The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea


*The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; †Department of Biochemistry and Molecular Biology, 5850 College Street, Dalhousie University, Halifax, NS, Canada, B3H 1X5; §Department of Microbiology, Universidad Miguel Hernandez, Apartado 18, 03550 San Juan de Alicante, Alicante, Spain

Contributed by W. F. Doolittle, October 19, 2005

Saturated thalassic brines are among the most physically demanding habitats on Earth: few microbes survive in them. *Salinibacter ruber* is among these organisms and has been found repeatedly in significant numbers in climax saltern crystallizer communities. The phenotype of this bacterium is remarkably similar to that of the hyperhalophilic Archaea (Haloarchaea). The genome sequence suggests that this resemblance has arisen through convergence at the physiological level (different genes producing similar overall phenotype) and the molecular level (independent mutations yielding similar sequences or structures). Several genes and gene clusters also derive by lateral transfer from (or may have been laterally transferred to) haloarchaea. *S. ruber* encodes four rhodopsins. One resembles bacterial proteorhodopsins and three are of the haloarchaeal type, previously uncharacterized in a bacterial genome. The impact of these modular adaptive elements on the cell biology and ecology of *S. ruber* is substantial, affecting salt adaptation, bioenergetics, and photobiology.

Until recently, halophilic archaea (haloarchaea) were thought to be the only cells capable of thriving in saltern crystallizers. These impoundments contain ~37% NaCl, at the limits of tolerance for this environmental factor. Further concentration of thalassic (seawater-derived) hypersaline water leads to precipitation of magnesium salts and sterility. Fluorescent *in situ* hybridization indicates that one crystallizer morphotype, well defined large rods, corresponds to a bacterium of the *Rhodothermus* cluster (1), within the Bacteroides/Chlorobi group. This organism represents 10–20% of the cells in climax crystallizer communities (spring and summer in temperate latitudes). Representative strains (as defined by 16S rRNA sequences) have been isolated from the same environment and described as the previously uncharacterized genus and species *Salinibacter ruber* (2).

The closest cultivated relative of *S. ruber* (henceforth *Salinibacter*) is *Rhodothermus marinus* (89% 16S rRNA sequence similarity), a slightly halophilic thermophile isolated from marine hot springs (2). *Salinibacter* displays many remarkable similarities to haloarchaea, one being a very high concentration of potassium in the cytoplasm (3). This property is associated, as in haloarchaea, with a high content of acidic amino acids and a low content of hydrophobic residues in bulk protein, necessary for protein solubility at such high ionic strength (4). Cell integrity requires high salt concentrations in both cases, and growth only occurs at ≥2 M NaCl. Both *Salinibacter* and the haloarchaea are aerobic heterotrophs that exploit the large stock of organic nutrients produced in previous stages of seawater concentration, mostly by the green alga *Dunaliella*, and they use a similar range of organic compounds as carbon and energy sources (5). Like haloarchaea, *Salinibacter* contains a high proportion of carotoids in its membrane, producing red colonies of similar appearance (2). However, the pigment found in *Salinibacter* (sali-nixinanthin) is a C-40 acyl glycoside carotenoid chemically related to the carotenoids found in *R. marinus* rather than the C-50 bacterioruberins known from haloarchaea (6).

The common features of *Salinibacter* and haloarchaea could have arisen through convergence at the physiological level (different genes producing similar overall phenotype, as in the above-mentioned case of membrane carotenoids) or the molecular level (independent mutations yielding similar sequences or structures). Alternatively, genes may have been shared by lateral gene transfer (LGT) between ancestors of these hyperhalophiles, which we here define as organisms for whom saturated brines, containing >5 M NaCl, are a natural habitat.

