Genomic and molecular ecological studies on thermophilic hydrogenogenic carboxydotrophs

Kimiho OMAE

2020

Table of contents

Chapter 1
General introduction

Carbon monoxide (CO) is well known as highly toxic, colorless, tasteless, and odorless gas called a “silent killer” (Rose *et al.*, 2017). The toxicity mainly comes from its high affinity to many ferrous heme-containing proteins, including hemoglobin, myoglobin, and cytochrome oxidase, which results in the inhibition of these proteins (Wu & Wang, 2005). It is known that CO is toxic not only for human, but also for many microorganisms (Wareham *et al.*, 2018). Chemical compounds called CO-releasing molecules, which are transition metal complexes and locally deliver CO to cells and tissues, have been developed as potential therapeutic antimicrobial agents against pathogenic microorganisms (Nobre *et al.*, 2007). Growth inhibition by atmospheric CO gas has also been reported in sulfate-reducing and acetogenic microorganisms, and methanogenic archaea (Davidova *et al.*, 1994; Parshina *et al.*, 2005a). In microbial cells, CO inhibits the intracellular respiratory heme enzymes (Davidge *et al.*, 2009) and the iron-sulfur enzymes of tricarboxylic acid (TCA) cycle (Carvalho *et al.*, 2019). Generation of reactive oxygen species is considered as another potential toxicity caused by CO (Tavares *et al.*, 2011).

Meanwhile, some microorganisms can grow by using CO in metabolic process (Ragsdale, 2004; Xavier *et al.*, 2018). These microorganisms possess CO dehydrogenase (CODH), which catalyzes the interconversion of CO and CO_2 as follows: $CO + H_2O \leftrightarrow CO_2 + 2H^+ + 2e^-$ (Dobbek *et al.*, 2001). CODHs are divided into two distinct families: an aerobic molybdenum- and copper-containing CODH encoded by *cox* genes (Hille *et al.*, 2015; King & Weber, 2007), and an anaerobic nickel-containing CODH encoded by *cooS*

gene (Aragão *et al.*, 2008; Inoue *et al.*, 2019a). One prominent CO-related metabolic process by microorganisms is anaerobic carbon fixation mechanism called the Wood-Ljungdahl pathway (WLP) (or reductive acetyl coenzyme A [acetyl-CoA] pathway) (Ragsdale, 2004). Of the six carbon fixation pathways encountered in prokaryotes (Fuchs, 2011), the WLP is the sole pathway which involves CO as an intermediate and has been proposed to be one of the oldest because of its linear nature, chemical simplicity, favorable energetics, and occurrence among both bacteria and archaea (Peretó, 2012; Xavier *et al.*, 2018). A CODH/acetyl-CoA synthase (ACS) complex is responsible for the final steps of the WLP, where anaerobic-type CODH reduces CO_2 to CO and provide the source of the carbonyl group of acetyl-CoA to ACS (Ragsdale, 2004). ACS combines the carbonyl moiety with a methyl group, which is donated by another CO_2-fixing branch of the WLP (the methyl branch), and CoA to form acetyl-CoA as a building block (Ragsdale, 2004). The CODH/ACS homologs are widely conserved in both prokaryotic domains (Adam *et al.*, 2018).

CO serves as an energy source for microorganisms as well as an inorganic carbon source (Diender *et al.*, 2015; Mörsdorf *et al.*, 1992; Robb & Techtmann, 2018). CO has one of the lowest-redox potentials ($E^{0\prime}$ [CO/CO_2] = −520 mV) in inorganic compounds available in the biosphere (Thauer *et al.*, 1977). Therefore, oxidation reaction of CO to CO_2 catalyzed by CODH provides energy-rich electrons which can be used for the generation of a proton motive force across the cytoplasmic membrane leading to the formation of ATP (Mörsdorf *et al.*, 1992). While aerobic-type CODH can utilize O_2 and nitrate as terminal electron acceptors, various compounds such as CO_2, sulfate and Fe(III) can be used as terminal electron acceptors for energy conservation by anaerobic-type CODH, which results in physiological diversity of anaerobic CO-oxidizing

microorganisms: acetogens, methanogens, sulfate reducers and iron reducers (Diender *et al.*, 2015; Robb & Techtmann, 2018). In addition, CO can directly reduce low potential electron carrier like ferredoxin (Fd), a low-molecular-weight, iron-sulfur center-containing protein, generating powerful reducing power which drives energy conservation and carbon fixation (Schuchmann & Müller, 2014).

Of physiologically diverse anaerobic CO utilizing microorganisms, hydrogen-producing CO oxidizers (hydrogenogenic carboxydotrophs) have remarkable features. The microbial hydrogenogenic CO oxidizing activity is achieved by membrane associated CO-oxidizing and H_2-producing system, which is characterized in mesophilic photosynthetic bacteria for the first time (Ensign & Ludden, 1991; Wakim & Uffen, 1983). In this system, anaerobic-type CODH (hereafter, I call it simply the CODH), electron-transferring ferredoxin like protein (CooF) and membrane-bound H_2-evolving Group 4c [NiFe] hydrogenase (Søndergaard *et al.*, 2016), which are encoded in a CO-induced operon, form a complex (Fox *et al.*, 1996a, b). In the complex, the CODH carries out the primary oxidation of CO and passes the resulting reducing-equivalents via the CooF to the membrane-bound Group 4c [NiFe] hydrogenase, where protons are reduced to H_2 and the proton motive force is possibly generated at the same time (Bonam & Ludden, 1987; Ensign & Ludden, 1991; Soboh *et al.*, 2002). Similar mechanism is also predicted in other hydrogenogenic carboxydotrophs harboring a gene cluster comprised of CODH and Group 4a (Sant'Anna *et al.*, 2015) or 4b (Lee *et al.*, 2008; Lim *et al.*, 2010) [NiFe] hydrogenase which is iron-translocating (Schoelmerich & Müller, 2019). Today, the CO-oxidizing and H_2-producing systems is recognized as the combination of CODH and membrane-bound H_2-evolving Group 4 [NiFe] hydrogenases which are correctively called energy-converting hydrogenase (ECH) (Techtmann *et al.*, 2012). The CODH/ECH

complex, which is often found in a single gene cluster, might comprise the simplest respiration machinery with an ATP synthase (Schoelmerich & Müller, 2019), because it achieves energy conservation without any extra electron carrier nor terminal electron acceptor (i.e. it only requires water).

CO is a trace gas in the atmosphere which occurs at a mixing ratio of 0.06-0.15 ppm (Mörsdorf *et al.*, 1992), but the total budget of atmospheric CO is 2,000–3,000 Tg/year (Khalil & Rasmussen, 1990; Mörsdorf *et al.*, 1992; Schade & Crutzen, 1999), and about a half of the CO emissions are from natural chemical processes (e.g., photochemical and thermochemical degradation of organic matter) and biological processes (e.g., production by microorganisms, leaves, roots, and animals) in volcanic, fresh water, marine, and terrestrial environments (Conrad, 1996; Conte *et al.*, 2019; Khalil & Rasmussen, 1990; King & Weber, 2007; Mörsdorf *et al.*, 1992). Therefore, CO-dependent H_2 production by hydrogenogenic carboxydotrophs is considered a "safety valve" to reduce toxic CO and supply H_2 which is an energy source for H_2-utilizing microbial communities (Techtmann *et al.*, 2009). In addition, H_2 fuel has now expanded to human society as a zero emission fuel and CO is contained in syngas or industrial waste gases (Dürre & Eikmanns, 2015). Therefore, hydrogenogenic carboxydotrophy has potential for biotechnological applications to produce H_2 from CO in syngas or industrial waste gases.

Since 1970s, phylogenetically diverse hydrogenogenic carboxydotrophs have been reported. *Rhodospirillum rubrum* ATCC 11170 is a mesophilic photosynthetic proteobacterium which is isolated from fresh water (Dashekvicz & Uffen, 1979), and is a model of hydrogenogenic carboxydotroph (Ensign & Ludden, 1991; Fox *et al.*, 1996a, b). In the dark, *R. rubrum* can grow under ~100% CO in head space with doubling time of

5–8 h (Kerby *et al.*, 1995). *Carboxydothermus hydrogenoformans* Z-2901 is another model of hydrogenogenic carboxydotroph isolated from a terrestrial hot spring (Svetlichny *et al.*, 1991). In contrast to *R. rubrum*, *C. hydrogenoformans* Z-2901 is a thermophilic bacterium of phylum Firmicutes and grows rapidly under 100% CO with doubling time of 2 h at 70–72°C (Svetlichny *et al.*, 1991). Its genome harbors five CODH gene clusters including CODH–ACS (carbon fixation), CODH–ECH (energy conservation) and CODH–*cooF* (reducing power generation), which may in part explain how this species is able to grow so much more rapidly on CO than many other species (Wu *et al.*, 2005). From deep-sea hydrothermal vents, *Thermococcus onnurineus* NA1 was isolated (Bae *et al.*, 2006), which is a hyper-thermophilic archaeon of phylum Euryarchaeota, and grows hydrogenogenically under 100% CO at 80°C (Kim *et al.*, 2013; Lee *et al.*, 2008). Biotechnological application for hydrogen production from synthetic gas has advanced in this isolate (Kim *et al.*, 2013). Including those representatives, 23 isolates of hydrogenogenic carboxydotrophic archaea and bacateria in five phyla, 16 genera, and 22 species had been described before 2010 (Diender *et al.*, 2015; Sokolova *et al.*, 2009; Techtmann *et al.*, 2009). Seventeen of these isolates are (hyper-)thermophiles isolated from hot springs, deep-sea hydrothermal fields or bioreactors, most of which are comprised of phylum Firmicutes (nine genera 12 species 13 isolates) (Diender *et al.*, 2015; Sokolova *et al.*, 2009; Techtmann *et al.*, 2009).

Under those circumstances, our laboratory achieved isolation of multiple novel thermophilic hydrogenogenic carboxydotrophs mainly from terrestrial hot springs in Japan in 2010s (Fukuyama *et al.*, 2017, 2019; Inoue *et al.*, 2019b; Yoneda *et al.*, 2012, 2013a). In particular, we successfully isolated novel species of the genus *Carboxydothermus*, *C. pertinax* Ug1, from a terrestrial acidic hot spring of Unagi-onsen

in southern Kyushu Island (Yoneda *et al.*, 2012). Furthermore, we successfully isolated *Calderihabitans maritimus* KKC1 from a submerged marine caldera (Kikai Caldera) on the southern coast of Kyushu Island, which is the first thermophilic hydrogenogenic carboxydotroph isolated from a marine sediment core and represents new genus of Firmicutes (Yoneda *et al.*, 2013a). Therefore, I revealed the novel CO metabolism of hydrogenogenic carboxydotroph by *de novo* genomic sequencing and analysis of *C. maritimus* KKC1 (In Chapter 2). *C. maritimus* KKC1 is a thermophilic endospore-forming bacterium, which might have survived in the mesophilic marine sediment by forming endospore and be an ancient bacterium. Also, *C. maritimus* KKC1 couples CO oxidation with reduction of sulfur- or iron-compounds during hydrogenogenic growth (Yoneda *et al.*, 2013a) as reported in other thermophilic hydrogenogenic carboxydotrophs (Fukuyama *et al.*, 2018; Toshchakov *et al.*, 2018), suggesting that these microorganisms have versatile cellular metabolic processes driven by CO. However, knowledge for CO metabolism by hydrogenogenic carboxydotrophs is limited, largely due to the small number of genomic information. The novel genomic information of *C. maritimus* KKC1, which was unexpectedly isolated from mesophilic environment, provided insight into the CO metabolism of these microorganisms.

After isolation of *C. maritimus* KKC1, other studies also reported thermophilic hydrogenogenic carboxydotrophs from mesophilic environments like soil, marine and lake sediments (Fukuyama *et al.*, 2017; Inoue *et al.*, 2019b; Mohr *et al.*, 2018), suggesting unexpectedly wide distribution of these microorganisms. In addition, isolation-based study itself arose cultivation bias resulting in limited lineages of hydrogenogenic carboxydotrophic isolates, largely due to the strong selective pressure by CO. Although it is predicted that still unknown thermophilic hydrogenogenic carboxydotrophs with

novel CO metabolism exist in environments, ecological information on diversity and distribution of these microorganisms is not sufficient. Therefore, I developed the molecular ecological techniques for exploration of hydrogenogenic carboxydotrophs with avoiding cultivation bias to elucidate ecology of these microorganisms. To enable exploration of thermophilic hydrogenogenic carboxydotrophic isolates by microbial community analysis (16S amplicon sequencing), I constructed a reference database of these microorganisms with revealing distribution of CODH–ECH gene clusters upon 16S rRNA phylogeny by bioinformatics-based analysis (In Chapter 3). This analysis also identified 46 overlooked potential hydrogenogenic carboxydotrophs harboring CODH–ECH gene clusters, whose hydrogenogenic CO oxidizing activities have never been reported so far, from increasing genomic database. Furthermore, I designed new primers for PCR amplification of CODH genes of CODH–ECH gene clusters for the first time to evaluate thermophilic hydrogenogenic carboxydotrophs including unknown species by culture-independent way (In Chapter 4). These studies correctively suggested that thermophilic hydrogenogenic carboxydotrophs including unknown species widely distribute in a variety of environment.

Chapter 2

Genomic analysis of *Calderihabitans maritimus* KKC1, a thermophilic hydrogenogenic carboxydotrophic bacterium isolated from marine sediment

Summary

Calderihabitans maritimus KKC1 is a thermophilic, hydrogenogenic carboxydotroph isolated from a submerged marine caldera. Here, I describe the *de novo* sequencing and feature analysis of the *C. maritimus* KKC1 genome. Genome-based phylogenetic analysis confirmed that *C. maritimus* KKC1 was most closely related to the genus *Moorella* which includes well-studied acetogenic members. Comparative genomic analysis revealed that, like *Moorella*, *C. maritimus* KKC1 retained both the CO_2-reducing Wood-Ljungdahl pathway (WLP) and energy-converting hydrogenase (ECH)-based modules activated by reduced ferredoxin, but it lacked the HydABC and NfnAB electron-bifurcating enzymes required for ferredoxin reduction in autotrophic acetogenic growth. Furthermore, *C. maritimus* KKC1 harbored six genes encoding carbon monoxide dehydrogenase (CODH) that can reduce ferredoxin via CO oxidation, whereas *Moorella* possessed only two CODH genes. My analysis revealed that three CODH genes formed known gene clusters in other microorganisms, i.e., CODH–ACS (which contained a frameshift mutation), CODH–ECH, and *cooF*–CODH–flavin adenine dinucleotide-nicotinamide adenine dinucleotide oxidoreductase, while the other three had novel genomic contexts. One of the three novel CODH genes was flanked by 2-oxoglutarate:ferredoxin oxidoreductase, which is a CO_2-fixing enzyme in reverse tricarboxylic acid (RTCA) cycle, and expected

to comprise novel carbon fixation pathway where CO is incorporated into RTCA cycle. Sequence composition analysis indicated that these CODH genes likely evolved from a common ancestor. Collectively, these data suggest that *C. maritimus* KKC1 may be highly dependent on CO as a low-potential electron donor to directly reduce ferredoxin and may be more suited to carboxydotrophic growth compared to the acetogenic growth observed in *Moorella*, which show adaptation at a thermodynamic limit.

Introduction

Carbon monoxide (CO) is a potent electron donor that can serve as an energy and carbon source for thermophilic carboxydotrophs (CO-oxidizing microbes) (Diender *et al.*, 2015; Mörsdorf *et al.*, 1992; Robb & Techtmann, 2018). CO utilization requires specific carbon monoxide dehydrogenases (CODHs) to catalyze the reversible reaction $CO + H_2O \leftrightarrow CO_2 + 2H^+ + 2e^-$ (Dobbek *et al.*, 2001). CODHs from anaerobic microorganisms possess a nickel-containing reaction center (Ni-CODHs) (Dobbek *et al.*, 2001; Svetlitchnyi *et al.*, 2004), whereas aerobic-type CODHs contain a highly conserved molybdenum-based active site (Hille et al., 2015; King & Weber, 2007). From thermodynamic considerations, CO oxidation can be coupled to the reduction of most redox-active cofactors (Oelgeschläger & Rother, 2008). A number of diverse physiological anaerobic carboxydotrophs and CO oxidizers have been described, such as acetogens, methanogens, sulfate reducers, iron reducers, and hydrogenogens (Sokolova *et al.*, 2009), many of which possess multiple Ni-CODH genes (Sokolova & Lebedinsky, 2013; Techtmann *et al.*, 2012; Wu *et al.*, 2005).

Recent comprehensive genomic database surveys reveal that 2–6% of bacterial

and archaeal genomes contain at least one CODH gene (Inoue *et al.*, 2019a; Techtmann *et al.*, 2012). CODHs are subdivided into the Cdh type, almost all of which are found in archaea, and the CooS type, which are more frequent in bacteria (Inoue *et al.*, 2019a; Techtmann *et al.*, 2012). While CooS-type CODHs contain one [Ni-Fe-S] cluster (C cluster) and two [4Fe-4S] clusters (B and D clusters), Cdh types harbor two additional [4Fe-4S] clusters (E and F clusters) (Gencic *et al.*, 2010) and generally show relatively low homology to CooS-type CODHs. The functions of CODHs have often been predicted from other genes located in close proximity to CODH genes (genomic context) (Inoue *et al.*, 2019a; Techtmann *et al.*, 2012). CODHs can be divided into four functional groups according to their genomic context (Techtmann *et al.*, 2012): (i) within an acetyl coenzyme A (acetyl-CoA) synthase (ACS) gene cluster, (ii) adjacent to an energy-converting hydrogenase (ECH) gene cluster, (iii) adjacent to a CooF gene but not an ECH gene cluster, and (iv) other than types i to iii.

CODHs in category i form CODH/ACS complexes that catalyze the reduction of CO_2 to CO and acetyl-CoA synthesis in the final step of the Wood-Ljungdahl pathway (WLP) (Ragsdale, 2004). These complexes are widespread in CO-oxidizing and non-CO-oxidizing anaerobes that employ the WLP, such as acetogens (Drake *et al.*, 2002; Ragsdale, 1997), methanogens (Ferry, 1999; Ladapo & Whitman, 1990; Stupperich *et al.*, 1983), sulfate reducers (Schauder *et al.*, 1988; Spormann & Thauer, 1988), and thermophilic hydrogenogenic carboxydotrophs (Svetlitchnyi *et al.*, 2004). Cdh-type CODHs fall exclusively in the type i category, while type ii CODH genes cluster with those encoding ECH, whose presence is therefore considered a feature of hydrogenogenic carboxydotrophs. Three types of CODH–ECH gene clusters are known in bacteria and archaea as mentioned in Chapter 1. One corresponds to the *coo* (CO-oxidizing) gene

cluster found in *Carboxydothermus hydrogenoformans* (Wu *et al.*, 2005) and *Rhodospirillum rubrum* (Fox *et al.*, 1996a, b) and includes genes encoding CO-induced hydrogenase, which is phylogenetically classified as Group 4c [NiFe] hydrogenase (Søndergaard *et al.*, 2016). The second is found in *Caldanaerobacter subterraneus* subspecies (Sant'Anna *et al.*, 2015) and clusters with the *hyf*/*hyc*-type ECH genes long known as the [NiFe] hydrogenase module (Group 4a) of formate hydrogen lyase complexes (Søndergaard *et al.*, 2016). The last is found in *Thermococcus* archaea and comprises a CODH–Group 4b [NiFe] hydrogenase–cation/proton antiporter gene cluster (Lee *et al.*, 2008; Lim *et al.*, 2010). CODH genes of type iii are believed to encode an CODH responsible for generating electrons during CO oxidation and transferring them to CooF, which in turn relays them to various redox reactions. Members of group iv are "lone" CODH genes, in that they are not found in a genomic context with known CO metabolism-related genes.

Some thermophilic hydrogenogenic carboxydotrophs, such as *C. hydrogenoformans* and *C. subterraneus* subsp. *pacificus*, can propagate on high concentrations of CO as the sole carbon and energy sources (Oelgeschläger & Rother, 2008). Thermophilic hydrogenogenic carboxydotrophs have been studied extensively as models of CO metabolism, and a genomic study revealed that *C. hydrogenoformans* possessed five distinct CODH genes, one each of types i, ii, and iv and two of type iii (Wu *et al.*, 2005). In contrast, the genome of *C. subterraneus* subsp. *pacificus* includes only one CODH gene cluster of type ii (Sant'Anna *et al.*, 2015). As mentioned above, type ii CODH gene clusters in both organisms are distinct even though they exhibit physiology similar to that of thermophilic hydrogenogenic carboxydotrophs. Thus, the presence of highly divergent CODH gene cluster combinations prompts fundamental

questions on their function, evolution, and origin.

Here, I describe the *de novo* sequencing and feature analysis of the *Calderihabitans maritimus* KKC1 genome. *C. maritimus* KKC1 is a hydrogenogenic carboxydotrophic thermophile isolated from a sediment core sample taken from a submerged marine caldera (Yoneda *et al.*, 2013a). *C. maritimus* KKC1 belongs to the member of thermophilic endospore-forming anaerobic bacteria in the family Thermoanaerobacteraceae of the phylum Firmicutes and is the first hydrogenogenic carboxydotroph isolated from marine sediment (Yoneda *et al.*, 2013a). Taking account into the mesophilic environment of isolation site (note that temperature in the bottom core site was 15.3 °C), where the growth of *C. maritimus* KKC1 will not occur (Yoneda *et al.*, 2013a), it is assumed that *C. maritimus* KKC1 survived by forming endospore in the sediment which might be apart from original high-temperature habitat. According to 16S rRNA phylogenetic analysis, *C. maritimus* KKC1 showed only <91% identity with the closest *Moorella* species suggesting the isolate represents a novel genus (Yoneda *et al.*, 2013a). In addition, *Moorella stamsii* and *Moorella thermoacetica* strain AMP are reported to be hydrogenogenic carboxydotrophs like *C. maritimus* KKC1 (Alves *et al.*, 2013; Jiang *et al.*, 2009), but most *Moorella* strains are known homoacetogens. *M. thermoacetica* is the type species for the genus and is a well-known model of acetogenic bacteria that can grow autotrophically using H_2 plus CO_2 or CO to produce acetate via the WLP (Drake & Daniel, 2004; Fontaine *et al.*, 1942). Therefore, I compared the overall genomic features of *C. maritimus* KKC1 (a hydrogenogenic carboxydotroph) to those of acetogenic *M. thermoacetica* ATCC 39073 and *Moorella perchloratireducens* An10 and analyzed CODH gene clusters to gain insight into the physiological and phylogenetic differences between *C. maritimus* KKC1 and *Moorella* groups.

Materials and Methods

Bacterial strains, genome sequencing, and assembly

C. maritimus KKC1 was isolated and maintained in our laboratory at 65°C in hypotonic artificial seawater (hASW) medium under a 100% CO atmosphere (Yoneda *et al.*, 2013a). Genomic DNA was extracted by the NaOH method as previously described (Yoneda *et al.*, 2013a) and sequenced by Fasmac Co. Ltd., (Kanagawa, Japan) using MiSeq, NexteraXT, and TruSeq DNA sample preparation kits (Illumina, San Diego, CA, USA). I obtained 4,553,796 150-bp paired-end reads; those displaying a Phred score above Q20 for 80% of the bases were quality filtered using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). This yielded 2,835,116 reads, which were then assembled with Velvet version 1.2.10 program (Zerbino & Birney, 2008).

Open reading frames (ORFs) prediction and annotation

To predict ORFs in the *C. maritimus* KKC1 genome, I employed Glimmer version 3.02 program (Delcher *et al.*, 2007), which uses Markov's interpolated models, and GeneMarkS version 4.29 program (Besemer, 2001), followed by a manual curation process. After the ORFs were determined, protein sequences were further analyzed by BLASTp searches against nonredundant protein sequences in the National Center for Biotechnology Information (NCBI) (Sayers *et al.*, 2019), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2019), and clusters of orthologous groups (COG) databases (Tatusov *et al.*, 2001). tRNA and rRNA were predicted using tRNA Scan-SE version 1.3.1 program (Schattner *et al.*, 2005) and RNAmmer version 1.2 program (Lagesen *et al.*, 2007), respectively.

Phylogenetic analysis based on 16S rRNA, housekeeping genes, and GSS

I retrieved the 16S rRNA gene sequences of the Thermoanaerobacteraceae family from the Reference Sequence Database in NCBI (RefSeq). The sequences were aligned using MUSCLE version 3.8.31 program (Edgar, 2004), and gap positions were removed automatically using trimAL version 1.4 program (Capella-Gutiérrez *et al.*, 2009). Phylogenetic reconstructions were performed by the maximum-likelihood (ML) method using PhyML version 3.1 program (Guindon *et al.*, 2010) and visualized with MEGA version 6.06 package (Tamura *et al.*, 2011). Robustness of the topology of the phylogenetic trees was evaluated by bootstrap analysis based on 100 runs.

For genome-wide phylogenetic analysis, I collected 29 publicly available genomes of Thermoanaerobacteraceae members: from NCBI, *Ammonifex degensii* KC4 (NC_013385), *Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4 (NC_003869), *Carboxydothermus hydrogenoformans* Z-2901 (NC_007503), *Moorella thermoacetica* ATCC 39073 (NC_007644), *Tepidanaerobacter acetatoxydans* Re1 2011 (NC_015519), *Tepidanaerobacter acetatoxydans* Re1 2013 (NC_019954), *Thermoacetogenium phaeum* DSM 12270 (NC_018870), *Thermoanaerobacter brockii* subsp. *finnii* Ako-1 (NC_014964), *Thermoanaerobacter italicus* Ab9 (NC_013921), *Thermoanaerobacter mathranii* subsp. *mathranii* A3 (NC_014209), *Thermoanaerobacter pseudethanolicus* ATCC 33223 (NC_010321), *Thermoanaerobacter* sp. X513 (NC_014538), *Thermoanaerobacter* sp. X514 (NC_010320), and *Thermoanaerobacter wiegelii* Rt8.B1 (NC_015958); from Integrated Microbial Genomes (IMG) (Chen *et al.*, 2019), *Caldanaerobacter subterraneus* subsp. *pacificus* DSM 12653 (647533123), *Caldanaerobacter subterraneus* subsp. *yonseiensis* KB-1 (2563367176), *Caldanaerobius polysaccharolyticus* DSM 13641 (2510065085), *Carboxydothermus ferrireducens* DSM

11255 (2510065088), *Desulfovirgula thermocuniculi* DSM 16036 (2524023160), *Moorella perchloratireducens* An10 (2506520025), *Moorella thermoacetica* Y72 (2582580993), *Thermoanaerobacter ethanolicus* CCSD1 (645058764), *Thermoanaerobacter ethanolicus* JW 200 (2503538027), *Thermoanaerobacter indiensis* BSB-33 (2517287027), *Thermoanaerobacter kivui* DSM 2030 (2576861811), *Thermoanaerobacter siderophilus* SR4 (2509276025), *Thermoanaerobacter* sp. strain A7A, *Thermoanaerobacter* sp. strain X561 (645058760), *Thermoanaerobacter thermocopriae* JCM 7501 (2546825535), and *Thermoanaerobacter thermohydrosulfuricus* WC1 (2517572224).

I retrieved the amino acid sequences corresponding to the genes for ribosome recycling factor (*frr*), transcription elongation factor (*nusA*), 50S ribosomal protein L2 (*rplB*), 50S ribosomal protein L27 (*rpmA*), and elongation factor Ts (*tsf*) from the genomes of the Thermoanaerobacteraceae species listed above. The sequences were aligned and trimmed as described above. Concatenated alignments of five genes were then used to build an ML tree using PhyML (bootstrap = 100).

