Archaeal genomics: **Do archaea have a mixed heritage?**

W. Ford Doolittle and John M. Logsdon, Jr.

A third complete archaeal genome sequence, replete with eukaryote-like genes for replication, transcription and translation, has appeared. The sequence also shows bacteria-like features. It is time to come to grips with this evidence for a mixed heritage.

Address: Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7.
E-mail: ford@is.dal.ca

Current Biology 1998, 8:R209–R211
http://biomednet.com/elecref/09609822008R0209

 ISSN 0960-9822

Completed microbial genome sequences are like books that we have paid for and brought home from the shop, but have not had time to read. Dauntingly, there are already a dozen of them on the shelf, according to the website (http://www.tigr.org) of The Institute for Genome Research. Three of these complete genome sequences are for archaeal (archaebacterial) species, the latest being that of *Archaeoglobus fulgidis*, published recently by Klenk *et al.* [1].

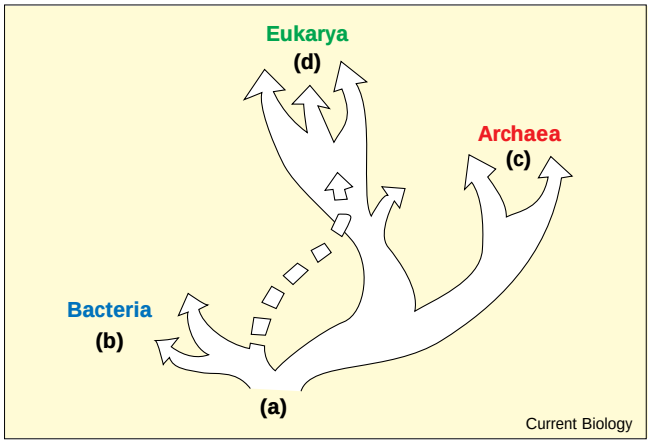
Readers may recall that, in 1993, the otherwise obscure archaeal genus *Archaeoglobus* was the subject of a scary paper entitled "Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs" [2]. These pressure-tolerant organisms, with temperature optima between 80°C and 90°C, grow organo-heterotrophically on a variety of carbon and energy sources, reducing sulphate to sulphide, and "souring" oil wells. Although this makes *Archaeoglobus* a sort of petrochemical pathogen, it is not the economic impact of this or the other two archaea for which complete genome sequences have been published — the methanogens *Methanococcus jannaschii* [3] and *Methanobacterium thermoautotrophicum* [4] — that makes them especially interesting. Rather, it is their thermophily, their unique energy metabolism, and (in the main) their pivotal position in our current understanding of cellular evolution.

Readers will surely recall the arguments in support of the special evolutionary position of archaebacteria [5], so we shall be brief in justifying our assertion. Archaea were 'discovered' by Woese and Fox in the late 1970s, in the course of their construction of a universal evolutionary tree from ribosomal RNA sequence information. A large amount of additional molecular sequence data, and many phenotypic peculiarities, show that archaea are phylogenetically coherent — that is, that they are not polyphyletic, although they may be paraphyletic. Attempts to root the 'universal tree' using duplicated genes for proteins involved in one of the cell's most fundamental processes, translation, indicate that archaea are the sister group of eukarya. That is, the deepest branch in the universal tree separates bacteria from a lineage which later diverged into archaea and eukaryotes, or more precisely, the nuclear-cytoplasmic component of eukaryotes (Figure 1).

This view suggests that some of the cellular and molecular features that distinguish eukaryotes from bacteria arose in the common archaeal/eukaryotic branch and may be present as 'eukaryote-like' traits in today's archaea. Much of the excitement in the archaeal genome sequence papers is indeed about such features. These papers reveal replication, transcription and translation machinery that are strikingly 'eukaryotic' in complexity and the sequence of their components. Archaea, like eukaryotes and not bacteria, use B-type DNA polymerases for replication, and many other proteins present at the archaeal replication fork seem 'eukaryotic'. Most ribosomal proteins, and a host of translation initiation factors, appear eukaryote-like. In various combinations, the three genomes bear genes clearly homologous to those encoding the eukaryotic transcription factors TFIIB, TFIIC, TFIID, TFIIE, TFIIS and TIP49. It has also recently been shown [6] that *Mc. jannaschii* carries homologs of bacterial sigma-70 proteins, so archaeal transcription may in fact have a mixed bacterial–eukaryotic character.

By contrast, many archaeal genes with products that are involved in intermediary metabolism look *purely* bacterial.

