Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press



The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota

Tomoyuki Kosaka, Souichiro Kato, Takefumi Shimoyama, et al.

Genome Res. published online January 24, 2008
Access the most recent version at doi:10.1101/gr.7136508

P<P Published online January 24, 2008 in advance of the print journal.

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © 2008, Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Letter

The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota

Tomoyuki Kosaka,[1,2,5] Souichiro Kato,[1] Takefumi Shimoyama,[1] Shunichi Ishii,[1,5] Takashi Abe,[3] and Kazuya Watanabe[1,4,6]

[1]*Laboratory of Applied Microbiology, Marine Biotechnology Institute, Kamaishi, Iwate 026-0001, Japan;* [2]*PRESTO, JST, Kawaguchi, Saitama 332-0012, Japan;* [3]*Department of Bioscience, Nagahama Institute of Bio-science and Technology, Nagahama, Shiga 526-0829, Japan;* [4]*Hashimoto Light Energy Conversion Project, ERATO, JST, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

The anaerobic biodegradation of organic matter is accomplished by sequential syntrophic catabolism by microbes in different niches. *Pelotomaculum thermopropionicum* is a representative syntrophic bacterium that catalyzes the intermediate bottleneck step in the anaerobic-biodegradation process, whereby volatile fatty acids (VFAs) and alcohols produced by upstream fermenting bacteria are converted to acetate, hydrogen, and carbon dioxide (substrates for downstream methanogenic archaea). To reveal genomic features that contribute to our understanding of the ecological niche and evolution of *P. thermopropionicum*, we sequenced its 3,025,375-bp genome and performed comparative analyses with genomes of other community members available in the databases. In the genome, 2920 coding sequences (CDSs) were identified. These CDSs showed a distinct distribution pattern in the functional categories of the Clusters of Orthologous Groups database, which is considered to reflect the niche of this organism. *P. thermopropionicum* has simple catabolic pathways, in which the propionate-oxidizing methylmalonyl–CoA pathway constitutes the backbone and is linked to several peripheral pathways. Genes for most of the important catabolic enzymes are physically linked to those for PAS-domain-containing regulators, suggesting that the catabolic pathways are regulated in response to environmental conditions and/or global cellular situations rather than specific substrates. Comparative analyses of codon usages revealed close evolutionary relationships between *P. thermopropionicum* and other niche members, while it was distant from phylogenetically related sugar-fermenting bacteria. These analyses suggest that *P. thermopropionicum* has evolved as a syntrophy specialist by interacting with niche-associated microbes.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to GenBank under accession no. AP009389.]

The anaerobic methanogenic biodegradation of organic matter is an important process in the global carbon cycle and is widely distributed among different habitats (e.g., soil, sediment, rumen, and digesters). Extensive studies on the anaerobic microbiota in these habitats have revealed that this process is accomplished by sequential catabolic reactions (syntrophic interactions) of microbes in different niches (Fig. 1) (Schink 1997). In addition, the genomes of some representative microbes have recently been sequenced (Fig. 1), providing us with opportunities to reveal their hidden ecological roles (McInerney et al. 2007; Stolyar et al. 2007). The anaerobic microbiota is thus considered to be a model for studying evolutionary mechanisms for niche speciation in prokaryotic communities.

Pelotomaculum thermopropionicum is a Gram-positive thermophilic bacterium that ferments fatty acids and alcohols (e.g., propionate, butyrate, lactate, propanol, and ethanol), metabolic products of upstream fermenting bacteria (FB) under syntrophic associations with methanogenic archaea (MA) (Fig. 1) (Imachi et al. 2002). Metabolites of *P. thermopropionicum* (acetate, hydrogen, formate, and carbon dioxide) serve as growth substrates for MA, and efficient consumption of these metabolites (particularly hydrogen and formate) by MA is indispensable for *P. thermopropionicum* to make fermentation thermodynamically feasible (Schink 1997). Owing to this lifestyle, bacteria that share niches similar to *P. thermopropionicum* are called syntrophic bacteria (SB) (Fig. 1). Since SB catalyze the intermediate bottleneck step in the anaerobic biodegradation process (Kaspar and Wuhrmann 1978), their activities and physiology are of practical interest. Another area of interest is how SB thrive by utilizing the limited energy conserved by syntrophic fatty-acid oxidation (Thauer et al. 1977); for instance, the Gibbs free energy change in syntrophic propionate oxidation is approximately −25 kJ/mol, which is less than the energy needed for synthesizing 1 mol of ATP (Jackson and McInerney 2002). Due to the difficulties associated with laboratory axenic cultivation, however, information regarding SB is relatively scarce. Complete genome sequences of several FB and MA have already been published, while the complete genome of one SB, *Syntrophus aciditrophicus*, has only recently been reported for the first time (McInerney et al. 2007). *P. thermopropionicum* is distinct from *S. aciditrophicus* in phylogeny, available substrates, and growth temperature (Imachi et al. 2002). Furthermore, *P. thermopropionicum* is considered important for anaerobic biodegradation in terms of its ability to oxidize propionate under

