

Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea

Shijulal Nelson-Sathi^a, Tal Dagan^b, Giddy Landan^{a,b}, Arnold Janssen^c, Mike Steel^d, James O. McInerney^e, Uwe Deppenmeier^f, and William F. Martin^{a,1}

^aInstitute of Molecular Evolution, ^bInstitute of Genomic Microbiology, ^cMathematisches Institut, Heinrich Heine University, 40225 Düsseldorf, Germany; ^dBiomathematics Research Centre, University of Canterbury, Private Bag 4800, Christchurch, New Zealand; ^eDepartment of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland; and ^fInstitute of Microbiology and Biotechnology, University of Bonn, 53115 Bonn, Germany

Edited* by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved October 25, 2012 (received for review May 29, 2012)

Archaeobacterial halophiles (Haloarchaea) are oxygen-respiring heterotrophs that derive from methanogens—strictly anaerobic, hydrogen-dependent autotrophs. Haloarchaeal genomes are known to have acquired, via lateral gene transfer (LGT), several genes from eubacteria, but it is yet unknown how many genes the Haloarchaea acquired in total and, more importantly, whether independent haloarchaeal lineages acquired their genes in parallel, or as a single acquisition at the origin of the group. Here we have studied 10 haloarchaeal and 1,143 reference genomes and have identified 1,089 haloarchaeal gene families that were acquired by a methanogenic recipient from eubacteria. The data suggest that these genes were acquired in the haloarchaeal common ancestor, not in parallel in independent haloarchaeal lineages, nor in the common ancestor of haloarchaeans and methanosarcinales. The 1,089 acquisitions include genes for catabolic carbon metabolism, membrane transporters, menaquinone biosynthesis, and complexes I–IV of the eubacterial respiratory chain that functions in the haloarchaeal membrane consisting of diphytanyl isoprene ether lipids. LGT on a massive scale transformed a strictly anaerobic, chemolithoautotrophic methanogen into the heterotrophic, oxygen-respiring, and bacteriorhodopsin-photosynthetic haloarchaeal common ancestor.

Halophilic archaeobacteria (Haloarchaea) require concentrated salt solutions for survival and can inhabit saturated brine environments such as salt lakes, the Dead Sea, and salterns (1). In rRNA and phylogenomic analyses of informational genes, Haloarchaea always branch well within the methanogens (2–4). Haloarchaea can thus be seen as deriving from methanogen ancestors, but the physiology of methanogens and halophiles could hardly be more different. Methanogens are strict anaerobes, most species are lithoautotrophs that use electrons from H₂ to reduce CO₂ to methane (obligate hydrogenotrophic methanogens), thereby generating a chemiosmotic ion gradient for ATP synthesis in their energy metabolism, although some species can generate methane from reduced C₁ compounds, or acetate in the case of acetoclastic forms (5–7). Their carbon metabolism involves the Wood–Ljungdahl (acetyl-CoA) pathway of CO₂ fixation (5–7). In contrast, Haloarchaea are obligate heterotrophs that typically use O₂ as the terminal acceptor of their electron transport chain, although many can also use alternative electron acceptors such as nitrate in addition to light harnessing via a bacteriorhodopsin-based proton pumping system (8). The evolutionary nature of that radical physiological transformation from anaerobic chemolithoautotroph to aerobic heterotroph is of interest.

Many individual reports document that lateral gene transfer (LGT) from eubacteria was involved in the origin of at least some components of haloarchaeal metabolism. These include the operon for gas vesicle formation, which allows Haloarchaea to remain in surface waters (9), the newly identified methylaspartate cycle of acetyl-CoA oxidation (10), various components of the haloarchaeal aerobic respiratory chain (11–18), and proteins

involved in the assembly of FeS clusters (19). The sequencing of the first haloarchaeal genome over a decade ago identified some eubacterial genes that possibly could have been acquired by lateral gene transfer (11, 20), and whereas substantial data that would illuminate the origin of haloarchaeal physiology have accumulated since then, those data have not been subjected to comparative evolutionary analysis. Investigating the role of the environment in haloarchaeal genome evolution, Rhodes et al. (21) recently showed that Haloarchaea are indeed far more likely to acquire genes from other halophiles, but they did not address the issues at the focus of our present investigation, namely: How many eubacterial acquisitions are present in haloarchaeal genomes? How was the physiological transformation of methanogens to Haloarchaea affected by LGT? Do those acquisitions trace to the haloarchaeal common ancestor as a single acquisition or not?

To discern whether the eubacterial genes in haloarchaeal genomes are the result of multiple independent transfers in individual lineages or the result of a single ancient mass acquisition, here we have analyzed 10 sequenced haloarchaeal genomes—*Haloarcula marismortui* (22), *Halobacterium salinarum* (23), *Halobacterium* sp. (20), *Halomicrobium mukohataei* (24), *Haloquadratum walsbyi* (25), *Halorhabdus utahensis* (26), *Halorubrum lacusprofundi* (27), *Natrialba magadii* (28), *Natronomonas pharaonis* (29), and *Haloterrigena turkmenica* (30)—in the context of 65 other archaeobacterial and >1,000 eubacterial reference genomes.

Results and Discussion

We first clustered the 172,531 proteins encoded in the chromosomes of 75 archaeobacterial genomes into families using the standard Markov cluster (MCL) procedure (31) yielding 16,061 protein families. Comparison with 1,078 completely sequenced eubacterial genomes delivered 1,479 protein families that are present in at least two Haloarchaea and contain archaeobacterial and eubacterial homologs (Fig. 1A). Gene trees for the protein families were reconstructed using maximum likelihood inference (Methods).

