Conservation of ribosomal protein gene ordering in 16 complete genomes

WANG Ning (王 宁)[1], CHEN Runsheng (陈润生)[1]
& WANG Yongxiong (Wing Hung Wong, 王永雄)[2]

1. Protein Engineering Lab, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China;
2. Department of Statistics, University of California at Los Angeles, CA 90095-1554, USA
Correspondence should be addressed to Chen Runsheng and Wang Yongxiong

Received May 19, 1999

Abstract The organization of ribosomal proteins in 16 prokaryotic genomes was studied as an example of comparative genome analyses of gene systems. Hypothetical ribosomal protein-containing operons were constructed. These operons also contained putative genes and other non-ribosomal genes. The correspondences among these genes across different organisms were clarified by sequence homology computations. In this way a cross tabulation of 70 ribosomal proteins genes was constructed. On average, these were organized into 9—14 operons in each genome. There were also 25 non-ribosomal or putative genes in these mainly ribosomal protein operons. Hence the table contains 95 genes in total. It was found that: (i) the conservation of the block of about 20 r-proteins in the L3 and L4 operons across almost the entire eubacteria and archaebacteria is remarkable; (ii) some operons only belong to eubacteria or archaebacteria; (iii) although the ribosomal protein operons are highly conserved within domain, there are fine variations in some operons across different organisms within each domain, and these variations are informative on the evolutionary relations among the organisms. This method provides a new potential for studying the origin and evolution of old species.

Keywords: comparative genome analysis, operon, ribosomal proteins, gene ordering.

With the availability of more and more complete genomes, researchers now have the chance to study how genes are organized in different organisms at the genomic level[1]. There have been some recent attempts to infer phylogenetic relationship between organisms by comparing their gene orders as a supplement to comparisons based on conserved sequences. In this work, we study the organization of ribosomal proteins (r-proteins) in 16 complete genomes as an example of comparative genome analyses of gene systems. These organisms include 12 eubacteria and 4 archaebacteria. The r-proteins are ancient and universal, and, at least for prokaryotes, the analysis of the organizations of the large number of r-proteins within the genomes has the potential to be a novel source of information on the origin and evolutionary relations among organisms.

1 Datasets

Our data were from the published databank. There were totally 16 complete genomes in

GenBank of NCBI before starting our work. These 16 complete genomes include 12 eubacteria and 4 archaebacteria. The 12 eubacteria are: *Aquifex aeolicus* (*A. aeo*)[2], *Borrelia burgdorferi* (*B. bur*)[3], *Treponema pallidum* (*T. pal*)[4], *Chlamydia trachomatis* (*C. tra*)[1)], *Escherichia coli* (*E. coli*)[5], *Haemophilus influenzae Rd* (*H. inf*)[6], *Helicobacter pylori* (*H. pyl*)[7], *Mycoplasma genitalium* (*M. gen*)[8], *Mycoplasma pneumoniae* (*M. pneu*)[9], *Bacillus subtilis* (*B. sub*)[10], *Mycobacterium tuberculosis* (*M. tub*)[11], *Synechocystis PCC*6803 (*S. pcc*)[12]. The remaining 4 archaebacteria are: *Methanococcus jannaschii* (*M. jan*)[13], *Methanobacterium thermoautotrophicum* (*M. the*)[14], *Archaeoglobus fulgidus* (*A. ful*)[15], and *Pyrococcus horikoshii* (*P. hor*)[16]. Table 1 provides the names, phyla sources, and genome sizes of the 16 organisms.

