Phylogeny and evolution of the family *Ectothiorhodospiraceae* based on comparison of 16S rRNA, *cbbL* and *nifH* gene sequences

Tatjana P. Tourova,1 Elizaveta M. Spiridonova,2 Ivan A. Berg,2 Natalia V. Slobodova,3 Eugenia S. Boulygina3 and Dimitry Yu. Sorokin1,4

Correspondence
Tatjana P. Tourova
ttour@biengi.ac.ru

1Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia
2Department of Microbiology, Moscow State University, Moscow, Russia
3Center Bioengineering, Russian Academy of Sciences, Moscow, Russia
4Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

The occurrence of genes encoding nitrogenase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) was investigated in the members of the family *Ectothiorhodospiraceae*. This family forms a separate phylogenetic lineage within the *Gammaproteobacteria* according to 16S rRNA gene sequence analysis and mostly includes photo- and chemoautotrophic halophilic and haloalkaliphilic bacteria. The *cbbL* gene encoding the large subunit of 'green-like' form I RubisCO was found in all strains, except the type strains of *Alkalispirillum mobile* and *Arhodomonas aquaeolei*. The *nifH* gene encoding nitrogenase reductase was present in all investigated species of the phototrophic genera *Ectothiorhodospira*, *Halorhodospira* and *Thiorhodospira*, but not of the genus *Ectothiorhodosinus*. Unexpectedly, *nifH* fragments were also obtained for the chemotrophic species *Thioalkalispira microaerophila* and *Alkalilimnicola halodurans*, for which diazotrophic potential has not previously been assumed. The *cbbL*-*, *nifH*- and 16S rRNA gene-based trees were not highly congruent in their branching patterns since, in the 'RubisCO' and 'nitrogenase' trees, representatives of the *Ectothiorhodospiraceae* are divided in a number of broadly distributed clusters and branches. However, the data obtained may be regarded as evidence of the monophyletic origin of the *cbbL* and *nifH* genes in most species within the family *Ectothiorhodospiraceae* and mainly corresponded to the current taxonomic structure of this family. The *cbbL* phylogeny of the chemolithoautotrophic sulfur-oxidizers *Thioalkalivibrio nitratireducens* and *Thioalkalivibrio paradoxus* and the nitrifier *Nitrooccus mobilis* deviated significantly from the 16S-rRNA gene-based phylogeny. These species clustered with one of the duplicated *cbbL* genes of the purple sulfur bacterium *Allochromatium vinosum*, a member of the family *Chromatiaceae*.

INTRODUCTION

The family *Ectothiorhodospiraceae* forms a separate phylogenetic lineage within the *Gammaproteobacteria* according to 16S rRNA gene sequence analysis and mostly includes halophilic and haloalkaliphilic bacteria. Originally, this family was created to accommodate purple sulfur bacteria that deposit elemental sulfur outside the cell. Later, it was divided into the moderately halo(alkali)philic species of the genus *Ectothiorhodospira* and the extremely halo(alkali)-philic species of the genus *Halorhodospira* (Imhoff & Suling, 1996). More recently, two new genera, *Thiorhodospira* (Bryantseva et al., 1999) and *Ectothiorhodosinus* (Gorlenko et al., 2004), made up of moderately haloalkaliphilic purple sulfur bacteria isolated from soda lakes, were added to the family. All these phototrophic bacteria prefer to grow anaerobically in the light using reduced sulfur compounds as electron donors. At the same time, representatives of several genera with aerobic chemotrophic metabolism are closely related to the anaerobic phototrophs of the *Ectothiorhodospiraceae*, namely the heterotrophs *Arhodomonas* (Adkins et al., 1993) and *Alkalispirillum* (Rijkenberg et al., 2001), the facultative autotroph *Alkalilimnicola* (Yakimov et al., 2001; Sorokin et al., 2006; Hoeft et al., 2007), the obligately chemolithoautotrophic nitrifier *Nitrooccus* (Watson & Waterbury, 1971; Teske et al., 1994) and chemolithotrophic sulfur-oxidizing...
bacteria of the genera *Thioalkalivibrio* (Sorokin et al., 2001, 2002a, b; Banciu et al., 2004) and *Thioalkalispira* (Sorokin et al., 2002c). This raises several interesting questions on the evolution of the family.

The modern taxonomy of the *Ectothiorhodospiraceae* follows their phylogenetic relationships based on 16S rRNA gene sequences (Fig. 1). The gene encoding the 16S rRNA is widely used as a universal molecular marker for phylogenetic reconstructions and taxonomy because it has been assumed that intraspecies variation and horizontal transfer of this gene were low. Some housekeeping protein-encoding genes, e.g. *gyrB*, *recA* and *rpoB* (Dauga, 2002; Holmes et al., 2004), have been used as additional molecular markers for phylogenetic studies of different bacterial groups. Phylogenetic reconstructions based on sequence analyses of the 16S rRNA gene and housekeeping protein-encoding genes usually correlate quite well, but sometimes an additional analysis of functional genes is necessary in order to clarify uncertain cases. The functional genes responsible for key metabolic properties can also be used as alternative molecular markers. Sequence analysis of functional genes might help to resolve difficult taxonomic problems as well as to clarify the evolution of corresponding metabolic pathways.