[5]Present address: Institute for Biological Resources and Function, Advanced Industrial Science and Technology, Higashi, Tsukuba, Ibaraki 305-8566, Japan.
[6]Corresponding author.
E-mail kazuya.watanabe@mbio.jp; fax 81-193-26-5781.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.7136508.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

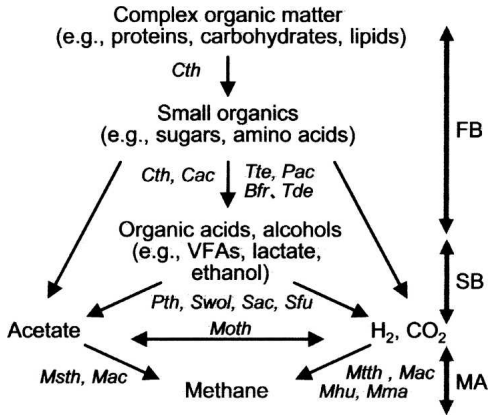


Figure 1. Metabolic interactions in the anaerobic biodegradation process. Italic letters indicate representative organisms in the anaerobic microbiota, whose genomes have been determined; (*Pth*) *Pelotomaculum thermopropionicum*; (*Moth*) *Moorella thermoacetica*; (*Tte*) *Thermoanaerobacter tengcongensis*; (*Swol*) *Syntrophomonas wolfei*; (*Cth*) *Clostridium thermocellum*; (*Cac*) *Clostridium acetobutylicum*; (*Sac*) *Syntrophus aciditrophicus*; (*Sfu*) *Syntrophobacter fumaroxidans*; (*Pac*) *Propionibacterium acnes*; (*Bfr*) *Bacteroides fragilis*; (*Tde*) *Treponema denticola*; (*Msth*) *Methanosaeta thermophila*; (*Mtth*) *Methanothermobacter thermautotrophicus*; (*Mhu*) *Methanospirillum hungatei*; (*Mma*) *Methanococcus maripaludis*; (*Mac*) *Methanosarcina acetivorans*. For more information about these organisms, refer to Supplemental Table S1.

methanogenic conditions, because a large fraction (>30%) of methane is produced from complex organic matter via propionate (Glissmann and Conrad 2000).

Here we report the complete genome sequence of *P. thermopropionicum*. We focused our analyses to reveal genomic features that shall further our understanding of the niche and evolution of this organism. A particular focus was made to reveal evolutionary interactions of this organism with other members in the anaerobic microbiota. As mentioned above, the genome sequences of microbes sharing different niches in the anaerobic microbiota are currently available, which facilitated comparative genomic analyses in the context of ecological interactions (see Fig. 1 and Supplemental Table S1). We show evidence for the active interaction of *P. thermopropionicum* with other niche members during its evolution into a syntrophy specialist.

Results and Discussion

General genome features

The genome of *P. thermopropionicum* consists of a circular chromosome of 3,025,375 bp with an average G+C content of 53.0%. Primary characteristics of the genome are summarized in Supplemental Table S2, while its organization is shown in Supplemental Figure S1. The start codon of *dnaA* at a GC-skew transition point was defined as the starting point of the genome (Supplemental Fig. S1). In silico analysis predicted 2920 coding sequences (CDS) that covered 86.3% of the genome. Among them, biological functions can be assigned to 1637 CDSs (56.1%) based on homologies to known proteins or protein domains/motifs, 397 (13.6%) are homologous to conserved proteins whose functions are poorly characterized, and 886 (30.3%) have no homologous counterparts in the databases. These ratios are similar to those reported for other bacteria (Supplemental Table S3).

The *P. thermopropionicum* genome displays two clear GC-skew transitions that likely correspond to the DNA replication origin and terminus (Supplemental Fig. S1). Similar to other *Firmicutes* bacteria (Bao et al. 2002; Rocha 2004; Wu et al. 2005), a large portion of the predicted CDSs (78.6%) are located on the leading strand (Supplemental Fig. S1). However, this bacterium does not possess PolC, a DNA polymerase used to synthesize leading strands in *Bacillus subtilis* (Rocha 2004), and DnaE is therefore thought to replicate both strands in *P. thermopropionicum*.

Functional categories of CDSs

The distribution patterns of the predicted CDSs to the functional categories of the Clusters of Orthologous Groups (COG) database (Tatusov et al. 1997) were compared between *P. thermopropionicum* and other organisms listed in Supplemental Table S1 (Supplemental Table S3). It has been reported that large genomes are disproportionately enriched in regulation and secondary metabolism genes and depleted in protein translation, DNA replication, cell division, and nucleotide metabolism genes compared to medium and small-sized genomes (Konstantinidis and Tiedje 2004). Furthermore, archaeal genomes are known to encode significantly more genes for energy production, coenzyme metabolism, and the poorly characterized category than bacterial genomes (Konstantinidis and Tiedje 2004). Based on these findings, it has been suggested that a COG distribution trend could reflect the lifestyles of prokaryotes (Konstantinidis and Tiedje 2004). Our analyses reveal that the COG distribution pattern of *P. thermopropionicum* is in part similar to that of the archaeal average rather than the bacterial average (Supplemental Table S3). For example, *P. thermopropionicum* and the archaeal average are similar in percentages of CDSs categorized into "energy production and conversion," "carbohydrate transport and metabolism," and "coenzyme metabolism." This trend is also observed for other SB (*Sac* and *Swol* in Supplemental Table S3), although *Sfu* is somewhat distinct from the other SB. We consider that this trend may be related to the large genome size of *Sfu* (Supplemental Table S1). It is also important to note that, among these SB, *Sfu* is the only organism that is capable of growing on sulfate as an electron acceptor (Harmsen et al. 1998). Based on these data, we consider that the CDS-distribution biases in SB reflect their niche where they grow on limited types of substrates (mostly organic acids and alcohols), which accounts for the low content of "carbohydrate transport and metabolism" genes, and under energy-limited conditions, which results in a high content of "energy production and conversion" genes.