Of 1,479 trees, 1,089 (73%) uncovered Haloarchaea as monophyletic and rooting within (or branching next to) eubacterial rather than archaeobacterial homologs (Fig. 1B). For 414 of these trees, no homologs at all were detected in nonhalophilic

Author contributions: T.D., U.D., and W.F.M. designed research; S.N.-S. and T.D. performed research; S.N.-S., G.L., A.J., M.S., and J.O.M. analyzed data; and S.N.-S., G.L., and W.F.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: bill@hhu.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209119109/-DCSupplemental.

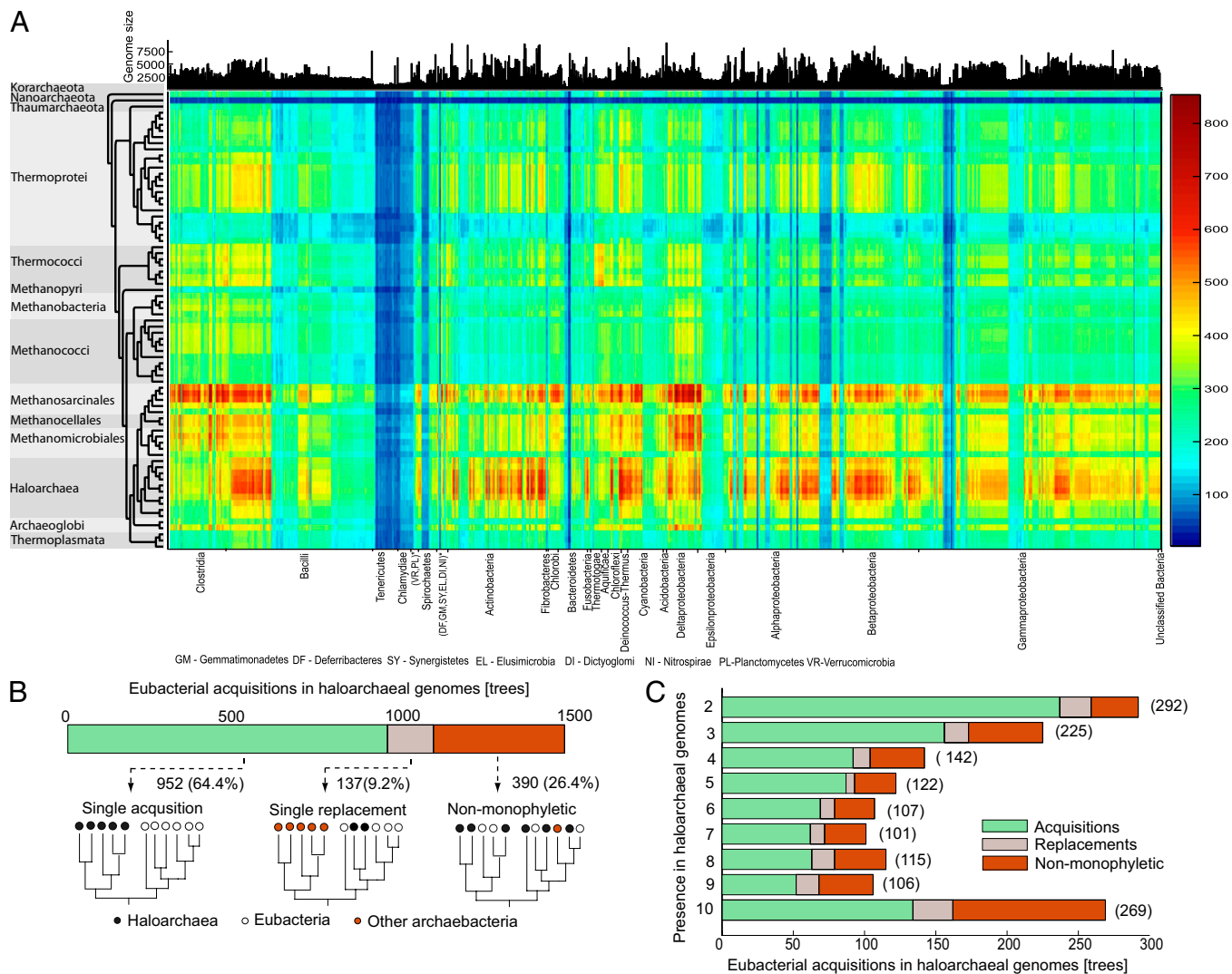


Fig. 1. (A) Number of shared genes between 1,078 bacterial genomes and 75 archaeobacterial genomes. (B) Types of phylogenetic trees obtained with respect to the relationship of Haloarchaea, nonhalophilic archaea, and eubacterial genes. (C) Types of phylogenetic trees detailed by the number of haloarchaeal taxa.

archaeobacteria. An additional 538 families had only very distant homologs (E values $>10^{-10}$ or amino acid identity $<30\%$) in some nonhalophilic archaeobacteria, together we designate these 952 cases as “acquisitions.” An additional 137 genes yielded trees in which Haloarchaea branch within eubacteria to the exclusion of readily detectable archaeobacterial homologs, we designate these genes as “replacements”; acquisitions and replacements we designate collectively as “imports” (Fig. 1B). The 390 cases of Haloarchaea nonmonophyly included 76 trees in which one haloarchaeon branched deviantly and 105 trees in which the Haloarchaea were split into two groups of two or more species. Because LGT is common in prokaryotes (32, 33), among haloarchaeans in particular (21), these 181 gene trees could well depict secondary transfers into or from the Haloarchaea.

