The phylogenetic placement of the non-phototrophic, Gram-positive thermophile ‘Thermobaculum terrenum’ and branching orders within the phylum ‘Chloroflexi’ inferred from gene order comparisons

Takashi Kunisawa

Department of Applied Biological Sciences, Science University of Tokyo, 2641 Yamasaki, Noda 278-8510, Japan

The phylogenetic position of an anaerobic, non-spore-forming thermophile ‘Thermobaculum terrenum’ was investigated on the basis of gene order data from completely sequenced bacterial genomes. Gene order data can be an excellent source of phylogenetic information. Shared gene arrangements are unlikely to have arisen by chance convergence. They are likely to reflect common ancestry. ‘Thermobaculum terrenum’ was found to share three gene arrangements that are present uniquely in genomes of members of the phylum ‘Chloroflexi’, indicating convincingly that ‘Thermobaculum terrenum’ is a member of this phylum. Branching orders within the phylum ‘Chloroflexi’ were inferred by identifying monophyletic groups of species, which were circumscribed by characteristic gene arrangements. The branching orders thus inferred were in good agreement with previously reported phylogenies based on single 16S rRNA gene sequences and on multiple protein sequences. The gene order comparisons revealed a close phylogenetic affinity of ‘Thermobaculum terrenum’ to Sphaerobacter thermophilus and Thermomicrobium roseum.

INTRODUCTION

‘Thermobaculum terrenum’ was originally isolated from an extreme thermal soil in Yellowstone National Park, USA (Botero et al., 2004). The initial phylogenetic analysis of 16S rRNA gene sequences placed this anaerobic, non-spore-forming hyperthermophilic bacterium as being most closely related to the phyla ‘Chloroflexi’ and ‘Thermomicrobia’ (Botero et al., 2004). The phylum ‘Thermomicrobia’ was later emended to be a class within the phylum ‘Chloroflexi’ on the basis of more sequence data from the 16S rRNA genes (Hugenholtz & Stackebrandt, 2004). This phylogenetic replacement has been supported by a recent phylogenomic analysis of Thermomicrobium roseum based on a concatenated alignment of 31 housekeeping gene products (Wu et al., 2009). The genome of ‘Thermobaculum terrenum’ ATCC BAA-798 was recently completely sequenced (GenBank accession no. NC_013525, S. Lucas and others); the genomic sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/). This bacterium is annotated as an unclassified bacterium in the NCBI genome database. Thus, it is unclear whether ‘Thermobaculum terrenum’ is a member of the phylum ‘Chloroflexi’, as in the case of Thermomicrobium roseum, or whether this species constitutes its own phylum.

Phylogenetic relationships have commonly been inferred based on comparisons of sequence data from a single protein (or gene). Recently, it has widely been accepted that genome trees reconstructed based on a concatenated alignment of multiple protein sequences are less susceptible to stochastic errors than those based on a single protein or gene and provide more reliable phylogenetic information (e.g. Delsuc et al., 2005). Identification of insertion/deletion sequences characteristically conserved among a group of species (signature sequences) provides another line of phylogenetic information that is obtained independently of the genome tree method (Gupta, 1998, 2010). Signature proteins that are present in a specific group of bacteria also provide phylogenetic information (see references cited in Gupta, 2010). The other approach is to compare the order of genes among different genomes (Sankoff et al., 1992). If genome rearrangements occur at random positions on a bacterial genome harbouring more than 1000 genes, then the probability of finding gene B next to a given gene A is approximately 1/1000. Thus, the shared gene adjacency reflects common ancestry rather than chance convergence. Although gene order is generally not a conserved property, restricted portions of bacterial
genomes show gene arrangements that have been conserved between distantly related phyla. In previous papers, I have described four sets of gene arrangements that are found uniquely in a group of the major bacterial phyla, ‘Firmicutes’, ‘Actinobacteria’, ‘Thermotogae’, ‘Deinococcus–Thermus’, ‘Fusobacteria’ and ‘Chloroflexi’ (Kunisawa, 2006, 2010). In contrast, other phyla, such as the ‘Proteobacteria’, ‘Chlorobi’, ‘Verrucomicrobia’ and ‘Acidobacteria’, show alternative gene arrangements. From these comparisons, the phylum ‘Chloroflexi’, for instance, is inferred to be grouped within the former group of phyla. It is to be noted that, unlike the genome tree and indel methods, gene order comparison does not rely on multiple sequence alignments and therefore can provide an independent measure of the reliability of evolutionary trees obtained using aligned sequences. The accuracy of multiple sequence alignment methods is to be noted.