Most representatives of the family *Ectothiorhodospiraceae* are obligate or facultative autotrophs. It has been shown that phototrophic species of *Ectothiorhodospira* and *Halorhodospira* (Imhoff, 2006) and chemolithotrophic species of *Thioalkalivibrio* (Sorokin et al., 2001, 2002a, b; Banciu et al., 2004) and *Thioalkalispira* (Sorokin et al., 2002c) have ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) activity. RubisCO is the key enzyme of the Calvin–Benson–Bassham reductive pentose phosphate cycle. Two different RubisCO forms are known. The large catalytic subunit of form I is encoded by *cbbL*, while the only subunit of form II is encoded by *cbbl*. The *cbbl* gene is in turn divided into ‘green-like’ and ‘red-like’ types. The ‘green-like’ *cbbL* gene has recently been found in the genomes of *Ectothiorhodospira shaposhnikovii* (Spiridonova et al., 2004), *Thioalkalivibrio* species (Tourova et al., 2005) and the *Alkalispirillum–Alkalilimnicola* group (Orenland et al., 2002; Sorokin et al., 2006).

The capacity for nitrogen fixation has been shown for most purple phototrophic bacteria, including some members of the *Ectothiorhodospiraceae* (Moshkovskii et al., 1971; Imhoff, 2005). During the process of biological nitrogen fixation, the enzyme nitrogenase catalyses the ATP-dependent reduction of dinitrogen to ammonia. Nitrogenase consists of two component metalloproteins, the iron (Fe) protein (encoded by *nifH*) and the molybdenum–iron (MoFe) protein (encoded by *nifD* and *nifK*). The Fe protein mediates the coupling of ATP hydrolysis to interprotein electron transfer, and its gene is highly conserved among closely related micro-organisms. This means that *nifH* can be used to study relationships among diazotrophic bacteria. Nevertheless, the occurrence and phylogeny of *nifH* in representatives of the *Ectothiorhodospiraceae* has not yet been investigated. The only exception is *Halorhodospira halophila* strain BN 9629, where the nitrogenase operon has recently been cloned and characterized (Tsuihiji et al., 2006).

The aim of the present study was to extend the evolutionary analysis of the family *Ectothiorhodospiraceae* based on the comparison of the 16S rDNA gene-based phylogeny with phylogenies of genes encoding the key functional enzymes RubisCO (*cbbL*) and nitrogenase (*nifH*).

**METHODS**

**Bacterial strains.** Strains used in this study are listed in Table 1. Type strains of the genera *Alkalispirillum* and *Arhodomonas* were obtained from the DSMZ. *Nitrooccus mobilis* ATCC 25380 was kindly provided by Dr E. Speck (Universität Hamburg, Germany). Biomass of these bacteria was used directly for DNA extraction. Other strains were from the culture collections of the Institute of Microbiology and of the Department of Microbiology of Moscow State University (KM MGU). The cultures were maintained as described previously (Imhoff, 2005; Bryantseva et al., 1999; Sorokin et al., 2001, 2002a, b, c).

**DNA isolation and PCR amplification.** DNA extraction and purification were performed as described previously (Boulygina et al., 1995, 1999). DNA sequences were determined by the Sanger method as described previously (Boulygina et al., 1995, 1999). Positive PCR amplification was indicated by a band of approximately 1400 bp in some species (Sorokin et al., 2001, 2002a, b).

**Fig. 1.** Phylogeny of the family *Ectothiorhodospiraceae* based on analyses of 16S rDNA gene sequences. Strains for which RubisCO gene sequences were determined are underlined. The tree topography and evolutionary distances are given by the neighbour-joining method with Jukes and Cantor distances. Numbers at nodes indicate percentage bootstrap values for the clade of this group in 1000 replications. Only values above 75% were considered significant. Bar, 5% sequence divergence.
et al., 2002). The cbbL and nifH gene fragments were amplified using specially developed and previously tested primers (Spiridonova et al., 2004; Marusina et al., 2001). PCR products were purified from low-melting-point agarose using the Wizard PCR Preps kit (Promega).

**Cloning and sequencing of the PCR fragments.** Purified PCR products were cloned using the pGEM-T vector system (Promega). Plasmid DNA was extracted and purified using the Wizard MiniPrep kit (Promega). Clones containing appropriately sized inserts were sequenced from universal M13 forward and reverse primers (Sambrook et al., 1989). Sequencing was performed with an ABI 3730 sequencer using the Big Dye Terminator v. 3.1 sequencing reaction kit (Applied Biosystems).

**Phylogenetic analysis.** Preliminary analysis of the new sequences was performed via the NCBI BLAST server (http://www.ncbi.nlm.nih.gov/BLAST/). Nucleotide and inferred amino acid sequences were aligned with sequences from GenBank using CLUSTAL W (Thompson et al., 1994).

Genetic distances were calculated using Kimura’s two-parameter method (Kimura, 1980). To obtain synonymous and non-synonymous distances, a method of Nei & Gojobori (1986) was applied to the various sequences of protein-encoding genes using WET software (J. Dopazo; http://www.tdi.es). Phylogenetic trees were reconstructed using four different algorithms: neighbour-joining (Saitou & Nei, 1987) in the TREECONW program package (Van de Peer & De Wachter, 1994) and maximum-parsimony (Fitch, 1971), distance matrix (Fitch & Margoliash, 1967) and maximum-likelihood (Felsenstein, 1981) using PHYLIP 3.5c software (Felsenstein, 1993). Relative synonymous codon usage (RSCU) values of the cbbL and nifH genes were calculated using CodonW software (J. Peden; http://codonw.sourceforge.net). To investigate the major codon usage trends in different species, CodonW was used to carry out a correspondent analysis. Each gene produced a point in the codon space, the positions of which suggested a possible codon usage bias.

**RESULTS**

**Detection and phylogenetic analysis of RubisCO genes**

Primers specific to the genes encoding ‘red-like’ form I and form II of RubisCO did not amplify corresponding gene fragments in any of the investigated strains listed in Table 1. On the other hand, using the specific primer set for the ‘green-like’ cbbL gene, PCR products of about 750 bp were obtained from the DNAs of all strains, except the type strains of *Alkalispirillum mobile* and *Arhodomonas*.

