

Evolutionary genomics: *Thermotoga* heats up lateral gene transfer

John M. Logsdon, Jr. and David M. Faguy

The complete sequence of the bacterium *Thermotoga maritima* genome has revealed a large fraction of genes most closely related to those of archaeal species. This adds to the accumulating evidence that lateral gene transfer is a potent evolutionary force in prokaryotes, though questions of its magnitude remain.

Address: Department of Biochemistry and Molecular Biology,
Dalhousie University, Halifax, Nova Scotia, B3H 4H7, Canada.
E-mail: dfaguy@unm.edu

Current Biology 1999, 9:R747–R751

0960-9822/99/\$ – see front matter
© 1999 Elsevier Science Ltd. All rights reserved.

Prokaryotes exchange genes on a regular basis, especially under highly selective conditions (such as in the presence of antibiotics). These DNA transfer events have been appreciated by biologists for some time now. Indeed, it was recognized early on that this process could have profound implications for the evolution of microbes and our ability to trace their history, yet most microbiologists maintained that an evolutionary classification of microorganisms, while difficult, was possible. In the last decade, reports of lateral — or horizontal — gene transfers have been steadily rising. From the profusion of recent articles on the topic [1–13], observations of lateral gene transfer have seemingly reached fever pitch, largely catalyzed by the sequencing and analyses of complete genomes from numerous diverse prokaryotes. Lateral gene transfer has clearly caught the attention of biologists, but despite this excitement important questions on the prevalence and impact of the process remain unanswered.

The latest addition to the lateral gene transfer fray comes from the genome-sequencing crew at The Institute for Genomic Research (TIGR) [1], who have determined the complete genome sequence of *Thermotoga maritima*, a hyperthermophilic bacterium that may be one of the deepest-branching lineages within the Bacteria. The most interesting feature of this genome is the surprisingly high proportion of open reading frames — putative protein-encoding genes — that most closely resemble genes, not from Bacteria, but instead from the other prokaryotic domain, the Archaea.

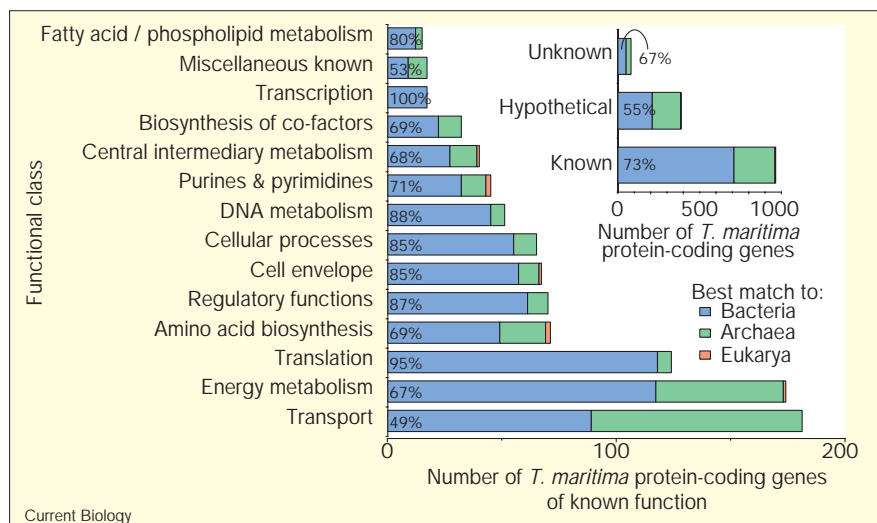
Nelson *et al.* [1] found that a full 24% of the *T. maritima* open reading frames — 451 of 1877 — are most similar to archaeal genes (Figure 1). This fraction of archaeal-like open reading frames is nearly twice that of another hyperthermophilic — and likely deep-branching — bacterium, *Aquifex aeolicus* [14], which previously held the record for