Ribosome

P. thermopropionicum has 2 rRNA operons and 51 tRNA genes that represent all 20 amino acids. We found that each of the two 23S rRNA genes (3483 bp in length) carried two specific intervening sequences (IVS) with terminal palindrome structures at positions 593–700 (IVS1) and 1328–1754 (IVS2) from the 5′-end, which was revealed by alignments of 23S rRNA genes of microbes appearing in Supplemental Table 1 (see Supplemental Fig. S2). Fragmentation of 23S rRNA into three parts was confirmed by electrophoresis of an RNA extract from a *P. thermopropionicum* culture (Supplemental Fig. S2). In addition to *P. thermopropionicum*, these two IVSs were also found only in *Syntrophomonas wolfei,* another SB affiliated with *Firmicutes* (Supplemental Fig. S2), whose IVS1 and IVS2 are 58% and 57%, respectively, identical to those of *P. thermopropionicum* (Supplemental Fig. S2). A BLAST search revealed that the IVS1 sequences of these organisms were highly

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

homologous to partial sequences of 16S rRNA genes, while no homologous sequence was found for IVS2. These results provide evidence for the close evolutionary relationship between *P. thermopropionicum* and *S. wolfei*, while the origin of IVS2 remains a mystery. It is also not clear why these two bacteria use fragmented 23S rRNA. It has been reported that fragmentation facilitates rapid degradation of rRNA molecules in *Salmonella* under non-growing conditions (Ralph and McClelland 1993). Therefore, we deduce that these SB may efficiently recycle fragmented rRNA molecules under energy-limited conditions.

Metabolism

Among the 1637 function-assigned CDSs of *P. thermopropionicum*, 761 are organized into metabolic pathways, which reveal the potentials of this bacterium for catabolism, energy conservation, and biosynthesis (Fig. 2).

Under syntrophic association with hydrogenotrophic methanogens, *P. thermopropionicum* utilizes propionate, lactate, and several alcohols, such as ethanol, propanol, and butanol (Imachi et al. 2002). In monocultures, it can ferment pyruvate and fumarate, with fumarate also serveing as an electron acceptor in the presence of propionate, ethanol, or lactate (Imachi et al. 2002). The genome analyses revealed that the propionate-oxidizing methylmalonyl–CoA pathway (Kosaka et al. 2006) constitutes the backbone of its catabolism, which is linked to branch pathways for propanol, lactate, and ethanol oxidation and those for the biosynthetic pathways (Fig. 2). Based on the results of metabolite analyses (Imachi et al. 2002), the methylmalonyl–CoA pathway is considered to be reversibly operated depending on the substrates (see Supplemental Fig. S3); for example, 4 mol of pyruvate are converted to 3 mol of acetate and 1 mol of propionate via the methylmalonyl–CoA pathway. Butanol oxidation (catalyzed by alcohol dehydrogenases) is independent from the methylmalonyl–CoA pathway and generates butyrate as the final product. This organism therefore needs another organic compound, such as acetate, as a carbon source when growing on butanol (Supplemental Fig. S3). Relevant transporters for carboxylic acids have also been identified (Fig. 2).

During the oxidation of 1 mol of propionate, reducing equivalents are discharged at three steps (Fig. 2). Among them, succinate dehydrogenase reduces menaquinone (the major quinone of this organism) (Imachi et al. 2002), and malate dehydrogenase reduces NAD, while pyruvate/ferredoxin oxidoreductase reduces ferredoxin. The genome analyses reveal that four hydrogenases and two formate dehydrogenases are present in *P. thermopropionicum* (Supplemental Table S4), among which two hydrogenases and one formate dehydrogenase have twin-arginine translocation-signal sequences (Rodrigue et al. 1999) and are considered to be transported to the periplasmic side of the cytoplasmic membrane. Hydrogenases corresponding to these three electron-discharging steps were predicted (Supplemental Table S4). A previous culture experiment showed that *P. thermopropionicum* did not produce formate for exhausting reducing equivalents (Imachi et al. 2002), while recent data showed

