lacking both elongation factors and other proteinic components (Gavrilova *et al.*, 1976; Spirin, 1986) strongly supports the possibility of an older ancestral protein synthesis apparatus prior to the emergence of elongation factors.

7. Concluding Remarks

The properties of a minimal cell can be approached in two different but complementary directions. One possibility involves the laboratory synthesis of encapsulated cell-like systems which may eventually metabolize, multiply and adapt (Szostak *et al.*, 2001). An alternative approach involves the study of extant min-

TABLE II
Genetic redundancies in small genomes of endosymbionts and obligate parasites [a]

Proteome	Genome sizes (kb)	Number of ORFs	Number of redundant sequences	% of redundancy
Mycoplasma genitalium	580	480	52	10.83
Mycoplasma pneumoniae	816	688	134	19.47
Buchnera sp. APS	640	574	67	11.67
Ureaplasma urealyticum	751	611	105	17.18
Chlamydia trachomatis	1000	895	60	6.71
Chlamydia muridarum	1000	920	60	6.52
Chlamydophila pneumoniae J138	1200	1070	148	13.83
Rickettsia prowazekii	1100	834	49	5.87
Rickettsia conorii	1200	1366	189	13.83
Treponema pallidum	1100	1031	78	7.56

[a] Genome sizes, complete proteomes, and the number of ORFs were all retrieved from NCBI http://www.ncbi.nlm.nih.gov.

imal genomes in order to describe cells with decreasing degrees of complexity. As discussed here, the small values of DNA content found in widely separated microbial species do not represent a primitive trait, but are in fact the outcome of polyphyletic sequence losses that have occurred in recent clades. Thus, they are excellent laboratory models to study the properties of the genetic and metabolic repertoire of minimal cells, but the information they provide on their evolutionary predecessors, specially those that may have existed during Archaean times, is rather limited. Primitive cells were probably endowed not only with less genes, but also with less complex sequences and simpler mechanisms of gene expression.

As discussed here, an examination of the distribution of DNA content of Archaea and Bacteria complements other genomic approaches, even if our conclusions are hindered by the nature of the available information. All known organisms share a core of highly conserved, genetically-encoded features, a significant portion of which corresponds to the translation machinery and is maintained even in highly streamlined genomes such as those of Table 1. However, our methodology is hindered by the fact that prokaryotes with similar genome sizes can have very different complements of genes. Regardless of one's definition of life, the size and content of the minimal gene set required for life will be strongly determined by the environment of the minimum cell itself. The search for minimal living systems under highly permissive conditions should thus be complemented with the search for free-living prokaryotes with genomes smaller than those of *H. halmophila*, in order to understand the minimum gene content for sustaining viability. The existence of extremely reduced 55S mitochondrial ribosomes in *Caenorhadbditis elegans*

(Mears *et al.*, 2002), as compared to its 70S prokaryotic counterpart, suggest that other organisms may exist with novel or reduced version of the essential molecular machinery. Whether such prokaryotes exist or not is not yet known, but the current cut-off values of genome size distribution curves (Figure 1) suggest that considerable attention should be given to the search for similar free-living prokaryotes and the sequencing of their genomes.

The experimental efforts to define the essential genes required for life under highly permissive conditions have shown mutant *M. genitalium* populations with 265 to 350 genes can growth and divide under laboratory conditions (Hutchinson *et al.*, 1999). Extrapolation of these results to the early evolution of life may help us to understand some of the essential characteristics, but additional efforts are required for a proper understanding of the evolutionary transition between putative RNA-cells and full-flegged DNA/protein cells. Insights into such intermediate stages are provided by analysis of genetic redundancy (Table II) and by the experimental evidence reviewed here that has demonstrated that under *in vitro* conditions protein synthesis can take place even in the absence of some of its molecular components. Indeed, the selection and maintenance of laboratory strains in which paralogous copies of highly conserved genes such as those encoding the two elongation factors involved in protein synthesis would be substituted by one single, less-specific catalyst appear to be feasible with the available experimental techniques.