Single Ancestral Acquisition. Are the 1,089 eubacterial imports in haloarchaeal genomes the result of a single ancestral acquisition or multiple parallel acquisitions? Monophyly alone does not completely decide the issue, because it is possible that a bacterial gene could be acquired recently in one haloarchaeal lineage and then passed around to other Haloarchaea by LGT. Such a process could, in principle, also generate monophyly for imported genes in a phylogenetic tree. However, in that case, individual

gene trees for imported genes would be very different from one another as opposed to the case of single acquisition, where trees for imports should be the same due to vertical inheritance from the haloarchaeal common ancestor. Moreover, trees for ancestrally acquired eubacterial imports should not only be similar to each other, they should also be similar to trees for endogenous haloarchaeal genes that are shared only with other archaeobacteria, which we call recipient genes. There are 364 haloarchaeal recipient genes that are present as single copies in all 10 Haloarchaea sampled and 109 haloarchaeal imports that are present as single copies in all 10 Haloarchaea (Fig. 2A), providing comparable tree sets. To avoid oversampling, the *H. salinarum* and the *Halobacterium* sp. genomes were condensed to one genome, because they share almost exactly the same genes and would have skewed the test by enhancing the congruence of the two sets.

Comparing the distributions of phylogenetic splits observed in the 364 recipient trees and the 109 imported trees containing all 10 (condensed to 9) Haloarchaea shows that the two sets exhibit a very similar phylogenetic signal (Fig. 2B). The six most common splits in the two sets of trees are identical and comprise 51% and 46% of the splits in the two sets, respectively. Moreover, these six splits exactly define the haloarchaeal phylogeny