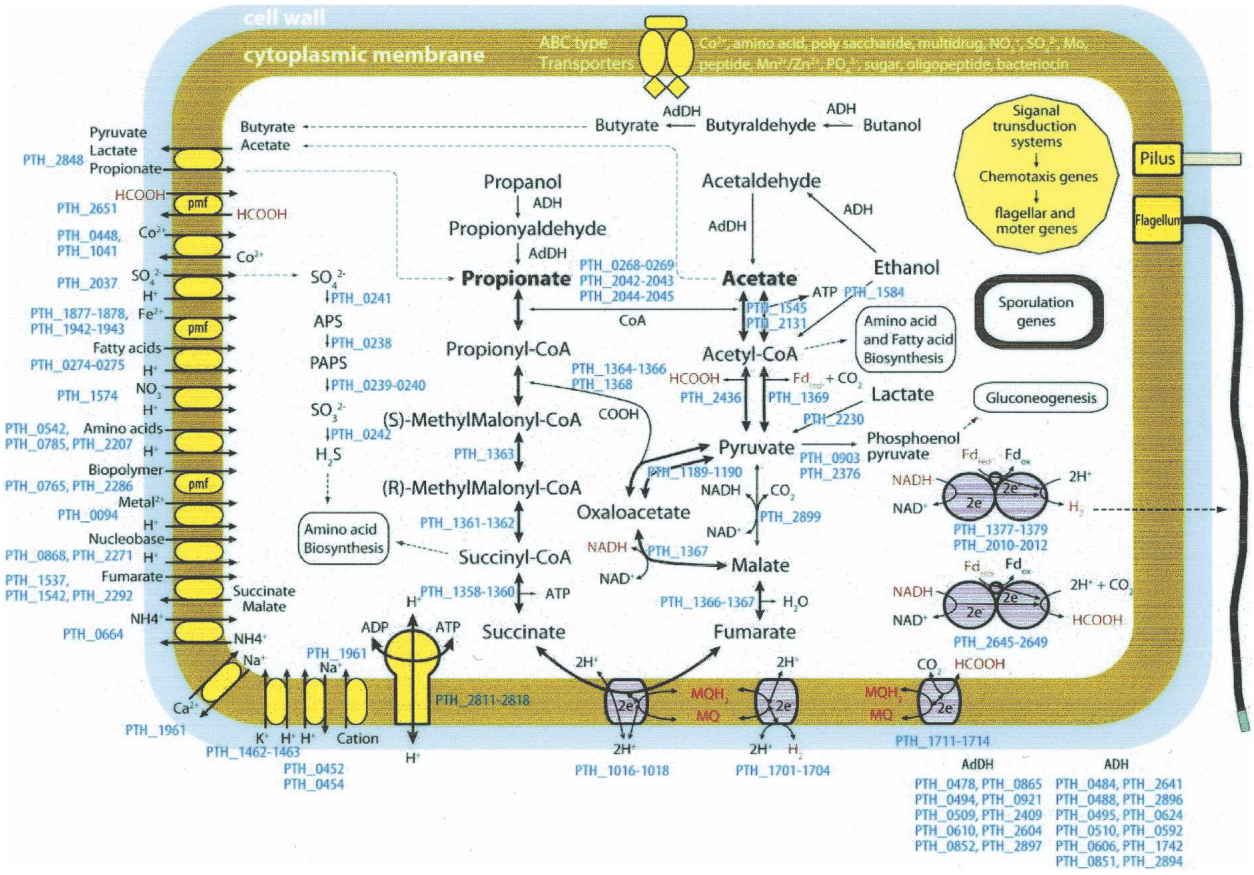


Figure 2. Overview of the metabolism of *P. thermopropionicum*. Substrates of alcohol dehydrogenases (ADH) and aldehyde dehydrogenases (AdDH) are not identified, and they are listed *below* the cell. (MQ) Menaquinone.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

that formate was temporarily produced when it grew on ethanol under syntrophic association with a hydrogenotrophic methanogen (data not shown). We speculate that formate serves as a temporal pool of electrons that is, in turn, used to generate hydrogen.

During the oxidation of propionate, 1 mol of ATP is generated by substrate-level phosphorylation at the succinyl-CoA synthase step (Fig. 2). We consider that this ATP is partially used by F_0F_1-ATPase to produce a proton gradient across the cytoplasmic membrane that is necessary for the energy-consuming conversion of succinate to fumarate (Supplemental Fig. S3). In contrast, when this organism grows on fumarate, pyruvate, or ethanol, fumarate reductase (the reverse reaction of succinate dehydrogenase) produces the proton gradient, which is, in turn, used to produce ATP (Supplemental Fig. S3).

P. thermopropionicum has a gene cluster that encodes transporters and enzymes for sulfur anions (Supplemental Table S5). Its gene content is similar to those annotated as assimilatory sulfate-reducing gene clusters in *Clostridium acetobutylicum* (AE001437) and other bacteria (Rückert et al. 2005). On the other hand, gene clusters considered to be necessary for dissimilatory sulfate reduction, such as those for transmembrane electron transport complexes (Haveman et al. 2004), are not found in *P. thermopropionicum*. This result is consistent with previous physiological observations showing that this bacterium could not utilize sulfate as an electron acceptor (Imachi et al. 2002). Genome analyses also showed that the other electron-accepting systems (those for oxygen, nitrate, and metal respiration) were also not present in *P. thermopropionicum*, indicating that this organism is a specialist for syntrophic association with methanogens.

