Further refinement of the phylogeny of the *Halobacteriaceae* based on the full-length RNA polymerase subunit B′ (*rpoB′*) gene

Hiroaki Minegishi,1 Masahiro Kamekura,2 Takashi Itoh,3 Akinobu Echigo,1 Ron Usami1,4 and Tetsuo Hashimoto5

1Bio-Nano Electronics Research Center, Toyo University, Kawagoe-shi, Saitama, Japan
2Halophiles Research Institute, Noda-shi, Chiba, Japan
3Japan Collection of Micro-organisms, RIKEN BioResource Center, Wako-shi, Saitama, Japan
4Graduate School of Interdisciplinary New Science, Toyo University, Kawagoe-shi, Saitama, Japan
5Institute of Biological Sciences, University of Tsukuba, Tsukuba-shi, Ibaraki, Japan

A considerable number of species of the *Halobacteriaceae* possess multiple copies of the 16S rRNA gene that exhibit more than 5 % divergence, complicating phylogenetic interpretations. Two additional problems have been pointed out: (i) the genera *Haloterrigena* and *Natrinema* show a very close relationship, with some species being shown to overlap in phylogenetic trees reconstructed by the neighbour-joining method, and (ii) alkaliophilic and neutrophilic species of the genus *Natrialba* form definitely separate clusters in neighbour-joining trees, suggesting that these two clusters could be separated into two genera. In an attempt to solve these problems, the RNA polymerase B′ subunit has been used as an additional target molecule for phylogenetic analysis, using partial sequences of 1305 bp. In this work, a primer set was designed that consistently amplified the full-length RNA polymerase B′ subunit gene (*rpoB′*) (1827–1842 bp) from 85 strains in 27 genera of the *Halobacteriaceae*. Differences in sequence length were found within the first 15 to 31 nt, and their downstream sequences (1812 bp) were aligned unambiguously without any gaps or deletions. Phylogenetic trees reconstructed from nucleotide sequences and deduced amino acid sequences by the maximum-likelihood method demonstrated that multiple species/strains in most genera individually formed cohesive clusters. Two discrepancies were observed: (i) the two species of *Natronolimnibius* were placed in definitely different positions, in that *Natronolimnibius innermongolicus* was placed in the *Haloterrigena*/*Natrinema* cluster, while *Natronolimnibius baerhuensis* is closely related to *Halostagnicola larsenii*, and (ii) *Natronorubrum tibetense* was segregated from the three other *Natronorubrum* species in the protein tree, while all four species formed a cluster in the gene tree, although supported by a bootstrap value of less than 50 %. The six *Haloterrigena* species/strains and the five species of *Natrinema* formed a large cluster in both trees, with *Halopiger xanaduensis* and *Nln. innermongolicus* located in the cluster in the protein tree and *Nln. innermongolicus* in the gene tree. *Hpg. xanaduensis* broke into the cluster of the genus *Halobiforma*, instead of the *Haloterrigena*/*Natrinema* cluster, in the gene tree. The six *Natrialba* species formed a tight cluster with two subclusters, of neutrophilic species and alkaliophilic species, in both trees. Overall, our data strongly suggest that (i) *Nln. innermongolicus* is a member of *Haloterrigena*/*Natrinema*, (ii) *Nrr. tibetense* might represent a new genus and (iii) the two genera *Haloterrigena* and *Natrinema* might constitute a single genus. As more and more novel species and genera are proposed in the family *Halobacteriaceae*, the full sequence of the *rpoB′* gene may provide a supplementary tool for determining the phylogenetic position of new isolates.

**Abbreviations:** ML, maximum-likelihood; NJ, neighbour-joining.

The GenBank/EMBL/DDBJ accession numbers for the *rpoB′* gene sequences determined in this study are AB477130–AB477222 and AB478421 and those of the 16S rRNA gene sequences determined in this study are AB477223–AB477234 and AB477970–AB477986, as detailed in Supplementary Table S1.

