Genomics 98 (2011) 370–375

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae

Hiromi Nishida [a,*], Teruhiko Beppu [b], Kenji Ueda [b]

[a] Agricultural Bioinformatics Research Unit, Graduate School of Agricultural and Life Sciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[b] Life Science Research Center, College of Bioresource Sciences, Nihon University, Fujisawa, Japan

ARTICLE INFO

Article history:
Received 2 June 2011
Accepted 1 August 2011
Available online 7 August 2011

Keywords:
Whole-genome comparison
Dictyoglomus
Bacterial systematics
Coprothermobacter proteolyticus
Thermotogae

ABSTRACT

The anaerobic thermophilic bacterial genus *Dictyoglomus* is characterized by the ability to produce useful enzymes such as amylase, mannanase, and xylanase. Despite the significance, the phylogenetic position of *Dictyoglomus* has not yet been clarified, since it exhibits ambiguous phylogenetic positions in a single gene sequence comparison-based analysis. The number of substitutions at the diverging point of *Dictyoglomus* is insufficient to show the relationships in a single gene comparison-based analysis. Hence, we studied its evolutionary trait based on whole-genome comparison. Both gene content and orthologous protein sequence comparisons indicated that *Dictyoglomus* is most closely related to the phylum Thermotogae and it forms a monophyletic group with *Coprothermobacter proteolyticus* (a constituent of the phylum Firmicutes) and Thermotogae. Our findings indicate that *C. proteolyticus* does not belong to the phylum Firmicutes and that the phylum Dictyoglomi is not closely related to either the phylum Firmicutes or Synergistetes but to the phylum Thermotogae.

1. Introduction

The domain Bacteria has 26 phyla in the present bacterial systematics [1]. The phylum Dictyoglomi, one of the 26 phyla, consists of the single genus *Dictyoglomus*. This genus was established based on the identification of *Dictyoglomus thermophilum*, which was isolated because of its ability to produce thermostable amylase in a screening study [2]. Currently, the genus *Dictyoglomus* consists of 2 species *D. thermophilum* and *D. turgidum*. Evidence has indicated that this group of bacteria is characterized by unique cell morphology and the ability to produce various enzymes capable of degrading biopolymers [2–6].

The phylum Dictyoglomi has been recognized as a constituent of the anaerobic Gram-positive bacteria, but its exact phylogenetic position is unclear. The 16S rRNA gene comparison-based phylogenetic analysis by Love et al. [7] showed that *D. thermophilum* is related to the phylum Thermotogae. Meanwhile, Rees et al. [8] indicated that it is closely related to *Anaerobaculum* (belonging to the phylum Synergistetes), and Takai et al. [9] showed that it is clustered with *Thermoanaerobacter* and *Thermoanaerobacterium* (belonging to the phylum Firmicutes). Recently, Wagner and Wiegel [10] showed phylogenetic relationships among 127 thermophilic anaerobes on the basis of 16S rRNA sequence comparison. In that phylogenetic tree, *Dictyoglomus* clusters not with Thermotogae but Chloroflexi. Furthermore, different phylogenetic relationships from other gene comparison-based analyses have provided contrasting information. For example, Gibbs et al. [4] showed that xylanase of *D. thermophilum* has a similar structure to that of the clostridial bacterium *Caldicellulosiruptor saccharolyticus*. In a phylogenetic tree based on PPi-dependent phosphofructokinase comparison, *D. thermophilum* did not cluster with *Thermotoga martima* [11]. Besides, in a phylogenetic tree based on reverse gyrase comparison, *Thermoanaerobacter* and *Caldicellulosiruptor*, but not *Dictyoglomus*, clustered with *Thermotoga* [12].