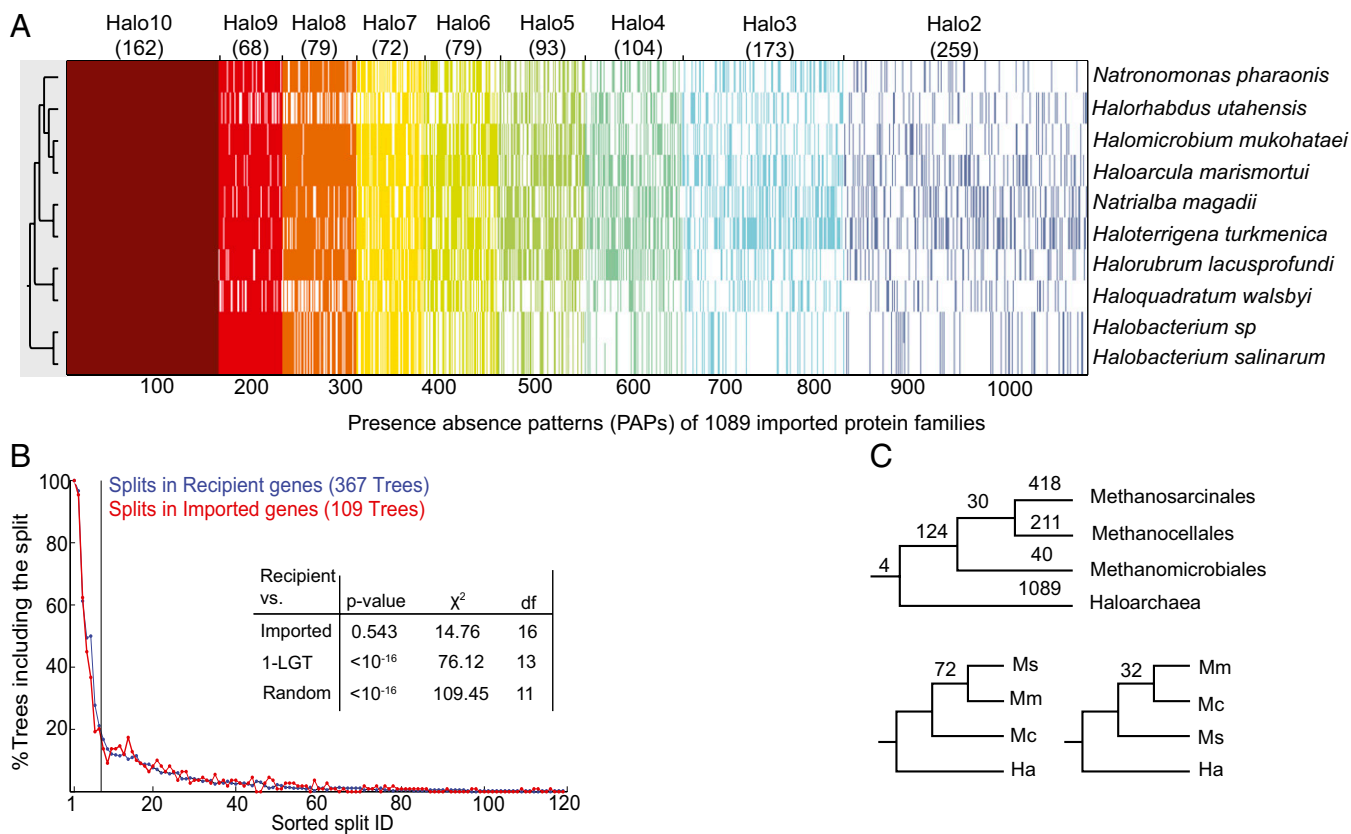


Fig. 2. Eubacterial genes in Haloarchaea. (A) Distribution of eubacterial imports present in at least two Haloarchaea. (B) Histogram of phylogenetic splits in imported and recipient trees. (Inset) Statistical test supporting single acquisition of imported eubacterial genes into the haloarchaeal common ancestor (*Methods*). df, degrees of freedom. Note that incompatible split frequency correlates with topological distance to the reference tree ($P = 7 \cdot 10^{-13}$ for recipient genes, $r = 0.76$; $P = 7 \cdot 10^{-19}$ for the imports), as expected for phylogenetic errors but not for competing biological signals (*SI Text*). (C) Numbers of eubacterial acquisitions and replacements in the ancestors of haloarchaea (Ha), methanosarcinales (Ms), methanomicrobiales (Mm), and methanocellales (Mc) shown for the reference topology in Fig. 1A and for the alternative topologies with respect to the Ms/Mm/Mc branching order (for unlabeled branches, the number of imports is identical with that shown for the reference topology; numbers of acquisitions and replacements are given in *SI Text*). Note that the uncultured methanogenic archaeon RC-1 is not classified with methanocellales in GenBank taxonomy, but it branched with *Methanocella paludicola* SANA in our reference topology, for which reason it was treated as an Mc member here. The frequency distributions of eubacterial imports across genomes and functional categories for haloarchaea is given in *Table S1*. The numbers of acquisitions and replacements, respectively, for the numbers of imports shown in C are 4: (4, 0); 124: (101, 23); 30: (22, 8); 418: (373, 45); 211: (141, 70); 40: (30, 10); 1,089: (952, 137); 72: (58, 14); and 32: (17, 15). For the methanogens in C, all species names and corresponding frequency distributions for functional categories are given in *Table S3*.

generated by 56 universally distributed archaeobacterial genes (Fig. 1A, Left). To test the statistical significance of this evidence in favor of single acquisition, we used a goodness-of-fit test to compare the distributions of topologies in the two tree sets. The null hypothesis is that the two samples of trees are drawn from the same distribution, whereas the alternative hypothesis is that the two samples differ in their distribution (*Methods*). The test's P value was 0.543, meaning that the null hypothesis could not be rejected (Fig. 2B, Inset). To complement this result, we examined two additional sets of trees. One set consisted of 109 random trees, and the second consisted of the observed 109 single copy imported gene trees subject to one LGT rearrangement (one random prune and graft operation) each. The latter case of one LGT rearrangement represents the slightest possible LGT-induced deviation from the null hypothesis of single acquisition in the haloarchaeal common ancestor followed by vertical evolution. Both sets were tested against the recipient trees and found to be significantly different (P values $<<10^{-16}$; Fig. 2B, Inset), strongly rejecting the one LGT rearrangement per gene case.

When we include the 53 multiple copy genes that are present in all 10 genomes, the one LGT rearrangement per gene is also excluded, although the significance (P values $<10^{-8}$, see *SI Text*) drops. That drop is expected, however, because horizontal

transfer, not duplication, drives the expansion of gene families in prokaryotes (34), hence the inclusion of multicopy genes preferentially includes genes for which LGT is more prevalent. We note that the goodness-of-fit test does not exclude one LGT for each gene, up to 34% of the 109 single-copy recipient trees can accept a single random prune and graft operation without the test rejecting the one LGT rearrangement case for the recipient set as a whole. However, for the 162 genes that are present in all 10 genomes, the possibility that the majority of imported genes are monophyletic because of import into one of the haloarchaeal lineages and subsequent passing around of the same gene can be excluded.

For the imports present in eight or fewer haloarchaeal genomes, excluding the (perhaps unlikely) possibility that monophyly is not due to acquisition in the haloarchaeal common ancestor but to lineage-specific acquisition and subsequent spread, is more difficult, mainly for reasons of sample size. The goodness-of-fit test based on split distributions cannot be used because few comparisons yield identical leaf sets for import vs. recipient trees. For the ≤ 8 -species cases, we therefore developed a less direct test, comparing the sets of recipient and import trees via their phylogenetic compatibility with the recipient trees for the 10 haloarchaeal species (*Methods*). Here, too, the null hypothesis of common

ancestry for import and recipient genes could not be rejected in any of the ≤ 8 -species cases, although the acquire-and-spread scenario was also not rejected for the 4-, 5-, and 6-species single and multiple copy cases (268 imports total; *SI Text*). Given that (i) the conventional interpretation of monophyly is presence in the common ancestor, that (ii) the 151 eight and seven species cases reject the acquire-and-spread scenario (*SI Text*) as an alternative explanation of monophyly, and that (iii) the data that most directly address the acquire-and-spread scenario—the 162 eubacterial imports present in all 10 genomes—most strongly reject it (Fig. 2*B*), the simplest interpretation of monophyly for the 1,089 imports is that their origin traces to a single acquisition in the haloarchaeal common ancestor followed mainly by vertical descent and widespread differential loss, with some subsequent LGT among haloarchaea (21, 32, 33), notably for multicopy genes (34), not being excluded.

Methanogens Are Affine for Eubacterial Genes. As seen in Fig. 1*A*, not only the 10 Haloarchaea, but also the five Methanosarcinales (Ms), the two Methanocellales (Mc), and the five Methanomicrobiales (Mm) sampled share many genes with eubacteria, raising the question of when these imports entered these methanogen lineages. Repeating our phylogenetic analyses for these groups (Fig. 2*C*) reveals that merely four eubacterial imports (three predicted membrane proteins and a glycosyl transferase) can be traced to their common ancestor, and that these are present in at most 6 of the 22 descendant genomes. Whereas 124 imports can be traced to the Ms/Mc/Mm common ancestor, these imports are also sparsely distributed, with only two (COG1032, an FeS-oxidoreductase and COG1387, histidinol phosphatase) being present in all 12 descendant methanogens. This contrasts to the 1,089 haloarchaeal imports that are specific to the haloarchaeal lineage, 162 of which (15%) have been retained in all 10 haloarchaeans sampled. The Ms, Mc, and Mm lineages have—like the haloarchaea—acquired hundreds of eubacterial genes, but the crucial observation is that they have remained strict anaerobes, and they have furthermore remained obligatory methanogenic (5–7). In stark contrast, the halophiles became aerobic heterotrophs and lost methanogenesis altogether. Collectively, the data point to a very different nature of the gene acquisition process in the halophiles and methanogens sampled here.

