MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Mar. 1997, p. 90–104 Vol. 61, No. 1
0146-0749/97/$04.00+0

Evolutionary Divergence and Salinity-Mediated Selection in Halophilic Archaea†

PATRICK P. DENNIS[1]* AND LAWRENCE C. SHIMMIN[2]

Department of Biochemistry & Molecular Biology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3 Canada,[1] *and Human Genetics Center, The University of Texas Houston Health Science Center, Houston, Texas 77225*[2]

INTRODUCTION: ENVIRONMENTS, STRATEGIES, AND ORGANISMS

The proteins of halophilic (literally salt-loving) archaea are highly adapted and magnificently engineered to function in a milieu containing between 2 and 5 M inorganic salt. Because of this extreme and unfriendly saline environment, halophilic proteins and their encoding genes represent a valuable repository and resource for reconstructing and visualizing processes of natural selection and adaptive evolution.

Hypersaline environments, although often rich in radiant energy and organic nutrients, are subject to exploitation by a relatively small number of closely related microbial species (reviewed in reference 11). The presence of a high concentration of solute ions is generally devastating to proteins and other macromolecules: (i) it causes aggregation or structural collapse of proteins because of enhancement of hydrophobic interactions; (ii) it interferes with essential electrostatic interaction within or between macromolecules because of charge shielding; and (iii) it reduces the availability of free water below that required to sustain essential biological processes because of salt ion hydration (reviewed in references 12 and 31).

A number of different eubacterial and eucaryotic microorganisms have evolved the ability to exploit moderately saline environments. In general, they use an ion-pumping mechanism to maintain a low (physiological) intracellular solute concentration and thereby protect macromolecules and essential processes from the adverse effects of salt ions (11, 31). Many of these organisms also produce and accumulate biologically compatible solutes such as glycine betaine or ectine to partially buffer the osmotic imbalance. These strategies are effective in environments up to about 1.5 M in salt; above that concentration, they are ineffective and energetically unfavorable.

A small group of halophilic archaea are capable of exclusive exploitation of hypersaline environments that are between 2 and 5 M in salt (11). Remarkably, these extremely halophilic organisms have abandoned the strategy of maintaining an osmotic gradient across the cell envelope. Instead, adaptation is achieved (i) by balancing the extracellular ionic strength against the intracellular ion concentration (where K^+ is the predominant cation) and (ii) by modifying soluble intracellular proteins and other macromolecules to function in this high-ionic-strength milieu (3, 6, 12, 21, 24, 32).

The evolutionary modifications required to reengineer a protein so that it becomes halophilic seem to involve the introduction of additional acidic (glutamic and aspartic acid) residues onto the surface of the protein. Acidic residues are more highly hydrated than other amino acids and can coordinate the organization of a hydrated salt ion network at the surface of the protein (7). Replacements with acidic and other hydrophilic residues also reduce protein hydrophobicity and help prevent the structural collapse or aggregation (salting out) of the protein (6, 12, 32). Moreover, certain acidic residues can be used to form salt bridges with strategically positioned basic (lysine and arginine) residues. Salt bridges, positioned so that they are protected from shielding by solute ions, provide structural rigidity and are important determinants in the stabilization of the three-dimensional structure of the halophilic protein (5). In solutions of low ionic strength, halophilic proteins often denature and unfold due to charge repulsion.

Environments inhabited by halophilic archaea contain Na^+, K^+, Mg^{2+}, and Ca^{2+} as the predominant cations and Cl^-, SO_4^{2-}, and CO_3^{2-} as the major anions (22, 26). The concentrations and relative proportions of these ions vary in different localities and, within a given locale, fluctuate with time due to solubilization-precipitation or dilution-evaporation processes. Extreme halophiles achieve a near balance in the overall in-

* Corresponding author. Mailing address: Department of Biochemistry & Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, B.C. V6T 1Z3, Canada. Phone: (604) 822-5975. Fax: (604) 822-5227. E-mail: patrick.p.dennis@unixg.ubc.ca.

† This paper is dedicated to Al Matheson, who introduced us to the wonder of halophilic archaea.

tracellular cation concentration with that of the external environment by concentrating predominantly K^+ and extruding Na^+; chloride is the predominant intracellular anion. The accumulation of K^+ rather than Na^+ is important because the K^+ ion hydrates less water than the Na^+ ion does.

In this review, the sequence divergence of halophilic archaeal protein-encoding genes has been analyzed and compared to the divergence of homologous nonhalophilic eubacterial protein-encoding genes. The comparisons indicate that halophilic genes accumulate two to three times as many nonsynonymous nucleotide substitutions as the homologous nonhalophilic eubacterial genes. We suggest that many of these substitutions, resulting in amino acid replacements, represent "evolutionary tinkering" at sites which influence the hydrophobic and surface hydration properties of the proteins. Moreover, we propose that fluctuations in environmental salinity provide the driving force for fixation of many of these nucleotide substitutions. In a constant saline environment, a slightly deleterious mutation is likely to be eliminated from the population, whereas in a fluctuating saline environment, an existing mutation which was previously deleterious may on occasion become advantageous and have an increased probability of achieving fixation in the population.

METHODS OF ANALYSIS

Organisms, Genes, and Alignments

In this analysis, protein-encoding gene sequences from six species representing three genera of halophilic archaea (*Halobacterium cutirubrum*, *Halobacterium* sp. strain GRB, *Haloferax volcanii*, *Haloferax mediterranei*, *Haloarcula hispanica*, and *Haloarcula marismortui*) have been compared. Based upon 16S rRNA sequences, these three genera are believed to have diverged from a common halophilic ancestor in an ill-defined order about 600 million years ago (17; also, see below).

The gene sequences used in the analysis encode either the L11, L1, L10, and L12 ribosomal proteins or Mn/Fe superoxide dismutase (SOD) proteins. Ribosomal protein genes are single copy and essential in all three genera, whereas *sod* genes are duplicated in *Halobacterium* and *Haloferax* and single copy in *Haloarcula* and presumably are not essential (1, 8, 9, 14–16, 27, 28). Complete *sod* gene sequences are available from *Halobacterium cutirubrum* (*sod*, *slg*), *Halobacterium* sp. strain GRB (*sod*, *slg*), *Haloferax volcanii* (*sod1*, *sod2*), and *Haloarcula marismortui* (*sod*). Partial sequences of the paralogous (duplicated) *sod* genes from *Haloferax mediterranei* (*sod1*, *sod2*) and the single-copy gene from *Haloarcula hispanica* (*sod*) were obtained by PCR amplification. The primers used were oPD62 (TGGCAYCACGACACCCACCAYCA; forward primer, codons 30 to 37) and oPD63 (GTCCCAGTC GAY(GCA)ACCTCGAAGAA; reverse primer, codons 189 to 182). These primers amplify the region between codons 38 and 181; this represents about 70% of the coding region of halophilic *sod* genes.

To visualize normal evolutionary processes in the absence of salinity-mediated selection, homologous L11 and L1 ribosomal protein genes from three enteric bacteria, *Escherichia coli*, *Serratia marcescens*, and *Proteus vulgaris*, were analyzed in parallel (19, 29). *Proteus* and *Serratia* are believed to have diverged from *Escherichia* about 350 and 200 million years ago, respectively (see below). These were the only homologous sequences in the databases that had a sufficiently small number of synonymous nucleotide substitutions per synonymous site (K_s) to make appropriate comparisons with the halophilic ribosomal protein sequences.

TABLE 1. Organisms and gene sequences used in the analysis of nucleotide substitution rates in halophilic archaea

Gene(s)[a] and organism	Abbreviation	Accession no.	Reference(s)
L11, L1, L10, L12			
Haloferax volcanii	*Hfvo*	X58924	28
Halobacterium cutirubrum	*Hbcu*	X15078	27
Haloarcula marismortui	*Hama*	X51430	1
E. coli	*Eco*	V00339	19
S. marcescens	*Sma*	X12584	29
P. vulgaris	*Pvu*	X12585	29
sod			
Halobacterium cutirubrum	*Hbcu*	J04956	16
		M26502	15
Halobacterium sp. strain GRB	*Hb.GRB*	M97483	8, 9
		M97484	8, 9
Haloferax volcanii	*Hfvo*	M97486	8, 9
		M97487	8, 9
Haloferax mediterranei	*Hfme*	U78907	This work
		U78908	This work
Haloarcula marismortui	*Hama*	M97485	8, 9
Haloarcula hispanica	*Hahi*	U78906	This work
M. avium	*Mav*	X81384	33
M. fortuitum	*Mfo*	X81385	33
M. gordonae	*Mgo*	X81386	33
M. intracellulare	*Min*	X81387	33
M. kansasii	*Mka*	X81388	33
M. scrofulaceum	*Msc*	X81389	33

[a] The database accession numbers for small-subunit rRNA sequences are as follows: *Hfvo*, K00421; *Hfme*, D11107; *Hbcu*, K02971; *Hama*, M27042, M27043; *Hahi*, U68541; *Eco*, V00384; *Sma*, M59106, *Pvu*, J01874; *Mfo*, X52933; *Msc*, X52924; *Mav*, X52918; *Mgo*, X52923; *Min*, X52927; *Mka*, X15916; *Methanobacterium thermoautotrophicum* (*Mth*), Z37156; *Vibrio cholerae* (*Vch*), X76337.

Similarly, homologous Fe/Mn SOD-encoding genes from six species of the genus *Mycobacterium* were subjected to comparative analysis (33). These mycobacterial species diverged from a common ancestor during the last 250 million years (23). The mycobacterial sequences were obtained by PCR amplification of a 489-nucleotide region within the *sod* gene (33). This region is homologous to 80% of the coding region. The database accession numbers for the nucleotide and deduced amino acid sequences used in this study are listed in Table 1. All ribosomal protein and SOD nucleotide and amino acid sequences were aligned manually. The alignments are in general uncomplicated and unambiguous. However, regions of doubtful alignment, particularly around deletion and insertion events appearing in one or more of the sequences, were excluded from the analysis.

Quantitation of Synonymous and Nonsynonymous Nucleotide Substitutions

The aligned sequences were compared in all possible pairwise combinations, and the number of synonymous nucleotide substitutions per synonymous site (K_s) and the number of nonsynonymous substitutions per nonsynonymous site (K_a) were estimated by the method of Li (13). In a table of the genetic code (with translation termination codons excluded), 28% of all possible sites for substitution are synonymous and 72% are nonsynonymous. In any given sequence, the exact percentages of synonymous and nonsynonymous sites vary depending on the amino acid composition and codon utilization. For calculation of the K_s and K_a values in pairwise comparisons, the number of synonymous and nonsynonymous sites was taken as the average of the values for the two sequences. Moreover, the K_s and K_a values as calculated contain a correction for the

transition-transversion bias and, most importantly, a correction for multiple substitutions at the same site by using the Jukes-Cantor model (10, 13). To compare multiple pairs of sequences with different or uncertain divergence times, the ratio $K_a/(K_s + K_a)$ has been calculated. This ratio is (on a per-site basis) the fraction of total substitutions that are nonsynonymous; when comparing two sequence pairs, differences due to divergence times cancel in the ratio.

Accurate estimations of K_s and K_a are often difficult to obtain because synonymous substitutions are much more frequent than nonsynonymous substitutions and because synonymous sites are far less prevalent than nonsynonymous sites (28% versus 72% in the standard genetic code). This means that synonymous sites can become saturated with substitutions before a significant number of substitutions have occurred at nonsynonymous sites. We have attempted to minimize these difficulties by choosing sequences where the estimated values of K_s are almost always below 1.0 and never exceed 1.602. There is no indication that the higher K_s values are grossly inaccurate or have biased the interpretation of the data in any significant way.