To date, whole-genome sequencing has been carried out on more than 1500 bacterial organisms, including *D. thermophilum* (http://www.ncbi.nlm.nih.gov/genomeprj/59439) and *D. turgidum* (http://www.ncbi.nlm.nih.gov/genomeprj/59177) by the J. Craig Venter Institute (http://www.jcvi.org/) and the DOE Joint Genome Institute (http://www.jgi.doe.gov/) respectively. Based on this information, we carried out whole-genome comparison analyses and clarified the phylogenetic relationships between *Dictyoglomus* and other bacteria. Whole-genome comparison is a powerful tool that enables more accurate phylogenetic evaluation than that based on a single gene sequence comparison [13–15]. We applied the 2 major strategies of whole-genome comparison studies, i.e. gene content comparison-based analysis and orthologous protein sequence comparison-based analysis [16].

2. Methods

To determine the phylogenetic relationships between *Dictyoglomus* and representative bacteria, we compared the gene contents from the

* Corresponding author.
E-mail address: hnishida@iu.a.u-tokyo.ac.jp (H. Nishida).

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.08.001

Table 1
List of bacterial names used in this study.

Acidobacterium capsulatum
Agrobacterium tumefaciens
Alteromonas macleodii
Aminobacterium colombiense
Anaerococcus prevotii
Aquifex aeolicus
Azorhizobium caulinodans
Bacillus subtilis
Bacteroides thetaiotaomicron
Borrelia burgdorferi B31
Brachyspira hyodysenteriae
Caldicellulosiruptor bescii
Caldicellulosiruptor saccharolyticus
Campylobacter jejuni NCTC 11168
Candidatus Phytoplasma asteris
Carboxydothermus hydrogenoformans
Chlamydia trachomatis D/UW-3/CX
Chlorobium chlorochromatii
Chloroflexus aurantiacus
Clostridium acetobutylicum
Coprothermobacter proteolyticus
Coxiella burnetii RSA 493
Cyanothece sp. ATCC 51142
Deferribacter desulfuricans
Dehalococcoides ethenogenes
Deinococcus radiodurans
Denitrovibrio acetiphilus
Desulfatibacillum alkenivorans
Desulfitobacterium hafniense DCB-2
Desulfohalobium retbaense
Dictyoglomus thermophilum
Dictyoglomus turgidum
Elusimicrobium minutum
Escherichia coli K-12 MG1655
Eubacterium eligens
Fervidobacterium nodosum
Fibrobacter succinogenes
Fusobacterium nucleatum
Gemmatimonas aurantiaca
Geobacter sulfurreducens
Gloeobacter violaceus
Haemophilus influenzae Rd KW20
Halothermothrix orenii
Heliobacterium modesticaldum
Helicobacter pylori 26695
Kosmotoga olearia
Leptospira interrogans serovar lai 56601
Leptotrichia buccalis
Mesoplasma florum
Moorella thermoacetica
Mycoplasma genitalium
Myxococcus xanthus
Natranaerobius thermophilus
Neisseria meningitidis MC58
Nitrosomonas europaea ATCC 19718
Nostoc sp. PCC 7120
Opitutus terrae
Pelotomaculum thermopropionicum
Persephonella marina
Petrotoga mobilis
Pirellula staleyi
Prochlorococcus marinus CCMP1375
Pseudomonas aeruginosa PO1
Ralstonia solanacearum
Rhodobacter sphaeroides 2.4.1
Rhodospirillum rubrum
Rickettsia prowazekii
Sebaldella termitidis
Shewanella oneidensis
Sphingomonas wittichii
Streptobacillus moniliformis
Streptomyces griseus
Symbiobacterium thermophilum
Syntrophobacter fumaroxidans
Syntrophomonas wolfei
Thermanaerovibrio acidaminovorans
Thermoanaerobacter tengcongensis
Thermobaculum terrenum
Thermodesulfovibrio yellowstonii
Thermomicrobium roseum
Thermomonospora curvata
Thermosipho melanesiensis
Thermotoga maritima
Thermus thermophilus
Trichodesmium erythraeum
Ureaplasma parvum ATCC 700970
Veillonella parvula
Vibrio cholerae N16961
Xylella fastidiosa 9a5c

89 representative bacteria in this analysis (Table 1). Ortholog cluster analysis among the above 89 bacteria was performed using the Microbial Genome Database for Comparative Analysis (MBGD; http://mbgd.nibb.ac.jp/) [17]. The analysis (minimum cluster size 2) provided a gene presence/absence data matrix (14,732 gene groups × 89 organisms) that served as the basis for a distance matrix between all pairs of the 89 organisms. The distance was calculated from the different ratios between the presence/absence patterns of the 14,732 genes. On the basis of the distance matrix, a neighbor-joining tree was reconstructed using MEGA software version 5 [18]. The bootstrap was performed with 1000 replicates.

