Hypothesis

The emergence of major cellular processes in evolution

Christos Ouzounis[a,*], Nikos Kyrpides[b,**]

[a] *AI Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493, USA*
[b] *IMBB, FORTH, Heraklion, Greece*

Received 29 April 1996; revised version received 4 June 1996

Abstract The phylogenetic distribution of divergently related protein families into the three domains of life (archaea, bacteria and eukaryotes) can signify the presence or absence of entire cellular processes in these domains and their ancestors. We can thus study the emergence of the major transitions during cellular evolution, and resolve some of the controversies surrounding the evolutionary status of archaea and the origins of the eukaryotic cell. In view of the ongoing projects that sequence the complete genomes of several Archaea, this work forms a testable prediction when the genome sequences become available. Using the presence of the protein families as taxonomic traits, and linking them to biochemical pathways, we are able to reason about the presence of the corresponding cellular processes in the last universal ancestor of contemporary cells. The analysis shows that metabolism was already a complex network of reactions which included amino acid, nucleotide, fatty acid, sugar and coenzyme metabolism. In addition, genetic processes such as translation are conserved and close to the original form. However, other processes such as DNA replication and repair or transcription are exceptional and seem to be associated with the structural changes that drove eukaryotes and bacteria away from their common ancestor. There are two major hypotheses in the present work: first, that archaea are probably closer to the last universal ancestor than any other extant life form, and second, that the major cellular processes were in place before the major splitting. The last universal ancestor had metabolism and translation very similar to the contemporary ones, while having an operonic genome organization and archaean-like transcription. Evidently, all cells today contain remnants of the primordial genome of the last universal ancestor.

Key words: Cellular process; Evolution; Last universal ancestor

1. Introduction

From the ongoing genome projects on model organisms, which span all major domains of the tree of life, we can now examine the phylogenetic distribution of a collection of literally thousands of protein families. Some of them are present universally, some are thought to be present in two domains only, and finally some seem to be unique to a particular domain. The mere presence or absence of a protein family can be indicative of the existence of an entire cellular process and consequently the functional potential of a cellular type as well as the origins of biochemical and genetic processes during evolution. Protein families, being universally present or not, allow the derivation of important conclusions about the origins of molecular functions during cellular evolution.

2. Protein sets as phylogenetic traits

Here, the PROSITE pattern sets [1] are presented as a random sample of protein families with known function. Although the data is far from complete, and given the biases introduced in this collection, this is the first time that this database of protein families is studied from a phylogenetic viewpoint. We will combine these data to support predictions about cellular evolution from previous work. The enormous amount of information and its classification imposes the establishment of a new nomenclature. Therefore, to present the protein families that are distinct in various domains of life, we need a new classification scheme. The scheme can be used as a predictive framework for the presence or absence of a protein family, and contributes towards a better understanding of cellular evolution.

We propose the following simple names: archaean, bacterial and eukaryotic proteins represent those that are currently known to be exclusively present in those domains. Families that are present in all domains are defined as universal. Proteins common to two but absent from the third domain acquire the prefix 'un-' for that domain. Thus, general transcription factors TFIIB, TFIID and TFIIS common in archaea and eukaryotes [2] can be called un-bacterial, transcriptional regulator YB1 from bacteria and eukaryotes [3] can be called un-archaean, and finally N-4 cytosine-specific DNA methylases from bacteria and archaea [4] can be called un-eukaryotic. It should be noted that until the full genome of at least some model organisms from each domain is known, any such definition can only be tentative, and the presence of a single member of a particular protein family in one domain can change its definition. For example, the eukaryotic transcription factor TFIIB was found to be un-bacterial [5]. However, from the protein families that are *already* known to be universal, we can draw important conclusions about the last universal ancestor.

3. The common intersection of the three domains

We have chosen to examine the occurrence of PROSITE protein families in each domain, in order to gain insights on cellular evolution. Even at this high level of observation (ignoring, for example, sequence similarity relationships and the implied dendrograms), and despite some approximations

*Corresponding author. Fax: (1) (415) 8593735.
E-mail: ouzounis@ai.sri.com

**Present address:* Department of Microbiology, University of Illinois at Urbana-Champaign, 131 Burrill Hall, 407 South Goodwin Avenue, Urbana, IL 61801, USA.