having the largest (by far) fraction of archaeal-like genes observed in a bacterial species. The high fraction of archaeal-like genes is found in the *T. maritima* genome even though the comparisons included the previously determined *A. aeolicus* genome, though the converse is not true. Indeed, others have made strong claims for “massive gene exchange” between *A. aeolicus* and archaeal thermophiles [8], yet it appears that the extent of archaeal genes in *T. maritima* is even greater. There is little doubt that *T. maritima* is a member of the Bacteria, and over half of its genes (though only just) appear bacterial in origin. Although many of the archaeal-like *T. maritima* genes appear to be involved in metabolic functions, such as transport and energy metabolism, it is perhaps surprising that at least some are involved in such presumably more general (‘core’) functions as transcription and gene regulation (Figure 1).

Consistent with the view that with these archaeal-like genes arose as a result of rampant lateral gene transfer during the evolution of the *T. maritima* genome, Nelson *et al.* [1] observed that substantial regions of the genome have a DNA base composition significantly different than the rest of the genome. This may indicate that the genes in these regions were transferred *en masse*. In further support of an origin of these regions by lateral gene transfer, the authors suggest — though with no statistical support — that the archaeal-like genes are clustered in these areas. Curiously, some of these regions contain a series of 30 base-pair repeats that are very similar in structure and base composition to repeats found in Archaea and some (especially thermophilic) Bacteria. But as these repeats were originally reported in (archaeal) mesophilic halophiles [15], and a similar repeat structure is found in *Escherichia coli* [16], their relevance for lateral gene transfer is unclear.

The genome sequence of *T. maritima*, like all completed genomes of hyperthermophiles (to date, mostly Archaea), contains significant numbers of genes classed as ‘unknown’ or ‘hypothetical’ because their closest sequence matches are to genes of unknown or hypothetical function, respectively [1]. It is likely that a number of these genes will turn out to be specific to hyperthermophiles, whether by common ancestry and loss in other lineages or by lateral gene transfer. This is borne out by the *T. maritima* data: of the 108 genes matching *only* genes in other hyperthermophiles, 93 are in the ‘hypothetical’ function class, roughly 23% of all ‘hypothetical’ proteins encoded in the genome. Most relevant, perhaps, to the question of lateral gene transfer is that a significant fraction of these ‘hypothetical’ genes in *T. maritima* are archaeal-like (Figure 1, inset).

Figure 1



Distribution of genes in *T. maritima* by functional class (using values and class assignments from [1]). The number of genes in *T. maritima* which most closely resemble (by BLAST) known genes from Bacteria, Archaea, and Eukarya are shown in blue, green and red, respectively. In each case, the percentage of genes showing a best match to genes from Bacteria is specified. The inset shows the same distribution for all genes in *T. maritima* that have any match in the non-redundant protein database: the 'known' class is the sum of the functional classes shown in the main graph; the 'unknown' class is all those showing a match to a gene with unknown function; the 'hypothetical' class is all those showing a match to a gene with an inferred or hypothetical function.

Evidence for lateral gene transfer?

Although *T. maritima* is not the first genome that appears to have a mosaic origin, what is striking is the large fraction of its genes (almost 25%) which appear specifically related to another domain. If most prokaryotic organisms experienced lateral gene transfer of this magnitude, the very concept of a prokaryotic lineage would be called into question. Similarly, results reported last year by Lawrence and Ochman [5] indicate that approximately 18% of the genes in the *E. coli* genome are derived from lateral gene transfers, although it is unclear how comparable these lateral gene transfers are to more distant ones, such as those inferred from the *T. maritima* genome sequence. In any case, these data, taken in sum, are prompting the deconstruction of prokaryotic molecular systematics [2,3,12]. But before throwing out the organismal trees, we should ask if there are explanations — other than lateral gene transfer — for at least some of the *T. maritima* cases.

