Phylogenetic relationships within the order Halobacteriales inferred from 16S rRNA gene sequences

André-Denis Girard Wright*

CSIRO Livestock Industries, Centre for Environment and Life Sciences, Private Bag 5, Wembley, Western Australia 6913, Australia

Phylogenetic relationships within the halophilic archaea were inferred from comparisons of the 16S rRNA gene sequences from 61 strains, representing 18 genera with validly published names within the order Halobacteriales. Trees produced using distance-matrix (least-squares and neighbour-joining) methods affirm with strong bootstrap support that the members of the order Halobacteriales are a monophyletic group. Ten genera were supported as monophyletic groups [Halocarcina, Halobiforma, Halococcus, Haloferax, Halorubrum, Halosimplex (multiple sequences from a single strain), Natralba, Natriinema, Natronococcus and Natronorubrum] and eight genera were represented by a single strain (Halobacterium, Halobaculum, Halogeometricum, Halomicrobium, Halorhabdus, Halosimplex, Natronobacterium and Natronomonas). The genus Haloterrigena was always paraphyletic, and the phylogenetic placement of and sister groups to Halobacterium and Natronomonas could not be resolved. Both treeing methods failed to resolve the deep branching patterns within the order Halobacteriales and the relationships between the major clades. Additional representation from the currently monospecific genera and/or the use of other macromolecules may be required to resolve the relationships within the order Halobacteriales.

The order Halobacteriales consists of a large group of aerobic microbes which live and grow in hypersaline (i.e. brine) environments such as the Great Salt Lake and the Dead Sea. Members of this order are the most halophilic organisms known to exist, thriving in environments 10 times saltier than seawater; hence, collectively they are referred to as the ‘extreme halophiles’. High salinity is toxic to most cells. However, the extreme halophiles are well adapted to their hypersaline environment, as evidenced by their predominance in these habitats, requiring at least 1·5 M NaCl for growth and 3·5–4·5 M NaCl for optimal growth (Grant et al., 2001). To prevent dehydration and maintain osmotic balance with their surroundings, the extreme halophiles have a high intracellular concentration of salts, as high as the NaCl concentration in their immediate environment.

The extreme halophiles are very diverse, existing as cocci, rods, flat discs (Halococcus), squares (Haloferax), rectangles (Halobacterium) and triangles (Halocarcina) (Grant et al., 2001). They are either Gram-negative or Gram-positive, and some are motile. Some extreme halophiles are chemoheterotrophs that can grow on sugars, carboxylic acids, alcohols and amino acids, while others are photosynthetic autotrophs that use a special pigmented membrane protein called bacteriorhodopsin (Grant et al., 2001; Oren, 2002) instead of chlorophyll to photosynthesize. Bacteriorhodopsin uses all light except for violet light, making these cells appear purple. The red coloration of hypersaline soda lakes is due to the abundance of carotenoid pigments found in the cell membranes of other halobacteria (Oren, 1994).

Limited phenotypic data often do not provide a clear means of differentiating the extreme halophiles or determining their taxonomic affinities with each other. Deciphering the evolutionary history of the extreme halophiles is essential to understand their adaptation to extreme environments and for the exploration of their novel metabolic activities. For this report, I tested the monophyly of and the phylogenetic relationships within the order Halobacteriales using 16S rRNA gene sequences from 61 extreme halophiles whose names have been validly published. This examination, involving 18 different genera, represents an extensive phylogenetic analysis of the extreme halophiles.

The 61 halophilic archaeal 16S rRNA gene sequences examined in this paper were downloaded from the GenBank database. Because the extreme halophiles are closely related to the methanogenic archaea (i.e. methanogens) and some thermoacidophiles (i.e. Thermoplasma), 12 methanogens and two thermoacidophiles were used as the outgroup (Watrous & Wheeler, 1981) to the extreme halophiles. A
total of 75 sequences were globally aligned using the Dedicated Comparative Sequence Editor (DCSE) program (de Rijk & de Wachter, 1993) and further refined by eye.

The phylogenetic software package PHYLIP (version 3.62c) (Felsenstein, 2004) was used to calculate the sequence similarity and evolutionary distances between pairs of nucleotide sequences using the Kimura two-parameter correction model (Kimura, 1980). Distance-matrix trees were then constructed using the Fitch and Margoliash least-squares (LS) method (Fitch & Margoliash, 1967) and the neighbour-joining (NJ) method (Saitou & Nei, 1987). Support values for internal branches for LS and NJ data were bootstrap resampled independently 1000 times.