PII S0014-5793(96)00631-X

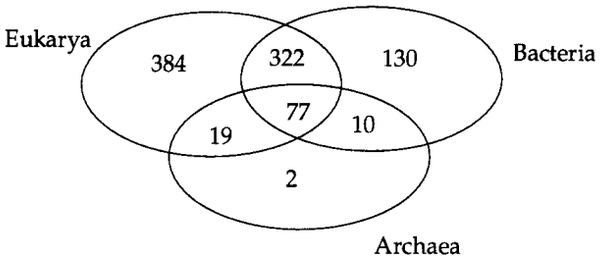

Fig. 1. The three sets of 944 protein pattern families and their distribution. Numbers represent counts of actual patterns from PROSITE and may be slightly more than the corresponding proteins, as more than one pattern can be used for the description of a family. The acute lack of data from archaea is evident. The classification of non-universal families can only be provisional; however, the universal families may have been present in the last universal ancestor. The full listing of the patterns according to this split is available on the World Wide Web at the following URL: http://www.ai.sri.com/~ouzounis/prosplit.html. The PROSITE release 13.0 has been used. The database was parsed and the TAX-RANGE fields for archaea, eukaryotes and bacteria were counted, using programs written in Common Lisp. For detailed information about the database, refer to PROSITE home page at the following URL: http://expasy.hcuge.ch/sprot/prosite.html.

(viruses and phages not considered, plastid families are classified as eukaryotic although of bacterial origin, some families known but not listed), certain domain relationships are clearly emerging (Fig. 1). For example, there are 384 eukaryotic protein families, 130 bacterial and only 2 archaean. The latter number is small compared to bacterial proteins because there has not been a major public release of a significantly large portion of the genome of an archaean species.

While the un-archaean families amount to 322, there are 19 un-bacterial families and 10 un-eukaryotic ones. This seems to support the notion that archaea may be closer to eukaryotes than to bacteria [6,7]. Yet, with findings such as the presence of the transcriptional activator Lrp in an archaean genome, a typical bacterial protein involved in transcription which is believed to be eukaryotic-like in archaea [8] point to the possibility that archaea may not be definitely closer to a particular domain, but rather of mosaic nature [9].

At the other extreme, the 77 universal protein families are also of a small number, and of a very unique nature: they are mostly either metabolic enzymes participating in various metabolic pathways (Table 1) or translation-associated proteins, such as ribosomal proteins and amino-acyl tRNA synthetases. We will discuss below what are the implications of the universal families for molecular evolution.

4. The last universal ancestor

Some principles that arise from the present status of sequence analysis of molecular data and its classification, can shed light to the evolution of the major cellular types. Archaea, with an apparent genome structure very much like bacteria [10], but with content that bears resemblance to both eukaryotic and bacterial genomes [9], seem to be an intermediate form. Consequently, archaea seem to hold the secrets for the origin of eukaryotic cells and, until a genome of a primitive eukaryote becomes available [11], they constitute the best model for the study of eukaryotic life forms. It is not clear, for instance, whether operonic organization is common and widespread in primitive eukaryotes [12]. The elements common to all domains are mainly metabolic enzymes and translation-associated proteins, pointing to the existence of these families in the last universal ancestor, before the split of the three domains. We can thus state that two of the first major classes of cellular processes were precursor biosynthesis and transfer of information from RNA to proteins. A relationship between these two processes may still exist as a relic of a primordial state – this issue has been addressed elsewhere [13–15].

Intelligent systems that represent metabolic processes [16] such as EcoCyc [17,18] have been developed, and can be used to retrieve the pathways in which these enzymes are involved. The metabolic enzymes that are universally present participate in major biochemical pathways and are generally very conserved and easily identifiable (Table 1). From the mere presence of these enzymes, we can then predict that the full set of their counterparts participating in the same pathways, may be present, at the domain level (Table 1). For example, although three glycolytic enzymes are known to be universal, the prediction would be that, at least at the domain level, the other seven enzymes will be found later. Such a case is pyruvate kinase, which has been identified in archaea [19], but its sequence is incomplete and thus absent from PROSITE. Species exceptions, of course, can occur: for example, while the TCA cycle seems to be a very ancient pathway, three of its enzymes have not been found in the first completely sequenced genome of a free-living organism, *Haemophilus influenzae* [20,21]. Using EcoCyc, we can perform exhaustive analyses to find out which pathways are absent in a given organism with a completely sequenced genome [21].