Acknowledgements

The suggestions of Dr. Cesar Hernandez and the assistance of Mlle. Ana Maria Velasco are gratefully acknowledged. A.L. is an Affiliate of the NSCORT (NASA Specialized Center for Research and Training) in Exobiology at the University of California, San Diego.

References

Anantharaman, V., Koonin, E. V. and Aravind, L.: 2002, Comparative Genomics and Evolution of Proteins Involved in RNA Metabolism, *Nucleic Acid Res.* **30**, 1427–1464.

Bergthorsson, U. and Ochman, H.: 1995, Heterogeneity of Genome Sizes among Natural Isolates of *Escherichia coli, J. Bact.* **177**, 5784–5789.

Böhlke, K., Pisani, F. M., Rossi, M. and Antranikian, G.: 2002, Archaeal DNA Replication: Spotlight on a Rapidly Moving Field, *Extremophiles* **6**, 1–14.

Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. and Stanhope, M. J.: 2001, Universal Trees Based on Large Combined Protein Sequence Datasets, *Nature Genetics* **28**, 281–285.

Casjens, S.: 1998, The Diverse and Dynamic Structure of Bacterial Genomes, *Annu. Rev. Genet.* **32**, 339–377.

Delaye, L., Vázquez, H. and Lazcano, A.: 2001, The Cenancestor and its Contemporary Biological Relics: The Case of Nucleic Acid Polymerases, in J. Chela-Flores, T. Owen and F. Raulin (eds.), *First Steps in the Origin of Life in the Universe*, Kluwer Academic Publishers, Dordrecht, pp. 223–230.

Delaye, L. and Lazcano, A.: 2000, RNA-Binding Peptides as Molecular Fossils, in J. Chela-Flores, G. Lemerchand and J. Oró (eds.), *Origins from the Big-Bang to Biology: Proceedings of the First Ibero-American School of Astrobiology*, Kluwer Academic Publishers, Dordrecht, pp. 285–288.

Delaye, L., Becerra, A. and Lazcano, A.: 2002, The Nature of the Last Common Ancestor, in L. Ribas de Pouplana (ed.), *The Genetic Code and the Origin of Life*, Landes Bioscience, Georgetown, in press.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.-T., Wu, X., Reith, M., Cavalier-Smith, T., and Maier, U.-G.: 2001, The Highly Reduced Genome of an Enslaved Algal Nucleus, *Nature* **410**, 1091–1096.

Dyson, F. J.: 1985, *Origins of life*, Cambridge University Press, Cambridge.

Edgell, R. D. and Doolittle, W. F.: 1997, Archaea and the Origin(s) of DNA Replication Proteins, *Cell* **89**, 995–998.

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. and Kelley, J. M., *et al.*: 1995, The Minimal Gene Complement of *Mycoplasma Genitalium, Science* **270**, 397–403.

Frick, D. N. and Richardson, C. C.: 2001, DNA Primases, *Annu. Rev. Biochem.* **70**, 39–80.

Ganti, T.: 1997, Biogenesis Itself, *J. Theoret. Biol.* **187**, 583–593.

Gavrilova, L. P., Kostiashkina, O. E., Koteliansky, V. E., Rutkevitch, N. M. and Spirin, A. S.: 1976, Factor-Free (Non-Enzymic) and Factor-Dependent Systems of Translation of Polyuridylic Acid by *Escherichia coli* Ribosomes, *J. Mol. Biol.* **101**, 537–552.

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.

Herdman, M.: 1985, The Evolution of Bacterial Genomes, in T. Cavalier-Smith (ed.), *The Evolution of Genome Size*, John Wiley and Sons, New York, pp. 37–68.

Holt, G., Krieg, R., Sneath, A., Staley, T. and Williams, T.: 1994, *Bergey's Manual of Determinative Bacteriology*, 9th Edition, Williams and Wilkins.

Huber, H., Hohn, M. J., Rachel R., Fuchs, T., Wimmer, C. V. and Stetter, K.: 2002, A New Phylum of Archaea Represented by a Nanosized Hyperthermophilic Symbiont, *Nature* **417**, 63–67.