Estimation of Divergence Times

Divergence time for organisms used in this study were estimated from small-subunit 16S rRNA sequences. These sequences were assumed to diverge at a constant and uniform rate of 1% per 50×10^6 years (18). Phylogenetic distance trees based on 16S rRNA sequences from the halophiles, enteric bacteria, and mycobacteria used are illustrated in Fig. 1. Among halophilic genera, 16S rRNA sequences differ by about 11 to 13% (17). From this, we conclude that the three genera diverged from each other about 600×10^6 years ago. The precise branching order is uncertain, although *Haloferax* appears to represent the deepest branch. Within the genus *Haloferax*, separation of the two species, *H. volcanii* and *H. mediterranei*, occurred about 85×10^6 years ago. The situation within *Haloarcula* is somewhat more complicated. Two 16S rRNA genes have been cloned and characterized from *H. marismortui*. These *rrnA* and *rrnB* 16S sequences differ at 5% of the nucleotide positions (17). The *rrnB* 16S sequence is more similar (98.4% identity) to the single 16S sequence available from *H. hispanica* than is the *rrnA* 16S sequence (96.0% identity); based on the *rrnB* sequence, we estimate the divergence time for the two species to be about 80×10^6 years ago. Within *Halobacterium*, only the *H. cutirubrum* 16S gene has been sequenced. *Halobacterium* sp. strain GRB and *H. cutirubrum* are very closely related by a number of criteria and, along with *H. halobium*, are generally considered to be substrains of the type species *H. salinarium.*

The enteric bacterial 16S rRNA tree is uncomplicated. The 16S sequences of *E. coli* and *S. marcescens* are about 4% divergent from each other and about 7% divergent from *P. vulgaris*. From this, we conclude that *Escherichia-Serratia* diverged from *Proteus* about 350×10^6 years ago and from each other about 200×10^6 years ago.

The situation within the mycobacteria is also complicated because a number of different culture collection entries (ATCC versus DSM) have been given the same species designation. The 16S rRNA and *sod* database entries for *M. fortuitum* (ATCC 6841), *M. scrofulaceum* (ATCC 19981), and *M. kansasii* (DSM 43224) were apparently derived from the same isolates (23, 32). The 16S rRNA sequences are about 4% and 1.5% divergent; from this, we conclude that *M. fortuitum* diverged about 200×10^6 years ago and that *M. scrofulaceum* and *M. kansasii* diverged from each other about 75×10^6 years ago.

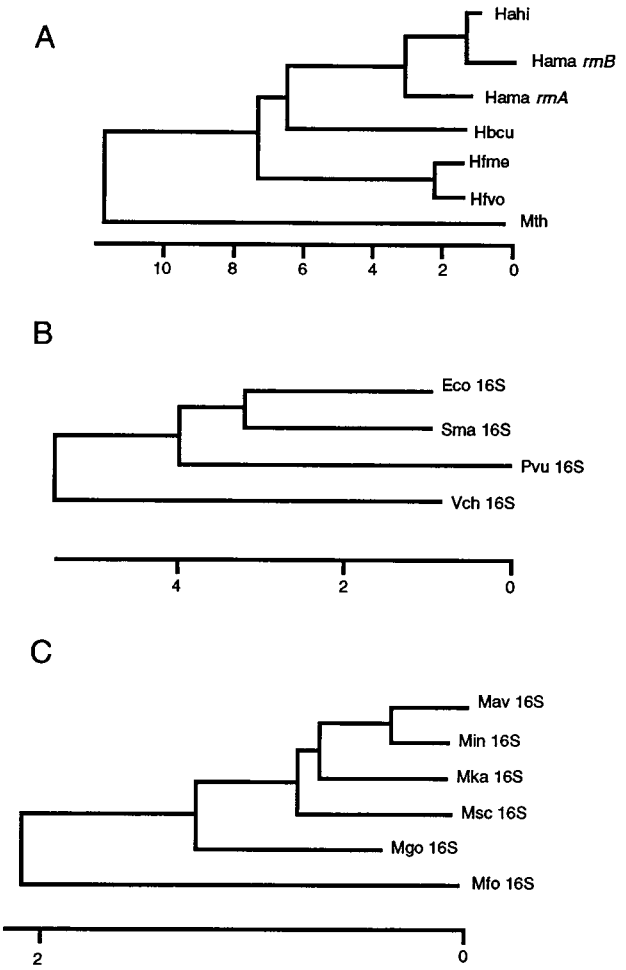


FIG. 1. Phylogenetic trees of 16S rRNA sequences. The sequences of 16S rRNA genes from organisms used in this study were aligned and analyzed by the neighbor-joining method (25). Species abbreviations and data accession numbers are presented in Table 1. The three groups of organisms are halophiles (A), enteric bacteria (B), and mycobacteria (C). The outgroups used for rooting the trees are, respectively, *Methanobacterium thermoautotrophicum* (Mth), *Vibrio cholerae* (Vch), and *Mycobacterium fortuitum* (Mfo). The scale at the bottom of each tree is percent divergence; here we have assumed that 16S rRNA sequences diverge at a rate of 1% per 50×10^6 years (18).

In spite of the well-defined 16S rRNA phylogeny for these three species, the *sod* gene from *M. kansasii* represents a conundrum. The *sod* gene is the most divergent within the mycobacterial *sod* gene tree and exhibits a much higher level of nonsynonymous substitution than do other mycobacterial *sod* genes (see below). We cannot explain this apparent discrepancy. For the remaining mycobacterial species (*M. avium*, *M. intracellulare*, and *M. gordonae*), correspondence between *sod* and 16S gene sequences is somewhat uncertain, although their branching positions within the two trees are congruent. The database accession numbers for all 16S rRNA sequences used to estimate divergence times are given as a footnote in Table 1.

ANALYSIS OF HALOPHILIC ARCHAEAL AND EUBACTERIAL GENE SEQUENCES

Ribosomal Protein Gene Sequences

The essential L11, L1, L10, and L12 ribosomal proteins are functionally conserved and highly specialized components of a complex translational apparatus (20). The availability of nu-

cleotide sequences from three halophilic genera, *Haloferax volcanii*, *Halobacterium cutirubrum*, and *Haloarcula marismortui*, makes it possible to directly visualize the divergence that has occurred within genes that encode proteins with a conserved and essential function since their separation from a common ancestral sequence about 600 million years ago. In the analysis, a conservative approach was followed; all initiation and termination codons, all codon positions that are not represented in one or more of the sequences and, most importantly, all codons within regions of ambiguous alignment were excluded from consideration (Fig. 2). For example, in the most extreme case, only 294 of 356 codons in the halophilic L10 alignment were considered. All codons beyond position 297 were excluded because the large number of apparent deletions, coupled with the high proportion of acidic residues, makes homology at any given position far from certain. Thus, the numbers of nucleotide substitutions or amino acid replacements analyzed are underestimates of the actual number of differences.

At the nucleotide level and over the regions analyzed, these halophilic ribosomal protein genes are 71 to 81% identical, and at the amino acid level the corresponding proteins are 65 to 81% identical (Fig. 2). The most striking feature in these comparisons is the high proportion of nonsynonymous nucleotide substitutions. Over the regions analyzed, 47% (414 of 885) of all observed nucleotide substitutions in the alignment are nonsynonymous and result in amino acid replacements in the encoded protein (Fig. 2). This is reflected in a transition-to-transversion ratio of considerably less than 1 (on average, about 0.5) and in an elevated proportion of substitutions occurring at the first and second codon positions. The most extreme example of this feature is between codons 65 and 75 of the L12 genes, where there are on average three substitutions per codon position and very little amino acid sequence conservation. Normally, for closely related genes encoding proteins with conserved function in enteric bacteria, the vast majority of nucleotide substitutions (80 to 99%) are synonymous changes at the third codon position and the transition-to-transversion ratio is around 2 (18).

To quantitate these features, K_s and K_a have been computed for all pairwise combinations of sequences (Table 2). These parameters contain adjustments for transition-transversion bias and for multiple substitutions at the same site by using the Jukes-Cantor correction (13). To account for differences or uncertainties in the divergence times between the pairwise combinations of sequences, the ratio $K_a/(K_a + K_s)$ was calculated. These calculations indicate that on a per-site basis and for an aggregate of all four halophilic proteins, K_a represents 14.4 to 16.7% of all nucleotide substitutions in these halophilic ribosomal protein genes (Table 2).

On their own, values of K_s, K_a, and even the ratio $K_a/(K_s + K_a)$ are relatively meaningless. Genes accumulate synonymous and nonsynonymous substitutions at characteristic rates that can vary substantially from one gene to another. To bring the K_s and K_a values for halophilic L11, L1, L10, and L12 ribosomal protein genes into perspective, the nucleic acid sequence database was searched for nonhalophilic homologous sequences from organisms that have diverged from each other during the last 600 million years. Sequences for L11 and L1 ribosomal protein genes from three enteric bacteria, *E. coli*, *S. marcescens*, and *P. vulgaris*, were recovered. The L10 and L12 genes from the last two species have unfortunately not been sequenced. Based on 16S rRNA sequences, *Proteus* and *Serratia* are believed to have diverged from *Escherichia* about 350 and 200 million years ago, respectively. No other entries suitable for comparison were recovered.

At the nucleotide level, the enteric L11 and L1 genes are 80 to 90% identical, and at the amino acid level, the corresponding proteins are 88 to 95% identical (Fig. 3). There is a nearly twofold variation in the pairwise K_s and K_a values because of the more recent divergence between *S. marcescens* and *E. coli*. However, the value of the ratio $K_a/(K_a + K_s)$, in which differences in divergence time cancel, indicates that only 6.3 to 8.3% of the substitutions on a per-site basis and for the aggregate of the L11 and L1 sequences are nonsynonymous (Table 2). Thus, in halophilic ribosomal protein genes, the ratio $K_a/(K_a + K_s)$ is more than twofold greater than that observed in nonhalophilic but homologous bacterial genes. That is, halophilic ribosomal protein genes contain an inordinately high proportion of nonsynonymous nucleotide substitutions, which result in greater numbers of amino acid replacements in the encoded proteins.

Superoxide Dismutase Gene Sequences

Is this excessive abundance of nonsynonymous nucleotide substitutions unique to these particular halophilic ribosomal protein genes, or is this a more general characteristic of halophilic genes? To address this, seven complete halophilic Mn/Fe SOD gene sequences from *Halobacterium cutirubrum* (two nonidentical genes), *Halobacterium* sp. strain GRB (two nonidentical genes), *Haloferax volcanii* (two virtually identical genes), and *Haloarcula marismortui* (one gene) and three partial *sod* gene sequences (codons 38 to 181) from *Haloferax mediterranei* (two nonidentical genes) and *Haloarcula hispanica* (one gene) were subjected to quantitative reanalysis. The gene and protein alignment is unambiguous and has been published previously (4, 8). It contains 198 internal codon positions that are common to all seven complete gene sequences. The three additional partial PCR generated sequences match the alignment perfectly between codons 38 and 181, with the exception that the *Haloferax mediterranei sod2* sequence contains two single-codon insertions between alignment positions 92 and 93 and positions 116 and 117. For purposes of comparison, we have analyzed the nine halophilic *sod* gene sequences for synonymous and nonsynonymous nucleotide substitutions between codon positions 38 and 181; within this region, the *sod1* and *sod2* genes from *Haloferax volcanii* are absolutely identical, and therefore only the *sod1* sequence was included. This region represents about 70% of the coding region of these genes. At the nucleotide and amino acid levels, the sequences are, respectively, 72 to 99% and 66 to 99% identical. In the alignment, 50% (137 of 274) of the nucleotide substitutions are nonsynonymous. From analysis of the 36 different pairwise comparisons, the average value of $K_a/(K_a + K_s)$ indicates that on a per-site basis, K_a represents 18.7% of the total substitutions in these halophilic *sod* genes (Table 3). Virtually identical results are obtained when the seven complete *sod* genes are analyzed over the 198 common codon positions; K_a represents 17.9% of the total substitutions.