In this work, the phylogenetic placement of ‘Thermobaculum terrenum’ is investigated on the basis of gene order comparison of completely sequenced bacterial genomes. According to the NCBI Genome Database, at the time of writing, 11 genomic sequence data of six genera, ‘Dehalococcoides’, Roseiflexus, Chloroflexus, Herpetosiphon, Sphaerobacter, Thermomicrobiurn, are available within the phylum ‘Chloroflexi’. If one identifies gene arrangements that are shared characteristically by ‘Chloroflexi’ and ‘Thermobaculum terrenum’ genomes but are not found elsewhere, then ‘Thermobaculum terrenum’ is likely to be a member of the phylum ‘Chloroflexi’. I describe three such gene arrangements uniquely found in genomes of members of the phylum ‘Chloroflexi’, which serve to define and circumscribe that phylum. I also attempt to reconstruct branching orders within the phylum ‘Chloroflexi’ by identifying a monophyletic group of species from gene transfer-suggesting arrangements (Kunisawa, 2001, 2003). The branching orders thus inferred from gene order comparisons are consistent with the recently reported genome tree based on a concatenated alignment of sequences from species of the phylum ‘Chloroflexi’ (Wu et al., 2009), although ‘Thermobaculum terrenum’ is not included in their analysis. The gene order comparisons indicate that ‘Thermobaculum terrenum’ is most closely related to members of the genera Sphaerobacter and Thermomicrobiurn.

METHODS

Gene order data. Complete genomic sequence data of bacteria were downloaded from the ftp server at NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). Gene order data for 590 bacterial genomes were prepared as described in a previous paper (Kunisawa, 2001), in which a substantial portion of the genomes compared in this study was listed. The downloaded genomic sequence data of the phylum ‘Chloroflexi’ were as follows: ‘Dehalococcoides ethenogenes’ 195 (GenBank accession no. CP00027, Seshadri et al., 2005), ‘Dehalococcoides’ sp. CBDB1 (NC_007356, Kube et al., 2005) and BAV1 (NC_009455, A. Copeland and others), Roseiflexus castenholzii DSM 13941T (CP000804, A. Copeland and others) and Roseiflexus sp. RS-1 (NC_009523, A. Copeland and others), Chloroflexus auranticus J-10-f (NC_010175, A. Copeland and others), Chloroflexus aggregans DSM 9485T (NC_011831, S. Lucas and others), Chloroflexus sp. Y-400-f (NC_012032, S. Lucas and others), Herpetosiphon aurantiacus ATCC 23779T (NC_009972, A. Copeland and others), Sphaerobacter thermophilus DSM 20745T (NC_013523, S. Lucas and others) and Thermomicrobium roseum DSM 5159T (NC_011959, Wu et al., 2009). Most of these sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/).

The computer search identified three gene arrangements that were present uniquely in species of the phylum ‘Chloroflexi’ and ‘Thermobaculum terrenum’. The first example was found around the nusB gene (COG0781) encoding a transcription termination protein. In Fig. 1 the gene arrangements of the species of the phylum ‘Chloroflexi’ are compared with those of other species from the phyla ‘Thermotogae’, ‘Ddictyoglomi’, ‘Deinococcus–Thermus’, ‘Fusobacteria’, ‘Actinobacteria’ and ‘Firmicutes’. Genes are represented by their four digit COG number and common gene names are shown at the bottom of the gene alignment. Homologous genes are vertically aligned and contiguous genes are connected by dashes, while split segments are interrupted by forward slashes. For the