We compared orthologous protein sequences from the above 89 bacteria. From the gene presence/absence data matrix, 44 proteins (ArgS, Frr, Gcp, HisS, InfB, LepA, NusA, ObgE, PheS, PheT, PrfA, RplA, RplB, RplC, RplD, RplE, RplF, RplK, RplL, RplN, RplO, RplP, RplQ, RplR, RplT, RplU, RplV, RplX, RpmA, RpsB, RpsC, RpsE, RpsH, RpsJ, RpsK, RpsM, RpsS, SecA, SecY, SmpB, Tsf, UvrB, UvrC, and YbeY) were extracted as orthologous proteins. Thus, we constructed 44 multiple alignments using MUSCLE [19]. Then, a concatenated multiple alignment of the 44 multiple alignments was generated. The phylogenetic analysis was performed on the basis of 8869 amino acid sites without the gap/insertion sites. A neighbor-joining tree was reconstructed using MEGA software version 5 [18]. The bootstrap was performed with 1000 replicates. The rate variation among sites was considered to have a γ-distributed rate ($\alpha = 1$). The other default parameters (e.g., Poisson distance) were not changed. In addition, A maximum likelihood tree was reconstructed using MEGA version 5 [18]. The JTT model was used as the amino acid substitution model. The nearest neighbor interchange was used as the maximum likelihood heuristic method. The γ-distributed rate was considered and number of discrete gamma categories was three.

3. Results and discussion

In the gene content comparison-based phylogenetic tree, the phylum Firmicutes did not form a monophyletic lineage (Fig. 1). A marked discrepancy was observed with regard to the position of mycoplasmas that clustered to *Rickettsia* (belonging to the subphylum Alphaproteobacteria), *Chlamydia* (belonging to the phylum Chlamydiae), and *Borrelia* (belonging to the phylum Spirochetes) with 100% bootstrap support. This is rather consistent with the observation by Wolf et al. [16], indicating that parasitic Alphaproteobacteria, parasitic Gammaproteobacteria, Chlamydiae, Spirochetes, and mycoplasmas are clustered because of the considerable numbers of genes lost during evolution. Another inconsistency with respect to Firmicutes was the close relation of *Anaerococcus*, *Eubacterium*, and *Veillonella* to the phylum Fusobacteria (Fig. 1). In addition, *Coprothermobacter proteolyticus*[20] was also excluded from the large cluster of Firmicutes. Although these inconsistencies were observed with respect to Firmicutes, the analysis showed distinct clustering of the phyla Dictyoglomi with Synergistetes and Thermotogae with 98% bootstrap support (Fig. 1).

In the course of evolution, the gene content of the bacterial genome has changed via gene acquisition and loss, along with adaptation to each environment [15]. Probably, the above inconsistency with regard to