**Materials and Methods**

The genome of *S. ruber* strain M31T DSM13855 (2) was sequenced by the random shotgun method, with cloning, sequencing, and assembly as described in ref. 7. Briefly, one small insert (2–3 kb) and one medium insert plasmid library (8–10 kb) were constructed by random nebulization and cloning of genomic DNA. Each library was sequenced to an estimated 4-fold coverage, and the sequences were assembled by using TIGR ASSEMBLER (8) or CELERA ASSEMBLER (9). All sequence and physical gaps were closed as described in ref. 7. ORFs were identified by using GLIMMER (10), and those shorter than 90 bp, as well as some of those with overlaps, were eliminated. Membrane proteins were detected by a hidden Markov method (11). Automated gene annotation was performed as described in ref. 7. The sequences have been submitted to GenBank [accession nos. CP000159 (chromosome) and CP000160 (plasmid)]. The *Salinibacter* automated genome annotation is available at http://cmer.tigr.org/tigr-scripts/CMR/CMRHomePage.cgi.

**Results**

*Genome and Proteome Characteristics.* The genome of strain M31T DSM 13855, the type strain of *S. ruber*, comprises a 3,551,823-bp chromosome of high G+C content (66.29%) and a 35,505-bp plasmid (57.9% G+C), containing 2,934 and 33 ORFs, respectively. Table 1, which is published as supporting information on the PNAS web site, shows a comparison of general genomic features with those of *Chlorobium tepidum* TLS (*Salinibacter*’s closest sequenced relative), *Halobacterium* sp. NRC-1 (an ar-

Conflict of interest statement: No conflicts declared.

Abbreviations: IS, insertion sequences; LGT, lateral gene transfer; SR, sensory rhodopsin.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database [accession nos. CP000159 (chromosome) and CP000160 (plasmid)].

To whom correspondence should be addressed. E-mail: ford@dal.ca.

© 2005 by The National Academy of Sciences of the USA
chael halophile), and *Bacillus halodurans* (a halotolerant bacterium). Previous analysis of halophilic archaeal genomes (such as *Halobacterium* sp. NRC-1, *Haloarcula marismortui* and *Halofex volcanii*) has suggested a bipartite organization, with a high overall G+C content but smaller replicons and chromosomal islands of lower G+C (12, 13), the low G+C regions bearing most of the insertion sequences (IS) and phage-related elements (12). In *Salinibacter*, chromosomal low G+C islands are also enriched in such elements (Fig. 5, which is published as supporting information on the PNAS web site). The first of the low G+C islands is located 250 kb from the inferred origin of replication and contains 15 transposases and five prophage components (including three glycosyl-transferases). The second, spanning over 55 kb, contains 12 transposases, several prophage-related ORFs and, unexpectedly, a number of ORFs involved in DNA metabolism, replication, and recombination. The total number of such candidate trees examined for specific support of transfer or from *Salinibacter* and the indicated genome is shown by the gray bars in Fig. 2. In the pie charts, the fraction of these trees actually supporting LGT from the indicated genome to *Salinibacter* is shown in blue, those supporting LGT from *Salinibacter* to that genome in chartreuse, those supporting LGT of uncertain direction in green, and those not supporting an LGT are shown in red. Phylogenetic reconstruction may be a more reliable method for identifying LGTs, although applicable only when adequate taxa are available and a rooting is possible. For each of these 10 genomes, we examined maximum likelihood trees generated by PHYML (18) with an assumed Bacteria/Archaea rooting in which *Salinibacter* was the deepest branching member of a clade containing that genome. The total number of such candidate trees examined for specific support of transfer to or from *Salinibacter* and the indicated genome is shown by the gray bars in Fig. 2. In the pie charts, the fraction of these trees actually supporting LGT from the indicated genome to *Salinibacter* is shown in blue, those supporting LGT from *Salinibacter* to that genome in chartreuse, those supporting LGT of uncertain direction in green, and those not supporting an LGT are shown in red. Phylogeny, like BLAST, identified haloarchaea as the most frequent donors or recipients in LGTs involving *Salinibacter*, although the total number of apparent transfers between *Salinibacter* and haloarchaea appears to be modest. (And aside from the haloarchaea and *Rhodopirellula baltica*, many LGTs indicated by BLAST are not supported by trees.) Where appropriate in the following sections of this report, we highlight individual instances in which LGT between these groups nevertheless has likely been a key factor in adaptation to the harsh saline environment.