Hutchinson, 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.

Islas, S., Velasco, A. M., Becerra, A., Delaye, L. and Lazcano, A.: 2003, Hyperthermophily and the Origin and Earliest Evolution of Life, *Inter. Microbiol.*, submitted.

Itaya, M.: 1995, An Estimation of the Minimal Genome Size Required for Life, *FEBS Letters* **362**, 257–260.

Jay, D. and Gilbert, W.: 1987, Basic Protein Enhances the Encapsulation of DNA into Lipid Vesicles: Model for the *Formation* of Primordial Cells, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1978–1980.

Joyce, G. F.: 2002, The Antiquity of RNA-Based Evolution, *Nature* **418**, 214–221.

Kim, J. R., Kang, B. S., Ko, J. H., Park, J. S., Kim, S. J., Bai, G. H., Chung, T. H., Nam, K. S., Choi, Y. K., Choi, I. S., Chung, T. W., Lee, Y. C. and Kim, C. H.: 1996, Genomic Heterogeneity in Clinical Strains of *Mycobacterium tuberculosis, M. terraecomplex, M. gordonae, M. avium-intracellulare complex*, and *M. fortuitum* by Pulsed-Field Gel Electophoresis, *J. Biochem. Mol. Biol.* **29**, 569–573.

Koonin, E. V.: 2000, How Many Genes Can Make a Cell: The Minimal-Gene-Set Concept, *Annu. Rev. Genomics Human Genet.* **1**, 99–116.

Kyrpides, N., Overbeek, R. and Ouzonis, C.: 1999, Universal Protein Families and the Functional Content of the Last Universal Common Ancestor, *J. Mol. Evol.* **49**, 413–423.

Lazcano, A.: 2001, El último ancestro común, in E. Martínez Romero y J. Martínez Romero (eds.), *Microbios en Línea*, UNAM, México, pp. 421–429.

Luisi, P. L., Oberholzer, T. and Lazcano, A.: 2002, The Notion of a DNA Minimal Cell: A General Discourse and some Guidelines for an Experimental Approach, *Helv. Chim. Acta.* **85**, 1759–1777.

Marshall, E.: 2002, Genetics: Venter Gets Down to Life Basics, *Science* **298**, 1701.

Martin-Didonet, M., Chubatsu, S. L., Souza, M. E., Klein, A. M., Rego, G. M., Rigo, U. L., Yates, G. M. and Pedrosa, O.: 2000, Genome Structure of the Genus *Azospirillum, J. Bacteriol.* **182**, 4113–4116.

Mears, J. A., Cannone, J. J., Stagg, S. M., Gutell, R. R., Agrawal, R. K. and Harvey, S. C.: 2002, Modeling a Minimal Ribosome Based on Comparative Sequence Analysis, *J. Mol. Biol.* **321**, 215–234.

Mellado, E., Garcia, M. T., Roldan, E., Nieto, J. J. and Ventosa, A.: 1998, Analysis of the Genome of the Gram-Negative Moderate Halophiles *Halomonas* and *Chromohalobacter* by Using Pulsed-Field Gel Electrophoresis, *Extremophiles* **2**, 435–438.

Mira, A., Ochmanm H. and Moran, N. A.: 2001, Deletional Bias and the Evolution of Bacterial Genomes, *Trends Genet.* **17**, 589–596.

Morowitz, H. J.: 1967, Biological Self-Replicating Systems, *Prog. Theor. Biol.* **1**, 35–58.

Morowitz, H. J.: 1992, *The Beginnings of Cellular Life*, Yale University Press, New Haven.

Morowitz, H. J. and Wallace, D. C.: 1973, Genome Size and the Life Cycle of the Mycoplasma, *Ann. N.Y. Acad. Sci.* **225**, 62–73.

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.

Nissen, P., Hansen, J., Ban., N., Moore, P. B. and Steitz, T. A.: 2000, The Structural Basis of Ribosome Activity in Peptide Bond Synthesis, *Science* **289**, 920–930.