To compute the similarity between genomes of Thermoanaerobacteraceae, I calculated the corresponding genomic similarity score (GSS) (Moreno-Hagelsieb & Ha, 2008). This measurement is based on the sum of bit scores of shared orthologs. These are determined by the all-versus-all BLASTp search using protein sets and are normalized against the sum of bit scores of the compared genes against themselves (self-bit score). I used protein sets for each genome with coverage of 70% of both genes, with an E value of $1 \times e^{-5}$ at an effective database size of 10^7. The GSS ranged from 0 to 1, and the maximum score was obtained when two proteomes were identical. The neighbor-joining tree was built using a GSS distance matrix (Alcaraz *et al.*, 2010).

Phylogenetic analysis of CooS genes

I retrieved CooS amino acid sequences by BLASTp searches against RefSeq proteins. I also used some sequences of the Ni-CODH phylogenetic tree from Techtmann *et al*. (Techtmann *et al.*, 2012) as references. The sequences were aligned and trimmed, and used to build an ML tree (bootstrap = 100) as described above.

Horizontal gene transfer analysis of CooS genes

I calculated tetranucleotide frequencies of coding sequences (CDSs) with the length of greater than or equal to 500 bp from *C. maritimus* KKC1 and its genome (all contigs were catenated, and sequence gaps "N" were removed). Sequences were extended with their reverse complements. The observed frequencies of all 256 possible tetranucleotides were computed for these sequences. I calculated Euclidean distances of tetranucleotide frequencies of CDSs to that of whole genome and evaluated the significance of distances of CooSs.

Accession number(s)

The draft genome sequence generated in this study has been deposited in the DNA Data Bank of Japan (DDBJ) database under accession numbers BDGJ01000001 to BDGJ01000223.

Results

General features of the *C. maritimus* KKC1 genome and subsequent phylogenetic analysis

Overall, draft assemblies of the *C. maritimus* KKC1 genome yielded 223 contigs with an average GC content of 47%. The draft genome was approximately 3.1 Mbp, and a total of 3,509 CDSs were identified (Table 2-1). *C. maritimus* KKC1 possessed a single copy of 16S and 23S rRNA genes, two 5S rRNA genes (each of which was on different contigs), and a total of 48 tRNA genes coding for all 20 amino acids.

Table 2-1. General features of the genomes from *C. maritimus* KKC1, *M. thermoacetica*, and *M. perchloratireducens*

Parameter	Value for:		
	***C. maritimus* KKC1**	***M. thermoacetica* ATCC 39073 (NC_007644.1)**	***M. perchloratireducens* An10 (2506520025)**
Genome size (bp)	3,064,849	2,628,784	3,307,499
G+C content (%)	47	55.8	53.8
No. of:			
CDSs	3,509	2,463	3,349
rRNAs	4	3	3
tRNAs	48	51	52
No. (%) of genes in COG	2,287 (65.2)	1,953 (79.3)	2,518 (75.2)
No. of contigs	223	1	133
Source	This study	RefSeq	IMG

C. maritimus KKC1 can grow heterotrophically on pyruvate, lactate, fumarate, glucose, fructose, and mannose with thiosulfate as an electron acceptor under an N_2 atmosphere (Yoneda *et al.*, 2013a). Metabolic pathways predicted by KEGG analysis revealed that the *C. maritimus* KKC1 genome encoded a complete glycolytic pathway and an incomplete tricarboxylic acid (TCA) cycle that lacked citrate synthase (present in

Moorella species) and malate dehydrogenase (Fig. 2-1AB). It also possessed one gene encoding a nicotinamide adenine dinucleotide (NAD)-dependent malic enzyme (EC 1.1.1.38) (Fig. 2-1C), which is responsible for linking the TCA cycle to glycolysis by catalyzing the interconversion of malate and pyruvate (Meyer & Stülke, 2013). In addition, *C. maritimus* KKC1 maintained the fructose utilization pathway driven by the phosphoenolpyruvate-dependent phosphotransferase system (Fraenkel & Vinopal, 1973) and l-lactate dehydrogenase (Fig. 2-1D). Therefore, I suggest that when *C. maritimus* KKC1 utilizes lactate, fumarate, and fructose, these compounds are converted into pyruvate. Pathways for mannose metabolism were not predicted by my analysis of the *C. maritimus* KKC1 genome; hence, the underlying mechanism remains unclear. *C. maritimus* KKC1 utilized the WLP for autotrophy, but genes encoding key enzymes for other known carbon fixation pathways, such as RuBisCO and 4-hydroxybutyryl-CoA dehydratase, were not found.

As reported previously, the *C. maritimus* KKC1 is most closely related to *Moorella* species on the basis of 16S rRNA phylogenetic analysis (Yoneda *et al.*, 2013a) (Fig. 2-2A). The *C. maritimus* KKC1 genome showed a relatively low GC content (47%) compared to those of *M. thermoacetica* and *M. perchloratireducens* (55.8 and 53.8%, respectively). I conducted phylogenetic analyses based on five housekeeping genes and the genomic similarity score (GSS), which confirmed that *Moorella* species were the most closely related to *C. maritimus* KKC1 (Fig. 2-2BC). In both phylogenetic trees, the sister group of the *C. maritimus* KKC1 and *Moorella* clades included known hydrogenogenic carboxydotrophs, such as *Carboxydothermus*. This was particularly true of the ML tree of housekeeping genes, as indicated by the high bootstrap replica value.

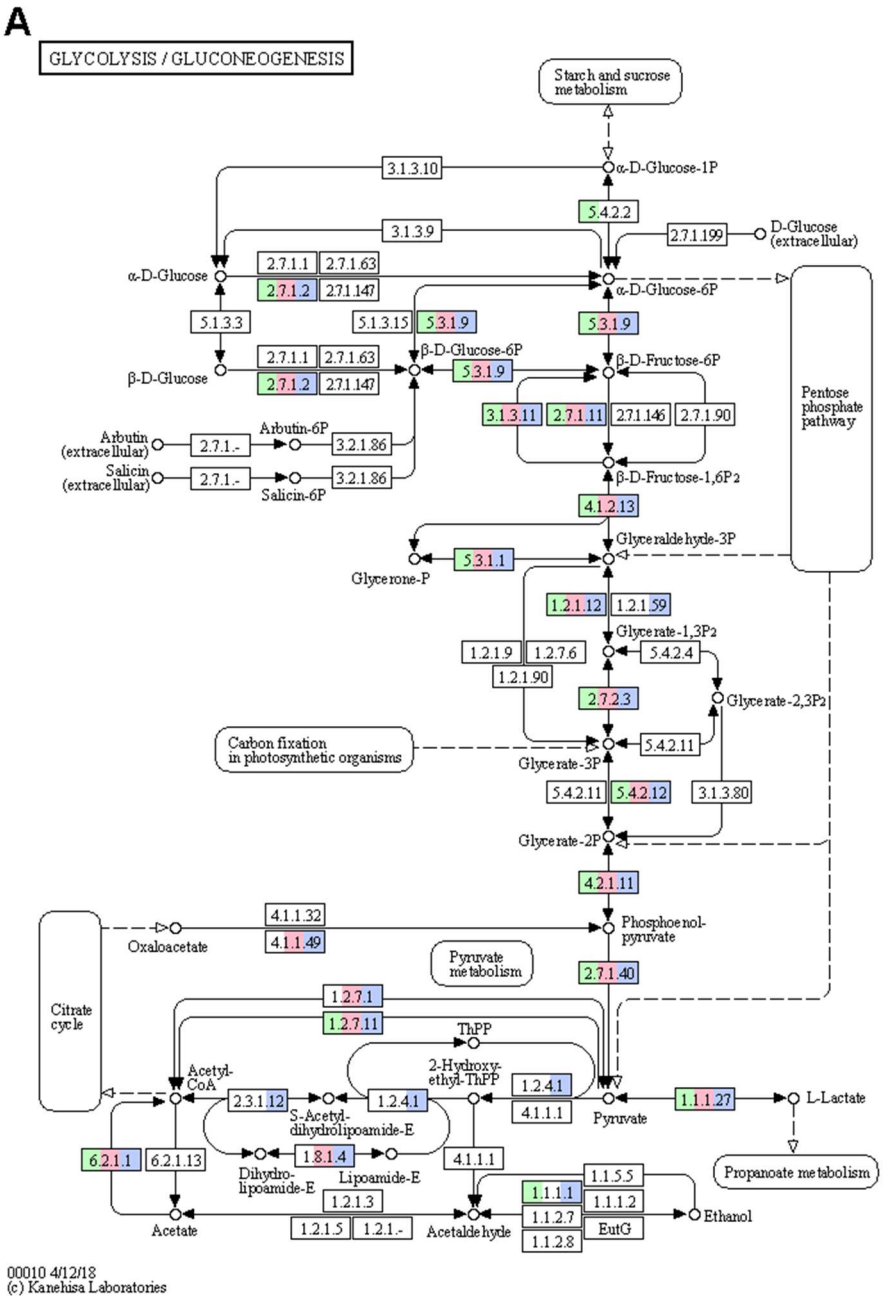


Fig. 2-1. Comparison of metabolic pathways in *C. maritimus* KKC1, *M. thermoacetica*, and *M. perchloratireducens*. (A) glycolysis and gluconeogenesis. (B) TCA cycle. (C) pyruvate metabolism. (D) fructose and mannose metabolism. (E) carbon fixation pathways in prokaryotes. Green, *C. maritimus*; red, *M. thermoacetica*; blue, *M. perchloratireducens*. The empty box indicates that there are no ORFs assigned. Pathway maps were generated by KEGG.

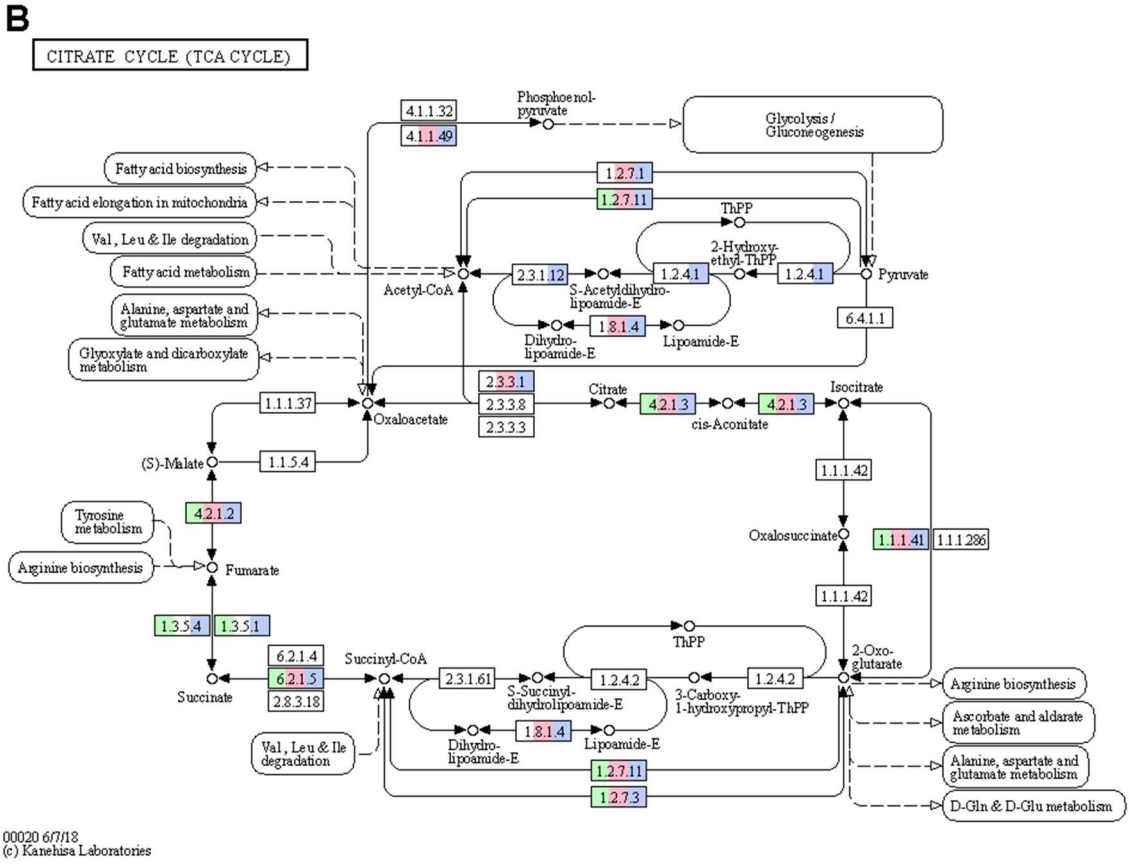


Fig. 2-1 Continued.

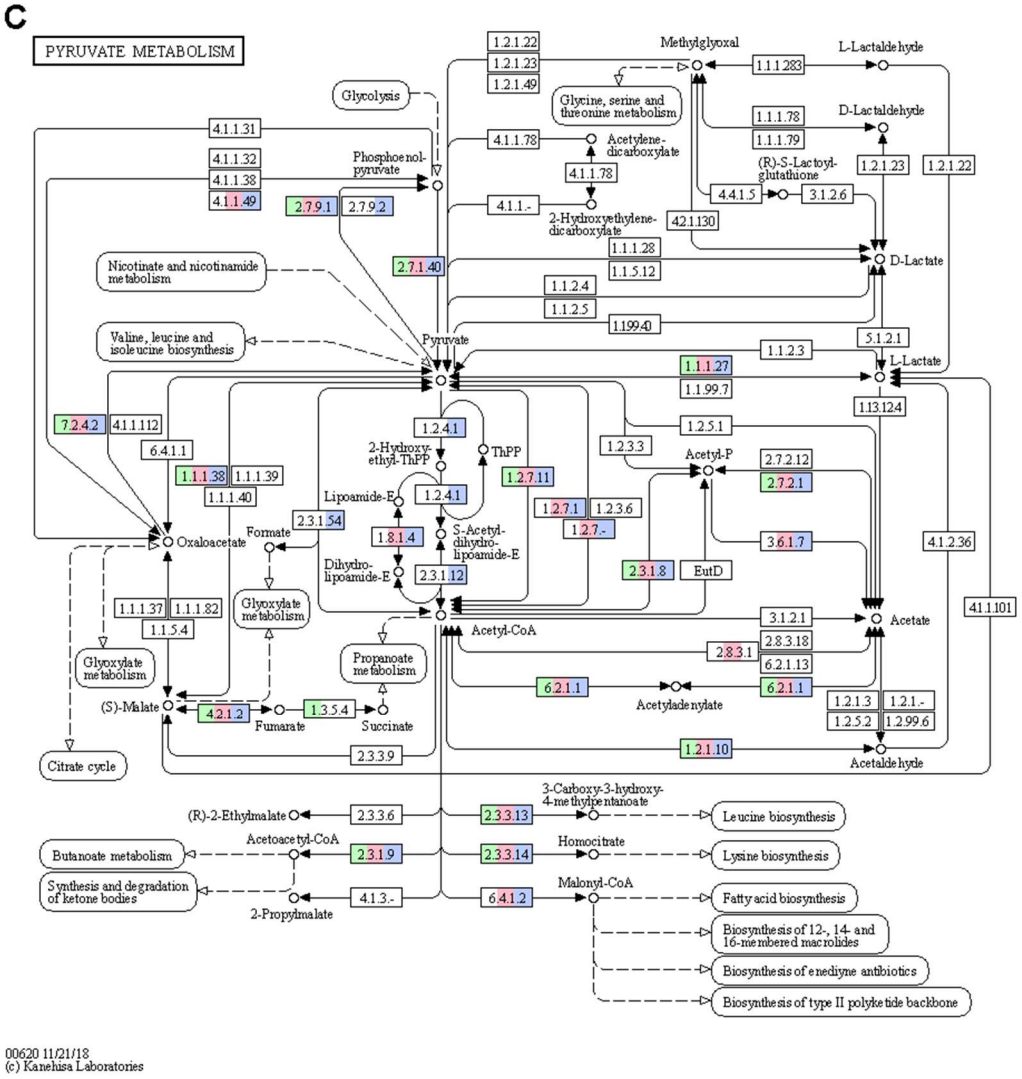


Fig. 2-1 Continued.

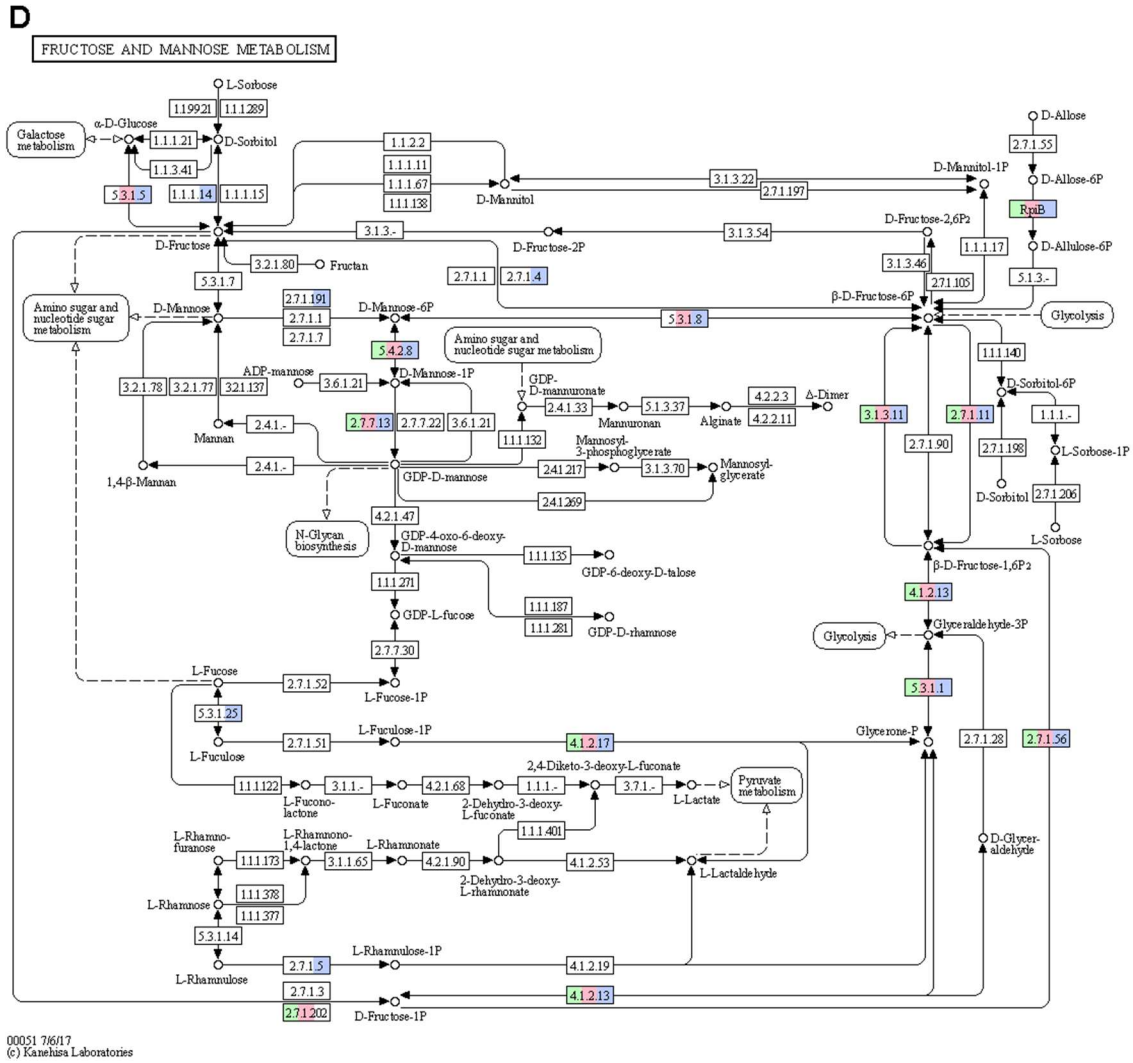


Fig. 2-1 Continued.

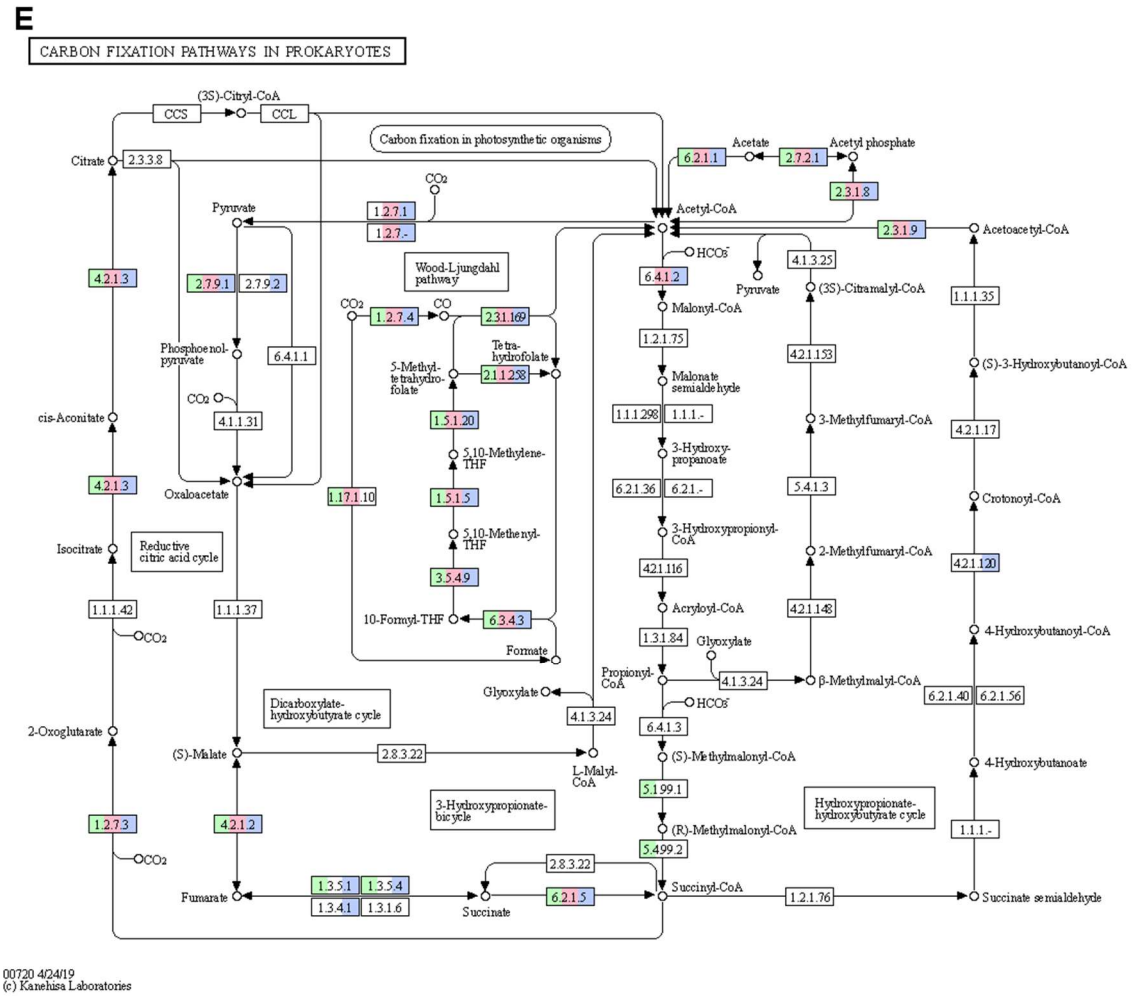


Fig. 2-1 Continued.

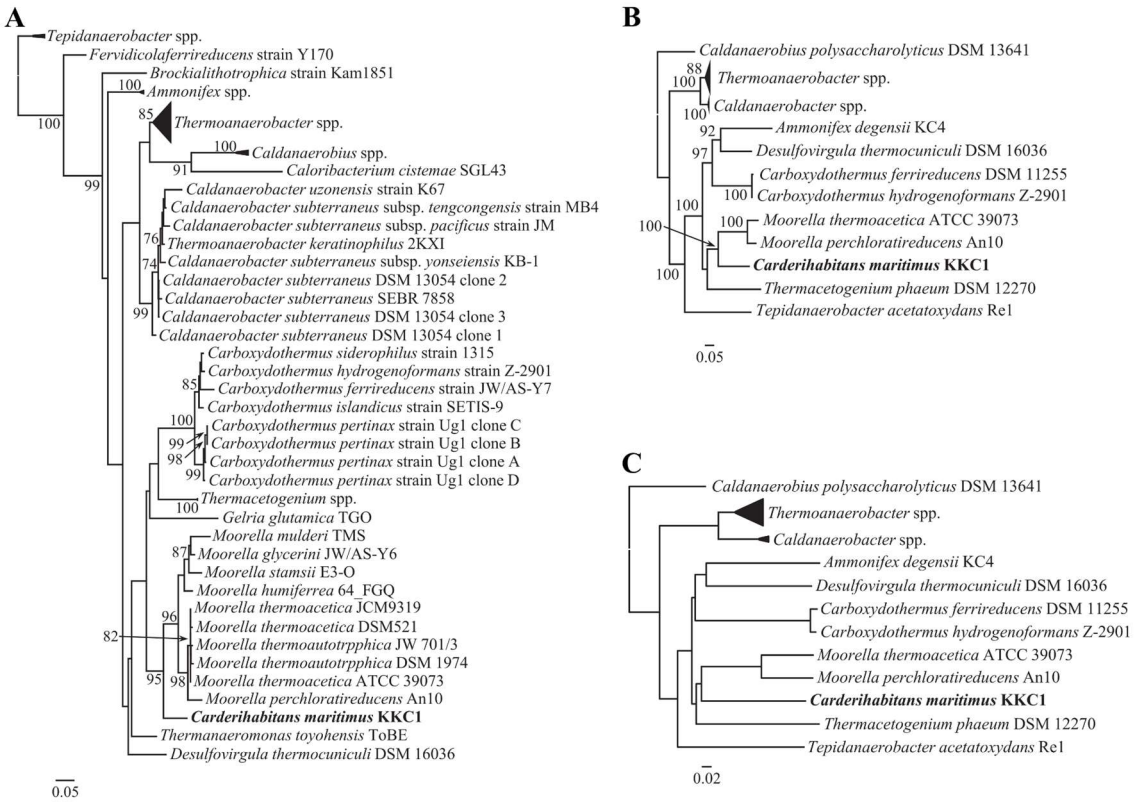


Fig. 2-2. Phylogenetic reconstruction of Thermoanaerobacteraceae. (A) ML phylogenetic analysis using 16S rRNA. (B) ML phylogenetic analysis of five concatenated housekeeping genes. Only bootstrap support values (out of 100 runs) greater than or equal to 70 are shown in both panels A and B. (C) GSS distance matrix plotted as a neighbor-joining tree. *C. maritimus* KKC1 is indicated in bold font.