Signal transduction and regulation

A positive correlation between the number of genes involved in signal transduction and genome size has been revealed (Ulrich et al. 2005). According to regression models based on previously sequenced bacterial genomes (Ulrich et al. 2005), a 3.0-Mb-genome is estimated to encode 75 one-component and 15 two-component systems, and it has been reported that genomes of *Firmicutes* bacteria roughly follow this rule (Wu et al. 2005). In the case of *P. thermopropionicum*, however, more signal-transduction systems (165 one-component and 22 two-component systems) were identified, suggesting that this organism broadly regulates its cellular functions. When concerning input domains of these signal-transduction systems, PAS is the most frequently present (Supplemental Table S6), and GAF is the second. HTH_3, HD, AraC, LysR, and GGDEF are frequently detected output domains (Supplemental Table S6).

PAS domains function as signal input modules in proteins of diverse organisms, which sense oxygen, redox potential, light, and other environmental stimuli (Taylor and Zhulin 1999). One way in which prokaryotic PAS proteins sense the environment is by detecting changes in the electron transport system (Taylor and Zhulin 1999). In *P. thermopropionicum*, genes for PAS domain-containing putative regulators are physically linked to those for important catabolic enzymes (Supplemental Table S7). In particular, the *mmc* operon, which encodes most of the enzymes in the propionate-oxidizing methymalonyl–CoA pathway, is considered to be regulated by a PAS-domain-containing regulator homologous to RocR of *B. subtilis* (Calogero et al. 1994; Kosaka et al. 2006). These findings suggest that the catabolic pathways of *P. thermopropionicum* are mostly regulated in response to environmental conditions and/or global cellular situations rather than specific substrates. We deduce that this regulation strategy should be suitable for *P. thermopropionicum* to utilize a limited repertoire of growth substrates by means of the simple catabolic pathway. In addition, the abundant presence of GAF-domain-containing signal-transduction systems suggests that cyclic nucleotides may also play important roles in *P. thermopropionicum* (Martinez et al. 2002).

Comparative genomics

In addition to organisms appearing in Figure 1, some representative organisms (also listed in Supplemental Table S1) were also included in our comparative analyses. Nucleotide sequence-based phylogeny (mostly based on 16S rRNA gene sequences) has been considered as a primary index for describing evolutionary relationships among prokaryotes (Olsen et al. 1994). Based on this idea, SB are divided into two groups (Fig. 3A), namely, those affiliated with the phyla *Firmicutes* (represented by *Pth*, *Swol*, and *Moth*) and *Deltaproteobacteria* (*Sac* and *Sfu*). *Pth* is affiliated with the order *Clostridiales* in the *Firmicutes* together with *Chy*, *Moth*, *Tte*, *Swol*, *Cth*, and *Cac* (Fig. 3A).

Recent accumulation of genome sequences of diverse prokaryotes has provided new approaches that facilitate genome-scale analyses of their relationships (Bapteste et al. 2004), including gene-content analysis (Snel et al. 1999). In our gene-content analysis (Fig. 3B), close relationships of *Pth* with *Chy*, *Swol*, and *Moth* are revealed, while *Tte*, *Cth*, and *Cac* form another coherent cluster, showing niche-dependent separation of the *Clostridiales* members. Unexpectedly, this trend was also observed in the codon-usage analysis, whose results are shown in the self-organization map (SOM) (Fig. 3C). In this map, CDSs (presented as dots) with similar codon usages are situated in close proximity to each other, and these CDSs are considered to be evolved in similar organisms (Kanaya et al. 2001; Abe et al. 2005). This analysis thus provides us with evolutionary information that is independent of encoded functions. Previously, phylogenetically related organisms were shown to have similar codon-usage patterns (Kanaya et al. 2001; Abe et al. 2005). In our SOM analysis (Fig. 3C), however, CDSs of *Pth* were situated close to those of *Chy*, *Moth*, and *Swol* but very distant from those of *Tte*, *Cth*, and *Cac* (sugar-fermenting bacteria; see Fig. 1). *Chy* was originally known as a CO-utilizing hydrogenogen, while it later came to be known that it also grows by fermenting organic acids (Henstra and Stams 2004), suggesting its potential for syntrophic interaction with methanogens. In addition, CDSs of *Sac* (SB affiliated with *Deltaproteobacteria*) are also shown to be close to those of the *Pth*-related *Clostridiales* members (Fig. 3C), suggesting possible evolutionary interactions between them. It is likely that they have frequently exchanged genetic materials. CDSs of another SB (*Sfu*) are situated somewhat distantly from those of the other SBs (Fig. 3C). We consider that this trend may be relevant to its large genome size (Supplemental Table S1), which suggests that this organism has broader metabolic potentials than the SB specialists. CDSs of *Pth* are also shown to be close to those of *Dde*; this organism is known to be a sulfate-reducing bacterium, although it can also grow on organic acids, e.g., lactate, under syntrophic association with methnogens (Stolyar et al. 2007). These analyses suggest that codon usage is influenced by the niche of the host organism, probably due to frequent genetic exchanges between members within the niche.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