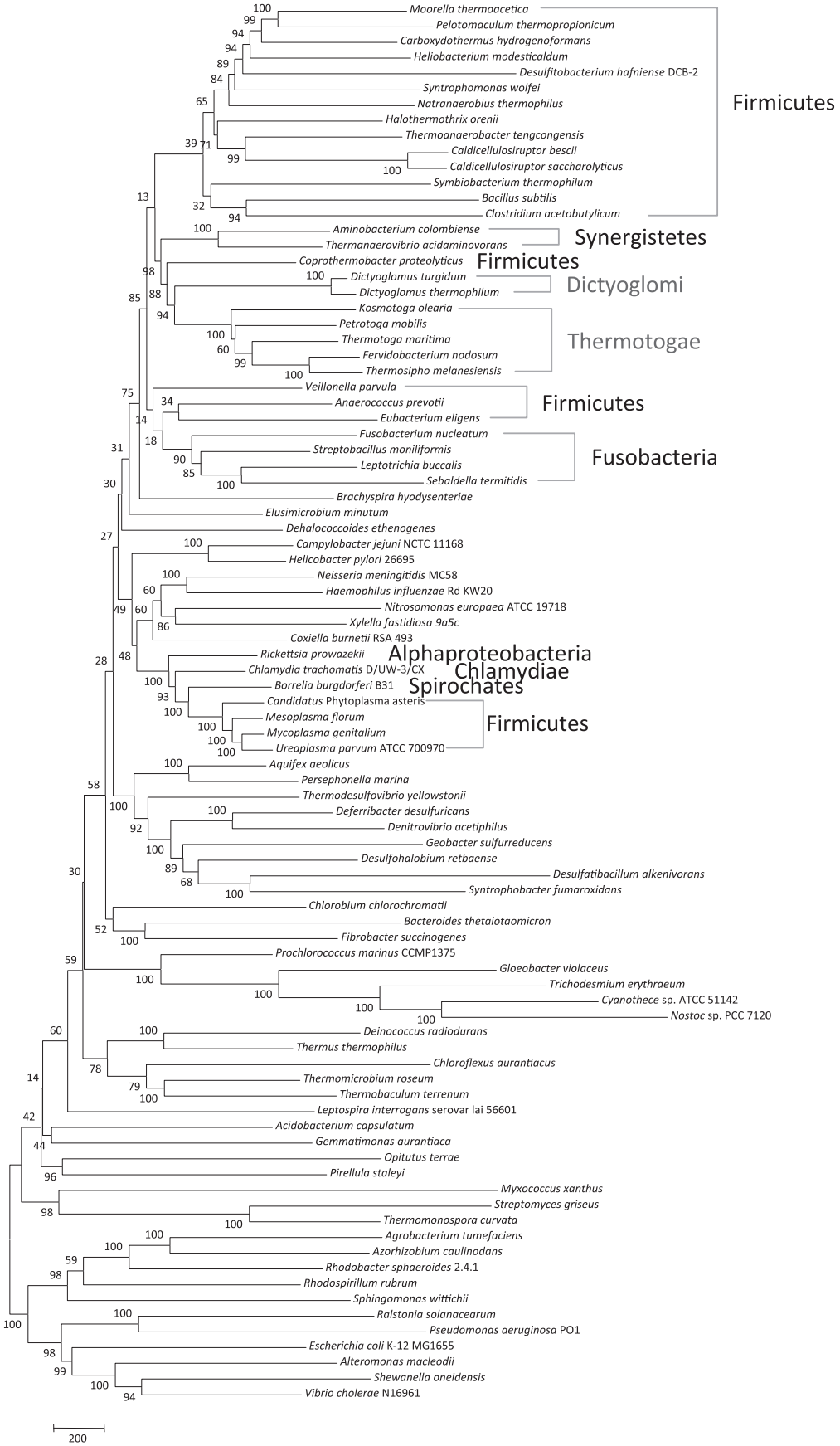

Fig. 1. Phylogenetic relationships on the basis of gene content comparisons among 89 bacteria. Ortholog cluster analysis (minimum cluster size 2) among the 89 bacteria was performed using the MBGD [17]. This analysis produced a gene presence/absence data matrix (14,732 gene groups × 89 organisms), which was used to generate a distance matrix between all pairs of the 89 bacteria. Based on the distance matrix, a neighbor-joining tree was reconstructed using MEGA software version 5 [18]. The bootstrap was performed with 1000 replicates. The bar indicates a 200-gene difference.

