Occurrence, phylogeny and evolution of ribulose-1,5-bisphosphate carboxylase/oxygenase genes in obligately chemolithoautotrophic sulfur-oxidizing bacteria of the genera *Thiomicrospira* and *Thioalkalimicrobium*

Tatjana P. Tourova,1 Elizaveta M. Spiridonova,2 Ivan A. Berg,2 Boris B. Kuznetsov3 and Dimitry Yu. Sorokin1,4

1Institute of Microbiology, Russian Academy of Sciences, p-t 60-letiya Oktyabrya, 7/2, Moscow, Russia
2Department of Microbiology, Moscow State University, Moscow, Russia
3Centre Bioengineering, Russian Academy of Sciences, Moscow, Russia
4Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

The occurrence of the different genes encoding ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), the key enzyme of the Calvin–Benson–Bassham cycle of autotrophic CO₂ fixation, was investigated in the members of the genus *Thiomicrospira* and the relative genus *Thioalkalimicrobium*, all obligately chemolithoautotrophic sulfur-oxidizing Gammaproteobacteria. The *cbbL* gene encoding the ‘green-like’ form I RubisCO large subunit was found in all analysed species, while the *cbbM* gene encoding form II RubisCO was present only in *Thiomicrospira* species. Furthermore, species belonging to the *Thiomicrospira crunogena* 16S rRNA-based phylogenetic cluster also possessed two genes of green-like form I RubisCO, *cbbL-1* and *cbbL-2*. Both 16S-rRNA- and *cbbL*-based phylogenies of the *Thiomicrospira–Thioalkalimicrobium–Hydrogenovibrio* group were congruent, thus supporting its monophyletic origin. On the other hand, it also supports the necessity for taxonomy reorganization of this group into a new family with four genera.

INTRODUCTION

The *Thiomicrospira* genus includes obligately chemolithoautotrophic sulfur-oxidizing bacteria (SOB) isolated from saline, mostly marine, habitats and belonging to the Gammaproteobacteria (Robertson & Kuenen 1999; Brinkhoff et al., 1999c), except *Thiomicrospira dentrificans* which belongs to the Epsilonproteobacteria and needs to be reclassified. Currently, ten species are recognized in this genus. It also includes some not yet validly described strains. The two major features discriminating this genus from other SOB are salt tolerance and high growth rates. In fact, *Thiomicrospira crunogena* is known as the fastest-growing mesophilic chemolithoautotroph (Jannasch et al., 1985). According to our experience with SOB in salt lakes, halophilic *Thiomicrospira* sp. outcompeted *Halothiobacillus* spp. with the same salinity profile at micro-oxic conditions and with sulfide instead of thiosulfate as a substrate (unpublished data).

Recent phylogenetic 16S rRNA-based analysis and descriptions of new SOB related to *Thiomicrospira* (Brinkhoff et al., 1999a, b; Takai et al., 2004) have demonstrated (i) that the *Thiomicrospira* genus is heterogeneous, containing at least two different groups, clustering either with *Thiomicrospira pelophila* (Kuenen & Veldkamp, 1972; Wood & Kelly, 1989, 1993) or with *Thiomicrospira crunogena*, and (ii) that the genus itself is a member of a bigger group of closely related SOB. In particular, the halooalkaliphilic SOB of the genus *Thioalkalimicrobium* (Sorokin et al., 2001, 2002) and the hydrogen-oxidizing genus *Hydrogenovibrio* (Nishihara et al., 1991) are members of *Thiomicrospira pelophila* and *Thiomicrospira crunogena* clusters, respectively. Moreover, the SOB symbionts of marine clams *Bathymodiolus* and *Calyptogena* are firmly related to the whole *Thiomicrospira–Thioalkalimicrobium–Hydrogenovibrio* group (‘Thiomicrospira group’). This evidence of their divergence clearly demands reorganization of the genus *Thiomicrospira* into

Abbreviations: RSCU, relative synonymous codon usage; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; SOB, sulfur-oxidizing bacteria.
two genera based on \textit{Thiomicrospira pelophilia} and \textit{Thiomicrospira crunogena} clusters, and creation of a new family ‘\textit{Thiomicrospiraceae}’, which would include at least five separate genera: the two \textit{Thiomicrospira}-based genera, \textit{Thioalkalimicrobium}, \textit{Hydrogenovibrio} and, perhaps, the SOB symbionts of marine clams.

