Archaebacterial genomes: eubacterial form and eukaryotic content

Patrick J Keeling, Robert L Charlebois and W Ford Doolittle

Dalhousie University, Halifax and University of Ottawa, Ottawa, Canada

Since the recognition of the uniqueness and coherence of the archaebacteria (sometimes called Archaea), our perception of their role in early evolution has been modified repeatedly. The deluge of sequence data and rapidly improving molecular systematic methods have combined with a better understanding of archaebacterial molecular biology to describe a group that in some ways appears to be very similar to the eubacteria, though in others is more like the eukaryotes. The structure and contents of archaebacterial genomes are examined here, with an eye to their meaning in terms of the evolution of cell structure and function.

Current Opinion in Genetics and Development 1994, 4:816–822

Introduction

As molecular phylogenetics advances, it provides evolutionary biologists with a framework on which to assemble a coherent picture of the history of life. Two unexpected discoveries that emerged from sequence comparisons have had particularly far-reaching effects on our view of early evolution; both concern the archaebacteria. The first, on the basis of ribosomal RNA (rRNA) sequence information, and now extensively supported by a variety of other data, was the finding that the prokaryotes are actually composed of two distinct groups, eubacteria and archaebacteria, as distant from one another as either are from the eukaryotes [1]. The second, equally exciting discovery is that the eubacteria branched first from the universal tree, so that the archaebacteria and eukaryotes seem to be sister groups [2]. The significance of this second conclusion (which, although increasingly popular, does need confirmation from additional data sets) is that it allows us to propose specific arguments about the nature of both the prokaryotic ancestor of the eukaryotes and the last common ancestor of all extant life on the basis of characters shared between pairs of taxa.

As our understanding of the archaebacteria on the molecular level develops, it is becoming clear that they share a number of features in overall genome design and gene organization with the eubacteria. These traits can then be said most parsimoniously to be ancestral to all life. Many other characters, particularly those involved with gene expression, are shared between the archaebacteria and eukaryotes; these can be either primitive or derived, but in both cases help us to understand the order of evolutionary events at or near the origin of the eukaryotes. Detailed reviews of archaebacterial molecular biology are available in two recent books [3,4], so in this review we will focus on more recent discoveries and try to assimilate these observations with phylogeny.

Archaebacterial genome design

At present, seven archaebacterial genomes have been physically mapped. In each case, the genome is composed of a single circular chromosome that falls within the size range observed in the eubacteria [5–11,12••,13•] (Table 1). In several instances, extrachromosomal elements have also been identified [14]; in some halophiles, plasmid and megaplasmid (up to 700 kb) DNA makes up a substantial portion of the genome [15]. Halophile plasmids and chromosomes bear scores of transposable elements, which give rise to high frequencies of genetic instability as revealed by spontaneous mutations and local genetic rearrangements detectable through Southern analysis [16]. This phenomenon can be observed on the time scale of laboratory experiments, but comparative mapping indicates that in evolutionary time, chromosomal order is remarkably conserved in halobacteria (P López-García, A St Jean, R Amils, R Charlebois, unpublished data; N Hackett, personal communication). Forces must exist that can effectively oppose the powerful destabilizing pressures present in the halobacterial genome.

Archaebacterial and eubacterial genomes are similar in size and structure, both being small, compact, and circular. If the rooting of the universal tree by Iwabe [2] is correct, it is reasonable to suppose that this design is ancestral to all extant life, although we know too few of the details of genome construction to be certain that the trait could not have evolved twice independently. To describe the nature of the primitive genome more accurately, we

Abbreviation

snoRNA—small nucleolar RNA.

 ISSN 0959-437X

Table 1. Archaebacterial genome sizes as determined by mapping.

Species	Genome size	Method	References
Methanococcus voltae	1900	PFG	[7]
Halobacterium halobium	2400	PFG	[10]
Halobacterium sp. GRB	2472	Cloning	[12••]
Haloferax volcanii	4140	Cloning	[6]
Haloferax mediterranei	3840	PFG	[8,13•]
Thermococcus celer	1890	PFG	[5]
Sulfolobus acidocaldarius	2760	PFG	[9]

Summary of archaebacterial genomes that have been mapped. Each of these genomes is within the range of genome size documented in the eubacteria and all finished maps are circular. PFG, pulsed-field gel electrophoresis.

need more comparative data and a better understanding of archaebacterial genomes from a functional standpoint; for instance, the mechanisms of replication initiation, termination, and chromosomal segregation need to be examined.

Archaebacterial gene organization

In general, the structure of archaebacterial genes closely resembles that of eubacteria [15,17]. They are organized into co-transcribed units, or operons, that are expressed as long non-capped mRNAs with only short bacterial-like poly(A) tails [18]. Protein-coding genes of the archaebacteria are not disrupted by introns of any type, other than some spliced at the protein level [19,20], and to date no intron resembling the nuclear spliceosomal type has been identified. Introns have been found in 16S, 23S, and tRNA genes; these are of a novel type, dependent upon a defined secondary structure at the intron/exon boundary and an unidentified enzymatic activity [21]. Some of these introns have been found to contain open reading frames related to homing endonucleases first characterized in group I self-splicing introns of mitochondria and eubacteriophages [22•].