Details of strains used in this study and accession numbers of deposited sequences, alternative subtrees used in the analyses of Table 1, phylogenetic trees based on 16S rRNA gene sequences generated by maximum-likelihood and neighbour-joining methods and a table showing similarities of *rpoB′* gene sequences are available as supplementary material with the online version of this paper.
INTRODUCTION

The extremely halophilic, aerobic members of the Archaea are classified within the family Halobacteriaceae, order Halobacteriales in the class Halomicrobia of the phylum Euryarchaeota (http://www.the-iscp.org/taxa/halobacterlist.htm). At the time of writing, the family Halobacteriaceae comprised 27 genera: Haladaptatus (Hap.), Halalkalicoccus (Hac.), Haloarcula (Har.), Halobacterium (Hbt.), Halobaculum (Hbl.), Halobiforma (Hbf.), Halococcus (Hcx.), Haloferax (Hfx.), Halogeometricum (Hgm.), Halomicrobium (Hmc.), Halopiger (Hpg.), Haloplanus (Hpn.), Haloquadratum (Hqr.), Halorhabdus (Hrd.), Halorubrum (Hrr.), Halosarcina (Hsn.), Halosimplex (Hsx.), Halostagnicola (Hst.), Haloterrigena (Htg.), Halorubrum (Hrb.), Natrinema (Nnm.), Natronobacterium (Nbt.), Natronococcus (Ncc.), Natronolimnobius (Nlh), Natronomonas (Nnm.) and Natronorubrum (Nrr).

Until a few years ago, it was commonly believed that members of the Halobacteriaceae live only in hypersaline environments such as salt lakes, saline soils, subterranean salt deposits and solar salterns. More recently, it has been suggested that strains of the Halobacteriaceae can grow within saline microniches in non-saline environments, and a novel species, Halosarcina pallida, was isolated from a spring with a low salt concentration (Savage et al., 2007, 2008). Strains of the Halobacteriaceae display physiological and morphological variations. The cells are rods, cocci or flattened triangles or squares. They can be acidophilic (Minegishi et al., 2008), neutrophilic or alkalophilic. Many strains grow in a simple medium with sucrose and glutamate, or in a medium containing pyruvate/glycerol and ammonia, while some strains require 16 amino acids for growth. Most strains lyse instantaneously in hypotonic solution, but some strains survive in 0.5 % salt solution for several days (Fukushima et al., 2007).

Significant intragenomic 16S rRNA gene sequence heterogeneity of 2–5 % is a rather common feature of species of the Halobacteriaceae, as first demonstrated by Myl Wagneram & Dennis (1992). Significant heterogeneities have been detected in species of Haladaptatus, Haloarcula, Halobaculum, Halomicrobium, Halorubrum and Halosimplex. Intragenomic heterogeneity of more than 9 % (Cui et al., 2009) may make phylogenetic interpretation of 16S rRNA gene sequences very complicated (Boucher et al., 2004).

Two problems have been pointed out in the taxonomy of the Halobacteriaceae based on 16S rRNA gene sequences: (i) the status of the genera Haloterrigena and Natrinema (Tindall, 2003) and (ii) possible separation of alkalophilic and neutral species of the genus Natrinema (Xu et al., 2001). The genera Haloterrigena and Natrinema were created at almost the same time by Ventosa et al. (1999) and McGinity et al. (1998), respectively. The genus Natrinema included two species, Natrinema pellirubrum and Natrinema pallidum, and a closely related strain, GSL-11. The genus Haloterrigena was established with Haloterrigena turkmenica as the type species, with strain GSL-11 as a strain of Htg.

turkmenica, suggesting a close relationship between the two genera. As the numbers of species of the two genera increased, some species were shown to overlap each other in phylogenetic trees reconstructed by the neighbour-joining (NJ) method. Nnm. pellirubrum and Nnm. pallidum were positioned between Haloterrigena strains (Xin et al., 2000; Romano et al., 2007). Another problem is that alkalophilic and neutrophilic species of the genus Natrinema formed definitely separate clusters in trees reconstructed by the NJ method, suggesting that these two clusters could be separated into two genera (Xu et al., 2001; Itoh et al., 2005).