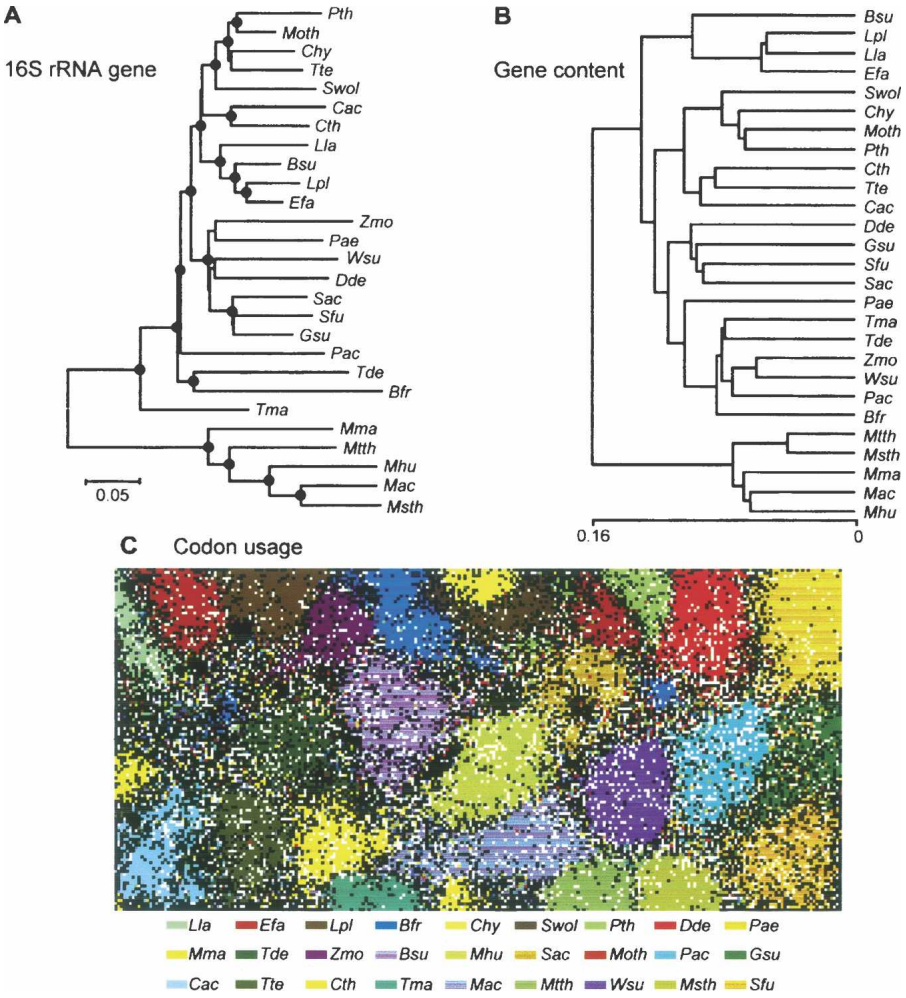


Figure 3. Evolutionary relationships among the organisms analyzed in this study. (*A*) A neighbor-joining tree based on 16S rRNA gene sequences. (Black circles) Branch nodes supported by the Bootstrap analysis (>70%). (*B*) A UPGMA tree based on gene contents. (*C*) A self-organization map (SOM) based on codon usages. Each point represents one CDS, and CDSs of one organism are presented with the same color. Italic letters indicate organism names; (*Chy*) *Carboxydothermus hydrogenoformans*; (*Lla*) *Lactococcus lactis*; (*Lpl*) *Lactobacillus plantarum*; (*Efa*) *Enterococcus faecalis*; (*Bsu*) *Bacillus subtilis*; (*Wsu*) *Wolinella succinogenes*; (*Dde*) *Desulfovibrio desulfuricans*; (*Gsu*) *Geobacter sulfurreducens*; (*Zmo*) *Zymomonas mobilis*; (*Pae*) *Pseudomonas aeruginosa*; (*Tma*) *Thermotoga maritima*. Refer to the legend of Figure 1 for other organisms and to Supplemental Table S1 for more information about these organisms.

We also found that CDSs of *P. thermopropionicum* formed several small areas in the SOM (Fig. 3C), suggesting a mixed nature for its genome probably owing to recent genetic exchanges. We listed 259 CDSs in *P. thermopropionicum* (Supplemental Table S8), which show atypical codon usage and are considered to have been acquired relatively recently. Representatives of these CDSs are those for hydrogenase, succinate dehydrogenase, CoA transferase, chemotaxis/flagellum proteins, and transporters, while many other CDSs encode hypothetical proteins (Supplemental Table S8). Among them, genes for succinate dehydrogenase are considered to have been transferred from Deltaproteobacterial organisms. Donor organisms of these 259 CDSs are predicted to be relatives of *Moth* (77 CDSs), *Dde* (31), *Swol* (28), *Mac* (25), *Chy* (19), *Sfu* (15), *Sac* (13), and others, suggesting that *P. thermopropionicum* has actively interacted with these organisms during its evolution. One such organism, *Mac*, is a methanogenic archaea, which suggests that *P. thermopropionicum* has also been affected by syntrophic partners to some extent. We conducted phylogenetic analyses of some representative CDSs exhibiting atypical codon usage and found that they also exhibited atypical phylogenetic relationships (those different from the 16S rRNA phylogeny) (Supplemental Fig. S4). For example, CDSs encoding subunits of succinate dehydrogenase were closely related to those in *Deltaproteobacteria*. The phylogenetic analyses are thus considered to support the results of codon-usage analysis. From these comparative genome analyses, it is revealed that *P. thermopropionicum* has evolved by interacting with microbes that are associated with the niche of SB.