The sharing of transcription factors with eukaryotes only points out to the clear possibility that translation could have actually preceded transcription [22]. If this hypothesis was true, and additional evidence comes to support the existence of translation before transcription, then it would even be possible to narrow down the time of the transition from the RNA to the DNA world. However, given the sharing of some major components between all domains, and in particular the similarities between DNA-dependent RNA polymerases [7], the hypothesis is not supported. Being of common origin, transcription in the three domains followed independent ways, with bacteria being most unique, while archaea and eukaryotes having more common components.

5. Contemporary genomes: organization and transcription

Thus, we are confronted with the puzzling observation that archaea have eukaryotic-like transcription machinery, while they operate through a bacterial-like genomic organization. Here we pose the following question: *why is genome organization independent from transcription*? To rephrase the question, we can ask: why do bacteria have a unique transcription machinery? Or, in other words, why is transcription (as exemplified by the corresponding protein families) *not* universal?

A possible answer is that transcription originated after the splitting of the major domains, which may also mean that the DNA world was devised concurrently at the time of the splitting. However, DNA photolyases [23] have been identified in archaea, pointing out that DNA synthesis was present before the deep split. Other authors, focusing on other DNA metabolizing enzymes (such as polymerases and topoisomerases)

arrived at the same conclusion [24]. Thus, the difference could only be attributed to the generation time of bacteria (faster than any domain by orders of magnitude) that made them most different from their ancestral state, relative to the other domains [25]. In that sense, the eukaryotic/archaean (un-bacterial) mode of transcription may be closest to the original common state.

In general, only basic transcription is proven different in bacteria as compared to the other domains. In addition, the general and/or specific transcription factors are in most cases different between eukaryotes and bacteria – with few exceptions, like the cold-shock domain, present in eukaryotic factor YB1 [3]. A commonly cited reason is that bacteria have genomes that are turned on most of the time so they need more repressor-type regulators and some terminators. Indeed, the majority of bacterial promoters are repressible [26]. A future challenge for systems that represent biochemical information is to extend their representations to account for macromolecular biochemistry such as transcription or translation (Collado-Vides and Karp, unpublished).

The eukaryotic mode of transcription requires more complex and expensive mechanisms that regulate the activation of particular genes that are mostly silenced and hidden (tightly packed) in huge genomes. Therefore, we can predict that the complex transcription factors of eukaryotes will not be identified in any of the other two domains, since most probably they have been invented only after (and possibly due to) the expansion of the eukaryotic ancestor.

Finally in archaea, as far as general transcription is concerned, no clear patterns can yet emerge. Archaean genomes contain eukaryotic basic transcription factors [2]. However, since archaea retained (or re-discovered) operonic organization, we can predict that at least some operon-type transcriptional regulators will be found in these genomes. The identification of an Lrp-like regulator in the genome of *Pyrococcus furiosus* supports this prediction [8].

Summarizing the above, and using our previously proposed terminology, we can make the following predictions, with regard to transcription:

1. operonic transcription factors are un-eukaryotic
2. basic transcription factors are un-bacterial
3. complex transcription factors are eukaryotic (only)

6. Transcriptional repression as a mechanism for genome evolution

Why is genome organization in eukaryotes so different? How was it possible to expand their genomes to such an extent? To address this issue, it is important to realize that a necessary condition for the development of large genomes would be efficient transcriptional repression, or what is familiarly known as packing. With this reasoning, particularly im-