Figure 1



Rooted universal tree of life [5]. The broken arrow indicates mitochondrial endosymbiosis.

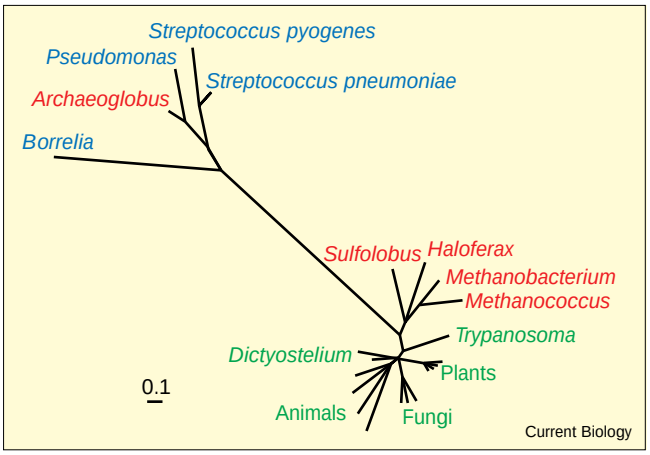
For instance, Koonin and collaborators [7], using the BLAST2 algorithm and new methods for amino-acid motif detection, find that about 37% (reading from their figure) of *Mc. jannaschii* proteins are found in all three domains — eukarya, archaea and bacteria — and as many as 26% are found otherwise only in bacteria, but only 5% are confined to archaea and eukarya. Curiously enough, people are starting to say the same thing about eukaryotes — that many more of their 'metabolic genes' are bacterial than one should expect if Figure 1 is correct [8]. How can this be true: how can both archaea and eukarya have 'bacterial' genes for metabolism?

If the same 'metabolic' genes are being addressed in each situation, this paradox must mean one of two things. First, it could be that Figure 1 is correct, but bacterial genes evolve (change in sequence) much more slowly than do archaeal or eukaryotic ones, so that the distances between (b) and (c) and between (b) and (d) in Figure 1 are each shorter than the distance between (c) and (d). Alternatively, Figure 1 could be wrong, and bacterial genes have been injected into archaeal or eukaryotic genomes, or both, sometimes adding to and sometimes replacing archaeal or archaea-like genes already there. Gene-by-gene phylogenetic analyses [9] may help in the choice here, but ancient paralogy (gene duplication) and differential loss can always confuse us.

Two things are clear about such injection or 'lateral transfer' of genes. First, there are individual instances in which it has surely occurred. In browsing the *Archaeoglobus* sequence, for instance, we noticed that the strongest BLAST 'hit' for the *Archaeoglobus* enzyme 3-hydroxy-3-methylglutaryl-coenzyme A (HMGCoA) reductase was against that of the bacterium *Pseudomonas mevalonii*. We happen to be familiar with HMGCoA reductase, having sequenced its gene from the archaea *Haloferax volcanii* and *Sulfolobus solfataricus* [10]. These latter two, and the *Mc. jannaschii* and *Mb. thermoautotrophicum* versions, are very similar to each other and to eukaryotic HMGCoA reductase genes (which is why archaea are sensitive to lovastatin, a drug used to treat hypercholesterolemia). What is critical — and can be known only from the complete genome sequence — is that *Archaeoglobus* does not have this archaeal/eukaryal version of HMGCoA reductase.

More important to the microbes themselves, HMGCoA reductase catalyses the crucial first specific step in the synthesis of the isoprenoid-based ether lipids, arguably the most unusual and characteristic feature of archaea as cells. The *Pseudomonas* type enzyme, on the other hand, has been thought of as degradative in character, a distant and functionally distinct homolog of the archaeal/eukaryal enzyme that is possibly confined to this species. (In fact, its distribution is patchy among bacteria: among the nine fully and twelve partially sequenced bacterial genomes, only those of *Borrelia burgdorferii*, *Streptococcus pyogenes* and *S. pneumoniae* seem to have it.) It almost beggars the imagination that a gene for an anabolic function so crucial to all archaea could be not only supplemented by, but replaced with, a distant catabolic homolog through 'lateral gene transfer'. But more economical interpretations of Figure 2 do not readily come to mind.

Figure 2



Phylogenetic tree of HMGCoA reductases from three domains. Bacteria are shown in blue, Archaea in red and Eukarya in green. Animals included are *Caenorhabditis*, *Schistosoma*, *Drosophila* and *Rattus*; fungi are *Saccharomyces*, *Schizosaccharomyces* and *Giberrella*; plants are *Arabidopsis* (two paralogs), *Hevea* and *Zea*. The distance tree was constructed from an alignment of 443 amino-acid residues, similar to that used by Bochar *et al.* [10], employing the PAM250 substitution matrix with PROTDIST and NEIGHBOR programs from PHYLIP v3.57.