The obligate autotrophy of the ‘\textit{Thiomicrospira} group’ representatives is based on the high activity of the Calvin–Benson–Bassham cycle of inorganic carbon assimilation with ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) as the key enzyme. Although the potential for heterotrophy has been currently claimed for three strains of the \textit{Thiomicrospira crunogena} cluster (Takai et al., 2004), the evidence given is not very convincing. Interestingly, \textit{Thiomicrospira denitrificans}, a member of the \textit{Epsilonproteobacteria}, uses the reductive tricarboxylic acid cycle for autotrophic CO\textsubscript{2} assimilation (Hugler et al., 2005). This correlates with the anaerobic nature of this bacterium, which separates it from other \textit{Thiomicrospira} species.

Rubisco exists in two distinct forms. Form I Rubisco is composed of eight large subunits and eight small subunits (L\textsubscript{4}S\textsubscript{4}), which are encoded by the \textit{cbbL} and \textit{cbbS} genes, respectively. This form is widely distributed in CO\textsubscript{2}-fixing organisms, including all higher plants, algae, cyanobacteria and many autotrophic bacteria. Form I Rubisco is divided into two major groups, termed ‘green-like’ and ‘red-like’ (Delwiche & Palmer, 1996). Form II Rubisco is composed of large subunits only (L\textsubscript{4}), encoded by the \textit{cbbM} gene. This form is far restricted to several phototrophic purple bacteria, aerobic and facultatively anaerobic chemautotrophic bacteria, and dinoflagellates. In addition to these well-recognized forms, two novel types, forms III and IV Rubisco, have recently been revealed after the complete genome sequencing of some archaea and bacteria.

Some bacteria have been found to possess more than one set of Rubisco genes. \textit{Cupriavidus necator} H16 (formerly \textit{Alcaligenes eutrophus} H16) possesses two sets of almost identical genes which encode the red-like form I enzyme (Kusian et al., 1995), \textit{Allochromatium vinosum} (formerly \textit{Chromatium vinosum}) has two sets of divergent genes which encode the green-like form I enzyme (Viale et al., 1989) and \textit{Rhodobacter azotoformans} has two sets of genes which encode both green- and red-like form I enzymes (Uchino & Yokota, 2003). Furthermore, \textit{Halothiobacillus neapolitanus} (formerly \textit{Thiobacillus neapolitanus}), \textit{Thiomonas intermedia} (formerly \textit{Thiobacillus intermedium}), \textit{Thiobacillus denitrificans}, \textit{Rhodobacter sphaeroides} and \textit{Rhodobacter capsulatus} have genes for both form I and form II enzymes (Gibson & Tabita, 1977a, b; Shively et al., 1986; English et al., 1992; Stoner & Shively, 1993; Paoli et al., 1995). \textit{Acidithiobacillus ferrooxidans} (formerly \textit{Thiobacillus ferrooxidans}) and \textit{Hydrogenovibrio marinus} have three different sets of Rubisco genes, two encoding the green-like form I enzyme, whereas the third one encodes a form II enzyme (Kusano et al., 1991; Heinhorst et al., 2002; Yaguchi et al., 1994; Nishihara et al., 1998).

In this work the phylogenetic diversity and evolution of the Rubisco genes of some \textit{Thiomicrospira} and all \textit{Thioalkalimicrobium} species has been analysed with the aim of obtaining additional insight into the relatedness of different species and their clusters within this SOB group.

**METHODS**

**Bacterial strains.** Three type strains of the genus \textit{Thiomicrospira} were obtained from DSMZ. The cultures were grown on mineral medium with thiosulfate as a substrate as described by Brinkhoff et al. (1999a), except for \textit{Thiomicrospira pelophilia} where 100 \mu g vitamin B\textsubscript{12} 1\textsuperscript{-1} was added. Three type strains belonging to the \textit{Thioalkalimicrobium} genus were maintained in our laboratory and grown on a mineral medium with thiosulfate at pH 10, as described previously (Sorokin et al., 2001).

**DNA isolation and PCR amplification.** DNA extraction was performed as described previously (Boulygina et al., 2002). PCR was used to amplify the fragments of bacterial genes encoding the large subunit of red- and green-like form I Rubisco (\textit{cbbL}) and form II Rubisco (\textit{cbbM}) using specially developed and previously tested primer pairs (Spiridonova et al., 2004). PCR products were analysed by electrophoresis in 1-0% agarose gel stained with ethidium bromide and documented by the BioDoc Analyse System (Biometa). PCR products were purified through low-melting-point agarose using Wizard PCR Prep 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 using the Big Dye Terminator v.3.1 sequencing reaction kit (Applied Biosystems).