Six partial Mn/Fe *sod* gene sequences from a group of related mycobacterial species were recovered from a database search and used as a comparison (32). These eubacterial *sod* sequences generated by PCR contain 163 codon positions, align end-to-end without deletion or insertion, and correspond approximately to codons 30 to 190 in the halophilic *sod* genes. At the nucleotide and amino acid levels, the sequences are respectively 79 to 93% and 81 to 97% identical. In the alignment, 32% (52 of 162) of the nucleotide substitutions are nonsynonymous. Moreover, the average value of the ratio $K_a/(K_a + K_s)$ from the 15 pairwise comparisons indicates that K_a on a per-site basis represents only 7.9% of the substitutions in the mycobacterial *sod* genes (Table 3). This value is more than twofold below the corresponding value of 18.7% for halophilic

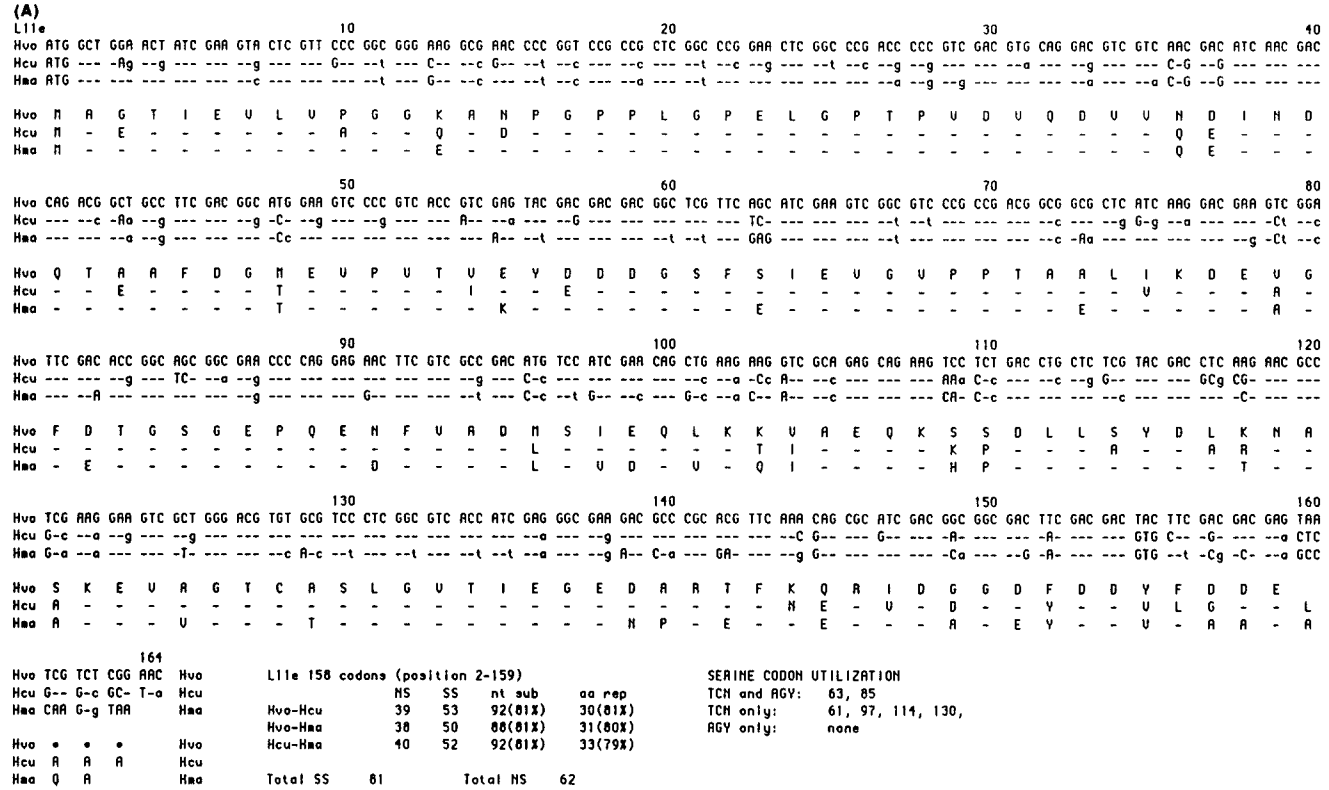
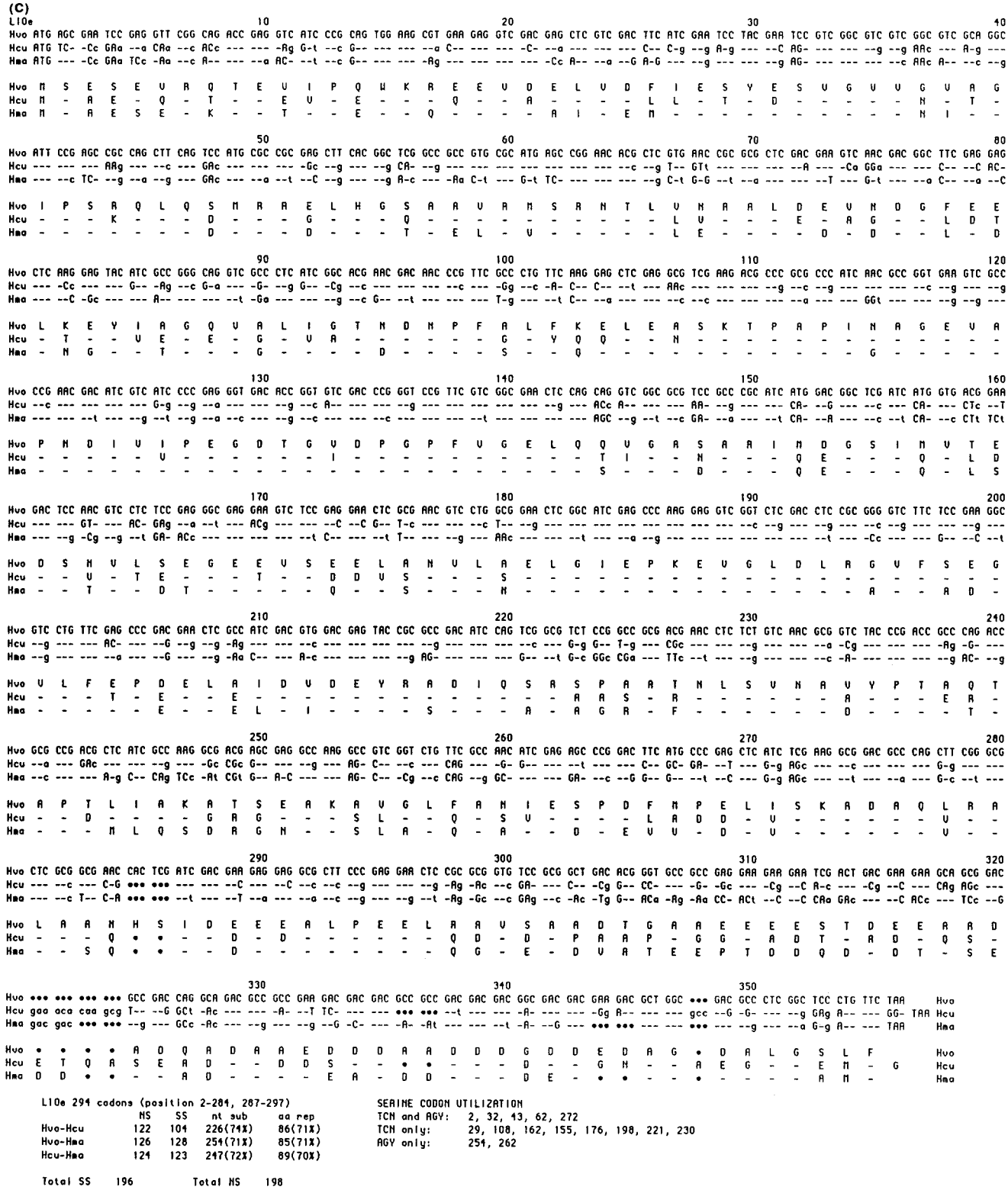