Genomic comparison between *C. maritimus* KKC1 and *Moorella* species

Recent studies revealed that *M. thermoacetica* is an "ECH-acetogen" (Schuchmann & Müller, 2014) that utilizes two metabolic modules, the CO_2-reducing WLP and the energy-conserving ECH-based module energized by reduced ferredoxin. *M. thermoacetica* possesses HydABC and NfnAB, which catalyzes the endergonic reduction of low-potential ferredoxin with H_2 by flavin-based electron bifurcation (Schuchmann & Müller, 2014). It also possesses pyruvate:ferredoxin oxidoreductases (PFORs) or CODHs that generate reduced ferredoxin, an actual "energy equivalent," by pyruvate or CO

oxidation, respectively (Diender *et al.*, 2015; Schuchmann & Müller, 2014), and can utilize versatile energy sources in acetogenic growth (Pierce *et al.*, 2008). *M. perchloratireducens* can grow on CO, methanol, pyruvate, glucose, fructose, cellobiose, mannose, xylose, and pectin, but no growth is observed on H_2 plus CO_2 (Balk *et al.*, 2008). The products from substrate utilization are acetate, CO_2, and H_2. On the other hand, *C. maritimus* KKC1 is a hydrogenogenic carboxydotroph that can grow on CO with production of H_2 in a medium containing ferric citrate (10 mg/L) as the sole organic compound. Acetogenic growth on H_2 plus CO_2 has not been observed in *C. maritimus* KKC1. It can grow heterotrophically on pyruvate, lactate, fumarate, glucose, fructose, and mannose with thiosulfate as an electron acceptor under an N_2 atmosphere but not without any electron acceptors (Yoneda *et al.*, 2013a). I performed a functional classification of ORFs from *C. maritimus* KKC1, *M. thermoacetica*, and *M. perchloratireducens* by BLAST search against COGs (Fig. 2-3). The number of ORFs assigned to COG categories related to central metabolic pathways (C, energy production and conversion; E, amino acid transport and metabolism; G, carbohydrate transport and metabolism) varied substantially between *C. maritimus* KKC1 and *Moorella* species, as described below.

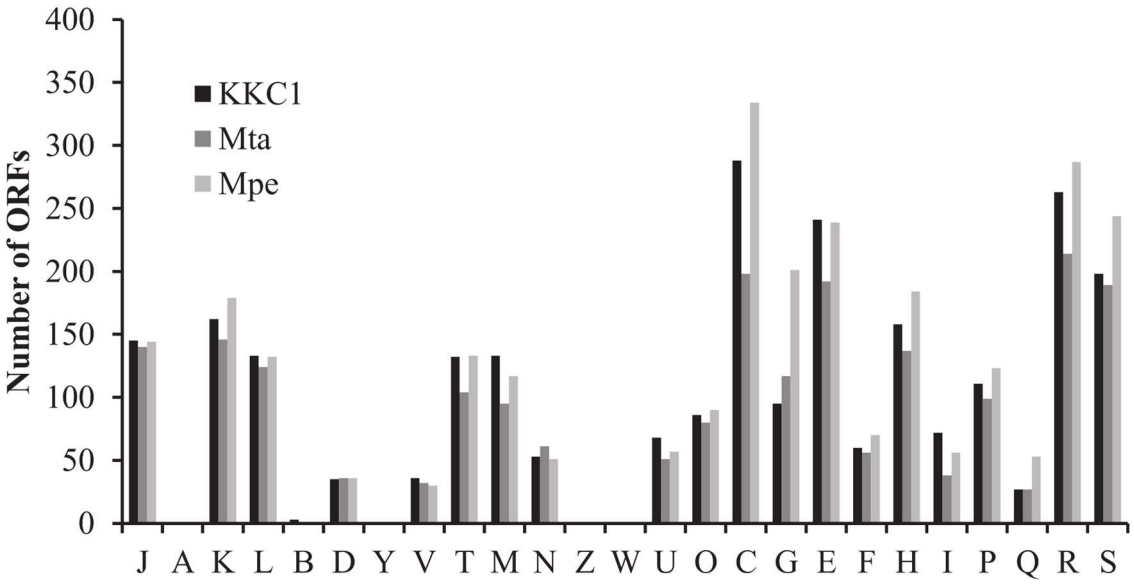


Fig. 2-3. Numbers of ORFs associated with the general COG functional categories. KKC1, *C. maritimus* KKC1 (3,509 ORFs); Mta, *M. thermoacetica* (2,463 ORFs); Mpe, *M. perchloratireducens* (3,349 ORFs). Functional categories: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; Y, nuclear structure; and Z, cytoskeleton

While each *Moorella* species possessed one *hyc*/*hyf*-type ECH gene cluster, the *C. maritimus* KKC1 genome contained two ECH complexes: one *coo*-type ECH (forming CODH–ECH) and one *hyc*/*hyf*-type ECH clustered with a formate dehydrogenase gene (*fdoG*). The structures of the *hyc*/*hyf*-type ECH gene clusters from *C. maritimus* KKC1, *M. thermoacetica*, and *M. perchloratireducens* were very similar, but only the *hyf*/*hyc*-type ECH from *M. perchloratireducens* lacked a formate dehydrogenase gene and

clustered with *cooS* (Fig. 2-4, 2-5). The *C. maritimus* KKC1 genome conserved a complete WLP, like *M. thermoacetica* (although *cooS* within the ACS gene cluster was frameshifted, as mentioned below), while *M. perchloratireducens* lacked formate dehydrogenase (Fdh), which catalyzes the first CO_2 fixation step in the WLP (Fig. 2-1). Unlike for *Moorella* species, no HydABC and NfnAB homologs were found in the *C. maritimus* KKC1 genome, consistent with its failure of acetogenic growth on H_2 and CO_2 (Yoneda *et al.*, 2013a). The authentic PFOR of *M. thermoacetica* is encoded in Moth_0064 and contains three domains, the α, γ, and β subunits, annotated as COG0674, COG1014, and COG1013, respectively (Pierce *et al.*, 2008). A Moth_0064 homolog was found in the *M. perchloratireducens* genome but not in that of *C. maritimus* KKC1 (Table 2-2). Although *C. maritimus* KKC1 possessed six sets of genes annotated as COG0674, COG1014, and COG1013, these were more similar to 2-oxoglutarate (α-ketoglutarate):ferredoxin oxidoreductase (KFOR) than to PFOR, according to KEGG orthology annotation.

Remarkably, *C. maritimus* KKC1 harbored six CooS genes with conserved residues linked to metal clusters in its genome (Dobbek *et al.*, 2001; Ragsdale, 2004) (Fig. 2-6). Functional types of the six CooS genes were affiliated to each of types i, ii, and iii and three of type iv, although *cooS* within the *cooS*-ACS type i was frameshifted. I also detected the simultaneous transcription of all six CooS genes in *C. maritimus* KKC1 during carboxydotrophic growth by reverse transcription-PCR (RT-PCR) (data not shown). Moreover, both *Moorella* species possessed only two *cooS* clusters (Fig. 2-4), one engaged in the WLP (i.e., type i CODH). The other *cooS* clusters were types iv and ii in *M. thermoacetica* and *M. perchloratireducens*, respectively. I discuss the details of the CooSs in the following sections.

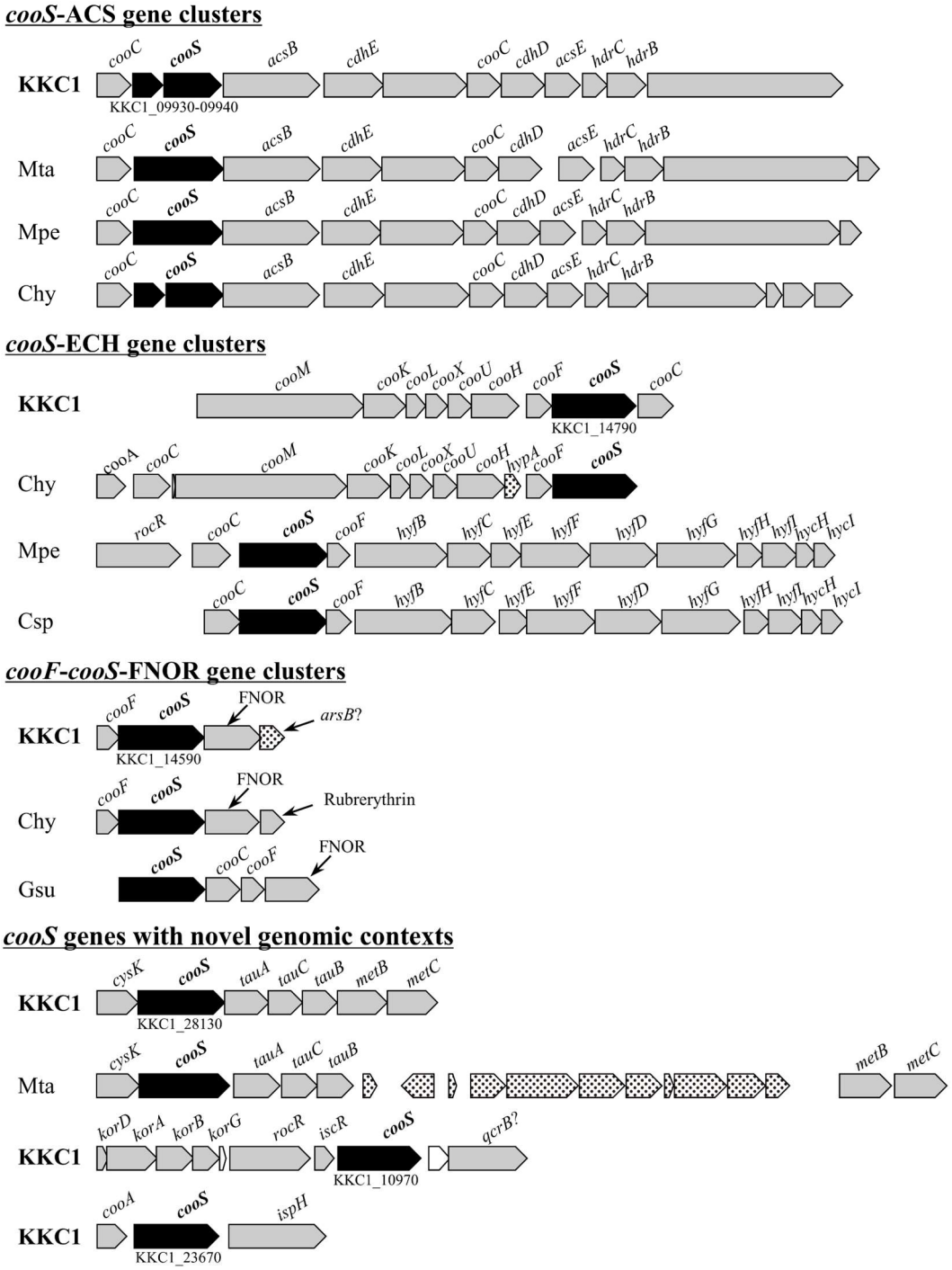


Fig. 2-4. Schematic representation of CooS gene clusters from *C. maritimus* KKC1, *M. thermoacetica*, *M. perchloratireducens*, *C. subterraneus* subsp. *pacificus*, and *C. hydrogenoformans*. KKC1, *C. maritimus*; Mta, *M. thermoacetica*; Mpe, *M. perchloratireducens*; Csp, *C. subterraneus* subsp. *pacificus*; Chy, *C. hydrogenoformans*; Gsu, *Geobacter sulfurreducens*. Black, *cooS*; dots, inserted genes; gray, other functional proteins.

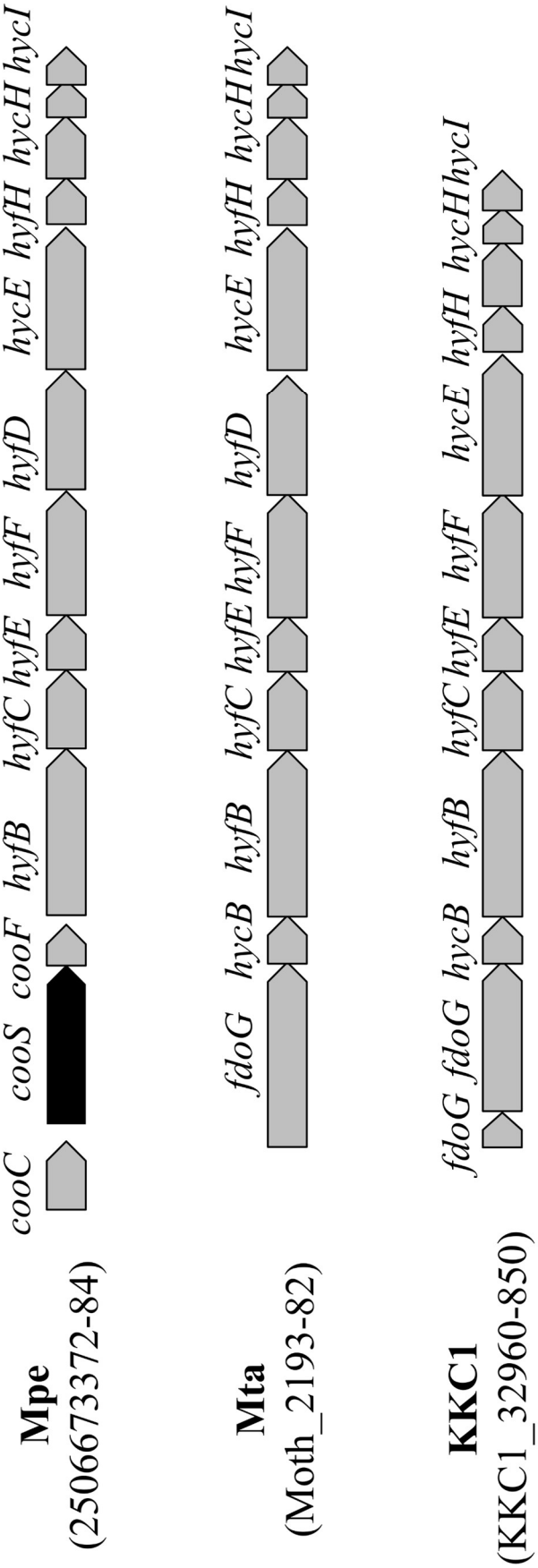


Fig 2-5. Schematic representation of *hyf*/*hyc*-type ECH gene clusters from *C. maritimus* KKC1, *Moorella thermoacetica* and *Moorella perchloratireducens*. KKC1, *C. maritimus*; Mta, *M. thermoacetica*; Mpe, *M. perchloratireducens*. Black, *cooS*; gray, other functional protein.

Table 2-2. CDSs made up of COGs found in pyruvate:ferredoxin oxidoreductase.

KKC1[a]				Mta[a]				Mpe[a]			
Locus tag	COG	K number[b]	Subunit name[b]	Locus tag	COG	K number[b]	Subunit name[b]	Locus tag[c]	COG	K number[b]	Subunit name[b]
KKC1_06220	COG0674	K00174	korA	Moth_0033	COG0674	K00174	korA	2506673660	COG1146	K00176	korD
KKC1_06230	COG1013	K00175	korB	Moth_0034	COG1013	K00175	korB	2506673661	COG1014	K00177	korC
KKC1_06240	COG1014	K00177	korC	Moth_0035	COG1014	K00177	korC	2506673662	COG1013	K00175	korB
				Moth_0036	COG1149	K00176	korD	2506673663	COG0674	K00174	korA
KKC1_10900	COG1146	K00176	korD								
KKC1_10910	COG0674	K00174	korA	Moth_1983	COG1149	K00176	korD	2506674734	COG0674	K00174	korA
KKC1_10920	COG1013	K00175	korB	Moth_1984	COG0674	K00174	korA	2506674735	COG1013	K00175	korB
KKC1_10930	COG1014	K00177	korC	Moth_1985	COG1013	K00175	korB	2506674736	COG1146	K00176	korD
				Moth_1986	COG1014	K00177	korC	2506674737	COG1014	K00177	korC
KKC1_14030	COG0674	K00174	korA								
KKC1_14040	COG1013	K00175	korB					2506675637	COG2221	K00176	korD
KKC1_14050	COG1146	K00176	korD					2506675638	COG0674	K00174	korA
KKC1_14060	COG1014	K00177	korC					2506675639	COG1013	K00175	korB
								2506675640	COG1014	K00177	korC
KKC1_15100	COG1146	K00176	korD								
KKC1_15110	COG0674	K00174	korA								
KKC1_15120	COG1013	K00175	korB								
KKC1_15130	COG1014	K00177	korC								
KKC1_21110	COG0674	K00174	korA								
KKC1_21120	COG1013	K00175	korB								
KKC1_21130	COG1014	K00177	korC								
KKC1_21140	COG1146	K00176	korD								

Table 2-2. Continued

KKC1_27530	COG1146	K00176	korD								
KKC1_27540	COG1014	K00177	korC								
KKC1_27550	COG1013	K00175	korB								
KKC1_27560	COG0674	K00174	korA								
				Moth_0064[d]	COG0674	K00169	porA	2506673623[e]	COG0674	K00169	porA
					COG1014	K00172	porG		COG1014	K00172	porG
					COG1013	K00170	porB		COG1013	K00170	porB
				Moth_0376	COG1014	K00172	porG	2506674258	COG1144	K00172	porG
				Moth_0377	COG1144	K00171	porD	2506674259	COG0674	K00169	porA
				Moth_0378	COG0674	K00169	porA	2506674260	COG1013	K00170	porB
				Moth_0379	COG1013	K00170	porB				
								2506674776	COG1014	K00172	porG
				Moth_1591	COG1013	K00170	porB	2506674777	COG1144	K00171	porD
				Moth_1592	COG0674	K00169	porA	2506674778	COG0674	K00169	porA
				Moth_1593	COG1014	K00172	porG	2506674779	COG1013	K00170	porB
				Moth_1921	COG1013	K00170	porB	2506675171	COG1014	K00172	porG
				Moth_1922	COG0674	K00169	porA	2506675172	COG1013	K00170	porB
				Moth_1923	COG1144	K00171	porD	2506675173	COG0674	K00169	porA
				Moth_1924	COG1014	K00172	porG	2506675174	COG1144	K00171	porD
KKC1_00050	COG1014	K00180	iorB	Moth_0934	COG1014	K00180	iorB	2506674290	COG1014	K00180	iorB
KKC1_16920	COG1014	K00180	iorB	Moth_2276	COG1014	K00180	iorB	2506675897	COG1014	K00180	iorB

[a] KKC1, *Calderihabitans maritimus* KKC1; Mta, *Moorella thermoacetica*; Mpe, *Moorella perchloratireducens*.
[b] Description of KEGG Orthology. Kor, 2-oxoglutarate (α-ketoglutarate):ferredoxin oxidoreductase; por, pyruvate:ferredoxin oxidoreductase; ior, indolepyruvate:ferredoxin oxidoreductase.
[c] Locus tag of *M. perchloratireducens* is represented by IMG Gene ID.
[d] Moth_0064 is a fusion of three COGs, COG0674, COG1014, COG1013
[e] This CDS showed 89% similarlity with Moth_0064.

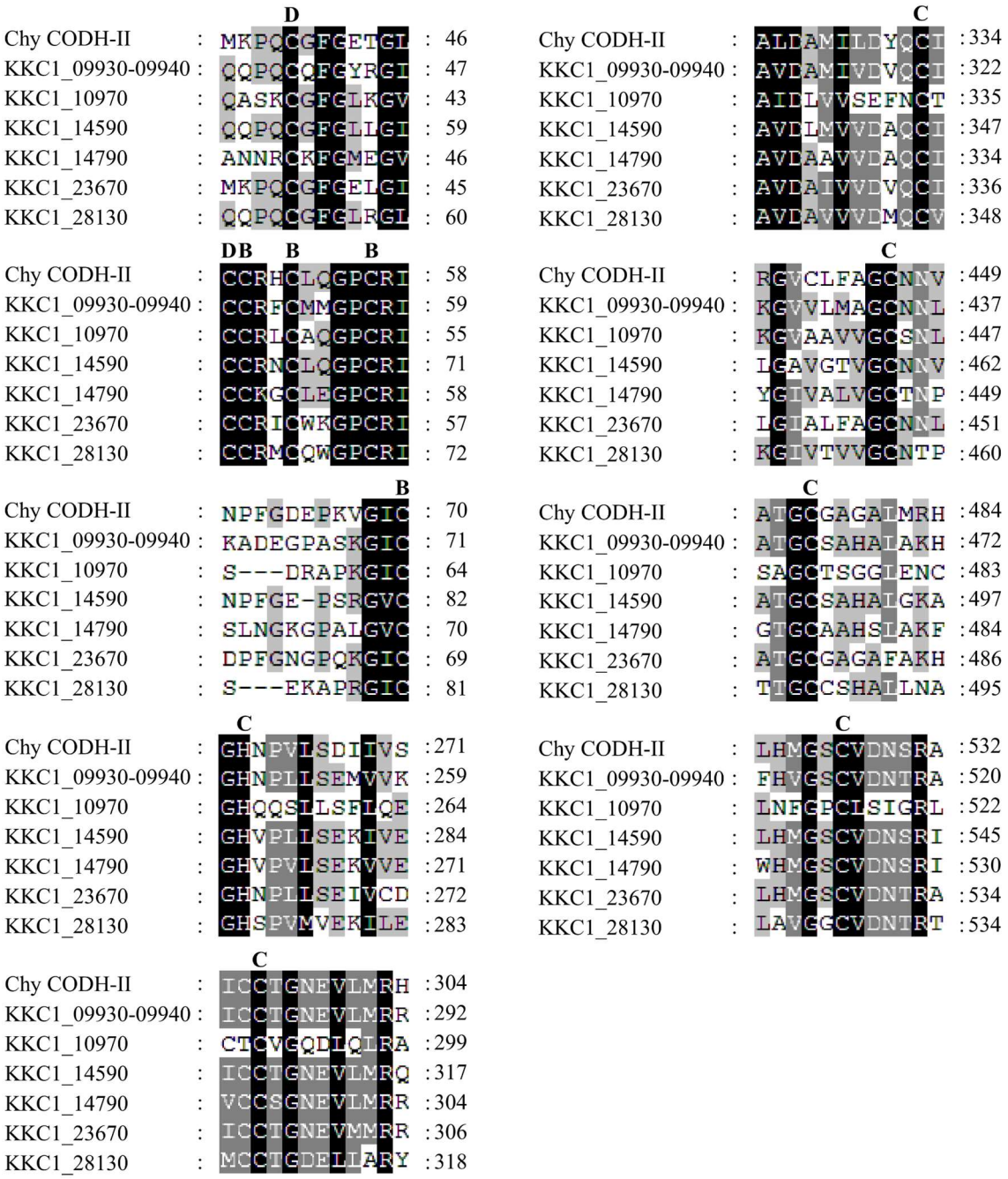


Fig2-6. Alignment of amino acid sequences for CooSs from *C. hydrogenoformans* (CODH-II) and *C. maritimus* KKC1. Conserved residues are indicated with a black background. Residues conserved in >80% and >60% of the proteins examined are indicated with a white type on a dark gray background and by a black type on a light gray background, respectively. Residues linking Clusters B, D, and C are indicated by black type on top.

Genomic contexts of the six CooS genes in *C. maritimus* KKC1

Of the six CooS genes from *C. maritimus* KKC1, three presented already-known genomic contexts in other microorganisms: the type i *cooS*-ACS, type ii *cooS*-ECH, and type iii *cooF-cooS*-FNOR gene clusters. These are almost identical to the *cooS-I*, *III*, and *IV* clusters of *C. hydrogenoformans*, respectively (Wu *et al.*, 2005), but with some variation (Fig. 2-4). The sequence of the *cooS* gene within the *cooS*-ACS gene cluster in *C. maritimus* KKC1 was split into two ORFs (KKC1_09930 and KKC1_09940) owing to a frameshift. The *cooS*-ECH gene cluster of *C. maritimus* KKC1 was a *coo* type. However, unlike in *C. hydrogenoformans*, a homolog of *cooA*, encoding a heme-containing regulator of the *coo* operon, was not found upstream of the ECH gene cluster (*cooMKLXUH*) (Fig. 2-4). The *cooS-IV* gene (type iii) from *C. hydrogenoformans* forms an operon with *cooF* and the genes encoding FAD-NAD oxidoreductase (FNOR) and rubrerythrin-like protein, which is thought to play a role in reactive oxygen species detoxification (Wu *et al.*, 2005). In *C. maritimus* KKC1, *cooS* (KKC1_14590) in the *cooF-cooS*-FNOR gene cluster lacked a gene encoding rubrerythrin (Fig. 2-4). Similar gene clusters consisting of sequential genes putatively coding for CooS, CooF, and FNOR have been found in some sulfate reducers (e.g., *Geobacter sulfurreducens*), thermophilic fermenting bacteria, and Clostridium species (Geelhoed *et al.*, 2016) (Fig. 2-4).

The other three *cooS* genes of *C. maritimus* KKC1 (KKC1_28130, KKC1_10970, and KKC1_23670) were found in novel genomic contexts. The *cooS* gene in KKC1_28130 was associated with those encoding cysteine synthase A (CysK), a putative ABC transport system with domains similar to those of TauABC, cystathionine γ-synthase (MetB), and β-lyase (MetC) (Fig. 2-4). The genomic context of KKC1_28130 was similar to that of type iv *cooS* genes found in *M. thermoacetica* (Fig. 2-4). CysK

catalyzes the formation of L-cysteine and acetate from O-acetyl-L-serine and sulfide (Mino & Ishikawa, 2003). TauABC is required for the utilization of taurine as an organic sulfur source when inorganic sulfur is not available (van der Ploeg *et al.*, 1996). MetB and MetC catalyze consecutive trans-sulfuration reactions in the biosynthesis of methionine (Lill & Mühlenhoff, 2005). Three copies of *cysK* and two copies of TauABC genes were found in the genome of *C. maritimus* KKC1, whereas *metBC* was found only in the proximity of KKC1_28130.

The *cooS* gene in KKC1_10970 clustered with those encoding the KFOR δ, α, β, and γ subunits (KorDABG; KKC1_10900 to KKC1_10930), which are one of the six sets of genes putatively encoding KFOR as described above (Table 2-2), and two putative transcriptional regulators (RocR and IscR; KKC1_10940 and KKC1_10950) (Fig. 2-4). KFOR is a TCA cycle-related enzyme that catalyzes the oxidative decarboxylation of 2-oxoglutarate and the reverse reaction (succinyl-CoA carboxylation) in autotrophic bacteria that fix CO_2 by the reductive TCA (RTCA) cycle (Shiba *et al.*, 1982, 1985).

The *cooS* gene in KKC1_23670 clustered with those encoding CooA (KKC1_23660) and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (IspH; KKC1_23680). In the *C. maritimus* KKC1 genome, KKC1_23660 was the sole *cooA* homolog that conserves the His-82 residue (the axial ligands of the Fe[III] and Fe[II] hemes) in CooA from *C. hydrogenoformans* (Inagaki *et al.*, 2005). When searching the upstream regions of CooS genes, I identified CooA-binding sites (5′-TGTCA-N_6-CGACA) previously reported in *R. rubrum* (He *et al.*, 1996), 95-bp downstream of the CooA gene (KKC1_23660) and 85-bp upstream of the *cooS*-ECH gene cluster (KKC1_14720-800). IspH catalyzes the terminal step of the nonmevalonate route, a biosynthetic pathway for isopentenyl diphosphate and dimethylallyl diphosphate, which

are universal precursors for all isoprenoids or terpenes (e.g., steroids and carotenoids) in living organisms (Adam *et al.*, 2002; Seemann *et al.*, 2006). In particular, quinones in the electron transport chain, such as ubiquinone and menaquinone, or polyprenols, including the carbohydrate carrier bactoprenol from eubacteria, represent ubiquitous bacterial isoprenoids (Rohmer, 1999).

Phylogenetic analysis of CooSs

Comprehensive phylogenetic analysis of CooS genes revealed the presence of six distinct clades (Techtmann *et al.*, 2012). Following previously described criteria (Techtmann *et al.*, 2012), CooSs encoded in KKC1_09930-40 (from *cooS*-ACS), KKC1_14790 (from *cooS*-ECH), KKC1_14590 (from *cooF*-*cooS*-FNOR), and KKC1_23670 (in the proximity of *cooA* and *ispH*) were classified as clade F (Fig. 2-7). In contrast, CooSs encoded in KKC1_28130 (in the proximity of *cysK*, *tauACB*, and *metBC*) and KKC1_10970 (in the proximity of *korDABG*) were classified as clade B and clade C, respectively.