With the incredible amount of data present in a complete genome, it is now common for bioinformaticians to describe each gene by its closest match in the database (usually using the BLAST program). While this practice is certainly useful as a first cut, it can lead to unwarranted conclusions. Caution should be exercised in interpreting the results of overall similarity scores, especially when used to specify the phylogenetic source of individual genes. A simplified schematic of this problem is shown in Figure 2a. A gene from some organism, number 5 say, when included in a phylogenetic tree, clearly belongs with a particular group (Bacteria, blue) and not another (Archaea, green) when the tree is appropriately rooted (with Eukarya, red). Yet, when this tree is either rooted incorrectly — for example, in the absence of an outgroup,

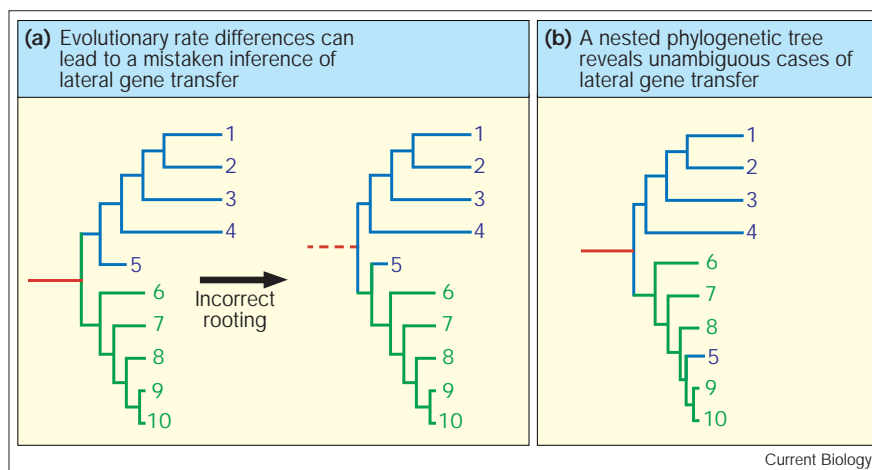
using a midpoint rooting — or is unrooted, it can appear that organism 5 is not in the Bacteria, but instead groups with Archaea. This effect would be exacerbated by unequal evolutionary rates [10].

As *T. maritima* is a plausible candidate for being a representative of one of the deepest bacterial lineages, this scenario is certainly possible for some of the genes thought to be derived by lateral gene transfer. In any case, it is easy to see how incorrect inferences of lateral gene transfer can arise. In the absence of additional supporting data — most clearly a well-supported tree in which the lateral gene transfer recipient is nested within the donor lineage (Figure 2b) — inferences of lateral gene transfer from such distance comparisons (such as BLAST scores), regardless of their sheer numbers, are really hypotheses in need of further testing.

Nelson *et al.* [1] did perform a phylogenetic analysis on 33 homologous gene families with members from *T. maritima*, and report that, in this small subset, a “majority of genes” showed no lateral gene transfer between Archaea and Bacteria. These analyses revealed significant differences between different gene trees within the Bacteria, suggesting that gene duplication, loss and/or lateral gene transfer (within Bacteria) are important in the evolution of the *T. maritima* genome. In light of the incongruencies observed in their phylogenetic trees, Nelson *et al.* [1] favor overall pairwise similarity comparisons of complete genomes as “an alternative to single gene phylogenetic analyses”. Unfortunately though, it is such phylogenetic trees that can provide the best possible evidence for and substantiation of lateral gene transfers — particularly those between Bacteria and Archaea. In any case, the verified archaeal lateral gene transfers (some unknown fraction of

Figure 2

An illustration of how phylogenetic trees can be used to infer lateral gene transfer. In these hypothetical distance trees – with branch lengths shown proportional to divergence – Bacteria (organisms 1–5) are shown in blue, Archaea (organisms 6–10) in green, and the outgroup (Eukarya) root in red. (a) The two trees in this panel have the same topology, but while the one on the left is correctly rooted, the one on the right is incorrectly rooted (dashed line). This incorrect rooting could lead to an erroneous inference of an Archaeal lateral gene transfer for organism 5 (a deep-branching member of Bacteria). This effect is enhanced by the fact that rates of sequence evolution differ: as organism 5 and the Archaea are evolving more slowly than Bacteria, the evolutionary distance between 5 and the other Bacteria (organisms 1–4) is two-fold higher than that between 5 and the Archaea. As BLAST scores indicate such pairwise distances (and not the overall tree drawn from them), the ‘best match’ criterion used by Nelson *et al.* [1] would indicate for this case that organism 5 has an ‘archaeal-like’ gene. In addition to probably being deep-branching