(A)
Lile
Hvo ATG GCT GGA ACT ATC GAA GTA CTC GTT CCC GGC GGG AAG GCG AAC CCC GGT CCG CCG CTC GGC CCG GAA CTC GGC CCG ACC CCC GTC GAC GTG CAG GAC GTC GTC AAC GAC ATC AAC GAC
Hcu ATG --- -Ag --g --- --- --g --- --- G-- --t --- C-- --c G-- --t --c --- --c --- --t --c --g --- --t --c --g --g --- --- --a --- --g --- --- C-G --G --- --- ---
Hma ATG --- --- --- --- --- --c --- --- --- --t --- G-- --c --- --t --c --- --a --- --t --- --- --- --- --- --a --g --g --- --- --- --a --- --a C-G --G --- --- ---
Hvo M A G T I E V L V P G G K A N P G P P L G P E L G P T P V D V Q D V V N D I N D
Hcu M - E - - - - - - - - - Q - D - Q E - - -
Hma M - - - - - - - - - - - E - Q E - - -
Hvo CAG ACG GCT GCC TTC GAC GGC ATG GAA GTC CCC GTC ACC GTC GAG TAC GAC GAC GAC GGC TCG TTC AGC ATC GAA GTC GGC GTC CCG CCG ACG GCG GCG CTC ATC AAG GAC GAA GTC GGA
Hcu --- --c -Ac --g --- --- --g -C- --g --- --g --- --- A-- --a --- --G --- --- --- --- TC- --- --- --- --t --t --- --- --- --c --- --g G-g --a --- --- -Ct --c
Hma --- --- --a --g --- --- --- -Cc --- --- --- --- --- --- A-- --t --- --- --- --t --t --- GAG --- --- --- --t --- --- --- --- --c -Aa --- --- --- --- --g -Ct --c
Hvo Q T A A F D G M E V P V T V E Y D D D G S F S I E V G V P P T A A L I K D E V G
Hcu - - E - - - - T - - - - - I - - E - - - - - - - - - - - - - - - - V - - - A -
Hma - - - - - - - T - - - - - - K - - - - - - - - E - - - - - - - - - - E - - - - A -
Hvo TTC GAC ACC GGC AGC GGC GAA CCC CAG GAG AAC TTC GTC GCC GAC ATG TCC ATC GAA CAG CTG AAG AAG GTC GCA GAG CAG AAG TCC TCT GAC CTG CTC TCG TAC GAC CTC AAG AAC GCC
Hcu --- --- --g --- TC- --a --g --- --- --- --- --- --- --g --- C-c --- --- --- --- --- --c --a -Cc A-- --c --- --- --- AAa C-c --- --c --g G-- --- --- GCg CG- --- ---
Hma --- --A --- --- --- --- --g --- --- --- G-- --- --- --t --- C-c --t G-- --c --- G-c --a C-- A-- --c --- --- --- CA- C-c --- --- --- --c --- --- --- -C- --- ---
Hvo F D T G S G E P Q E N F V A D M S I E Q L K K V A E Q K S S D L L S Y D L K N A
Hcu - - - - - - - - - - - - - - L - - - - - - T I - - - - K P - - - A - - A A - -
Hma - E - - - - - - - - D - - - - L - V D - V - Q I - - - - - H P - - - - - - - - T - -
Hvo TCG AAG GAA GTC GCT GGG ACG TGT GCG TCC CTC GGC GTC ACC ATC GAG GGC GAA GAC GCC CGC ACG TTC AAA CAG CGC ATC GAC GGC GGC GAC TTC GAC GAC TAC TTC GAC GAC GAG TAA
Hcu G-c --a --g --- --g --- --- --- --- --- --- --- --- --- --- --a --- --g --- --- --- --- --- --C G-- --- G-- --- -A- --- --- -A- --- --- GTG C-- -G- --- --a CTC
Hma G-a --a --- --- -T- --- --- --c A-c --t --- --t --- --t --t --a --- --g A-- C-a --- GA- --- --g G-- --- --- --- -Ca --- --G -A- --- --- GTG --t -Cg -C- --a GCC
Hvo S K E V A G T C A S L G V T I E G E D A R T F K Q R I D G G D F D D Y F D D E
Hcu A - N E - V - D - - Y - - V L G - - L
Hma A - - - V - - - T - - - - - - - - - N P - E - - E - - - A - E Y - - V - A A - A
164
Hvo TCG TCT CGG AAC Hvo
Hcu G-- G-c GC- T-a Hcu
Hma CAA G-g TAA Hma
Hvo • • • Hvo
Hcu R A A Hcu
Hma Q A Hma
Lile 158 codons (position 2-159)
NS SS nt sub aa rep
Hvo-Hcu 39 53 92(81%) 30(81%)
Hvo-Hma 38 50 88(81%) 31(80%)
Hcu-Hma 40 52 92(81%) 33(79%)
Total SS 81 Total NS 62
SERINE CODON UTILIZATION
TCN and AGY: 63, 85
TCN only: 61, 97, 114, 130,
AGY only: none
(B)
Lle
Hvo ATG GCA GAC ACA ••• ATA GTT GAC GCA GTC TCT CGC GCT CTC GAC GAG GCA CCC GGG CGG AAC TTC CGC GAA ACG GTT GAC CTC GCC GTG AAC TTG CGA GAC TTA GAT CTT AAC GAC CCG
Hcu ATG --- --- -Ac GAT --- -Aa --G --c --a G-- --- --a --t --G --T --c --a CA- --- --- --- --t --g --- --a --- --- --a --c --- c-- --c --- c-c --c --c --- --- ---
Hma ATG --- --t CAg GAA --- -Ag A-- --- --- --g --- --a --- --G --C --- --t -A- --- --t --- --- --- --- --c --- --- --t --- --- c-- --c --- --- --- --- --- --- ---
Hvo M A D T • I V D A V S R A L D E A P G R N F R E T V D L A V N L R D L D L N D P
Hcu M - - N D - E E - - A - - - E D - - Q -
Hma M - - Q E - E N - - - - - - E D - - E -
Hvo TCG AAG CGT GTC GAC GAG AGC ATC GTG CTC CCG TCT GGC ACC GGC CAG GAC ACC CAG ATT GTG GTG TTC GCG ACC GGT GAA ACG GCC ••• ••• GCA GAG GAT GCC GCC GAC GAA GTT CTC
Hcu --- C-a --a --- --- --- G-- G-- --- --g --- --g --- --- --a --- --g --g --- --- --c --- --t --- --a GA- --c --- --c --g GTT CGC --g --C --c -T- --t --- --C --c ---
Hma --- --C --c --- --- --- TC- G-- --- --t --t G-- --- --- --t --- --g --- ACt --- --- --c --- --c GAg --c --- --c --- CTT CGT --c --- --A -T- --a --- --C --a ---
Hvo S K R V D E S I V L P S G T G Q D T Q I V V F A T G E T A • • A E D A A D E V L
Hcu - Q - - - - G V - - - - - - - - E - - - - - - - - D - - - - V R - D - V - - D - -
Hma - N - - - - - V - - - A - - - - E - T - - - - - - E - - - - L R - - E V - - D - -
Hvo GGG CCC GAT GAG CTC GAA GAC TTC GGC GAC GAC ACC GAT GCG GCG AAA GAC CTT GCC GAC GAG ACC GAC TTC TTC GTT GCC GAG GCT GGA CTG ATG CAG GAC ATC GGT CGC TAC CTC GGT
Hcu GAc GAG GGc --C --- AGC --- C-- -Ca --- --- --- --c --c --- --g --t --c --a --- --- --g --- --- --- --g --g --a --a CCc A-- --- --- --- --- -Tg G-t GCg --- ---
Hma GAc GAG GGg --a --- --g --A C-- --t -G- --- GA- --c --c --c --g --- --- --- --t --C --t --- --- --- A-- --g --- AAg --- --- --- --- --- --- --- --- --- --- --g
Hvo G P D E L E D F G D D T D A A K D L A D E T D F F V A E A G L M Q D I G A Y L G
Hcu D E G D - S - L A - - - - - - - - - - - - - - - - - - A P M - - - - V G A - -
Hma D E G - - - E L - G - D - - - - - - - - - D - - - - I - - K - - - - - - - - - -
Hvo ACC GTG CTC GGT CCC CGT GGT AAG ATG CCG ACC CCG CTT CAG CCC GAC GAC GAC GTC GTC GAG ACG GTG AAC CGG ATG AAG AAC ACG GTG CAG CTC CGC TCG CGT GAC CGT CGG ACG TTC
Hcu CAa --- --t --- --g --c --g --a --- --- --- --c --- --- --- --- --- --- --- --C --a --c --- --c --- --a --- --c --- --- --- A-- --- AGc --c --- --c --- --- ---
Hma --- --c --- --- --g --- --g --- --- --- GAa --- --c G-C --- --- --- --- --- --- --- GTc A-c G-a --c --- --a --- --c --- --- --- --- AGc G-g --A --g --a --- ---
Hvo T V L G P A G K M P T P L Q P D D D V V E T V N A M K N T V Q L A S A D A A T F
Hcu Q - D T - - - - - - - - - - I - - - - - - - -
Hma - - - - - - - - - - E - - D - - - - - - - - V I E - - - - - - - - - - G E - - - -
Hvo CAC ACG CGC GTC GGC GCG GAC GAC ATG ACG CCC GAC GAA ATC GCG GAG AAC ATC GAC GTC ATC GTC CGT CGT CTC GAA GCG ACG CTC GAA AAG GGC CCG CTC AAC ATC GAT TCC GTC TAC
Hcu --- --- --- --- --- --- --G --- --- TCc G-- --G --C --- --c AGC --- --- --- --- --- A-g --- --g --g C-C --- -Ac --- --- --a --- --- --g --- G-g --c --- --- ---
Hma --- --- --g --- --- --- --G --- --- TCc G-- --G A-T --- --- --C --- --- --- --t --- C-- --c --- --g C-C --- GAc --- --g --- --- --- --- --- --- --- --c A-t --- ---
Hvo H T R V G A D D M T P D E I A E N I D V I V R A L E A T L E K G P L N I D S V Y
Hcu - - - - - - E - - S A E D - - S - - - - - M - - - H - N - - - - - - - U - - - -
Hma - - - - - - E - - S A E M - - D - - - - - L - - - H - D - - - - - - - - - T - -
210
Hvo GTG AAG ACG ACG ATG GGG CCG TCC GTG GAG GTG CCC GCA TGA Hvo
Hcu --- --- --a --- --- --- --t G-- --- --- --t G-- TAg Hcu
Hma --- --- --- --- --- --- --t G-g A-- --- --t G-- TG- Hma
Hvo V K T T M G P S V E V P A Hvo
Hcu - - - - - - - A - - - A Hcu
Hma - - - - - - - A M - - A Hma
Lle 208 codons (position 2-4, 6-69, 72-212)
NS SS nt sub aa rep
Hvo-Hcu 58 83 141(77%) 44(79%)
Hvo-Hma 64 66 134(79%) 45(78%)
Hcu-Hma 56 76 132(79%) 42(80%)
Total SS 123 Total NS 89
SERINE CODON UTILIZATION
TCN and AGY: 47, 154
TCN only: 11, 41, 52, 170, 198
AGY only: none

FIG. 2.



L10e 294 codons (position 2-284, 287-297)

	NS	SS	nt sub	aa rep
Hvo-Hcu	122	104	226(71%)	86(71%)
Hvo-Hma	126	128	254(71%)	85(71%)
Hcu-Hma	124	123	247(72%)	89(70%)

Total SS 196 Total NS 198

SERINE CODON UTILIZATION

TCN and AGY: 2, 32, 43, 62, 272

TCN only: 29, 108, 162, 155, 176, 198, 221, 230

AGY only: 254, 262

FIG. 2—*Continued.*

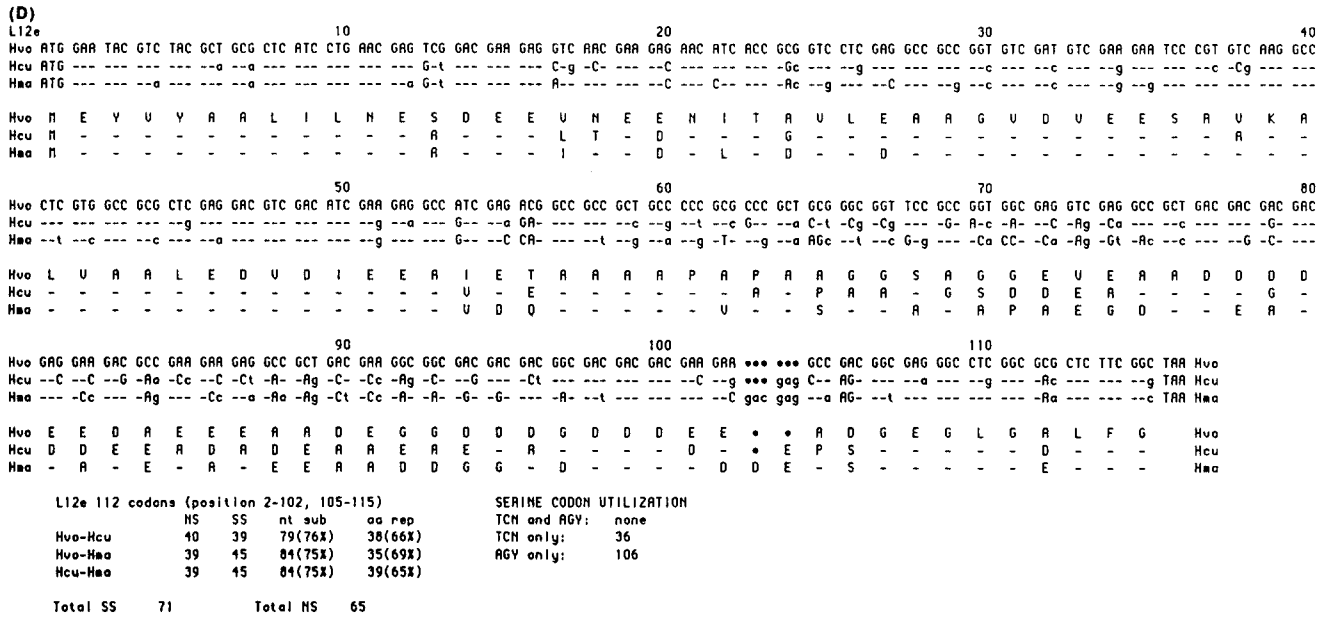


FIG. 2. Alignment of the halophilic L11, L1, L10, and L12 nucleic acid and protein sequences. The sequences of the L11 (A), L1 (B), L10 (C), and L12 (D) genes and protein from *Haloferax volcanii* (Hvo), *Halobacterium cutirubrum* (Hcu), and *Haloarcula marismortui* (Hma) are aligned. Dashes in the nucleic acid and protein alignments represent identities with the illustrated *Haloferax volcanii* sequences. Amino acids are indicated by the standard single-letter designation. In the nucleic acid alignments, capital letters in the *Halobacterium cutirubrum* and *Haloarcula marismortui* sequences represent nonsynonymous substitutions whereas lowercase letters represent synonymous substitutions when compared to the *Haloferax volcanii* sequences. Where a particular codon contains multiple substitutions and the third-position change is ambiguous, it has been indicated in lowercase letters and counted for purposes of illustration as synonymous. Solid circles represent positions of nucleotide or amino acid deletion. At the end of each alignment, the number of synonymous (SS) and nonsynonymous (NS) substitutions, the total number of nucleotide substitutions (nt sub; equal to NS + SS), and the total number of amino acid replacements (aa rep) for pairwise comparisons are indicated. The values in parentheses are the percent identities between the nucleic acid or protein sequences. For each gene, the codon positions where Ser appears in at least two of the three sequences are categorized according to serine codon utilization: (TCN and AGY) or (TCN) or (AGY). At the very bottom, the total numbers of synonymous (total SS) and nonsynonymous (total NS) substitutions in the *Halobacterium cutirubrum* and *Haloarcula marismortui* sequences when compared to the *Haloferax volcanii* sequence are indicated.

sod genes. The reason for the anomalously high frequency of nonsynonymous substitutions (K_a of 12.5%) in the *sod* gene from *M. kansasii* is unknown. The average value of $K_a/(K_a + K_s)$ for *M. kansasii* when compared to all other strains is 0.126, whereas values excluding *M. kansasii* are between 0.028 and 0.067.