Clade F CooSs encoded within the *cooS*-ACS, *cooS*-ECH, and *cooF*-*cooS*-FNOR gene clusters from *C. maritimus* KKC1 showed 71%, 82%, and 68% identity with respect to their counterparts in *Desulfotomaculum kuznetsovii* (WP_013822590.1), *Thermosinus carboxydivorans* (WP_007288856.1), and *Thermincola potens* (WP_013120796.1), respectively, and formed subclades with each one from *C. hydrogenoformans*. However, the CooS encoded in KKC1_23670 did not form a subclade with Thermoanaerobacteraceae and instead formed a subclade together with *Thermodesulfobacterium* (phylum Thermodesulfobacteria), *Desulfotomaculum*, *Desulfosporosinus* (order Clostridiales, phylum Firmicutes), *Desulfurispirillum* (phylum

Chrysiogenetes), *Paenibacillus* (class Bacilli, phylum Firmicutes), and *Pelosinus* (class Negativicutes, phylum Firmicutes). The clade B CooS (encoded in KKC1_28130) was phylogenetically close to the type iv *cooS* of *M. thermoacetica* (76% identity), which presented a similar genomic context (Fig. 2-4), forming the most deeply branched members of clade B (Fig. 2-7). Clade C CooS (encoded in KKC1_10970) had 70% identity with counterparts from *Desulfotomaculum acetoxidans* (WP_015758381.1). Both CooSs were phylogenetically distinct from those from members of Thermoanaerobacteraceae.

Type i CooS genes from *M. thermoacetica* and *M. perchloratireducens* clustered in the same subclade in clade F, which is constituted with only type i CooS genes (Fig. 2-7). The type ii CooS gene (IMG Gene ID 2506673373) clustered with the *hyf*/*hyc*-type ECH gene cluster in *M. perchloratireducens* was phylogenetically distinct from those of *M. thermoacetica* or *C. maritimus* KKC1 but formed the same subclade with type iii *cooS-II* (clade F) from *C. hydrogenoformans* (Fig. 2-7).

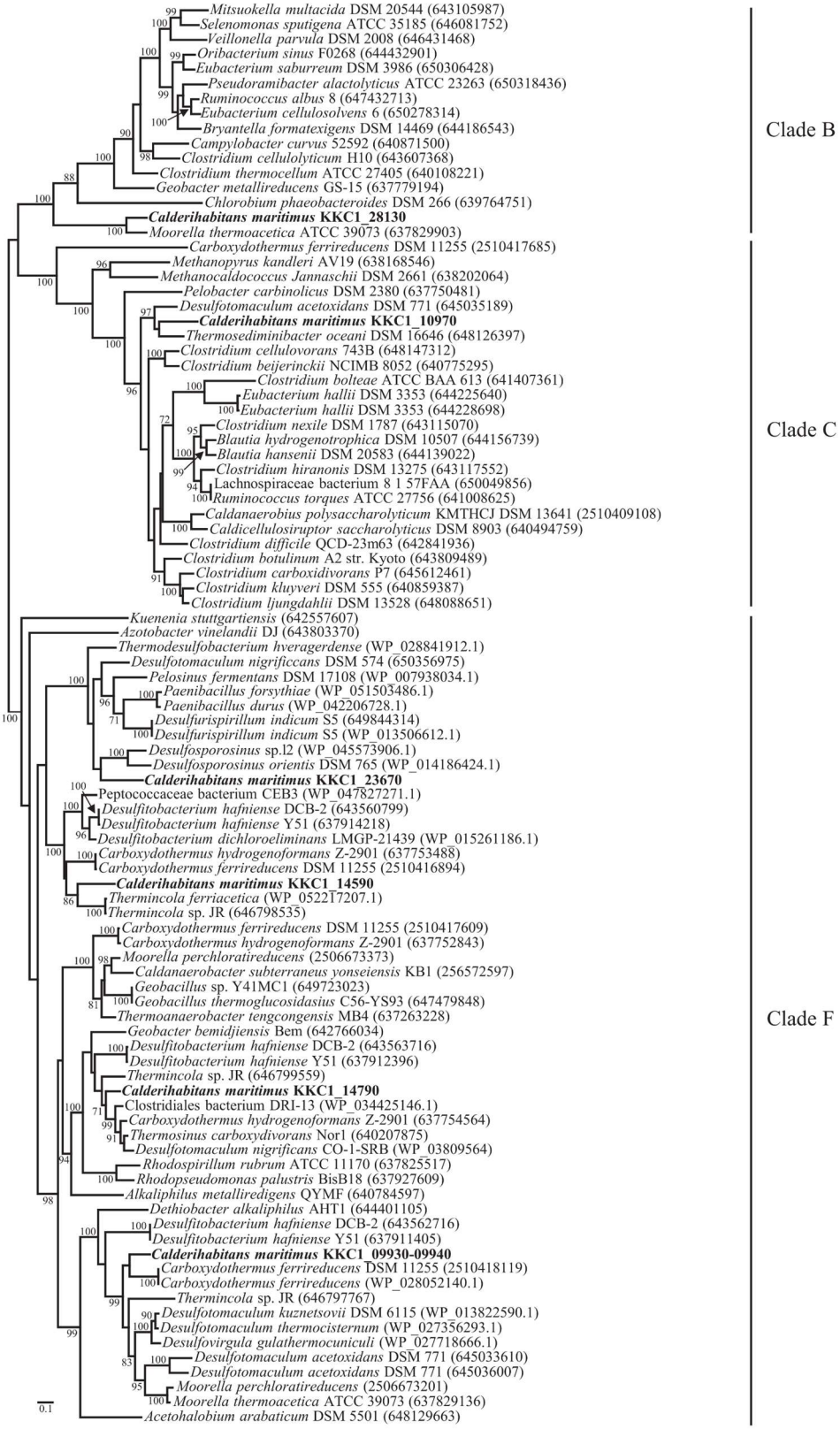


Fig. 2-7. ML phylogenetic tree of CooSs. CooSs from *C. maritimus* KKC1 are indicated in bold font. Only bootstrap support values (out of 100 runs) equal to or greater than 70 are shown.

Horizontal gene transfer analysis of six CooSs from *C. maritimus* KKC1

To determine whether *cooS* was obtained by horizontal gene transfer, I performed a simple test for sequence composition (Fig. 2-8). In this test, I calculated Euclidean distances between CDS tetranucleotide frequencies and the whole genome and evaluated the significance of distances of CooSs. As a general rule, horizontally transferred DNA fragments exhibit the oligonucleotide composition of the species they are derived from, and the screening of local variations of oligonucleotide composition along genomes is expected to reveal regions of interest where horizontally transferred genes might be located (Dufraigne *et al.*, 2005). A study predicted that the average proportion of horizontally transferred genes per genome was ~12% of all CDSs, ranging from 0.5% to 25% depending on the prokaryotic lineage (11% in *Bacillus subtilis* 168 [Firmicutes]) (Nakamura *et al.*, 2004). Therefore, I used 75% (corresponding to a distance of 0.03024) as a loose threshold for the detection of horizontally transferred CooSs. Accordingly, the distances of four *cooS* genes, KKC1_14790 (from *cooS*-ECH), KKC1_28130 (in the proximity of *cysK*, *tauACB*, and *metBC*), KKC1_10970 (in the proximity of *korDABG*), and KKC1_23670 (in the proximity of *cooA* and *ispH*), to the *C. maritimus* KKC1 genome were from 0.02004 to 0.02277, whereas the other two *cooS* genes, KKC1_09930-40 (from *cooS*-ACS) and KKC1_14590 (from *cooF-cooS*-FNOR), showed slightly higher values (0.0290 and 0.0279, respectively). However, distance values for all *cooS* genes from *C. maritimus* KKC1 were below the threshold, suggesting that all *cooS* genes descended from a common ancestor.

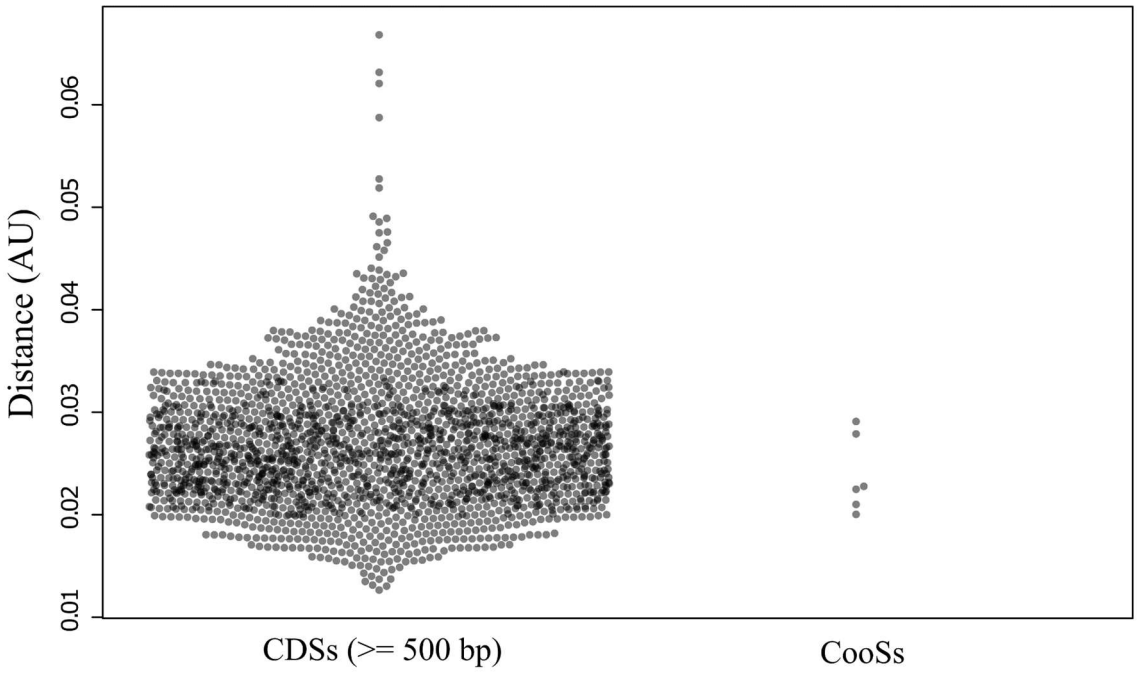


Fig. 2-8. Distances of CDS tetranucleotide frequency to that of the whole *C. maritimus* KKC1 genome. Left column: CDSs greater equal than 500 bp. Right column: six *cooS* genes (KKC1_09930-40, KKC1_10970, KKC1_14590, KKC1_14790, KKC1_23670, KKC1_28130). Distances are expressed in arbitrary units (AU).

Discussion

The similar branching pattern observed by phylogenetic analyses of 16S rRNA, five housekeeping genes, and GSS (Fig. 2-2) indicates that the thermophilic, hydrogenogenic carboxydotroph *C. maritimus* KKC1 and members of the genus *Moorella*, one of the most studied groups of acetogenic bacteria, evolved from a common ancestor. Both *M. thermoacetica* and *C. maritimus* KKC1 possessed the CO_2-reducing WLP and the energy-conserving ECH-based module energized by reduced ferredoxin. However, in contrast to *M. thermoacetica*, *C. maritimus* KKC1 lacked the electron-bifurcating enzymes HydABC and NfnAB. HydABC couples the simultaneous endergonic reduction of ferredoxin with

H_2 to the exergonic reduction of NAD^+ with H_2 (Schuchmann & Müller, 2014), and NfnAB catalyzes the reduction of two $NADP^+$ molecules with one NADH and one reduced ferredoxin to generate two NADPH molecules, which are required for the reduction of CO_2 to acetate in *M. thermoacetica* (Huang *et al.*, 2012). Because the potential of CO is lower than that of ferredoxin, reduction of ferredoxin by oxidation of CO may not need electron bifurcation in *C. maritimus* KKC1. The frameshift mutation in *cooS* within *cooS*-ACS, which has been reported in *C. hydrogenoformans* (Wu *et al.*, 2005), was also found in *C. maritimus* KKC1. Even so, *C. maritimus* KKC1 is known to produce a small amount of acetate during hydrogenogenic growth under a CO atmosphere (Yoneda *et al.*, 2013a). According to a study by Svetlitchnyi and colleagues (Svetlitchnyi *et al.*, 2004) which suggests that CooS may be unnecessary for operation of the WLP, the frameshift mutation in *cooS* (KKC1_09930-40) within the *cooS*-ACS gene cluster might not affect pathway function.

M. thermoacetica is able to catalyze the near-stoichiometric conversion of glucose to 3 mol of acetate using PFOR, which couples the glycolytic pathway to the WLP (Drake & Daniel, 2004; Furdui & Ragsdale, 2000). In contrast, *C. maritimus* KKC1 cannot grow on glucose (and other organic compounds) without electron acceptors, but it can grow with electron acceptors such as thiosulfate, resulting in a small amount of acetate (Yoneda *et al.*, 2013a). Genomic analysis of *C. maritimus* KKC1 revealed that it lacks genes encoding authentic PFOR, which is conserved across *Moorella* species. From a thermodynamic perspective, acetogenesis from glucose is less effective in supporting growth than anaerobic respiration using electron acceptors except for CO_2 (Drake & Daniel, 2004). Therefore, it is assumed that the lack of authentic PFOR in *C. maritimus* KKC1 might result in a survival strategy different from that of *M. thermoacetica*, which

can thrive where no electron acceptors (except for CO_2) are available. The small production of acetate during heterotrophic growth with thiosulfate by *C. maritimus* KKC1 might be explained by the presence of six gene sets encoding putative KFORs, because KFOR shows significant similarity with PFOR and some KFORs show broad specificity for pyruvate and 2-oxoglutarate (Fukuda & Wakagi, 2002). In this case, why *C. maritimus* KKC1 cannot grow acetogenically on glucose without electron acceptors using KFORs is unknown. One possibility is that the reaction efficiency of KFOR is lower than that of PFOR in oxidation of pyruvate, but further studies are needed to understand the mechanism of heterotrophic growth of *C. maritimus* KKC1.

The highest number of *cooS* genes ever reported in a single genome is five (*cooS-I* to *-V*) in *C. hydrogenoformans* (Wu *et al.*, 2005). Thus, *C. maritimus* KKC1 harboring six CooS genes (five *cooS* genes conserving all residues linked to metal clusters (Inoue *et al.*, 2013) and one frameshifted *cooS* within the *cooS*-ACS gene cluster) possessed the most CooS genes of microorganisms with sequenced genomes. As described above, the simultaneous transcription of five *cooS* genes in *C. maritimus* KKC1 during carboxydotrophic growth was observed, and all might contribute to its CO metabolism. Three of the six *cooS* genes formed *cooS*-ACS (type i), *cooS*-ECH (type ii), and *cooF-cooS*-FNOR (type iii) gene clusters. On the other hand, the other three type iv *cooS* genes (KKC1_28130, KKC1_10970, and KKC1_23670) exhibit an uncharacterized genomic context. Although the *cooS* gene in KKC1_23670 clustered with a CooA homolog, KKC1_28130 and KKC1_10970 were not flanked by any genes with obvious roles in CO-related processes. However, the genomic context of *cooS* in KKC1_10970 is interesting, because the KFOR encoded upstream of KKC1_10970 is a redox enzyme that requires ferredoxin and produces (or consumes) CO_2. Therefore, an interaction between

CooS and KFOR could result in a novel CO fixation pathway where CooS oxidizes CO to produce CO_2 and reduced ferredoxin, which could then be used to produce 2-oxoglutarate by KFOR. Sequence composition analysis of the six *cooS* genes from *C. maritimus* KKC1 showed that their distances to the whole genome were not exceedingly high (Fig. 2-8), suggesting that they could descend from a common ancestor.

M. perchloratireducens is phylogenetically and physiologically similar to *M. thermoacetica* but cannot grow acetogenically on H_2 plus CO_2, unlike *M. thermoacetica* (Balk *et al.*, 2008). This might be explained by the lack of formate dehydrogenase (Fdh), which fixes CO_2 to formate in the first step of the WLP. Although hydrogenogenic carboxydotrophic growth has never been reported, *M. perchloratireducens* possessed a *cooS-hyf*/*hyc*-type ECH gene cluster that might form from replacement of the *fdoG-hycB* by *cooC-cooS-cooF* in *hyf*/*hyc*-type ECH gene clusters conserved in *M. thermoacetica* and *C. maritimus* KKC1 (Fig. 2-5). It appears that the assembly of the CooS gene and *hyf*/*hyc*-type ECH gene cluster might occur in the course of *M. perchloratireducens* evolution to efficiently generate energy by CO oxidation and proton translocation with hydrogen production. In addition, the origin of the CooS gene from *cooS-hyf*/*hyc*-type ECH might be the same as for type iii *cooS-II* from *C. hydrogenoformans*, implying that the common ancestor of *Moorella* species and *C. maritimus* KKC1 may have harbored a *cooS-II* homolog, which might have been lost by *C. maritimus* KKC1 and *M. thermoacetica* during evolution, while *M. perchloratireducens* retained it in the *cooS-hyf*/*hyc*-type ECH.

In conclusion, *de novo* genome sequencing and analysis of the hydrogenogenic carboxydotroph *C. maritimus* KKC1 revealed its genomic contents largely different from those of acetogenic *Moorella* species, despite their phylogenetic similarity. Both species

utilize energy-converting ECH-based modules that require the low-potential electron carrier ferredoxin. The lack of bifurcating enzymes and authentic PFOR and the presence of six copies of *cooS* genes in the *C. maritimus* KKC1 genome suggested that the organism may be highly dependent on CO as an electron donor, which can directly reduce ferredoxin, and more adaptive to carboxydotrophic growth than the acetogenic growth observed in *Moorella* species. Remarkably, *C. maritimus* KKC1 harbored the novel CODH gene cluster which might be responsible for CO-derived carbon fixation via incomplete RTCA cycle and represent a novel CO metabolism. Therefore, the *C. maritimus* KKC1 genome might reveal its survival strategy of reliance on the energy-rich substrate CO, whereas the genomes of *Moorella* species show an adaptation at the thermodynamic limit (Schuchmann & Müller, 2014). Thus, *C. maritimus* KKC1 may serve as a model for understanding the evolution and adaptation of CO metabolism.

In this Chapter, I revealed the novel CO metabolism in the novel hydrogenogenic carboxydotrophic isolate from marine sediment, *C. maritimus* KKC1, which implies that the exploration of unknown hydrogenogenic carboxydotrophs in environments is important for understanding the diversity of CO metabolisms of these microorganisms. In Chapter 3 and 4, I developed molecular ecological techniques for exploration of hydrogenogenic carboxydotrophs in environments.

Chapter 3

Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs revealed by microbial community analysis in sediments from multiple hydrothermal environments in Japan

Summary

In hydrothermal environments, carbon monoxide (CO) utilization by thermophilic hydrogenogenic carboxydotrophs may play an important role in microbial ecology by reducing toxic levels of CO and providing H_2 for fueling microbial communities. I evaluated thermophilic hydrogenogenic carboxydotrophs by microbial community analysis. First, I analyzed the correlation between carbon monoxide dehydrogenase (CODH)–energy-converting hydrogenase (ECH) gene cluster and taxonomic affiliation by surveying an increasing genomic database. I identified 71 genome-encoded CODH–ECH gene clusters, including 46 whose owners were not reported as hydrogenogenic carboxydotrophs. I identified 13 phylotypes showing > 98.7% identity with these taxa as potential hydrogenogenic carboxydotrophs in hot springs. Of these, Firmicutes phylotypes such as *Parageobacillus*, *Carboxydocella*, *Caldanaerobacter*, and *Carboxydothermus* were found in different environmental conditions and distinct microbial communities. The relative abundance of the potential thermophilic hydrogenogenic carboxydotrophs was low. Most of them did not show any symbiotic networks with other microorganisms, implying that their metabolic activities might be low.

3. Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs

Introduction

Hydrothermal systems, where geothermally heated water is expelled through fissures in the Earth's crust, are located both on land and under the sea. It is now well known that a wide variety of microorganisms, called thermophiles or hyperthermophiles, can prevail and even thrive in such high-temperature environments. The pioneering studies by Brock and his colleagues (Bott & Brock, 1969; Brock, 1967; Brock & Darland, 1970) at the Yellowstone National Park hot springs established that these organisms grow at near boiling temperatures. Furthermore, a research study led by Pace using molecular phylogenetic techniques demonstrated the high abundance of unidentified thermophilic bacteria and archaea and their remarkable phylogenetic diversity in pink filaments and sediments in the same area (Barns *et al.*, 1994, 1996; Reysenbach *et al.*, 1994).

In recent years, microorganisms that can utilize carbon monoxide (CO) have been found from the hydrothermal area. To date, the list of known thermophilic anaerobic CO-utilizing microorganisms includes acetogenic bacteria (*Moorella thermoacetica*, for instance), sulfate-reducing bacteria (*Desulfotomaculum carboxydivorans*), methanogenic archaea (*Methanothermobacter thermautotrophicus*), and hydrogenogenic bacteria as well as various archaea, such as *Carboxydothermus hydrogenoformans*, *Thermosinus carboxydivorans*, and *Thermococcus* AM4 (Techtmann *et al.*, 2009). Of these, hydrogenogenic bacteria and archaea (collectively designated thermophilic hydrogenogenic carboxydotrophs) are thought to play a key ecological role by virtue of providing a 'safety valve' for reducing toxic levels of CO and supplying H_2 for fueling H_2-dependent microbial community processes (Techtmann *et al.*, 2009). In general, the ability of hydrogenogenic carboxydotrophy is linked to the presence of CO dehydrogenase (CODH)–energy-converting hydrogenase (ECH) gene cluster in genomes.

This cluster is believed to be horizontally transferred between the representatives of separate taxa (Techtmann *et al.*, 2012). So far, 28 phylogenetically diverse thermophilic anaerobic hydrogenogenic CO-utilizing archaea and bacteria have been reported (Sokolova *et al.*, 2009). Most of them (23 species) are members of the phylum Firmicutes.

In addition to their basic isolation and identification, there are several ecological studies on thermophilic hydrogenogenic carboxydotrophs (Brady *et al.*, 2015; Kochetkova *et al.*, 2011; Yoneda *et al.*, 2015). Notably, a radio isotopic study suggests that the majority of CO is oxidized to CO_2 (120 μmol/L of sediment per day) by microbial activities in the hot springs of Uzon Caldera (Kamchatka) (Kochetkova *et al.*, 2011). Thermophilic hydrogenogenic carboxydotrophs of the genera *Carboxydocella* and *Dictyoglomus* have also been isolated from the same environment (Kochetkova *et al.*, 2011). A quantitative polymerase chain reaction (qPCR) analysis targeting the CODH gene, which encodes a key enzyme involved in CO oxidation, suggests that the *Carboxydothermus* species, which is the most studied thermophilic carboxydotrophic species, is distributed in a wide range of hydrothermal environments despite its relatively low population size (≤ 0.000795% of the total bacterial population) (Yoneda *et al.*, 2015). In addition, using the stable isotope probing (SIP) method by ^{13}CO DNA, *Thermincola*, *Desulfotomaculum*, and *Carboxydocella* species were all detected and enriched at geothermal sites, although they are present at < 1% in the original communities (Brady *et al.*, 2015). While there is evidence for the temporal dominance of the *Carboxydothermus* species (~10% of bacterial population) in hydrothermal environments (Yoneda *et al.*, 2013b), thermophilic hydrogenogenic carboxydotrophs are generally considered to occur in low abundance in the environments.

However, these ecological studies on thermophilic hydrogenogenic

carboxydotrophs had a few limitations. Because the sequences of CODH genes are highly diverse, it was difficult to design universal primers that could amplify a wide range of CODH genes from different taxa (Yoneda *et al.*, 2013b). SIP is effective for identifying CO-utilizing microorganisms in the environment (Brady *et al.*, 2015); however, cultivation bias could be observed. In addition, the previous CO-SIP study was limited to a few neutral pH hot springs (Brady *et al.*, 2015). On the other hand, 16S metagenomics is a culture-independent and high-throughput technique, which is applicable for exploring diverse thermophilic hydrogenogenic carboxydotrophs and co-occurring microorganisms. The number of available microbial genome sequences has vastly increased thanks to recent advances in next-generation sequencing technology, using which CODH genes were detected in some species that had never been reported to show hydrogenogenic carboxydotrophic growth (Inoue *et al.*, 2019a; Mohr *et al.*, 2018). However, the correlation between the presence of CODH–ECH gene cluster and taxonomic affiliation has not been well understood. Here, I performed a comprehensive survey of a current prokaryotic genomic database and revealed the phylogenetic distribution of CODH–ECH gene clusters across prokaryotes. Next, I performed 16S rRNA gene amplicon (V3/V4 region) sequencing analysis on 100 sediment samples from a wide variety of hydrothermal and mesophilic environments in Japan and unveiled the distribution patterns of these “potential hydrogenogenic carboxydotrophs”.

Materials and Methods

Sample collection and DNA extraction

I collected a total of 100 sediment samples (17.5–99.0 °C; pH 2.2–8.9; oxidation-

reduction potential [ORP] −262 ~ +449 mV) from terrestrial hydrothermal and mesophilic environments in Japan from May 2014 to March 2017 (Tables 3-1, 3-2). The sampling sites included 76 on southern Kyushu Island (Kagoshima prefecture), 14 on Northern Kyushu Island (Oita prefecture), five on the eastern Izu peninsula (Shizuoka prefecture), and five on the southern Izu peninsula (Shizuoka prefecture). At the Unagi-onsen hot spring (southern Kyushu Island), I collected a total of 65 samples in May 2014, May 2015, November 2015, and December 2016 as a previous study suggested that *Carboxydothermus* species are abundant in this environment (Yoneda *et al.*, 2013b). In addition, we previously isolated the *Carboxydocella* strains ULO1 and JDF658 at Unagi-ike lake and the Jiunji-onsen hot spring, respectively (Fukuyama *et al.*, 2017). Temperature was measured using a TX10 digital thermometer (Yokogawa, Tokyo, Japan) with a type K temperature probe (Yokogawa, Tokyo, Japan) at each sampling site. The pH and ORP of the sediment pore water were measured using an HM-31P portable pH meter (DKK-TOA, Tokyo, Japan) with pH (GST-2729C; DKK-TOA, Tokyo, Japan) or ORP (PST-2729C; DKK-TOA, Tokyo, Japan) electrodes. Sediment samples were collected using 50 mL plastic tubes filled with pore water, put into plastic bags with AnaeroPouch-Anaero (Mitsubishi Gas Chemical, Tokyo, Japan), and immediately sealed to minimise contact with oxygen. The samples were then packed in a cooler box with ice, transported to the laboratory, and stored at –80 °C until use. DNA was extracted from 0.5 g of sediment material using an Extrap Soil DNA Kit Plus ver. 2 (Nippon Steel and SUMIKIN Eco-Tech, Tokyo, Japan) following the manufacturer's instructions. During the homogenizing step, I used a bead beater-type homogenizer, Beads Crusher μT-12 (Taitec, Koshigaya, Japan), at a speed of 3200 r min^{-1} for 60 s. The extracted DNA was stored at –30 °C until use.