RESULTS

Phylogenetic placement of the genus ‘Thermobaculum’

The computer search identified three gene arrangements that were present uniquely in species of the phylum ‘Chloroflexi’ and ‘Thermobaculum terrenum’. The first example was found around the nusB gene (COG0781) encoding a transcription termination protein. In Fig. 1 the gene arrangements of the species of the phylum ‘Chloroflexi’ are compared with those of other species from the phyla ‘Thermotogae’, ‘Ddictyoglomi’, ‘Deinococcus–Thermus’, ‘Fusobacteria’, ‘Actinobacteria’ and ‘Firmicutes’. Genes are represented by their four digit COG number and common gene names are shown at the bottom of the gene alignment. Homologous genes are vertically aligned and contiguous genes are connected by dashes, while split segments are interrupted by forward slashes.
maximum alignment, open reading frames are represented simply as dots, and 'n' is denoted when the corresponding gene is not identified. For example, the *pepQ* (COG0006) gene is located immediately upstream of the *efp* gene (COG0231) in *Kosmotoga olearia*, while in another member of the phylum 'Thermotogae', *Thermotoga maritima*, the *pepQ* gene is not adjacent to the *efp* gene and is located elsewhere on its genome. The *xseB* gene (COG1722) is not identified in the genome of *K. olearia*, and 'n' is indicated at the corresponding position in the gene alignment. As can be seen in Fig. 1, in those phyla shown, i.e. 'Thermotogae', 'Dictyoglomi', 'Deinococcus–Thermus', 'Fusobacteria', 'Actinobacteria' and 'Firmicutes' (superphylum 1), a conserved gene cluster (*accC*-yqhY-nusB) is found between the *efp* (COG0231) and *folD* (COG0190) genes. In order to show the degree of gene order conservation, the number of species showing the indicated or similar order and the number of genomes compared are indicated as the numerator and denominator, respectively, in the parentheses following the phylum name. In contrast to the arrangements depicted in Fig. 1, the other group of phyla, i.e. 'Aquificae', 'Nitrosporae', 'Acidobacteria', 'Verrucomicrobia' and 'Proteobacteria' (superphylum 2), the *nusB* gene tends to be located immediately downstream of the *ribH* gene (COG0054), as described in previous papers (Kunisawa, 2006, 2010). Some of the species of the phylum 'Chloroflexi' show the adjacency *accC*-nusB as seen in superphylum 1. Based on this gene adjacency, the phylum 'Chloroflexi' has been inferred to be a member of superphylum 1 (Kunisawa, 2006, 2010). However, the gene location of *nusB* is different between the phylum 'Chloroflexi' and other members of superphylum 1. In species of the phylum 'Chloroflexi', the entire cluster *accC*-yqhY-nusB or its fragments, *accC*-nusB or *nusB* alone, are present between the *fabG* (COG1028) and *acpP* (COG0236) genes, while in other superphylum 1 species, the *fabG* and *acpP* genes are adjacent to each other. Since the *nusB* gene is present in a single copy in almost all the bacterial genomes, the most straightforward explanation of this observation may be that the *accC*-yqhY-*nusB* cluster was translocated in evolution from the region adjacent to the *efp* and *folD* genes to a different region neighbouring the *fabG* and *acpP* genes within a common ancestral genome of the species of the phylum 'Chloroflexi'. In 'Thermobaculum terrenum', the *accC*-nusB arrangement is present between the

**Fig. 1.** Gene arrangements suggesting a transfer of the *nusB* (COG0781) gene. Genes are represented by the four-digit COG number and common gene names are shown at the bottom. Homologous genes are vertically aligned. Contiguous genes are connected by '-', while split segments are interrupted by '/'. For the maximum alignment, open reading frames are represented simply as '.', and 'n' denotes cases where the corresponding gene is not identified. Gene orders are shown for all the members of the phylum 'Chloroflexi', whereas conserved orders are illustrated for only two members belonging to other phyla. The transferred genes are boxed. Note that 'Thermobaculum terrenum' shows a gene arrangement similar to those present in the species of the phylum 'Chloroflexi'.