Table 1
The universal *metabolic enzyme* families and patterns

PROSITE pattern	Enzyme family	Pathway	Gene	min
MDH	Malate dehydrogenase	TCA cycle, other	mdh	72.78
GAPDH	G3PD dehydrogenase C	Glycolysis	gapC	31.8
GLD_DEHYDROGENASE	NADP glutamate dehydrogenase	Amino acid interconversion	gdhA	39.62
P5CR	Pyrroline-5-carboxylate reductase	Proline biosynthesis	proC	8.795
COX1	Cytochrome O ubiquinol oxidase I	Electron transport	cyoB	9.736
SOD_MN	Superoxide dismutase (Mn)	Oxygen protection	sodA	88.38
	Superoxide dismutase (Fe)	Oxygen protection	sodB	37.32
SUMT_1/SUMT_2	Siroheme synthase subunit?	Siroheme biosynthesis	CysG	75.26
GATASE_TYPE_1	Glutamine amidotransferase component II	Folate biosynthesis	pabA	75.09
	Carbamoyl-P synthase small chain	Arginine/Pyrimidine biosynthesis	carA	0.642
	GMP synthase	Purine biosynthesis	guaA	56.45
	CTP synthase	Pyrimidine biosynthesis	pyrG	62.6
	HisH amidotransferase	Histidine biosynthesis	hisH	45.05
	Anthranilate synthase component II	Tryptophan biosynthesis	trpD	28.38
AA_TRANSFER_CLASS_I	Aromatic amino acid aminotransferase	Phe, Tyr and Leu biosynthesis	tyrB	91.92
	Aspartate aminotransferase	Amino acid interconversion	aspC	21.2
PGLYCERATE_KINASE	Phosphoglycerate kinase	Glycolysis	pgd	66.06
ADENYLATE_KINASE	Adenylate kinase	Nucleotide metabolism	adk	10.76
ATPASE_ALPHA_BETA	ATP synthase α	ATP synthesis	atpA	84.41
	ATP synthase β	ATP synthesis	atpD	84.36
IGPS	Indole-3-glycerol phosphate synthase	Tryptophan biosynthesis	trpC	28.35
CITRATE_SYNTHASE	Citrate synthase	TCA cycle	gltA	16.32
ENOLASE	Enolase	Glycolysis	eno	62.58
TRP_SYNTHASE_ALPHA	Tryptophan synthase α	Tryptophan biosynthesis	trpA	28.3
TRP_SYNTHASE_BETA	Tryptophan synthase β	Tryptophan biosynthesis	trpB	28.32
D_ALA_DEHYDRATASE	Delta-aminolevulinic acid dehydratase	Porphyrin biosynthesis	hemB	8.46
PGM_PMM	Phosphoglucomutase	Sugar metabolism	pgm	15.46
	Phosphomannomutase	GDP-mannose biosynthesis	cpsG	unknown
GLNA_1/GLNA_ATP	Glutamine synthetase	Amino acid interconversion	glnA	87.42
ARGININOSUCCIN_SYN	Argininosuccinate synthase	Urea cycle, Arginine biosynthesis	argG	71.39
CPSASE_1/2	Carbamoyl-P synthase large chain	Arginine/Pyrimidine biosynthesis	carB	0.667
	Biotin carboxylase	Long-chain fatty acid synthesis	accC	73.27

The PROSITE identifier is followed by description, the metabolic pathway, gene name in *E. coli* and finally the location (in minutes) on the *E. coli* chromosome, where available; the table is approximately sorted by PROSITE pattern numbering. Information compiled from the EcoCyc knowledge base [17,18], during our efforts to include phylogenetic data for the enzymes participating in central metabolic pathways. The knowledge base is accessible on the World Wide Web at the following URL: http://www.ai.sri.com/ecocyc/browser.html.

portant was the discovery that the core histone fold is found not only in eukaryotic histones, but also in two subunits from the transcription factor CBF (A and C) while at the same time it seems to have been present in Archaea, before the origins of eukaryotes [27]. These findings suggest that nucleosomal packing is an artifact of extreme transcriptional repression and its origins lie in archaea and possibly in transcription regulation [27].

Therefore, it is possible that eukaryotes could expand their genomes because they discovered a system that could keep them silent. If eukaryotic packing indeed originates from transcription, then the family of archaean histone-like proteins [28] were the prime means for the creation of large genomes. This 'genome silencing' process in the dawn of the eukaryotic cell would be accompanied by novel families of complex transcription factors, whose sole function would be to identify the corresponding promoters and tag them for transcription activation.

An additional feature that allowed the increase of complexity of eukaryotic genomes was the capacity of performing simultaneous initiation of DNA replication at many origins, as opposed to bacteria, which possess a single origin of replication [29]. It will be of particular interest to decipher the corresponding mechanism in archaea in the near future.

7. Archaea: mosaic genomes?

Archaea display a mosaic pattern of form and function that stems from their ancient nature and not from their intermediate character [25]. There is a number of important arguments which support this (not widely held) view.

First, genome organization in archaea resembles the simple form of bacteria, with small, compact, circular genomes and operonic clusters [10,30,31]. The content of the genomes, however, seems to contain some proteins that are common to eukaryotic ones and are absent from bacteria [9,31].

Second, un-archaean families are many, and mostly arise from the lack of data for archaean protein families, and their relationships to the other two domains. Once a complete genome from archaea becomes available, this number is expected to diminish significantly (Fig. 1). On the other hand, archaean families come from uniqueness, since data from bacterial and eukaryotic genomes are abundant. This uniqueness is an important issue and should be emphasized as much as their mosaicity [25] as it argues for the ancestral nature of this domain: the less unique families are found when an archaean genome is complete, the closer archaea will be to the last universal ancestor.