Table 3-1. Sediment sample description

Sample ID	Location	Latitude, Longitude	Sampling date (yyyy-mm)	Depth (cm)
Kyushu Island, Japan (hot spring sediments)				
1405_UN_A1_D	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	4.5
1405_UN_A1_M	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	2.5
1405_UN_A1_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 1.5
1405_UN_A2_D	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	4.5
1405_UN_A2_M	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	2.5
1405_UN_A2_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 1.5
1405_UN_B1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1405_UN_B2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1405_UN_B3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1405_UN_B4	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1405_UN_B7	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1405_UN_B8	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1405_UN_R2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2014-05	< 3
1505_UN_A1_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B4	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B6	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B7	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_B8	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_C1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_C2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_C3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3

Table 3-1. Continued

1505_UN_C4	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_C5	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_R1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1505_UN_R2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-05	< 3
1511_UN_A1_C	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	40
1511_UN_A1_D	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	9
1511_UN_A1_M	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	6
1511_UN_A1_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_A2_C	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	30
1511_UN_A2_D	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	9
1511_UN_A2_M	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	6
1511_UN_A2_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_B1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_B2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_B3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_B4_C	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	40
1511_UN_B4_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_C1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_D1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_D3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_D4_C	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	30
1511_UN_D4_S	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1511_UN_R3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2015-11	< 3
1612_UN_A1_1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	40
1612_UN_A1_2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	30
1612_UN_A1_3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	20
1612_UN_A1_4	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	10
1612_UN_A1_5	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	< 3
1612_UN_B1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	< 3

Table 3-1. Continued

1612_UN_B7_1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	55
1612_UN_B7_2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	45
1612_UN_B7_3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	35
1612_UN_B7_4	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	25
1612_UN_B7_5	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	15
1612_UN_B9_1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	65
1612_UN_B9_2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	50
1612_UN_B9_3	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	35
1612_UN_B9_4	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	20
1612_UN_B9_5	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	< 3
1612_UN_E1	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	< 3
1612_UN_E2	Unagi-onsen (hot spring); Ibusuki, Kagoshima	31°13′ 41″ N, 130°36′ 47″ E	2016-12	< 3
1612_KR_A1	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_A2	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_A3	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_B1	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_B2	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_C1	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_C2	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_D1	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_D2	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1612_KR_E1	Yamanoshiro-onsen (hot spring); Kirishima, Kagoshima	31°54′ 51″ N, 130°49′ 46″ E	2016-12	< 3
1703_KM_1	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_2-1_03	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_2-1_07	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	7
1703_KM_2-2	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_3_05	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_3_20	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	20
1703_KM_3_40	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	40

Table 3-1. Continued

1703_KM_4	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_5	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_6	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_7	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_8_05	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
1703_KM_8_10	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	10
1703_KM_9	Komatsu-jigoku (hot spring); Kusugun Kokonoemachi, Oita	33°06′ 26″ N, 131°11′ 22″ E	2017-03	< 3
Izu peninsula, Japan (hot spring sediments)				
1501_IZ_ATGH01	hot spring; Kamogun Higashiizucho, Shizuoka	34°48′ 55″ N, 139°04′ 18″ E	2015-01	< 3
1501_IZ_ATGM02	drain; Kamogun Higashiizucho, Shizuoka	34°49′ 02″ N, 139°04′ 07″ E	2015-01	< 3
1501_IZ_BZ05	Benzainoyu (hot spring); Kamogun Higashiizucho, Shizuoka	34°48′ 59″ N, 139°04′ 13″ E	2015-01	< 3
1501_IZ_BZ06	Benzainoyu (hot spring); Kamogun Higashiizucho, Shizuoka	34°48′ 59″ N, 139°04′ 13″ E	2015-01	< 3
1501_IZ_BZ07	Benzainoyu (hot spring); Kamogun Higashiizucho, Shizuoka	34°48′ 59″ N, 139°04′ 13″ E	2015-01	
1501_IZ_JD	Jiunji (drain); Kamogun Minamiizucho, Shizuoka	34°38′ 54″ N, 138°52′ 00″ E	2015-01	< 3
1501_IZ_JU	Jiunji (drain); Kamogun Minamiizucho, Shizuoka	34°38′ 54″ N, 138°52′ 00″ E	2015-01	< 3
1501_IZ_SM05	hot spring; Kamogun Minamiizucho, Shizuoka	34°38′ 59″ N, 138°51′ 27″ E	2015-01	< 3
1501_IZ_SM07	hot spring; Kamogun Minamiizucho, Shizuoka	34°38′ 55″ N, 138°51′ 26″ E	2015-01	< 3
1501_IZ_KD01	Yakushinoyu (hot spring); Kamogun Kawazucho, Shizuoka	34°44′ 39″ N, 138°59′ 04″ E	2015-01	< 3
Kyushu Island, Japan (lake sediment)				
1612_UN_UL	Unagi-ike (lake); Ibusuki, Kagoshima	31°13′ 35″ N , 130°36′ 37″ E	2016-12	< 3

Table 3-2. Concentration of extracted DNA, sequencing, clustering statics, and alpha diversity indices

Sample ID	Temperature (°C)	pH	ORP (mV)	Salinity (%)	Concentration of DNA (ng/g sediment)	Original paired-end reads	Reads after QC	Prokaryotic reads	Prokaryotic OTUs	Chao1	Shannon	Simpson
Kyushu Island, Japan (hot spring sediments)												
1405_UN_A1_D	65.2	2.59	-130	n.m.	260	221219	192043	181734	301	223.0	2.331	0.778
1405_UN_A1_M	65.6	2.65	-127	n.m.	506	160545	140181	133054	266	187.8	2.329	0.815
1405_UN_A1_S	65.2	2.75	-102	n.m.	325	169789	150646	139173	371	255.5	2.803	0.875
1405_UN_A2_D	65.1	2.88	-23	n.m.	3083	52823	47952	44850	192	185.9	3.372	0.946
1405_UN_A2_M	52.3	2.91	-18	n.m.	1440	134228	121805	113709	251	181.6	3.339	0.946
1405_UN_A2_S	41.4	3.05	223	n.m.	3548	161792	146806	132987	439	368.9	3.848	0.931
1405_UN_B1	94	3.66	-140	n.m.	938	168663	149584	131999	173	163.6	2.296	0.820
1405_UN_B2	99	2.2	-88	n.m.	526	123304	102283	96729	107	65.2	1.912	0.798
1405_UN_B3	67	3.56	426	n.m.	495	81040	71050	67996	172	196.6	2.942	0.913
1405_UN_B4	63.7	2.94	-87	n.m.	1028	41266	37031	35239	230	232.6	2.714	0.840
1405_UN_B7	61.8	4.89	-218	n.m.	1515	113124	101466	82452	312	275.3	3.065	0.907
1405_UN_B8	97.1	3.65	-102	n.m.	227	218728	190809	180316	308	230.6	2.663	0.849
1405_UN_R2	46.1	3.47	-130	n.m.	2955	180988	167742	150457	373	263.2	2.823	0.876
1505_UN_A1_S	51.2	5.2	-54	n.m.	593	514741	373261	314614	386	278.6	2.827	0.900
1505_UN_B1	84.8	5.3	-148	n.m.	473	571792	308981	246657	631	504.1	2.720	0.789
1505_UN_B2	84.9	5.65	-133	n.m.	n.d.	62176	41263	34781	385	400.0	2.711	0.799
1505_UN_B3	46.4	4.81	-90	n.m.	431	647383	448848	387792	251	151.0	1.478	0.601
1505_UN_B4	66	4.91	-121	n.m.	332	505750	361180	309314	416	305.8	3.479	0.945
1505_UN_B6	39.4	5.54	-33	n.m.	624	634131	372764	316964	664	503.5	3.316	0.923
1505_UN_B7	48.4	4.85	206	n.m.	737	646641	461262	398919	416	243.6	2.956	0.904
1505_UN_B8	80.9	5.07	-123	n.m.	1080	317415	220321	183807	318	262.7	2.853	0.851
1505_UN_C1	80	4.85	140	n.m.	795	161418	112449	90397	536	469.6	3.387	0.912
1505_UN_C2	93.1	4.63	-118	n.m.	n.d.	83276	54762	45264	172	149.1	2.159	0.718
1505_UN_C3	58.7	4.57	-125	n.m.	167	236695	163085	132740	253	200.6	2.877	0.873

Table 3-2. Continued

1505_UN_C4	95.8	4.37	-174	n.m.	n.d.	15801	9268	7743	60			
1505_UN_C5	54.6	4.54	-62	n.m.	n.d.	193328	133094	111801	415	325.0	2.675	0.809
1505_UN_R1	33.8	5.83	277	n.m.	2633	698236	426056	221063	764	606.1	3.619	0.889
1505_UN_R2	48.4	5.83	108	n.m.	855	605043	425210	337733	730	446.1	3.323	0.930
1511_UN_A1_C	83.9	5.08	-130	n.m.	n.d.	7855	4402	3422	120			
1511_UN_A1_D	63.5	5.44	18	n.m.	179	93774	64053	48893	324	317.8	3.379	0.929
1511_UN_A1_M	53	5.53	55	n.m.	484	112126	76938	60517	317	272.6	3.120	0.896
1511_UN_A1_S	48	5.56	145	n.m.	173	125770	85979	65641	384	338.3	3.151	0.899
1511_UN_A2_C	89.1	3.59	-89	n.m.	n.d.	6442	2968	2371	54			
1511_UN_A2_D	70.9	4.68	-68	n.m.	n.d.	5117	1710	1393	42			
1511_UN_A2_M	60.6	4.77	-41	n.m.	23	92623	58731	43710	134	116.5	1.934	0.720
1511_UN_A2_S	52.9	5.39	10	n.m.	495	90223	60140	44173	320	327.4	3.426	0.931
1511_UN_B1	81.3	4.7	135	n.m.	319	149571	90284	68544	312	264.3	3.357	0.933
1511_UN_B2	96.2	4.33	449	n.m.	248	370115	253605	206331	304	210.3	2.896	0.898
1511_UN_B3	75.9	4.24	309	n.m.	10	229180	143543	118214	113	75.0	1.083	0.399
1511_UN_B4_C	94.9	3.65	-50	n.m.	15	232777	118963	87190	159	141.9	2.115	0.688
1511_UN_B4_S	60.6	4.14	151	n.m.	424	34512	22592	18125	208	225.0	2.762	0.844
1511_UN_C1	91.3	3.8	-108	n.m.	81	155755	103895	81930	399	321.2	3.053	0.833
1511_UN_D1	41.2	3.59	413	n.m.	435	140249	79786	58827	338	288.6	2.742	0.837
1511_UN_D3	66.5	4.64	-121	n.m.	336	184298	125610	102651	194	129.8	1.897	0.730
1511_UN_D4_C	94.3	3.31	-96	n.m.	45	140444	89892	71530	177	158.6	2.248	0.829
1511_UN_D4_S	67.6	4.78	276	n.m.	198	95200	59087	46282	154	163.0	2.228	0.758
1511_UN_R3	46.2	4.03	-113	n.m.	1144	341559	224784	183392	403	289.3	3.125	0.913
1612_UN_A1_1	95.6	5.32	-234	0	n.d.	320	188	153	38			
1612_UN_A1_2	96.9	5.23	-201	0	n.d.	3626	2669	2462	110			
1612_UN_A1_3	96.1	5.05	-185	0	n.d.	4806	3892	3690	58			
1612_UN_A1_4	92.8	5.09	-150	0	n.d.	1269	725	656	26			
1612_UN_A1_5	88.2	4.04	-239	0	104	163225	130251	103809	390	348.3	3.937	0.966
1612_UN_B1	92.6	2.64	164	0	420	200744	174422	164923	476	354.3	3.012	0.912

Table 3-2. Continued

1612_UN_B7_1	94	5.5	-160	0	n.d.	6647	5401	5114	71			
1612_UN_B7_2	93.1	5.9	-227	0	n.d.	501	340	256	45			
1612_UN_B7_3	90.3	5.14	-262	0	78	85410	74702	70244	367	318.6	1.454	0.537
1612_UN_B7_4	82.7	4.74	-92	0	n.d.	1953	1299	993	45			
1612_UN_B7_5	76.1	3.73	-60	0	22	64247	55549	51332	69	60.2	1.551	0.658
1612_UN_B9_1	89.2	3.67	-100	0	n.d.	3133	2172	1903	87			
1612_UN_B9_2	70.1	3.65	-51	0	n.d.	2250	1462	1353	58			
1612_UN_B9_3	60.1	3.61	-40	0	n.d.	5834	5049	4745	75			
1612_UN_B9_4	48.6	4.27	-17	0	32	113501	98876	94821	151	122.7	1.194	0.419
1612_UN_B9_5	35.5	4.84	155	0	1429	137544	115062	75817	438	475.0	3.974	0.948
1612_UN_E1	84.6	4	-157	0	271	145063	124502	119082	120	79.1	1.128	0.524
1612_UN_E2	68.1	3.61	151	0	413	165409	142986	112993	569	442.3	4.003	0.952
1612_KR_A1	69.5	2.62	310	0	506	111982	97764	88943	132	153.7	1.624	0.675
1612_KR_A2	75.2	2.62	-23	0	270	117971	102995	93963	205	198.7	2.827	0.906
1612_KR_A3	74	2.53	-13	0	147	106489	93170	88561	100	94.0	1.149	0.505
1612_KR_B1	82.8	3.12	-41	0	405	135679	117074	107985	216	150.2	2.218	0.759
1612_KR_B2	72.5	2.67	262	0	244	114607	100028	92287	122	130.0	1.263	0.522
1612_KR_C1	87.7	2.41	0	0	79	180943	160748	151386	86	67.0	1.432	0.692
1612_KR_C2	88.7	2.78	-45	0	119	184953	163045	156340	147	81.3	0.834	0.350
1612_KR_D1	88.6	3.5	-179	0	101	79737	69519	64539	119	90.9	1.774	0.744
1612_KR_D2	63.4	3.19	-80	0	143	196456	171183	150908	311	247.0	3.313	0.921
1612_KR_E1	64.4	4.08	-134	0	1279	179637	152804	132974	169	149.1	2.075	0.758
1703_KM_1	61.1	5.41	-34	n.m.	2888	14525	9137	7229	243			
1703_KM_2-1_03	n.m.	n.m.	n.m.	n.m.	1631	84879	53525	41340	542	531.2	4.651	0.980
1703_KM_2-1_07	n.m.	n.m.	n.m.	n.m.	881	83277	52460	38434	465	459.2	4.119	0.964
1703_KM_2-2	61.6	2.54	130	n.m.	82	94557	63392	54013	202	159.5	2.339	0.834
1703_KM_3_05	n.m.	n.m.	n.m.	n.m.	73	21349	11178	9065	74			
1703_KM_3_20	n.m.	n.m.	n.m.	n.m.	n.d.	12517	603	160	23			
1703_KM_3_40	n.m.	n.m.	n.m.	n.m.	n.d.	2307	192	107	24			

Table 3-2. Continued

1703_KM_4	65.8	2.24	157	n.m.	398	36162	19005	16359	146	162.3	2.989	0.904
1703_KM_5	71.6	2.77	286	n.m.	140	45141	31244	26176	165	180.3	2.960	0.892
1703_KM_6	71	5.6	-179	n.m.	1710	36419	26812	19036	108	104.2	2.567	0.855
1703_KM_7	80.9	2.34	40	n.m.	146	18611	9658	7793	70			
1703_KM_8_05	n.m.	n.m.	n.m.	n.m.	303	27478	19263	14242	138	160.9	2.446	0.833
1703_KM_8_10	n.m.	n.m.	n.m.	n.m.	72	15637	8876	5826	86			
1703_KM_9	75.1	2.6	54	n.m.	68	22796	11903	9810	54			
Izu peninsula, Japan (hot spring sediments)												
1501_IZ_ATGH01	73.5	8.5	121	0	n.d.	49115	45323	28823	75	86.4	1.366	0.582
1501_IZ_ATGM02	68.2	8.4	189	0.1	235	20001	18248	14146	125	162.0	2.503	0.832
1501_IZ_BZ05	80.1	8.5	-7	0.1	n.d.	67427	59075	16992	92	94.5	1.787	0.707
1501_IZ_BZ06	78.5	8.5	-22	0.2	n.d.	7961	6663	1775	41			
1501_IZ_BZ07	78.3	8.5	63	0.2	n.d.	3115	2349	762	47			
1501_IZ_JD	60.1	7.7	259	0.9	533	183670	165434	98280	1171	1072.6	4.562	0.968
1501_IZ_JU	78.5	7.8	156	1	663	74488	66002	39452	1069	1115.4	3.857	0.857
1501_IZ_SM05	63.2	7.9	81	2.4	1223	77505	67130	26283	518	546.5	4.008	0.959
1501_IZ_SM07	61.4	8.3	94	2	1545	188828	167270	67058	378	370.5	3.397	0.933
1501_IZ_KD01	70.4	8.9	-30	0	825	56458	48662	30657	115	124.6	3.186	0.933
Kyushu Island, Japan (lake sediment)												
1612_UN_UL	17.5	7.37	75	n.m.	1868	122928	101844	71662	4737	4323.5	7.068	0.996

n.m., not measured; n.d., not detected.

16S rRNA gene amplification and sequencing

The V3/V4 region of bacterial and archaeal 16S rRNA genes was amplified with the following prokaryotic universal primer sets (Takahashi *et al.*, 2014): forward (5′-CCTACGGGNBGCASCAG-3′) and reverse (5′-GACTACNVGGGTATCTAATCC-3′) with added overhanging adapter sequences at each 5′-end according to the 16S metagenomic sample preparation guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). Each sample was amplified with KAPA HiFi HotStart ReadyMix (2X) (KAPA Biosystems, Cape Town, South Africa) according to the manufacturer's instructions. Paired-end (PE, 2×300 nucleotides) sequencing was performed with an Illumina MiSeq (MiSeq Reagent kit v3) and followed the manufacturer's run protocols (Illumina, Inc., San Diego, CA, USA).

16S rRNA gene sequence processing and statistical analyses

Primer-binding regions were removed by trimming 17 and 21 nt sequences from the 5′ ends of the forward and reverse reads without adapter regions, respectively, with VSEARCH version 2.6.0 (Rognes *et al.*, 2016). The reads were further processed by trimming low-quality regions from the sequences with Trimmomatic version 0.36 (SLIDINGWINDOW: 50:20) (Bolger *et al.*, 2014). Using VSEARCH, the paired-end reads were joined and de-multiplexed, and a further round of quality control was conducted to remove sequences shorter than 200 nt as well as those containing ambiguous bases (N) or bases with a quality score below 20. Chimeric 16S rDNA sequences were detected using the UCHIME algorithm in the USEARCH package implemented within

VSEARCH. The SILVA 132 SSU Ref Nr99 (Quast *et al.*, 2013), a comprehensive, quality checked data sets of small subunit rRNA sequences, was used as a reference for chimera detection. Operational taxonomic units (OTUs) were defined as clusters of sequences that were not singletons (unique sequences that are present exactly once in each sample) with 98.7% similarity using VSEARCH. Then, taxonomic classification of individual OTU was performed with the stand-alone SINA version 1.2.11 aligner (Pruesse *et al.*, 2012) using the SILVA 132 SSU Ref Nr99 database as a reference. The non-prokaryotic OTUs (i.e., eukaryote and unclassified domain) were then removed. OTU abundance was estimated by adding prokaryotic singleton reads using the global alignment search option of VSEARCH (--usearch_global --id 0.987), to increase sensitivity. Prior to community analysis, samples with less than 10,000 sequences were omitted (leaving 77 samples) in the beta-diversity patterns. The resulting OTU abundance tables were rarefied to an even number of sequences per sample to ensure equal sampling depth (14,146 sequences per sample) using the vegan package (Oksanen *et al.*, 2017) of the R software (R Core Team, 2016). Alpha and beta diversity analyses were then performed with the phyloseq (McMurdie & Holmes, 2013) and vegan packages of the R software.

Database search for CODH–ECH gene clusters

The amino acid sequences corresponding to CODHs were obtained from the Reference Sequence Database (RefSeq) in National Center for Biotechnology Information (NCBI) (December 2018) (Sayers *et al.*, 2019) through a BLASTp search using *C. hydrogenoformans* CooSI (ABB14432.1) subunit as a query. Low-scoring and short-length hits (bit score < 200, amino acid length < 550) including hybrid-cluster proteins and partial fragments were excluded from the data set. Then, coding sequences (CDSs)

within 20 CDSs upstream and downstream of the CODH gene locus were annotated by clusters of orthologous groups (COGs) (Tatusov, 2001) through RPS-BLAST search (E value < 10^{-6}) using NCBI Conserved Domain Database (Marchler-Bauer *et al.*, 2002). Of these, I identified CODH genes with ECH small and large subunits (COG3260 and COG3261, respectively) as CODH–ECH gene clusters.

Phylogenetic analyses

I retrieved the reference 16S rRNA gene sequences that were equal or longer than 1,000 nt and did not include N from the genomes of prokaryotes possessing CODH–ECH gene clusters and those that were classified into the same genera as them via the RefSeq genome database. To obtain a non-redundant data set for phylogenetic analysis, retrieved sequences were trimmed into V3/V4 region identical to the amplicons and clustered with 100% similarity using VSEARCH. The sequences were aligned using MAFFT version 7.402 (Katoh & Standley, 2013). Maximum-likelihood phylogenetic trees were calculated using FastTree version 2.1.9 (Price *et al.*, 2010) with an approximate-maximum-likelihood method using the GTR + GAMMA model. Robustness of the topology of the phylogenetic trees was evaluated by local bootstrap values based on 1000 re-samples. The tree was imported into the iTOL online tool (Letunic & Bork, 2016) for visualization.

Exploring the co-occurrence of thermophilic hydrogenogenic carboxydotrophs and other microorganisms

Based on the OTU read numbers, a network of phylotype co-occurrence was produced with a minimum Spearman correlation coefficient of 0.8 using R. I retrieved and have presented the smaller networks, including phylotypes, related to the thermophilic

hydrogenogenic carboxydotrophs identified in my phylogenetic analysis.

Results and Discussion

Sample profiles and overview of 16S rRNA gene amplicon sequencing

I collected 100 sediment samples from geographically distant areas in Japan, including Kyushu Island and the Izu Peninsula (Table 3-3; additional data are provided in Tables 3-1, 3-2). Except for a single sample from Unagi-ike lake, which has a moderate environment (17.5 °C; pH 7.37; ORP, + 75 mV), all the samples were collected from geothermally heated hydrothermal environments (33.8–99.0°C). Although the in situ environmental conditions of the sampling sites were variable, the hot springs on Kyushu Island had an acidic pH (average pH 4.1 ± 1.1 [sd]; measurable sites, n = 82), whereas those on the Izu Peninsula were neutral or weakly alkaline (pH 8.3 ± 0.4; n = 10).

16S rRNA gene amplicon sequencing analysis generated 8,531,132 bacterial and archaeal quality-controlled sequences from the 100 samples, with a range of 107–398,919 sequences (average, 85,311 sequences) per sample (Table 3-2). A total of 9,394 prokaryotic OTUs were defined at the 98.7% similarity level, and 23–4,737 OTUs (average, 299 OTUs) were observed in each sample (Table 3-2). Diversity analysis using rarefied 77 samples with equal or greater than 10,000 sequences revealed that microbial communities in the sampled hot springs showed much lower alpha diversity than those in the moderate environment (Unagi-ike lake; Fig. 3-1), indicating that high temperature imposed constraints on community properties as observed in other studies (Sharp *et al.*, 2014).

Furthermore, beta diversity analysis revealed apparent differences between the

acidic hot springs on Kyushu Island and the neutral or weak alkaline environments on Izu Peninsula and Unagi-ike lake (Fig. 3-2). At the domain level, microbial communities in the acidic hot springs were dominated by archaea, whereas those in the neutral or weak alkaline environments were dominated by bacteria (Fig. 3-3). The phylotypes that shared 100% identity with *Vulcanisaeta souniana* (phylum Crenarchaeota; OTU_1) and *Thermus thermophilus* (phylum Deinococcus-Thermus; OTU_20) were notably prominent in the acidic hot springs and neutral or weak alkaline environments, respectively. *V. souniana* is a heterotrophic anaerobic hyperthermophilic crenarchaeon found in hot springs that grows optimally at 85–90 °C and pH 4.0–4.5 (Itoh *et al.*, 2002). In contrast, *T. thermophilus* is an extremely thermophilic bacterium also found in hot springs, but its optimal growth occurs at 65–72 °C and pH 7.5 (Oshima & Imahori, 1974). Although the major phylotypes were the same in each acidic and neutral or weak alkaline environments, my non-metric multidimensional scaling analysis using rarefied 77 samples with greater equal than 10,000 sequences shows that microbial community compositions vary across each sampling sites (Fig. 3-2).

Table 3-3. Summary of samples

Sampling area and time point	Sampling date	Numbers of samples	Temperature (°C)	pH	ORP (mV)	Salinity (%)
1405_Unagi	May 2014	13	41.4 ~ 99.0	2.2 ~ 4.9	−218 ~ +426	n.m
1505_Unagi	May 2015	15	33.8 ~ 95.8	4.4 ~ 5.8	−174 ~ +277	n.m
1511_Unagi	November 2015	19	41.2 ~ 96.2	3.3 ~ 5.6	−130 ~ +449	n.m
1612_Unagi	December 2016	18	35.5 ~ 96.9	2.6 ~ 5.9	−262 ~ +164	n.m
1612_Kirishima	December 2016	10	63.4 ~ 88.7	2.4 ~ 4.1	−179 ~ +310	0
1703_Komatsu	March 2017	14	61.1 ~ 80.9	2.2 ~ 5.6	−179 ~ +286	n.m
1501_Eastern_Izu	January 2015	5	68.2 ~ 80.1	8.4 ~ 8.5	−22 ~ +189	0.0 ~ 0.2
1501_Southern_Izu	January 2015	5	60.1 ~ 78.5	7.7 ~ 8.9	−30 ~ +259	0.0 ~ 2.4
1612_Unagi-ike_lake	December 2016	1	17.5	7.37	75	n.m

n.m. not measured

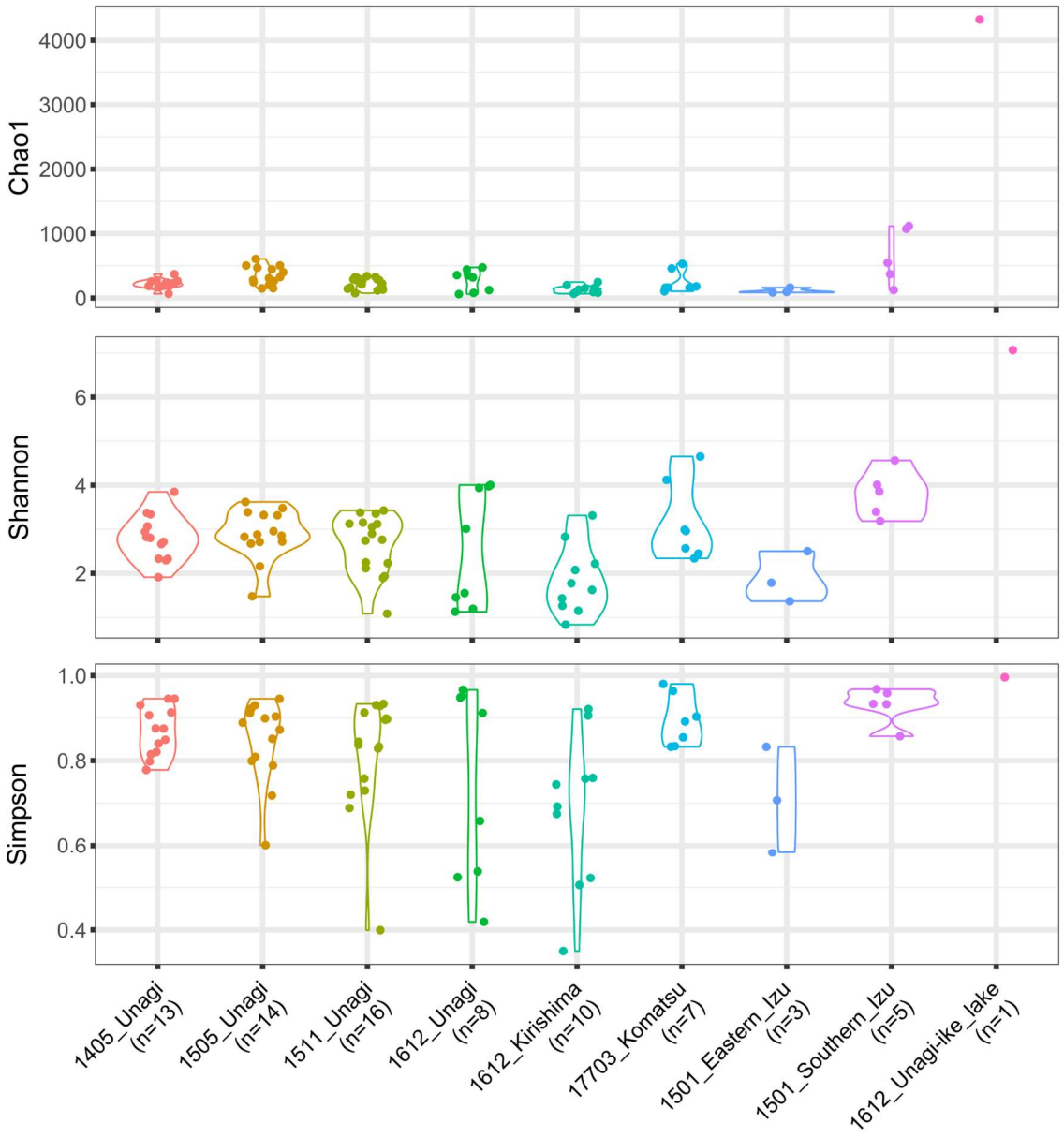


Fig. 3-1. Alpha diversity using rarefied samples. The upper, middle, and bottom panels display the Chao1, Shannon, and Simpson indices, respectively. Samples from different areas and periods were plotted separately.