When Grant & Larsen (1989) proposed a new taxonomic structure for the extreme halophiles, six genera were recognized. To date, the names of 18 genera have been validly published (Schoop, 1935; Elazari-Volcani, 1957; Tindall et al., 1984; Torreblanca et al., 1986; Kamekura & Dyall-Smith, 1995; Oren et al., 1995, 2002; McGenity & Grant, 1995; Kamekura et al., 1997; McGenity et al., 1998; Montalvo-Rodriguez et al., 1998; Ventosa et al., 1999; Xu et al., 1999; Waino et al., 2000; Hezayen et al., 2002; Vreeland et al., 2002). Despite the geographical and habitat diversity from which each of these halophiles was isolated and the evolutionary pressure to adapt to their extreme environments, distance-matrix treeing methods strongly supported (i.e. 100 %) the conclusion that the extreme halophiles were a monophyletic group (Fig. 1). Monophyly of the order Halobacteriales was also supported by a recent study of five diverse halophiles using a low-pass shotgun sequencing approach (Goo et al., 2004). In that study, strong similarities in characteristics associated with environmental exposure to UV radiation and hypersalinity were also affirmed (Goo et al., 2004).

This treatise also closely resembled previous studies by Grant et al. (2001) and Boucher et al. (2004), where they examined 16S rRNA gene sequences from a large representation of extreme halophiles, including uncultivated and unknown strains. In comparison, this study used only taxa whose names were validly published, did not include any environmental clones or unknown strains and was derived from an independent alignment. As part of this study, pairwise distance data also revealed that the greatest genetic divergence within the order Halobacteriales was 18-9 %, between Halococcus morrhuae and Halorubrum distributum JCM 9100T. In addition, the mean genetic divergence over all possible pairs of halophilic archaeal 16S rRNA gene sequences was 12-4 % (variance = 0-14 %; SD = 0-38 %). In comparison, the greatest genetic divergence within the methanogenic archaea was 34-2 %.

The LS and NJ distance-matrix methods produced nearly identical trees, except for an unsupported clade in the NJ tree consisting of three distinct branches: the Halobacterium–Halomicrobium branch, the Halosimplex–Halorhabdus branch and the Natronomonas branch. As in previous examinations (Grant et al., 2001; Boucher et al., 2004), this study was unable to resolve the deep branching patterns within the extreme halophiles. As a result, a large polytomy was created. The LS tree revealed a polytomy consisting of six major lineages, whereas the NJ tree revealed a polytomy consisting of eight major lineages. The difference in the number of major lineages was because of unresolved relationships amongst the Natronomonas clade, the Halorhabdus–Halosimplex clade and the Halomicrobium–Haloarcula clade. These unresolved branches may reflect both ‘hard’ and ‘soft’ polytomies. A hard polytomy results from the absence of data to resolve branching dichotomously and may be interpreted as multiple speciation events occurring at the same time. On the other hand, a soft polytomy reflects uncertainty resulting from the incongruence among two or more fully resolved, equally parsimonious cladograms. Additional representation from the eight genera represented by a single strain (Halobacterium, Halobaculum, Halogeometricum, Halomicrobium, Halorhabdus, Halosimplex, Natronobacterium and Natronomonas) and/or the use of other macromolecules may be required to resolve the relationships among the extreme halophiles.

All trees supported the monophyly of 10 genera: Haloarcula (100 %), Halobiforma (100 %), Halococcus (100 %), Haloferax (100 %), Halorubrum (100 %) Halosimplex (100 %), Natraliba (86 % LS, 60 % NJ), Natrinema (77 %), Natronococcus (65 %) and Natronorubrum (68 % LS, 87 % NJ). Moreover, all trees revealed a large, strongly supported clade (100 %) consisting of 18 species, representing seven different genera (Halobiforma, Haloterrigena, Natraliba, Natrinema, Natronobacterium, Natronococcus and Natronorubrum). In addition, Haloterrigena thermotolerans PR-5T was always more closely related to species belonging to the genus Natrinema than to Haloterrigena turkenvenca VKM B-1734T. Even the genetic distance between Htg. thermotolerans PR-5T and Htg. turkenvenca VKM B-1734T (d = 3-7 %) was greater than the divergence between Htg. thermotolerans PR-5T and Natrinema pellirubrum NCIMB 786T (d = 2-3 %) and the divergence between Nnm. pellirubrum NCIMB 786T and Natrinema versiforme XF10T (d = 2-7 %). However, paraphyly of the genus Haloterrigena was not unexpected, given the assertion of Tindall (2003) that the assignment of strains GSL-11 and JCM 9743 to the species Htg. turkenvenca may be unreliable as a result of errors in the original dataset.

Within the Haloarcula clade, the sequence of Haloarcula vallismortis strain A did not pair with the sequence deposited under GenBank accession no. D50851, representing an unidentified strain of Har. vallismortis. Instead, Har. vallismortis A was closer to operon rrnB of Haloarcula marismortui ATCC 43049T, whereas the unidentified strain of Har. vallismortis was closer to Haloarcula hispanica ATCC 33960T. Examination of their 16S rRNA gene sequences revealed that Har. vallismortis A was 99-6 % similar to operon rrnB (i.e. 6 bp difference) and 98-8 % similar to
operon rrnC (17 bp difference) of *Har. marismortui* ATCC 43049\(^\text{T}\). In contrast, *Har. vallismortis* A was 2-9 % divergent from *Har. vallismortis*, which in turn was 3-3 % divergent from its nearest neighbour *Har. hispanica* ATCC 33960\(^\text{T}\). These data suggest that the assignment of strain A to the species *Har. vallismortis*, and not *Har. marismortui*, may be incorrect.