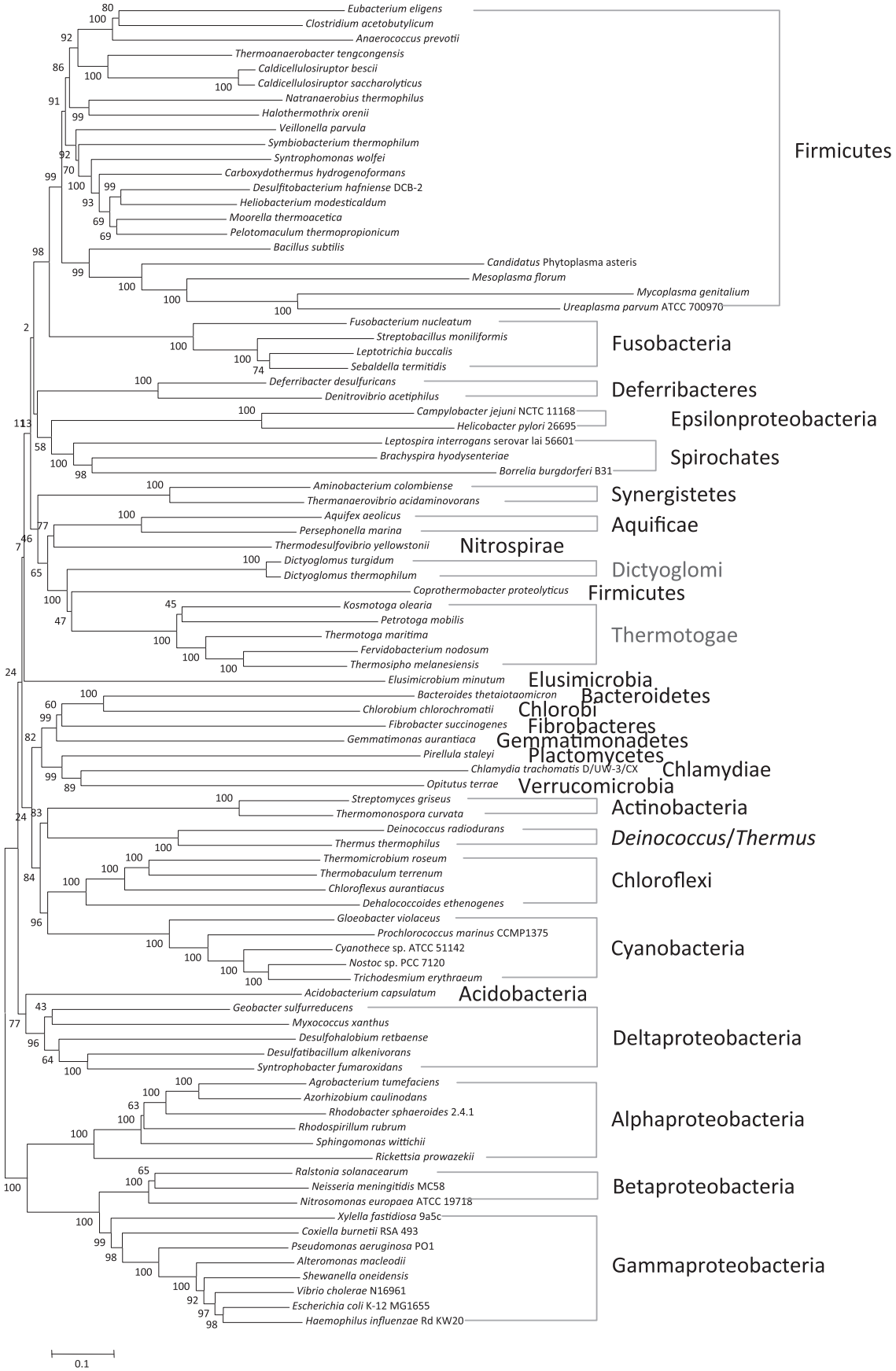

Fig. 2. Neighbor-joining tree on the basis of 44 orthologous protein sequence comparisons among 89 bacteria. We constructed 44 multiple alignments using MUSCLE [19]. Then, a concatenated multiple alignment of the 44 multiple alignments was generated. The phylogenetic analyses were performed on the basis of 8869 amino acid sites without the gap/insertion sites. The neighbor-joining tree was reconstructed using MEGA software version 5 [18]. The bootstrap was performed with 1000 replicates. The rate variation among sites was considered to have a γ-distributed rate ($\alpha = 1$). The other default parameters (e.g. Poisson distance) were not changed. The bar indicates a 10% difference.

