Bacterial Origin for the Isoprenoid Biosynthesis Enzyme HMG-CoA Reductase of the Archaeal Orders Thermoplasmatales and Archaeoglobales

Yan Boucher,* Harald Huber,† Stéphane L’Haridon,‡ Karl O. Stetter,† and W. Ford Doolittle* 

*Canadian Institute for Advanced Research, Dalhousie University, Halifax, Nova-Scotia, Canada; †Lehrstuhl für Mikrobiologie, Universität Regensburg, Regensburg, Germany; and ‡Station Biologique, Centre National de la Recherche Scientifique, Roscoff, France

The enzyme 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase or HMGR) fulfills an essential role in archaea, as it is required for the synthesis of isoprenoid ethers, the main component of archaeal cell membranes. There are two clearly homologous but structurally different classes of the enzyme, one found mainly in eukaryotes and archaea (class 1), and the other found in bacteria (class 2). This feature facilitated the identification of several cases of interdomain lateral gene transfer (LGT), in particular, the bacterial origin for the HMGR gene from the archaeon Archaeoglobus fulgidus. In order to investigate if this LGT event was recent and limited in its scope or had a broad and long-term impact on the recipient and its related lineages, the HMGR gene was amplified and sequenced from a variety of archaea. The survey covered close relatives of A. fulgidus, the only archaean known prior to this study to possess a bacterial-like HMGR; representatives of each main euryarchaeal group were also inspected. All culturable members of the archaeal group Archaeoglobales were found to display an HMGR very similar to the enzyme of the bacterium Pseudomonas mevalonii. Surprisingly, two species of the genus Thermoplasma also harbor an HMGR of bacterial origin highly similar to the enzymes found in the Archaeoglobales. Phylogenetic analyses of the HMGR gene and comparisons to reference phylogenies from other genes confirm a common bacterial origin for the HMGRs of Thermoplasmatales and Archaeoglobales. The most likely explanation of these results includes an initial bacteria-to-archaea transfer, followed by another event between archaea. Their presence in two divergent archaean lineages suggests an important adaptive role for these laterally transferred genes.

Introduction

As genomics unveil the sequences of ever more prokaryotic genomes, our views about the nature of evolutionary forces leading to microbial diversification and speciation are changing (Doolittle 1999; Lawrence and Roth 1999). A notable proportion of the genes found in prokaryotic genomes appear to have been acquired by lateral gene transfer (LGT) from phylogenetically distant lineages (Doolittle and Logsdon 1998; Nelson et al. 1999). However, few comparative studies allow us to pinpoint sources or directions of transfers with precision. The gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR or HMG-CoA reductase, E.C.1.1.1.34) provides an ideal system with which to study LGT in detail, for several reasons. First, there are two homologous classes of the enzyme, one found mainly in eukaryotes and archaea (class 1) and the other found in bacteria (class 2) (Bochar, Staufacher, and Rodwell 1999). Identification of the class by sequence is unambiguous, as the two classes share less than 20% amino acid identity, and enzymes within a class are ~40% identical. It is thus very easy to detect events of LGT between bacteria and archaea and identify the donor and the recipient of the transfer. One very clear cut case of LGT was indeed identified from the complete genome sequence of Archaeoglobus fulgidus. This hyperthermophilic archaean harbors a bacterial class 2 HMGR instead of the class 1 enzyme which is found in other archaea. In some bacteria, most eukaryotes, and all archaea, HMGR is an essential enzyme and thus cannot be simply lost but only displaced. Displacement is the acquisition of a functionally homologous gene followed by the loss of the ancestral gene, so an enzyme is always present to fulfill the specific function. This simplifies the study of the evolutionary history of HMGR, as no period in which the enzyme is absent comes to confuse this history.

Not only is A. fulgidus HMGR of bacterial origin, but it also shows striking amino acid identity (61%) with the HMGR of the soil proteobacterium Pseudomonas mevalonii (Boucher and Doolittle 2000). These two homologous enzymes likely play very different metabolic roles in their respective cells. In all archaea (like A. fulgidus), HMGR catalyzes the first committed step of the mevalonate pathway (reductive decylation of HMG-CoA to mevalonate). This central biosynthetic pathway leads to isopentenyl pyrophosphate, which is, with its isomer dimethylallyl pyrophosphate, the universal precursor for the synthesis of isoprenoids (Qureshi and Porter 1981). Isoprenoids are quantitatively very important in archaea. The latter’s membrane lipids, which are a unique and characteristic taxonomic feature of this domain, are made up of glycerol (or more complex polyols) ether linked to isoprenoid alcohols (De Rosa, Gambacorta, and Gliozzi 1986). In contrast, the proteobacteria P. mevalonii does not use its HMGR in the biosynthesis of isoprenoids, but to biodegrade mevalonate by oxidative acylation. It can use this compound as the sole source of carbon and energy (Schier and Rodwell 1989).