The degree of similarity between archaebacterial and eubacterial operons or operon clusters is best exemplified by a number of archaebacterial operons that contain the same genes in the same order as their eubacterial homologs (reviewed in [23]; see Fig. 1). These include the RNA polymerase operon and ribosomal protein gene clusters, which are homologous to the *Escherichia coli* L11 and L10 clusters (four of four genes in the same order in two archaebacteria and *E. coli* [24,25••]), the spectinomycin operon (for which *Methanococcus vannielii* and *Sulfolobus acidocaldarius* show eleven and nine, respectively, of the eleven *E. coli* genes in the same order, although the archaebacteria each have three 'extras', which are homologs of eukaryotic ribosomal protein genes [24,26]), the S10 operon (seven genes in a stretch of seven in the same order in halobacteria and *E. coli* [27]), and the streptomycin operon (four genes in the order S12–S7–EF-G–EF-Tu for both *M. vannielii* and *E. coli* [28]). In one case, the *M. vannielii* L1 operon, the similarity can be extended to the regulation of expression, as it has been shown that the L1 protein in *M. vannielii* represses translation of its own mRNA by binding to a site that resembles its binding site on the 23S rRNA [25••]. The same autoregulation is seen in *E. coli*, differing only in the position of the binding site on the mRNA.

Enough data may not be available to say that gene order is conserved over even longer distances in eubacterial and archaebacterial chromosomes, but enough is known to begin to assemble some intriguing correlations. The spectinomycin operon is immediately downstream of the S10 operon in both *E. coli* and *M. vannielii* [17], and the streptomycin operon is followed by the S10 operon in *Thermotoga maritima* and by the S10 protein (the first reading frame in the S10 operon) in three different archaebacteria [28], as well as in the cyanobacterium *Spirulina platensis* [29]. The current best guess is that in the ancestor of archaebacteria and eubacteria, these three operons followed one another in the order streptomycin–S10–spectinomycin, with the genes in each being similar to those found commonly in contemporary operons.

The transition to eukaryotes

A particularly close affinity between archaebacteria and eukaryotes has been long suspected, but for the most part on vague grounds — resistance to certain drugs, lack of peptidoglycan, or the occurrence of tRNA introns [30]. Recent work, however, has elucidated a handful of molecular similarities that argue unambiguously for a shared inheritance with the eukaryotes, three of which are reviewed here.

One of the strongest similarities is in transcription mechanisms, now known to differ fundamentally from the homologous eubacterial process. Evidence for this similarity has been accumulating for a number of years, beginning with the observation that archaebacterial RNA polymerases are multisubunit complexes, unlike the simple three or four subunit eubacterial enzyme [31,32], and that the consensus archaebacterial promoter bears a closer resemblance to the eukaryotic TATA box than to eubacterial promoter sequences [33]. The depth of this similarity was further revealed when the existence of an archaebacterial homolog of the eukaryotic transcription factor TFIIB was discovered by a database search [34,35•], and by two reports of a gene that encodes a protein closely resembling both TFIIS and a subunit of eukaryotic RNA polymerases I and II [36•,37•]. Moreover, work by two independent investigators has now shown that archaebacterial genomes also encode TFIID, or the TATA-binding protein [38••,39••], and that TFIID recognizes archaebacterial promoters and directs the binding of TFIIB [39••]. Unlike their eubacterial counterparts, eukaryotic RNA polymerases alone cannot recognize promoters; instead, the assembly of the