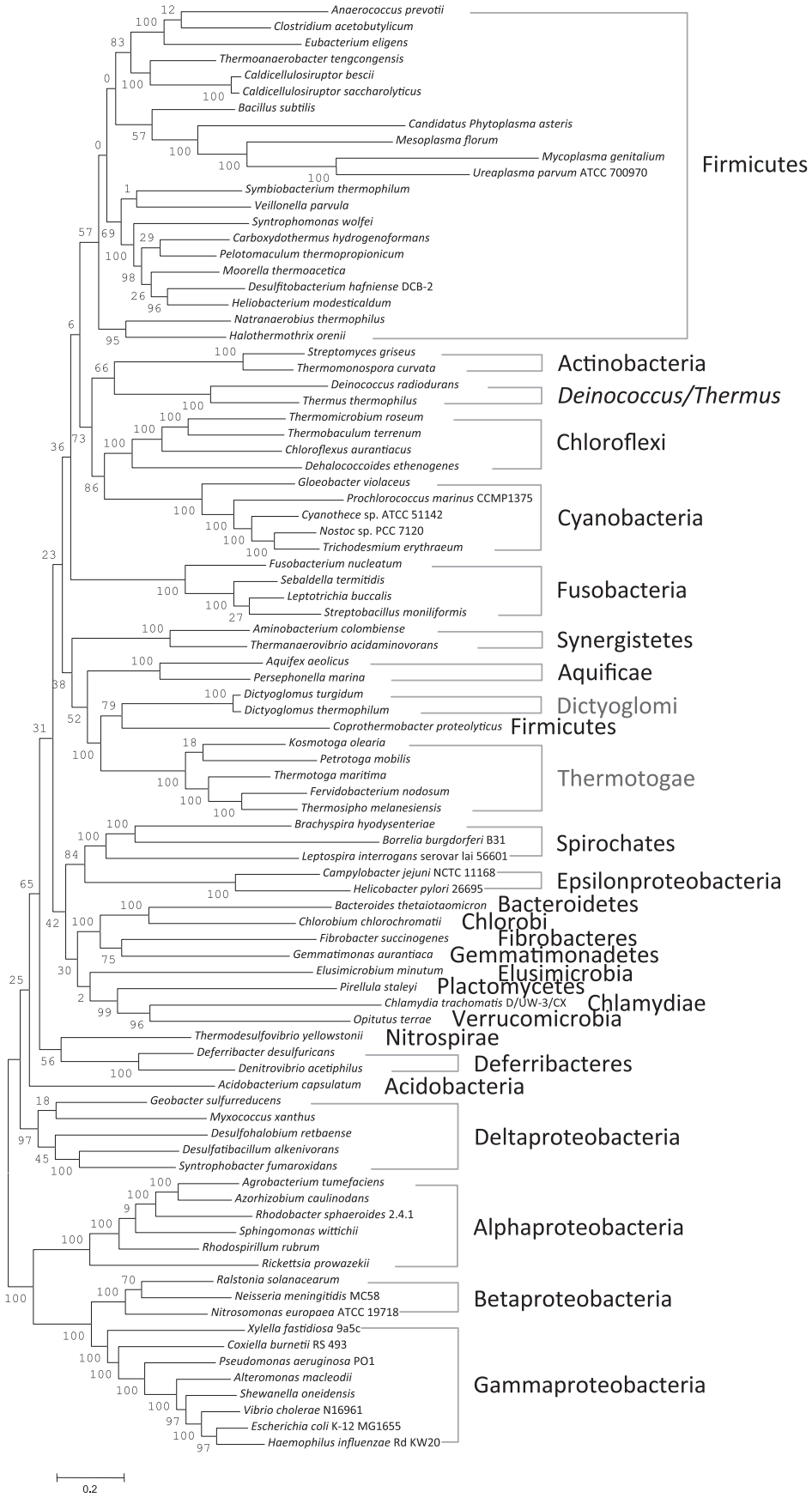

Fig. 3. Maximum likelihood tree on the basis of 44 orthologous protein sequence comparisons among 89 bacteria. MEGA version 5 [18] was used. The JTT model was used as the amino acid substitution model. The ML heuristic method is the nearest neighbor interchange. The bootstrap was performed with 100 replicates. The γ-distributed rate was considered and number of discrete gamma categories was three. As the result, the gamma was 1.5828; the discrete rates were 0.2978 (1.7%), 0.8115 (98.3%), and 1.8907 (0%).