Key words: HMG-CoA reductase, isoprenoid, lateral gene transfer, gene displacement, Thermoplasma, Archaeoglobus.
This is thus a case in which an essential anabolic archaeal enzyme was displaced by a biodegradative bacterial homolog sharing little amino acid sequence identity (less than 20% between HMGR classes).

We decided to investigate this metabolically important case of interdomain LGT. A survey of close relatives of *A. fulgidus* and of other archaea that live in similar environments was performed, searching for other events of LGT of the HMGR gene. Among the archaea surveyed were the Archaeoglobales (*Archaeoglobus veneeficus, Archaeoglobus lithothrophicus, Archaeoglobus profundus,* and *Ferroglobus placidus*) and the Thermococcales (*Thermococcus litoralis, Thermococcus stetteri,* and *Thermococcus celer*). Although occupying very different ecological niches, two species of acidophilic Thermoplasmatales (*Thermoplasma volcanium* and *Thermoplasma acidophilum*) were also inspected in order to obtain representation of all major euryarchaeal groups. Some of the few bacteria known to possess biochemical activity requiring HMGR were also surveyed.

Small ribosomal subunit (SSU) genes were sequenced from all species from which an HMGR gene sequence had been obtained. The large number of SSU sequences available for archaea and its widespread use as a phylogenetic marker make this gene ideal for the construction of a reference phylogeny. HMGR phylogenies were compared with SSU and other phylogenetic markers to identify cases of LGT and try to reconstruct the evolutionary history of the HMGR gene in archaea.

### Materials and Methods

#### Strains Used and DNA Extraction

For archaea, total genomic DNA was extracted from frozen cell pellets following the protocol of Charbonnier et al. (1995). For bacteria, genomic DNA was isolated from fresh cell cultures as described in Wilson (1998). The archaeal strains used are listed in table 1. The bacterial strains are from the following sources: *Staphylococcus carnosus* TM300 (Friedrich Gotz, Tubingen University, Germany), *M. mevalonii* (V. W. Rodwell, Purdue University), *Chloroflexus aurantiacus* j-10 (Reidun Sirevåg, University of Oslo, Norway).

#### Polymerase Chain Reaction (PCR), Cloning, and Sequencing

The amplification of 600–950 bp of the HMGR gene from genomic DNA was carried out in two steps: (1) amplification with class-specific degenerate primers designed with the help of an alignment of all available HMGR gene sequences, and (2) amplification with degenerate primers designed from the partial fragments previously amplified (5’ end) and *A. fulgidus* complete HMGR gene sequence (3’ end). This second round of amplification was only performed for archaea that displayed a class 2 HMGR to obtain more complete sequences for phylogenetic analysis (Archaeoglobales and Thermoplasmatales). The HMGR primer sequences were as follows: class 1 forward (HMGR.F2: 5’-CCAT-TGGCTACNACNGARGG-3’); HMGR.F3: 5’-GACGC-
For the Archaeoglobales, the internally transcribed spacer between the SSU and LSU genes were also sequenced for all Archaeoglobales species and ranged from 228 to 259 bp and all contained a tRNA-Ala gene of 72 bp in length with a UGC anticodon.

Part of the HMGR gene was amplified and sequenced from bacterial species known from biochemical studies to use the mevalonate pathway: *Staphylococcus aureus* and *C. aurantiacus* (Hefer et al. 1993; Horbach, Sahm, and Welle 1993). The HMGR gene was also isolated from species known to be found with *A. fulgidus* in nature (*T. celer, T. stetteri, T. litoralis, and Pyrococcus occultum*). These species were all found to harbor the expected class of HMGR (respectively, class 2 for the bacteria and class 1 for the archaea), and no attempt was made to obtain a larger proportion of the gene from these organisms. The HMGR genes of four Archaeoglobales (*A. profundus, A. lithothrophicus, A. veneficus* and *F. placidus*) and two Thermoplasmatales (*T. volcanium* and *T. acidophilum*) were first partially amplified (around 600 bp) and found to be of class 2 instead of the usual class 1 for archaea. A more substantial part of the HMGR gene was ampliﬁed and sequenced from the same cell pellet. To conﬁrm their origin, the HMGR gene fragments sequenced were radioactively labeled and used as probes for Southern hybridization on the genomic DNA from which they had been ampliﬁed.