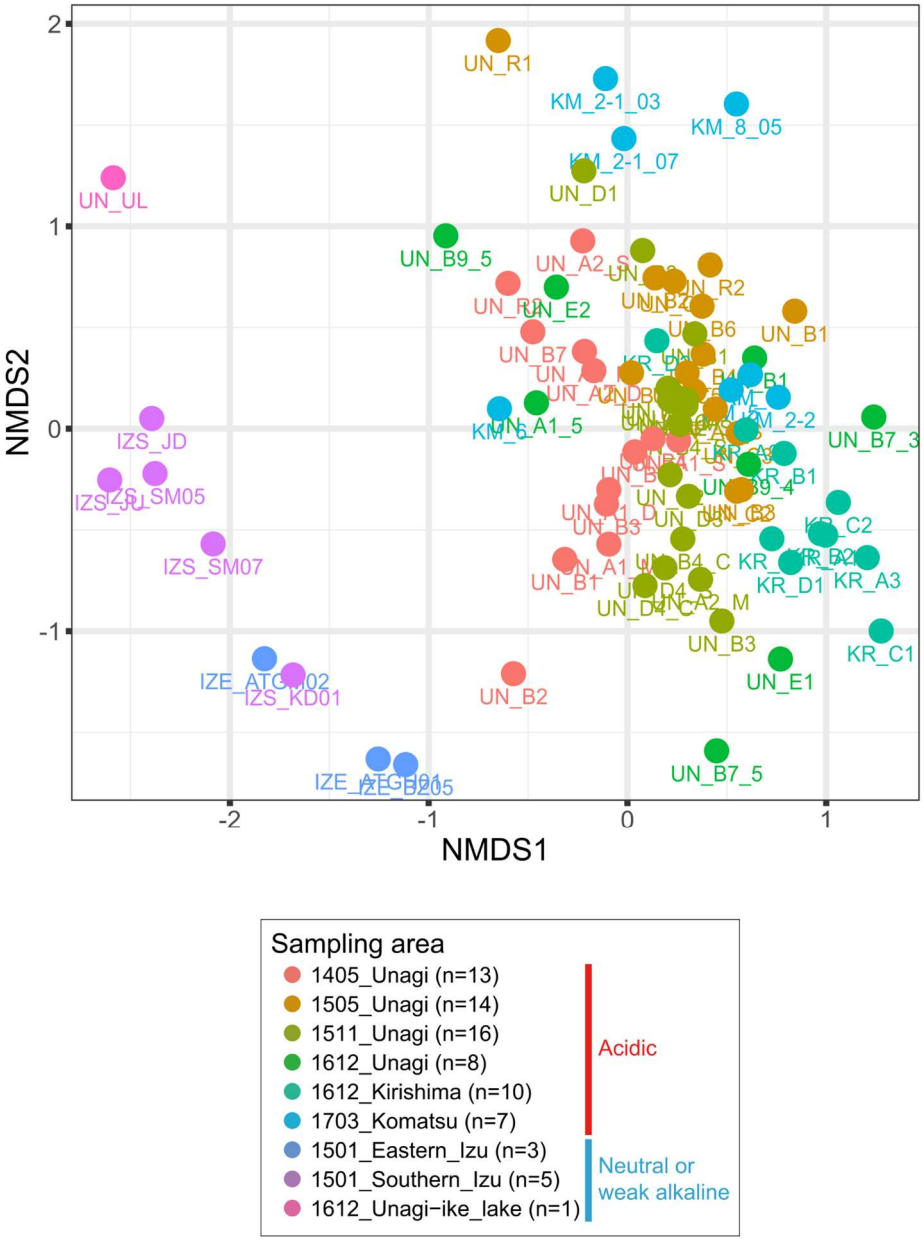


Fig. 3-2. Non-metric multidimensional scaling analysis based on Bray–Curtis dissimilarity. Plot colors illustrate sampling area and period

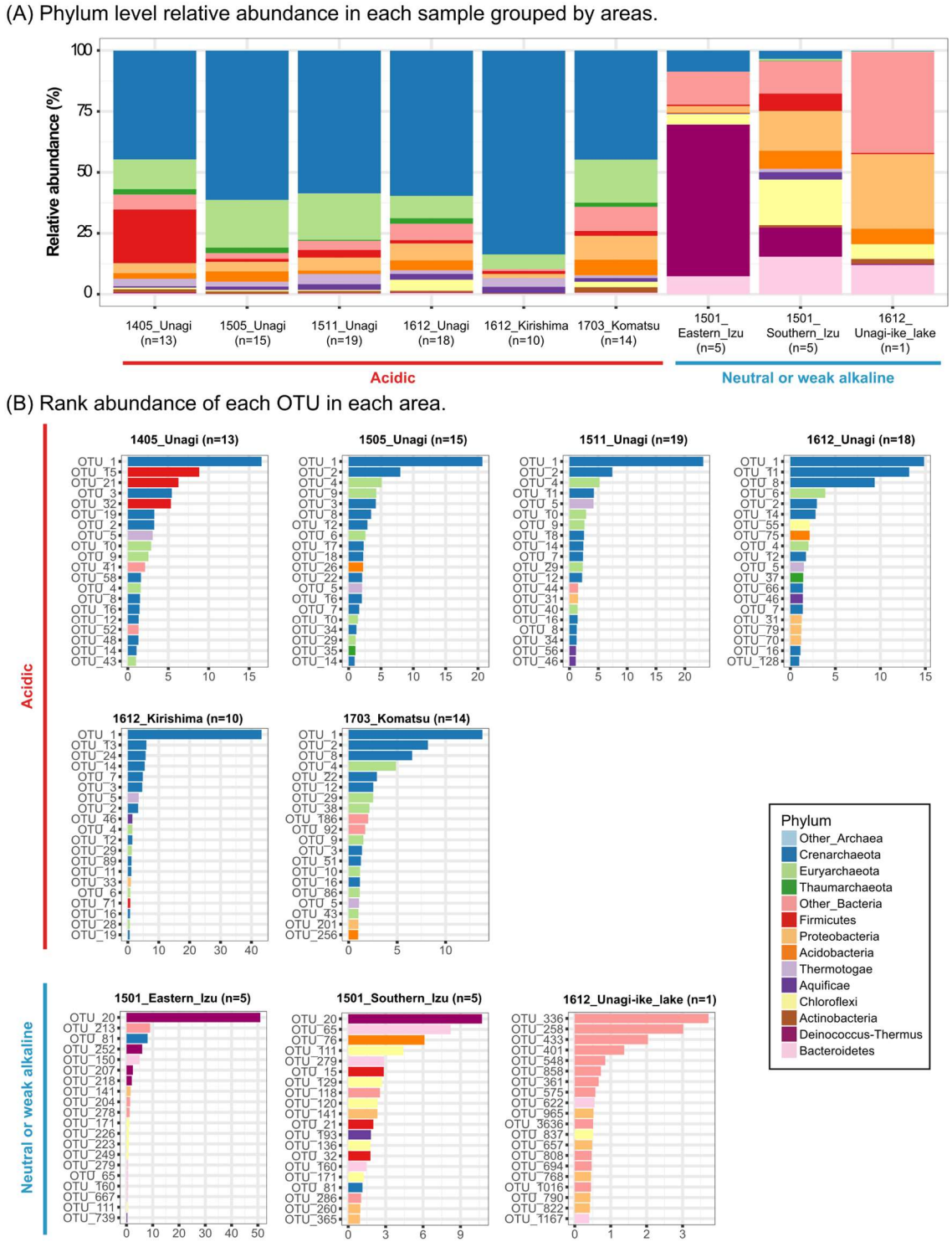


Fig. 3-3. Microbial taxonomic compositions in each area and period. (A) Phylum level community composition of each area. (B) The proportion of the 20 most predominant OTUs in each community. Bar colours indicate the phyla to which individual groups were assigned.

CODH–ECH gene clusters found in prokaryotic genomes

The previous study examined CODHs and their genomic context in 2,887 microbial genomes and revealed 185 genomes that encoded at least one CODH gene (Techtmann *et al.*, 2012). Of these, 12 genomes possessed CODH–ECH gene clusters. However, by December 2018, the number of sequenced microbial genome entries in the RefSeq genome database had reached 142,909, and novel thermophilic hydrogenogenic carboxydotrophs had been reported. Therefore, I searched CODH–ECH gene clusters in the current RefSeq database and examined their taxonomic information. I identified 71 genomes encoding CODH–ECH gene clusters, which include 40 thermophile genomes (14 genera), 25 mesophile genomes (12 genera), and six unclassified microbial genomes (Table 3-4). All mesophilic members were classified into the phylum Proteobacteria, which included phototrophic bacteria or sulfate-reducing bacteria, whereas thermophilic members were phylogenetically diverse and classified into the phylum Crenarchaeota, Euryarchaeota, and Firmicutes. Of these 71 genomes, 46 have never been reported as hydrogenogenic carboxydotrophs (Table 3-4), and the presence of CODH–ECH gene clusters in 22 genomes was reported for the first time in this study (Table 3-4).

Conservation patterns of CODH–ECH gene clusters were different in each genus (Fig. 3-4). I classified these genera into three groups: (1) the CODH–ECH gene clusters and the hydrogenogenic carboxydotrophy ability were well conserved; (2) a portion of members conserved the CODH–ECH gene clusters; and (3) genera that we could not classify into (1) nor (2) because of inadequate availability of genomic information. *Thermincola*, *Carboxydocella*, *Carboxydothermus*, and *Caldanaerobacter* were classified into the group (1). In most cases, the phylogeny of CODH genes was corresponding to their taxonomic phylogeny in this group (Adam *et al.*, 2018; Fukuyama

et al., 2018; Toshchakov *et al.*, 2018), suggesting that the CODH–ECH gene clusters descended from the common ancestors of each genus. The genus *Carboxydothermus* has been one of the most studied models of thermophilic carboxydotrophy, and the members of this genus possess four or five CODH genes (Fukuyama *et al.*, 2018). The comparative genomic analysis in *Carboxydothermus* revealed that the CODH–ECH gene clusters were conserved in the members except for *C. pertinax*, which lacked only the CODH (CODH-I) unit of CODH–ECH gene cluster and *Carboxydothermus ferrireducens*, which lacked the whole CODH–ECH gene cluster (Fukuyama *et al.*, 2018). *C. ferrireducens* can grow carboxydotrophically, but is not hydrogenogenic (Slobodkin *et al.*, 2006). On the other hand, *C.pertinax* can grow by hydrogenogenic carboxydotrophy (Yoneda *et al.*, 2012), and it is suggested that *C. pertinax* could couple alternative CODH (CODH-II) to the distal ECH (Fukuyama *et al.*, 2018). *C. pertinax* was the only isolate that could grow by hydrogenogenic carboxydotrophy without the CODH–ECH gene cluster. *C. subterraneus* subspecies can oxidize CO and possess CODH–ECH gene clusters, whose structures are very similar (Sant'Anna *et al.*, 2015). However, phylogenetic reconstruction of CODH genes revealed that CODH genes from *C. subterraneus* have distinct evolutionary histories. It is suggested that replacement of CODH gene occurred by a horizontal gene transfer event in *C. subterraneus* subsp. *tengcongensis* and *C. subterraneus* subsp. *yonseiensis* (Sant'Anna *et al.*, 2015). *Thermococcus*, *Thermofilum*, *Thermoanaerobacter*, *Moorella*, *Desulfotomaculum*, *Desulfosporosinus*, *Parageobacillus*, and members of the phylum Proteobacteria were classified into group (2). Because most species of *Thermococcus*, *Thermofilum*, *Thermoanaerobacter*, *Desulfotomaculum*, and *Desulfosporosinus* did not possess the CODH–ECH gene clusters, it was suggested that CODH–ECH gene clusters might have been obtained by a portion of the members in a

horizontal gene transfer event. In fact, this cluster is believed to be horizontally transferred between the representatives of separate taxa (Techtmann *et al.*, 2012). In the genus *Moorella*, *Moorella stamsii* and *Moorella glycerini* possessed identical CODHs that were flanked by ECH gene clusters. *Moorella* sp. Hama-1 and *Moorella thermoacetica* DSM 21394, which formed a different subclade from *M. stamsii* and *M. glycerini*, also possessed a similar CODH–ECH gene cluster. However, it was revealed that the other 11 *M. thermoacetica* strains did not possess the CODH–ECH gene cluster (Table 3-4). *M. thermoacetica* might be an acetogenic carboxydotroph rather than being hydrogenogenic, as reported previously (Pierce *et al.*, 2008; Schuchmann & Müller, 2014), and only strain DSM 21394 might be hydrogenogenic. *Parageobacillus thermoglucosidasius* is the only facultative anaerobic bacillus among the thermophilic hydrogenogenic carboxydotrophic species (Mohr *et al.*, 2018). Although other *Parageobacillus* species did not possess the CODH–ECH gene cluster, *P. thermoglucosidasius* possesses a CODH–ECH gene cluster that is phylogenetically related to those of *Moorella* and *Caldanaerobacter* (Mohr *et al.*, 2018). Unlike *M. thermoacetica*, all 10 genomes of *P. thermoglucosidasius* have conserved the CODH–ECH gene clusters (Table 3-4), and hydrogenogenic carboxydotrophy might be an important trait for this species. The other species, *Thermanaeromonas toyohensis*, *Thermosinus carboxydivorans*, *Calderihabitans maritimus*, and uncultured archaea and bacteria (Candidatus Korarchaeota archaeon MDKW, Clostridium bacterium DRI-13, and Rhizobiales bacterium) were classified into the group (3).

Table 3-4. Prokaryotes possessing CODH–ECH gene clusters

Organism	Hydrogenogenic carboxydotrophic growth	Isolation source
Crenarchaeota (thermophilic)		
Thermofilum carboxyditrophus 1505	Yes (Sokolova *et al*., 2009)	Water and mud (Sokolova *et al*., 2009)
Euryarchaeota (thermophilic)		
Thermococcus barophilus CH5	Yes (Kozhevnikova *et al*., 2016)	Deep-sea hydrothermal fields (Kozhevnikova *et al*., 2016)
Thermococcus barophilus MP	Yes (Kozhevnikova *et al*., 2016)	Deep-sea hydrothermal fields (Kozhevnikova *et al*., 2016)
Thermococcus guaymasensis DSM 11113	n.r	Hydrothermal vent sediment (Canganella *et al*., 1998)
Thermococcus onnurineus NA1	Yes (Bae *et al.*, 2006)	Deep-sea hydrothermal fields (Bae *et al.*, 2006)
Thermococcus paralvinellae ES1	n.r	Active hydrothermal vent chimneys (Hensley *et al*., 2014)
Thermococcus sp. AM4	Yes (Sokolova *et al*., 2004b)	Active chimney (Sokolova *et al*., 2004b)
Firmicutes (thermophilic)		
Parageobacillus thermoglucosidasius B4168	n.r	Food (Berendsen *et al*., 2016)
Parageobacillus thermoglucosidasius C56-YS93	n.r	Hot spring (Brumm *et al*., 2015)
Parageobacillus thermoglucosidasius DSM 2542[a]	Yes (Mohr *et al*., 2018)	Soil (Suzuki *et al*., 1983)
Parageobacillus thermoglucosidasius GT23	n.r	Casein pipeline (SAMN04532072[b])
Parageobacillus thermoglucosidasius NBRC 107763	n.r	n.r
Parageobacillus thermoglucosidasius NCIMB 11955	n.r	n.r
Parageobacillus thermoglucosidasius TG4	Yes (Inoue *et al*., 2019a)	Marine sediment (Inoue *et al*., 2019b)
Parageobacillus thermoglucosidasius TM242	n.r	n.r
Parageobacillus thermoglucosidasius TNO-09.020	n.r	Dairy factory (Zhao *et al*., 2012)
Parageobacillus thermoglucosidasius Y4.1MC1	n.r	Hot spring (Brumm *et al*., 2015)
Carboxydocella sp. JDF658	Yes (Fukuyama *et al*., 2017)	Open-air stream from a hot spring well (Fukuyama *et al*., 2017)
Carboxydocella sp. ULO1	Yes (Fukuyama *et al*., 2017)	Sediment of a maar lake (Fukuyama *et al*., 2017)
Carboxydocella sporoproducens DSM 16521	Yes (Slepova *et al.*, 2006)	Hot spring (Slepova et al., 2006)
Carboxydocella thermautotrophica 019	Yes (Toshchakov *et al*., 2018)	Thermal field (Toshchakov *et al*., 2018)
Carboxydocella thermautotrophica 041	Yes (Sokolova *et al*., 2002)	Terrestrial hot vent (Sokolova *et al*., 2002)
Desulfosporosinus sp. OL	n.r	n.r
Desulfotomaculum nigrificans CO-1-SRB	Yes (Parshina *et al*., 2005b)	Anaerobic bioreactor sludge (Sokolova *et al*., 2009)
Thermincola ferriacetica Z-0001	Yes (Zavarzina *et al*., 2007)	Ferric deposits of a terrestrial hydrothermal spring (Zavarzina *et al*., 2007)
Thermincola potens JR	Yes (Wrighton *et al*., 2008; Byrne-Bailey *et al*., 2010)	Thermophilic microbial fuel cell (Wrighton *et al*., 2008)

Table 3-4. Continued.

Caldanaerobacter subterraneus subsp. *pacificus* DSM 12653	Yes (Sokolova *et al*., 2001; Fardeau *et al*., 2004)	Oilfields (Fardeau *et al*., 2004)
Caldanaerobacter subterraneus subsp. *tengcongensis* MB4	n.r	Oilfields (Fardeau *et al*., 2004)
Caldanaerobacter subterraneus subsp. *yonseiensis* KB-1	n.r	Oilfields (Fardeau *et al*., 2004)
Calderihabitans maritimus KKC1	Yes (Yoneda *et al*., 2013a)	Submerged marine caldera (Yoneda *et al*., 2013a)
Carboxydothermus hydrogenoformans Z-2901	Yes (Svetlichny *et al*., 1991)	Hot swamp (Svetlichny *et al*., 1991)
Carboxydothermus islandicus SET	Yes (Novikov *et al*., 2011)	Hot spring (Novikov *et al*., 2011)
Moorella glycerini NMP	n.r	Underground gas storage (Slobodkin *et al*., 1997)
Moorella sp. Hama-1	n.r	Thermophilic anaerobic digestion reactor (Harada *et al*., 2018)
Moorella stamsii DSM 26271	Yes (Alves *et al*., 2013)	Anaerobic sludge (Alves *et al*., 2013)
Moorella thermoacetica DSM 21394	Yes (Jiang *et al*., 2009)	Anaerobic bioreactors (Jiang *et al*., 2009)
Thermanaeromonas toyohensis ToBE	n.r	Geothermal aquifer in mine (Mori *et al*., 2002)
Thermoanaerobacter sp. YS13	n.r	Geothermal hot spring (Peng *et al*., 2016)
Thermosinus carboxydivorans Nor1	Yes (Sokolova *et al*., 2004a)	Hot spring (Sokolova *et al*., 2004a)
Proteobacteria (mesophilic)		
Rhodopseudomonas palustris BisB18	n.r	River sediment (Oda *et al*., 2008)
Pleomorphomonas carboxyditropha SVCO-16	n.r	Anaerobic sludge (Esquivel-Elizondo *et al*., 2018)
Pseudovibrio sp. POLY-S9	n.r	Intertidal marine sponge (Alex & Antunes, 2015)
Pseudovibrio sp. Tun.PSC04-5.I4	n.r	Tunicate symbiont (marine) (SAMN04515695[b])
Rhodospirillum rubrum ATCC 11170	Yes (Kerby *et al*., 1992)	Fresh water (Munk *et al*., 2011)
Rhodospirillum rubrum F11	Yes (Singer *et al*., 2006)	n.r
Desulfovibrio bizertensis DSM 18034	n.r	Marine sediment (Haouari *et al*., 2006)
Pseudodesulfovibrio piezophilus C1TLV30	n.r	Wood falls at deep sea (Khelaifia *et al*., 2011)
Geobacter bemidjiensis Bem	n.r	Subsurface sediments (Nevin *et al*., 2005)
Geobacter pickeringii G13	n.r	Kaolin clays (Shelobolina *et al*., 2007)
Ferrimonas futtsuensis DSM 18154	n.r	Sediment (Nakagawa *et al*., 2006)
Ferrimonas kyonanensis DSM 18153	n.r	Alimentary tract of littleneck clams (Nakagawa *et al*., 2006)
Ferrimonas sediminum DSM 23317	n.r	Coastal sediment (Ji *et al*., 2013)
Shewanella sp. M2	n.r	Antarctic deep-sea sediments (SAMN10397594[b])
Shewanella sp. R106	n.r	Antarctic deep-sea sediments (SAMN10397511[b])
Citrobacter amalonaticus Y19	Yes (Oh *et al*., 2008)	Anaerobic wastewater sludge digester (Jung *et al*., 1999)
Salmonella enterica subsp. *enterica* serovar Montevideo 50262	n.r	n.r

Table 3-4. Continued.

Salmonella enterica subsp. *enterica* serovar Montevideo 50270	n.r	n.r
Salmonella enterica subsp. *enterica* serovar Senftenberg 50263	n.r	n.r
Salmonella enterica subsp. *enterica* serovar Senftenberg 50264	n.r	n.r
Salmonella enterica subsp. *enterica* serovar Senftenberg 50265	n.r	n.r
Salmonella enterica subsp. *enterica* serovar Senftenberg 50271	n.r	n.r
Salmonella enterica subsp. *enterica* serovar Senftenberg 50272	n.r	n.r
Salmonella enterica subsp. *enterica* serovar Senftenberg SS209	n.r	Hatchery of broiler chickens (Grépinet *et al*., 2012)
Photobacterium marinum AK15	n.r	Sediment (Srinivas *et al*., 2013)
Uncultured		
Candidatus Korarchaeota archaeon MDKW	n.r	Hot springs metagenomes (SAMN10094317[b])
Clostridiales bacterium DRI-13	n.r	Subglacial ecosystem (SAMN02745515[b])
Rhizobiales bacterium AFS016371	n.r	n.r
Rhizobiales bacterium AFS041951	n.r	n.r
Rhizobiales bacterium AFS049984	n.r	n.r
Rhizobiales bacterium AFS089140	n.r	n.r

n.r. not reported

[a] Two genomes are available for this strain in the database.

[b] When there is no appropriate reference work, NCBI BioSample accessions are shown if available.

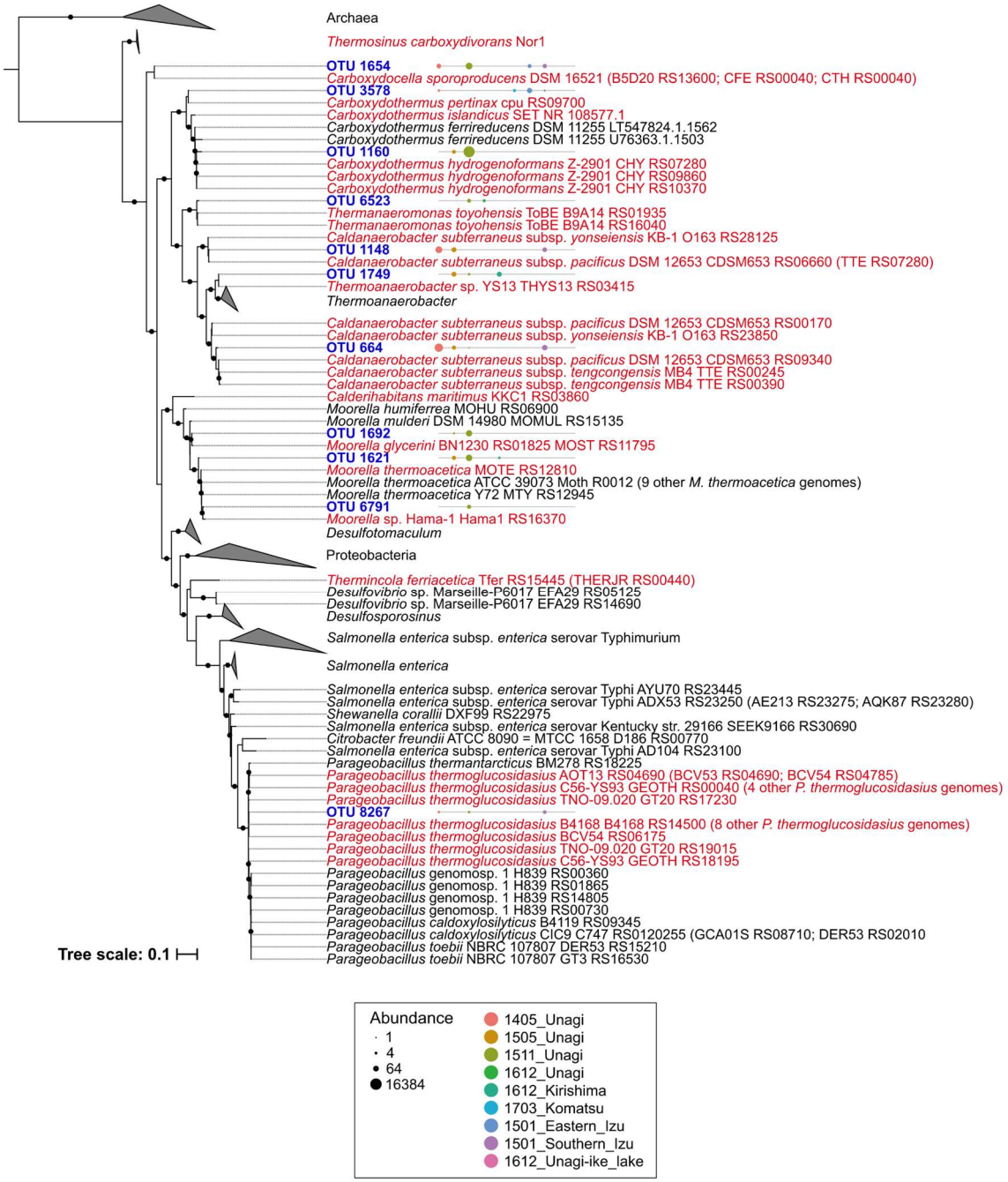


Fig. 3-4. Phylogenetic reconstruction of potential thermophilic hydrogenogenic carboxydotrophic phylotypes of the phylum Firmicutes. Other but identical sequences to their leaves are shown in parenthesis (only one sequence per genome are shown). The phylotype sequences obtained in this study are expressed by 'OTU' prefix. Microorganisms possessing CODH–ECH gene clusters and *C. pertinax* (cpu_RS09700) are shown in red font. Nodes supported by a bootstrap value greater than 80% are indicated by black circles. The bubble plots which are shown at the right of OTUs display the distribution pattern of each phylotype. Abundance is indicated by the number of amplicon reads in each sample

Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs

In the 16S amplicon sequencing analysis, I revealed that the representative sequences of 13 phylotypes showed > 98.7% identity with known thermophilic hydrogenogenic carboxydotrophs or microorganisms possessing CODH–ECH gene clusters, and 10 phylotypes were members of the phylum Firmicutes (Fig. 3-4, Table 3-5). Of these, the representative sequences of OTU_1654 and OTU_3578 were identical to *Carboxydocella* species and *C. pertinax*, respectively, and OTU_664 and OTU_1148 showed 98.8% and 99.5% identities with *C. subterraneus* subspecies, respectively. They were members of group (1). It should be noted that the abundant phylotype OTU_1160 showed 97.7% identity with *Carboxydothermus* species, all of which possess multi CODH gene clusters. The phylotypes that were close to *Thermofilum carboxyditrophus* 1505 (OTU_1051, identity = 99%), *M. thermoacetica* DSM 21394 (OTU_1621, identity = 98.8%; OTU_6791, identity = 99.1%), *M. glycerini* DSM 26271 or *M. stamsii* NMP (OTU_1692, identity = 99.3%), *Thermoanaerobacter* sp. YS13 (OTU_1749, identity = 100%), *Thermococcus barophilus* (OTU_1816, identity = 99%), *T. toyohensis* ToBE (OTU_6523, identity = 99.3%), and *P. thermoglucosidasius* (OTU_8267, identity = 100%), were members of group (2) hydrogenogenic carboxydotrophs, suggesting that these phylotypes are also potential thermophilic hydrogenogenic carboxydotrophs. I also found that OTU_1000 showed 99% identity with Candidatus Korarchaeota archaeon MDKW, whose genome was assembled from Washburn Hot Spring metagenome.