Third, as far as the other categories are concerned, the question of mosaicity will be resolved by the number of un-eukaryotic families that will persist, after the completion of the first eukaryotic genome (e.g. yeast). The more un-eukaryotic families remain, after compilation of the proteins from a complete archaean genome, the closer archaea will come to bacteria. Most un-eukaryotic and all un-bacterial families, however, can be checked whether they are genuine elements of archaean genomes, and not due to lack of data. For example, archaea have restriction endonucleases [4], an un-eukaryotic family, which may never be found to be universal. Also, many un-bacterial families are known to exist in archaea; for example eukaryotic histones [27], absent from the genome of *Haemophilus influenzae* (unpublished observations).

Finally, the most striking pattern of mosaicity in archaea comes from the proteins involved in genome packing, and maybe transcription and translation. Archaea have both HU-like [32] (absent from eukaryotes) and eukaryotic histone-like proteins [28] – clearly absent from bacteria. The extent of participation of the above families in packing archaean genomes remains unclear at this point. Interestingly, both of these protein families contain members which are involved in transcription: IHF is homologous to HU proteins [33] and CBF subunits A and C are homologous to eukaryotic core histones [27]. Also, some of the components of translation in archaea are shared with bacteria or eukaryotes [25].

One valid explanation for mosaicity, therefore, seems to be the ancient nature of archaea. This domain contains components from both other domains, because it is closest to the last universal ancestor and not a later derivative. Bacterial genome organization and eukaryotic transcription are closest to the original form as they are both shared with archaea, while bacterial transcription and eukaryotic genome organization seem to be evolved forms of these processes, unique to the corresponding domains.

8. A scenario for the evolution of cellular processes

We envision that the distribution of protein families will help us reconstruct the history of the major transitions in cellular evolution, very much like ribosomal RNA served as a guiding tool for the first establishment of the phylogenetic order and relationships of the three domains in the tree of life [34].

An orderly progression of events, in a very tentative scale, and proposed here for the first time, can be as follows:

1. metabolism and translation *preceded* transcription
2. transcription preceded packing (transcriptional repression)
3. development of packing enabled the expansion of genomes
4. expansion of genomes gave rise to new transcription factors

In words, the basic transcription was retained in archaea and eukaryotes, and lost in bacteria due to fast replication, while genome organization was changed in eukaryotes due to a new packing system through archaean-like transcription (repressive system, called nucleosomal packing) thus enabling genome expansion. This event changed the genome organization for the first time and created the need for new transcription factors unique in eukaryotes to regulate this vast, expanding genome. *The last universal ancestor most probably had basic molecular components in metabolism and translation very similar to the contemporary ones, while having a operonic genome organization and archaean-like transcription.*

What is a counterintuitive fact, however, is that even in the most advanced genomes, there is a component of primitive origins, that of the universal families. It would be interesting to estimate how much these families have diverged and multiplied. If we assume a three-fold multiplication by gene duplication on average, and with an estimated primitive genome of 1 MB, then modern eukaryotes may carry remnants of a few Mb in their genomes, from their last universal ancestor. As genomes expanded, this component becomes a small percentage of the total genome, and could be used as an indication

of the evolutionary status of a taxon. This fraction contains some of the most important proteins that are unique for life, major metabolic enzymes and translation-associated proteins. All additional components were added to regulate this prime set of molecules that are most essential for life. Finally, it is apparent that the last universal ancestor was a fairly complex entity, and that molecular evidence must be supplemented by other biochemical and geochemical data to advance our knowledge back in time.

Acknowledgements: We acknowledge the effort by Amos Bairoch, his colleagues and the PROSITE curators to compile this valuable resource of patterns and protein families. Also, the efforts of the sequencing communities that provide us with such a wealth of data. Communication between the authors was facilitated through Internet. Comments and criticisms by Amos Bairoch (University of Geneva, Switzerland), A. Economou (Dartmouth College Medical School, NH, USA), Elsa Maniataki (IMBB, Greece) and particularly Peter Karp (SRI International, CA, USA) are gratefully acknowledged. C.O. is a recipient of a fellowship from the Human Frontiers Science Program Organization.