The second thing that is clear about lateral transfer is so obvious that it is often overlooked. None of the sequenced genomes has the same complement of genes. *Archaeoglobus* (and its close relatives), for instance, has a panoply of genetic determinants pertinent to its heterotrophic and sulphate-reducing habits that are not found in the other sequenced archaeal genomes. Klenk *et al.* [1] note that, while 80% of *Archaeoglobus*' genes for replication, transcription and translation are present in *Mc. jannaschii*, only 35% of its genes for intermediary metabolism are so shared. This will be the general pattern for all sufficiently unrelated genomes, and it can be accounted for in three general ways. The first is that new genes are constantly being invented. Within-genome duplications are indeed known for all sequenced genomes and provide a potential source for new genes, but we must remember that many of the genes that *Archaeoglobus* does not share with *Mc. jannaschii* are in fact homologs of genes in other organisms (often bacteria). Otherwise, Klenk *et al.* [1] could not have assigned names to them.

The second possibility is that the last common ancestor — the organism at point (a) in Figure 1 — had ancestral versions

of all the genes now found in any organism, and the subsequent history of genomes has been one of loss. Not only is such a 'mother-of-all genomes' hypothesis bizarre, it provides a most unparsimonious interpretation of distributions such as that exhibited by HMGCoA reductase, when we remember that only three of eleven completely sequenced prokaryotic genomes have any version of this gene. The third, and by elimination most likely, interpretation is that lateral transfer is not just a molecular phylogenetic nuisance supported in evidence by a few anecdotal cases. Instead it is a major force, at least in prokaryotic evolution. Our trees are gene trees, not organism trees.

If we are not to give up molecular phylogenetics of prokaryotes altogether, we need an explicit way of recognizing this. The view implicit in archaeal genome sequence papers (and the editorials which accompany them) is that genes for replication, transcription and translation machinery comprise the blue-print for a sort of enduring cellular hardware, by which genes for other (more traditionally biochemical) functions can be read as software. Because the hardware endures, we can equate its evolutionary history with that of cellular lineages, as in Figure 1. This is of course very much a molecular biologist's view. Were cell biologists and biochemist running the genomics show, they might favor a different perspective. The community needs a more open debate.

References

1. Klenk H-P, Clayton RA, Tomb, J-F, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey ER, Peterson JD, *et al.*: The complete genome sequence of the hyperthermophilc, sulphate-reducing archaeon ***Archaeoglobus fulgidis***. *Nature* 1997, 390:364-370.
2. Stetter KO, Huber R, Blochl E, Kurr M, Eden RD, Fielder M, Cash H, Vance I: Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* 1993, 354:743-745.
3. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, *et al.*: Complete genome sequence of the methanogenic archaeon, ***Methanococcus jannaschii***. *Science* 1996, 273:1058-1073.
4. Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, *et al.*: Complete genome sequence of ***Methanobacterium thermoautotrophicum*** DELTA H: functional analysis and comparative genomics. *J Bacteriol* 1997, 179:7135-7155.
5. Doolittle WF, Brown JR: Tempo, mode, the progenote and the universal root. *Proc Natl Acad Sci USA* 1994, 91:6721-6728.
6. Kyrpides NC, Ouzounis CA: Bacterial σ^{70} transcription factor DNA-binding domains in the archaeon ***Methanococcus jannaschii***. *J Mol Evol* 1997, 45:706-707.
7. Koonin EV, Mushegian AR, Galperin MY, Walker DR: Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol Microbiol* 1997, 25:619-637.
8. Feng DF, Cho G, Doolittle RF: Determining divergence times with a protein clock: update and reevaluation. *Proc Natl Acad Sci USA* 1997, 94:13028-13033.
9. Brown JR, Doolittle WF: Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Revs* 1997, 61:456-502.
10. Bochar DA, Brown JR, Doolittle WF, Klenk HP, Lam W, Schenk ME, Stauffacher CV, Rodwell VW: 3-hydroxy-3-methylglutaryl coenzyme A reductase of ***Sulfolobus solfataricus***: DNA sequence, phylogeny, expression in Escherichia coli of the hmgA gene, and purification and kinetic characterization of the gene product. *J Bacteriol* 1997, 179:3632-3638.