Table 1 List of 16 organisms' name, phyla source and genome size

Name of genome (abbreviation)	Source	Size of genome
Aquifex aeolicus (*A. aeo*)	Eubacteria; Aquificales; Aquificaceae; Aquifex.	1 551 335 bp
Borrelia burgdorferi (*B. bur*)	Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.	910 724 bp
Treponema pallidum (*T. pal*)	Eubacteria; Spirochaetales; Spirochaetaceae; Treponema.	1 138 011 bp
Chlamydia trachomatis (*C. tra*)	Eubacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	1 042 519 bp
Escherichia coli (*E. coli*)	Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	4 639 221 bp
Haemophilus influenzae (*H. inf*)	Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.	1 830 138 bp
Helicobacter pylori (*H.pyl*)	Eubacteria; Proteobacteria; epsilon subdivision; Helicobacter.	1 667 867 bp
Mycoplasma genitalium (*M. gen*)	Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Mycoplasmas and walled relatives; Mycoplasmatales; Mycoplasmataceae; Mycoplasma.	580 073 bp
Mycoplasma pneumoniae (*M.pneu*)	Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Mycoplasmas and walled relatives;Mycoplasmatales; Mycoplasmataceae; Mycoplasma.	816 394 bp
Bacillus subtilis (*B. sub*)	Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Bacillaceae; Bacillus.	4 214 814 bp
Mycobacterium tuberculosis (*M. tub*)	Eubacteria; Firmicutes; Actinomycetes; Mycobacteria; Mycobacteriaceae; Mycobacterium.	4 411 529 bp
*Synechocystis PCC*6803 (*S. pcc*)	Eubacteria; Cyanobacteria; Chroococcales; Synechocystis.	3 573 470 bp
Methanococcus jannaschii (*M. jan*)	Archaea; Euryarchaeota; Methanococcales; Methanococcaceae; Methanococcus.	1 664 970 bp
Methanobacterium thermoautotrophicum (*M. the*)	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae; Methanobacterium.	1 751 377 bp
Archaeoglobus fulgidus (*A. ful*)	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.	2 178 400 bp
Pyrococcus horikoshii (*P. hor*)	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.	1 738 505 bp

1) Stephens, R. S., Kalman, S., Lammel, C. J. et al., Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*, unpublished.

2 Methods

Among the 16 complete genomes, almost each one has about 50 r-proteins, and some of them form operons. For the *E. coli* genome, many documented promoter and transcription start sites are given in the GenBank annotation and in these cases we construct the operons according to the given annotations. However, such documented sites are not available for the other genomes.

2.1 Ascertaining operon's structure in 15 genomes

We found that some r-protein-containing operons often include some non-ribosomal proteins. In this work we use the gene names and inter-gene distances given in GenBank annotations as a guide to construct hypothetical r-protein-containing operons in the 15 bacterial genomes of table 1. The general rule we used is to classify two genes having the same orientation and an inter-gene distance less than 70 bp as belonging to the same operon.

2.2 Finding homologous genes by using dynamic programming algorithm

In GenBank, due to naming confusion, some homologous genes, especially those corresponding to putative coding regions, were given different names in different organisms. For example, the gene *ylqC* is a hypothetical protein which is given the names *ylqC*, BB0696, and TP0906 respectively in the genomes of *B. sub*, *B. bur* and *T. pal*. In order to ensure the reliability of our result, we performed pair-wise (amino-acid based) sequence alignment to clarify the correspondences among all such genes. Similarity of amino acid sequences was computed with the ALIGN program (version 2.0)[17]. If their sequence similarity is higher than 25%, the genes in different genomes are regarded as homologous genes[18].

2.3 Comparing gene ordering of each operon in 16 genomes

If sequence similarity shows that several successive genes in one genome are homologous to several successive genes in another genome, and if the order and orientation of these genes are identical in the two genomes, then this block of genes is potentially an operon. If this structure is conserved over many genomes, then it is very likely to represent a conserved operon. The operon is named from the first r-protein in it.

3 Results

Table 2 presents the conserved hypothetical operons containing ribosomal proteins genes. These operons cover 70 r-proteins genes and 25 non-ribosomal genes. The organization of these 95 genes in each of the 16 genomes is given in the table. The leftmost column gives the gene names (mainly according to *B. sub* annotation). On average, these genes are organized into 9—14 operons in each genome. The main findings are as follows.