### Table 1. Comparison between the G+C contents (mol%) of the cbbL and nifH genes and the full genome in the family Ectothiorhodospiraceae

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genome</th>
<th>cbbL</th>
<th>cbbL (3rd position)</th>
<th>nifH</th>
<th>nifH (3rd position)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ectothiorhodospira mobilis</em></td>
<td>Triöper 8112T = DSM 237T</td>
<td>63.3</td>
<td>65.8</td>
<td>88.8</td>
<td>65.0</td>
</tr>
<tr>
<td><em>Ectothiorhodospira shaposhnikovi</em></td>
<td>Kondratieva N1T = DSM 243T</td>
<td>62.3</td>
<td>64.0</td>
<td>83.5</td>
<td>65.0</td>
</tr>
<tr>
<td><em>Ectothiorhodospira vacuolata</em></td>
<td>BN 9512T = DSM 2111T</td>
<td>63.2</td>
<td>64.7</td>
<td>85.6</td>
<td>64.6</td>
</tr>
<tr>
<td><em>Ectothiorhodospira halalkaliphila</em></td>
<td>BN 9903T = ATCC 51935T</td>
<td>63.5</td>
<td>66.3</td>
<td>88.4</td>
<td>61.6</td>
</tr>
<tr>
<td><em>Ectothiorhodospira marismortui</em></td>
<td>EG-1T = DSM 4180T</td>
<td>65.0</td>
<td>68.6</td>
<td>95.6</td>
<td>ND</td>
</tr>
<tr>
<td><em>Halorhodospira abdelmaleki</em></td>
<td>BN 9840T = DSM 2110T</td>
<td>63.8</td>
<td>64.7</td>
<td>85.0</td>
<td>64.3</td>
</tr>
<tr>
<td><em>Halorhodospira halochloris</em></td>
<td>BN 9850T = DSM 1059T</td>
<td>52.9</td>
<td>62.7</td>
<td>77.5</td>
<td>60.9</td>
</tr>
<tr>
<td><em>Halorhodospira halophilia</em></td>
<td>BN 9632T = DSM 244T</td>
<td>68.4</td>
<td>67.4</td>
<td>93.3</td>
<td>68.0</td>
</tr>
<tr>
<td><em>Thiorhodospira sibirica</em></td>
<td>Gorlenko A12T = ATCC 700588T</td>
<td>56.7</td>
<td>58.9</td>
<td>70.5</td>
<td>56.2</td>
</tr>
<tr>
<td><em>Ectothiorhodospirinaeus</em></td>
<td>Gorlenko M9T = DSM 15479T</td>
<td>57.5</td>
<td>56.2</td>
<td>57.7</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalislira microaerophila</em></td>
<td>ALEN 1T = DSM 14784T</td>
<td>58.9</td>
<td>56.2</td>
<td>56.9</td>
<td>60.3</td>
</tr>
<tr>
<td><em>Thioalkalivibrio versutus</em></td>
<td>AL2T = DSM 13738T</td>
<td>63.7</td>
<td>65.0</td>
<td>82.0</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio denitrificans</em></td>
<td>ALJD T = DSM 13742T</td>
<td>62.9</td>
<td>67.0</td>
<td>89.4</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio thiocyanoxidans</em></td>
<td>ARh 2T = DSM 13532T</td>
<td>66.2</td>
<td>65.4</td>
<td>82.9</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio halophilia</em></td>
<td>HL 17T = DSM 15791T</td>
<td>65.1</td>
<td>65.8</td>
<td>85.0</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio jannaschii</em></td>
<td>ALM 2T = DSM 14478T</td>
<td>63.7</td>
<td>65.1</td>
<td>83.7</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio nitratireducens</em></td>
<td>ALEN 2T = DSM 14787T</td>
<td>64.8</td>
<td>66.5</td>
<td>85.5</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio nitratii</em></td>
<td>ALJ 12T = DSM 13741T</td>
<td>62.1</td>
<td>65.4</td>
<td>82.5</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio paradoxus</em></td>
<td>ARh 1T = DSM 13531T</td>
<td>65.6</td>
<td>66.1</td>
<td>84.6</td>
<td>NA</td>
</tr>
<tr>
<td><em>Thioalkalivibrio thiocyanoxidans</em></td>
<td>ArhD 1T = DSM 16954T</td>
<td>63.7</td>
<td>67.9</td>
<td>91.2</td>
<td>NA</td>
</tr>
<tr>
<td><em>Alkalimirillum mobile</em></td>
<td>DSM 12769T</td>
<td>66.2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><em>Alkalimirillum mobile</em> AGDZ</td>
<td>66.6</td>
<td>70.7</td>
<td>96.0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><em>Alkalimirillum mobile Z-0008</em></td>
<td>66.5</td>
<td>70.7</td>
<td>96.5</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><em>Alkalimirillum mobile ALPS2</em></td>
<td>66.6</td>
<td>71.0</td>
<td>96.9</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><em>Alkalimimnica halodurans</em></td>
<td>34AlcT = DSM 13718T</td>
<td>65.6</td>
<td>70.3</td>
<td>95.6</td>
<td>66.3</td>
</tr>
<tr>
<td><em>Alkalimimnica sp. AHN 1</em></td>
<td>65.0</td>
<td>69.7</td>
<td>93.8</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><em>Nitrooccus mobilis</em> ATCC 25380T</td>
<td>61.2</td>
<td>63.9</td>
<td>79.1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><em>Arhodomonas aquaeolei</em> DSM 8974T</td>
<td>67.0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

ND, Not determined; NA, not applicable.
Clones prepared from the cbbL PCR fragments yielded a single sequence-type for all species. Results of BLAST analysis revealed a high level of similarity between the newly determined nucleotide sequences and the other cbbL sequences available in GenBank, confirming their affiliation to the same family of genes.