**Physiology.** *Salinibacter* has a full complement of fermentative components and genes involved in transport and degradation of organic compounds. A chitinase, two cellulases, one amylase, a pectinase, and several proteases and lipases were among the polymer-degrading enzymes encoded on the chromosome. Additionally, the plasmid encodes a 1,4-β-cellobiosidase homolog. Genes related to the transport and metabolism of glyceral and glycine betaine (the most common compatible solutes found in more moderate halophiles and halotolerant species) were also present. The broad degradative abilities fit with the expected complexity of the organic pool present in the crystallizer as a result of the lysis of biomass produced at lower salinities. Contrary to a previous suggestion (19), glycolysis appears to take place through an Embden-Meyerhoff pathway and not a modified Entner-Doudoroff pathway.

Like the haloarchaea, *Salinibacter* has all of the genes for a

---

**Fig. 1.** Normalized distribution of pI values at 0.2 intervals for predicted ORFs in *Haloarcula marismortui* (purple), *Halobacterium* sp. NRC-1 (red), *Salinibacter* (blue), *C. tepidum* (green), and *B. fragilis* (cyan). Predicted pI values of the proteins were calculated by using NeuroGadgets Bioinformatics Web Service.
complete tricarboxylic acid cycle and a cytochrome c-containing respiratory chain. It possesses a succinate dehydrogenase gene cluster (SRU.0484–0487) very similar to Actinobacteria and a second succinate dehydrogenase flavoprotein subunit, sdhA (SRU.2444), most similar to sdhA2 from *H. marismortui*. Similarly, comparison of respiratory chain proteins in *Salinibacter* indicates there are two clusters of cytochrome c oxidase subunit I and II genes, the first with coxA1 and coxB1 (SRU.2099 and SRU.2100) and the second with coxA2 and coxB2 (SRU.0314 and 0313). Phylogenetic analyses demonstrate coxA1 and coxB1, which are also next to the genes for subunits III (SRU.2098) and IV (SRU.2097), are related to those of *R. marinus*, whereas coxB2 and coxA2 are clearly most closely related to their haloarchaea homologs (Fig. 6, which is published as supporting information on the PNAS web site). This second gene cluster may have been imported from the haloarchaea, a shared adaptation to the microoxic conditions often associated with hypersalinity, perhaps functioning in a novel respiratory electron transfer pathway. In support of this notion, a nosZDF gene cluster and an oxygen-sensitive Cpr/Fnr-type transcription regulator (20) are found near to coxA2 and coxB2 in *Salinibacter*. The nos cluster encodes a nitrous oxide reductase similar to that found in a variety of proteobacteria (21, 22) and in *H. marismortui*. Phylogenetic analyses of NosZ (SRU.0308) and NosD (SRU.0310) demonstrate high similarity between *Salinibacter* and *H. marismortui*; however, the exact relationship remains unresolved (Fig. 7, which is published as supporting information on the PNAS web site).

*Salinibacter* may have other unique characteristics with respect to microoxic adaptation. Unlike haloarchaea but like *R. marinus* (23), it possesses components of a cbb3-type cytochrome oxidase (ccoO and ccoN) (SRU.0323 and SRU.0322). Cytochrome c oxidases of the cbb3 type have a very high affinity for O2 and allow respiration to continue under low-O2 levels (24). Both ccoO and ccoN were acquired likely by LGT from a β-proteobacterium. Strangely, *Salinibacter* does not possess a ccoP gene, thought to be an essential subunit of cbb3-type cytochrome oxidase (25). However, next to ccoON is a gene encoding a cytochrome c protein (SRU.0324), which also appears to have been transferred from a β-proteobacterium and may substitute for CcoP function in *Salinibacter*.

**Haloadaptation.** A cluster of 19 genes (Fig. 3) includes K+ uptake/efflux systems and cationic amino acid transporters of crucial importance to a hyperhalophilic lifestyle. This “hyper-salinity island” is mosaic in nature, apparently pieced together from a variety of bacterial and archaeal sources. More than one-half of the genes are most similar to their haloarchaea homologs, including the cationic amino acid transporters, one trkH gene, and three trkA homologs. The Trk system is responsible for the uptake of K+, where TrkH is the membrane bound translocating subunit and TrkA is a cytoplasmic membrane surface protein that binds NAD+ (26). The existence of multiple trkA genes in *Salinibacter* suggests complex regulation of the Trk system, a feature likely shared with haloarchaea. This complexity is increased by the presence of yet an additional TrkAH system in the hypersalinity island that is most similar to that found in Firmicutes (Fig. 3).