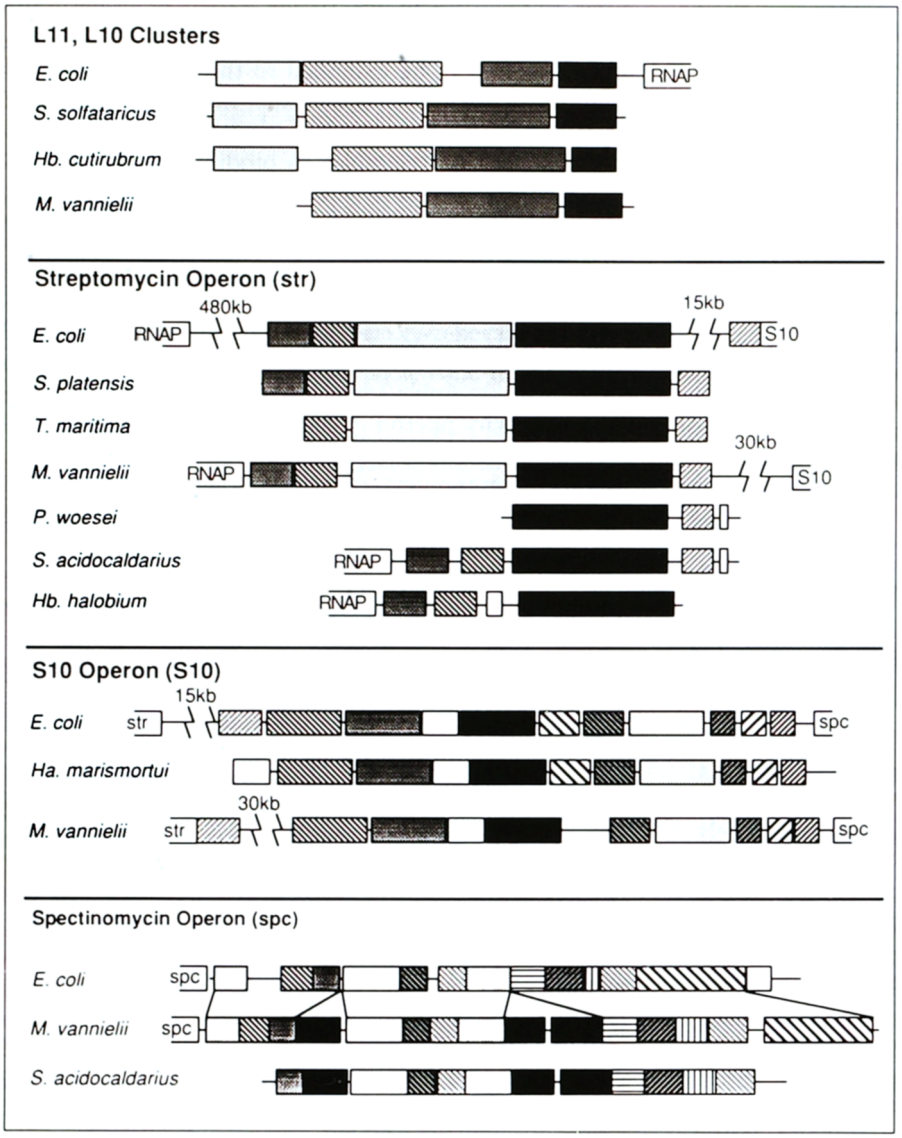

Fig. 1. Summary of conserved gene order between eubacteria and archaebacteria in four operons. Homologous genes are indicated by shading, the actual gene names have not been included for simplicity, but can be found in three recent reviews [17,23,28]. In the case of the spectinomycin operon, the archaebacterial genes with only eukaryotic homologs are shaded black. Where the position of the cluster relevant to other clusters is known, it is shown as an open box on the end, and distances are given where they are not closely linked [17,23,24,25••,26–29,57].

TFIID/TFIIB complex is responsible for efficiently directing the polymerase to the promoter site *in vivo*. It appears that this may also be the case in the archaebacteria; if so, it represents a significant divergence from eubacterial transcription machinery.

A similar message can be drawn from current work on the mechanism for processing the rRNA primary transcript for which, once again, the archaebacteria appear to use a system homologous to the eukaryotic processing pathway and distinct from that of eubacteria. The 16S and 23S rRNA genes of archaebacteria are flanked by repeats that can form a bulge-helix-bulge motif and which have a demonstrated role in transcript processing that is neither eubacterial nor eukaryotic in nature [40]. In the 16S gene of *S. acidocaldarius*, however, it is now known that the bulge-helix-bulge motif is not necessary in defining the 5′ processing site, but that the processing is actually carried out by an (as yet) unidentified enzymatic complex that requires an RNA component [41••], now recognized to be U3, the small nucleolar RNA (snoRNA) used by eukaryotes in the homologous process (S Potter, P Dennis, personal communication). Furthermore, in eukaryotes U3 forms a close association with fibrillarin, and this also appears to be the case in *Sulfolobus*, as U3 precipitates from cell extract when exposed to anti-fibrillarin antibody (S Potter, P Dennis, personal communication). Independent support for this comes from the discovery of fibrillarin in two methanogens [42••], which are related only distantly to *Sulfolobus*. This argues that this pathway is common to all archaebacteria, rather than only the crenarchaeota, which have an rRNA operon structure unlike other archaebacteria or eubacteria, and more like eukaryotes [43–47] (Fig. 2).

Arguably the best example of a feature shared between the archaebacteria and eukaryotes is ubiquitin-directed proteolysis by the proteasome, as this system has no known eubacterial homolog. The eukaryotic proteasome is a large multisubunit complex that recognizes polypeptides bound covalently to ubiquitin and degrades them. Both the proteasome and ubiquitin have been characterized in the genus *Thermoplasma* [48,49,50••], in which