References

[1] Bairoch, A. and Bucher, P. (1994) Nucleic Acids Res. 22, 3583–3589.
[2] Klenk, H.-P. and Doolittle, W.F. (1994) Curr. Biol. 4, 920–922.
[3] Wolffe, A.P. (1994) BioEssays 16, 245–251.
[4] Nolling, J., van Eeden, F.J., Eggen, R.I. and de Vos, W.M. (1992) Nucleic Acids Res. 20, 6501–6507.
[5] Ouzounis, C. and Sander, C. (1992) Cell 71, 189–190.
[6] Keeling, P.J. and Doolittle, W.F. (1995) Proc. Natl. Acad. Sci. USA 92, 5761–5764.
[7] Langer, D., Hain, J., Thuriaux, P. and Zillig, W. (1995) Proc. Natl. Acad. Sci. USA 92, 5768–5772.
[8] Kyrpides, N.C. and Ouzounis, C.A. (1995) Trends Biochem. Sci. 20, 140–141.
[9] Ouzounis, C.A., Kyrpides, N.C. and Sander, C. (1995) Nucleic Acids Res. 23, 565–570.
[10] Woese, C.R. (1987) Microbiol. Rev. 51, 221–271.
[11] Sogin, M.L. (1991) Curr. Opin. Genet. Dev. 1, 457–463.
[12] Zorio, D.A.R., Cheng, N.N., Blumenthal, T. and Spieth, J. (1994) Nature 372, 270–273.
[13] Kyrpides, N.C. and Ouzounis, C.A. (1993) J. Theor. Biol. 163, 373–392.
[14] Ouzounis, C.A. and Kyrpides, N.C. (1994) J. Theor. Biol. 167, 373–380.
[15] Kyrpides, N.C. and Ouzounis, C.A. (1995) J. Mol. Evol. 40, 564–569.
[16] Karp, P. and Mavrovouniotis, M. (1994) IEEE Expert 9, 11–21.
[17] Karp, P. and Riley, M. (1993) In: First International Conference on Intelligent Systems for Molecular Biology (Hunter, L., Searls, D. and Shavlik, J., Eds.), pp. 207–215. AAAI Press, Bethesda, MD.
[18] Karp, P., Riley, M., Paley, S. and Pellegrini-Toole, A. (1996) Nucleic Acids Res. 24, 32–46.
[19] Potter, S. and Fothergill-Gilmore, L.A. (1992) FEMS Microbiol. Lett .73, 235–239.
[20] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995) Science 269, 496–512.
[21] Casari, G., Andrade, M.A., Bork, P., Boyle, J., Daruvar, A., Ouzounis, C., Schneider, R., Tamames, J., Valencia, A. and Sander, C. (1995) Nature 376, 647–648.
[22] Karp, P., Ouzounis, C. and Paley, S. (1996) In: Fourth International Conference on Intelligent Systems for Molecular Biology (States, D.J., Agarwal, P., Gaasterland, T., Hunter, L. and Smith, R.F., Eds.) AAAI Press, St. Louis, MO.
[23] Sogin, M.L. (1994) In: Early Life on Earth, Nobel Symposium No. 84 (Bengtson, S., Ed.) pp. 181–192. Columbia University Press, New York.
[24] Tako, M., Kobayashi, T., Oikawa, A. and Yasui, A. (1989) J. Bacteriol. 171, 6323–6329.
[25] Forterre, P., Benanchenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. and Labedan, B. (1992) BioSystems 28, 15–32.
[26] Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Pühler, G. and Klenk, H.-P. (1988) Eur. J. Biochem. 173, 473–482.
[27] Collado-Vides, J., Magasanik, B. and Gralla, J.D. (1991) Microbiol. Rev. 55, 371–394.
[28] Ouzounis, C.A. and Kyrpides, N.C. (1996) J. Mol. Evol. 42, 234–239.
[29] Sandman, K., Perler, F.B. and Reeve, J.N. (1994) Gene 150, 207–208.
[30] Szathmáry, E. and Maynard Smith, J. (1995) Nature 374, 227–232.
[31] Brown, J.R., Daniels, C.J. and Reeve, J.N. (1989) CRC Crit. Rev. Microbiol. 16, 287–338.
[32] Keeling, P.J., Charlebois, R.L. and Doolittle, W.F. (1994) Curr. Opin. Genet. Dev. 4, 816–822.
[33] DeLange, R.J., Williams, L.C. and Searcy, D.G. (1981) J. Biol. Chem. 256, 905–911.
[34] Schmid, M.B. Cell (1990) 63, 451–453.
[35] Olsen, G.J. and Woese, C.R. (1993) FASEB J. 7, 113–123.