The recruitment of several nonhaloarchaeal hypersalinity genes may have involved an island-encoded IS transposase. In *Escherichia coli*, IS1 has two overlapping ORFs, insA and insB, and translational frameshift results in the production of InsAB, a functional IS1 transposase (27). There are four insA-like genes in *Salinibacter* and a similar expression mechanism may exist in this bacterium. Genes with significant similarity to the *Salinibacter* insA have been found only in cyanobacteria, *Parachlamydia*, and members of two archaeal orders, Methanosarcinales and Sulfolobales. This patchy phylogenomic distribution of the IS1 transposase is mirrored in the distribution of the KefB K+ efflux system proteins (28) and the Na-K-Cl cotransport proteins (29) present in...
matches in BLAST sequence similarity searches: haloarchaea, red; cyanobacteria, green; methanogenic archaea, yellow; firmicutes, blue.

until very recently (see below), there has been no information other and to the haloarchaeal SRs, specifically SRI. In genes, 83 kb apart on the genome and distantly related to each until the present work. Indeed, like Halobacterium sp. NRC-1 and Haloarcula marismortui, and unlike any other characterized bacterium, Salinibacter contains four rhodopsin genes (Fig. 4). Three of the Salinibacter rhodopsin groups with haloarchaeal rhodopsins in phylogenetic reconstructions and might be inferred, from the nature of neighboring genes in the Salinibacter genome, to have similar functions. Salinibacter SRU.2780 forms a basal member of the haloarchaeal halorhodopsin clade and, like them, is immediately upstream of an oxidoreductase gene (possibly a regulator of pump activity) (Fig. 4E). A Na+/H+ antiporter is located only three genes away in Salinibacter and a Na+/substrate cotransporter is adjacent. In both archaean genomes, trkA (a potassium symporter gene) is found nearby. This observation, together with relatively high sequence similarity, supports the same function, as an inward-directed chloride pump, for the Salinibacter protein. Salinibacter has two presumptive sensory rhodopsin (SR) genes, 83 kb apart on the genome and distantly related to each other and to the haloarchaeal SRs, specifically SRI. In Halobacterium, two SRs (SRI and SRII) interact to produce a color-sensitive phototactic behavior. SRI is only synthesized under low oxygen tension (as are the proton pump BR and the chloride pump HR), mediating the attraction to orange light that drives the ion pumps (35). SRII is produced at high oxygen tensions when the bioenergetics of the cell are driven by the respiratory chain and mediates a photophobic response to blue light (35). In haloarchaeae, the signaling function of both SRs depends on their tight association with transducers that block their potential ion transport activity and transform them into photosensors (36).

Fig. 3. A schematic representation of the hypersalinity island identified in the genome of Salinibacter. Genes are color-coded with respect to their closest matches in BLAST sequence similarity searches: haloarchaea, red; cyanobacteria, green; methanogenic archaea, yellow; firmicutes, blue.