Conclusions

The complete genome of *P. thermopropionicum* reveals its niche strategy: that the simple catabolic pathways and their PAS-dependent regulation should be suitable for the energy-limited syntrophic lifestyle of this organism. This is supported by its gene repertoire, which appears to be optimized for this lifestyle. We suggest that syntrophic organisms have adapted such a strategy through close interactions with microbes associated within their niche. In the codon-usage analysis, close evolutionary interactions among niche members were also observed for different groups of microbes, suggesting the importance of niche interactions in the genome evolution and niche speciation of prokaryotes. This idea should be further evaluated by analyzing a large data set of prokaryotic genomes.

Methods

Genome sequencing

The draft-genome materials and information (Kosaka et al. 2006) were used to complete the genome (Dragon Genomics). A physical map was constructed by optical mapping (OpGen), and physical linkages of contigs were analyzed by further sequencing of plasmid and fosmid libraries described previously (Kosaka et al. 2006). Gaps were closed by sequencing of gap-spanning PCR products. The genome was finally assembled into a single contig with average coverage of 11.1. The mean accuracy of the final sequence was 99.995%.

Annotation

The rRNA-encoding regions were identified by a BLASTN program using the ribosomal sequences of *Clostridium acetobutyli-*

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

cum, while tRNA-coding regions were predicted by the tRNAScan-SE program (Lowe and Eddy 1997). CDSs were predicted by CRITICA (Badger and Olsen 1999) and Glimmer 2 (Delcher et al. 1999) programs and manually identified by comparative analyses using the genome annotation system (Dragon Genomics). Annotation of the predicted CDSs was performed by a conserved-domain search against the COG database in the NCBI database with an E-value threshold of 0.01 (Marchler-Bauer et al. 2005) and the InterPro search (Apweiler et al. 2001). Protein domains and motifs were searched using the Pfam (Sonnhammer et al. 1998) and SMART (Schultz et al. 1998) databases. Membrane proteins and signal peptides were predicted by the SOSUI program (Hirokawa et al. 1998).

Comparative genomics

The 16S rRNA gene sequences of organisms listed in Supplemental Table S1 were aligned using the MEGA program version 3.1 (Kumar et al. 2004), and a phylogenetic tree was constructed using the neighbor-joining method (Saitou and Nei 1987). For analyzing relationships of organisms based on gene contents, we first constructed protein families according to the method of Enright et al. (2002, 2003). All sequences of CDSs in the organisms listed in Supplemental Table S1 were compared with each other using BLASTP with an *E*-value threshold of 1×10^{-15}. This E-value was determined by empirical examination. A similarity matrix was then calculated based on the BLASTP output; the BLASTP output was parsed by the TribeMCL part of the package to produce a matrix of protein similarities (BLASTP *E*-values). A resulting symmetric similarity table was converted into a Markov matrix and clustered into protein families using the TribeMCL algorithm (Enright et al. 2002, 2003). A dendrogram was constructed using the UPGMA algorithm. For analyzing codon-usage similarities, codon usages of all CDSs in the organisms listed in Supplemental Table S1 were analyzed, and their similarities were estimated according to reference (Kanaya et al. 2001; Abe et al. 2005), and CDSs were clustered onto a two-dimensional map based on similarities using the self-organization map (SOM) program (Kanaya et al. 2001).

Acknowledgments

We thank Yasuo Igarashi and Hiroshi Ikenaga for continuous encouragement; Fusako Numazaki, Midori Sato, Hiromi Awabuchi, and Reiko Hirano for technical support; and Robert Kanaly and Greg Newton for critical reading of the manuscript. This work was supported by New Energy Development Organization (NEDO).

References

Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S., and Ikemura, T. 2005. Self-Organizing Map (SOM) unveils and visualizes hidden sequence characteristics of a wide range of eukaryote genomes. *DNA Res.* **12:** 281–290.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., et al. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29:** 37–40.

Badger, J.H. and Olsen, G.J. 1999. CRITICA: Coding Region Identification Tool Invoking Comparative Analysis. *Mol. Biol. Evol.* **16:** 512–524.

Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., et al. 2002. A complete sequence of the *T. tengcongensis* genome. *Genome Res.* **12:** 689–700.

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W.F. 2004. Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* **12:** 406–411.

Calogero, S., Gardan, R., Glaser, P., Schweizer, J., Rapoport, G., and Debarbouille, M. 1994. RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. *J. Bacteriol.* **176:** 1234–1241.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27:** 4636–4641.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30:** 1575–1584.

Enright, A.J., Kunin, V., and Ouzounis, C.A. 2003. Protein families and TRIBES in genome sequence space. *Nucleic Acids Res.* **31:** 4632–4638.

Glissmann, K. and Conrad, R. 2000. Fermentation pattern of methanogenic degradation of rice straw in anoxic paddy soil. *FEMS Microbiol. Ecol.* **31:** 117–126.