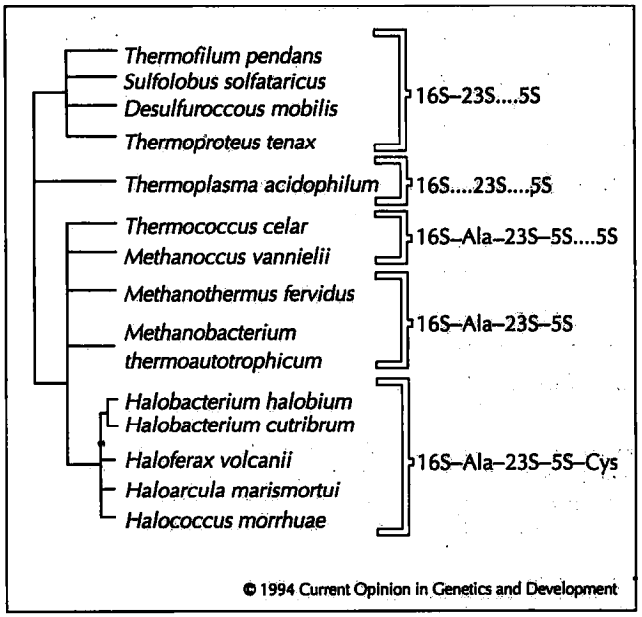

Fig. 2. Schematic diagram of archaebacterial phylogeny showing the different classes of rRNA operons. The crenarchaeota, represented here by *T. pendans, S. solfataricus, D. mobilis,* and *T. tenax,* have a separately transcribed 5S gene, as for eukaryotes. The euryarchaeota, which cover the remaining archaebacteria, maintain a 16S–23S–5S operon. The intergenic spacers of the euryarchaeota also contain lineage-specific tRNA genes, the halophiles being defined by an additional $tRNA^{Cys}$ at the extreme 3′ end of the operon. *Thermococcus celer* and *Methanococcus vannielii* have an operon typical of the euryarchaeota, but have additional independently expressed 5S genes (they are shown grouping together, but this trait may have evolved independently). *Thermoplasma,* as usual, has a unique organization [40,43–47].

their activity has been demonstrated to resemble the eukaryotic complex [51••]. The *Thermoplasma* proteasome is much simpler, consisting of only two subunits, rather than the 12–20 found commonly in eukaryotes. Moreover, the two subunits are similar in sequence, a product of duplication and divergence, a process that has apparently continued in the eukaryotes leading to the many contemporary subunits of the eukaryotic complex [52••]. As such, the two *Thermoplasma* subunits can be thought of as representing an ancestral-like complex and give us an opportunity to study the process of gene family evolution in the eukaryotes. The relatively simple subunit structure of the archaebacterial proteasome also makes it an ideal model for the difficult task of understanding the proteolytic activity of the complex, and progress has already been made in this area [51••].

Curiously, all efforts to identify proteasomes in archaebacterial genera other than *Thermoplasma* have failed resoundingly [52••]. Whether this is actually because of their absence, or only resistance to detection by the various techniques employed, is not certain, but the unique presence of proteasomes in *Thermoplasma* in the archaebacteria would raise some interesting issues.

Hanging ornaments on the universal tree

Besides simply lacking a nucleus, the archaebacteria share much of their gene and genome structure with the eubacteria, arguing for the utility of the concept 'prokaryote'. They also have a number of molecular characters in common with the eukaryotes, however, seemingly at the exclusion of the eubacteria, so it is imperative to be sure of the place of the archaebacteria in the universal tree. Conveniently, the most strongly supported topology of the universal tree, that which groups the archaebacteria and eukaryotes as sister taxa, is in agreement with most aspects of archaebacterial biology and can be used as a model from which to draw some interesting conclusions.

Earlier, we pointed out that, given this topology, anything found in both the archaebacteria and eubacteria can be said to have been inherited from the last common ancestor of all cells, even if it is now absent from eukaryotes. So, it seems likely that the universal ancestor was a complex prokaryotic cell with a circular DNA genome organized into operons with well developed regulatory systems and contents, which are still maintained in some cases. The genome of this cell included highly effective replication and expression systems, as well as a full complement of metabolic enzymes, and possibly a complex locomotary system resembling the rotary motor found in eubacteria and archaebacteria (the nature of the actual flagellar filaments is questionable [53••]). This ancestor is often referred to as a 'progenote', a term coined originally to define a very different organism, one that lacked all these things [54]. Given what we currently believe about the last common ancestor, it is clear that referring to it as a progenote is misleading.

It is not possible to use the same logic to fix the time of origin of features found only in archaebacteria and eukaryotes, as it is conceivable that these traits are themselves primitive. One can, however, actually begin to consider the order of events that took place in the development of eukaryotes without having to assume that the majority of complex characters present in only the archaebacterial and/or eukaryotic lineage are derived. Even the deepest branching protist lineages are significantly different from eubacteria in many aspects of cell structure and physiology, making it hard to imagine how such a rapid and thorough transition could take place. By identifying the archaebacterial homologs of systems thought to be strictly eukaryotic, we can reduce the number of novel developments necessary to account for the transition to eukaryotic cells, simplifying our explanation of a difficult step in cellular evolution. The presence of proteasomes, U3-based rRNA processing, and eukaryotic-like transcription in the archaebacteria, for instance, argue that these features had all evolved prior to the divergence of archaebacteria. Similarly, the apparent absence of other cellular