Rhodopsins and Retinol. The discovery of the proton pump bacteriorhodopsin in the haloarchaea Halobacterium salinarum in the early 1970s launched a new fruitful field of bioenergetics and biophysics, and this protein and its relatives were seen as a defining invention of the archaea (30). More recently, rhodopsin-based photobiology has been found in other groups of prokaryotes (31, 32) and in unicellular eukaryotes (33, 34). The possibility that Salinibacter might also have rhodopsin genes, derived from the haloarchaea with which it lives, was one of the initial motivations behind the present work. Indeed, like Halobacterium sp. NRC-1 and Haloarcula marismortui, and unlike any other characterized bacterium, Salinibacter contains four rhodopsin genes (Fig. 4). Three of the Salinibacter rhodopsin groups with haloarchaeal rhodopsins in phylogenetic reconstructions and might be inferred, from the nature of neighboring genes in the Salinibacter genome, to have similar functions. Salinibacter SRU.2780 forms a basal member of the haloarchaeal halorhodopsin clade and, like them, is immediately upstream of an oxidoreductase gene (possibly a regulator of pump activity) (Fig. 4E). A Na+/H+ antiporter is located only three genes away in Salinibacter and a Na+/substrate cotransporter is adjacent. In both archaean genomes, trkA (a potassium symporter gene) is found nearby. This observation, together with relatively high sequence similarity, supports the same function, as an inward-directed chloride pump, for the Salinibacter protein. Salinibacter has two presumptive sensory rhodopsin (SR) genes, 83 kb apart on the genome and distantly related to each other and to the haloarchaeal SRs, specifically SRI. In Halobacterium, two SRs (SRI and SRII) interact to produce a color-sensitive phototactic behavior. SRI is only synthesized under low oxygen tension (as are the proton pump BR and the chloride pump HR), mediating the attraction to orange light that drives the ion pumps (35). SRII is produced at high oxygen tensions when the bioenergetics of the cell are driven by the respiratory chain and mediates a photophobic response to blue light (35). In haloarchaeae, the signaling function of both SRs depends on their tight association with transducers that block their potential ion transport activity and transform them into photosensors (36). Until very recently (see below), there has been no information about Salinibacter photobiology and the function of the two genes homologous to SRs still can be inferred only from the genomic context. Nevertheless, there seems little doubt that SRU.2780 is a photosensor, because the next ORF, 86 nucleotides downstream, is an Htr1 transducer, as in Halobacterium sp.
The second putative SR in *Salinibacter* (SRU.2511) is also tightly linked to nearby signal transduction genes (Fig. 4C), pointing to a photoresponse function and potentially complex photobehavior in *Salinibacter*. Notably, although there is no doubt of the relatively close relationships among the sensory rhodopsins of *Salinibacter* and the haloarchaea, the transducers that are tightly linked at the genetic (and presumably functional) level are not closely related to those of haloarchaea (Fig. 8, which is published as supporting information on the PNAS web site). Their closest relatives outside the *Salinibacter* genome are transducers of other bacteria. Thus, physiological convergence between *Salinibacter* and haloarchaea likely has been effected through independent gene recruitment processes.

Downstream and tightly linked to the first *Salinibacter* SR gene discussed above (Fig. 4D) (SRU.2579), there is a typically bacterial flagellar cluster that might be coregulated with it: SR function in phototaxis would be useful only when cells are actively motile; members of the Bacteroides/Chlorobi group typically lack flagella. However, a polar flagellum has been observed in *Salinibacter*’s closest relative *R. marinus* (37). Phylogenetic analyses of the genes within this flagellar gene cluster indicate a mosaic structure, with many genes showing a close affinity to the *S. solfataricus* proteobacteria. Perhaps the flagellum was a recent acquisition by *Salinibacter*/Rhodothermus. With respect to motility, a general, although not universal, character of the Bacteroides group is gliding motility. Studies in *Flavobacterium johnsoniae* have identified a suite of genes required for the gliding motility phenotype (gldABDFGH) (38). *Salinibacter* possesses homologs of the gldA, gldF, and gldG genes (SRU.1249–SRU.1251), which are thought to form an ATP-binding cassette transporter in *F. johnsoniae*, but it lacks the critical lipoprotein components (gldB, gldD, and gldH) that are required for gliding.

The fourth *Salinibacter* rhodopsin gene (SRU.1500) appears to be most closely related to those found in cyanobacteria and forms a clade with organisms that encode diverse rhodopsin genes, which structurally related to the major carotenoid pigment found in *R. marinus* (6). The gene encoding xanthorhodopsin is linked to two ORFs (crtY and crtO) (SRU.1501 and SRU.1502) that code for proteins responsible for important steps in both carotenoid and vitamin A synthesis (Fig. 4B). Lycopene β-cyclase (CrtY) is required for the formation of β-carotene, a precursor to both retinal and subsequent carotenoid production, whereas β-carotene ketolase (CrtO) is likely responsible for an important step in salixinanthin biosynthesis (40, 41).