Firmicutes reflects the specific adaptation history of some group of this phylum; gene content-based analysis accounts for such genome reconstitution in a phylogenetic evaluation. In contrast to this characteristic of gene content-based analysis, the orthologous protein sequence comparison-based analysis is expected to be less influenced by such genetic alterations and to show principal evolutionary relationships. Hence, we also carried out this analysis on the same set of bacterial genomes.

In the orthologous protein sequence comparison-based phylogenetic trees, well-defined major bacterial groups, including Firmicutes, formed each monophyletic lineage (Figs. 2 and 3). The integration of the phylum Firmicutes into a unique lineage supports the aforementioned view that the results of orthologous protein sequence-based analyses represent primary evolutionary relationships, compared to the results obtained by gene content-based analyses. The orthologous protein sequence-based analysis also demonstrated the close relationships between the phyla Dictyoglomi and Thermotogae; hence, we reasonably conclude that the phylum Dictyoglomi is evolutionary correlated with Thermotogae and diverted from Firmicutes and Synergistetes. In the 16S rRNA sequence comparison-based phylogenetic tree, the phylogenetic relationships at the early stage of bacterial evolution are uncertain due to very low bootstrap support (Additional file 1). Thus, the 16S rRNA sequence comparison is not useful to show the phylogenetic relationships among the phyla Dictyoglomi, Firmicutes, Synergistetes, and Thermotogae.

An inconsistency found in the above orthologous protein comparison-based phylogenetic trees (Figs. 2 and 3) was the position of *C. proteolyticus*. In the present bacterial taxonomic system, the genus *Coprothermobacter* is classified into the phylum Firmicutes. In the 16S rRNA gene comparison-based analysis, it is a constituent of Firmicutes [10]. However, the three phylogenetic trees drawn in this study demonstrated that it clusters not to Firmicutes but to Dictyoglomi and Thermotogae (Figs. 1, 2, and 3), indicating that *C. proteolyticus* is not a constituent of Firmicutes but represents another taxonomic group mostly closely related to Dictyoglomi. Further, the deep branching point of the cluster containing *Coprothermobacter*, Dictyoglomi, and Thermotogae has diverged from other lineages at an early stage of evolution.

It is noteworthy that monophyletic lineage formation by the above 3 taxa (*Coprothermobacter*, Dictyoglomi, and Thermotogae) was observed in only 4 individual neighbor-joining trees with weak bootstrap supports (LepA, RplE, RplV, and SecA; Additional file 2) out of the 44 trees of orthologous proteins used in the analysis. Most orthologous proteins do not have enough resolution to show the correct phylogenetic relationships among the three taxa. Thus, multiple substitutions have occurred at the same site during the evolution. These results support that *Coprothermobacter*, Dictyoglomi, and Thermotogae diverged from a common ancestor at an early stage of bacterial evolution. We suggest that the exact evolutionary relationships of the three taxonomic groups cannot be clearly depicted by comparing a single gene and that whole-genome comparison analyses are useful for this purpose.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2011.08.001.

References

[1] A. Oren, The Phyla of Prokaryotes—Cultured and Uncultured, in: A. Oren, R.T. Papke (Eds.), Molecular Phylogeny of Microorganisms, Caister Academic Press, Norfolk, 2010, pp. 85–108.

[2] T. Saiki, Y. Kobayashi, K. Kawagoe, T. Beppu, *Dictyoglomus thermophilum* gen. nov., sp. nov., a chemoorganotrophic, anaerobic, thermophilic bacterium, Int. J. Syst. Bacteriol. 35 (1985) 253–259.