---

**Table 1.** Gene content of the region including the *accC*-yqhY-nusB cluster in various species. The *accC*-yqhY-nusB cluster is boxed.**

<table>
<thead>
<tr>
<th>Species</th>
<th>COG Numbers</th>
<th>Gene Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermotoga maritima</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Kosmotoga olearia</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Dictyoglomi</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Deinococcus-thermus</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Thermus thermophilus HB1</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Fusobacteria</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Fusobacterium nucleatum</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Leptotrichia buccalis</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Mycobacterium tuberculosis</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Atopobium parvulum</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Firmicutes</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Thermotoga maritima</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Thermoanaerobacter sp. X514</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>0006/0231/</td>
<td><em>pepQ</em></td>
</tr>
</tbody>
</table>

---

**Legend:** *pepQ* (COG0006) gene is located immediately upstream of the *efp* gene (COG0231) in *Kosmotoga olearia*, while in another member of the phylum 'Thermotogae', *Thermotoga maritima*, the *pepQ* gene is not adjacent to the *efp* gene and is located elsewhere on its genome. The *xseB* gene (COG1722) is not identified in the genome of *K. olearia*, and 'n' is indicated at the corresponding position in the gene alignment. As can be seen in Fig. 1, in those phyla shown, i.e. 'Thermotogae', 'Dictyoglomi', 'Deinococcus–Thermus', 'Fusobacteria', 'Actinobacteria' and 'Firmicutes' (superphylum 1), a conserved gene cluster (*accC*-yqhY-nusB) is found between the *efp* (COG0231) and *folD* (COG0190) genes. In order to show the degree of gene order conservation, the number of species showing the indicated or similar order and the number of genomes compared are indicated as the numerator and denominator, respectively, in the parentheses following the phylum name. In contrast to the arrangements depicted in Fig. 1, the other group of phyla, i.e. 'Aquificae', 'Nitrosporae', 'Acidobacteria', 'Verrucomicrobia' and 'Proteobacteria' (superphylum 2), the *nusB* gene tends to be located immediately downstream of the *ribH* gene (COG0054), as described in previous papers (Kunisawa, 2006, 2010). Some of the species of the phylum 'Chloroflexi' show the adjacency *accC*-nusB as seen in superphylum 1. Based on this gene adjacency, the phylum 'Chloroflexi' has been inferred to be a member of superphylum 1 (Kunisawa, 2006, 2010). However, the gene location of *nusB* is different between the phylum 'Chloroflexi' and other members of superphylum 1. In species of the phylum 'Chloroflexi', the entire cluster *accC*-yqhY-*nusB* or its fragments, *accC-nusB* or *nusB* alone, are present between the *fabG* (COG1028) and *acpP* (COG0236) genes, while in other superphylum 1 species, the *fabG* and *acpP* genes are adjacent to each other. Since the *nusB* gene is present in a single copy in almost all the bacterial genomes, the most straightforward explanation of this observation may be that the *accC*-yqhY-*nusB* cluster was translocated in evolution from the region adjacent to the *efp* and *folD* genes to a different region neighbouring the *fabG* and *acpP* genes within a common ancestral genome of the species of the phylum 'Chloroflexi'. In 'Thermobaculum terrenum', the *accC-*nusB arrangement is present between the
fabG and acpP genes just as in the species of the phylum ‘Chloroflexi’, indicating convincingly that ‘Thermobaculum terrenum’ is a member of the phylum ‘Chloroflexi’.