members of the Bacteria, *T. maritima* and *A. aeolicus* apparently share with some Archaea relatively slow rates of gene sequence evolution (the rRNA tree in [10] illustrates this). Although not shown, misleading associations can be also made between unrelated, but

rapidly evolving genes. (b) The position of organism 5 here – within the Archaea – illustrates a case of unambiguous lateral gene transfer from Archaea (especially if organism 5 lacks the bacterial version of the gene). An example of this is given in [17].

those identified by pairwise distance methods) will actually underestimate the total amount of lateral gene transfer that has contributed to the *T. maritima* genome, as their numbers would not include those lateral gene transfers with other Bacteria.

Using phylogenetic reasoning similar to that illustrated in Figure 2, Kyrpides and Olsen [10] recently challenged the use of simple pairwise comparisons as the basis for inferring lateral gene transfer. Instead, they argued that many such shared genes could be explained by common – vertically inherited – ancestry. In particular, they disputed claims of “massive” lateral gene transfer between *A. aeolicus* and thermophilic Archaea made by Aravind *et al.* [8], who reported that of 1512 *A. aeolicus* open-reading frames, 246 (16.2%) were “reliable best hits” to Archaea. In response, Aravind *et al.* [11] continue to favor lateral gene transfer as the best overall explanation. But their further analyses indicate that, from 97 of these archaeal-like open reading frames (those which were sampled in enough species), only 54 genes – comprising 3.7% of *A. aeolicus* open reading frames – clearly “fall within the archaeal group”. This more precise criterion demonstrates that, although lateral gene transfers can be uncovered by percentage similarity scores, these ‘best-hit’ analyses can exaggerate its extent.

Is it possible that the unusually numerous gene similarities which both *A. aeolicus* and *T. maritima* share with Archaea reflect the fact that these are deep-branching

Bacteria that have retained more genes (or have evolved more slowly) from their shared common ancestor with Archaea than have other lineages? Some of the recent results from Aravind *et al.* [11] – specifically that 43 of the 97 *A. aeolicus* open reading frames, each with highest similarity to Archaea, did not group within Archaea – actually appear most consistent with the vertical “common ancestry” hypothesis of Kyrpides and Olsen [10]. Hence, the high fraction of ‘archaeal-like’ genes in *T. maritima* (and to a lesser extent in *A. aeolicus*) could indicate that these organisms are truly deep branches on the bacterial tree (such as Figure 2a), perhaps even implying that *T. maritima* diverged prior to *A. aeolicus*. These possibilities are certainly not inconsistent with current views on bacterial phylogeny. In fact, TIGR’s Nelson was quoted in *Science* as saying that “It was impossible to say whether *Aquifex* or *Thermotoga* was more ancient” [3]. At present there is not enough evidence to decide between lateral gene transfer and vertical evolution as the primary cause of the high proportion of ‘archaeal-like’ genes in *T. maritima* (and in *A. aeolicus*), even though there are examples that can be clearly explained by one or the other process. However, vertical ancestry would become an increasingly untenable explanation for most cases if further evidence of extensive lateral gene transfer is found between additional diverse Bacteria and Archaea.

Our caution about lateral gene transfer should not be taken as dismissal of the process, just a healthy skepticism of its magnitude. Certainly there are numerous well-documented

cases of lateral gene transfer in prokaryotes, including exchanges between Bacteria and Archaea [17]. One particularly striking example was discovered recently in the course of a partial genome analyses of the archaeon *Pyrococcus furiosus*. This revealed a 16 kilobase region sharing over 99% identity to a segment of the *Thermococcus litoralis* genome (also an archaeon); similar segments are not found in the complete genome sequences of the closely related species *Pyrococcus abyssi* or *Pyrococcus horikoshii* (Frank Robb, personal communication). Interestingly, this region, which is flanked by transposons, shows all the signs of transposon-mediated lateral gene transfer, and thus may be a fortuitous snapshot of a lateral gene transfer in progress.