Harmsen, H.J., van Kuijk, B.L., Plugge, C.M., Akkermans, A.D., de Vos, W.M., and Stams, A.J. 1998. *Syntrophobacter fumaroxidans* sp. nov., a syntrophic propionate-degrading sulfate-reducing bacterium. *Int. J. Syst. Bacteriol.* **48:** 1383–1387.

Haveman, S.A., Greene, E.A., Stilwell, C.P., Voordouw, J.K., and Voordouw, G. 2004. Physiological and gene expression analysis of inhibition of *Desulfovibrio vulgaris* Hildenborough by nitrite. *J. Bacteriol.* **186:** 7944–7950.

Henstra, A.M. and Stams, A.J. 2004. Novel physiological features of *Carboxydothermus hydrogenoformans* and *Thermoterrabacterium ferrireducens*. *Appl. Environ. Microbiol.* **70:** 7236–7240.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. 1998. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14:** 378–379.

Imachi, H., Sekiguchi, Y., Kamagata, Y., Hanada, S., Ohashi, A., and Harada, H. 2002. *Pelotomaculum thermopropionicum* gen. nov., sp. nov., an anaerobic, thermophilic, syntrophic propionate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* **52:** 1729–1735.

Jackson, B.E. and McInerney, M.J. 2002. Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature* **415:** 454–456.

Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., and Ikemura, T. 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): Characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene* **276:** 89–99.

Kaspar, H.F. and Wuhrmann, K. 1978. Kinetic parameters and relative turnovers of some important catabolic reactions in digesting sludge. *Appl. Environ. Microbiol.* **36:** 1–7.

Konstantinidis, K.T. and Tiedje, J.M. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci.* **101:** 3160–3165.

Kosaka, T., Uchiyama, T., Ishii, S., Enoki, M., Imachi, H., Kamagata, Y., Ohashi, A., Harada, H., Ikenaga, H., and Watanabe, K. 2006. Reconstruction and regulation of the central catabolic pathway in the thermophilic propionate-oxidizing syntroph *Pelotomaculum thermopropionicum*. *J. Bacteriol.* **188:** 202–210.

Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* **5:** 150–163.

Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25:** 955–964.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., et al. 2005. CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33:** D192–D196.

Martinez, S.E., Beavo, J.A., and Hol, W.G. 2002. GAF domains: Two-billion-year-old molecular switches that bind cyclic nucleotides. *Mol. Interv.* **2:** 317–323.

McInerney, M.J., Rohlin, L., Mouttaki, H., Kim, U., Krupp, R.S., Rios-Hernandez, L., Sieber, J., Struchtemeyer, C.G., Bhattacharyya, A., Campbell, J.W., et al. 2007. The genome of *Syntrophus aciditrophicus*: Life at the thermodynamic limit of microbial growth. *Proc. Natl. Acad. Sci.* **104:** 7600–7605.

Olsen, G.J., Woese, C.R., and Overbeek, R. 1994. The winds of (evolutionary) change: Breathing new life into microbiology. *J. Bacteriol.* **176:** 1–6.

Ralph, D. and McClelland, M. 1993. Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. *Proc. Natl. Acad. Sci.* **90:** 6864–6868.

Rocha, E.P. 2004. The replication-related organization of bacterial genomes. *Microbiology* **150:** 1609–1627.

Rodrigue, A., Chanal, A., Beck, K., Müller, M., and Wu, L.F. 1999. Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *J. Biol. Chem.*

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

274: 13223–13228.
Rückert, C., Koch, D.J., Rey, D.A., Albersmeier, A., Mormann, S., Puhler, A., and Kalinowski, J. 2005. Functional genomics and expression analysis of the *Corynebacterium glutamicum fpr2-cysIXHDNYZ* gene cluster involved in assimilatory sulphate reduction. *BMC Genomics* **6:** 121. doi: 10.1186/1471-2164-6-121.
Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.
Schink, B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61:** 262–280.
Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci.* **95:** 5857–5864.
Snel, B., Bork, P., and Huynen, M.A. 1999. Genome phylogeny based on gene content. *Nat. Genet.* **21:** 108–110.
Sonnhammer, E.L., Eddy, S.R., Birney, E., Bateman, A., and Durbin, R. 1998. Pfam: Multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* **26:** 320–322.
Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J.A., and Stahl, D.A. 2007. Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.* **3:** 92.
Tatusov, R., Koonin, E., and Lipman, D. 1997. A genomic perspective on protein families. *Science* **278:** 631–637.
Taylor, B.L. and Zhulin, I.B. 1999. PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63:** 479–506.
Thauer, R.K., Jungermann, K., and Decker, K. 1977. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* **41:** 100–180.
Ulrich, L.E., Koonin, E.V., and Zhulin, I.B. 2005. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* **13:** 52–56.
Wu, M., Ren, Q., Durkin, A.S., Daugherty, S.C., Brinkac, L.M., Dodson, R.J., Madupu, R., Sullivan, S.A., Kolonay, J.F., Haft, D.H., et al. 2005. Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet.* **1:** e65. doi: 10.1371/journal.pgen.0010065.

Received September 10, 2007; accepted in revised form December 7, 2007.