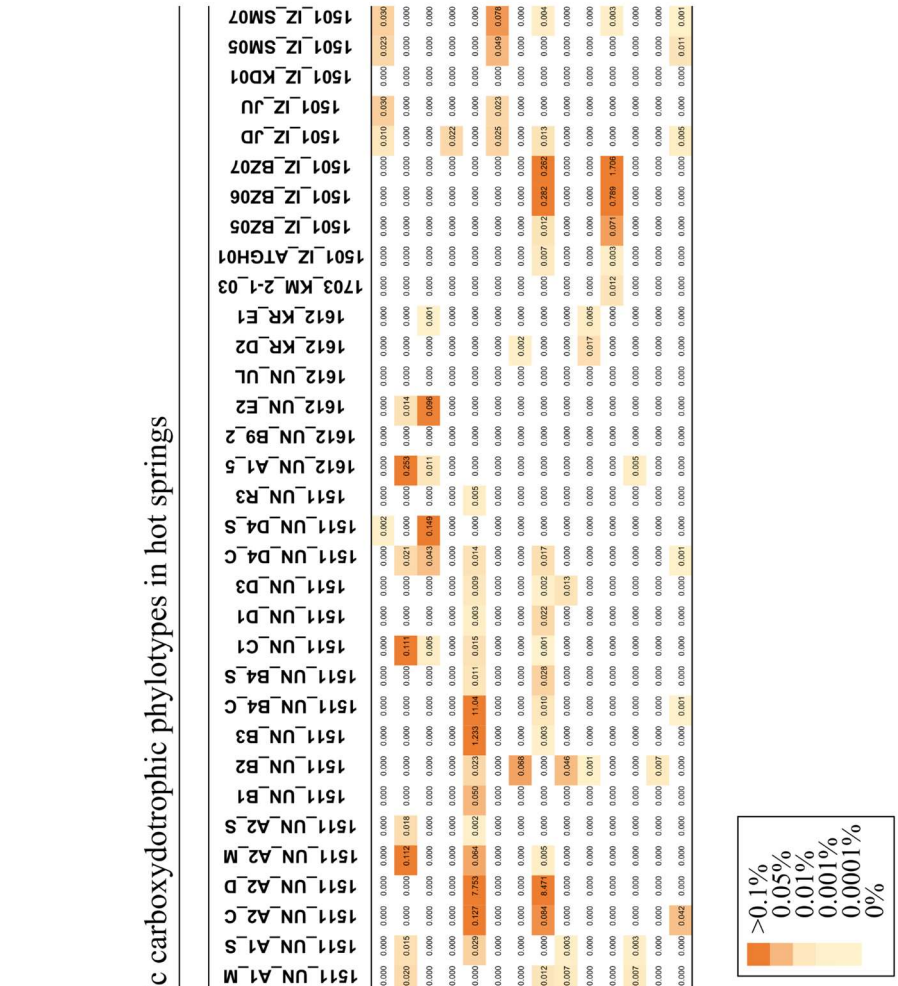
The 13 phylotypes of potential thermophilic hydrogenogenic carboxydotrophs were detected in 45 samples (Tables 3-5, 3-6). Of these, OTU_1654 (*Carboxydocella*), OTU_664 (*C. subterraneus*), OTU_1148 (*C. subterraneus*), OTU_3578 (*C. pertinax*), and OTU_8267 (*P. thermoglucosidasius*) were detected in 7 to 21 samples and widely

distributed in geographically distinct areas (both Kyushu Island and the Izu Peninsula) that showed different environmental conditions and microbial community structures (Tables 3-5, 3-6). OTU_1000, uncultured archaeon phylotype, was also detected widely from 11 samples. The distribution of OTU_1051 (*T. carboxyditrophus*), OTU_1692 (*M. glycerini* or *M. stamsii*), OTU_1749 (*Thermoanaerobacter* sp. YS13), OTU_6523 (*T. toyohensis*), OTU_1621 (*M. thermoacetica* DSM 21394), and OTU_6791 (*M. thermoacetica* DSM 21394) was limited to hot springs in Kyushu Island (mainly in Unagi-onsen in May 2015, November 2015, and December 2016). OTU_1816, the phylotype of *T. barophilus* that was isolated from a deep-sea hydrothermal vent (Marteinsson *et al.*, 1999), was uniquely detected in the saline hot springs in the Izu Peninsula (Tables 3-5, 3-6).

Table 3-5. Potential hydrogenogenic carboxydotrophic phylotypes in hot springs

OTU	Closest hit		
	Locus_tag	Taxon	Identity
OTU_664	CDSM653_RS09340	*Caldanaerobacter subterraneus* subsp. *pacificus* DSM 12653	98.8
OTU_1000	D6D85_RS06615	Candidatus Korarchaeota archaeon MDKW	99.0
OTU_1051	TCARB_RS08035	*Thermofilum carboxyditrophus* 1505	100.0
OTU_1148	O163_RS28125	*Caldanaerobacter subterraneus* subsp. *yonseiensis* KB-1	99.5
OTU_1160	CHY_RS09860	*Carboxydothermus hydrogenoformans* Z-2901	97.7
OTU_1816	TBCH5v1_RS05400	*Thermococcus barophilus* CH5	99.0
OTU_1621	MOTE_RS12810	*Moorella thermoacetica* DSM 21394	98.8
OTU_1654	B5D20_RS13600	*Carboxydocella sporoproducens* DSM 16521	100.0
OTU_1692	BN1230_RS01825	*Moorella glycerini* NMP	99.3
OTU_1749	THYS13_RS07610	*Thermoanaerobacter* sp. YS13	100.0
OTU_3578	NR_113201.1	*Carboxydothermus pertinax* Ug1	100.0
OTU_6523	B9A14_RS16040	*Thermanaeromonas toyohensis* ToBE	99.3
OTU_6791	MOTE_RS12810	*Moorella thermoacetica* DSM 21394	99.1
OTU_8267	B4168_RS14500	*Parageobacillus thermoglucosidasius* B4168	100.0

Table. 3-6. Distribution pattern of potential hydrogenogenic carboxydotrophic phylotypes in hot springs



3. Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs

In most cases, the phylotypes of potential thermophilic hydrogenogenic carboxydotrophs showed a relative abundance of < 0.1% (Tables 3-5, 3-6). Previous studies also suggested that Firmicutes carboxydotroph abundance in hydrothermal environments is usually low (Brady *et al.*, 2015; Yoneda *et al.*, 2015). However, the phylotypes of *C. subterraneus* (OTU_664), *Carboxydocella* (OTU_1654), *C. pertinax* (OTU_3578), and *Carboxydothermus* phylotype (OTU_1160) exhibited a relative abundance of > 0.1% in nine samples (Tables 3-5, 3-6). In particular, I found that the relative abundance of OTU_1654 reached 8.47% per sample at the 1511_UN_A2_D site (70.9 °C, pH 4.68). OTU_1160 was abundant in Unagi-onsen in November 2015, and its relative abundance reached 7.75% and 11% at the 1511_UN_A2_D and 1511_UN_B4_C (94.9 °C, pH 3.65) sites, respectively. However, we could not identify whether the phylotypes, whose relative abundance exceeded 0.1% were growing in these environments, because six of the nine sites showed higher temperature or lower pH than the growth conditions for the isolates of *C. subterraneus* subspecies (50–80 °C, pH 4.5–9.0) (Fardeau *et al.*, 2004), *Carboxydocella* species (40–70 °C, pH 6.2–8.0) (Slepova *et al.*, 2006; Sokolova *et al.*, 2002; Toshchakov *et al.*, 2018), and *Carboxydothermus* species (40–78 °C, pH 4.6−8.6) (Novikov *et al.*, 2011; Svetlichny *et al.*, 1991; Yoneda *et al.*, 2012) (Fig. 3-5). The other three sites including 1511_UN_A2_D showed moderate environmental conditions, where the growth could occur (Fig. 3-5), but the DNA yields from these sites were low (< 15 ng/g sediment). Firmicutes members of *Carboxydothermus*, *Carboxydocella*, and *Caldanaerobacter* are reported to be able to form endospore (Kim *et al.*, 2001; Slepova *et al.*, 2006; Wu *et al.*, 2005). Notably, these groups possessed the genes for endospore formation. It was speculated that DNAs of these phylotypes might persist in such environments longer than those of non-spore-forming

Carboxydotrophs have been suggested to be functionally important, because they mediate a 'currency exchange' between CO and hydrogen in hydrothermal environments (Techtmann *et al.*, 2009). For example, symbiotic interactions have been observed between *C. hydrogenoformans* and thermophilic sulfate reducers in culture, wherein the carboxydotroph provides protection from CO toxicity, whereas H_2 is removed by sulfate reduction, thus reducing end-product inhibition (Parshina *et al.*, 2005a). I investigated the co-occurrence of the potential thermophilic hydrogenogenic carboxydotrophs and other microorganisms using non-parametric Spearman correlations of phylotype presence/absence across all sampling sites. Among the phylotypes present in at least seven sites, networks between OTU_664 and four uncultured microorganisms, and between OTU_1000 and two uncultured bacteria were identified with a Spearman correlation coefficient > 0.8 (Table 3-7). There seem to be no specific symbiotic interactions between most of the potentially hydrogenogenic carboxydotrophic phylotypes and other microorganisms at these sampling sites.

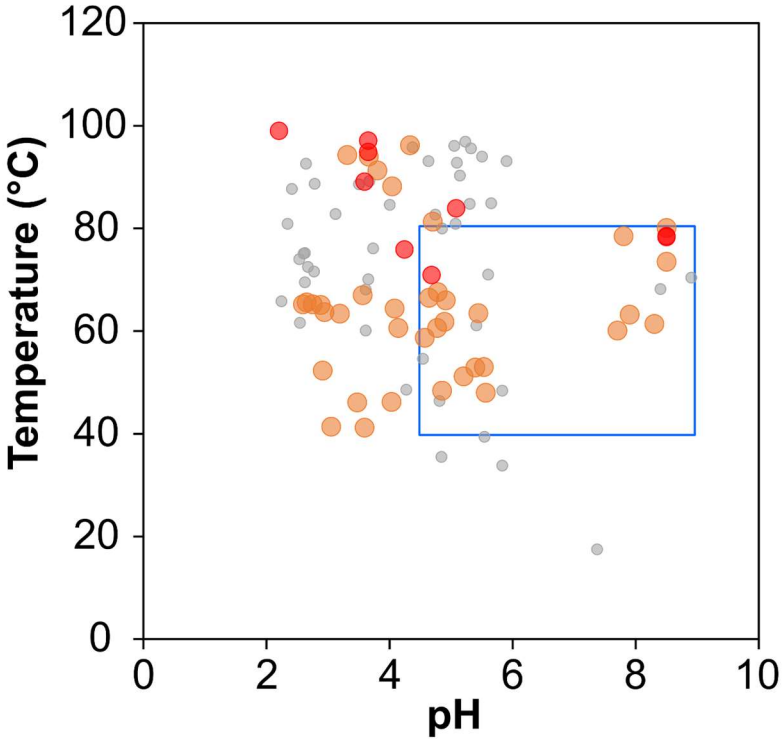


Fig. 3-5. Temperature and pH measured in hot springs. The sampling sites where any potential thermophilic hydrogenogenic carboxydotrophic phylotypes of the Phylum Firmicutes (OTU_664, OTU_1148, OTU_1160, OTU_1621, OTU_1654, OTU_1692, OTU_1749, OTU_3578, OTU_6523, OTU_6791 and OTU_8267) were detected are shown in orange circles. The sites where the relative abundance of *C. subterraneus* (OTU_664), *Carboxydocella* (OTU_1654), *C. pertinax* (OTU_3578) or *Carboxydothermus* phylotype (OTU_1160) exceeded 0.1% are shown in red circles. The other sites are shown in gray circles. The growth range of *Caldanaerobacter*, *Carboxydocella* and *Carboxydothermus* (40-80°C, pH 4.5-9.0) are indicated by blue square.

Table 3-7. Co-occurrence of potential hydrogenogenic carboxydotrophic phylotypes and other microorganisms using non-parametric Spearman correlations of phylotype presence/absence across all sampling sites

OTU	Co-occurring phylotypes		Spearman correlations
	OTU	Closest hit (Identity)	
OTU_664	OTU_21	Firmicutes; *Tepidanaerobacter* anaerobic bacterium TOL (100%)	0.806
OTU_664	OTU_32	Firmicutes; Clostridia D8A-2 uncultured bacterium (96.5%)	0.807
OTU_664	OTU_130	Firmicutes; Clostridia D8A-2 uncultured bacterium (96.8%)	0.883
OTU_664	OTU_425	Proteobacteria; *Silanimonas* uncultured bacterium (100%)	0.874
OTU_1000	OTU_79	Proteobacteria; uncultured *Thermodesulfobacterium* sp. (95.6%)	0.803
OTU_1000	OTU_165	Thermosulfidibacteraeota; *Thermosulfidibacter* uncultured bacterium (96.5%)	0.843

3. Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs

A microbial population whose relative abundance is < 0.1% is called 'rare biosphere' and contributes to a persistent microbial seed bank, which is a collection of dormant microorganisms that can respond to favorable environmental conditions (Lynch & Neufeld, 2015). Endospore formation has an important role for dormancy as well as microbial dispersal (Hubert *et al.*, 2009; Lynch & Neufeld, 2015; Müller *et al.*, 2014; Zeigler, 2014). It was considered that Firmicutes members of the potential thermophilic hydrogenogenic carboxydotrophs found in a variety of hot springs (in most case, as rare biosphere) might form endospores in extreme environmental conditions and have a strategy of microbial seed bank dynamics. The result that most of the potential hydrogenogenic carboxydotrophs did not show any symbiotic networks with other microorganisms also might support the speculation that metabolic activities of these members are low in extreme environments.

Conclusion

This study explored the distribution, diversity, and ecology of thermophilic carboxydotrophs across various hydrothermal environments using microbial community analysis. First, I searched CODH–ECH gene clusters in the current microbial genomic database and revealed 71 genomes encoding CODH–ECH gene clusters. Of these, 46 were genomes whose carriers have never been reported as hydrogenogenic carboxydotrophs. In a microbial community analysis, I identified 13 phylotypes that showed > 98.7% identity with thermophilic members of these taxa. Of these, 10 phylotypes were members of the phylum Firmicutes, and *Parageobacillus*, *Carboxydocella*, *Caldanaerobacter*, and *Carboxydothermus* phylotypes were found across geographically distant hot springs with different environmental conditions,

wherein distinct microbial community structures were formed. Although the relative abundance of the *Carboxydothermus* and *Carboxydocella* phylotypes was greater than 1% at some sites, most of the potentially thermophilic hydrogenogenic carboxydotrophs were usually rare biospheres, whose relative abundances were < 0.1%. They might be in dormant states in extreme environmental conditions. Although symbiotic interactions between hydrogenotrophic microorganisms and hydrogenogenic carboxydotrophs have been suggested (Parshina *et al.*, 2005a), no symbiotic interaction was identified between most of these phylotypes and other microorganisms in my study, leading to the speculation that thermophilic hydrogenogenic carboxydotrophic species might not be active in these environments. However, the previous sediment incubation and cultivation studies have shown that *Carboxydothermus* and *Carboxydocella* species respond to the presence of CO and actively grow (Brady *et al.*, 2015; Kochetkova *et al.*, 2011; Yoneda *et al.*, 2012, 2015). There is also evidence that an unusually high-density population (equivalent to 9.45×10^5 cells/g sediment) of *Carboxydothermus* is present in Unagi-onsen hot springs (Yoneda *et al.*, 2013b), suggesting that they are viable in the environment. While further studies such as transcription analysis are needed to better understand the ecological function of thermophilic hydrogenogenic carboxydotrophs, the present study provides essential information concerning their distribution and diversity in a variety of volcanic environments.

Chapter 4

Diversity analysis of thermophilic hydrogenogenic carboxydotrophs by carbon monoxide dehydrogenase amplicon sequencing using new primers

Summary

The microbial hydrogenogenic carbon monoxide (CO) oxidizing activity achieved by a membrane associated CO-oxidizing and H_2-producing machinery, which is comprised of carbon monoxide dehydrogenase (CODH)/energy converting hydrogenase (ECH) complex, is considered as an important metabolic process in microbial community. In this Chapter, to address diversity of thermophilic hydrogenogenic carboxydotrophs by culture-independent way, I designed six new degenerate primers, which effectively amplified CODH genes associated with ECH (CODHech) of phylum Firmicutes. Amplicon sequencing by these primers in two hot spring sediments with or without incubation under CO gas to enrich endogenous hydrogenogenic carboxydotrophs identified at least six lineages of CODHech genes. The lineages similar to the CODHech genes of previously known hydrogenogenic carboxydotrophs had increased in the CO-enriched samples in which hydrogen productions were observed. This strongly suggested that the new primers detected CODHech genes of active thermophilic hydrogenogenic carboxydotrophs of which abundance in natural habitat was small. The new primers also identified at least two novel lineages of CODHech genes in enriched samples, which might be derived from uncultured hydrogenogenic carboxydotrophs. The new primers I provided here enable us to evaluate diversity of thermophilic hydrogenogenic

carboxydotrophs in vast variety of environments and will pave the way for revealing ecology of these microorganisms.

Introduction

Hydrogenogenic carboxydotrophs can grow by coupling oxidation of carbon monoxide (CO) with H_2 production (Robb & Techtmann, 2018). While CO occurs in wide variety of environments by natural chemical processes and biological processes (Conrad, 1996; Conte *et al.*, 2019; Khalil & Rasmussen, 1990; King & Weber, 2007; Mörsdorf *et al.*, 1992), it is toxic for many microorganisms and inhibits growth (Carvalho *et al.*, 2019; Davidge *et al.*, 2009; Davidova *et al.*, 1994; Nobre *et al.*, 2007; Parshina *et al.*, 2005a; Tavares *et al.*, 2011). Therefore, it is predicted that CO-dependent H_2 production by hydrogenogenic carboxydotrophs is an important metabolic process to reduce toxic CO and supply H_2 which is an energy source for H_2-utilizing microbial communities (Techtmann *et al.*, 2009). The CO-oxidizing and H_2-producing systems is achieved by the coupling of carbon monoxide dehydrogenase (CODH) and membrane-bound H_2-evolving Group 4 [NiFe] hydrogenases which are correctively called energy-converting hydrogenase (ECH) (Søndergaard *et al.*, 2016; Techtmann *et al.*, 2012). These CODH- and ECH-related genes are often found in a single gene cluster in the hydrogenogenic carboxydotroph (Inoue *et al.*, 2019a; Sokolova & Lebedinsky, 2013; Techtmann *et al.*, 2012), and predicted to encode a CODH/ECH complex, which might comprise the simple respiration machinery with an ATP synthase (Schoelmerich & Müller, 2019).

In Chapter 3, the bioinformatics-based analysis identified the 71 genomes of potential hydrogenogenic carboxydotrophs, which harbor CODH–ECH gene cluster,

from ~140,000 prokaryotic genomes (Sayers *et al.*, 2019). Except for six uncultivated members, 33 Firmicutes, 25 Proteobacteria, six Euryarchaeota and one Crenarchaeota strains are included in these potential hydrogenogenic carboxydotrophs. By using the 16S rRNA gene sequences retrieved from the 71 genomes as a reference, we can address these potential hydrogenogenic carboxydotrophs in increasing sequence data by microbial community analysis. However, we cannot define hydrogenogenic carboxydotroph by the 16S rRNA gene sequence, because both strains with and without hydrogenogenic carboxydotrophy occur at a same species (see Chapter 3). Also, we cannot access phylogenetically novel hydrogenogenic carboxydotrophs other than the 71 reference genomes by 16S-targeted amplicon sequencing in microbial community data sets. Therefore, amplicon sequencing by primers targeting the CODH gene of CODH–ECH gene cluster (hereafter called "CODHech gene") is desired for exploring novel hydrogenogenic carboxydotrophs.

Phylogenetically, ~2,000 CODH genes in currently available genomic sequence databases fall into seven clades (Clades A~G) (Inoue *et al.*, 2019a; Techtmann *et al.*, 2012). On the other hand, like CODHech genes, functions of CODHs are often predicted from other genes located in close proximity to themselves (i.e. genomic context) (Inoue *et al.*, 2019a; Techtmann *et al.*, 2012). For example, CODH within an acetyl coenzyme A (acetyl-CoA) synthase (ACS) gene cluster is predicted to be engaged as CO_2-fixing machinery via the Wood-Ljungdahl pathway (Ragsdale, 2004), while CODH adjacent to a CooF gene, which encodes an electron carrier (Fox *et al.*, 1996a), might oxidize CO and transfer electrons to the CooF to produce reducing power. Generally, functions of CODH genes are correlated with the associated gene clusters rather than phylogeny, and several functionally similar CODHs are found in different clades as follows: CODHs in CODH–

ACS, CODH–CooF and CODH–ECH gene clusters are found in Clades A/E/F, Clades C/E/F and Clades E/F, respectively (Inoue *et al.*, 2019a; Techtmann *et al.*, 2012). However, CODH genes with the same function in each clade are phylogenetically related and form subclades within the clade, suggesting both horizontal gene transfer and vertical transmission have driven the remarkable divergence of CODHs (Techtmann *et al.*, 2012).

The high diversity of CODH prevents us from designing universal primer for PCR amplification. To our knowledge, only one study designed degenerate primers to amplify CODH genes for diversity analysis, which were designed to match 27% of the CODH genes in microbial genomes (Matson *et al.*, 2011). The PCR amplification by these primers revealed that CODH genes which are associated with ACS or CooF in Clades C and E are distributed in the hindgut of lower termites and the wood roach (Matson *et al.*, 2011) and deep subseafloor sediments (Hoshino & Inagaki, 2017). However, CODHech genes have never been amplified by these primers. In this Chapter, to address diversity of hydrogenogenic carboxydotrophs, I designed new degenerate primers which effectively amplified CODHech genes for the first time. To cover major taxa of hydrogenogenic carboxydotroph which is predicted to widely distribute in environments (Chapter 3), the new primers were designed to amplify the 34 CODHech genes of the members of phylum Firmicutes (hereafter called "FirmiCODHech genes") which were derived from 12 genera 20 species including one uncultivated strain and formed three subclades within Clades E and F. Amplicon sequencing by these primers in hot spring sediment samples with or without incubation under CO gas to enrich endogenous hydrogenogenic carboxydotrophs identified at least six lineages of CODHech genes within the three subclades. Of these lineages, at least two were revealed as the CODH genes which might be derived from uncultured hydrogenogenic carboxydotrophs.

Materials and Methods

Identification and classification of CODHech genes

The 71 CODH–ECH gene clusters were identified from ~140,000 prokaryotic genomes in the Reference Sequence Database (RefSeq) Database in National Center for Biotechnology Information (NCBI) (December 2018) (Sayers *et al.*, 2019) as described in Chapter 3. In addition, I performed phylogenetic analyses on CODH and ECH catalytic subunit genes for classification of these CODH–ECH gene clusters and primer design. The 1,558 CODH proteins including 47 encoded in the 71 CODHech genes were obtained as described in Chapter 3. I curated the CODH-encoding genomes from the NCBI assembly database (December 2018) by searching 'feature_table' for CODH protein accessions (Sayers *et al.*, 2019), which identified 5,311 CODH genes in 3,050 prokaryotic genomes. The CODH proteins were aligned with the MAFFT using the E-INS-I method (Katoh & Standley, 2013). The multiple sequence alignment (MSA) was subsequently trimmed using the trimAl version 1.4.1 program with a gap-threshold value of 0.9 (Capella-Gutiérrez *et al.*, 2009). Phylogenetic tree was then constructed using the FastTree version 2.1.11 program (Price *et al.*, 2010) with an approximate-maximum-likelihood method using the WAG model. Robustness of the topology of the phylogenetic trees was evaluated by local bootstrap values based on 1000 re-samples. Phylogenetic classifications of CODHs were performed according to previous studies (Inoue *et al.*, 2019a). The tree were visualized using iTOL version 5.2 software (Letunic & Bork, 2016).

For retrieval of ECH catalytic subunit genes, the amino acid sequences of Group 4 [NiFe]-hydrogenase catalytic subunit homologs were obtained from the RefSeq Database in NCBI (December 2018) through a BLASTp search (E value ≤ 0.001) in the BLAST+ using following representative proteins in HydDB (Søndergaard *et al.*, 2016) as

queries: *Escherichia coli* HycE (WP_014639275.1, Group 4a), *E. coli* HyfG (WP_014641051.1, Group 4a), *Pyrococcus abyss* MchD (WP_010868591.1, Group 4b), *Carboxydothermus hydrogenoformans* CooH (WP_011344721.1, Group 4c), *Pyrococcus furiosus* MbhL (WP_011012581.1, Group 4d), *Methanosarcina barkeri* EchA (WP_011305188.1, Group 4e), *Desulfosporosinus orientis* EhfE (WP_014183752.1, Group 4f), *Thermosphaera aggregans* MahB (WP_013129492.1, Group 4g), *Methanothermobacter marburgensis* EhaO (WP_013295617.1, Group 4h) and *Methanothermobacter marburgensis* EhbN (WP_013296415.1, Group 4i). Obtained protein sequences were classified by using HydDB classifier to select Group 4 [NiFe]-hydrogenase (Søndergaard *et al.*, 2016). Further round of quality control was performed by MSA using the MAFFT with the FFT-NS-2 method (Katoh & Standley, 2013) with discarding sequences which lacked conserved cysteine residues required to ligate H_2-binding metal centers (L1 and L2 motifs) (Vignais & Billoud, 2007). I curated the Group 4 [NiFe]-hydrogenase catalytic subunit-encoding genomes as described above, resulting in 3,464 Group 4 [NiFe]-hydrogenase catalytic subunit proteins encoded in 50,441 genes of 38,046 prokaryotic genomes. Phylogenetic tree was constructed as described above with modification, where MSA was performed by MAFFT with the FFT-NS-2 method and trimming and tree-construction were performed with default settings of trimAl and FastTree, respectively.