Preliminary Analysis of the SSU and HMGR Gene Sequences

The SSU genes from *A. fulgidus, F. placidus,* and *T. acidophilum* were identical to those already present in the database. The *A. veneficus* SSU gene was found to different from the published sequence for the same strain. The original sequence has been found to originate from another strain of Archaeoglobus, which was separated from an enrichment culture of *A. veneficus* SNP6 with optical tweezers. GenBank entries have been updated, and the *A. veneficus* SNP6 SSU sequence from this study replaces the old entry (accession number Y10011).

**Archaeoglobus lithothrophicus, A. profundus, T. volcanium, and Pseudomonas mevalonii** SSU genes are novel sequences. The internally transcribed spacers between the SSU and LSU genes were also sequenced for all Archaeoglobales species and ranged from 228 to 259 bp and all contained a tRNA-Ala gene of 72 bp in length with a UGC anticodon.

Sequence Alignment and Phylogenetic Analysis

For phylogenetic analysis of HMGR, sequences were retrieved from the NCBI web site (http://www.ncbi.org). Preliminary sequence data from the unfinished genomes of *S. aureus, Enterococcus faecalis, Streptococcus pyogenes,* and *Streptococcus pneumoniae* were analyzed.
were obtained from the Institute for Genomic Research web site (http://www.tigr.org). Methanosarcina mazei (http://www.g2l.bio.uni-goettingen.de/), Pyrobaculum aerophilum (http://www.tree.caltech.edu/bes.html), and Halobacterium sp. NRC-1 (http://zdna.micro.umass.edu/haloweb/) HMGR sequences were provided by their respective genome-sequencing projects. The retrieved amino acid sequences and the new sequences from this study were aligned using CLUSTAL W (Thompson, Higgins, and Gibson 1994). The alignment was subsequently edited manually to remove gaps and ambiguous characters. The numbers of sites used in the HMGR alignments were as follows: 203 for class 1 and class 2, 253 for Archaeoglobales and Thermoplasmatales, and 335 for archaeal class 1.

To build a reference archaeal phylogeny and look at the relationship between Thermoplasmatales and Archaeoglobales, different phylogenetic markers were chosen. This choice was based on the availability of sequences from most archaeal groups. The selected molecules were as follows: SSU and LSU, archaeal release factor (aRF1), elongation factors 1 and 2 (ef1/EF-Tu and ef2/EF-G), and RNA polymerase subunits β and β’ (rpoB and rpoC). Markers that were protein-coding genes were aligned at the amino acid level, all available archaeal sequences were included, and the trees were rooted with crenarchaeotes. For SSU, an alignment of representative archaeal sequences was obtained from the European Small Subunit Ribosomal RNA Database (Van de Peer et al. 2000). New sequences were added to this alignment and local adjustments were performed manually. The alignment was also edited to exclude gaps and ambiguous characters. A total of 1,124 sites were used in the archaeal SSU alignment, which included sequences from 66 taxa, covering all euryarchaeal groups and including representative crenarchaeotes. A subset alignment containing only representatives of the Archaeoglobales and Thermoplasmatales was also assembled (1,372 sites used). The LSU alignment was obtained (De Rijk et al. 2000) and edited (2,218 sites used) by the same procedure. For comparison with the archaeal class 1 HMGR phylogeny, for which only nine taxa were available, the SSU and LSU alignments were reduced to include only the same nine taxa, plus A. fulgidus and T. acidophilum.

To evaluate the influence of taxon sampling on the SSU phylogeny, a random sampling analysis was performed (Van de Peer et al. 1994). In this random taxon sampling analysis, the SSU alignment was broken down into 10 different groups composed of four to eight taxa each based on recognized phylogenetic groups: Thermoplasmatales, Archaeoglobales, Thermococcales, Methanosaetales, Methanomicrobiales, Methanococcales, Methanobacteriales, Halobacteriales, and Methanopyrales (there was only one sequence in this group, which was from Methanopyrus kandleri), and Crenarchaeota (eight representative crenarchaeotes were chosen). A program based on (Van de Peer et al. 1994) was written to randomly choose one taxon from each of these groups and then produce a neighbor-joining (NJ) Jukes-Cantor distance tree (10 taxa, rooted with the crenarchaeal taxon). The procedure was repeated 1,000 times, giving a set of SSU trees, each including one randomly chosen taxon from each archaeal group. A consensus tree was built based on the frequency of occurrence of different groups of taxa in the 1,000 SSU trees.