**Phylogenetic analysis.** The preliminary analysis of the new sequences was done using BLAST from the NCBI server (www.ncbi.nlm.nih.gov/blast/). The nucleotide and inferred amino acid sequences were aligned with sequences from GenBank using CLUSTAL W (Thompson et al., 1994). 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).

The relative synonymous codon usage (RSCU) values of the Rubisco genes were calculated using CODONW software (John Pedersen, www.molbiol.ox.ac.uk/cu). To investigate the major trends in codon usage in different species, CODONW was used to carry out a correspondent analysis. This resulted in a point in the codon space for each species, the positions of which sometimes suggested codon usage bias.

Levels of synonymous (dS) and non-synonymous (dN) nucleotide diversities were calculated with the YN00 program (PAML package: Yang, 2000) using the method of Yang & Nielsen (2000).

**RESULTS**

**Detection and amplification of the Rubisco genes**

Using the specific primer set for the green-like \textit{cbbL} gene, fragments of about 800 bp were amplified from the DNAs of
the type strains of *Thiomicrospira crunogena*, *Thiomicrospira kuenenii*, *Thiomicrospira pelophila*, *Thioalkalimicrobium aerophilum*, *Thioalkalimicrobium cyclicum* and *Thioalkalimicrobium sibiricum*. The use of the primer set specific for red-like *cbbL* gave negative results for all species. The application of the *cbbM*-amplifying primer set resulted in the appearance of a PCR product of about 800 bp only for *Thiomicrospira* species.

The clones prepared from the *cbbL* PCR fragments yielded a single sequence-type for all three *Thioalkalimicrobium* species and *Thiomicrospira pelophila*, but two sequence-types for *Thiomicrospira crunogena* and *Thiomicrospira kuenenii*. The clones prepared from the *cbbM* PCR fragments yielded only one sequence-type for all *Thiomicrospira* species.

The results of the BLAST analysis indicated that all newly obtained sequences belonged to the Rubisco family. *Thiomicrospira crunogena* and *Thiomicrospira kuenenii* possessed three genes (*cbbL-1*, *cbbL-2* and *cbbM*), *Thiomicrospira pelophila* possessed two genes (*cbbL* and *cbbM*) and the *Thioalkalimicrobium* species possessed only a single Rubisco gene (*cbbL*).

**Phylogenetic analysis of the *cbbL* sequences**

The nucleotide and amino acid sequences of *cbbL* were aligned, the positions with gaps and ambiguous sequences were removed and the remaining 738 nt and 246 aa were used for further phylogenetic analysis. The phylogenetic trees constructed by neighbour-joining (Fig. 1), maximum-parsimony, Fitch–Margoliash and maximum-likelihood algorithms (data not shown) were similar for each method (data not shown) were similar for each method with minor exceptions both for nucleotide-based and amino-acid-based phylogenetic trees. In both nucleotide- and amino-acid-based trees, *cbbL-2* of *Thiomicrospira crunogena* and *Thiomicrospira kuenenii*, and *cbbL* of *Thiomicrospira pelophila* and three *Thioalkalimicrobium* species formed a single cluster with high bootstrap values (100 and 97 % based on nucleotides and amino acids, respectively) with the *cbbL-2* of the marine hydrogen-oxidizing SOB *Hydrogenovibrio marinus*. Similarly, *cbbL-1* of *Thiomicrospira crunogena* and *Thiomicrospira kuenenii* formed a single cluster with high bootstrap values (100 % for both nucleotide- and amino-acid-based trees) with *cbbL-1* of *Hydrogenovibrio marinus*.

However, for the ‘*Thiomicrospira* group’ as a whole, the trees based on nucleotides and amino acids differed in their topologies and in the lengths of the branches (compare Fig. 1a and b). In the nucleotide-based tree, *cbbL-1* and *cbbL-2* formed a single group with an 87 % bootstrap value, clustering with *rbcL* of the cyanobacterium *Prochlorococcus marinus* with an 84 % bootstrap value, with very long branching. The closest neighbour of this cluster was the major cluster of Rubisco genes of the cyanobacteria *Prochlorothrix hollandica*, *Synechococcus* sp. PCC 6301, *Synechococcus* sp. PCC 7002 and *Anabaena* sp. PCC 7120 with a 92 % bootstrap value. In the tree based on amino acids, *cbbL-1* and *cbbL-2* of the ‘*Thiomicrospira* group’ formed two separate clusters with uncertain branching, whereas *Prochlorococcus marinus* clustered with *Synechococcus* sp. WH7803 at a high bootstrap value (99 %) with almost equal branch length. Also, the relatedness of the ‘*Thiomicrospira* group’ and the major cluster of cyanobacteria was not detected in the latter tree.