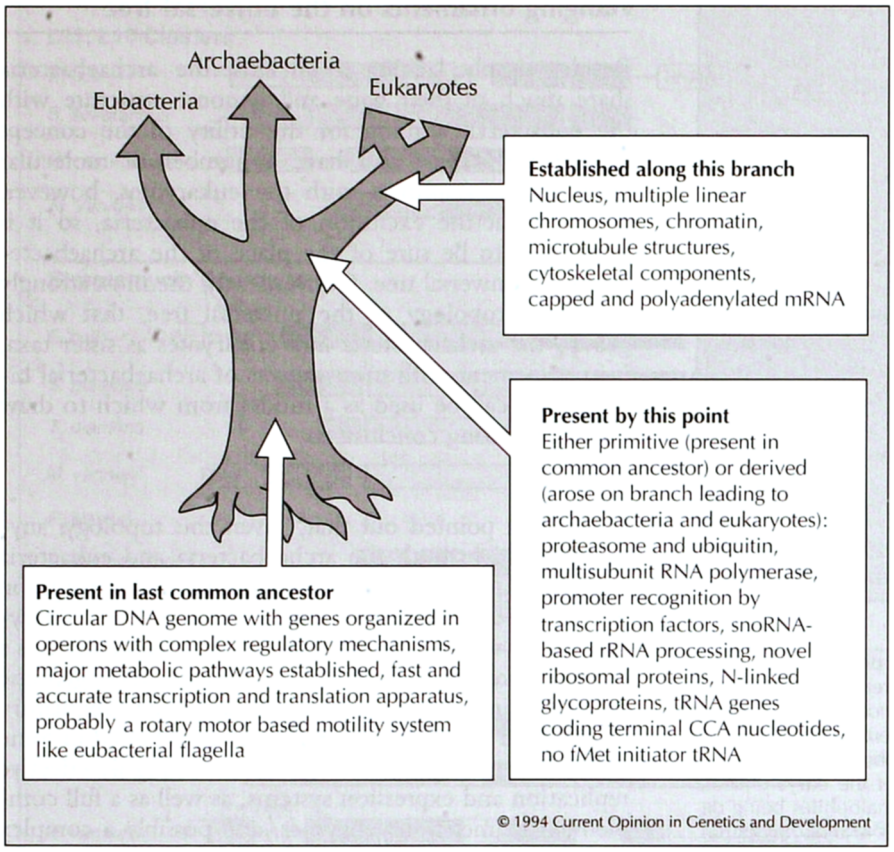

Fig. 3. Universal tree as determined by duplicated gene phylogeny [2]. A hypothetical order of events is shown for three major steps leading to the origin of eukaryotes. For the universal ancestor and the ancestor of eukaryotes and archaebacteria, we can say only that the various traits common to both resulting lineages had developed by that point, but we cannot say exactly when. For the traits common only to eukaryotes, it is most parsimonious to argue that they developed sometime after the divergence of archaebacteria and eukaryotes and before the divergence of the first eukaryotic lineages.

elements, such as microtubules and other cytoskeletal structures, cap-dependent translation initiation or multiple linear chromosomes (to name a few), implies that these evolved in the eukaryotes after the divergence of the archaebacteria. Even this short list puts some large-scale changes into perspective. In the case of gene expression, the eukaryotic system did not arise by concurrent changes to both translation and transcription, as the latter was already established in the prokaryotic ancestor of eukaryotes (a summary of this and other evolutionary inferences is presented in Fig. 3).

Conclusions

Eukaryotic cells are defined by their possession of many novel and complex features. As it is doubtful that many of these evolved *de novo*, their antecedants will probably be found somewhere in the prokaryotes, and if the archaebacteria really are the closest living relatives of the eukaryotes, then one would expect them to contain homologs of the vast majority of eukaryotic genes. This has already been seen to be the case for several integrated systems (some of which may also prove to be present in the eubacteria). In other instances, in which a complex system appears to have evolved after the divergence of eukaryotes and archaebacteria, it is still possible to uncover the prokaryotic origin of this trait by identifying the divergent, but distantly related, homologs of individual elements of this system.

A few interesting cases can be made for the recognition of these distant homologs, exemplified by actin and tubulin, two cytoskeletal proteins for which putative eubacterial relatives have been identified [55,56]. These assertions can be tested by identifying archaebacterial homologs of the relevant eubacterial proteins. The relatively short phylogenetic distance between the archaebacteria and the eukaryotes would lead to the prediction that the archaebacterial sequence would show a more obvious similarity to the novel eukaryotic gene than do any of the eubacterial sequences.

Another important observation is that an archaebacterial gene is often related equally to two or more different eukaryotic genes that perform different functions [37•,39••]. If this is a general condition, then the apparent slow rate of divergence in many archaebacterial sequences [38••] may allow the identification of new families or perhaps new members of existing families of genes and will certainly clarify the genealogy of recognized eukaryotic gene families by providing an external reference.

A major task of future genomic studies will be to explain changes in function and, thus, the differences between prokaryotic and eukaryotic cell and molecular biology in terms of changes in gene sequence and number. The sisterhood of the archaebacteria and eukaryotes makes the former, not the eubacteria, the appropriate model for defining the 'starter kit' of genes available to the first eukaryotic cells.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
- • of special interest
- •• of outstanding interest

1. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc Natl Acad Sci USA* 1977, **74**:5088–5090.

2. Iwabe N, Kuma K-I, Hasegawa M, Osawa S, Miyata T: **Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes.** *Proc Natl Acad Sci USA* 1989, **86**:9355–9359.

3. Pfeifer F, Palm P, Schleifer K-H (Eds): *Molecular biology of Archaea.* Stuttgart: Gustav Fischer Verlag; 1994.

4. Kates M, Kushner DJ, Matheson AT (Eds): *The biochemistry of Archaea.* Amsterdam: Elsevier; 1993.

5. Noll K: **Chromosome map of the thermophilic archaebacterium *Thermococcus celer.*** *J Bacteriol* 1989, **171**:6720–6725.