3.1 Strong conservation of operons in both eubacteria and archaebacteria

From table 2, the conservation of the block of about 20 r-proteins in the L3 and L4 operons across almost the entire eubacterial and archaebacterial domains is remarkable. This strongly

Table 2 Organization of hypothetical r-protein containing operons in 16 organisms

	Organisms in eubacteria											Organisms in archaebacteria			
Gene name	*A. aeo*	*B. bur*	*T. pal*	*C. tra*	*E. coli*	*H. inf*	*H. pyl*	*M. gen*	*M. pneu*	*B. sub*	*M. tub*	*S. pcc*	*M. jan*	*M. the*	*A. ful*	*P. hor*
rpL28					√	√					√					
rpL33	√	√	√		√	√	√			√	√	√				
secE	—	√	√	√	√	√	—			√	√	√				
nusG	√	√	√	√	√	√	√			√	√	√		√		
rpL11	√	√	√	√	√	√	√	√	√	√	√	√		√		
rpL1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL10	√	√	√	√	√	√	√	pth	pth	√	√	√	√	√	√	√
rpL12	√	√	√	√	√	√	√	yacA	yacA	√	√	√	√	√	√	√
ybxB	—	—	—	—	—	—	—	nadph	nadph	√			rpoA	rpoA	rpoA	rpoA
rpoB	√	√	√	√	√	√	√	lgt	lgt	√	√		L30E	L30E	L30E	L30E
rpoC	√	√	√	√	√	√	—			√	√		nusA	nusA	nusA	nusA
rpS12	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpS7	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
fusG	—	—	—	√	√	√	√	√	√	√	—	√	—	√	√	
rpS6	√	√	√	√	√	√	√	√	√	√	√					
ssb	√	√	√	—	priB	priB	√	√	√	√	√					
rpS18	√	√	√	√	√	√	√	√	√	√	√					
rpL9	—	√	√	√	√	√	—	√	√		√			√		
rpS10	√	√	√	√	√	√	√	√	√	√	√	√		√	√	√
rpL3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL4	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL23	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	—
rpL2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpS19	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL22	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpS3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL16	√	√	√	√	√	√	√	√	√	√	√	√	—	—	—	—
rpL29	—	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpSUI1	—	—	—	—	—	—	—	—	—	—	—	—	√	√	—	—
hpy	—	—	—	—	—	—	—	—	—	—	—	—	√	√	√	√
rpS17	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√

(*To be continued on the next page*)

(*Continued*)

Gene name	Organisms in eubacteria											Organisms in archaebacteria			
	A. aeo	*B. bur*	*T. pal*	*C. tra*	*E. coli*	*H. inf*	*H. pyl*	*M. gen*	*M. pneu*	*B. sub*	*M. tub*	*S. pcc*	*M. jan*	*M. the*	*A. ful*	*P. hor*
rpL14	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL24	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpS4E	—	—	—	—	—	—	—	—	—	—	—	—	√	√	√	√
rpL5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpS14	√	√	√	—	√	√	√	√	√	√	√	—	—	√	√	—
rpS8	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL6	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL32E	—	—	—	—	—	—	—	—	—	—	—	—	√	√	√	√
rpL19E	—	—	—	—	—	—	—	—	—	—	—	—	√	√	√	√
rpL18	√	√	√	√	√	√	√	√	√	√	√	√	√	—	√	√
rpS5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL30	√	√	—	—	√	√	—	—	—	√	√	—	√	√	√	√
rpL15	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
secY	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
adk	√	—	—	—	—	—	—	√	√	√	√	√	√	√		√
map	√	—	—	—	—	—	√	√	√	√	√	—				
infA	√	—	—	—	—	—	√	√	√	√	√	√				
rpL36	√	√	√	—	√	√	√	√	√	√	√	√		√→3		
rpS13	√	√	√	√	√	√	√	√	√	√	√	√	√→S4E	√→S4E	√→S4E	S4E→1
rpS11	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpS4	√	—	—	—	√	√	√	—	—	—	√	—	—	—	—	—
rpoA	√	√	√	√	√	√	√	√	√	√	√	√	rpoD	rpoD	rpoD	rpoD
rpL17	√	√	√	√	√	√	√	√	√	√	√	√	L18E	√	L18E	√
truA	—	—	—	—	—	—	—	—	—	—	—	√	—	—	—	—
rpL13	√	√	√	√	√	√	√	√	√	√	√	√	√	—	√	√
rpS9	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
rpL31	—	—	—	—	—	—	—	—	—	—	—	√	—	—	—	—
infC		√	√	√	√	√	√	√	√	√	√	radC				
rpL35		√	√	√	√	√	√	√	√	√	√	√				
rpL20		√	√	√	√	√	√	√	√	√	√	√				
rpL21		√	√	√	√	√	√	√	√	√	√	√				
ysxB		√	√	—	—	—	—	√	√	√	—	—				
rpL27		√	√	√	√	√	√	√	√	√	√	√				