The results of *cbbL* analysis within the ‘*Thiomicrospira* group’ correlated with its phylogenetic clustering based on 16S rRNA analysis (Fig. 2). In this tree, *Thiomicrospira pelophila* formed a cluster with *Thioalkalimicrobium* species (‘*Thiomicrospira pelophila cluster*’), whereas *Thiomicrospira crunogena* and *Thiomicrospira kuenenii* formed another cluster with *Hydrogenovibrio marinus* (‘*Thiomicrospira crunogena* cluster’). In compliance with these data, single *cbbL* genes of the *Thioalkalimicrobium* species were closely related (88-9 % nucleotide similarity and 97-2 % amino acid identity) to a single *Thiomicrospira pelophila* *cbbL* gene, whereas the *cbbL-2* and *cbbL-1* genes of *Thiomicrospira crunogena* and *Thiomicrospira kuenenii* were closer to *cbbL-2* (87-7–90-1 % nucleotide similarity and 96-8–98-0 % amino acid identity) and *cbbL-1* (85-2–92-0 % nucleotide similarity and 89-6–93-0 % amino acid identity) of *Hydrogenovibrio marinus*, respectively. However, a closer relatedness of *cbbL* genes of this whole group to cyanobacterial *rbcL* genes contradicts its phylogenetic position within the *Gammaproteobacteria* in a 16S-rRNA-based tree.

**Phylogenetic analysis of *cbbM***

The nucleotide and amino acid sequences of *cbbM* were aligned, the positions with gaps and ambiguous sequences were removed and the remaining 777 nt and 259 aa were used for the phylogenetic analysis. The topologies of the nucleotide-based and amino-acid-based phylogenetic trees constructed by neighbour-joining (Fig. 3), maximum-parsimony, Fitch–Margoliash and maximum-likelihood algorithms (data not shown) were similar for each method with minor exceptions.

In the nucleotide-based tree, the *cbbM* gene of *Thiomicrospira* species and of *Hydrogenovibrio marinus* formed a single cluster (at 100 % bootstrap value). This cluster fell into the radiation of thioautotrophic Beta- and Gammaproteobacteria belonging to the genera *Thiobacillus*, *Acidithiobacillus*, *Halothiobacillus* and *Thiomonas*. In the amino-acid-based tree, only *Thiomicrospira pelophila* and *Thiomicrospira crunogena* remained in the thioautotrophic *cbbM* cluster, whereas *Thiomicrospira kuenenii* and *Hydrogenovibrio marinus* with almost identical amino acids sequences (98-1 %) formed a separate branch with an uncertain branching point position.

**Nucleotide composition and codon usage**

Genes acquired by horizontal transfer often have atypical GC content, codon bias and repetitive elements (Medigue et al., 1991). Therefore, it was interesting to compare the
The phylogenetic position of the ‘Thiomicrospira’ group species in cbbL molecular trees based on sequence analysis of (a) nucleotides and (b) translated amino acids. The sequences determined in this study are marked by bold type. 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 the nodes indicate the percentage bootstrap values for the clade of this group in 1000 replications (the values for the maximum-parsimony method are given in parentheses).
GC content and codon usage of the RubisCO genes within the ‘Thiomicrospira group’ to detect the role of possible gene transfer in their evolution. The total GC content of all analysed RubisCO genes was close to the genomic GC content for each species of the group (45.1–48.9 mol% against 42.0–49.6 mol%, respectively). The GC3 content (third position of codons) of RubisCO gene sequences (30.3–37.0 mol%) was lower than the total GC content of the RubisCO genes and the overall genomic GC content.