6. Charlebois RL, Schalkwyk LC, Hofman JD, Doolittle WF: **Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2.** *J Mol Biol* 1991, **222**:509–524.

7. Sitzmann J, Klein A: **Physical map of the *Methanococcus voltae* chromosome.** *Mol Microbiol* 1991, **5**:505–513.

8. López-García P, Abad JP, Smith C, Amils R: **Genomic organization of the halophilic archaeon *Haloferax mediterranei*: physical map of the chromosome.** *Nucleic Acids Res* 1992, **20**:2459–2464.

9. Kondo S, Yamagishi A, Oshima T: **A physical map of the sulfur-dependent archaebacterium *Sulfolobus acidocaldarius* 7 chromosome.** *J Bacteriol* 1993, **175**:1532–1536.

10. Bobovnikova Y, Ng W-L, Dassarma S, Hackett NR: **Restriction mapping the genome of *Halobacterium halobium* strain NRC-1.** *Syst Appl Microbiol* 1994, **16**:597–604.

11. Antón J, López-García P, Abad JP, Smith C, Amils R: **Alignment of genes and *Swa*I restriction sites to the *Bam*HI genomic map of *Haloferax mediterranei.*** *FEMS Microbiol Lett* 1994, **117**:53–60.

12. •• St Jean A, Trieselmann BA, Charlebois RL: **Physical map and set of overlapping cosmid clones representing the genome of the archaeon *Halobacterium sp.* GRB.** *Nucleic Acids Res* 1994, **22**:1476–1483.
The second construction of an ordered set of cosmids representing the genome of a halophile (the other is *Haloferax volcanii*). This will be an important tool in future investigations into comparative genomics and genome stability.

13. • López-García P, Abad JP, Smith C, Amils R: **Genome analysis of different *Haloferax mediterranei* strains using pulsed-field gel electrophoresis.** *Syst Appl Microbiol* 1993, **16**:310–321.
Develops a strain-screening protocol for *Hf. mediterranei* using a previously established map to predict and compare fingerprints. Also shows a fair degree of fingerprint polymorphism between strains.

14. Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Lanzendorfer, Kristjansson JK: **Screening for *Solfolobales*, their plasmids and viruses in Icelandic solfataras.** *Syst Appl Microbiol* 1994, **16**:609–624.

15. Schalkwyk L: **Halobacterial genes and genomes.** In *The biochemistry of Archaea.* Edited by Kates M, Kushner DJ, Matheson AT. Amsterdam: Elsevier; 1993:467–496.

16. Sapienza C, Rose MR, Doolittle WF: **High-frequency genomic rearrangements involving archaebacterial repeat sequence elements.** *Nature* 1982, **299**:182–185.

17. Palmer JR, Reeve JN: **Structure and function of methanogen genes.** In *The biochemistry of Archaea.* Edited By Kates M, Kushner DJ, Matheson AT. Amsterdam: Elsevier; 1993:497–534.

18. Brown JW, Reeve JN: **Polyadenylated, noncapped RNA from the archaebacterium *Methanococcus vannielii.*** *J Bacteriol* 1985, **162**:909–917.

19. Perler FB, Comb DG, Jack WE, Moran LS, Qiang B, Kucera RB, Benner J, Slatko BE, Nwankwo DO, Hempstead SK *et al.*: **Intervening sequences in an Archaea DNA polymerase gene.** *Proc Natl Acad Sci USA* 1992, **89**:5577–5581.

20. Xu M-Q, Southworth MW, Mersha FB, Hornstra LJ, Perler FB: ***In vivo* protein splicing of purified precursor and the identification of a branched intermediate.** *Cell* 1993, **75**:1371–1377.

21. Nieuwlandt DT, Carr MB, Daniels CJ: ***In vivo* processing of an intron-containing archael tRNA.** *Mol Microbiol* 1993, **8**:93–99.

22. • Dalgaard JZ, Garrett RA, Belfort M: **A site-specific endonuclease encoded by a typical archaeal intron.** *Proc Natl Acad Sci USA* 1993, **90**:5414–5417.
This demonstration that an intron-encoded open reading frame in *Desulfurococcus* is an endonuclease homologous to those found in certain homing introns provides clear evidence that these endonucleases are themselves mobile elements and argues that some archaebacterial introns can also home.

23. Ramírez C, Köpke AKE, Yang D-C, Boeckh T, Matheson AT: **The structure, function and evolution of archaeal ribosomes.** In *The biochemistry of Archaea.* Edited By Kates M, Kushner DJ, Matheson AT: Amsterdam. Elsevier; 1993:439–466.

24 Shimmin LC, Newton CH, Ramírez C, Yee J, Downing WL, Louie A, Matheson AT, Dennis PP: **Organization of genes encoding the L11, L1, L10, and L12 equivalent ribosomal proteins in eubacteria, archaebacteria, and eukaryotes.** *Can J Microbiol* 1989, **35**:164–170.