Lateral gene transfer events may not be rare in *P. furiosus*, as other genes (and entire operons) are present in the *P. furiosus* genome but absent in closely related *Pyrococcus* species [7]. The case of *P. furiosus* and *T. litoralis* also illustrates the importance of lateral gene transfer for acquiring new functions. Close relatives of *P. furiosus* do not have the gene for maltose transport, which is contained in the segment shared between *T. litoralis* and *P. furiosus*. It is not hard to imagine that such a gene could be highly advantageous, with its resulting selection leading to maintenance in the *P. furiosus* population of a relatively rare interspecies genetic transfer.

Are there 'core' genes refractory to lateral gene transfer?

Returning to *T. maritima*, it is notable that most of its 'archaeal-like' genes for which a function has been identified are involved with energy metabolism or transport of macromolecules (Figure 1). It is reasonable to assume that the acquisition of these sorts of genes conferred a strong selective advantage. So while it is clear that lateral gene transfer is a real mechanism for the acquisition of new functions — even though its magnitude is only now being revealed from complete genome analyses — it is not at all clear what role lateral gene transfer has played in the evolution of the 'core' genes that code for proteins with more universal functions, such as transcription factors and other proteins concerned with processing the information content of the genome.

Figure 1 clearly shows a pattern that has been observed in many genomes: genes involved in some functional classes, such as energy metabolism, more frequently show evidence of lateral gene transfers than those in other functional classes, such as transcription and translation. Jain *et al.* [6] have recently proposed that 'informational' genes are more likely to be components of large, complex systems which are more refractory to lateral gene transfer; they refer to this as "the complexity hypothesis". Complete genome analyses have given this proposition much support.

On the other hand, if members of large, complex systems generally evolve more slowly, it is possible that they might

actually be *easier* to replace with a homolog from a distant relative. Support for this comes from classic *in vitro* studies in which components from different species were shown to assemble into functional ribosomes [18]. Even stronger evidence has come recently from Asai *et al.* [19], who showed that ribosomal DNA genes in *E. coli* could be replaced with homologs from *Salmonella typhimurium* or *Proteus vulgaris* with little apparent effect. Certainly, it is possible to transfer (and replace) genes involved in a large, macromolecular complex, such as the ribosome, but the relevance of this 'artificial' lateral gene transfer to evolutionary history is not obvious.

So what if there is lateral gene transfer?

One might infer from our discussion that surprisingly little is known about lateral gene transfer. We know it has occurred, but we do not yet know its prevalence. And, as the saying goes, a little knowledge can be a dangerous thing. It is worth remembering that before the advent of genomics, or even gene sequencing, microbiologists constructed a taxonomy of microorganisms that remains relatively robust. Genomic data has not yet contradicted the general placement of an organism — *T. maritima* is still a bacterium — and probably never will. But if lateral gene transfer is proved to be as rampant as suggested by Nelson *et al.* [1] and others [2,8,13], then prokaryotic phylogeny (evolutionary history) is in serious danger of becoming mere taxonomy (similarity grouping). For one to say *T. maritima* is a 'bacterium', while correct, may not express much about its evolutionary history — just as saying Tiger Woods is an 'American' expresses little about his genealogy. In any case, the fact of lateral gene transfer, whatever its magnitude, compels us to include it in our concept of evolution — especially for prokaryotes.

Acknowledgements

The authors would like to extend our gratitude to Ford Doolittle for numerous free-ranging discussions on lateral gene transfer and for his encouragement of dissent. Thanks are also extended to Doolittle lab members for discussion and suggestions on the manuscript. This work was supported, in part, by a MRC (Canada) grant to W.F. Doolittle. J.M.L. and D.M.F. were also supported by postdoctoral fellowships from NIH and NSERC (Canada), respectively.