The nucleotide sequences obtained for RubisCO gene fragments as well as the deduced amino acid sequences of the corresponding proteins were aligned with analogous sequences of the 'green-like' cbbL from GenBank. Positions with gaps and ambiguous sequences were removed and the remaining 720 nucleotide and 240 amino acid positions were used for further phylogenetic analysis. The topologies of the phylogenetic trees constructed on the basis of these alignments were similar for all the methods used, the neighbour-joining (Fig. 2), maximum-parsimony, distance-matrix and maximum-likelihood (data not shown) methods.

Similar to results published previously (Watson & Tabita, 1997), the topologies obtained did not correspond to those of the ‘ribosomal’ tree. According to 16S rRNA gene sequence analysis, all members of the family Ectothiorhodospiraceae form a monophyletic clade within the Gammaproteobacteria, consisting of several phylogenetic clusters and branches corresponding to different genera (Fig. 1). However, the cbbL-based trees suggest that the Ectothiorhodospiraceae is divided into independent clusters and branches with positions in the trees dependent on the set of chosen reference sequences (Fig. 2). These intrafamily divisions correlated only partially with the 16S rRNA gene-based tree topology. Different species of the genera Halorhodospira and Alkalilimnicola formed monophyletic clusters with high bootstrap values (100 and 98–100 %, respectively, based on nucleotides and amino acids), and the monotypic genus Thioalkalivibrio formed a separate branch unrelated to other members of the family in both trees. However, the genus Ectothiorhodospira realistically formed a single cluster (with 91 % bootstrap value) in the ‘RubisCO’ tree based on nucleotides only, whereas in the tree based on amino acids this cluster was unstable (31 % bootstrap value), depending on the set of chosen reference sequences. Moreover, in the ‘RubisCO’ trees, the monotypic genus Ectothiorhodosinus formed a separate branch unrelated to Ectothiorhodospira species.

The most dramatic discrepancies between the topologies of the ‘ribosomal’ and ‘RubisCO’ trees were observed in the case of the genus Thioalkalivibrio. According to the RubisCO gene analysis, Thioalkalivibrio is not a monophyletic genus (Tourova et al., 2005). The Thioalkalivibrio species formed three independent clusters in the nucleotide-based ‘RubisCO’ tree (with 74–100 % bootstrap values) and two clusters (34–100 % bootstrap values) and one branch in the amino acid-based tree. This division correlated only partially with the inner structure of the single ‘ribosomal’ Thioalkalivibrio cluster. Only the cluster combining Thioalkalivibrio denitrificans and Thioalkalivibrio thiocyanodenitrificans (the deepest Thioalkalivibrio subcluster according to the ‘ribosomal’ tree) demonstrated a distant relationship to Ectothiorhodospira in the nucleotide-based ‘RubisCO’ tree (75 % bootstrap value). At the same time, the phylogenetic position was strongly supported in both ‘RubisCO’ trees only for the cluster Thioalkalivibrio nitratireducens / Thioalkalivibrio paradoxus. This cluster grouped with Nitrococcus mobilis with high bootstrap values (98 and 100 %, respectively, based on nucleotides and amino acids) and with one of the duplicated cbbL genes of Allochromatium vinosum (with 100 and 70 % bootstrap values, respectively, for nucleotide and amino-acid-based trees). The position of Thiorthodospira sibirica in both RubisCO trees was unexpected, because it grouped together with some representatives of the Alphaproteobacteria with strong bootstrap support (respectively 98 and 97 % based on nucleotides and amino acids).

Detection and phylogenetic analysis of nitrogenase reductase genes

PCR products of the expected size (about 450 bp) were obtained with the nifH-specific primers using the DNAs of all investigated species of the phototrophic genera Ectothiorhodospira, Halorhodospira and Thiorthodospira, but not of the genus Ectothiorhodosinus. Unexpectedly, nifH fragments were also obtained for the chemotrophic species Thioalkalispira microaerophila and Alkalilimnicola halodurans, for which diazotrophic potential was not suspected previously. Preliminary screening in the GenBank database demonstrated that all newly determined nucleotide sequences belong to the nifH gene family. Moreover, the nifH gene fragment obtained from the type strain of Hlr. halophila was almost identical to the analogous fragment of the previously sequenced nifH gene from Hlr. halophila BN 9629 (GenBank accession no. AB189641).

The nucleotide sequences of nifH gene fragments as well as the deduced amino acid sequences of the corresponding proteins were aligned, positions with gaps and ambiguous sequences were removed and the remaining 444 nucleotide and 148 amino acid positions were used for further phylogenetic analysis. The topologies of the phylogenetic trees constructed on the basis of these alignments were similar for all methods used: the neighbour-joining (Fig. 3), maximum-parsimony, distance-matrix and maximum-likelihood (data not shown) methods.

The topologies of the nifH-based trees correlated only partially with the ‘ribosomal’ tree. Similar to the cbbL-based trees, species of the family Ectothiorhodospiraceae did not form a monophyletic clade. Moreover, in the case of nifH, the Ectothiorhodospira species (as well as Halorhodospira representatives) were clustered together only in the nucleotide-based tree (with 55 and 96 % bootstrap values, respectively). On the other hand, the nifH-based phylogeny of the monotypic genus Thiorthodospira demonstrated a clear relationship of this bacterium to Ectothiorhodospira mobilis, Ectothiorhodospira shaposhnikovii and Ectothiorhodospira vacuolata (80 % bootstrap value for the
Fig. 2. Phylogenetic positions of species of the Ectothiorhodospiraceae in cbbL molecular trees based on analysis of nucleotide sequences (a) and translated amino acid sequences (b). Sequences of species of the Ectothiorhodospiraceae determined in this study are marked in bold. Underlining marks sequences belonging to species of the family Ectothiorhodospiraceae. Tree topography and evolutionary distances are given by the neighbour-joining method with Jukes and Cantor (for nucleotides) and Poisson (for amino acids) corrections. Numbers at nodes indicate percentage bootstrap values for the clade of this group in 1000 replications. Only values above 75% were considered significant. Bars, 10% (a) or 5% (b) sequence divergence.
amino-acid-based tree) and was similar to the 16S rRNA gene-based phylogeny, although it differed from the cbbL-based one.