25. •• Hanner M, Mayer C, Köhrer C, Golderer G, Gröbner P, Piendl W: **Autogenous translational regulation of the ribosomal MvaI1 operon in the archaebacterium *Methanococcus vanielii.*** *J Bacteriol* 1994, **176**:409–418.
Feedback regulation at the level of translation is characteristic of eubacterial ribosomal protein operons. Here, it was shown that an archaebacterial operon uses the same system as its homolog in the eubacteria.

26. Auer J, Spicker G, Böck A: **Organization and structure of the *Methanococcus* transcriptional unit homologous to the *Escherichia coli* 'spectinomycin operon'.** *J Mol Biol* 1989, **209**:21–36.

27. Arndt E, Kromer W, Hatakeyama T: **Organization and nucleotide sequence of a gene cluster coding for eight ribosomal proteins in the archaebacterium *Halobacterium marismortui.*** *J Biol Chem* 1990, **265**:3034–3039.

28. Zillig W, Palm P, Klenk H-P, Langer D, Hudepohl U, Hain J, Lanzendorfer M, Holz H: **Transcription in Archaea.** In *The biochemistry of Archaea.* Edited by Kates M, Kushner DJ, Matheson AT. Elsevier: Amsterdam; 1993:367–391.

29. Sanangelantoni AM, Tiboni O: **The chromosomal location of genes for elongation factor Tu and ribosomal protein S10 in the cyanobacterium *Spirulina platensis* provides clues to the ancestral organization of the *str* and S10 operons in prokaryotes.** *J Gen Microbiol* 1993, **139**:2579–2584.

30. Zillig W: **Eukaryotic traits in archaebacteria.** *Ann NY Acad Sci* 1981, **503**:78–81.

31. Pühler G, Leffers H, Gropp F, Palm P, Klenk H-P, Lottspeich F, Garrett RA, Zillig W: **Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome.** *Proc Natl Acad Sci USA* 1989, **86**:4569–4573.

32. Klenk H-P, Palm P, Lottspeich F, Zillig W: **Component H of the DNA-dependent RNA polymerase of Archaea is homologous to a subunit shared by the three eukaryal nuclear RNA polymerases.** *Proc Natl Acad Sci USA* 1992, **89**:407–410.

33. Reiter W-D, Hudepohl U, Zillig W: **Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro.*** *Proc Natl Acad Sci USA* 1990, **87**:9509–9513.

34. Ouzounis C, Sander C: **TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes.** *Cell* 1992, **71**:189–190.

35. Creti R, Londei P, Cammarano P: **Complete nucleotide sequence of an archaeal (*Pyrococcus woesei*) gene encoding a homolog of eukaryotic transcription factor IIB (TFIIB).** *Nucleic Acids Res* 1993, **21**:2942.
An unidentified open reading frame, recognized originally by Ouzounis and Sander to be TFIIB [34], is sequenced fully here.

36. Langer D, Zillig W: **Putative *tfIIs* gene of *Sulfolobus acidocaldarius* encoding an archaeal transcription elongation factor is situated directly downstream of the gene for a small subunit of DNA-dependant RNA polymerase.** *Nucleic Acids Res* 1993, **21**:2251.
Initial report of a putative TFIIS gene in archaebacteria.

37. Kaine BP, Mehr IJ, Woese CR: **The sequence, and its evolutionary implications, of a *Thermococcus celer* protein associated with transcription.** *Proc Natl Acad Sci USA* 1994, **91**:3854–3856.
Report from a second archaebacteria of a gene homologous to the *Sulfolobus* TFIIS. This analysis concludes that this gene is actually not TFIIS, but rather more similar to subunits of eukaryotic RNA polymerases I and II. This situation will probably become increasingly common as more archaebacterial genes are shown to have multiple eukaryotic homologs.

38. Marsh TL, Reich CI, Whitlock RB, Olsen GJ: **Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes.** *Proc Natl Acad Sci USA* 1994, **91**:4180–4184.
This paper presents the sequence of an archaebacterial TFIID gene from a random genome sequencing project. TFIID is a tandem dimer, and sequence analysis of the two halves of the *Thermococcus* gene show that the duplication occurred before the divergence of the archaebacteria and that the rate of change in the archaebacterial sequence is less than in the eukaryotes, suggesting that archaebacterial proteins will be useful in understanding the evolution of gene families.

39. Rowlands T, Baumann P, Jackson SP: **The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria.** *Science* 1994,**264**:1326–1329.
The sequence of TFIID from *Pyrococcus* is determined and the activity of the initiation complex (specifically TFIIB/TFIID) described to be much like eukaryotes. This provides definitive evidence for the homology of the archaebacterial and eukaryotic transcription mechanisms.

40. Garrett RA, Aagaard C, Andersen M, Dalgaard JZ, Lykke-Andersen J, Phan HTN, Trevisanato S, Østergaard L, Larsen N, Leffers H: **Archaeal rRNA operons, intron splicing and homing endonucleases, RNA polymerase operons and phylogeny.** *Syst Appl Microbiol* 1994, **16**:680–691.

41. Durovic P, Dennis PP: **Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium *Sulfolobus acidocaldarius*: similarities to eukaryotic rRNA processing.** *Mol Microbiol* 1994, **13**:229–242.
Analysis of 23S and 16S rRNA processing in *Sulfolobus* shows that the former requires the typical bulge-helix-bulge structure, whereas the 16S RNA does not. The mature 5′ end of the 16S is defined by a putative RNA-containing endonuclease, and it is speculated that this is related to the eukaryotic U3-dependent processing system.