References

1. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, *et al.*: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 1999, **399**:323-329.
2. Doolittle WF: Phylogenetic classification and the universal tree. *Science* 1999, **284**:2124-2129.
3. Pennisi E: Is it time to uproot the tree of life? *Science* 1999, **284**:1305-1307.
4. Lake JA, Jain R, Rivera MC: Mix and match in the tree of life. *Science* 1999, **283**:2027-2028.
5. Lawrence JG, Ochman H: Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* 1998, **95**:9413-9417.
6. Jain R, Rivera MC, Lake JA: Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* 1999, **96**:3801-3806.

7. Maeder DL, Weiss RB, Dunn DM, Cherry JL, Gonzalez JM, DiRuggiero J, Robb FT: Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* 1999, 152:1299-1305.
8. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV: Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* 1998, 14:442-444.
9. Wolf YI, Aravind L, Koonin EV: Rickettsiae and Chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends Genet* 1999, 15:173-175.
10. Kyrpides NC, Olsen GJ: Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet* 1999, 15:298-299.
11. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV: Reply. *Trends Genet* 1999, 15:299-300.
12. Martin W: Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* 1999, 21:99-104.
13. Wolf YI, Aravind L, Grishin NV, Koonin EV: Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res* 1999, 9:689-710.
14. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, et al.: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 1998, 392:353-358.
15. Mojica FJ, Ferrer C, Juez G, Rodriguez-Valera F: Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol Microbiol* 1995, 17:85-93.
16. Nakata A, Amemura M, Makino K: Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J Bacteriol* 1989 171:3553-3556.
17. Doolittle WF, Logsdon JM, Jr.: Archaeal genomics: do archaea have a mixed heritage? *Curr Biol* 1998, 8:R209-R211.
18. Nomura M, Held WA: Reconstitution of ribosomes: studies of ribosome structure, function and assembly. In *Ribosomes*. Edited by Nomura M, Tissières A, Lengyel P. Plainview, NY: Cold Spring Harbor Laboratory Press; 1974.
19. Asai T, Condon C, Voulgaris J, Zaporozhets D, Shen B, Al-Omar M, Squires C, Squires CL: Construction and initial characterization of *Escherichia coli* strains with few or no intact chromosomal rRNA operons. *J Bacteriol* 1999, 181:3803-3809.

If you found this dispatch interesting, you might also want to read the **October 1999** issue of

Current Opinion in Microbiology

which included the following reviews, edited by **Howard Ochman**, on **Genomics**:

Gene transfer, speciation, and the evolution of bacterial genomes

Jeffrey G Lawrence

Codon usage and lateral gene transfer in *Bacillus subtilis*

Ivan Moszer, Eduardo PC Rocha and Antoine Danchin

Evolution of the linear DNA replicons of the *Borrelia* spirochetes

Sherwood Casjens

Origins of mitochondria and hydrogenosomes

Siv GE Andersson and Charles G Kurland

Archaeal genomics

Terry Gaasterland

Yeast genome evolution in the post-genome era

Cathal Seoighe and Kenneth H Wolfe

Genome archeology leading to the characterization and classification of transport proteins

Milton H Saier Jr

the same issue also included the following reviews, edited by **Patrice Courvalin** and **Julian Davies**, on **Antimicrobials**:

Recent developments in macrolides and ketolides

Daniel TW Chu

What's new in the antibiotic pipeline

Ving J Lee, George H Miller and Morimasa Yagisawa

Resistance gene capture

Dean A Rowe-Magnus and Didier Mazel

The biological cost of antibiotic resistance

Dan I Andersson and Bruce R Levin

Antibiotic use in humans and bacterial resistance

Didier Guillemot

Aminoglycoside-modifying

Gerard D Wright

The full text of *Current Opinion in Microbiology* is in the BioMedNet library at

<http://BioMedNet.com/cbiology/mcr>