Thus, there was a need for other targets for chemotaxy and complementary molecular markers for the phylogeny of the Halobacteriaceae (Wright, 2006). Recently, the DNA-dependent RNA polymerase subunit β (in bacteria) or B (in archaea) gene has become popular as a phylogenetic marker (Dahllo¨f et al., 2000; Adékambi et al., 2003; Korczak et al., 2004; Case et al., 2007). In members of the Halobacteriaceae, RNA polymerase subunit B has been shown to be split into smaller subunits, B′ and B′′ (Leffers et al., 1989). The subunits are arranged in the order H, B′′, B′ and A′. Subunit B′ is one of the important components of the transcription apparatus and the gene (rpoB′) is a single-copy conserved gene, highly constrained to evolve at a reasonably slow rate. The fact that only a single copy of each subunit is present in all bacteria and archaea is a tremendous benefit over the 16S rRNA gene for phylogenetic analyses (Acinas et al., 2004; Cilia et al., 1996).

The previous work on rpoB′ sequencing (1305 bp) of 23 strains of the Halobacteriaceae by Walsh et al. (2004) demonstrated that the RpoB′ protein sequence may be an appropriate alternative phylogenetic marker to the 16S rRNA gene. They demonstrated that the gene provided a similar degree of phylogenetic resolution as the 16S rRNA gene, yet does not suffer from the problem of paralogy, by showing that compositional bias of the nucleotide and amino acid sequences did not affect their phylogenetic analyses. Subsequently, Enache et al. (2007) accumulated 17 more rpoB′ gene sequences (1305 bp, as reported by Walsh et al., 2004). Phylogenetic analysis demonstrated that the rpoB′- and RpoB′-based phylogenies were mostly congruent with the 16S rRNA gene-based phylogeny, but some incongruence was also observed. However, with some strains, the primers were not able to amplify the rpoB′ gene, or primers for sequencing the amplified genes did not work.

The phylogenetic analyses by Walsh et al. (2004) and Enache et al. (2007) consistently recovered a monophyletic group, clade I, with high bootstrap support, which was embedded within a collection of less well-resolved lineages in trees based on sequences of the rpoB′ gene, the deduced RpoB′ protein and the 16S rRNA gene. Clade I consisted of the genera Halobiforma, Haloterrigena, Natrinema, Natrinema, Natronobacterium, Natronococcus, Natronolimnobius and...
**METHODS**

**Strains and DNA extraction.** Strains used in this study are listed in Supplementary Table S1, available in IJSEM Online. At the time that we finished this study, the genus *Haloterrigena* had not been described, and the authors were not able to obtain strains of several species of a few genera. The strains were cultivated in 3 ml of the media recommended for each strain by the Japan Collection of Microorganisms. Cells were harvested by centrifugation and suspended in TEN buffer (10 mM Tris/HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl) and 0.3 g glass beads were added. The cells were broken by shaking for 20 min on a vortex mixer at maximum speed. Nucleic acids were extracted by phenol/chloroform treatment and ethanol precipitation.

**PCR amplification and sequencing of the rpoB gene.** An approx. 2 kbp segment from the 3’ end of the rpoB gene to the 5’ end of the rpoA gene was amplified by PCR using the following 50 μl mixture: 5 μl 10 x ExTaq PCR buffer, 5 μl dNTPs (2.5 mM each), 1 μl forward primer, 1 μl reverse primer (100 μM each), 36.5 μl distilled water, 0.5 μl ExTaq polymerase (TaKaRa) and 1 μl template DNA (about 200 ng DNA). A primer set, HrpoB2-1420F (5’-TGTGGGCTNGTGAAGAACTT-3’ and HrpoA 153R (5’-GACGCCGACCTAGTGGAGGA-5’) (Hbt. salinarum 1454–1473; E. coli 1521–1540). The authenticity of this primer set has been proved by the accumulating genome sequences of strains of the *Halobacteriaceae*. Accession numbers are listed in Supplementary Table S1.