Fig. 2 shows two other gene arrangements shared uniquely by the species of the phylum ‘Chloroflexi’ and ‘Thermobaculum terrenum’. In this identification, computer search conditions on gene arrangements were relaxed to those of a-x-b in some genomes and c-x (or x-d) in other genomes. While the rpmE (COG0254) gene encoding ribosomal protein L31 is located between the rho (COG1158) and hemK (COG2890) genes in most of the members of the phyla ‘Firmicutes’ and ‘Actinobacteria’ as shown in Fig. 2a, the ribosomal protein gene is found immediately downstream of a gene cluster rplU (COG0261)-rpmA (COG0211) in the species of the phylum ‘Chloroflexi’ (Fig. 2a), suggesting a transfer of the rpmE gene in their common ancestor. Since the rplU-rpmA-rpmE gene arrangement is present in the genome of ‘Thermobaculum terrenum’, but is not found in other species, ‘Thermobaculum terrenum’ is most likely to be a member of the phylum ‘Chloroflexi’. Another example of such a gene transfer is inferred from the gene arrangements shown in Fig. 2b; while the mazG (COG1694) gene is present between the mfd (COG1197) and himA (COG0776) genes in some species of the phyla ‘Firmicutes’ and ‘Actinobacteria’, its orthologue is found downstream of the lepA (COG0481) gene or rpsT (COG0268) gene in the phylum ‘Chloroflexi’, accompanying the surA (COG0760) gene in some cases. Here, the rpsT gene is encoded on the complementary strand and is denoted with an asterisk. Thus, the mazG gene, probably together with the surA gene, is likely to have been transferred in a common ancestor of the species of the phylum ‘Chloroflexi’ with a reservation about H. aurantia-cus, in which the synteny is disrupted. As can be seen in Fig. 2b, the gene arrangement around the mazG gene in ‘Thermobaculum terrenum’ is identical or similar to those found in other members of the phylum ‘Chloroflexi’. This observation provides empirical evidence that ‘Thermobaculum terrenum’ should be included in the phylum ‘Chloroflexi’.

Branching orders within the phylum ‘Chloroflexi’

It is attempted here to infer branching orders among the species of the phylum ‘Chloroflexi’ on the basis of gene order comparisons.

(i) ‘Dehalococcoides’ as a deepest branching genus.

The genus ‘Dehalococcoides’ appears to be a most deeply branching genus among the completely sequenced species of the phylum ‘Chloroflexi’ on the basis of the following gene arrangements that are uniquely found only in a specific group of the ‘Chloroflexi’ species. As shown in Fig. 3a, the tatC gene (COG0805) coding for a Sec-independent protein secretion pathway component is located immediate downstream of the musB (COG0781)-acpP (COG0236) cluster in the species of the phylum ‘Chloroflexi’ other than those of the genus ‘Dehalococcoides’. In contrast, tatC orthologues are not located adjacent to the acpP gene, but are located downstream of the tatA (COG1826) gene in the genus ‘Dehalococcoides’. The adjacency of tatA-tatC can be recognized in various bacteria from the phyla ‘Firmicutes’, ‘Actinobacteria’ or ‘Deinococcus–Thermus’. Therefore, the

---

![Fig. 2. Gene arrangements present uniquely in the species of the phylum ‘Chloroflexi’. Suggested transfers of gene(s) in a common ancestor of the phylum ‘Chloroflexi’; (a) rpmE (COG0254), (b) surA (COG0760)–mazG (COG1694). The rpmE gene (COG0254) resides on the complementary strand and is denoted with an asterisk.](http://ijs.sgmjournals.org)
(a) Deinococcus-Thermus

Deinococcus geothermalis 0236/ 1829/ 0801
Thermus thermophilus HB27 0236/ 1829/ 0800
Actinobacteria

Mycobacterium tuberculosis 0236/ 1829/ 0800
Atopobium parvulum 0236/ 1829/ 0800
Firmicutes

Thermosnerobacter sp. X514 0236/ n n
Staphylococcus aureus 0236/ 1829/ 0800
Chloroflexi

Dehalococci des ethanogenes 0781–0236/ 1829/ 0800
Dehalococci des CDB01 0781–0236/ 1829/ 0800
Dehalococci des sp. BAV1 0781–0236/ 1829/ 0800