DNA sequences were used for phylogenetic analysis of the genes coding for HMGR, SSU, and LSU. All DNA analyses of the HMGR gene included only the first and second codon positions, as a large GC composition bias could be detected at the third position for some taxa ($\chi^2 = 35.16, df = 12, P < 0.0005$). DNA analyses were performed with PAUP* 4.04b (Swofford 1998) applying the heuristic-search option and using the TBR branch-swapping algorithm. Logdet distance trees were reconstructed taking into account the proportion of invariable sites. Maximum-likelihood analyses were performed under the transitional model allowing for invariable sites and among-site rate variation (TIM+I+G) substitution model as selected using MODELTEST (Posada and Crandall 1998). Parsimony analyses were also performed with PAUP*, with ACCTRAN character trait optimization selected. Quartet-puzzling maximum-likelihood and bootstrapped quartet-puzzling maximum-likelihood distance trees were reconstructed with TREE-PUZZLE and PUZZLEBOOT, respectively. PUZZLEBOOT was provided by Andrew J. Roger and Michael E. Holder (http://www.tree-puzzle.de). TREE-PUZZLE settings were as follows: HKY85 nucleotide substitution model, rate heterogeneity model with gamma-distributed rates over eight categories plus one invariable category, and the α parameter and amino acid frequencies estimated from the data. For the large euryarchaeal SSU data set (66 taxa), each archaeal group mentioned above was constrained as one moveable group for phylogenetic tree reconstruction. The confidence of nodes in DNA analyses was estimated by 1,000 bootstrap replicates (PAUP* and PUZZLEBOOT) or 1,000 quartet-puzzling steps (TREE-PUZZLE).

Phylogenetic analyses were performed on amino acid sequences for HMGR, ef1, ef2, rpoB, rpoC, and aRF1. For distance analyses, Fitch-Margoliash trees of PAM-based distances were made using PROTDIST and FITCH from the PHYLIP package, version 3.572 (Felsenstein 1993). Protein parsimony analyses were performed using PAUP*, again with ACCTRAN character trait optimization selected. Maximum-likelihood analyses were carried out using PROTL (quick-add search with 2,000 replicates, JTT-F amino acid substitution model) from the MOLPHY package, version 2.3 (Adachi and Hasegawa 1996), and TREE-PUZZLE 4.0 (JTT-F amino acid substitution model, rate heterogeneity model with gamma-distributed rates over eight categories plus one invariable category, and the α parameter and amino acid frequencies estimated from the data). PUZZLEBOOT was used with the same settings as TREE-PUZZLE to bootstrap quartet-puzzling maximum-likelihood distance trees. The confidence of nodes in amino acid analyses was estimated by 1,000 bootstrap replicates (PROTDIST and PUZZLEBOOT), 1,000 quartet-puzzling steps (TREE-PUZZLE), or RELL values (PROTMLO). The bootstrap replicates of PROTDIST
were generated using SEQBOOT and compiled in a consensus tree with CONSENSE.

When required, different tree topologies were compared using the Kishino-Hasegawa (KH) test, which was performed in TREE-PUZZLE. The trees used for input were obtained using PROTMML (quick-add search with 2,000 replicates, JTT-F amino acid substitution model) for amino acid sequences and by a parsimony heuristic search for the most-parsimonious trees in PAUP* (ACCTRAN character trait optimization selected) for DNA sequences. When investigating the possibility of the Thermoplasmatales and Archaeoglobales being sister taxa, the 100 most likely or most parsimonious trees obtained when these taxa were constrained to form a clade were compared with the 100 trees obtained without this constraint.

Results
Classification of Sequenced HMGR Genes

Based on sequence alignment with known HMGRs, the bacteria S. aureus and C. aurantiacus were found to harbor the expected class 2 HMGR. Of the different archaea sampled, those sharing an environment with A. fulgidus without being its phylogenetic neighbors (T. celer, T. stetteri, T. litoralis, and P. occultum) were found to have the standard archaeal class 1 HMGR (table 1). All close relatives of A. fulgidus (A. profundus, A. lithotrophicus, A. veneficus, and F. placidus) and the two Thermoplasmatales sampled (T. volcanium and T. acidophilum) harbored an unusual class 2 bacterial HMGR (table 1).

Pairwise sequence comparison with SIM (Huang and Miller 1991) revealed an unusually high similarity between the HMGRs of the Thermoplasmatales, the Archaeoglobales, and P. mevalonii. The amino acid sequence identities between HMGRs from different species of the three lineages ranged from 53% to 66%, with an average of 56.5% (SD = 4.3). Comparatively, the average amino acid identity was 41.1% (SD = 9.1) among class 2 HMGRs and 18.9% (SD = 2.3) between classes 1 and 2 (only one representative of a given genus was included in the calculations in both cases). The amino acid identity between different members of a genus for class 2 HMGRs averaged 70.8% (SD = 1.6) for Streptococcus, 79.1% (SD = 9.7) for Archaeoglobus, and 80.3% for Thermoplasma.