A three-dimensional X-ray crystal structure has been determined for the Mn SOD from the eubacterium *Thermus thermophilus* (30). Comparisons with this structure indicate that important amino acid residues involved in metal ion binding, formation of the hydrophobic active-site cavity, catalysis, and subunit interaction are immune to replacement in both the halophilic and mycobacterial SODs (4, 9). Most of the positions subject to replacement appear to be located at or near the surface of the protein.

Phylogeny Based upon *sod* Gene Sequences

The halophile and mycobacterial *sod* gene sequences have been used to construct a phylogenetic distance tree. Halophiles serve as the outgroup for mycobacteria, and mycobacteria serve as the outgroup for halophiles (Fig. 4). Within this tree, we have calculated the value of $K_a/(K_a + K_s)$ for each branch. The values are listed at the nodes and were determined by averaging the comparisons of all representatives on one branch above the node with all representatives on the other branch (Table 3).

Within the mycobacterial group, *M. kansasii* is anomalous in that it is the deepest branch and exhibits, when compared to all other mycobacteria, the highest $K_a/(K_a + K_s)$ ratio of 0.126. As pointed out above (Fig. 1), 16S rRNA analysis positions *M. kansasii* not at the base but rather high up in the tree above *M. scrofulaceum*. We cannot explain this anomaly. With the exception of *M. kansasii*, the *sod* gene divergence within halophiles appears to be remarkably uniform, with K_a (on a percent basis) representing 4 to 6% of the total nucleotide substitution.

The situation within the halophilic archaea is more interesting and complicated. First, the divergence of the three halophilic genera is, as in the 16S rRNA tree, poorly resolved, although *Halobacterium* is the deepest branch. Second, the variation in $K_a/(K_a + K_s)$ is remarkably variable, ranging from 10 to 30% within different regions of the tree. This implies (i) that the substitutions are not derived solely from a burst of divergence occurring during the inception of the halophilic lineage and consequent adaptation to extreme halophilic conditions and (ii) that there must be some fluctuation in the selection for fixation of nonsynonymous substitutions in the *sod* genes in different halophilic lineages. In contrast to the *sod* genes, halophilic ribosomal protein genes appear to accumulate synonymous and nonsynonymous substitutions at uniform rates in all three genera (Table 2).

Third, some halophilic species have a single *sod* gene whereas other species contain paralogous duplicated genes. The most parsimonious explanation for this would be that the last common halophilic ancestors had duplicated genes and one of the genes was subsequently lost from the *Haloarcula* lineage leading to *Haloarcula hispanica* and *Haloarcula marismortui*. As expected, the single-copy *sod* genes from these two species group together, implying that they are orthologs and were derived from a relatively recent speciation event. The

TABLE 2. Nucleotide sequence divergence within halophilic (archaeal) and enteric (eubacterial) ribosomal protein genes

Gene and species[a]	% Identity		No. of:		% NS/(NS+SS)	K_s[b]	K_a[b]	% $K_a/(K_s + K_a)$
	Amino acids	Nucleotides	SS	NS				
L11								
Hfvo Hbcu	81	81	53	39	42	0.78 ± 0.16	0.12 ± 0.02	13.3
Hfvo Hama	80	81	50	38	43	0.61 ± 0.10	0.12 ± 0.02	16.4
Hbcu Hama	79	81	52	40	44	0.70 ± 0.13	0.13 ± 0.02	15.7
Eco Pvu	88	81	58	21	27	0.85 ± 0.14	0.07 ± 0.02	7.6
Eco Sma	94	90	33	11	25	0.34 ± 0.07	0.04 ± 0.01	10.5
Pvu Sma	90	84	51	16	24	0.63 ± 0.11	0.06 ± 0.01	8.7
L1								
Hfvo Hbcu	79	77	83	58	41	1.05 ± 0.22	0.13 ± 0.02	11.0
Hfvo Hama	78	79	66	64	49	0.63 ± 0.11	0.15 ± 0.02	19.2
Hbcu Hama	80	79	76	56	42	0.85 ± 0.17	0.14 ± 0.02	14.1
Eco Pvu	89	81	98	32	25	0.81 ± 0.11	0.07 ± 0.01	8.0
Eco Sma	95	88	62	15	20	0.48 ± 0.07	0.03 ± 0.01	5.9
Pvu Sma	88	80	101	35	26	0.87 ± 0.12	0.08 ± 0.01	8.4
L10								
Hfvo Hbcu	71	74	104	122	54	1.06 ± 0.29	0.22 ± 0.02	17.2
Hfvo Hama	71	71	128	126	50	1.44 ± 0.48	0.23 ± 0.02	13.8
Hbcu Hama	70	72	123	124	50	1.48 ± 0.51	0.22 ± 0.02	12.9
L12								
Hfvo Hbcu	66	76	39	40	51	0.79 ± 0.28	0.18 ± 0.03	18.6
Hfvo Hama	69	75	45	39	46	0.91 ± 0.22	0.18 ± 0.03	16.5
Hbcu Hama	65	75	45	39	46	1.03 ± 0.35	0.19 ± 0.03	15.6
Total								
Hfvo Hbcu						0.92 ± 0.11	0.18 ± 0.01	16.4
Hfvo Hama						0.85 ± 0.08	0.17 ± 0.01	16.7
Hbcu Hama						1.01 ± 0.11	0.17 ± 0.01	14.4
Eco Pvu						0.82 ± 0.08	0.07 ± 0.01	7.9
Eco Sma						0.42 ± 0.05	0.03 ± 0.01	6.3
Pvu Sma						0.77 ± 0.08	0.07 ± 0.01	8.3

[a] Abbreviations are defined in Table 1.
[b] The standard errors for the estimates of K_s and K_a are given. K_s and K_a were determined by the method of Li (13).

duplicated *sod1* and *sod2* genes of *Haloferax volcanii* are virtually identical, differing only by the use of a TCA and AGC serine codon at position 2 and the deletion of a GAC aspartic acid codon from *sod2* at position 3. In spite of this nearly perfect gene sequence identity, the 5′- and 3′-flanking regions are unrelated except for a canonical promoter element and the two genes are transcriptionally regulated in different manners (8, 9). We have suggested that the perfect identity between the *sod1* and *sod2* genes of *Haloferax volcanii* between codons 4 and 200 is the result of a very recent gene conversion which homogenized the two coding sequences but left the flanking regions intact and undisturbed.

The paralogous *sod1* and *sod2* genes from *Haloferax mediterranei* are very divergent, and their common ancestor appears to pre-date the divergence of *Haloarcula* and *Haloferax*. The *sod1* gene from *Haloferax mediterranei* groups with the nearly identical *sod1* and *sod2* genes from *Haloferax volcanii*, whereas the *sod2* gene is ancient and deeply branching. The paralogous *sod* and *slg* genes from *Halobacterium* are less ancient and have a common ancestral sequence that existed well after *Halobacterium* separated from *Haloarcula* and *Haloferax*. It seems likely that this ancestral *sod-slg* gene sequence was generated by a conversion event similar to the recent event seen in *Haloferax volcanii*. Since the time of sequence homogenization, the *Halobacterium sod* and *slg* genes have once again diverged. Divergence is evident as recently as the separation of *Halobacterium cutirubrum* from *Halobacterium* sp. strain GRB; eight substitutions have occurred in the four genes (representing 2,400 nucleotides of sequence), and at least four of these are nonsynonymous. In contrast, only one substitution was detected in a total 1,064 nucleotides of noncoding 5′ and 3′ regions flanking the *sod* and *slg* genes (8).

Synonymous and Nonsynonymous Substitution Rates

It has been suggested that small-subunit rRNA sequences diverge at a nearly constant rate of about 1% per 50×10^6 years (18). By using 16S rRNA sequences as a chronometer, we estimate that *Halobacterium*, *Haloferax*, and *Haloarcula* diverged from a common ancestor about 600×10^6 years ago. If the ribosomal protein and *sod* genes diverged at the same time as the organism and if the situation has not been complicated by genetic exchange across species, it is possible to estimate the absolute rates of synonymous and nonsynonymous nucleotide substitution. For the *sod* and L11, L1, L10, and L12 ribosomal protein genes from the three halophilic genera, the rate of synonymous substitutions per synonymous site per 10^9 years is about 0.8 and the rate of nonsynonymous substitutions per nonsynonymous site is about 0.15 (Table 4). The corresponding bacterial rates based on mycobacterial *sod* and enteric r-protein genes are about 1.2 and 0.10. Thus, these halophilic protein-encoding sequences appear to be accumulating, in absolute time, fewer synonymous and more nonsynonymous nucleotide substitutions than are the homologous sequences in mycobacteria or enteric eubacteria.

It should be noted that estimates of the synonymous and nonsynonymous mutation rates are relevant to only the genes and organisms examined. It has been shown previously, by comparison of 22 different protein-encoding sequences (representing over 20 kbp) from *E. coli* and *Salmonella typhi-*