The nifH phylogeny of *Thioalkalispira microaerophila* and *Alkalimimicola halodurans* was of particular interest. In both ‘nitrogenase’ trees, they formed two separate branches with an uncertain branching point position, and were slightly related only to *Azoarcus* and *Azovibrio*, representing the Betaproteobacteria (bootstrap values did not exceed 56 %). Thus, their nifH genes do not have a clear relationship to the nifH of the other members of the Ectothiorhodospiraceae.

**Comparison of genetic distances among the cbbL, nifH and 16S rRNA genes**

Nucleotide substitutions within protein-encoding regions are divided into two classes: synonymous (silent), which are largely invisible to natural selection, and non-synonymous (resulting in amino acid replacement), which may be under strong selective pressure. Synonymous distances in the cbbL and nifH genes were examined for all possible combinations of the investigated species of the Ectothiorhodospiraceae. A significant correlation between the synonymous distances in the cbbL genes and those in the nifH genes was observed, with a correlation coefficient \((r)\) of 0.91. This result is in agreement with the assumption that the synonymous substitution rate is constant for many chromosomal genes in many organisms and that it can serve as a suitable molecular indicator of their evolution (Lawrence et al., 1991).

The total numbers of substitutions used for the calculation of genetic distances for cbbL and nifH allowed estimation of the ranges of genomic variation (Table 2). The intrafamily genetic distances between the cbbL, nifH and
16S rRNA gene sequences were up to 0.414, 0.295 and 0.134, respectively. This indicated a variability order \( cbbL > nifH > 16S \) rRNA gene.

The highest level of sequence divergence (0.282–0.414) was detected for the \( cbbL \) gene of *Trs. sibirica*. This was due to an increase in the number of non-synonymous substitutions. Interestingly, its *nifH* sequence was not so divergent from the *nifH* genes of other members of the *Ectothiorhodospiraceae* (Table 2).

On the basis of DNA–DNA hybridization analysis, Ventura *et al.* (2000) proposed to consider *Ect. vacuolata* and *Ectothiorhodospira marismortui* as junior synonyms of *Ect. shaposhnikovii* and *Ect. mobilis*, respectively. However, the similarities of *cbbL* gene sequences of the pairs *Ect. vacuolata*–*Ect. shaposhnikovii* and *Ect. marismortui*–*Ect. mobilis* were at the same level as for the pair *Ect. shaposhnikovii*–*Ect. mobilis*, and much higher than for different strains of *Alkalipirillum mobile* (Table 2). On the other hand, although the *nifH* gene sequences of *Ect. vacuolata* and *Ect. shaposhnikovii* were found to be very similar (0.010), the same (high) level of similarity was shown for the sequences of *Ect. shaposhnikovii* and *Ect. mobilis* (0.007). These data are in good agreement with the current taxonomy of the genus *Ectothiorhodospira*, where *Ect. vacuolata*, *Ect. shaposhnikovii*, *Ect. marismortui* and *Ect. mobilis* are considered as four different species, and they disagree with the reclassification proposed by Ventura *et al.* (2000).

**Table 2.** Ranges of genetic (Kimura), synonymous (syn) and non-synonymous (non-syn) distances calculated for 16S rRNA, *cbbL* and *nifH* genes of the family *Ectothiorhodospiraceae*.

<table>
<thead>
<tr>
<th>Range of variation</th>
<th>16S rRNA</th>
<th><em>cbbL</em></th>
<th><em>nifH</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kimura</td>
<td>syn</td>
<td>non-syn</td>
</tr>
<tr>
<td>Within:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ectothiorhodospiraceae</em></td>
<td>0.002–0.134</td>
<td>0.000–0.414</td>
<td>0.011–0.764</td>
</tr>
<tr>
<td><em>Ectothiorhodospira</em></td>
<td>0.009–0.042</td>
<td>0.063–0.133</td>
<td>0.147–0.303</td>
</tr>
<tr>
<td><em>Halorhodospira</em></td>
<td>0.015–0.068</td>
<td>0.143–0.150</td>
<td>0.394–0.499</td>
</tr>
<tr>
<td><em>Thioalkalivibrio</em></td>
<td>0.015–0.070</td>
<td>0.032–0.228</td>
<td>0.089–0.462</td>
</tr>
<tr>
<td><em>Alkalipirillum mobile</em></td>
<td>0.002–0.012</td>
<td>0.003–0.007</td>
<td>0.011–0.017</td>
</tr>
<tr>
<td>Between:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ect. shaposhnikovii</em> and <em>Ect. vacuolata</em></td>
<td>0.011</td>
<td>0.100</td>
<td>0.220</td>
</tr>
<tr>
<td><em>Ect. mobilis</em> and <em>Ect. marismortui</em></td>
<td>0.040</td>
<td>0.095</td>
<td>0.238</td>
</tr>
<tr>
<td><em>Ect. shaposhnikovii</em> and <em>Ect. mobilis</em></td>
<td>0.038</td>
<td>0.086</td>
<td>0.215</td>
</tr>
<tr>
<td><em>Ectothiorhodospira</em> and others</td>
<td>0.033–0.131</td>
<td>0.116–0.318</td>
<td>0.247–0.675</td>
</tr>
<tr>
<td><em>Halorhodospira</em> and others</td>
<td>0.052–0.134</td>
<td>0.134–0.335</td>
<td>0.284–0.737</td>
</tr>
<tr>
<td><em>Alkalipirillum</em> and <em>Alkalimnicola</em></td>
<td>0.011–0.019</td>
<td>0.050–0.074</td>
<td>0.176–0.238</td>
</tr>
<tr>
<td><em>Alkalipirillum–Alkalimnicola</em> and others</td>
<td>0.052–0.105</td>
<td>0.178–0.307</td>
<td>0.246–0.702</td>
</tr>
<tr>
<td><em>Thiorhodospira</em> and others</td>
<td>0.052–0.126</td>
<td>0.282–0.414</td>
<td>0.483–0.750</td>
</tr>
<tr>
<td><em>Thioalkalivibrio</em> and others</td>
<td>0.075–0.133</td>
<td>0.199–0.342</td>
<td>0.525–0.737</td>
</tr>
<tr>
<td><em>Ectothiorhodosinus</em> and others</td>
<td>0.033–0.123</td>
<td>0.213–0.342</td>
<td>0.610–0.764</td>
</tr>
<tr>
<td><em>Nitrooccus</em> and others</td>
<td>0.057–0.117</td>
<td>0.174–0.407</td>
<td>0.433–0.724</td>
</tr>
</tbody>
</table>