Phylogenetic Analysis of Thermoplasmatales and Archaeoglobales HMGRs

Phylogenetic trees based on amino acid sequences of all class 2 HMGRs rooted with representative class 1 enzymes were constructed with a variety of methods. The best maximum-likelihood distance tree is shown in figure 1. As expected, class 1 and class 2 genes formed well-separated clades. However, Archaeoglobales and Thermoplasmatales clustered with bacteria rather than with other archaea in the HMGR tree. In fact, the HMGRs from these two archaeal groups were much more similar to the P. mevalonii enzyme than any other bacterial homologs. The clustering of these three genes was observed in all phylogenetic reconstructions. Pseudomonas mevalonii also occupied a basal position relative to the Archaeoglobales and the Thermoplasmatales in all analyses in which only class 2 sequences were included (fig. 1). When the HMGR tree was rooted with class 1 enzymes, P. mevalonii still grouped with Archaeoglobales and Thermoplasmatales, but the branching order was unresolved in distance and quartet-puzzling maximum-likelihood analyses (fig. 1).

The Kishino-Hasegawa test was used to compare all 124 maximum-likelihood trees obtained by PROTMML (quick-add search with 2,000 replicates). The best tree showed the topology presented in figure 1. However, of the 124 trees compared, 10 did not show P. mevalonii as basal to the two archaeal clades and were not significantly worse than the best tree. These alternative trees either placed P. mevalonii between the Thermoplasmatales and the Archaeoglobales clades (five trees) or with either one of them (in three trees with the Thermoplasmatales and in two trees with the
Archaoglobales). The Thermoplasmatales clade was present in all trees, and the Archaeoglobales were monophyletic in all but 8 of the 124 best trees. In each of these eight cases, it is A. profundus, a notably long branch among Archaoabiales, that is probably attracted to the Thermoplasmatales or P. mevalonii.

Are Thermoplasmatales and Archaeoglobales Sister Taxa?

No previous phylogenetic studies have shown Thermoplasmatales and Archaeoglobales to be sister taxa. To test the possibility, however, we examined trees constructed with different phylogenetic markers (table 2). For each marker, the Kishino-Hasegawa test was used to compare the 100 most likely trees obtained by a PROML quick-add search with another 100 trees obtained by a similar search with Thermoplasmatales and Archaeoglobales constrained to be sister taxa. None of the markers gave Thermoplasmatales and Archaeoglobales as sister taxa in its best tree. However, several trees that placed these groups as monophyletic were not significantly worse than the best tree for most markers (table 2). Thermoplasmatales and Archaeoglobales were found as adjacent but distinct branches in the best trees of LSU and aRF1. No other marker indicated these clades as being adjacent in their best trees.

With the large taxon sampling of SSU and well-defined clades for archaea, it was possible to perform a random taxon sampling analysis, in which single taxa are randomly selected from each of several predefined clades, trees are constructed, and results are averaged over 1,000 repetitions (described in Van de Peer et al. 1994). None of the 1,000 trees showed the Thermoplasmatales and the Archaeoglobales as sister taxa. The only nodes recovered in a significant number of trees were the one grouping the Methanosarcinaceae and the Methanomicrobiales (1,000 trees), the node grouping these two clades with Halobacteriales (969 trees), and the position of the Methanopyrales as the deepest euryarchaeal branch (878 trees) (fig. 2A).

An interesting characteristic was found in one of the phylogenetic markers used, RNA polymerase subunit β, encoded by rpoB. The gene is split into two smaller genes encoding independent subunits (B’ and B") in all euryarchaeal taxa from which it has been sequenced so far except Thermococcales and Thermoplasmatales (Schleper et al. 1995). Crenarchaeal rpoBs are not split. This gene is thus split only in Archaeoglobales and not in Thermoplasmatales. This split can occur without altering the RNA polymerase function, as it has been introduced experimentally in Escherichia coli rpoB without removing the activity of the encoded polymerase (Severinov et al. 1996).

Phylogenetic Relationships Among Archaeoglobales

The phylogeny of the five different cultured Archaeoglobales species was determined using the SSU gene, also including sequences from six uncultured taxa (fig. 3). Both unrooted trees and trees rooted with the Thermoplasmatales gave a highly significant clustering.