Chloroflexi

Herpetosiphon aurantiacus 0781–0236/ n 0800
Roseiflexus sp. RS-1 0781–0236/ n 0800
Roseiflexus castenholzi 0781–0236/ n 0800
Chloroflexus aurantis 0781–0236/ n 0800
Chloroflexus aggregans 0781–0236/ n 0800
Chloroflexus sp. Y-400-f1 0781–0236/ n 0800
Sphaerobacter thermophilus 0781–0236/ n 0800
Thermomicrobium roseum 0781–0236/ n 0800
Thermobaculum termen 0781–0236/ n 0800

Chloroflexi

Thermotogae

Thermosnerobacter sp. X514 0468–2137/ n 1418/ n
Staphylococcus aureus 0468–2137/ n 1418/ n
Actinobacteria

Mycobacterium tuberculosis 0468–2137/ n 1418/ n
Atopobium parvulum 0468–2137/ n 1418/ n

Firmicutes

Dehalococci des ethanogenes 0468–2137/ n 1418–1692/ 2359/
Dehalococci des CDB01 0468–2137/ n 1418–1692/ 2359/
Dehalococci des sp. BAV1 0468–2137/ n 1418–1692/ 2359/

Firmicutes

Herpetosiphon aurantiacus 0468–2137/ n 1418–1692/ 2359/
Roseiflexus sp. RS-1 0468–2137/ n 1418–1692/ 2359/
Roseiflexus castenholzi 0468–2137/ n 1418–1692/ 2359/
Chloroflexus aurantis 0468–2137/ n 1418–1692/ 2359/
Chloroflexus aggregans 0468–2137/ n 1418–1692/ 2359/
Chloroflexus sp. Y-400-f1 0468–2137/ n 1418–1692/ 2359/
Sphaerobacter thermophilus 0468–2137/ n 1418–1692/ 2359/
Thermomicrobium roseum 0468–2137/ n 1418–1692/ 2359/
Thermobaculum termen 0468–2137/ n 1418–1692/ 2359/

(ii) The HRC clade. The present computer search suggested a monophyletic grouping of H. aurantiacus, members of the genera Roseiflexus and Chloroflexus (HRC clade), exclusive of the other species of the phylum ‘Chloroflexi’. As shown in Fig. 4, the purH gene (COG0138) encoding phosphoribosyl aminomimidazole carboxyl formyl formyltransferase and the hisI gene (COG0139) encoding phosphoribosyl-AMP cyclohydrolase are located adjacent to each other in this HRC clade. In contrast, the genes purH and hisJ are present separately at the 3′ region of the purM (COG0150) or purN (COG0299) gene and of the hisA (COG0106) or hisF (COG0107) gene, respectively, in the species of the genus ‘Dehalococci des’, S. thermophilus, Thermomicrobium roseum and ‘Thermobaculum termen’ as well as in some species from other phyla, the ‘Firmicutes’, ‘Actinobacteria’, ‘Fusobacteria’ and ‘Thermotogae’. The purH-hisJ arrangement is thus found uniquely in the HRC clade, which indicates that this gene adjacency has been created in a common ancestor of the HRC clade.

The gene arrangement purL (COG0047)-purQ (COG0046) or its reverse arrangement is found in the species of the genus ‘Dehalococci des’, H. aurantiacus, S. thermophilus, Thermomicrobium roseum and ‘Thermobaculum termen’ as well as in other bacteria from the phyla ‘Firmicutes’, ‘Actinobacteria’ and so on, while this adjacency is disrupted in the members of the genera Roseiflexus and Chloroflexus, as seen in Fig. 5a. Furthermore, the indicated arrangements in the genera Roseiflexus and Chloroflexus are not found.
Thermotoga
Thermotoga maritima 0034--0299 0130 0151--0150/ 0106--0107--0139
Thermotoga petrophila 0034--0299 0130 0151--0150/ 0106--0107--0139
Fusobacteria
Streptobacillus moniliformis 0034--0150--0299 0130--0151/ 0106--0107--0139
Sebalbella termiditis 0034--0150--0299 0130--0151/ n n n
Actinobacteria
Rubrobacter xylomoniphilus 0034--0150--0299 0130 0107--0139
Eggerthella lenta 0034--0150--0299 0106--0107--0139
Firmicutes
Syntrophomonas wolfei 0034--0150--0299 0130--0151/ 0106--0107--0139
Thermococcus sp. X014 0034--0150--0299 0106--0107--0139
Chloroflexi
Dehalococcoides ethenogenes 0034--0150--0299 0130 0106--0107--0139
Dehalococcoides sp. CCDBI 0034--0150--0299 0106--0107--0139
Dehalococcoides sp. BAV1 0034--0150--0299 0106--0107--0139
Sphaero bacter thermophillus / 0150--0130 0106--0107--0139
Thermomicrobi um roseum / 0150--0130 0106--0107--0139
Thermobaculum terr enum / 0299--0130 0106--0107--0139