ND, Not determined; NA, not applicable.

**Nucleotide composition and codon usage of the \( cbbL \) and \( nifH \) genes**

Genes in closely related species tend to be rather similar in their G+C content as well as in synonymous codon usage, in contrast to genes acquired by horizontal transfer, which often have atypical G+C content and codon usage bias (Medigue *et al.*, 1991). Therefore, it was interesting to compare the G+C content and codon usage of the RubisCO and nitrogenase reductase genes within the family *Ectothiorhodospiraceae* to detect the possible role of gene transfer in their evolution.

The total G+C content of all analysed \( cbbL \) and \( nifH \) gene fragments was close to the genomic G+C content for each species of the family (56.2–70.7 and 56.2–66.3 against 52.9–68.4 mol%, respectively; Table 1). The G+C \(_S\) content (in the third position of codons) of these genes (57.7–96.9 and 75.8–93.7 mol%) was higher than their total G+C content and the overall genomic G+C content; it is typical of G+C-biased micro-organisms that they preferentially use G or C in the third position of the codons (Ohtaka & Ishikawa, 1993).

The broadest intragenus range of G+C content variation in \( cbbL \) and \( nifH \) genes (62.7–68.6 and 60.9–68.0 mol%, respectively) was found for the *Halorhodospira* species, consistent with the variation of their genomic G+C content (52.9–68.4 mol%). Among the other representatives of the *Ectothiorhodospiraceae*, the lowest G+C contents of the *nifH* and \( cbbL \) genes (less than
60.3 mol%) were found for the monotypic genera *Ectothiorhodosinus*, *Thioalkalivibrio* and *Thiorhodospira*, and were also comparable to their overall genomic G+C content. Interestingly, among the species of the *Alkalilimnicola*–*Alkalilimnicola* group, the G+C content of *cbbL* differed significantly from the genomic G+C content and was the highest in the *Ectothiorhodosporaceae* (up to 71.0 mol%). At the same time, the G+C content and was the highest in the *cbbL* gene of *Alkalilimnicola halodurans* (which is the only representative of the *Alkalilimnicola*–*Alkalilimnicola* group that has a *nifH* gene) was similar to the overall genomic G+C content.

Codon usage analysis for the *cbbL* and *nifH* genes was carried out on the RSCU data. Correspondence analysis of the results (Fig. 4) identified the major trends in codon usage: the horizontal axis is associated with G+C content, whereas the vertical axis is correlated with the frequencies of codons ending in C or U versus A or G (Fennoy & Bailey-Serres, 1993). Codon usage analysis of *cbbL* genes is compatible with the formation of a common group on the plot for most representatives of the *Ectothiorhodospiraceae*: species of *Ectothiorhodospora*, *Halorhodosporia*, *Thioalkalivibrio* and the *Alkalilimnicola*–*Alkalilimnicola* group and *Nitrococcus mobilis* (Fig. 4). However, the codon usage of *Halorhodospora abdelmalekii* was similar to that of alphaproteobacteria. The *cbbL* codon usage bias of *Trs. sibirica*, *Thioalkalivibrio microaerophila* and *Ectothiorhodosporaceae mongolicus* was consistent with their G+C content bias.

The intrafamily codon usage pattern of the *nifH* genes on the RSCU correspondence analysis plot was similar to the pattern of the *cbbL* genes (Fig. 4). It appears that the codon usage bias in both *cbbL* and *nifH* suggests their origin from an intrafamily genome divergence rather than from lateral gene transfer.

**DISCUSSION**

The use of functional genes encoding key metabolic enzymes as molecular markers is becoming common practice in phylogenetic studies. The RubisCO phylogeny, in general, differs significantly from the traditional 16S rRNA gene-based phylogeny of autotrophic bacteria (Watson & Tabita, 1997). At the same time, the topology of the ‘nitrogenase’ tree exhibits significant correlation with reconstructions based on ‘ribosomal’ gene analysis (Achouak et al., 1999; Tourova et al., 2006). This divergence can be useful to resolve particular problems of the evolution of autotrophic and diazotrophic bacteria.