**Multiple sequence alignment and phylogenetic analyses.** Multiple alignments of the gene sequences were done using CLUSTAL_X version 2.0.9 (Larkin et al., 2007) and edited manually, if required in the case of 16S rRNA gene sequences, to remove gaps and ambiguously aligned characters. The rpoB gene sequences were translated into amino acid sequences by EnzymeX version 3.1 (http://mekentosj.com). Pairwise sequence similarities were calculated with GENEVX-MAC version 14.0.11 (GENETYX Corporation). ML analyses were performed with RAxML version 2.2.3, using the GTR + Γ and WAG + Γ models, respectively, for the nucleotide and amino acid sequence-based analyses (Stamatakis et al., 2005). Support values for internal branches of the ML tree were obtained by bootstrapping (1000 and 100 replicates for nucleotide and amino acid sequence-based analyses, respectively) using the GTR + Γ and WAG + Γ models, respectively (Stamatakis et al., 2005). The CONSENSE program in the PHYLIP package (http://evolution.gs.washington.edu/phylip.html) was used to calculate bootstrap values.

First, analyses of the rpoB gene and RpoB protein were performed by using 89 sequences. Second, based on the results of the first analyses, relationships among 34 species from clade I, which was redefined in this work (see Results), were analysed by using eight outgroup species. An optimal ML tree based on 16S rRNA gene sequences was also obtained by using the same taxon sampling.

Nucleotide and amino acid compositions for individual datasets were calculated and compared by a chi-squared test using the PUZZLE program (Schmidt et al., 2002). The approximately unbiased (AU) test (Shimodaira, 2002) in the CONSEL program (Shimodaira & Hasegawa, 2001) was used for statistical comparisons among alternative trees of interest. The significance level was set at P<0.05.

**RESULTS**

**Alignment of rpoB gene sequences**

Several sets of degenerate primers were designed based on an alignment of the sequences of RNA polymerase gene subunits B*, B’ and A of the four strains of the *Halobacteriaceae* whose genome sequences were available. Primers HrpoB2-1420F/HrpoA-153R were chosen because they consistently produced a single PCR product of the expected size (1.9 kb). The sequences upstream of the ATG corresponding to the N-terminal methionine and downstream of the termination codon were trimmed off, and the full-length rpoB sequences were translated into amino acid...
Full-length rpoB' phylogeny in the Halobacteriaceae

The diagram shows the phylogenetic relationships among various species within the Halobacteriaceae family, based on full-length rpoB sequences. The tree is rooted and includes several key species and strains, such as Halococcus marismortui, Halobacterium salinarum, and Halobacterium sp. Halobacterium sp. NRC-1. The phylogeny is supported by bootstrap values, with higher values indicating greater confidence in the branching order.

Clade I includes species such as Halobacterium salinarum and Halobacterium sp. NRC-1. Other notable species include Halococcus marismortui and Halobacterium salinarum, which are located at the base of the tree, indicating their basal position within the family.

The tree is further divided into various branches, each representing a different evolutionary lineage. The phylogenetic relationships are inferred from DNA sequence data, and the tree topology is consistent with previous studies on the family Halobacteriaceae.

This phylogenetic framework is crucial for understanding the evolutionary history and diversification of halobacteria, which are known for their unique ability to tolerate high salt concentrations and extreme environmental conditions.
The lengths of the 85 *rpoB* sequences determined in this study and the four derived from the genome sequences were 1827 bp (18 sequences), 1830 bp (59), 1833 bp (four), 1836 bp (four) and 1842 bp (one). The gene lengths of species of genera with multiple species were the same, except for the genus *Halococcus*, which varied from 1827 to 1842 bp (Supplementary Table S1). Alignment of the deduced amino acid sequences, however, demonstrated clearly that the differences among sequence lengths were not inside the RpoB coding sequence, but were concentrated within the first 15–31 nt. The downstream sequences [1812 bp, beginning with the sequence CGMGAGMC, or RD(E)A in amino acid sequences] were aligned unambiguously without any gaps or deletions until the 3′ termini.