Chloroflexi
Herpetosiphon aurantius 0130--0130
Roseiflexus sp. RS-1 / 0130--0130--0642
Roseiflexus castenholzi / 0130--0130--0642
Chloroflexus aurantiacus / 0130--0130--0642
Chloroflexus aggregans / 0130--0130--0642
Chloroflexus sp. Y-400-fl / 0130--0130--0642

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Gene Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>purF purM purN</td>
<td>hisA hisF hisN</td>
</tr>
</tbody>
</table>

**Fig. 4.** Gene adjacency of purH (COG0138)-his (COG0139) in the HRC clade.

(iii) The STT clade. The present comparisons suggest another monophyletic grouping of *S. thermophilus*, *Thermomicrobium roseum* and *Thermobaculum terrenum* (STT clade). As shown in Fig. 6a, in the members of this group a gene cluster clpA (COG0542)-sms (COG1066) is located immediately downstream of the rpsA gene (COG0539) and another cluster yacL (COG1855)-ispD (COG1211) is present at the 3' region of the COG0613 gene, namely, these two clusters are located separately on their genome. The adjacencies of clpA to rpsA and of yacL to COG0613 are only found in the 3' region of the *Chromatium* genome, suggesting a split into the two clusters in the lineage leading to the STT clade exclusive of other members of the phylum *Chloroflexi*.

Similarly, a unique gene adjacency ruvC (COG0817)-yqxC (COG1119) is likely to have been created in a common ancestor of the STT clade, as can be inferred from Fig. 6b. The gene arrangement yel (COG0217)-ruvC (COG0817) is found not only in the phyla *Actinobacteria* and *Dictyoglomi* but also in members of the genus *Dehalococcoides*, members of the genera *Roseiflexus* and *Chloroflexus*, and *Thermobaculum terrenum*, while in *S. thermophilus* and *Thermomicrobi um roseum* the yec (COG0217) gene is located separately from the yqxC (COG1119) gene and is present at the 3' region of the srmB gene (COG0513). From these observations, it is possible to infer that the gene adjacency between yec-ruvC and yqxC has been created in a common ancestor of the STT members and then the yec gene has been transferred to downstream of srmB in a common ancestor of *S. thermophilus* and *Thermomicrobi um roseum*. Besides the STT clade, the srmB-yec arrangement can also be seen in *Mycoplasma penetrans* as a very rare exceptional case; 15 other completely sequenced mycoplasma genomes do not show this arrangement. This move of the yec gene suggests that *Thermobaculum terrenum* is most closely related to *S. thermophilus* and *Thermomicrobi um roseum*.

Besides genes coding for proteins, a tRNA-specifying gene has been suggested as a candidate gene for transfer. A threonine-tRNA gene with the anticodon sequence GGT (Tggt) is present between a tyrosine-tRNA gene with the anticodon sequence GTA (Ygta) and an elongation factor encoding gene tufB (COG0050) in many bacteria, such as members of the phyla *Firmicutes*, *Actinobacteria*, and *Dictyoglomi* as shown in Fig. 7a. Although similar gene arrangements are found in most of the species of the phylum *Chloroflexi*, *S. thermophilus* and *Thermomicrobi um roseum* possess their own unique arrangement; Tggt is not found between Ygta and the tufB gene but is located upstream of another threonine-tRNA gene with the anticodon sequence CGT (Tcgt). These observations
suggest a transfer of Tggt in a common ancestor of S. thermophilus and Thermococcus roseum exclusive of the other species of the phylum ‘Chloroflexi’.