(*To be continued on the next page*)

(*Continued*)

	Organisms in eubacteria												Organisms in archaebacteria			
Gene name	*A. aeo*	*B. bur*	*T. pal*	*C. tra*	*E. coli*	*H. inf*	*H. pyl*	*M. gen*	*M. pneu*	*B. sub*	*M. tub*	*S. pcc*	*M. jan*	*M. the*	*A. ful*	*P. hor*
rpS15							√	√								
ATP-							√	√								
rpL28							√	√								
rpS16		√	√	√	√	√	√	√	√	√	√	√				
ylgC		√	√	—	—	—	√	—	—	√	√	—				
ylgE		√	√	—	√	√	√	—	—	√→1	√	—				
trmD		√	√	√	√	√	√	√	√	√	√	√				
rpL19		√	√	√	√	√	√	√	√	√	√→1	√				
rpL18E													√	√	√	√→1
reL13P													√	—	√	√
rpS9P													√	√	√	√
rpL24A													√	√→1	√	
rpL11													√	√	√	
rpS27A													√	√	√	
rpS24													√	√	√	
rpL34E													√	√		√
cmk													√	√		√
rpL14E													√	√		—
rpL31E													—	√	√	
rpL39E													√	√	√	
hyp													√	√	—	
hyp													√	√	√	
rpS19E													√	√	√	
rpL24E													√	√	√	—
rpS28E													√	√	√	√
rpL7AE													√	√	√	√
rpL44													√		√	
rpS27													√		√	
rpS3AE													√		√	
hyp													√		—	
rpS2P													√		—	
rpL15E													√		√	

Gene names were shown in the leftmost column, rp * indicated ribosomal protein. 16 organisms were arranged according to the phylogenetic relationship from left to right. √, gene matched; –, a deletion; →, insertion, number after it was the number of inserted genes; hyp, hypothetical protein, and double-lined break, the operon boundary, in rpS13 row, S4E in the column of organisms in archaebacteria meant ribosomal protein S4E inserted there.

supports that eubacteria and archaebacteria originated from one common ancestry.

3.2 Divergence of archaea and eubacteria is also evident

Also from table 2, it is clear that gene ordering in each operon is relatively conservative in each domain. For example, L16 is present in the L3 operon in every eubacteria genome but is missing from the corresponding operon in every archaeon genome. Conversely, L18E and L24E are present in the L3/L14 operon in each archaeon genome but missing in the eubacteria genomes.