**Phylogenetic trees based on 89 sequences of the *rpoB* gene and the RpoB protein**

Phylogenetic trees reconstructed from nucleotide sequences (Fig. 1) and deduced amino acid sequences (Fig. 2) by the ML method demonstrated that the sequences of the six genera outside clade I that were represented by multiple species/strains formed cohesive clusters individually: *Halococcus* (seven species), *Haloarcula* (seven species and two strains), *Halofexer* (nine species), *Halorubrum* (16 species), *Halobacterium* (two species and one strain) and *Halalkalicoccus* (two species).

The species of the problematic genera described in the Introduction, *Haloterrigena*, *Natrinema* and *Natrialba*, were included in clade I (indicated in Figs 1 and 2; Walsh et al., 2004; Enache et al., 2007). In these trees, species of the recently described genera *Halopiger*, *Halostagnicola* and *Halovivax* were also shown to constitute a monophyletic group with clade I with high bootstrap support (100 and 96 %, respectively, for Figs 1 and 2) and, thus, clade I was redefined to include these three genera, i.e. 34 species of 11 genera.

The six *Natrialba* species formed a tight cluster in the gene tree (Fig. 1) with two subclusters, of three neutrophilic species and three alkaliphilic species. The protein tree (Fig. 2) also reconstructed the same relationship, but bootstrap support for the monophyly of *Natrialba* and for the separation of the neutrophilic and alkaliphilic species was not high.

Two further discrepancies were apparent within clade I, in addition to the problems of *Haloterrigena/Natrinema* and *Natrialba*. (i) The two species of the genus *Natronolimnobius* were placed in definitely different positions and (ii) *Natronorubrum tibetense* was segregated from the other three species of the genus in the RpoB′ tree (Fig. 2). The interrelationships of the *Haloterrigena/Natrinema* cluster are described below.

**Trees for clade I**

By focusing on the subtree of the redefined clade I and eight outgroup species (of genera *Halococcus*, *Halalkalicoccus* and *Halobacterium*), the interrelationships within clade I were further analysed individually by the ML method using the sequences of the *rpoB* gene, the RpoB′ protein and the 16S rRNA gene (Fig. 3).

In the *rpoB* gene and RpoB′ phylogenies, two discrepancies between phylogeny and taxonomy were apparent within clade I. Firstly, *Natronolimnobius innermongolicus* was located within the *Haloterrigena/Natrinema* cluster, while *Natronolimnobius baerhuensis* was located as a sister group to *Halostagnicola larsenii*, although both had no clear bootstrap support. Secondly, *Nrr. tibetense* was segregated from the other three species of the genus in the RpoB′ analysis, while the four *Natronorubrum* species formed a cluster in the *rpoB*′ gene tree, although the bootstrap support was less than 50 %.

In the 16S rRNA gene tree (Fig. 3c), however, all species of the genera *Natronolimnobius* and *Natronorubrum* formed clusters with more than 90 % bootstrap support. On the other hand, neither *Haloterrigena nor Natrinema* was monophyletic; the 11 sequences of *Haloterrigena/Natrinema* strains formed a loose clade, though with no clear bootstrap support.

The chi-squared test performed by the PUZZLE program showed that no compositional bias was detected for the deduced amino acid sequences of the RpoB′ protein or the 16S rRNA gene sequences of the 42 species, while significantly different compositions were detected for *Hbt. salinarum* (P=0.0129) and *Nln. baerhuensis* (P=0.0288) in the *rpoB*′ gene analysis. *Hbt. salinarum* was one of the outgroup species in this analysis, and its base compositional bias may therefore not contribute much to the phylogeny of the ingroup. The significantly different base composition in *Nln. baerhuensis* could not be the reason for the failure of the two *Natronolimnobius* species to form a clade, since the amino acid-based RpoB′ phylogeny also failed to reconstruct the monophyly of *Natronolimnobius*.