Donor Lineages. The acquisition of >1,000 genes is reminiscent of massive gene acquisitions surrounding the origin of mitochondria (35, 36) or plastids (37, 38). From what donor were these genes acquired? Because bacterial chromosomes undergo gene influx and gene export over time, it is unlikely that any one contemporary bacterial lineage would emerge as the donor of all eubacterial genes in haloarchaeal chromosomes (36, 39). All of the higher level taxa sampled appear as the sole sister group to the haloarchaeal gene or appeared in a sister group of mixed phylogenetic composition, as one might expect due to frequent LGT among bacteria (Figs. S1 and S24). The most frequent apparent donor lineage was the actinobacteria with 131 occurrences as the sole taxon in the sister group to Haloarchaea and 169 occurrences in the mixed sister group cases, followed by α -proteobacteria (88 sole plus 97 mixed), γ -proteobacteria (51 sole plus 111 mixed), and δ -proteobacteria (53 sole plus 100 mixed).

Function of Imported Genes. Trees generated from 56 recipient genes present as a single copy in all archaeobacteria place the Haloarchaea branching from within the methanogens, but not specifically as sisters to the Methanosarcinales (Fig. 1). Rather, the Haloarchaea appear to have emerged from simpler and more primitive methanogens, ones that lack both cytochromes and methanophenazine (5). Methanogens that lack cytochromes and methanophenazine are capable only of H_2 -dependent methanogenesis, and have a single coupling site in their energy

metabolism (5, 40). Haloarchaea have a respiratory chain with several coupling sites (1). Methanogens are strict autotrophs and strict anaerobes (5), whereas Haloarchaea are heterotrophs and can use O_2 as their terminal acceptor. Thus, the essential metabolic functional units for transforming a methanogen into the haloarchaeal common ancestor are (i) membrane transporters for reduced carbon compounds; (ii) a heterotrophic carbon metabolism that directs the oxidation of organic substrates to support carbon and energy metabolism; (iii) a respiratory chain for terminal oxidation and chemiosmotic ion pumping; and (iv) genes for the synthesis of any additional cofactors required, for example menaquinone, the quinone universally present in all halophiles (41). Those four essential functional units are very clearly represented within the eubacterial imports in haloarchaeal genomes.

Among the 1,089 haloarchaeal imports from eubacteria almost half (482, 44%) of the imports are related to metabolism, with amino acid transport and metabolism (114) and energy conversion (95) being the most abundant classes, followed by inorganic ion transport and metabolism (86) (Table S1; Fig. S2*B* and *C*). Whereas methanogens without cytochromes grow on gases, which traverse membranes freely without transporters, Haloarchaea abound in eubacterial transporters: 157 of the acquired families are annotated as permease, importer, or transporter. Although the true substrate spectrum of these transporters is yet unknown, 49 trace to amino acid or carbohydrate metabolism (Tables S1 and S2), and they operate in a membrane consisting of typical archaeobacterial lipids (1).

Methanogens cannot use exogenous carbohydrates for growth (5, 42); their sugar synthetic pathways are anabolic, whereas carbon metabolism in Haloarchaea runs in the catabolic direction. For a methanogen to become heterotrophic, it needs to acquire the enzymes underpinning the heterotrophic lifestyle from a heterotrophic donor (43). Among the eubacterial genes imported into Haloarchaea are pyruvate kinase, glucose-6-phosphate isomerase, phosphoglyceromutase, 6-phosphogluconate dehydrogenase, the eubacterial type fructose 1,6-bisphosphatase, as well as genes for 2-keto-3-deoxy-6-phosphogluconate aldolase of the Entner–Doudoroff pathway. Eubacterial enzymes of pyruvate breakdown were also found, including two copies of pyruvate:ferredoxin oxidoreductase, and genes for pyruvate dehydrogenase complex E1 and E2 subunits.

Earlier studies showed that five haloarchaeal respiratory chain components are eubacterial acquisitions in two Haloarchaea (15). Fig. 3 shows that most of the 11 subunits of NADH dehydrogenase (complex I) are present in all 10 Haloarchaea. Complexes I–III require quinones. Haloarchaea possess the naphthoquinone menaquinone (41) and several of the imported genes are involved in menaquinone biosynthesis, including menA. Finally, among the imported genes, 26 are annotated as transcriptional regulators and 8 are annotated as chaperones, including members of the DnaJ family.