42. Amiri KA: **Fibrillarin-like proteins occur in the domain *Archaea*.** *J Bacteriol* 1994, **176**:2124–2127.
Genes from two methanogens were sequenced and shown to be very similar to eukaryotic fibrillarins. The presence of fibrillarins in these organisms not only supports the argument that archaebacterial rRNA processing is homologous to the process in eukaryotes, but also that this is generally true of archaebacteria and not just of the crenarchaeota.

43. Kjems J, Leffers H, Olesen T, Holz I, Garrett RA: **Sequence, organization and transcription of the ribosomal RNA operon and the downstream tRNA and protein genes in the archaebacterium *Thermophilum pendans*.** *Syst Appl Microbiol* 1990, **13**:117–127.

44. Kjems J, Leffers H, Garrett RA, Wich G, Leinfelder W, Böck A: **Gene organization, transcription signals and processing of the single ribosomal RNA operon of the archaebacterium *Thermoproteus tenax*.** *Nucleic Acids Res* 1987, **15**:4821–4835.

45. Reiter W-D, Palm P, Voos W, Kaniecki J, Grampp B, Schulz W, Zillig W: **Putative promoter elements for the ribosomal RNA genes of the thermoacidophilic archaebacterium *Sulfolobus sp.* strain B12.** *Nucleic Acids Res* 1987, **15**:5581–5595.

46. Leffers H, Kjems J, Østergaard L, Larsen N, Garrett RA: **Evolutionary relationships amongst archaebacteria. A comparative study of 23S ribosomal RNAs of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen.** *J Mol Biol* 1987, **195**:43–61.

47. Achenbach-Richter L, Woese CR: **The ribosomal gene spacer region in archaebacteria.** *Syst Appl Microbiol* 1988, **10**:211–214.

48. Zwickl P, Grziwa A, Pühler G, Dahlmann B, Lottspeich F, Baumeister W: **Primary structure of the *Thermoplasma* proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase.** *Biochemistry* 1992, **31**:964–972.

49. Zwickl P, Lottspeich F, Dahlmann B, Baumeister W: **Cloning and sequencing of the gene encoding the large (α-) subunit of the proteasome from *Thermoplasma acidophilum*.** *FEBS Lett* 1991, **278**:217–221.

50. Wolf S, Lottspeich F, Baumeister W: **Ubiquitin found in the archaebacterium *Thermoplasma acidophilum*.** *FEBS Lett* 1993, **326**:42–44.
A heroic effort to demonstrate that *Thermoplasma acidophilum* contains ubiquitin by systematically sequencing low molecular weight proteins.

51. Wenzel T, Baumeister W: ***Thermoplasma acidophilum* proteasome degrades partially unfolded and ubiquitin-associated proteins.** *FEBS Lett* 1993, **326**:215–218.
Shows that the *Thermoplasma* proteasome acts like its eukaryotic homolog. Also argues for the chaotropic effect of ubiquitin on the basis of observations from this *in vitro* system.

52. Pühler G, Pitzer F, Zwickl P, Baumeister W: **Proteasomes: multisubunit proteinases common to *Thermoplasma* and eukaryotes.** *Syst Appl Microbiol* 1994, **16**:734–741.
An intensive search for proteasome particles in archaebacteria other than *Thermoplasma* failed to detect anything by several techniques (as was also the case in several eubacteria), raising questions as to the position of *Thermoplasma* in the archaebacterial tree. Evolution of gene families is also examined; analysis of proteasome sequences reveals that the large number of eukaryotic subunits all arose from two original subunits and that the ancestral state is maintained in *Thermoplasma*.

53. Faguy DM, Jarrell KF, Kuzio J, Kalmokoff ML: **Molecular analysis of archaeal flagellins: similarity to the type IV pillin-transport superfamily widespread in bacteria.** *Can J Microbiol* 1994, **40**:67–71.
Shows that flagellin genes from the halophiles and methanogens are more closely related to eubacterial type IV pilin than to other flagellins. This correlates with other characteristics of archaebacterial flagellins and raises questions about flagellar evolution and biogenesis.

54. Woese CR: **Evolutionary questions: the 'progenote'.** *Science* 1990, **247**:789.

55. Lutkenhaus J: **FtsZ ring in bacterial cytokinesis.** *Mol Microbiol* 1993, **9**:403–409.

56. Bork P, Sander C, Valencia A: **An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins.** *Proc Natl Acad Sci USA* 1992, **89**:7290–7294.

57. Creti R, Citarella F, Tiboni O, Sanangelantoni A, Palm P, Cammarano P: **Nucleotide sequence of a DNA region comprising the gene for elongation factor 1-alpha (EF-1alpha) from the ultrathermophilic archaeote *Pyrococcus woesei*: phylogenetic implications.** *J Mol Evol* 1991, **33**:332–342.

PJ Keeling and WF Doolittle, Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, B3H 4H7, Nova Scotia, Canada.

RL Charlebois, Department of Biology, University of Ottawa, Ottawa, K1N 6N5, Ontario, Canada.