In order to address the three problems of *Natronolimnobius*, *Natronorubrum* and *Haloterrigena/Natrinema*, the tree of clade I was further analysed by using statistical tests. The AU test was performed to detect the significance of the log-likelihood differences between each optimal tree in the analysis of the RpoB′, *rpoB*′ gene or 16S rRNA gene datasets (Fig. 3) and five alternative trees of interest (see Table 1). (i) If *Nln. innermongolicus* was constrained to the branch leading to *Nln. baerhuensis* (tree 1 of Table 1), the log-likelihood difference was significant (P<0.01) in the RpoB′ protein analysis while, in the *rpoB*′ gene analysis, tree 1 was
Fig. 2. Optimal ML tree inferred from deduced RpoB’ protein sequences for the family Halobacteriaceae. The evolutionary model employed in the analysis was WAG+Γ. See legend to Fig. 1 for further details.

also unlikely, but statistical significance could not be detected for the log-likelihood difference between the tree in Fig. 3(a) and tree 1 (P=0.062). If *Nln. baerhuensis* was constrained to the *Nln. inermongolicus* branch (tree 2), the log-likelihood difference was significant in the analyses of the both the rpoB gene and the RpoB’ protein (P<0.001 and <0.01, respectively). These results suggest that *Nln. inermongolicus* was unlikely to be monophyletic with *Nln. baerhuensis*.

(i) If *Nrr. tibetense* was constrained to the common ancestor of the other three *Natronorubrum* species (tree 3) in the RpoB’ analysis, the log-likelihood difference of tree 3 from that of the tree in Fig. 3(b) was not significant (P=0.299), while, if the common ancestor of the other three *Natronorubrum* species was moved to a branch leading to *Nrr. tibetense* (tree 4), the log-likelihood difference was also not significant (P=0.186), suggesting that we could not exclude the possibility that *Nrr. tibetense* is monophyletic with the other *Natronorubrum* species, as is reconstructed by the rpoB’ gene analysis.

(ii) If constraints were made to separate the genera *Natrinema* and *Haloterrigena* clearly, as shown in Supplementary Fig. S1 (tree 5), the log-likelihood difference was highly significant (P<0.001) in all three analyses, demonstrating that neither of the genera *Natrinema* and *Haloterrigena* is monophyletic.

(iii) If constraints were made to separate the genera *Natrinema* and *Haloterrigena* clearly, as shown in Supplementary Fig. S1 (tree 5), the log-likelihood difference was highly significant (P<0.001) in all three analyses, demonstrating that neither of the genera *Natrinema* and *Haloterrigena* is monophyletic.

**DISCUSSION**

In this study, we have designed an excellent amplification primer set and eight sequencing primers that worked perfectly to determine full-length sequences of the rpoB’ genes of 85 strains of the Halobacteriaceae. The previous phylogenetic analyses by Walsh et al. (2004) and Enache et al. (2007) were performed on partial (1305 bp) rpoB’ gene sequences. Enache et al. (2007) pointed out a few important problems. In an rpoB’ gene tree reconstructed by the NJ method, the four species of the genus *Natrialba* analysed at that time formed a cluster, while, in an RpoB’ protein tree, the three alkaliphilic species of the genus *Natrialba* formed a tight group, while a neutrophilic species, *Natrialba asiatica*, formed a separate group with species of genera *Natronorubrum* and *Natronolimnobius*. Similar NJ trees reconstructed from 16S rRNA gene sequences had been published before (Xu et al., 2001; Itoh et al., 2005), suggesting a possible separation of the alkaliphilic and neutrophilic species into two genera. In the present study, we determined the full-length rpoB’ sequences of all six species of the genus *Natrialba*. In an rpoB’ NJ tree (not shown) reconstructed from the same dataset used for Fig. 1, a tight cluster was obtained consisting of the six species, while, in an RpoB’ protein NJ tree (not shown), the three alkaliphilic species formed a tight group and the three neutrophilic species formed a separate group on a branch with species of genera *Natronorubrum*. In all of our ML trees (Figs 1, 2 and 3), however, the monophyly of the genus *Natrialba* was reconstructed, but the support for the clade differed significantly between the analyses. High bootstrap support (more than 90%) was detected in the rpoB’ gene analysis (Fig. 1 and Fig. 3a). An ML tree based on 16S rRNA gene sequences (Supplementary Fig. S2) supported the monophyly of all species, while an NJ tree (Supplementary Fig. S3) gave different topologies of the alkaliphilic and neutrophilic species. Further data and analyses would be necessary to confirm the monophyly of *Natrialba* robustly.