3.3 Some variation in some operons across different organisms within each domain is informative on the evolutionary relations

Although the r-protein operons are highly conserved within domain, there are fine variations in some operons across different organisms within each domain, and these variations are informative on the evolutionary relations among the organisms. The followings are some examples:

i) The two mycoplasma *M. gen* and *M. pneu* have exactly the same structure for each of the operons in table 2, including *pth*, *yacA*, *nadph* and *lgt* four genes between L11 and S12 operons. While in other eubacteria *B. bur*, *T. pal*, *H. pyl*, *B. sub*, and *M. tub*, these four genes are substituted by L10 operon.

ii) *E. coli* and *H. inf* have exactly the same structure for the conserved r-protein operons. In particular, both have a substitution of *priB* for the *ssb* gene that is common to all other eubacteria in the S6 operon. Furthermore, there is complete concordance as regarding the presence/absence of the conserved hypothetical genes *ysbB*, *ylqE*, *ylqC* in the L21 and S16 operons. Thus from this table there is little doubt that these two bacteria are more closely related to each other than to the other organisms in the table.

iii) The structure of r-protein operons in the hydrogenobacterium *A. aeo* is markedly different from the other eubacteria. First of all, the r-protein genes L35 and L20, which exist in all other eubacteria genomes, are dispersed in the *A. aeo* genome. It is the same for the genes in the L21 operon and the S16 operon. Secondly, the two large operons L3 and L14 are either merged or adjacent to each other in all eubacteria genomes except in the *A. aeo* genome where they are more than 500 kb apart from each other. Interestingly, these two operons are also adjacent to each other in all 4 archaebacteria genomes. This is consistent with the assumption that Aquifex is representative of the deepest branching extant bacterial lineage.

iv) While *P. hor* clearly shares a common r-protein organization with three other archaebacteria, it lacks the L24A, L27A, L39E and L24E operons that are present in the other three genomes. This suggests that it may represent a deeper branch of the archea domain.

3.4 Operons are often clustered or merged in the genome and these clusters are often conserved

The followings are some examples:

i) In *A. aeo*, *B. bur*, *T. pal*, and *C. tra*, the L11 and L10 operons are merged and they are right next to the S12 operon.

ii) Also, the L3, L14 operons are merged in *B. bur*, *T. pal*, *C. tra*, *H. pyl*, *M. gen*, *M. pneu*,

B.sub, *M. tub*, and *S. pcc*.

iii) S13 operon is merged with L3 and L4 operons in other 10 eubacteria except *E. coli* and *H.inf*.

iv) In *B. sub*, 8 operons L33, L11, L10, S12, L3, L14, S13 and L13 are all clustered in a 47 kb region in the genome, and the ordering of these operons within the cluster is the same as that in *B. bur*.

v) In four genomes of archaebacteria, two large operons L3 and L14 also merged.

Finally, we note that this organization of r-protein genes does not seem to be present in eukaryotic genomes. In examining the completed genomes of *S. cerevisae* and *C. elegan*, we found that the r-proteins are dispersed or show only a minimal degree of clustering. This is not surprising as it is well known that genome organization in eukaryotes is very different from that in prokaryotes.

4 Discussions

The 16 bacterial organisms exhibit various degrees of relatedness. For example, the two mycoplasma *M. gen* and *M. pneu* are very closely related, in addition, *H. inf* is known to be closely related to *E. coli*. On the other hand, eubaceteria and archaebacteria have diverged very early on, and within the eubacteria, species such as *A. aeo*, *E. coli* and *S. pecc* are also very distantly related. These 16 organisms cover many branches of microbial life. Together they span a vast range of evolutionary distances. Thus it is interesting to see whether their evolutionary relations are reflected in the r-protein gene clusters.

4.1 Providing a new possibility for classifying different domains

Based on 16S ribosomal RNA, Woese divided the whole life into three domains[19], that is, dividing prokaryote into eubacteria and archae-bacteria. From table 2, it is clear that the organization of r-protein gene clusters is much more similar for two organisms within the same domain of life as compared to that across different domains. Some eubacteria operons, such as the L35, L21, S16 operons, seem to have no counterparts in archaebacterial genomes. Therefore, this method provides a new possibility for classifying different domains.