Because more than one copy of the 16S rRNA gene operon exists in some members of the genera *Haloarcula* (Mylvaganam & Dennis, 1992) and *Halosimplex* (Vreeland et al., 2002), all available 16S rRNA gene operons were examined from *Har. marismortui* ATCC 43049\(^\text{T}\), *Har. hispanica* ATCC 33960\(^\text{T}\) and *Halosimplex carlsbadense* ATCC BAA-75. The three 16S rRNA genes (rrnA, rrnB and rrnC) of *Har. marismortui* ATCC 43049\(^\text{T}\) (Baliga et al., 2004) were 5 % divergent from each other, but operon rrnA did not group with either operon rrnB or operon rrnC. Instead, operon rrnA of *Har. marismortui* ATCC 43049\(^\text{T}\) branched very early within the *Haloarcula* clade, well before the divergence of *Har. marismortui* ATCC 43049\(^\text{T}\) operons rrnB and rrnC. Interestingly, the rrnB operon of

![Fig. 1. Phylogenetic relationships of 61 extreme halophiles and 14 outgroup taxa (GenBank accession numbers in parentheses) were derived from 16S rRNA gene evolutionary distances produced by the two-parameter correction model (Kimura, 1980). Trees were constructed using the LS and NJ methods, and consensus trees for 1000 bootstrap resamplings of the dataset were computed independently. Bootstrap resamplings of 16S rRNA gene sequences are included as percentages at nodes. The LS bootstrap value is followed by the NJ bootstrap value, unless both values are the same, in which case only one is shown. Asterisks indicate bootstrap values below 50 %. Branches with less than 50 % bootstrap support by both methods were collapsed. Evolutionary distance is represented by the horizontal component separating species. Bar, 5 changes per 100 positions.](http://ijs.sgmjournals.org)
Har. marismortui ATCC 43049T was more closely related to Har. vallismortis strain A than to operon rrnC of Har. marismortui ATCC 43049T. In addition, two of the three 16S rRNA gene sequences of Har. hispanica ATCC 33960T (rrnA and rrnB) were 2-4% divergent from each other, whereas an additional operon of Har. hispanica ATCC 33960T (Arahal et al., 1996) was 2-3% divergent from Har. hispanica ATCC 33960T operon rrnA and 4-2% divergent from Har. hispanica ATCC 33960 operon rrnB. In all trees, the rrnB operon of Har. hispanica ATCC 33960T branched first, and the rrnA operon of Har. hispanica ATCC 33960T was more closely related to Haloarcula argentinensis arg-1T. Furthermore, Vreeland et al. (2002) identified three operons (A, B, C) of the 16S rRNA gene of Hxs. carlsbadense ATCC BAA-75T. However, Boucher et al. (2004) reported that operon B was a chimera of the other two operons. So, only operons A and C were analysed with two additional operons (rrnA, rrnB), available from GenBank, from Hxs. carlsbadense ATCC BAA-75T. Pairwise comparisons of these four operons indicated that operon A was 99-7% similar to operon rrnB (i.e. 5 bp different) and that operon C was 98% similar to operon rrnA (19 bp different). However, operon A was 6-2% divergent from operon C and operon rrnA was 6-9% divergent from operon rrnB. This intragenomic variability was sometimes greater than the divergence between some halophiles. Given this high variability, intragenomic heterogeneity is likely to have a significant impact on the phylogenetic reconstruction of the halophiles. However, the decision of which 16S rRNA gene parologue is 'correct' must await further examination.

Based on the GenBank database, the number of 16S rRNA gene sequences for the halobacteria has more than doubled over the past 3 years. This increased level of interest in the extreme halophiles probably stems, in part, from an increase in their potential biotechnological uses (Galinski & Tindall, 1992; Rodriguez-Valera, 1992; Ventosa & Nieto, 1995; Oren, 2002). The extreme halophiles have evolved proteins that are able to function in high-ionic-strength solutions (see Oren, 1999, 2000) and their highly evolved, complex set of biosynthetic and metabolic pathways could be exploited to enable other organisms to function under salty conditions. Deciphering the evolutionary history of the extreme halophiles is essential to understand their adaptation to extreme environments and for the exploration of their novel metabolic activities.

Acknowledgements

The author thanks Dr Lucy Skillman (CSIRO Livestock Industries, Perth, Australia) and the anonymous reviewers for critical comments on previous versions of the manuscript.

References


Phylogeny of the halophilic archaea