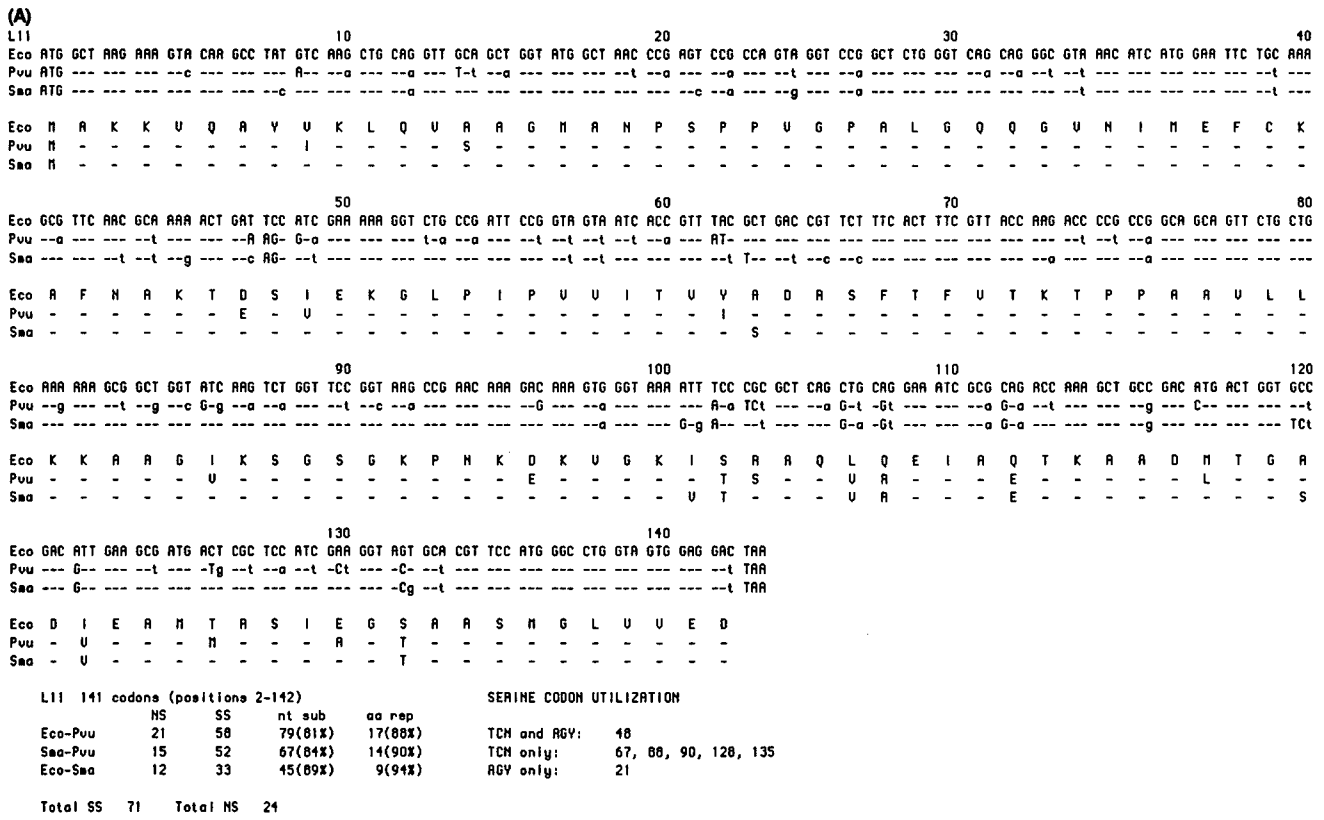


FIG. 3. Alignment of enteric L11 and L1 nucleic acid and protein sequences. The sequences of the L11 (A) and L1 (B) genes and proteins from *E. coli* (Eco), *P. vulgaris* (Pvu), and *S. marcescens* (Sma) are aligned. Dashes in the nucleic acid and protein alignments represent identities with the illustrated *E. coli* sequence. Other details are as indicated in the legend to Fig. 2.

murium, that the synonymous substitution rate can vary more than 20-fold among different genes (18); the average rate of synonymous substitution determined by Ochman and Wilson was about 6 to 8 per 10^9 years. Thus, the enteric L11 and L1 r-protein genes and the mycobacterial *sod* genes appear to exhibit a much lower synonymous substitution rate than do many eubacterial genes.

Polymorphic Positions and Serine Codon Utilization

Given the high proportion of nonsynonymous substitutions in halophilic protein-encoding genes, it is not surprising to see that the proportion of polymorphic amino acid positions within these proteins is substantially greater than in the eubacterial homologous protein. What is more interesting, however, is the proportion of polymorphic sites that contain both acidic (glutamic or aspartic) and nonacidic amino acid residues. For halophilic proteins, about 40% of the polymorphic sites contain at least one acidic residue, whereas for the homologous eubacterial proteins, only about 15% of the polymorphic site contains one or more acidic residues (Table 5). Acidic residues in halophilic proteins are known to be important determinants of protein hydrophobicity, surface hydration, and tertiary structure (5, 6, 12, 32).

Alignment positions containing serine residues are also extremely interesting and informative (2, 4). Serine is important in halophilic adaptation because of its compact size and borderline hydrophobic-hydrophilic properties (21). It is encoded by both the TCN and AGY families of codons; conversion between the two families requires at least two nonsynonymous nucleotide substitutions and presumably a nonserine protein intermediate. In both eubacteria and halophilic archaea, there appears to be no restriction as to which family is used, although TCN is somewhat more frequent than AGY. All positions in the *sod* and r-protein alignments where serine is present in at least two of the represented sequences have been examined for codon utilization. Surprisingly, at 35% of the sites where serine occurs in two or more of the aligned halophile sequences, both the TCN and AGY codons are used. In the corresponding bacterial alignments, only 9% of the positions containing multiple serines use both the TCN and AGY codons (Fig. 2 and 3; Table 6).

The nine alignment positions in the halophilic SODs that use both the TCN and AGY serine codons are illustrated in Table 7. Positions 2 and 24 are monomorphic for serine. In eubacterial SODs, these homologous positions often contain nonserine amino acid replacements. Apparently, serine has not always occupied these positions even in the halophilic proteins; that is, in some halophilic ancestral sequences a nonserine residue almost certainly occupied the position. The two different serine codon families now present presumably arose by two or more nonsynonymous substitutions of the ancestral codon. In the table of the genetic code, only the threonine (ACN) and cysteine (UGY) codons interconnect the two serine codon families by two nonsynonymous nucleotide substitutions. All other amino acid intermediates would require three or more nonsynonymous substitutions to make this connection. Alternatively, simultaneous substitutions at adjacent base pairs could interconnect the TCN and AGY codons without a nonserine intermediate. It seems unlikely that this infrequent process could explain all such instances.

FIG. 3—*Continued.*

Three of the serine polymorphic alignment positions contain at least one threonine. Position 62 contains four threonine and five serine residues. One possible explanation is that the TCN and AGY serine codons at this position arose by two independent single-step nonsynonymous substitutions from an ancestral ACN threonine codon (2). Position 87 contains six serine, one threonine, and two aspartic acid residues. Again, the threonine connects to either of the two serine codon families by a single nucleotide substitution. To generate an aspartic acid codon, at least two additional nonsynonymous substitutions are required in either the threonine or one of the two serine codons. Position 135 contains four serine, four alanine, and one threonine residues. The alanine and two serine codon families each connect to the ACN threonine codon family by a single nonsynonymous nucleotide substitution. At each of these three positions, the threonine codon is present in a lineage that branches deep within the halophilic *sod* tree (Fig. 4). This suggests that the threonine codons may have indeed been ancestral at each of these positions.

Positions 45, 121, 135, 154, and 200 are non-threonine-containing polymorphic positions. Each of these positions is complex in that there is no simple way to interconnect the three or more different codon families. It is again intriguing that two of these four positions, along with position 87, contain at least one acidic amino acid (5, 6, 7, 12, 32).

Model for the Evolution of Halophilic Proteins

When compared to bacterial homologs, the evolutionary divergence of halophilic SOD and ribosomal protein genes and probably other halophilic protein-encoding genes is anomalous in a number of respects. The halophilic sequences exhibit an inordinately high proportion (i) of nonsynonymous nucleotide substitutions that result in amino acid replacement in the encoded protein, (ii) of serine positions using both the TCN and AGY codon families, and (iii) of replacements involving acidic amino acids. In our model, we suggest that fluctuations in environmental salinity provide the driving force for fixation of these nonsynonymous substitutions in natural halobacterial

TABLE 3. Nucleotide sequence divergence within halophilic (archaeal) and mycobacterial (eubacterial) *sod* genes

Comparison[a]		% Identity		K_s	K_a	$K_a/(K_a + K_s)$
		Amino acids	Nucleotides			
Halophiles						
Hf sod1						
Hfvo sod1	*Hfme sod1*	91	88	0.481 ± 0.096	0.051 ± 0.013	0.096
Ha sod						
Hama	*Hahi*	95	93	0.219 ± 0.053	0.026 ± 0.009	0.106
Ha sod/Hf sod1						
Hahi	*Hfme sod1*	79	79	1.062 ± 0.292	0.144 ± 0.022	0.119
Hahi	*Hfvo sod1*	78	78	1.435 ± 0.911	0.157 ± 0.024	0.099
Hama	*Hfme sod1*	78	76	1.166 ± 0.272	0.157 ± 0.024	0.119
Hama	*Hfvo sod1*	77	77	1.606 ± 0.739	0.166 ± 0.024	0.094
Mean						0.108
Hfme sod2/Hfme sod1, Hfvo sod1, Ha sod						
Hfme sod2	*Hfme sod1*	63	71	0.933 ± 0.175	0.231 ± 0.030	0.198
Hfme sod2	*Hfvo sod1*	60	79	0.924 ± 0.184	0.254 ± 0.032	0.216
Hfme sod2	*Hahi*	71	72	1.161 ± 0.270	0.194 ± 0.027	0.143
Hfme sod2	*Hama*	67	70	1.250 ± 0.294	0.223 ± 0.029	0.151
Mean						0.177
Hb slg						
Hbcu slg	*HbGRB slg*	99	99	0.017 ± 0.014	0.007 ± 0.005	0.292
Hb sod						
Hbcu sod	*HbGRB sod*	99	99	0.020 ± 0.015	0.003 ± 0.003	0.130
Hb sod/slg						
Hbcu sod	*Hbcu slg*	88	90	0.218 ± 0.057	0.083 ± 0.017	0.276
Hbcu sod	*HbGRB slg*	86	90	0.223 ± 0.059	0.090 ± 0.018	0.288
HbGRB sod	*Hbcu slg*	87	91	0.192 ± 0.053	0.087 ± 0.017	0.312
HbGRB sod	*HbGRB slg*	86	89	0.205 ± 0.056	0.095 ± 0.018	0.317
Mean						0.298
Hb/Hf,Ha						
Hbcu sod	*Hfme sod1*	77	78	0.857 ± 0.172	0.162 ± 0.024	0.159
Hbcu sod	*Hfme sod2*	63	71	1.059 ± 0.245	0.233 ± 0.030	0.180
Hbcu sod	*Hfvo sod1*	78	79	0.790 ± 0.189	0.164 ± 0.024	0.172
Hbcu sod	*Hahi*	74	78	0.625 ± 0.130	0.198 ± 0.027	0.241
Hbcu sod	*Hama*	72	76	0.688 ± 0.138	0.216 ± 0.029	0.239
HbGRB sod	*Hfme sod1*	76	77	0.900 ± 0.185	0.167 ± 0.025	0.157
HbGRB sod	*Hfme sod2*	63	71	1.096 ± 0.274	0.233 ± 0.030	0.175
HbGRB sod	*Hfvo sod1*	70	79	0.837 ± 0.208	0.169 ± 0.025	0.168
HbGRB sod	*Hahi*	74	78	0.643 ± 0.137	0.198 ± 0.027	0.235
HbGRB sod	*Hama*	72	76	0.672 ± 0.134	0.216 ± 0.029	0.243
Hbcu slg	*Hfme sod1*	72	76	0.863 ± 0.174	0.188 ± 0.027	0.179
Hbcu slg	*Hfme sod2*	65	71	1.191 ± 0.306	0.216 ± 0.029	0.154
Hbcu slg	*Hfvo sod1*	71	77	0.897 ± 0.243	0.197 ± 0.027	0.180
Hbcu slg	*Hahi*	79	81	0.664 ± 0.141	0.150 ± 0.023	0.184
Hbcu slg	*Hama*	75	78	0.716 ± 0.139	0.175 ± 0.026	0.196
HbGRB slg	*Hfme sod1*	71	76	0.832 ± 0.166	0.191 ± 0.027	0.187
HbGRB slg	*Hfme sod2*	64	71	1.234 ± 0.344	0.225 ± 0.030	0.154
HbGRB slg	*Hfvo sod1*	69	77	0.842 ± 0.216	0.205 ± 0.028	0.196
HbGRB slg	*Hahi*	77	80	0.702 ± 0.150	0.159 ± 0.024	0.185
HbGRB slg	*Hama*	74	77	0.754 ± 0.146	0.184 ± 0.026	0.196
Mean						0.189
Overall mean						0.187
Mycobacteria						
Mav	*Min*	97	93	0.288 ± 0.066	0.017 ± 0.007	0.056
Msc	*Mav*	97	90	0.565 ± 0.116	0.016 ± 0.006	0.028
Msc	*Min*	95	90	0.469 ± 0.098	0.026 ± 0.008	0.053
Mean						0.041