4.2 Informative on evolutionary relationship

The organization of the r-protein operons contains useful information for the inference of phylogentic relation for organisms as closely related as *M. gen* and *M. pneu*, and as distantly related as *M. gen* and *A. aeo*. Since this information is not based on detailed alignment/comparison, the sequence of a highly conserved gene, it can serve as a new method for studying molecular evolution.

It has long been known from studies of *E. coli* that some ribosomal proteins genes are grouped into operons with other related genes such as *EF-G* and *rpo*. It is also known that some of these operons are clustered in the genome, and that a number of the r-proteins are involved in the

translation feedback regulation of their own synthesis[20]. It is therefore expected that some of these r-protein operons are conserved. The present analysis documents the degree of conservation of these operons. Using 12 complete eubacteria genomes covering many phyla of eubacteria, we identified all conserved r-protein operons. The conservation of the block of about 20 r-proteins in the L3 and L14 operons across almost the entire eubacterial and archaebacterial domains is remarkable. What is the reason for this conservation? At this time there is no clear explanation but the phenomenon may be a hint of long range co-regulation involving two or more operons.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 39392900, 39830070 & 19890380).

References

1. Koonin, E. V., Big time for small genomes, Genome Research, 1997, 7: 418.
2. Deckert, G., Warren, P. V., Gaasterland, T. et al., The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*, Nature, 1998, 392: 353.
3. Fraser, C. M., Casjens, S., Huang, W. M. et al., Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*, Nature, 1997, 390: 580.
4. Fraser, C. M., Norris, S. J., Weinstock, G. M. et al., Complete genome sequence of *Treponema pallidum*, the syphilis spirochete, Science, 1998, 281: 375.
5. Blattner, F. R., Plunkett III G., Bloch C. A. et al., The complete genome sequence of *Escherichia coli* K-12, Science, 1997, 277: 1453.
6. Fleischmann, R. D., Adams M. D., White, O. et al., Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, Science, 1995, 269: 496.
7. Tomb, J. F., White, O., Kerlavage, A. R. et al., The complete genome sequence of the gastric pathogen *Helicobacter pylori*, Nature, 1997, 388: 539.
8. Fraser, C. M., Gocayne, J. D., White, O. et al., The minimal gene complement of *Mycoplasma genitalium*, Science, 1995, 270: 397.
9. Himmelreich, R., Hilbert, H., Plagens, H. et al., Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*, Nucleic. Acids Res., 1996, 24: 4420.
10. Kunst, F., Ogasawara, N., Moszer, I. et al., The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, Nature, 1997, 390: 249.
11. Cole, S.T., Brosch, R., Parkhill, J. et al., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature, 1998, 393: 537.
12. Kaneko, T., Sato, S., Kotani, H. et al., Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential-coding regions, DNA Res., 1996, 3: 109.
13. Bult, C. J., White, O., Olsen, G. J. et al., Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*, Science, 1996, 273: 1058.
14. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C. et al., Complete genome sequence of methanobacterium thermoautotrophicum delta H: functional analysis and comparative genomics, J. Bacteriol., 1997, 179: 7135.
15. Klenk, H. P., Clayton, R. A., Tomb, J. et al., The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, Nature, 1997, 390: 364.
16. Kawarabayasi, Y., Sawada, M., Horikawa, H. et al., Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3, J. DNA Res., 1998, 5: 147.
17. Pearon, W. R., Rapid and sensitive sequence comparison with FASTP and FASTA, Methods Enzymol., 1990, 183: 63.
18. Sander, C., Schneider R., Database of homology-derived protein structures and the structural meaning of sequence alignment, Proteins, 1991, 9: 56.
19. Woese, C. R., *Archaebacteria*, Sci. Amer., 1981, 244: 98.
20. Nomura, M., Gourse, R., Bauthman, G., Regulation of the synthesis of ribosomes and ribosomal components, Ann. Rev. Biochem., 1984, 53: 75.