DISCUSSION

In this work the phylogenetic placement of ‘Thermobaculum terrenum’ was examined on the basis of gene order comparisons of completely sequenced bacterial genomes. Compared with the more common approach based on sequence alignment, an advantage underlying the present approach is in that gene transfers that occurred in a common ancestor of a group of bacteria can be identified and gene orders present uniquely in that group can be specified. If such gene-transfers are found, then they define and circumscribe that group. This is not always possible by drawing phylogenetic trees based on sequence alignment. Along this line, three such transfers were identified that are thought to have occurred in a common ancestor of species of the phylum ‘Chloroflexi’ (see Figs 1 and 2). It is to be noted here that the present method is valid when a foreign gene is introduced into the genome of a common ancestor of the species of the phylum ‘Chloroflexi’ by means of lateral gene transfer. Gene arrangements created by means of lateral transfer can also serve as genomic markers if they are present uniquely in a particular group of species.

A summary of the species branching pattern within the phylum ‘Chloroflexi’ obtained in this study based on the gene order comparisons is depicted in Fig. 8. The gene transfer events inferred are mapped onto the branching orders. The gene arrangements shown in Figs 1–7 are all consistently understood on the basis of the occurrence orders of the gene transfers and species branching. It is interesting to compare the branching orders deduced here with those obtained by the use of sequence alignments. Although bootstrap statistical supports were marginal, phylogenetic trees based on 16S rRNA sequences show that the species of the genus ‘Dehalococcoides’ (designated subclass 2) branched off first and then a split between the HRC clade (subclass 3) and the S. thermophilus/Thermococcus roseum group (subclass 3) occurred (Hugenholtz et al., 1998; Hugenholtz & Stackebrandt, 2004), which is in agreement with the branching orders illustrated in Fig. 8. Although sequences of ‘Thermobaculum terrenum’ were not included in the sequence analysis, recently reported phylogenomic trees based on a concatenated alignment (Wu & Eisen, 2008; Wu et al., 2009), show branching orders essentially identical to those inferred here. Thus, the present results provide another line of support for the branching orders that are weakly supported by the concatenation of sequence alignments.

The present analysis of gene orders revealed that ‘Thermobaculum terrenum’ is most closely related to S. thermophilus and Thermococcus roseum. This phylogenetic affinity of ‘Thermobaculum terrenum’ is also supported by a BLAST best-hit analysis. Since genes involved
in transcription or translation are thought to be less susceptible to lateral gene transfer (Jain et al., 1999), we selected 51 ribosomal protein sequences from 'Thermobaculum terrenum' and carried out BLAST searches against the KEGG Genes Database, which contains amino acid sequences translated from completely sequenced genomes (http://www.genome.jp/tools/blast/). Among the 51 sequences, 35 proteins showed the highest similarity scores...
with those from the members of the phylum ‘Chloroflexi’, and 12 sequences from members of the phylum ‘Firmicutes’ gained the highest scores. Among the 35 ‘Chloroflexi’ proteins, S. thermophilus showed the largest number of the best hits; 13 best-hit sequences were from S. thermophilus followed by seven sequences from Thermomicrobium proteus, S. thermophilus Chloroflexi and 12 sequences from members of the phylum ‘1952 International Journal of Systematic and Evolutionary Microbiology


Thermobaculum terrenum


ACKNOWLEDGEMENTS

In this work many unpublished genomic sequence data were used. The author would like to thank those who produced the genomic sequence data.

REFERENCES


Fig. 8. A summary diagram showing the relationships between gene transfers and species divergence. Note that a total of 12 gene transfers indicated are consistent with gene arrangements shown in Figs 1–7.