**Fig. 2.** The phylogenetic position of ‘Thiomicrospira group’ species in the 16S rRNA molecular tree. The sequences for which RubisCO genes were determined are marked by underlining. Tree topography and evolutionary distances are given by the neighbour-joining method with Jukes and Cantor distances. Numbers at the nodes indicate the percentage bootstrap values for the clade of this group in 1000 replications (the values for the maximum-parsimony method are given in parentheses).
Codon usage analysis was carried out on the RSCU data. Correspondence analysis of the results (Fig. 4) identified major trends in codon usage: the y axis is associated with GC3 (Musto et al., 1998), whereas the x axis is correlated with the frequencies of codons ending in C or U versus A or G (Fennoy & Bailey-Serres, 1993). The codon usage of all species of the 'Thiomicrospira group' was typical of AT-biased micro-organisms in which codons with an A or T in the third position are used preferentially (Ohtaka & Ishikawa, 1993). In general, the codon usage of all RubisCO genes of the 'Thiomicrospira group' was almost identical and differed from the RubisCO genes of other autotrophs in the RSCU correspondence analysis plot (Fig. 4). Thus analysis of codon usage did not show any intra-group bias that might be the result of gene transfer in these species.

Based on codon usage, the closest neighbours of the 'Thiomicrospira group' were Anabaena sp. PCC 7120 and its immediate relatives. Moreover, the unusual position of the other cyanobacterium Prochlorococcus marinus on the RSCU correspondence analysis plot correlated with its unusual clustering in the cbbL nucleotide-based phylogenetic tree.

**Synonymous and non-synonymous substitution analysis**

Since synonymous (silent) mutations are largely invisible to natural selection, whereas non-synonymous (amino-acid-changing) mutations may be under strong selective pressure, comparison of the rates of fixation of these two types of mutation provides a powerful tool for understanding
the mechanisms of DNA sequence evolution (Yang & Nielsen, 2000). Therefore, synonymous and non-synonymous nucleotide substitution rates ($d_S$ and $d_N$) and their ratio ($\omega = d_N/d_S$) for the nucleotide sequences of $cbbL$ and $cbbM$ genes of the ‘Thiomicrospira group’ were calculated (Table 1). $d_N$ was lower than $d_S$ in all cases with clear evidence of selective
constraint on amino acid replacements. This suggests that the RubisCO genes have evolved in all lineages under negative or purifying selection.

For cbbL-2 in the genomes of all species of the group under study the ω values within the Thiomicrospira pelophila and Thiomicrospira crunogena clusters (0.0140–0.0291 and 0.0132–0.0200, respectively; see Table 1) were half those between these clusters (0.0294–0.0646). The high ω value (0.3363) for the almost identical cbbL sequences for the pair Thioalkalimicrobium aerophilum–Thioalkalimicrobium sibiricum could be considered as an exception. The intra-cluster ω values for cbbL-1 of the Thiomicrospira crunogena cluster (0.0486–0.1135) were about 4- to 5-times higher than for cbbL-2. Moreover, the intra-cluster ω values for cbbL-1 of this group were even higher than the inter-cluster ω values for cbbL-2 between the Thiomicrospira pelophila and Thiomicrospira crunogena clusters, especially for the pair Thiomicrospira kuenenii–Hydrogenovibrio marinus (0.1135). This could be explained by the fact that intra-cluster dN values for cbbL-1 (0.0345–0.0624) were comparable with inter-cluster dN values for cbbL-2 (0.0524–0.0670), whereas dS values at the intra-cluster level were comparable for both cbbL-1 and cbbL-2 (0.3039–0.8602 and 0.3885–0.8611, respectively).

The ω value for cbbM genes of the pair Thiomicrospira kuenenii–Hydrogenovibrio marinus was relatively low (0.0513) because of the low dN (0.0117) and dS (0.2282) values (Table 1). The dN and dS values for cbbM genes of other combinations were within the ranges of 0.0720–0.2740 and 1.4505–2.3291, respectively, which is much higher compared with the values for cbbL-1 and cbbL-2. However, the ω value for the pair Thiomicrospira pelophila–Thiomicrospira crunogena was much lower (0.0309) than the ω values for other pairs (0.1192–0.1823). This is a result of a relatively low non-synonymous nucleotide substitution rate (dN = 0.0720) for this pair in contrast to a high synonymous nucleotide substitution rate (dS = 2.3291).

These results demonstrated that the synonymous and non-synonymous nucleotide substitution rates among the cbbL-1, cbbL-2 and cbbM genes of the ‘Thiomicrospira group’ were different. This might be explained by the proposal that the selection pressure for cbbL-2 was higher than for cbbL-1 and cbbM and, therefore, by the higher significance of RubisCO encoded by the cbbL-2 gene.