[3] S. Fukusumi, A. Kamizono, S. Horinouchi, T. Beppu, Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic thermophile, *Dictyoglomus thermophilum*, Eur. J. Biochem. 174 (1988) 15–21.

[4] M.D. Gibbs, R.A. Reeves, P.L. Bergquist, Cloning, sequencing, and expression of a xylanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46B.1 and activity of the enzyme on fiber-bound substrate, Appl. Environ. Microbiol. 61 (1995) 4403–4408.

[5] M.D. Gibbs, R.A. Reeves, A. Sunna, P.L. Bergquist, Sequencing and expression of a β-mannanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46B.1, and characteristics of the recombinant enzyme, Curr. Microbiol. 39 (1999) 351–357.

[6] S. Horinouchi, S. Fukusumi, T. Ohshima, T. Beppu, Cloning and expression in *Escherichia coli* of two additional amylase genes of a strictly anaerobic thermophile, *Dictyoglomus thermophilum*, and their nucleotide sequences with extremely low guanine-plus-cytosine contents, Eur. J. Biochem. 176 (1988) 243–253.

[7] C.A. Love, B.K.C. Patel, W. Ludwig, E. Stackebrandt, The phylogentic position of *Dictyoglomus thermophilum* based on 16S rRNA sequence analysis, FEMS Microbiol. Lett. 107 (1993) 317–320.

[8] G.N. Rees, B.K.C. Patel, G.S. Grassia, A.J. Sheehy, *Anaerobaculum thermoterrenum* gen. nov., sp. nov., a novel, thermophilic bacterium which ferments citrate, Int. J. Syst. Bacteriol. 47 (1997) 150–154.

[9] K. Takai, A. Inoue, K. Horikoshi, *Thermaerobacter marianesis* gen. nov., sp. nov., an aerobic extremely thermophilic marine bacterium from the 11000 m deep Mariana Trench, Int. J. Syst. Bacteriol. 49 (1999) 619–628.

[10] I.D. Wagner, J. Wiegel, Diversity of thermophilic anaerobes, Ann. N.Y. Acad. Sci. 1125 (2008) 1–43.

[11] Y.-H. Ding, R.S. Ronimus, H.W. Morgan, Sequencing, cloning, and high-level expression of the *pfp* gene, encoding a PPi-dependent phosphofructokinase from the extremely thermophilic eubacterium *Dictyoglomus thermophilum*, J. Bacteriol. 182 (2000) 4661–4666.

[12] C. Brochier-Armanet, P. Forterre, Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers, Archaea 2 (2007) 83–93.

[13] M.T. Alam, M.E. Merlo, E. Takano, R. Breitling, Genome-based phylogenetic analysis of *Streptomyces* and its relatives, Mol. Phylogenet. Evol. 54 (2010) 763–772.

[14] K. Oshima, H. Nishida, Phylogenetic relationships among mycoplasmas based on the whole genomic information, J. Mol. Evol. 65 (2007) 249–258.

[15] K. Oshima, K. Ueda, T. Beppu, H. Nishida, Unique evolution of *Symbiobacterium thermophilum* suggested from gene content and orthologous protein sequence comparisons, Int. J. Evol. Biol. 2011 (2011) 376831.

[16] Y.I. Wolf, I.B. Rogozin, N.V. Grishin, E.V. Koonin, Genome trees and the tree of life, Trends Genet. 18 (2002) 472–479.

[17] I. Uchiyama, T. Higuchi, M. Kawai, MBGD update 2010: toward a comprehensive resource for exploring microbial genome density, Nucleic Acids Res. 38 (2010) D361–D365.

[18] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. (in press), doi: 10.1093/molbev/msr121.

[19] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.

[20] B.M. Ollivier, R.A. Mah, T.J. Ferguson, D.R. Boone, J.L. Garcia, R. Robinson, Emendation of the genus *Thermobacteroides*: *Thermobacteroides proteolyticus* sp. nov., a proteolytic acetogen from a methanogenic enrichment, Int. J. Syst. Bacteriol. 35 (1985) 425–428.