![Fig. 4. Correspondence analysis of RSCU in species of the Ectothiorhodospiraceae for cbbL genes of the studied group (●), cbbL genes of other micro-organisms (○), nifH genes of the studied group (▲) and nifH genes of other micro-organisms (△). Abbreviations: Eno, Ect. mobilis; Esh, Ect. shaposhnikovii; Eva, Ect. vacuolata; Eha, Ect. haloalkaliphila; Ema, Ect. marismortui; Hab, Hlr. abdelmalekii; Hhc, Hlr. halochloris; Hha, Hlr. halophila; Trs, Trs. sibirica; Esm, Ectothiorhodosporaceae; Tmi, Thioalkalivibrio microaerophila; Tve, Thioalkalivibrio versutus; Tde, Thioalkalivibrio denitrificans; Thh, Thioalkalivibrio thiocyanoxidans; Tha, Thialkalivibrio halophilus; Tja, Thialkalivibrio jannaschii; Thn, Thialkalivibrio nitratireducens; Tns, Thialkalivibrio nitratis; Tpa, Thioalkalivibrio paradoxus; Ttc, Thialkalivibrio thiocyanodenitrificans; Amo1, Alkalilimnicola sp. AGDZ; Amo2, Alkalilimnicola sp. Z-0008; Amo3, Alkalilimnicola sp. ALPs2; Aha, Alkalilimnicola halodurans; Als, Alkalilimnicola sp. AHN 1; Nmo, Nitrococcus mobilis; AviA, Alc. vinosum rbcA; AviL, Alc. vinosum rbcL; Rca, Rhodobacter capsulatus; Rvs, Rhodovulum sulfidophilum; Tcl, Thioclava pacifica. Genes are plotted at their co-ordinates on the two axes produced by the analysis; the axes are described in Results.](image-url)
The cbbL-, nifH- and 16S rRNA gene-based trees reconstructed in this study for the species of the Ectothiorhodospiraceae were not highly congruent in their branching patterns. While all species formed a monophyletic clade in the 'ribosomal' tree, in the 'RubisCO' and 'nitrogenase' trees, this clade disintegrated into a number of broadly distributed clusters and branches. However, most of the differences between the trees were in areas of low bootstrap values. The low resolution of the deep branches in the cbbL- and nifH-based trees may be due to accelerated rates of sequence divergence or poor representation of the taxa in databases in comparison with 16S rRNA gene sequences. Therefore, considering the similar codon usage in the cbbL and nifH genes, these phylogenetic data may be regarded as evidence of the monophyletic origin of most cbbL and nifH genes within the family Ectothiorhodospiraceae.

Nevertheless, in some cases, the 'RubisCO'-based trees showed that relationships inside the Ectothiorhodospiraceae were inconsistent both with 'ribosomal' phylogeny and phenotypic properties. This may be due to either inaccuracy in the cbbL trees (phylogenetic construction bias) or occurrence of lateral gene transfer. The example is the phylogenetic position of Trs. sibirica: in both 'RubisCO' trees, it rooted with some alphaproteobacteria while, in the 'ribosomal' tree and in the 'nitrogenase' amino-acid-based tree, it clustered together with the species of the Ectothiorhodospiraceae. However, the G+C content of the cbbL gene of Trs. sibirica (58.9 mol%) was comparable to that of the nifH gene (56.2 mol%) and the total genome (56.7 mol%), but was significantly lower than in the available alphaproteobacterial gene sequences (64.3–65.8 mol%). The codon usage patterns of nifH and cbbL of Trs. sibirica were also similar and differed from the alphaproteobacterial pattern (Fig. 4). Therefore, the unusual rooting of Trs. sibirica on the 'RubisCO' tree might originate from a higher rate of non-synonymous nucleotide substitutions rather than from lateral gene transfer.

The cbbL phylogeny of the cluster including the chemolithoautotrophic sulfur-oxidizers Thioalkalivibrio nitratireducens and Thioalkalivibrio paradoxus and the nitrifier Nitrococcus mobilis is of particular interest. Phylogenetic analysis of both nucleotide and amino acid sequences showed that the RubisCO genes in this group have a common origin different from the origin of the analogous genes in other species of the family. These three species clustered with strong bootstrap support with one of the duplicated cbbL genes of the purple sulfur bacterium Alc. vinosum, a member of the family Chromatiaceae. One of the possible evolutionary mechanisms that could have taken place in this case is lateral gene transfer, which has presumably played a significant role in the evolution of the genes belonging to the RubisCO family (Delwiche & Palmer, 1996; Watson & Tabita, 1997). Similarities in codon usage (Fig. 4) and G+C content between the duplicated cbbL genes of Alc. vinosum (65.1–66.3 mol%) and genes of Thioalkalivibrio and Nitrococcus (63.9–66.5 mol%) do not contradict this suggestion. Interestingly, both of these Thioalkalivibrio species are morphologically similar to Allochromatium (large coccoid rods with sulfur inclusions) but strikingly different from the other Thioalkalivibrio species and members of the Ectothiorhodospiraceae in general (vibrio/spirilla that deposit elemental sulfur outside the cell). It is noteworthy that the formation of intracellular sulfur globules is a characteristic feature of members of the Chromatiaceae and Allochromatium in particular. Moreover, biochemical and recent genetic studies have demonstrated that a reverse dissimilatory sulfite reductase complex encoded by a large gene cluster is responsible for further oxidation of intracellular sulfur to sulfate (Dahl et al., 2005). Preliminary tests with primers specific for the Alc. vinosum reverse dsr gene complex indicated the presence of some of the genes in Thioalkalivibrio nitratireducens and Thioalkalivibrio paradoxus (C. Dahl, personal communication). Cells of Nitrococcus mobilis also have a spherical shape. The tubular membranes of Nitrococcus mobilis are quite similar to the internal photosynthetic membrane system of Thiooccus pfennigii, a member of the Chromatiaceae (Imhoff, 2005). These observations may indicate that there was an exchange of important genetic information between these different types of autotrophic bacteria. This case is just one of many examples of inconsistency between the taxonomic position determined on the basis of 16S rRNA gene sequences and other essential characteristics of the organism in question. Taking into consideration the striking structural similarity between species of the Chromatiaceae and these three representatives of the Ectothiorhodospiraceae (Thioalkalivibrio nitratireducens, Thioalkalivibrio paradoxus and Nitrococcus mobilis), it might be speculated that gene transfer between these phylogenetically distant (based on 16S rRNA gene analysis) organisms might have involved not only individual genes (i.e. cbbL) but also genetic blocks. The alternative and simpler assumption may be lateral transfer of the 16S rRNA gene. Although it is generally accepted that this is a rare event, such a possibility cannot be excluded (Tourova, 2003).