Table 2
Comparison of the Most Likely or Parsimonious Archaeal Trees of Different Phylogenetic Markers by the Kishino-Hasegawa (KH) Test

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. of Sites Used</th>
<th>No. of Taxa</th>
<th>KH Best Tree (log likelihood)</th>
<th>Trees Not Significantly Worse than Best Tree with TP and AG as Sister Taxa (log likelihood)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ef1</td>
<td>379 aa</td>
<td>11</td>
<td>-4,663.07</td>
<td>3 (-4,676.49 to -4,683.40)</td>
</tr>
<tr>
<td>ef2</td>
<td>684 aa</td>
<td>16</td>
<td>-11,954.78</td>
<td>4 (-11,971.65 to -11,997.06)</td>
</tr>
<tr>
<td>rpoB</td>
<td>976 aa</td>
<td>12</td>
<td>-13,708.58</td>
<td>12 (-13,732.23 to -13,750.49)</td>
</tr>
<tr>
<td>rpoC</td>
<td>1,080 aa</td>
<td>12</td>
<td>-15,351.32</td>
<td>34 (-15,360.47 to -15,396.61)</td>
</tr>
<tr>
<td>aRF1</td>
<td>335 aa</td>
<td>11</td>
<td>-5,396.24</td>
<td>21 (-5,404.18 to -5,425.82)</td>
</tr>
<tr>
<td>16S</td>
<td>1,124 nt</td>
<td>66</td>
<td>-13,514.43</td>
<td>61 (-13,519.24 to -13,538.03)</td>
</tr>
<tr>
<td>23S</td>
<td>2,218 nt</td>
<td>14</td>
<td>-15,218.4</td>
<td>1 (-15,237.79)</td>
</tr>
</tbody>
</table>

Note.—TP = Thermoplasmatasales; AG = Archaeoglobales; aa = amino acids; nt = nucleotides.

Fig. 2.—Phylogeny of the euryarchaeotes according to different genes. HMGR, SSU, and LSU trees are the best maximum-likelihood trees obtained in TREE-PUZZLE. (C) The node is recovered by all tree-reconstruction methods used (distance, maximum parsimony, maximum likelihood, quartet puzzling maximum likelihood, and quartet-puzzling maximum-likelihood distances). (D) The node is recovered by all tree-reconstruction methods used with a confidence of 90% or more. A. Representation of the frequency of occurrence of different groupings in the random taxon-sampling analysis on a total of 1,000 trees. B. Comparison of the LSU and class 1 HMGR phylogenies. C. SSU phylogeny.
Comparison of Phylogenies of the HMGR Gene and the Phylogenetic Markers SSU and LSU

To look for differences in their evolutionary histories, a phylogenetic tree of class 1 HMGRs was compared with trees of the commonly used phylogenetic markers SSU and LSU. Only taxa for which all three genes were available were included. Since no LSU sequence was available for *M. mazei*, we substituted a sequence from *Methanospirillum hungatei*, another representative of the large methanogen clade that includes *M. mazei*. LSU and SSU tree topologies were found to be slightly different from each other (fig. 2B and C). For the latter, however, the branching order of the groups varied substantially depending on the reconstruction method used, compared with a generally more stable topology with LSU (data not shown). Comparative analyses have shown that LSU seems less prone to variation due to changes in method and data set and is generally a more reliable phylogenetic marker (De Rijk 1995). LSU was thus chosen over SSU as a reference to which the archaeal HMGR phylogeny was to be compared.

Other than the fact that the HMGR tree was missing *A. fulgidus* and *T. acidophilum* (class 2 HMGRs were excluded from the analyses), its topology was found to be very similar to the LSU tree (fig. 2B).

Discussion

The common ancestor of archaea and eukaryotes almost certainly had a class 1 HMGR, since it is found in all but two archaeal groups (Thermoplasmatales and Archaeoglobales), as well as in all eukaryotes known to use the mevalonate pathway except for the diplomonad *Giardia lamblia* (which has a bacterial-like HMGR) (Boucher and Doolittle 2000). Knowing that class 1 HMGR is ancestral to archaea, it would be interesting to know how the Thermoplasmatales and Archaeoglobales acquired a bacterial-like class 2 enzyme. Two general evolutionary scenarios are discussed here.

Acquisition of the Bacterial-like HMGR in an Ancestral Euryarchaeote

In this scenario, the euryarchaeote ancestral to the Thermoplasmatales, the Archaeoglobales, and all later-branching lineages acquired a bacterial HMGR in addition to its endogenous class 1 enzyme. This acquisition would have been followed by differential losses in subsequently diverging lineages. The Thermoplasmatales and Archaeoglobales would have kept the class 2 enzyme and lost the class 1 homolog. All other lineages descending from the original recipient of the transfer would have kept only the class 1 enzyme or lost it and reacquired it later (fig. 4A and B).