Until recently, the manner by which retinal was produced in bacteria was unknown. A metagenomic study of proteorhodopsin-containing BAC clones from marine waters (42) indicate that these organisms may be synthesizing retinal by a mechanism first identified in haloarchaea, which have been shown to use bacteriorhodopsin-related protein (Brp) and its paralog, bacteriorhodopsin-related protein like homolog (Blh) to produce retinal from symmetrical cleavage of β-carotene (43). A divergent homolog of Blh (20% amino acid identity with the archaeal version) was found linked to proteorhodopsin on BAC clones (42), whereas in *Salinibacter*, homologs of both brp and blh are present in the genome (unlinked to rhodopsin genes) and exhibit close relation to those found in haloarchaea rather than marine bacteria.

**Discussion**

*Salinibacter* has adapted to hypersaline life in three general ways. First, *Salinibacter* must have modified the sequences of many of its proteins. Whole and partial proteome studies of haloarchaea show that these hyperhalophilic organisms have accommodated high internal ionic strength by replacing neutral amino acids with acidic ones (44). This situation must also be the case for *Salinibacter* because an acidic proteome is not a property of its relatives among the Bacteroides/Chlorobi group (Fig. 2). Although some genes imported from haloarchaea may have produced proteins already adapted in this way, these comprise only a fraction of *Salinibacter*’s genes. A comparative study of orthologous proteins in *Salinibacter* and haloarchaea and their closest nonhalophilic relatives should tell us much about convergent evolution at the level of protein structure.

Second, *Salinibacter* exhibits many cases of convergence at the level of physiology, in which proteins from different sources or with different original functions have been recruited to create a complex adaptation to hypersaline life that parallels a structure, pathway or behavior already known from haloarchaea. The mix-and-match of genes of haloarcheal and bacterial origin in the postulated phototaxis system will likely provide a good example of analogous complex collective functions served by nonhomologous component parts. The recent discovery that xanthorhodopsin (*Salinibacter*’s proteorhodopsin-like SRU.1500) functions in a light-harvesting complex with salixinanthin similarly provides a striking analogy with chlorophyll-based light-harvesting systems (39).

Third, some adaptations common among halophilic organisms have been passed between them by LGT. The rhodopsins shared by *Salinibacter* and the haloarchaea may be the best case in point. Overall, the specific functionally important residues in the *Salinibacter* rhodopsins match those of the haloarchaeal proteins to which the rhodopsins are homologous, and the two SR-transducer homologs both appear to be sensory rhodopsin photoreceptor transducers, like their haloarchaeal HtrI homologs, rather than chemoreceptor taxa proteins (John Spudich, personal communication). Unless we imagine that the last universal common ancestor had a full complement of rhodopsins, the presence of several identifiable classes of this protein in *Salinibacter* and haloarchaea is best explained by multiple events of LGT. Although proteorhodopsin-like genes have been described in several bacterial phyla (45), we note that *Salinibacter* is the only bacterium in which genes that cluster specifically with haloarchaeal rhodopsin genes are known. It might seem natural to assume that *Salinibacter* has derived these genes from archaea, by LGT, but we cannot be certain. *Salinibacter*’s two sensory rhodopsins appear to have derived from one of the two haloarchaeal sensory rhodopsin paralogs (as if after the duplication in that lineage), but its halorhodopsin diverged before the diversification of haloarchaea. Many of the other “haloarchaeal genes” found in *Salinibacter* could just as easily be “Salinibacterial genes” introduced by LGT into haloarchaeal genomes. Indeed, the respiratory chain and associated functions in haloarchaea are widely believed to have been of bacterial origin (46, 47), and *Salinibacter*’s ancestors could have been one source. Bacterial and archaeal cells have likely shared the saltern habitat for millennia, with ample chance to exchange genes. The notion of a “habitat genome” (or a pool of genes useful for adaptation under a specific set of environmental constraints) is appealing. Such a gene pool could be analogous (in an evolutionary time scale) to the gene pool in a metazoan (or plant) zygote from which different tissues extract the required components (48).
We thank Tom Cavalier-Smith for alerting us to the significance of the presence of genes for flagellar components. Sequencing and annotation were supported by National Science Foundation/U.S. Department of Agriculture Grant EF0333190. Additional bioinformatic analyses were supported by Canadian Institutes for Health Research Grant MOP4467, Genome Atlantic, and the Generic Architecture for Customised IP-Based IN Services over Hybrid VoIP and SS7 Project QLK3-CT-2002-02056 of the European Commission.