Another problem posed by Enache et al. (2007) concerns the genus *Natronorubrum*. They showed that *Nrr. tibetense* differed from the other two species included in their study in NJ trees. Our ML protein tree showed a similar topology (Fig. 1), but the four *Natronorubrum* species formed a
Fig. 3. ML trees inferred from *rpoB* gene (a), deduced RpoB protein (b) and 16S rRNA gene (c) sequences for clade I of the family Halobacteriaceae. Support for nodes in tree corresponds to bootstrap values for 100 and 1000 replicates for the amino acid and nucleotide sequence-based analyses, respectively; only values greater than 50% are displayed. The outgroup was composed of eight species: *Haloterrigena jeotgali*, *H. tibetensis*, *Halobacterium salinarum*, *Halococcus hamelinensis*, *Halococcus morrhuae*, *Halococcus qingdaonensis*, *Halococcus saccharolyticus* and *Halococcus salifodinae*. X, Y and Z represent the subtrees in which the genera *Haloterrigena* and *Natrinema* appear.
cluster in the rpoB' and 16S rRNA gene trees (Supplementary Figs S2 and S3). At present, it is not possible to draw any definite conclusions from the available dataset. However, it is noteworthy that our AU test-based analyses for rpoB' did not reject the monophyly of *Natronorubrum* (Table 1).

**Table 1.** AU tests of log-likelihood differences between each optimal tree and alternative trees of interest

The AU test was performed to detect the significance of log-likelihood differences between each optimal tree in analyses of the rpoB' gene, RpoB' protein or 16S rRNA gene dataset and five alternative trees of interest. Tree 1, *Nln. innermongolicus* was removed to the position of *Nln. baerhuensis*; tree 2, *Nln. baerhuensis* to *Nln. innermongolicus*; tree 3, *Nrr. tibetense* to the common ancestor of the other *Natronorubrum* species; tree 4, the common ancestor of the other *Natronorubrum* species to *Nrr. tibetense*; tree 5, the species of *Haloterrigena* were separated from those of *Natrinema*. Δ indicates the log-likelihood differences of alternative topologies from each optimal ML tree (Fig. 3). *P*-values of the AU test were estimated by CONSEL.
In the NJ gene and protein trees reconstructed by Enache et al. (2007), the type strains of the two species of the genus Natronolimnobius behaved as very close relatives (99.8 and 99.3 % similarity, respectively), although the similarity of their 16S rRNA gene sequences is 95.9 %. \textit{N. baerhuenis} has two known strains, JCM 12253\textsuperscript{T} and JCM 12254, while \textit{N. innermongolicus} consists of one strain, JCM 12255\textsuperscript{T}. In the present study, we prepared fresh DNA from new ampoules of JCM 12253\textsuperscript{T} and JCM 12255\textsuperscript{T}. The sequences obtained in our study clearly demonstrated that the sequence represented by GenBank accession no. AB295636 of \textit{N. innermongolicus} JCM 12255\textsuperscript{T} was wrong, and the two type strains, JCM 12253\textsuperscript{T} and JCM 12255\textsuperscript{T}, were statistically significantly split in individual analyses of RpoB’ protein (Fig. 3b and Table 1) and \textit{rpoB}’ gene (Fig. 3a and Table 1) sequences, suggesting strongly that the \textit{rpoB}’ (RpoB’) sequences of \textit{N. baerhuenis} and \textit{N. innermongolicus} are not monophyletic.