Continued on following page

TABLE 3—*Continued.*

Comparison[a]		% Identity		K_s	K_a	$K_a/(K_a + K_s)$
		Amino acids	Nucleotides			
Mgo	*Mav*	93	89	0.541 ± 0.115	0.034 ± 0.010	0.059
Mgo	*Min*	93	88	0.574 ± 0.121	0.041 ± 0.011	0.067
Mgo	*Msc*	93	88	0.528 ± 0.102	0.036 ± 0.010	0.064
Mean						0.065
Mfo	*Mav*	91	88	0.729 ± 0.190	0.046 ± 0.012	0.059
Mfo	*Min*	91	88	0.687 ± 0.170	0.046 ± 0.012	0.063
Mfo	*Msc*	92	87	0.618 ± 0.132	0.048 ± 0.012	0.072
Mfo	*Mgo*	94	88	0.631 ± 0.130	0.028 ± 0.009	0.042
Mean						0.059
Mka	*Mav*	82	80	0.772 ± 0.172	0.097 ± 0.017	0.112
Mka	*Min*	82	81	0.658 ± 0.137	0.097 ± 0.017	0.128
Mka	*Msc*	81	81	0.678 ± 0.135	0.096 ± 0.017	0.124
Mka	*Mgo*	82	79	0.750 ± 0.152	0.107 ± 0.018	0.125
Mka	*Mfo*	79	80	0.700 ± 0.154	0.114 ± 0.019	0.140
Mean						0.126
Overall mean						0.085

[a] Abbreviations are as in Tables 1 and 2. The halophile *sod* sequences were analyzed between codons 38 and 181. The mycobacterial sequences were generated by PCR amplification of an internal 489-nucleotide region within the *sod* gene (33). This region is homologous to the region in the halophilic *sod* genes from roughly codon 30 to codon 190.

populations. In a constant environment, selection and evolution results in near-optimal protein fitness. A nonsynonymous nucleotide substitution in a gene almost always produces a protein that is less fit, and the less fit substitution is almost always eliminated from the population. In an altered environment, a nonsynonymous substitution that was previously deleterious may suddenly become advantageous and therefore have an increased probability of reaching fixation in the pop-

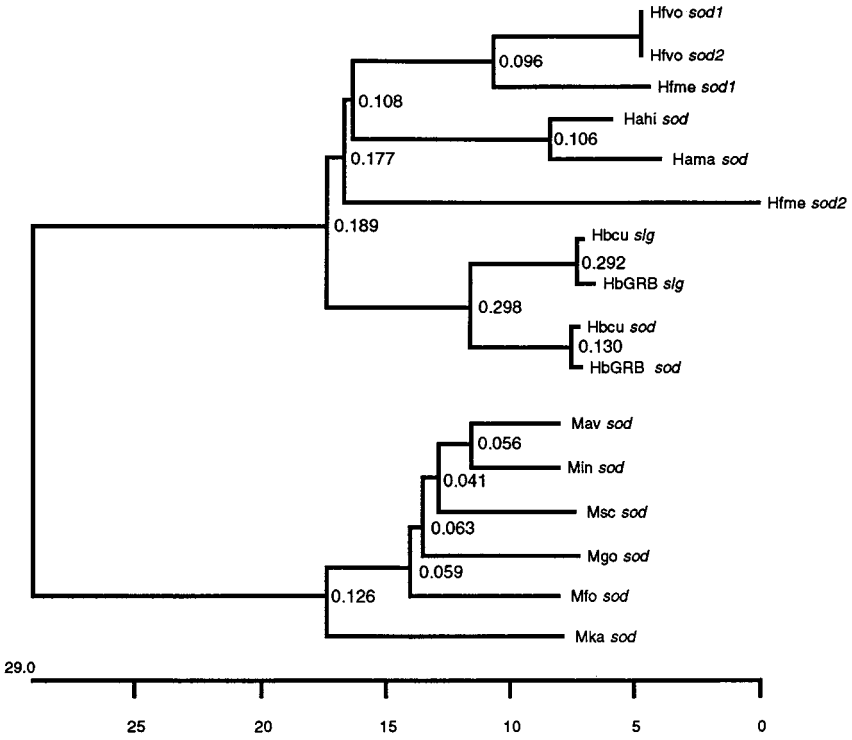


FIG. 4. Phylogenetic tree of *sod* genes. The sequences of halophile and mycobacterial *sod* genes were aligned and analyzed by the neighbor-joining method (25). Species abbreviations and database accession numbers are presented in Table 1. The average values of $K_a/(K_a + K_s)$—the fraction of the total nucleotide substitutions that are nonsynonymous on a per-site basis—are listed at each node and pertain to all sequences above the node. The values are taken from Table 3 and are calculated by comparing every sequence on one branch above the node with every sequence on the other branch above the node and averaging the values.

TABLE 4. Estimation of the rates of synonymous and nonsynonymous nucleotide substitution in halophilic and eubacterial r-protein and *sod* genes

Sequence	$K_s/10^9$ yr	$K_a/10^9$ yr
Halophilic *sod* genes[a]	0.88	0.16
Halophilic r-protein genes[b]	0.78	0.14
Mycobacterial *sod* genes[c]	1.24	0.10
Enteric r-protein genes[d]	1.11	0.09

[a] For this calculation, the *sod* gene of *Halobacterium cutirubrum*, the *sod1* gene of *Haloferax volcanii*, and the single-copy *sod* gene of *Haloarcula marismortui* were compared pairwise (data from Table 3). The divergence times were assumed to be 600×10^6 years between the three genera. To avoid potential complications with interpretation, only one *sod* gene from each organism was included in the analysis.

[b] For this calculation, the aggregate of the L11, L1, L10, and L12 ribosomal protein genes from *Halobacterium cutirubrum*, *Haloferax volcanii*, and *Haloarcula marismortui* (data from Table 2) were used. The divergence times were assumed to be 600×10^6 years between the three genera.

[c] For this calculation, the single orthologous *sod* genes from *M. fortuitum* and *M. scrofulaceum* were used. The divergence times were assumed to be 200×10^6 years between the two species.

[d] For this calculation, the aggregate of the L11 and L1 ribosomal protein genes from *E. coli*, *S. marcescens*, and *P. vulgaris* were used (data from Table 2). The divergence times were assumed to be 350×10^6 years for the *Proteus* split from *Escherichia* and *Serratia* and 200×10^6 years for the *Escherichia* split from *Serratia.*

ulation. The evolutionary outcome will probably be determined by the frequency and longevity of the environmental changes versus the evolutionary rate at which the substitutions are fixed. First, extremely slow environmental change may result in fundamental but similar changes to all proteins—the adaptation to extreme halophilism has resulted in a superabundance of acidic amino acid residues. Second, if environmental fluctuation occurs extremely rapidly, the fittest proteins are likely to be those that are functional over the widest range of conditions (although not necessarily the most efficient under any particular condition). Both of these evolutionary pathways exhibit a low nonsynonymous-to-synonymous-substitution ratio. Finally, when fluctuations occur at a rate similar to the rate of substitution, nonsynonymous substitutions leading to increased fitness under specific conditions may be fixed before the environment alters, thus leading to a high nonsynonymous-to-synonymous-substitution ratio. To understand how this scenario might apply to protein-encoding genes of halophilic archaea, it is useful to look more closely at environmental habitats and the types of replacements that confer, modify, alter, or modulate the halophilic "fitness" properties of a protein.

TABLE 5. Quantitation of the proportion of polymorphic amino acid positions that contain both acidic and non-acidic residues

Protein	Amino acid residues[a]	Polymorphic positions[b]	Polymorphic mixed acidic positions[c]	% Mixed acidic positions[d]
Halophilic SODs	144	63 (4)	26	41
Mycobacterial SODs	163	39 (3)	5	13
Halophilic L11, L1	366	86 (18)	31	37
Enteric L11, L1	373	45 (6)	8	18

[a] Number of positions in the respective alignments excluding gaps and initiation codons that were considered in the analysis.

[b] Number of positions where more than one amino acid is represented. Excluded from this number are positions that contain only glutamic acid and aspartic acid residues, since such replacements are usually considered to be interchangeable in halophilic proteins. The number of excluded glutamic acid-aspartic acid positions is given in parentheses.

[c] Number of polymorphic positions as defined in footnote 2 above that contain at least one glutamic acid or aspartic acid residue and at least one other non-acidic amino acid residue.

[d] Ratio of polymorphic mixed acidic positions (as defined in footnote *c*) to polymorphic position (as defined in footnote *b*) × 100.

TABLE 6. Serine codon utilization in halophilic and eubacterial genes

Organism and gene(s)	No. of codon positions using[a]:			
	TCN	AGY	TCN and AGY	% TCN and AGY[b]
Halophilic archaea				
sod	10	3	9	41
L11, L1	9	0	4	31
L10, L12	9	3	5	29
Total	28	6	18	35
Eubacteria				
sod	5	3	0	0
L11, L1	10	3	2	13
Total	15	6	2	9

[a] All positions in ribosomal protein and SOD alignments that contain two or more serine residues were analyzed for serine codon utilization. Positions were categorized as using TCN only, AGY only, or both TCN and AGY (Fig. 2 and 3). The nine positions in halophilic *sod* genes that use both TCN and AGY are given in Table 7.

[b] The percentage of positions containing multiple serines that utilize both TCN and AGY codons was calculated.

Hypersaline environments are between 2 and 5 M in total salt and contain Na^+, K^+, Mg^{2+}, and Ca^{2+} as the major cations and Cl^-, CO_3^{2-}, and SO_4^{2-} as the major anions (11, 22, 26). The relative proportions and concentration of these constituents vary dramatically from one locale to another; even within a single locale, composition and concentrations can fluctuate over time due to physical processes such as evaporation-dilution and precipitation-solubilization. On the one hand, decaying organic matter and sunlight make these environments rich in nutrients and energy; on the other hand, the adverse effects of high salinity on proteins and other macromolecules exclude most forms of biological exploitation. The lone exceptions to this are the halophilic archaea. These organisms thrive in hypersaline environments because they balance the internal salt concentrations with that of the external environment—they do not need to depend on energy-inefficient ion-pumping mechanisms to maintain an osmotic gradient across the cell envelope. These organisms concentrate predominantly K^+ as the intracellular cation. This is important because K^+ hydrates less water than Na^+; the major counteranion is Cl^-. In addition, the proteins of halophilic archaea have been adapted to function in high salt concentrations (12). These adaptations include a reduction in overall protein hydrophobicity and the accumulation of acidic amino acid residues. In halophilic proteins, most of the acidic residues map to the protein surface and are believed to coordinate a network of water and hydrated salt ions at the protein-solvent interface (5–7, 32). The negative charges of the carboxylates on the surface of the protein are shielded from each other by intervening solvent molecules. Indeed, in crystals of halophilic ferredoxin, water molecules are highly structured and tightly bound within the hydration shell by protein-water and water-water hydrogen bonds and by hydration of interspersed K^+ ions (7).