Conclusion

Were these 1,000 genes accrued in the haloarchaeal ancestor one by one or in a single mass acquisition? The former possibility is unlikely, because in the absence of corresponding interaction partners to form functional complexes, individual protein subunits of catabolic carbon metabolism, the respiratory chain, or cofactor biosynthesis lack selectable function, which would allow them to become fixed in a methanogenic recipient. This argues in favor of mass transfer of genes for the entire pathways and complexes over a short period of evolutionary time. The origin of Haloarchaea was thus an evolutionary leap that transformed a methanogenic host into an oxygen-respiring heterotroph—the founder haloarchaeon. A possible context of that cellular association is anaerobic syntrophy (44, 45), that is, a H_2 -producing heterotrophic bacterial donor in association with a H_2 -

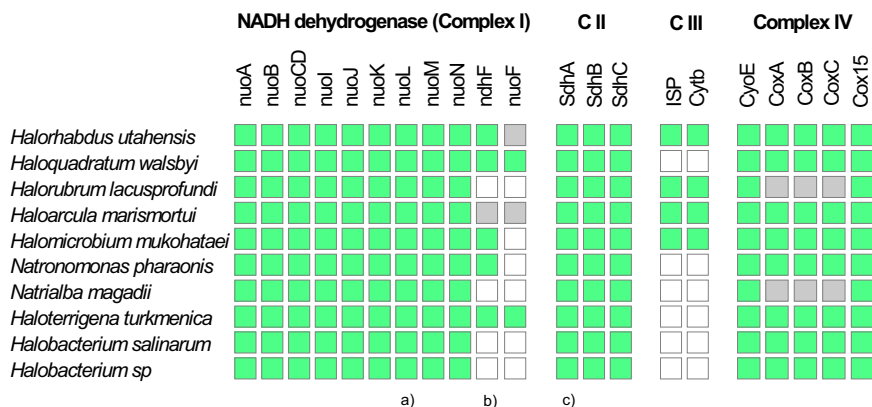


Fig. 3. Eubacterial respiratory chain components in Haloarchaea. Green boxes indicate presence of the gene in the corresponding Haloarchaea genome and that the gene is more similar to eubacterial than to archaeobacterial homologs in the corresponding phylogenetic trees. Gray boxes indicate that homologs can be detected in the corresponding genome by BLAST searches, but that the clustering procedure did not included them within the 16,061 archaeal clusters. White boxes indicate that no homolog was detected. (A) Haloarchaeal nuoL sequences are monophyletic but an additional paralogous copy is present in *Halorhabdus*. (B) *Salinibacter* has acquired a copy of ndhF from Haloarchaea, which are otherwise monophyletic. (C) Haloarchaeal sdhA sequences are monophyletic but additional paralogous copies of eubacterial origin are present in several genomes (see also Table S4).

dependent methanogenic recipient. Anaerobic syntrophy is common in nature and has been suggested as the selective force at the origin of eukaryotes (43, 46). If similar processes underlie the origin of haloarchaea and eukaryotes, why did Haloarchaea remain prokaryotic, whereas eukaryotes became complex? The main physiological difference between Haloarchaea and eukaryotes concerns the location of the bioenergetic membrane. In Haloarchaea it is the archaeobacterial plasma membrane (1). In eukaryotes it is the mitochondrial inner membrane—the key to eukaryote genome complexity (47). Mitochondria afforded ancient eukaryotes many orders of magnitude more energy per gene than their prokaryotic ancestors. That boost surmounted the energetic constraints imposed by reliance upon the cytoplasmic membrane as the source of chemiosmotic potential, thus allowing eukaryotic genomes and proteomes to expand freely, resulting in eukaryotic cell complexity (47). The Haloarchaea have long figured into issues of early microbial evolution (48). From the standpoint of genome chimaerism, they now appear to have undergone the very same physiological transformation as the eukaryotes, and the kind of gene transfer involved—from symbionts to the host chromosomes—is still ongoing in eukaryotic cells today (49). Haloarchaea remained prokaryotic because they failed to preserve a genome-containing bioenergetic organelle.

Methods

Data. Completely sequenced genomes of 1,153 microbial species were downloaded from the National Center for Bioinformatics Information (NCBI) website (www.ncbi.nlm.nih.gov). This includes 75 archaeobacterial genomes (version April 2010) and 1,078 eubacterial genomes (version September 2010). Taxonomic classification of the species was downloaded from the NCBI Taxonomy database (www.ncbi.nlm.nih.gov/Taxonomy/).

Clusters of Homologous Proteins. Clusters of homologous proteins were reconstructed from a total of 172,531 proteins encoded within the archaeal chromosomes. An all-against-all genomes BLAST (50) yielded 147,071 reciprocal best BLAST hits (rBBH) (51) using E value $<10^{-10}$ and $\geq 30\%$ amino acid identity as a threshold. Protein pairs were globally aligned using the Needleman–Wunsch algorithm with *needle* program (EMBOSS package) (52). A total of 137,022 protein pairs having global amino acid identities $\geq 30\%$ were clustered into protein families using the MCL algorithm (31) with default parameters. This yielded a total of 16,061 archaeal protein families of ≥ 2 proteins. The remaining 35,509 proteins were classified as singletons. Eubacterial homologs to archaeal proteins were found using an rBBH analysis as described above, which yielded 8,451 archaeal protein families having one or more eubacterial homologs. The functional classification of protein families was based on the eukaryotic orthologous groups database (KOG) database (53). Protein families that overlapped with KOG clusters were annotated to the same function as the matching KOG. The remaining protein families were manually classified by sequence similarity to known KOGs using the KOGnitor tool (<http://www.ncbi.nlm.nih.gov/COG/grace/kognitor.html>). The haloarchaeal respiratory chain component genes were identified from the Kyoto Encyclopedia of Genes and Genomes database (<http://www.genome.jp/kegg/>).