**DISCUSSION**

The usage of functional genes encoding key metabolic enzymes as molecular markers is becoming common practice in phylogenetic studies. In the case of RubisCO it has been shown that phylogenetic reconstructions based on its analysis differ significantly from the results of traditional 16S rRNA-based studies for autotrophic organisms in

---

**Table 1.** Synonymous and non-synonymous nucleotide substitution ratios among three sets of RubisCO genes of the ‘Thiomicrospira group’

<table>
<thead>
<tr>
<th>Species pair</th>
<th>cbbL-1</th>
<th></th>
<th></th>
<th>cbbL-2</th>
<th></th>
<th></th>
<th>cbbM</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Thioalkalimicrobium aerophilum and Thioalkalimicrobium sibiricum</td>
<td>dN</td>
<td>dS</td>
<td>dN/dS</td>
<td>dN</td>
<td>dS</td>
<td>dN/dS</td>
<td>dN</td>
<td>dS</td>
<td>dN/dS</td>
</tr>
<tr>
<td>Thioalkalimicrobium aerophilum and Thioalkalimicrobium cyclicum</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0055</td>
<td>0.0162</td>
<td>0.3363</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium aerophilum and Thioalkalimicrobium pelophila</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0055</td>
<td>0.3896</td>
<td>0.0140</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium aerophilum and Thiomicrospira crunogena</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0145</td>
<td>0.6886</td>
<td>0.0210</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium aerophilum and Thiomicrospira kuenenii</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0524</td>
<td>1.7508</td>
<td>0.0299</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium aerophilum and Hydrogenovibrio marinus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0550</td>
<td>1.3241</td>
<td>0.0416</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium cyclicum and Thioalkalimicrobium sibiricum</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0585</td>
<td>1.0473</td>
<td>0.0559</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium cyclicum and Thioalkalimicrobium pelophila</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0110</td>
<td>0.3885</td>
<td>0.0282</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium cyclicum and Thiomicrospira kuenenii</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0201</td>
<td>0.7407</td>
<td>0.0271</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium cyclicum and Thiomicrospira crunogena</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0585</td>
<td>1.3927</td>
<td>0.0420</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium cyclicum and Thiomicrospira kuenenii</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0611</td>
<td>1.6184</td>
<td>0.0378</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium cyclicum and Hydrogenovibrio marinus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0647</td>
<td>1.1346</td>
<td>0.0570</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium sibiricum and Thioalkalimicrobium pelophila</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0200</td>
<td>0.6863</td>
<td>0.0291</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium sibiricum and Thiomicrospira crunogena</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0583</td>
<td>1.9817</td>
<td>0.0294</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium sibiricum and Thiomicrospira kuenenii</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0610</td>
<td>1.3102</td>
<td>0.0465</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thioalkalimicrobium sibiricum and Hydrogenovibrio marinus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0645</td>
<td>0.9981</td>
<td>0.0646</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Thiomicrospira pelophila and Thiomicrospira crunogena</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0640</td>
<td>1.5795</td>
<td>0.0405</td>
<td>0.0720</td>
<td>2.3291</td>
<td>0.0309</td>
</tr>
<tr>
<td>Thiomicrospira pelophila and Thiomicrospira kuenenii</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0670</td>
<td>1.1564</td>
<td>0.0579</td>
<td>0.2695</td>
<td>1.9061</td>
<td>0.1414</td>
</tr>
<tr>
<td>Thiomicrospira pelophila and Hydrogenovibrio marinus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0661</td>
<td>1.1417</td>
<td>0.0579</td>
<td>0.2644</td>
<td>1.4505</td>
<td>0.1823</td>
</tr>
<tr>
<td>Thiomicrospira crunogena and Thiomicrospira kuenenii</td>
<td>0.0400</td>
<td>0.8226</td>
<td>0.0486</td>
<td>0.0090</td>
<td>0.6803</td>
<td>0.0132</td>
<td>0.2740</td>
<td>1.5779</td>
<td>0.1737</td>
</tr>
<tr>
<td>Thiomicrospira kuenenii and Hydrogenovibrio marinus</td>
<td>0.0345</td>
<td>0.3039</td>
<td>0.1135</td>
<td>0.0110</td>
<td>0.5477</td>
<td>0.0200</td>
<td>0.0117</td>
<td>0.2282</td>
<td>0.0513</td>
</tr>
<tr>
<td>Thiomicrospira crunogena and Hydrogenovibrio marinus</td>
<td>0.0624</td>
<td>0.8602</td>
<td>0.0726</td>
<td>0.0163</td>
<td>0.8611</td>
<td>0.0190</td>
<td>0.2675</td>
<td>2.2445</td>
<td>0.1192</td>
</tr>
</tbody>
</table>
general (Watson & Tabita, 1997) and for particular groups, for example haloalkaliphilic SOB of the genus *Thioalkalivibrio* (Tourova et al., 2005).