The biggest problem in clade I is the issue of \textit{Haloterrigena} and \textit{Natrinema}. In this paper, full-length \textit{rpoB}’ genes were amplified from all species of the two genera. The three ML trees based on our data showed that the six \textit{Haloterrigena} species/strains and the five \textit{Natrinema} species formed a loose, intercrossing cluster (Figs 1, 2 and 3). Species of \textit{Haloterrigena} and \textit{Natrinema} never formed individual clusters, suggesting that the species of the two genera might constitute a single genus. The trees also suggested that \textit{N. innermongolicus} should be incorporated into the \textit{Haloterrigena}/\textit{Natrinema} cluster if the species of the two genera are merged in future. It may be worth pointing out here that, in 16S rRNA gene NJ trees published in recent papers proposing novel species of the genera \textit{Haloterrigena} and \textit{Natrinema}, incomplete sequences as short as 1344 bp or sequences with 18–20 ambiguous bases have been used, giving a seemingly rational separation of the two genera. Our Supplementary Figs S2 and S3 reconstructed in this study, based on sequences longer than 1421 bp with very few ambiguous bases, clearly demonstrate the intercrossing between \textit{Haloterrigena} and \textit{Natrinema} species. A problematic observation is the wide range of G+C contents, from 59.8 mol\% in \textit{Htg. turkmenica} VKM B-1734\textsuperscript{T} (Ventosa et al., 1999) to 69.9 mol\% for the major DNA component (60.0 mol\% for the minor component) of \textit{Nnm. pellirubrum} (McGinty et al., 1998; Ross & Grant, 1985). This range of values far exceeds that commonly encountered within a single genus. It is noteworthy that the G+C content range of species of the genus \textit{Halobacterium} has recently been widened: \textit{Halobacterium piscisalii}, 65.5 mol\%; \textit{Hbt. salinarum}, 67.1–71.2 mol\% (Grant et al., 2001); \textit{Hbt. jilantiaense}, 64.2 mol\%; \textit{Hbt. noricense}, 54.5 mol\%. Future studies are needed to resolve this issue.

Quite recently, a challenging paper was published by Adékambi et al. (2008). They compared published values for DNA–DNA hybridization of strains of 230 bacterial species representative of 45 genera with similarity of \textit{rpoB} gene sequences (3400–4100 bp) retrieved from GenBank. They observed that \textit{rpoB} gene sequence similarity of less than 85.5 % correlated with membership of different genera. Our present phylogenetic analyses are based on gene sequences of \textit{rpoB}’ (approx. 1830 bp), a smaller subunit of RpoB in the \textit{Halobacteriaceae}. We calculated similarities of 1812 bp \textit{rpoB}’ gene sequences amongst species/strains within each genus (see Results). The genus \textit{Haloarcula}, with the highest similarities, more than 94.4 %, gave a very tight cluster with short branches in Figs 1 and 2. On the other hand, the cluster of the genus \textit{Halococcus} was very loose, with long branches within the cluster, reflecting similarities as low as 86–87 %. This may suggest that the species of the genus \textit{Halococcus} could be divided into different genera if differentiating characteristics between them are recognized. Calculation of similarities amongst the type species of 26 genera (Supplementary Table S2) showed that the highest value was obtained between \textit{Htg. turkmenica} and \textit{Nnm. pellirubrum}, suggesting very strongly that the two genera could be merged into a single genus.

According to the genome sequences, only single copies of the \textit{rpoB}’ and \textit{rpoB}’ genes are present in the nine strains of the \textit{Halobacteriaceae} mentioned in the Results. Similarities amongst the nine concatenated \textit{rpoB}’ and \textit{rpoB}’ gene sequences (3378 bp) were less than 86.2 %, the value between \textit{Har. marismortui} and \textit{Hmc. mukohatae}. The value of 86.2 % is very close to that of 85.5 % detected in bacterial strains (Adékambi et al., 2008), and may serve as a critical value for differentiation of genera of the \textit{Halobacteriaceae}. Full-length sequences of \textit{rpoB}’ and \textit{rpoB}’ genes of all strains of clade I may suggest clearer solutions to the long-standing problems in the family \textit{Halobacteriaceae} discussed above.

As more and more novel species and genera are proposed in the family \textit{Halobacteriaceae}, full-length sequences of the \textit{rpoB}’ gene may be useful as a supplementary tool in determining the phylogenetic position of new isolates.

REFERENCES


Fukushima, T., Usami, R. & Kamekura, M. (2007). A traditional Japanese-style salt field is a niche for haloarchaeal strains that can survive in 0.5 % salt solution. Saline Syst 3, 2.