Phylogenetic Trees. Protein families were aligned using MAFFT (multiple alignment using fast Fourier transform) (54), and trees were reconstructed using Phylml (55) with the best fitting model in individual trees as inferred by ProtTest3 (56) using the AIC measure. An archaeobacterial reference tree was reconstructed from a weighted concatenated alignment of 56 archaeobacterial single copy universal genes using Phylml with the IG+I+G model, which was the most frequent best fitting model, rooted using *Nanoarchaeota* and *Koarchaeota* as an outgroup. Trees of recipient genes were reconstructed from sequences of all 10 Haloarchaea and one nonhaloarchaeal sequence using the same procedure. For polarizing the direction of gene transfers, the root of Jain et al. (33) was used.

Reconstruction of Lateral Gene Transfer Events. Eubacterial acquisitions within halophilic archaeal genomes were identified by presence absence pattern (PAP) analysis and BLAST protein sequence similarity searches. Of the total 8,451 bacterial-like protein families in archaeobacteria 1,479 had ≥ 2 Haloarchaea species. Of these, 952 do not possess other nonhaloarchaeal homologs in the same families and correspond to unique acquisitions within Haloarchaea from eubacterial species. Archaeobacterial xenologous genes that were replaced by a eubacterial acquisition are expected to be more similar to their eubacterial ancestors than to their orthologs in other archaeobacterial species (57). Putative replaced halophilic proteins were identified by comparing the E value of their BBHs within eubacterial and archaeobacterial genomes. Proteins having a eubacterial BBH of lower E value than that of the archaeobacterial BBH were classified as putative acquisitions from eubacteria, corresponding to 527 protein families. All 1,479 protein families were aligned with their eubacterial homologs including the three best eubacterial hits per archaeobacterial protein (but excluding redundant eubacterial sequences), and phylogenies were reconstructed as described above. The trees were classified into groups by the branching topology of Haloarchaea and eubacteria using an in-house PERL script. A group is considered as monophyletic for Haloarchaea if there exists a bipartition (branch) in the tree that splits between Haloarchaea and the rest. Single eubacterial sequences branching with the haloarchaeal clade, and vice versa were tested manually. In each tree, the branch connecting the monophyletic Haloarchaea clade to the eubacteria serves to split the eubacteria clade into two groups, the nearest neighbor of Haloarchaea was assigned as described in Thiergart et al. (36).

Comparison of Tree Sets. Two sets of trees were compared using a χ^2 goodness-of-fit test (58), operating on a $2 \times m$ contingency table. The m cells were defined in an adaptive procedure as follows. The two samples were pooled together into a single set of size n , and the n trees converted into splits. Each split was ranked according to its frequency in the pooled split sets. Each tree was labeled by its lowest ranking split, and the pooled tree set was sorted by this label. Cells were defined as a collection of split ranks by sequential addition of split ranks from the sorted list, and creation of a new cell when the current cell included at least \sqrt{n} trees, resulting in $m \leq \sqrt{n}$ cells. In the last step, trees from the two sets were added to a $2 \times m$ contingency table based on their least ranked split. We have studied the adaptive cell procedure and goodness-of-fit testing in a series of permutation analyses, and the resulting χ^2 test proved to be an unbiased α -level test (SI Text, Table S5, and Figs. S3 and S4).

Phylogenetic Compatibility with a Reference Set. Two sets of trees were compared by their compatibility with a reference set of trees. Each n taxon tree was decomposed into its $(n-3)$ splits, and each split was scored by the

fraction of splits in the reference set that are phylogenetically compatible with it. The ($n-3$) split compatibility scores were averaged to produce a tree compatibility score. The distributions of the tree compatibility scores for the two sets of trees was compared using the Kolmogorov–Smirnov test (58) (*SI Text*).

ACKNOWLEDGMENTS. We thank Martin Embley and Dan Graur for critical comments on an earlier version of the manuscript. We thank the central

computing resources of the University of Düsseldorf for technical support. G.L. acknowledges financial support from the University of Düsseldorf rectorate, W.F.M. and T.D. are funded by the European Research Council and the German Ministry of Science and Education, J.O.M. is funded by the Science Foundation of Ireland, M.S. is funded by the Allan Wilson Centre and the Alexander von Humboldt Foundation, and A.J. and U.D. are funded by the German Research Foundation.