Under a differential loss scenario, since Thermoplasmatales and Archaeoglobales occupy distinct intermediate branches in the rooted tree of the euryarchaeotes, several events would be needed to account for the presence of a different class of HMGR in these two lineages and subsequent ones (fig. 4A). Also, for differential loss to be possible, there must have been two HMGRs encoded in the genome of the ancestral euryarchaeote. This situation would have had to be maintained long enough for at least two lineages to diverge (Thermoplasmatales and Archaeoglobales). Since there is no known case of the presence of both classes of
HMGR in any single genome, it is unlikely that the two enzymes would have persisted together on such a large timescale.

Displacement of the ancestral enzyme by a class 2 homolog followed by a reacquisition of a class 1 HMGR after divergence of the Archaeoglobales and Thermoplasmales also seems unlikely (fig. 4B). The LSU phylogeny of euryarchaeotes (assumed here to model the organismal phylogeny) matches the class 1 HMGR phylogeny very well (fig. 2B). This is incompatible with the displacement/reacquisition scenario postulated above. The most likely donor for a reacquisition of a class 1 HMGR being another archaeon, the recipient would show an affinity for other archaeal lineages, thus leading to a topology fairly different from the one obtained with LSU.

If Archaeoglobales and Thermoplasmales are neither sister taxa nor adjacent branches in the euryarchaeal tree, the possibilities mentioned above become even less appealing. Although the possibility that Archaeoglobales and Thermoplasmales are sister taxa cannot be formally excluded, the evidence presented here weighs heavily against it. None of the best maximum-likelihood trees for the seven different phylogenetic markers used here placed these two archaeal orders as sister taxa. Furthermore, in a random taxon sampling analysis performed on the SSU gene including representatives of all archaeal orders, none of the 1,000 trees obtained placed Archaeoglobales and Thermoplasmales as sister taxa. Also supporting the results obtained with the phylogenetic markers is the fact that the rpoB gene is split into two smaller genes in Archaeoglobales and not in Thermoplasmales. This situation would require an unlikely evolutionary event to have occurred twice if these two archaeal orders are to be sister taxa. It is, however, likely that Archaeoglobales and Thermoplasmales comprise consecutive branches within the euryarchaeal tree. Two of the markers, LSU and aRF1, did indeed place T. acidophilum and A. fulgidus as adjacent but distinct branches in their trees, regardless of the reconstruction method used.
Overall, if the bacterial-like class 2 HMGR displayed by the Thermoplasmatales and the Archaeoglobales was acquired by an ancestral euryarchaeote, the evolutionary history of this gene would be very complex. An unparsimonious scenario of differential losses and/or gene displacements followed by reacquisition would have to be used to explain the presence of a class 2 HMGR only in two euryarchaeal lineages (Thermoplasmatales and Archaeoglobales).

Direct Acquisition of the Bacterial-like HMGR by the Thermoplasmatales and the Archaeoglobales

The alternative possibility to the acquisition of a class 2 HMGR in an ancestral euryarchaeote is LGT directly to the Thermoplasmatales, the Archaeoglobales, or both. Very few events would be required to explain the presence of bacterial-like HMGRs in the Thermoplasmatales and Archaeoglobales under this scenario. Two displacement events would be required if they are distinct lineages, as the analysis presented earlier suggested (fig. 4C and D). Acquisition of a class 2 HMGR either would have proceeded independently for each lineage (fig. 4C) or would have occurred in one of the lineages, with the new gene being spread to the other by a second LGT event (fig. 4D). Independent acquisition by the two lineages seems to us less likely, because the high similarity of Thermoplasmatales and Archaeoglobales HMGRs (60.7% average amino acid identity, SD = 3.1) would require the displacement of the same gene occurring in two different clades to have involved very similar donors. Furthermore, if the Thermoplasmatales and Archaeoglobales enzymes did not originate from the same bacterial donor, the P. mevalonii homolog would then be more similar to the HMGR from one of the lineages and thus cluster with either one of them. On the other hand, if either the Archaeoglobales or the Thermoplasmatales lineage acquired the enzyme first and then transferred it to the other lineage, the potential donor, here P. mevalonii, would have a basal position in a phylogenetic tree relative to those lineages. This basal position of P. mevalonii is indeed what is observed in the HMGR phylogeny (fig. 1), supporting a single interdomain event, followed by an exchange between the archaean clades. However, a small number of most likely HMGR trees (10 out of 124) did not display P. mevalonii as basal and were not significantly worse than the best tree (Kishino-Hasegawa test). This means that we cannot eliminate two separate interdomain LGT events as a possibility, even if it is less likely than an interdomain event followed by a second event between archaea. Both possibilities nonetheless correspond to a common origin for the HMGR found in Thermoplasmatales and Archaeoglobales involving either a single donor or two closely related donors. The exact identity of the donor of the bacterial HMGR remains unknown. It is unlikely to be P. mevalonii itself, an aerobic mesophile that does not share a habitat with either the hyperthermophilic Archaeoglobales or the thermoacidophilic Thermoplasmatales and most likely acquired its own HMGR by LGT (Boucher and Doolittle 2000; Wilding et al. 2000).