The possibility of chemolithoheterotrophic aerobic growth has been shown for many representatives of the purple sulfur bacteria, from both the Chromatiaceae and the Ectothiorhodospiraceae, and some members of the Chromatiaceae can grow chemolithoautotrophically in the presence of oxygen, as colourless sulfur bacteria do (Kondratieva et al., 1976; Kämpf & Pfennig, 1980). In the case of Thioalkalivibrio species, which are currently classified as members of the Ectothiorhodospiraceae, it might be speculated that these alkaliophilic, aerobic, sulfur-oxidizing, chemolithoautotrophic bacteria represent direct aerobic descendants of the purple sulfur bacteria that have lost the genes responsible for photosynthesis. The results of the cbbL gene analysis did not contradict this suggestion as a whole, but they demonstrated that the putative
phototrophic ancestors might be different for some groups of *Thioalkalivibrio* species.

The genera *Alkalisspirillum* and *Alkalilimnicola* were originally described as non-phototrophic, aerobic, heterotrophic relatives of the *Ectothiorhodospiridae*–*Halorhodospiridae* group. However, it has been shown recently that some novel strains and the type strain of *Alkalilimnicola halodurans* are capable of lithoautotrophy and have *cbbL* genes (Oremland et al., 2002; Sorokin et al., 2006). The *cbbL*-based phylogenetic trees and codon usage analysis confirmed the relatedness of the *Alkalisspirillum–Alkalilimnicola* group to the *Ectothiorhodospiraceae*. However, in spite of the high DNA–DNA relatedness and the 16S rRNA gene and *cbbL* sequence similarity, they are significantly different in details of their autotrophic metabolism (Oremland et al., 2002; Sorokin et al., 2006). For example, hydrogen-based autotrophy was found only in two strains and could be lost easily during cultivation (Sorokin et al., 2006). This might indicate the location of the RubisCO genes on a plasmid (which may be lost during heterotrophic growth). Such a loss may be a reason for the absence (temporary or constant) of the *cbbL* genes in the genome of the type strain of *Alkalisspirillum mobile*.

The absence of RubisCO genes in *Arhodomonas aquaeolei* was in accordance with the original description of this bacterium as an obligate heterotroph (Adkins et al., 1993).

Nitrogen fixation is considered to be a characteristic property of purple sulfur bacteria, including the *Ectothiorhodospiraceae* (Moshkovskii et al., 1971; Imhoff, 2005). Moreover, nitrogenase-mediated hydrogen production was shown for representatives of this group (Chadwick & Irgens, 1991; Tsuihiji et al., 2006). Therefore, it is not surprising that *nifH* genes were detected in all phototrophic members of the *Ectothiorhodospiraceae* with the sole exception of *Ers. mongolicus*. This bacterium was isolated from a habitat enriched in organic compounds and had only a weak capacity for photoautotrophic growth (Gorlenko et al., 2004); growth under such conditions might also be accompanied by a loss of the nitrogenase genes. Since many members of the *Ectothiorhodospiraceae* do not have this gene and the topologies of the *nifH*-based trees correlate only partially with the topology of the 16S rRNA gene tree, *nifH* does not seem to be suitable for basic phylogenetic assessments. However, its analysis may help to understand the relationship of these bacteria, especially over small phylogenetic distances.

The occurrence of *nifH* genes in the aerobic chemotrophic *Alkalilimnicola halodurans* and *Thioalkalispira microaerophilica* is more difficult to interpret. Although analysis of these *nifH* genes confirmed their relatedness to *nifH* of other members of the *Ectothiorhodospiraceae*, measurement of nitrogenase activity in *Alkalilimnicola halodurans* and *Thioalkalispira microaerophilica* is necessary in order to confirm whether these genes are functional. Several examples of ancient altered genes (‘pseudogenes’), which can become non-functional but may still retain sufficient similarity to functional genes, are well documented (Ochman & Davals, 2006).

The conclusion that most of the *cbbL* and *nifH* genes within the family *Ectothiorhodospiraceae* have a monophyletic origin allows them to be used them for analyses of microbial communities in situ. The substantial limitation of *in situ* analyses based on functional genes is the poor representation of many taxa (including the *Ectothiorhodospiraceae*) in databases. This limits identification of the obtained sequences or leads to their misidentification. The database obtained in this study could provide good support for *in situ* studies (for example, in haloalkaline lakes, where members of the *Ectothiorhodospiraceae* thrive).

In general, molecular phylogenies based on a single gene may be misleading, because of the complexity of the evolutionary process. Thus, data from several genes encoding different cellular functions are more suitable for realistic phylogenetic reconstructions.

**ACKNOWLEDGEMENTS**

This work was supported by the Russian Foundation for Basic Research (grant 05-04-48064), by an NWO-RBR grant (047.011.2004.010) and by the Program on Molecular and Cell Biology of the RAS.

**REFERENCES**