- Oren A (2006) Life at high salt concentrations. *The Prokaryotes* 3:263–282.
- Makarova KS, Yutin N, Bell SD, Koonin EV (2010) Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat Rev Microbiol* 8(10):731–741.
- Kelly S, Wickstead B, Gull K (2011) Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc Biol Sci* 278(1708):1009–1018.
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic Crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 6(3):245–252.
- Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R (2008) Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 6(8):579–591.
- Thauer R (1998) Biochemistry of methanogenesis: A tribute to Marjory Stephenson. *Microbiol Aust* 144:2377–2406.
- Ferry JG (2010) How to make a living by exhaling methane. *Annu Rev Microbiol* 64:453–473.
- Oren A (2006) The order Halobacteriales. *The Prokaryotes: A Handbook on the Biology of Bacteria* (Springer, New York), 113–164.
- Li N, Cannon MC (1998) Gas vesicle genes identified in *Bacillus megaterium* and functional expression in *Escherichia coli*. *J Bacteriol* 180(9):2450–2458.
- Khomyakova M, Bükmez O, Thomas LK, Erb TJ, Berg IA (2011) A methylaspartate cycle in haloarchaea. *Science* 331(6015):334–337.
- Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S (2001) Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* 11(10):1641–1650.
- Lemos RS (2002) Quinol:fumarate oxidoreductases and succinate:quinone oxidoreductases: Phylogenetic relationships, metal centres and membrane. *Biochim Biophys Acta* 1553:1–13.
- Baymann F, Schoepp-Cothenet B, Lebrun E, van Lis R, Nitschke W (2012) Phylogeny of Rieske/cytb complexes with a special focus on the Haloarchaeal enzymes. *Genome Biol Evol* 4(8):720–729.
- van Ooyen J, Soppa J (2007) Three 2-oxoacid dehydrogenase operons in *Haloflexax volcanii*: Expression, deletion mutants and evolution. *Microbiology* 153(Pt 10):3303–3313.
- Boucher Y, et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37:283–328.
- Ichiki H, et al. (2001) Purification, characterization, and genetic analysis of Cu-containing dissimilatory nitrite reductase from a denitrifying halophilic archaeon, *Halorcula marismortui*. *J Bacteriol* 183(14):4149–4156.
- Pfeifer F, Griffing J, Oesterheld D (1993) The fdx gene encoding the [2Fe–2S] ferredoxin of *Halobacterium salinarum* (*H. halobium*). *Mol Gen Genet* 239(1–2):66–71.
- Bickel-Sandkötter S, Gartner W, Dane M (1996) Conversion of energy in halobacteria: ATP synthesis and phototaxis. *Arch Microbiol* 166(1):1–11.
- Boyd JM, Drevland RM, Downs DM, Graham DE (2009) Archaeal ApbC/Nbp35 homologs function as iron-sulfur cluster carrier proteins. *J Bacteriol* 191(5):1490–1497.
- Ng WV, et al. (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci USA* 97(22):12176–12181.
- Rhodes ME, Spear JR, Oren A, House CH (2011) Differences in lateral gene transfer in hypersaline versus thermal environments. *BMC Evol Biol* 11:199.
- Baliga NS, et al. (2004) Genome sequence of *Halorcula marismortui*: A halophilic archaeon from the Dead Sea. *Genome Res* 14(11):2221–2234.
- Pfeiffer F, et al. (2008) Evolution in the laboratory: The genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics* 91(4):335–346.
- Tindall BJ, et al. (2009) Complete genome sequence of *Halomicrobium mukohataei* type strain (arg-2). *Stand Genomic Sci* 1(3):270–277.
- Bolhuis H, et al. (2006) The genome of the square archaeon *Haloquadratum walsbyi*: Life at the limits of water activity. *BMC Genomics* 7:169.
- Anderson I, et al. (2009) Complete genome sequence of *Halorhabdus utahensis* type strain (AX-2). *Stand Genomic Sci* 1(3):218–225.
- Franzmann P, Stackebrandt E (1988) *CSA. Halobacterium lacusprofundi* sp. nov., a halophilic bacterium isolated from Deep Lake, Antarctica. *Syst Appl Microbiol* 11:20–27.
- Kamekura M, Dyal-Smith ML, Upasani V, Ventosa A, Kates M (1997) Diversity of alkaliphilic halobacteria: Proposals for transfer of *Natronobacterium vacuolatum*, *Natronobacterium magadii*, and *Natronobacterium pharaonis* to *Halorubrum*, *Natrialba*, and *Natronomonas* gen. nov., respectively, as *Halorubrum vacuolatum* comb. nov., *Natrialba magadii* comb. nov., and *Natronomonas pharaonis* comb. nov., respectively. *Int J Syst Bacteriol* 47(3):853–857.
- Falb M, et al. (2005) Living with two extremes: Conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res* 15(10):1336–1343.
- Saunders E, et al. (2010) Complete genome sequence of *Haloterrigena turkmenica* type strain (4k). *Stand Genomic Sci* 2(1):107–116.
- Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30(7):1575–1584.
- Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science* 284(5423):2124–2129.
- Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proc Natl Acad Sci USA* 96(7):3801–3806.
- Treangen TJ, Rocha EPC (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* 7(1):e1001284.
- Abhishek A, Bavishi A, Bavishi A, Choudhary M (2011) Bacterial genome chimaerism and the origin of mitochondria. *Can J Microbiol* 57(1):49–61.
- Thiergart T, Landan G, Schenk M, Dagan T, Martin WF (2012) An evolutionary network of genes present in the eukaryote common ancestor: Evidence for eukaryotic and mitochondrial origin. *Genome Biol Evol* 4(4):466–485.
- Lane CE, Archibald JM (2008) The eukaryotic tree of life: Endosymbiosis takes its TOL. *Trends Ecol Evol* 23(5):268–275.
- Deusch O, et al. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol* 25(4):748–761.
- Dagan T, Artzy-Randrup Y, Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* 105(29):10039–10044.
- Kaster A-K, Moll J, Thauer RK (2011) Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc Natl Acad Sci USA* 108(7):2981–2986.
- Collins MD, Jones D (1981) Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol Rev* 45(2):316–354.
- Siebers B, Schönheit P (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr Opin Microbiol* 8(6):695–705.
- Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392(6671):37–41.
- Schink B (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev* 61(2):262–280.
- Stams AJM, Plugge CM (2009) Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol* 7(8):568–577.
- Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440(7084):623–630.
- Lane N, Martin W (2010) The energetics of genome complexity. *Nature* 467(7318):929–934.
- Lake JA, et al. (1985) Eubacteria, halobacteria, and the origin of photosynthesis: The photocytes. *Proc Natl Acad Sci USA* 82(11):3716–3720.
- Wang D, Lloyd AH, Timmis JN (2012) Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc Natl Acad Sci USA* 109(7):2444–2448.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278(5338):631–637.
- Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276–277.
- Tatusov RL, et al. (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
- Katoh K, Misawa K, Kuma K-I, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059–3066.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704.
- Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.
- Deppenmeier U, et al. (2002) The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* 4(4):453–461.
- Zar JH (2010) *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, NJ), 5th Ed.