Oberholzer, T., Wick, R., Luisi, P. L. and Biebricher, C. K.: 1995, Enzymatic RNA Replication in Self-Reproducing Vesicles: An Approach to a Minimal Cell, *Biochem. Biophys. Res. Comm.* **207**, 250–257.

Olsen, G. and Woese, C. R.: 1996, Lessons from an Archaeal Genome: What Are We Learning From *Methanococcus jannaschii*?, *Trends Genet.* **12**, 377–379.

Oro, J. and Lazcano, A.: 1984, A Minimal Living System and the Origin of a Protocell, *Adv. Space Res.* **4**, 167–176.

Pohorille, A. and Deamer, D.: 2002, Artificial Cells: Prospects for Biotechnology, *Trends Biotech.* **20**, 123–128.

Riley, M. and Serres, M. H.: 2000, Interim Report on Genomics of *Escherichia coli, Annu. Rev. Microbiol.* **54**, 341–411.

Shao, Z., Mages, W. and Schmitt, R.: 1994, A Physical Map of the Hyperthermophilic Bacterium *Aquifex pyrophilus* Chromosome, *J. Bacteriol.* **176**, 6776–6780.

Shimkets, L. J.: 1998, Structure and Sizes of Genomes of the Archaea and Bacteria, in F. J. de Bruijn, J. R. Lupskin and G. M. Weinstock (eds.), *Bacterial Genomes: Physical Structure and Analysis*, Kluwer Academic Publishers, Boston, pp. 5–11.

Sowers, K. and Gunsaluz, R. R.: 1988, Plasmid DNA from Acetotrophic Methanogen, *Methanosarcina Acetivorans, J. Bacteriol.* **170**, 4979–4982.

Space Study Board/National Research Council: 1999, *Size Limits of Very Small Organisms*, National Research Council, National Academy of Sciences, Washington, D.C.

Spirin, A. S.: 1986, *Ribosome Structure and Protein Synthesis*, Benjamin/Cummings, Menlo Park, 414 pp.

Szostak, J. W., Bartel, D. P. and Luisi, P. L.: 2001, Synthesizing Life, *Nature* **409**, 387–390.

Tatusov, R. L., Koonin, E. V. and Lipman, D. J.: 1997, A Genomic Perspective on Protein Families, *Science* **278**, 631–637.

Tekaia, F., Dujon, B. and Lazcano, A.: 1999, Comparative Genomics: Products of the Most Conserved Protein-Encoding Genes Synthesize, Degrade, or Interact with RNA, *Abstracts of the 9th ISSOL Meeting*, San Diego, California, U.S.A., July 11–16, 1999, Abstract c4.6, 53 pp.

Tekaia, F., Yeramian, E. and Dujon, B.: 2002, Amino Acid Composition of Genomes, Lifestyles of Organisms, and Evolutionary Trends: A Global Picture with Correspondence Analysis, *Gene* **297**, 51–60.

Varela, F. J., Maturana, H. R. and Uribe, R.: 1974, Autopoiesis: The Organization of Living Systems, Its Characterization, and a Model, *Curr. Mod. Biol.* **5**, 187–196.

Walde, P., Goto, A., Monnard, P. A., Wessicken, M. and Luisi, P. L.: 1994, Oparin's Reaction Revisited: Enzymatic Synthesis of Poly(adenylic Acid) in Micelles and Self-Reproducing Vesicles, *J. Am. Chem. Soc.* **116**, 7541–7547.

Wallace, D. C. and Morowitz, H. J.: 1973, Genome Size and Evolution, *Chromosome* **40**, 121–126.

Woese, C. R.: 1983, The Primary Lines of Descent and the Universal Ancestor, in D. S. Bendall (ed.), *Evolution from Molecules to Man*, Cambridge Universiy Press, Cambridge, pp. 209–233.

Zhang, B. and Cech, T. R.: 1998, Peptidyl-Transferase Ribozymes: Trans Reactions, Structural Characterization and Ribosomal RNA-Like Features, *Chem. Biol.* **5**, 539–553.