In contrast to the latter case, phylogenetic analysis of the ‘Thiomicrospira group’ demonstrated a good correlation between 16S-rRNA- and RubisCO-based results. First of all, the analysis of the RubisCO genes showed a monophyletic origin of the group (including the previously studied genus *Hydrogenovibrio*), evident from the high probability of clustering of their RubisCO genes at the nucleotide level. The separate branching of *cbbl*-1 of the *Thiomicrospira crunogena* cluster and *cbbM* of the pair *Thiomicrospira kueneniai–Hydrogenovibrio marinus* at the amino acid level could be explained by an increased rate of non-synonymous nucleotide replacements in their RubisCO genes (Table 1).

A two-subgroup division of the ‘Thiomicrospira group’ based on 16S rRNA phylogeny correlated with form I RubisCO gene analysis. In particular, *cbbl*-1 genes were found only in the *Thiomicrospira crunogena* cluster, and the topology of *cbbl*-2- and 16S-rRNA-based trees was very similar. At the same time, subdivision of the group was not evident on the basis of the *cbbM* gene analysis, which might be a result of a different pathway of evolution of this RubisCO form and a different selection pressure on the form II enzyme for different species of the group, depending on their ecological niches. According to a recent hypothesis, form II might be the more ancient type of the enzyme, optimally functioning under anaerobic conditions and high CO2 concentration. In this case, form I can be considered as an aerotolerant descendant of form II (Watson & Tabita, 1997; Elsäied & Naganuma, 2001). Form II enzymes are conservative and uniform (Hernandez et al., 1996) in contrast to the more recent form I which has evolved into two types, green-like and red-like, according to their amino acid sequences. According to all the evidence presented above, it could be proposed that comparison of *cbbM* gene sequences allows us to trace only distant relatedness inside the ‘Thiomicrospira group’, but not more recent divergence of the group for two phylogenetic clusters. It is interesting to note that in the *cbbM*-based phylogenetic tree, the ‘Thiomicrospira group’ forms a monophyletic cluster with the other SOB of the genera *Acidithiobacillus*, *Halothiobacillus*, *Thiobacillus* and *Thiomonas*, which are currently assigned to the *Gamma*- and *Betaproteobacteria* based on 16S rRNA gene sequence analysis. It suggests a possible common origin of these chemolithoautotrophs with similar metabolism. In contrast, this common origin was not evident from the analysis of *cbbl* genes and this could be explained by lateral gene transfer of *cbbl* to the ancestor of the ‘Thiomicrospira group’ (see below).

Our data have increased the range of bacteria possessing multiple sets of *cbbl* genes. Among the reasons for the appearance of ‘multi-copy’ genes, duplication (with a probability of further selective loss of one copy) and lateral gene transfer are currently recognized. For almost identical copies, as in the case of two *cbbl* genes in *Cupriavidus necator* H16 and *Acidithiobacillus ferroxidans* Fe1, a recent duplication event is suggested (Kusano et al., 1991; Kusian et al., 1995). For the *cbbL* copies in *Allochromatium vinosum*, with significant sequence divergence but a common GC composition and codon usage, a more ancient duplication event is hypothesized (Viale et al., 1989; Kobayashi et al., 1991). On the other hand, the presence of two *cbbL* copies in *Acidithiobacillus ferroxidans* ATCC 23270 with significant differences in nucleotide sequence, GC ratios, and codon usage suggests lateral gene transfer as a mechanism of their origin (Heinhorst et al., 2002).

Analysis of the three sets of RubisCO genetic clusters present in the genome of the hydrogen-oxidizing member of the ‘Thiomicrospira group’, *Hydrogenovibrio marinus*, allows us to suggest the following method of their origin: the ancestor of this species, possessing the *cbbM* gene cluster, acquired *cbbL*-2 genes by lateral transfer, which, after duplication and rearrangement of other genes of the *cbbM* cluster, generated the *cbbL*-1 gene cluster (Yoshizawa et al., 2004).