Acidic residues are also used to form salt bridges with strategically positioned basic residues inside the protein, where they are protected from the shielding effects of salt ions. In halophilic malate dehydrogenase, all arginine resi-

TABLE 7. Codon positions within halophilic *sod* genes utilizing both TCN and AGY codons

Gene[a]	Amino acid and codon at position:								
	2	24	45	62	87	121	135	154	200
Hbcu sod									
aa	S	S	D	T	S	S	S	S	E
Codon	TCC	AGC	GAC	ACA	AGC	AGC	AGC	AGC	GAG
HbGRB sod									
aa	S	S	D	T	S	S	S	S	E
Codon	TCC	AGC	GAC	ACA	AGC	AGC	AGC	AGC	GAG
Hbcu slg									
aa	S	S	S	T	S	G	A	S	S
Codon	AGC	AGT	AGC	ACT	ACT	GGC	GCC	AGC	TCG
HbGRB slg									
aa	S	S	S	T	S	G	A	S	S
Codon	AGC	AGT	AGC	ACT	AGT	GGC	GCC	AGC	TCG
Hfvo sod1[b]									
aa	S	S	A	S	S	G	S	S	E
Codon	TCA	TCC	GCG	TCC	TCG	GGC	TCG	AGC	GAA
Hfme sod1									
aa			A	S	S	G	S	S	
Codon			GCT	TCC	TCG	GGT	TCG	TCC	
Hfme sod2									
aa			A	S	T	W	T	A	
Codon			GCT	TCG	ACG	TGG	ACG	GCC	
Hama sod									
aa	S	S	S	S	D	S	A	S	S
Codon	TCC	TCC	TCG	AGC	GAC	TCA	GCC	TCC	AGT
Hahi sod									
aa			S	S	D	S	A	S	
Codon			TCG	AGC	GAT	TCA	GCC	TCC	

[a] The alignment of halophilic *sod* genes has been published previously (4, 8). Because of two small internal deletions of one and three codons, only 199 of the 203 codon positions were used for analysis. The *Haloarcula hispanica sod* and *Haloferax mediterranei sod1* and *sod2* gene sequences are incomplete and cover only the region between codons 38 and 181. All alignment positions specifying two or more serine residues and utilizing both TCN and AGY codons are depicted. aa, amino acid.

[b] *Haloferax volcanii* contains two *sod* genes that are virtually identical within the coding region. The only differences are the use of AGC serine codon at position 2 and the deletion of the GAC aspartic acid codon from position 3 in the *sod2* gene. The *sod1* gene is represented here.

dues not involved in the active-site formation and catalysis are believed to be involved in ionic interaction with acidic residues; these internal salt bridges provide structural rigidity to the tetrameric protein (5). In halophilic ferredoxin, additional rigidity and stability are provided by the cysteine residues at positions 63, 68, 71, and 102, which serve as ligands to the iron atoms of the two-iron–two-sulfur redox cluster (7).

It is easy to imagine that in a halophilic environment of constant salt ion composition and concentration, evolution over time would achieve a near-optimal protein fitness. Once achieved, further amino acid replacements would be rare because they would usually be deleterious and diminish fitness. It is also apparent that optimal fitness (reflected in amino acid sequence) at one salt composition and concentration is likely to be substantially different from optimal fitness at a different salt composition and concentration. We propose that these differences and fluctuations in environmental salinity, both within and between locales, are providing the driving force for fixation of the excessive number of nonsynonymous substitutions in halophilic protein-encoding genes.

What types of amino acid replacement might be expected in protein-coding sequences subject to fluctuating salinity driven selection? Replacements that affect hydrophobicity and that tinker, adjust, and rearrange the negatively charged acidic residues that determine the surface hydration properties of the protein are anticipated. The number and distribution of acidic residues over the surface of the protein appear to play a crucial role in ensuring optimal coordination of the hydrated salt ion network (5, 7). Indeed, in the halophilic protein-encoding genes examined here, nearly half of the polymorphic positions involve acidic amino acid replacement. Serine is also an important determinant of hydrophobicity because of its compact size and borderline hydrophobic-hydrophilic character. Many serine positions are subject to tinkering and adjustment as evidenced by the use of both TCN and AGY codons at these single positions (Fig. 2 and 3; Tables 6 and 7).

PERSPECTIVES

The remarkable power of adaptive evolution and natural selection has permitted living organisms to extract nutrients and energy from a wide range of environmental habitats. In moderate environments, competition from other species is the most serious restriction to exploitation, whereas in extreme environments, physical conditions detrimental to essential bi-

ological processes are the more serious limitations. Hypersaline environments represent a unique and interesting challenge to biological exploitation because of the adverse effects of salt ions on the structure and function of proteins. A high concentration of salt reduces the availability of free water because of salt ion hydration and dramatically alters the nature of essential hydrophobic and electrostatic interactions on the solvent-exposed surface of proteins. To combat these almost insurmountable difficulties, new strategies evolved that resulted in the emergence of halophilic proteins, which retain essential structure and function in solutions containing up to 5 M salt, and permitted direct biological exploitation of hypersaline environments. Characterization of many different halophilic proteins reveals that they contain an inordinately high proportion of acidic amino acid residues when compared to homologous nonhalophilic proteins. From the two X-ray crystal structures now available for halophilic malate dehydrogenase and ferredoxin, we can begin to see for the first time how the carboxylates on acidic residues are used (i) to sequester, organize, and arrange a tight network of water and hydrated K^+ ions at the surface of the protein and (ii) to form internal salt bridges with basic amino acid residues to provide internal structural rigidity to the protein.

Analysis of the phylogenetic record for halophilic protein-encoding genes suggests that the balance between fluctuating environmental salinity and optimal protein structure-function is delicate and precarious. Halophilic genes exhibit an inordinately high proportion of nonsynonymous nucleotide substitutions, and many of the resulting amino acid replacements involve the addition, removal, relocation, or rearrangement of acidic residues. The physical and chemical nature of this phenomenon, described here as "tinkering," is not well understood. It seems possible that many of the principles that govern protein structure, folding, and surface hydration can be uncovered by combining the continuous variability in protein primary structure produced by evolution and natural selection with the powerful three-dimensional technique of crystallography.

ACKNOWLEDGMENTS

This work was supported by a grant from the Medical Research Council of Canada (MT6340) to P.P.D. P.P.D. is a fellow in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

REFERENCES

1. **Arndt, E., and C. Weigel.** 1990. Nucleotide sequence of the genes encoding the L11, L1, L10 and L12 equivalent ribosomal proteins from the archaea *Halobacterium marismortui*. Nucleic Acids Res. **18:**1285.
2. **Brenner, S.** 1988. The molecular evolution of genes and proteins: a tale of two serines. Nature (London) **334:**428–430.
3. **Christian, J. H. B., and J. A. Waltho.** 1962. Solute concentrations within cells of halophilic and non halophilic bacteria. Biochim. Biophys. Acta **65:**506–508.
4. **Dennis, P. P.** 1994. The divergence of halophilic superoxide dismutase gene sequences: molecular adaptation to high salt environments, p. 175–187. *In* B. Golding (ed.), Non-neutral evolution. Chapman & Hall, London, United Kingdom.
5. **Dym, O., M. Mevarech, and J. L. Sussman.** 1995. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. Science **267:**1344–1346.
6. **Eisenberg, H., M. Mevarech, and G. Zaccai.** 1992. Biochemical, chemical and molecular genetic aspects of halophilism. Adv. Protein Chem. **43:**1–62.
7. **Frolow, F., M. Hanel, J. L. Sussman, M. Mevarech, and M. Shoham.** 1996. Insights into protein adaptation to saturated salt environments from the crystal structure of a halophilic 2Fe 2S ferredoxin. Nat. Struct. Biol. **3:**452–458.
8. **Joshi, P., and P. P. Dennis.** 1993. Characterization of paralogous and orthologous members of the superoxide dismutase gene family from genera of the halophilic archaebacteria. J. Bacteriol. **175:**1561–1571.
9. **Joshi, P., and P. P. Dennis.** 1993. Structure, function, and evolution of the family of superoxide dismutase proteins from halophilic archaebacteria. J. Bacteriol. **175:**1572–1579.
10. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. M. Munro (ed.), Mammalian protein metabolism. Academic Press, Inc., New York, N.Y.
11. **Kushner, D. J.** 1985. The Halobacteriaceae, p. 171–214. *In* C. R. Woese and R. S. Wolfe (ed.), The bacteria, vol. 8. Academic Press, Inc., New York, N.Y.
12. **Lanyi, J.** 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. Bacteriol. Rev. **38:**272–290.
13. **Li, W.-H.** 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:**96–99.
14. **May, B. P., and P. P. Dennis.** 1987. Superoxide dismutase from the extremely halophilic archaebacterium *Halobacterium cutirubrum*. J. Bacteriol. **169:**1417–1422.
15. **May, B. P., and P. P. Dennis.** 1989. Evolution and regulation of the gene encoding superoxide dismutase from the archaebacterium *Halobacterium cutirubrum*. J. Biol. Chem. **264:**12253–12258.
16. **May, B. P., and P. P. Dennis.** 1990. Unusual evolution of a superoxide dismutase-like gene from the extremely halophilic archaebacterium *Halobacterium cutirubrum*. J. Bacteriol. **172:**3725–3729.
17. **Mylvaganam, S., and P. Dennis.** 1992. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui*. Genetics **130:**399–410.
18. **Ochman, H., and A. Wilson.** 1987. Evolution in bacteria: evidence for a universal rate in cellular genomes. J. Mol. Evol. **26:**74–86.
19. **Post, L., G. Strycharz, M. Nomura, H. Lewis, and P. Dennis.** 1979. Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **76:**1697–1701.
20. **Ramirez, C., L. C. Shimmin, C. H. Newton, A. T. Matheson, and P. P. Dennis.** 1989. Structure and evolution of the L11, L1, L10 and L12 equivalent ribosomal proteins in eubacteria, archaebacteria and eucaryotes. Can. J. Microbiol. **35:**234–244.
21. **Rao, J. K. M., and P. Argos.** 1981. Structural stability of halophilic proteins. Biochemistry **20:**6536–6543.
22. **Rodriguez-Valera, F.** 1993. Introduction to saline environments, p. 1–23. *In* R. H. Vreeland and L. I. Hochstein (ed.), The biology of halophilic bacteria. CRC Press, Inc., Boca Raton, Fla.
23. **Rogall, T., J. Wolters, T. Flohr, and E. Bottger.** 1990. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. Int. J. Syst. Bacteriol. **40:**323–330.
24. **Saenger, W.** 1987. Structure and dynamics of water surrounding biomolecules. Annu. Rev. Biophys. Biophys. Chem. **16:**93–114.
25. **Saiton, N., and M. Nei.** 1987. The neighbour joining method: a new method for constructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
26. **Sehgal, S. N., and N. E. Gibbons.** 1960. Effect of some metal ions on the growth of *Halobacterium cutirubrum*. Can. J. Microbiol. **6:**165–169.
27. **Shimmin, L. C., and P. P. Dennis.** 1989. Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. EMBO J. **8:**1225–1235.
28. **Shimmin, L. C., and P. P. Dennis.** 1996. Conserved sequence elements involved in regulation of ribosomal protein gene expression in halophilic archaea. J. Bacteriol. **178:**4737–4741.
29. **Sor, F., and M. Nomura.** 1987. Cloning and DNA sequence determination of the L11 ribosomal protein of *Serratia marcescens* and *Proteus vulgaris*: translational feedback regulation of *Escherichia coli* L11 operon by heterologous L1 proteins. Mol. Gen. Genet. **210:**52–59.
30. **Stallings, W. C., K. A. Pallridge, R. K. Strong, and M. L. Ludwig.** 1984. The structure of manganese superoxide dismutase from *Thermus thermophilus HB8* at 2.4Å resolution. J. Biol. Chem. **260:**16424–16432.
31. **Yancey, P., M. Clark, S. Hand, R. Bowlus, and G. Somero.** 1982. Living with water stress: evolution of osmolyte systems. Science **217:**1214–1222.
32. **Zaccai, G., F. Candrin, Y. Haik, N. Borokov, and H. Eisenberg.** 1989. Stabilization of halophilic malate dehydrogenase. J. Mol. Biol. **208:**491–500.
33. **Zolg, J. W., and S. Philippi-Schulz.** 1994. The superoxide dismutase gene, a target for detection and identification of mycobacteria by PCR. J. Clin. Microbiol. **32:**2801–2812.