More Exchange at the Species Level for the Archaeoglobales HMGR Gene?

The HMGR trees of Archaeoglobales, both DNA and amino acid, differ from the SSU tree with regard to the position of A. profundus and A. fulgidus (fig. 3). Duplication of all DNA isolation, PCR amplification, and DNA sequencing for each gene sequence obtained greatly reduces the risk of this inconsistency being caused by a mix-up of the HMGR/SSU sequences from different Archaeoglobus species. Furthermore, the sequences for HMGR and SSU from the A. fulgidus genome were confirmed here by PCR amplification of these genes from genomic DNA of the same strain as the one used in the genome sequencing project (VC-16). The inconsistency between the HMGR and the SSU trees of Archaeoglobales is thus due to either a gene exchange of SSU or HMGR among Archaeoglobales or a methodological problem with phylogenetic tree reconstruction. The branching position difference of A. fulgidus and A. profundus between the SSU and HMGR trees seems well supported by the different tree reconstruction methods (fig. 3). However, the HMGR phylogeny does not seem reliable because of an insignificant difference between the best HMGR tree for Archaeoglobales and an alternative topology that agrees with the SSU tree (as shown by the Kishino-Hasegawa test). Therefore, we cannot say with certainty whether or not LGT of HMGR has occurred between different Archaeoglobales species.

Gene exchange between closely related species is usually difficult to establish because of the lack of variability among sequences. Its frequency should, however, be higher than that between distant organisms, since less barriers and more vectors are present (Ochman, Lawrence, and Groisman 2000).

Essential Genes and Homologous Gene Displacement

LGT is often postulated to be unlikely for essential genes (Lawrence 1999). Two arguments are usually offered to support this assumption. The first is that the products of native genes have experienced long-term coevolution with the rest of the cellular machinery and are thus unlikely to be displaced. The presence of a bacterial-like HMGR in both the Archaeoglobales and the Thermoplasmatales shows that even a very divergent homolog can perform the essential function of its predecessor in a foreign genome. The second argument usually brought forward for limited frequency of LGT for essential genes is that the recipient taxa would already have a functional homolog. In the case presented here, although “accidental displacement” (where both homologs are more or less equivalent) is a possibility, the acquisition and maintenance of HMGRs from a common bacterial origin by two divergent archaeal lineages suggests an important adaptive role for these events of gene displacement.
One possible source of selective pressure that could have driven LGT in this case would be the presence of compounds that differentially inhibit the two classes of HMGR in the habitats of the Archaeoglobales and the Thermoplasmatasae. Naturally occurring statins (like mevinolin) are an example of such compounds, inhibiting class 1 HMGRs with an affinity four orders of magnitude greater than class 2 enzymes (inhibition constant \( K_i \) class 1 [5–15 nM] vs. \( K_i \) class 2 [0.15–0.20 mM]; Kim, Stauffacher, and Rodwell 2000). Archaea that harbored a class 2 HMGR in addition to their an- cestral enzyme would have been favored in such an environment. The selective pressure would have had to be present for a period sufficiently long to get the novel HMGR gene fixed in the population and to lose the an- cestral enzyme. As the selective pressure is very high on essential genes like HMGR, the fixation time could be relatively short. This hypothetical scenario illustrates the case with which an essential enzyme could be displaced by a homolog. It also suggests that gene displacement might be a common mode of evolution for essential genes, since these genes would be particularly sensitive to selective pressure and could only be relieved by an intragenomic adaptation (gene duplication, mutation, intragenic recombination) or gene displacement (intergenomic adaptation).

Acknowledgments

We thank J. B. Dacks, A. J. Roger, J. M. Archibald, and C. L. Nesbø for critical reading of the manuscript and helpful comments and suggestions. We are grateful to F. Gotz, R. Sirevaag, V. W. Rodwell, S. T. Fitz-Gibbon, A. Johann, S. P. Kennedy, and Göttingen Genomics Laboratory for providing prokaryotic strains, genomic DNA, and DNA sequences from unfinished genome sequencing projects. We also thank N. Hu for help with computer programming. This work was supported by the Medical Research Council of Canada. Y.B. was supported by the National Sciences and Engineering Council of Canada.

LITERATURE CITED


Felsenstein, J. 1993. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.


JULIAN P. ADAMS, reviewing editor

Accepted March 26, 2001