In general our results are consistent with such a scenario. Taking into consideration evolutionary distances and codon usage, it might be suggested that, among modern autotrophs, cyanobacteria and not other photo- or chemoautotrophic bacteria could be the most probable donors of the ‘Thiomicrospira group’ *cbbL*-2 gene pool. The presence of distant but undoubtedly related *cbbL*-1 and *cbbL*-2 genes, and their identical GC content and codon usage in investigated SOB species, suggest the occurrence of a gene duplication event in the ancestral form of the ‘Thiomicrospira group’. Recent evolution of the ancestor that acquired all three types of RubisCO genes has resulted in selective loss of the *cbbM* and *cbbL*-1 in its alkaliphilic descendants (genus *Thioalkalimicrobium*) and the *cbbL*-1 gene in *Thiomicrospira pelophila*.

Such a loss might be a result of different catalytic properties of various forms of RubisCO. In particular, immunoblotting analysis revealed different expression of the three types of RubisCO genes in *Hydrogenovibrio marinus* depending on CO2 content. *cbbM* is exclusively expressed at high CO2 content (15%), both *cbbM* and *cbbL*-1 are expressed at intermediate CO2 concentrations, while expression of the *cbbL*-2 gene starts at low levels of CO2, approaching its recent atmospheric content, when all three forms are present (Yoshizawa et al., 2004). This example demonstrates that possession of isoenzymes with slightly different metabolic properties, such as three different forms of RubisCO in *Hydrogenovibrio marinus*, gives the bacterium a certain ecological advantage and flexibility. Assuming the close similarity of the genes encoding these forms in *Hydrogenovibrio marinus* and the *Thiomicrospira* species of the *Thiomicrospira crunogena* cluster, it seems likely that these SOB species used the same mechanism of adaptation to environmental changes. Remarkably high chemolithoautotrophic growth rates common for this SOB cluster might be a result of the presence of several forms of this key enzyme with variable catalytic properties. Depending on conditions,
for example the ratio of CO₂ to O₂, one or another form might have been underexpressed and, eventually, even dropped completely, as probably happened within the haloalkaliphilic genus *Thioalkalimicrobium*, possessing a single form I RubisCO best adapted to modern atmospheric conditions (Yoshizawa et al., 2004). The anaerobic photosynthetic SOB *Allochromatium vinosum* probably represents an intermediate stage in such an evolutionary course, possessing two different *cbbL* copies, one of which is practically not expressed (Kobayashi et al., 1991).

The *cbbL-1* gene encoding RubisCO form I appears to be a reserve enzyme for the investigated SOB group. If this is so, the rate of non-synonymous nucleotide replacements in this gene must increase with decreased selection pressure. The complete lack of *cbbL-1* and *cbbM* in the genus *Thioalkalimicrobium* (obligate alkaliophiles) and *Thiomicropsira pelophila* (alkalitolerant bacterium) (see Sorokin & Kuenen, 2005) might have something to do with adaptation to high carbonate alkalinity: one of the possible explanations could be low actual CO₂ concentration at pH above 8.

**Taxonomic implications**

The use of phylogenetic analysis of genes other than those for 16S rRNA as molecular markers in bacterial taxonomy is not yet customary despite the obvious advantage of such information for genes encoding key metabolic enzymes vitally important for an organism’s survival. However, inclusion of additional molecular markers might help to solve some complicated taxonomic and evolutionary problems, such as the current example of the ‘*Thiomicropsira* group’. Both the 16S rRNA and RubisCO gene sequence analyses strongly support the necessity of taxonomic revision of this group, more specifically, dividing it into four genera within a new monophyletic family, the ‘*Thiomicropsiraceae*’. The genera *Thioalkalimicrobium* (Sorokin et al., 2001) and *Hydrogenovibrio* (Nishihara et al., 1991) are sufficiently separated from *Thiomicropsira* physiologically and genetically, but *Thiomicropsira* has to be divided into two genera based on the *Thiomicropsira crunogena* and *Thiomicropsira pelophila* clusters, while ‘*Thiomicropsira denitrificans*’, a member of the *Epsilonproteobacteria*, should certainly be removed from the group. The possibility of including the symbiotic SOB into this new family should also be considered, but this requires more data on their physiology and RubisCO-based phylogeny.

**ACKNOWLEDGEMENTS**

This work was supported by the Russian Foundation for Basic Research (grant 05-04-48064).

**REFERENCES**


expression of genes encoding the large and small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Chromatium vinosum*. *Gene* 97, 55–62.