Design of new CODH-targeted primers

Of the 71 CODHech genes identified, the 34 were derived from Firmicutes members. These 34 FirmiCODHech nucleotide sequences were aligned with the MAFFT, and conserved genetic regions were visualized by calculating the average ratio of dominant

base to all at each position in a 20 bases-length sliding window. I designed six primer sets for each target group, whose specificity was checked by *in silico* PCR using 'primersearch' program in the EMBOSS version 6.6.0 package allowing 10% mismatch (Rice *et al.*, 2000) (Table 4-1).

Collection of environmental samples

Two sediment samples of thermophilic environments, UN and JI, were collected from terrestrial hot springs of Unagi-onsen (temperature, 46.4°C; pH 2.9; ORP, 487 mV) located in the Kagoshima Prefecture (31°13′41″N., 130°36′47″E), Japan, and Jiunji-onsen (temperature, 60.1°C; pH 7.7; ORP, 259 mV) in the Shizuoka Prefecture (34°38′54″N., 138°52′00″E), Japan, in December 2012 and January 2015, respectively. Also, one sediment sample of mesophilic environment, which was used to prepare CODHech-mock community sample, was collected from Unagi-ike lake (temperature, 22.0°C; pH 7.9; ORP, 446 mV) located in the Kagoshima Prefecture (31°13′39″N., 130°36′35″E), Japan, in May 2018. The temperatures, pH and ORP of the sediment pore water were measured as described in Chapter 3, and the samples were packed in a cooler box with ice, transported to the laboratory, and stored at –80°C until use.

Table 4-1. List of the primer sets designed in this study to amplify FirmiCODHech genes

Primer set name	Target clade	Fw primer sequence (5' -> 3')	Rv primer sequence (5' -> 3')	Degeneracy	Expected amplicon size (bp)
E4a_p1	E4a	CCCAGAGCTTGAAGCTTTAGCC	CTACTAGCGCCGCTATACCAC	1	500
F4a_p1	F4a	GTGGTRGGCATCTGCTGYAC	GCGKRAYCTTGACGTTRTTGCA	64	490
F4a_p2	F4a	TGGATTACCAGTGCATCATGCCC	RAACCCGTGGCGCATGAGC	2	473
F4c1_p1	F4c1	GTCGTATYGATCCWTTTGGCAATGG	KTATAATCRGCMAGTGCTCCCTTTA	32	502
F4c1_p2	F4c1	GGSGTGCTGAAGGAAGATGC	RATTGCCTCRGCACTGAAMC	16	501
F4c2_p1	F4c2	GATGCWCAYACCATTGTGGCG	ATAATTCGGTWGACAAATACATTCCGGT	8	478

Enrichment of thermophilic hydrogenogenic carboxydotrophs

To enrich and analyze rare thermophilic hydrogenogenic carboxydotrophs, the sediment of JI was subjected to incubation under 10% CO at 65°C. Approximately 5.0 mL of sediment and pore water of the JI was placed in two glass vials (63.8 mL), which were then sealed with a butyl rubber stoppers. The gas phase of each vial was replaced by CO and N_2 at mixing rations of 10% v/v, and the vials were vigorously vortexed. After five days' incubation at 65°C, each enriched sample was collected and stored at –80°C until DNA extraction. The hydrogen production in each vial after five days' incubation was checked by using a GC-2014 gas chromatograph (Shimadzu, Kyoto, Japan) equipped with a thermal conductivity detector and a ShinCarbon ST packed column (Shinwa Chemical Industries, Kyoto, Japan). Argon was used as the carrier gas.

Preparation for CODHech-mock community sample

To evaluate the specificity and quantitativity of the new CODH-targeted primer sets, I prepared mock community sample containing the cells of four species of thermophilic hydrogenogenic carboxydotrophs of the phylum Firmicutes, which harbors F4c1 or F4a CODHech genes as follows: *C. hydrogenoformans* Z-2901, CHY_RS08505 (F4c1); *Carboxydocella* sp. ULO1, ULO1_RS08880 (F4c1); *Calderihabitans maritimus* KKC1, KKC1_RS06675 (F4c1); *Parageobacillus thermoglucosidasius* DSM 2542, AOT13_RS13420 (F4a). The culture of *C. hydrogenoformans* Z-2901 (= DSM 6008^T) was purchased from Deutsche Sammlung von Mikroorganismenund Zellkulturen (DSMZ). The culture of *P. thermoglucosidasius* DSM 2542 (= NBRC107763^T) was purchased from Biological Resource Center, National Institute of Technology and Evaluation (NBRC). Cells of *Carboxydocella* sp. ULO1 and *C. maritimus* KKC1 were

isolated and maintained in our laboratory (Fukuyama *et al.*, 2017; Yoneda *et al.*, 2013a). The cells of *C. hydrogenoformans* Z-2901 and *Carboxydocella* sp. ULO1 were grown in a modified DSM 507 medium (pH 7.0) (Fukuyama *et al.*, 2018), and the cells of *C. maritimus* KKC1 were grown in a NBRC 1251 medium (pH 7.5). These were grown under 100% CO gas at 65°C by using butyl rubber-stoppered bottles of 100 ml contained 50 ml medium. The cells of *P. thermoglucosidasius* DSM 2542 were grown in a NBRC 802 medium under aerobic condition at 65°C and 100 rotations per minute (rpm) by using a shaking erlenmeyer flask contained 100 ml medium. Rinsed cells of the four species were resuspended in filter-sterilized water containing 8 g/L of NaCl, mixed and added to 2 g of the lake sediment. The sample was stored at –80°C until DNA extraction.

DNA extraction

DNA was extracted from 0.5 g of the samples using an Extrap Soil DNA Kit Plus ver. 2 (Nippon Steel and SUMIKIN Eco-Tech, Tokyo, Japan) following the manufacturer's instructions. During the homogenizing step, I used a bead beater-type homogenizer, Beads Crusher μT-12 (Taitec, Koshigaya, Japan), at a speed of 3,200 rpm for 60 sec. The extracted DNA was stored at –30°C until use.

Quantification of CODHech genes

To reveal the composition, the CODHech gene sequences of *C. hydrogenoformans* Z-2901, *Carboxydocella* sp. ULO1, *C. maritimus* KKC1 and *P. thermoglucosidasius* DSM 2542 in the CODHech-mock community sample were quantified by qPCR. Specificity of each qPCR primer set designed was checked by using Primer-BLAST (Ye *et al.*, 2012). The reaction mixture contained 2 μL of the CODHech-mock community DNA template

with 12.5 μL of TB Green *Premix Ex Taq* II (Tli RNaseH Plus) (TaKaRa Bio, Shiga, Japan), according to the manufacturer's instructions. PCR amplification was performed using the Thermal Cycler Dice real-time system TP850 (TaKaRa). The cycling programs were as follows: 1 min at 95°C for initial denaturation; 38 cycles of 5 s at 95°C; 10 s at 55°C for CHY_RS08505 and ULO1_RS08880, 58°C for KKC1_RS06675, or 60°C for AOT13_RS13420; and 20 s at 72°C. Disassociation curves were created by gradually increasing temperature 60 to 95 °C after PCR cycle to verify amplification specificity. The qPCR standard curve for each targeted gene showed a log-linear relationship when a 10-fold dilution series of PCR products (from 10^1 to 10^7 copies/μL). All qPCR data represent the mean values of triplicate technical determinations.

PCR amplification and sequencing

The primer sets F4a_p1, F4c1_p1 and F4c1_p2 were tested for the CODHech-mock community DNA template, while all pairs of FCEPs were used for environmental and enrichment samples. Overhang adapters were appended at the 5′ end of each primer (forward overhang: 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, reverse overhang: 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) according to the Illumina 16S Metagenomic Sequencing Library Preparation guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). PCR reaction mixture contained 12.5 μL 2× KAPA HiFi HotStart ReadyMix (5 mM Mg^{2+}) (KAPA BIOSYSTEMS, Wilmington, MA, USA), 2.5 μL DNA template, and 5 μL of each primer (10 μM) for final volumes of 25 μL. PCR was performed in “touch down” mode: initial denaturation at 95°C for 3 min; 10 cycles

of 30 s denaturation at 95°C, 30 s annealing at 69–59°C (temperature decreased by 1°C per cycle during the first 10 cycles) and elongation for 30 s at 72°C; 28 cycles of 30 s denaturation at 95°C, 30 s annealing at 59°C and elongation for 30 s at 72°C; final elongation at 72°C for 5 min. The resulting PCR products were examined in 1.5% (w/v) agarose electrophoresis in 1× Tris-acetate EDTA buffer and stained with 3× GelGreen Nucleic Acid Gel Stain (Biotium, Fremont, CA, USA). The bands with the expected sizes were visualized on the Visi-Blue Transilluminator (UVP, Upland, CA, USA), excised and purified with the Wizard SV Gel and PCR CleanUp System (Promega, Madison, WI, USA). To distinguish reads from different PCR products, multiplex barcodes were attached to the amplicons according to Illumina's 16S library preparation guide. DNA concentration of library was determined by Qubit HS dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). The molarity was calculated according to Illumina's 16S library preparation guide. All amplicons were diluted to 1 nM and mixed. Further dilution yielded 12 pM final libraries. The sequencing was performed using the Illumina MiSeq platform with MiSeq V3 (2 × 300 bp) reagent kits (Illumina, San Diego, CA, USA) and with a spike-in of PhiX at 30% to serve as an internal control.

Sequence data processing and analyses

Adapter and primer-binding regions were trimmed from the 5' ends of the forward and reverse reads with the VSEARCH version 2.14.1 program (Rognes *et al.*, 2016). The reads were further processed by trimming low-quality regions from the sequences with the Trimmomatic version 0.36 program (SLIDINGWINDOW: 50:20) (Bolger *et al.*, 2014). Using the VSEARCH, the paired-end reads were joined, and a further round of quality control (QC) was conducted to remove sequences shorter than 200 nt as well as

those containing ambiguous bases (N) or bases with a quality score below 20 (--fastq_mergepairs --fastq_minmergelen 200 --fastq_maxns 0 --fastq_qminout 20). The merged sequences were pooled at each primer set and dereplicated to unique sequences by the VSEARCH (--derep_fulllength). I used stringent denoising strategy by the UNOISE3 algorithm (Edgar, 2016) implemented within the VSEARCH to cluster the remaining sequences into operational taxonomic units (OTUs), where the unique sequences with abundances of < 8 were discarded (--cluster_unoise --id 0.979 --minsize 8 --unoise_alpha 2.0). In OTU clustering, I applied 97.9% identity threshold for species classification, which was determined as described in the next section. Chimeric sequences were removed by UNOISE3 algorithm implemented within the VSEARCH (--uchime3_denovo –abskew 16). OTUs derived from CODH genes were selected by searching on the CODH protein dataset using the DIAMOND version 0.9.22 program (Buchfink *et al.*, 2015), where OTUs with bit score < 80 were discarded. Furthermore, the OTUs which were aligned within each primer target region were selected as 'CODH-OTUs'. For phylogenetic classification, the resulting CODH-OTU amino acid sequences were added to the existing multiple sequence alignment of the CODH proteins by the MAFFT with the E-INS-i method. The phylogenetic tree was constructed and visualized as described above.

Determination of an optimal CODHech gene identity threshold for species classification

To determine the optimal threshold in CODHech gene sequence identity for species classification, I calculated precision-recall and F-measure (Kim *et al.*, 2014; van Rijsbergen, 1979) using taxonomic information and pairwise sequence identities of the

71 CODHech gene sequences, and selected the identity threshold with the highest F-measure as the optimal species cut-off. First, I assigned taxonomic information of the Genome Taxonomy Database (GTDB) release 89 (Parks *et al.*, 2018), which is a standardized microbial taxonomy based on genome phylogeny, to the 71 genomes harboring CODH–ECH gene clusters. When there is no GTDB entry, genomes were additionally assigned to the GTDB taxonomy by using the GTDB-Tk version 2.2 program (Chaumeil *et al.*, 2019). For the four genomes of Rhizobiales bacteria without species level taxonomic information, the average nucleotide identity (ANI) between the genomes was calculated by using the FastANI version 1.1 program (Jain *et al.*, 2018), and the genomes with <95% ANI, which is a typical species ANI circumscription (Chaumeil *et al.*, 2019), were assigned as the same species. Full-length pairwise sequence identity was computed for each pair of the 71 CODHech gene sequences by using needleall in the EMBOSS (-gapopen 10.0 -gapextend 0.5) (Rice *et al.*, 2000), and the resulting value was noted along with the taxonomic relationship between the sequences, i.e., intraspecies or interspecies. A grid search approach was implemented to test all possible cutoff values between 80% and 100% sequence identity with a step-size of 0.1%. For each possible cutoff value, the number of sequences that were correctly (true positives [TP] and true negatives [TN]) and incorrectly (false positives [FP] and false negatives [FN]) placed were computed in the species-level comparisons. Precision and recall values were calculated as follows: Precision = TP/(TP + FP) and Recall = TP/(TP + FN). Afterwards, these values were used to calculate the F-measure, which is a harmonic mean of precision and recall and represents an accuracy of the test.

Results

The optimal CODHech gene identity threshold for species classification

Prior to development of the CODHech-targeted amplicon sequencing, I determined the optimal threshold in CODHech gene sequence identity for species classification which were used for OTU clustering and classification. I used genome taxonomic information which were assigned by GTDB, and pairwise sequence identities of the 71 CODHech gene sequences for calculation of F-measure (Fig. 4-1; Table 4-2). According to this analysis, pairwise sequence identity values between 94.5% and 97.9% marked the highest F-measure and were considered to be optimal (Fig. 4-1B). In this Chapter, I applied more stringent identity value (i.e. 97.9%) for OTU clustering cutoff. On the other hand, I assigned a CODH-OTU at the species level by using ≥94.5% pairwise identity with the closest CODHech gene. I applied the lower threshold in CODH-OTU classification to conservertively estimate whether it is derived from uncultured species.

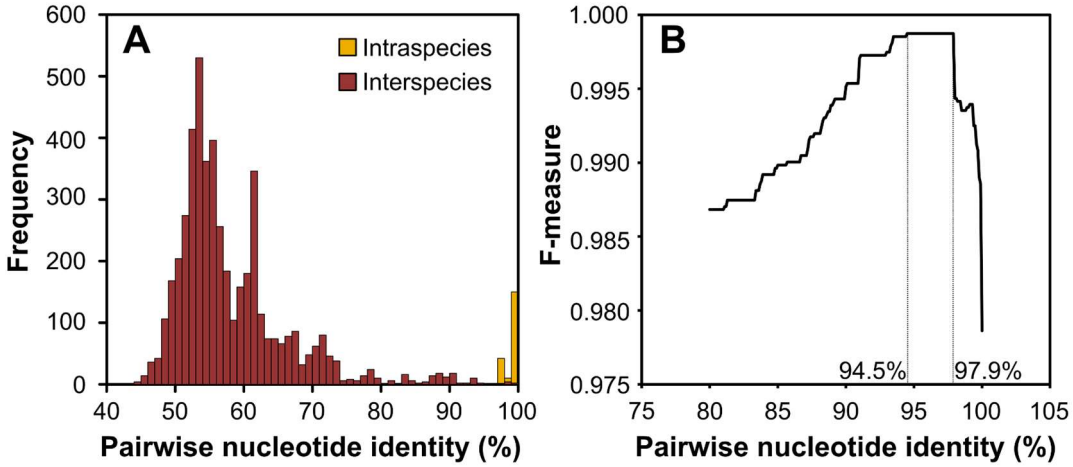


Fig. 4-1. Optimal thresholds for species classifications of CODHech genes. Pairwise sequence identities distribution (A) and F-measure (B) based on all CODHech genes. Each color represents pairwise nucleotide identity calculations between two strains belonging to the same species (yellow), and belonging to different species (red) in (A). The highest F-measure was found at 94.5–97.9% pairwise identity for species classification in (B).

Table 4-2. GTDB taxonomy of genomes harboring CODHech genes

Accession	NCBI taxonomy	GTDB taxonomy
GCF_003947435.1	*Candidatus* Methanodesulfokores washburnensis	*Candidatus* Methanodesulfokores washburnensis[a]
GCF_000813245.1	*Thermofilum carboxyditrophus* 1505	*Thermofilum carboxyditrophus*
GCF_000816105.1	*Thermococcus guaymasensis* DSM 11113	*Thermococcus guaymasensis*
GCF_000018365.1	*Thermococcus onnurineus* NA1	*Thermococcus onnurineus*
GCF_000151205.2	*Thermococcus* sp. AM4	*Thermococcus* sp000151205
GCF_000151105.2	*Thermococcus barophilus* MP	*Thermococcus_B barophilus*
GCF_001433455.1	*Thermococcus barophilus* CH5	*Thermococcus_B barophilus*
GCF_000517445.1	*Thermococcus paralvinellae* ES1	*Thermococcus_B paralvinellae*
GCF_000166075.1	*Parageobacillus thermoglucosidasius* Y4.1MC1	*Parageobacillus thermoglucosidasius*
GCF_000178395.2	*Parageobacillus thermoglucosidasius* C56-YS93	*Parageobacillus thermoglucosidasius*
GCF_000258725.1	*Parageobacillus thermoglucosidasius* TNO-09.020	*Parageobacillus thermoglucosidasius*
GCF_000648295.1	*Parageobacillus thermoglucosidasius* NBRC 107763	*Parageobacillus thermoglucosidasius*
GCF_000966225.1	*Parageobacillus thermoglucosidasius* DSM 2542	*Parageobacillus thermoglucosidasius*
GCF_001295365.1	*Parageobacillus thermoglucosidasius* DSM 2542	*Parageobacillus thermoglucosidasius*
GCF_001587555.1	*Parageobacillus thermoglucosidasius* B4168	*Parageobacillus thermoglucosidasius*
GCF_001651535.1	*Parageobacillus thermoglucosidasius* GT23	*Parageobacillus thermoglucosidasius*
GCF_001700985.1	*Parageobacillus thermoglucosidasius* NCIMB 11955	*Parageobacillus thermoglucosidasius*
GCF_001902495.1	*Parageobacillus thermoglucosidasius* TM242	*Parageobacillus thermoglucosidasius*
GCF_003865195.1	*Parageobacillus thermoglucosidasius* TG4	*Parageobacillus thermoglucosidasius*
GCF_000156275.2	*Caldanaerobacter subterraneus* subsp. *pacificus* DSM 12653	*Caldanaerobacter subterraneus*
GCF_000007085.1	*Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4	*Caldanaerobacter subterraneus*
GCF_000473865.1	*Caldanaerobacter subterraneus* subsp. *yonseiensis* KB-1	*Caldanaerobacter subterraneus*
GCF_000806225.2	*Thermoanaerobacter* sp. YS13	*Thermoanaerobacter uzonensis*
GCF_001936615.1	*Desulfosporosinus* sp. OL	*Desulfosporosinus* sp001936615
GCF_000012865.1	*Carboxydothermus hydrogenoformans* Z-2901	*Carboxydothermus hydrogenoformans*
GCF_001950325.1	*Carboxydothermus islandicus* SET	*Carboxydothermus islandicus*

Table 4-2. Continued.

GCF_000214435.1	*Desulfotomaculum nigrificans* CO-1-SRB	*Desulfotomaculum nigrificans*
GCF_002207765.1	*Calderihabitans maritimus* KKC1	*Calderihabitans maritimus*
GCF_003116935.1	*Moorella* sp. Hama-1	*Moorella* sp003116935
GCF_001373375.1	*Moorella glycerini* NMP	*Moorella stamsii*
GCF_002995805.1	*Moorella stamsii* DSM 26271	*Moorella stamsii*
GCF_001875325.1	*Moorella thermoacetica* DSM 21394	*Moorella thermoacetica_A*
GCF_900176005.1	*Thermanaeromonas toyohensis* ToBE	*Thermanaeromonas toyohensis*
GCF_000746025.1	Clostridiales bacterium DRI-13	DRI-13 sp000746025
GCF_000092945.1	*Thermincola potens* JR	*Thermincola ferriacetica*
GCF_001263415.1	*Thermincola ferriacetica* Z-0001	*Thermincola ferriacetica*
GCF_002049255.1	*Carboxydocella* sp. ULO1	*Carboxydocella thermautotrophica*
GCF_002049395.1	*Carboxydocella* sp. JDF658	*Carboxydocella thermautotrophica*
GCF_003047205.1	*Carboxydocella thermautotrophica* 19	*Carboxydocella thermautotrophica*
GCF_003054495.1	*Carboxydocella thermautotrophica* 41	*Carboxydocella thermautotrophica*
GCF_900167165.1	*Carboxydocella sporoproducens* DSM 16521	*Carboxydocella thermautotrophica*
GCF_000169155.1	*Thermosinus carboxydivorans* Nor1	*Thermosinus carboxydivorans*
GCF_900167065.1	*Desulfovibrio bizertensis* DSM 18034	*Desulfovibrio_O bizertensis*
GCF_000341895.1	*Pseudodesulfovibrio piezophilus* C1TLV30	*Pseudodesulfovibrio piezophilus*
GCF_000817955.1	*Geobacter pickeringii* G13	*Geobacter pickeringii*
GCF_000020725.1	*Geobacter bemidjiensis* Bem	*Geobacter_A bemidjiensis*
GCF_002770725.1	*Pleomorphomonas carboxyditropha* SVCO-16	*Pleomorphomonas* sp900095415
GCF_900466875.1	Rhizobiales bacterium AFS016371	*Neorhizobium* unculture_A[b]
GCF_900468955.1	Rhizobiales bacterium AFS041951	*Neorhizobium* unculture_B[b]
GCF_900469445.1	Rhizobiales bacterium AFS049984	*Neorhizobium* unculture_B[b]
GCF_900472805.1	Rhizobiales bacterium AFS089140	*Neorhizobium* unculture_B[b]
GCF_900104145.1	*Pseudovibrio* sp. Tun.PSC04-5.I4	*Pseudovibrio* sp900104145
GCF_001431305.1	*Pseudovibrio* sp. POLY-S9	*Pseudovibrio* sp900143565
GCF_000013745.1	*Rhodopseudomonas palustris* BisB18	*Rhodopseudomonas palustris_D*

Table 4-2. Continued.

GCF_000013085.1	*Rhodospirillum rubrum* ATCC 11170	*Rhodospirillum rubrum*
GCF_000225955.1	*Rhodospirillum rubrum* F11	*Rhodospirillum rubrum*
GCF_000981805.1	*Citrobacter amalonaticus* Y19	*Citrobacter_A amalonaticus_C*
GCF_000263295.1	*Salmonella enterica* subsp. *enterica* serovar Senftenberg SS209	*Salmonella enterica*
GCF_001276695.1	*Salmonella enterica* subsp. *enterica* serovar Montevideo 50262	*Salmonella enterica*
GCF_001276745.1	*Salmonella enterica* subsp. *enterica* serovar Senftenberg 50263	*Salmonella enterica*
GCF_001276765.1	*Salmonella enterica* subsp. *enterica* serovar Senftenberg 50264	*Salmonella enterica*
GCF_001276775.1	*Salmonella enterica* subsp. *enterica* serovar Senftenberg 50265	*Salmonella enterica*
GCF_001276825.1	*Salmonella enterica* subsp. *enterica* serovar Senftenberg 50271	*Salmonella enterica*
GCF_001276905.1	*Salmonella enterica* subsp. *enterica* serovar Montevideo 50270	*Salmonella enterica*
GCF_001276925.1	*Salmonella enterica* subsp. *enterica* serovar Senftenberg 50272	*Salmonella enterica*
GCF_000422645.1	*Ferrimonas futtsuensis* DSM 18154	*Ferrimonas futtsuensis*
GCF_000425405.1	*Ferrimonas kyonanensis* DSM 18153	*Ferrimonas kyonanensis*
GCF_900100175.1	*Ferrimonas sediminum* DSM 23317	*Ferrimonas sediminum*
GCF_003797165.1	*Shewanella* sp. R106	GCF_002836275.1
GCF_003855155.1	*Shewanella* sp. M2	GCF_002836275.1
GCF_000331515.1	*Photobacterium marinum* AK15	*Photobacterium marinum*

[a] NCBI taxonomy was used, because GTDB taxonomy was not assigned.

[b] Species-level taxonomy was manually assigned by calculating ANI.

New primer sets for PCR amplification of FirmiCODHech genes

Of the 71 CODHech genes identified in current genomic databases, the 34 were derived from the members of phylum Firmicutes which is a major taxon of hydrogenogenic carboxydotroph with wide distribution (Chapter 3). In this study, I designed new primers targeting these 34 FirmiCODHech genes. According to the phylogenetic analysis, the FirmiCODHech genes were found in three different subclades within both Clades E and F, which were named E4a, F4a and F4c according to the phylogenetic clade of the FirmiCODHech and the class of the associated ECH (Group 4 [NiFe] hydrogenase) catalytic subunit (Table 4-3, Fig. 4-2). While I couldn't identify the conserved regions which might be suitable for designing oligonucleotide primers for PCR amplification in the multiple sequence alignment of the 34 FirmiCODHech nucleotide sequences, conserved regions were found in each subclade (note that only the F4c had to be split into F4c1 and F4c2) (Fig. 4-3). I newly designed six primer sets for each target group, which collectively covered all the 34 FirmiCODHech genes (Table 4-1, Fig. 4-4). Note that the F4a_p2 and F4c1_p2 were designed as more specific primer sets with less degeneracy than F4a_p1 and F4c1_p1, respectively.

The specificity and quantitativity of these primer sets were evaluated by PCR amplification using the CODHech-mock community sample. The DNA extracted from the CODHech-mock community sample contained FirmiCODHech genes of *C hydrogenoformans* Z-2901 (F4c1), *Carboxydocella* sp. ULO1 (F4c1), *C. maritimus* KKC1 (F4c1) and *P. thermoglucosidasius* DSM 2542 (F4a) with the concentration of 36–760 copies/uL, which were quantified by qPCR using specifically designed primers with 93.7–97.7% efficiency and ≥0.998 R^2 values (Table 4-4, Figs. 4-4, 4-5). The three primer sets targeting F4c1 and F4a subclades, the F4c1_p1, F4c1_p2 and F4a_p1 (Fig. 4-4), were

tested for the PCR amplification, which resulted in the products with expected sizes. Sequencing of the PCR products produced >64,000 raw paired-end reads per primer set, and the sequence processing left >28,000 chimera- and noise-free merged reads per primer set (Table 4-5). The CODH-OTUs generated by the F4c1_p1, F4c1_p2 and F4a_p1 were identical to the targeting FirmiCODHech genes of each subclades (Fig. 4-5). Furthermore, compositions of these CODH-OTUs were comparable to the result of the quantification of the FirmiCODHech genes by qPCR (Fig. 4-5). These results collectively indicated that the newly designed primer sets specifically amplified each target and the amplicon sequencing by using them reflects the composition of CODHech genes in the sample. Noted that F4a_p1 generated the 42 extra CODH-OTUs, which accounted for ~60% of the reads, other than the CODH-OTU identical to target FirmiCODHech of *P. thermoglucosidasius* DSM 2542 (Table 4-5). These OTUs were not classified into F4a subclade, and phylogenetically novel forming new clades within Clades E and F (Fig. 4-6), suggesting that these ‘noisy CODH-OTUs’ were derived from unknown species in the sediment sample used for preparation of the CODHech-mock community sample.

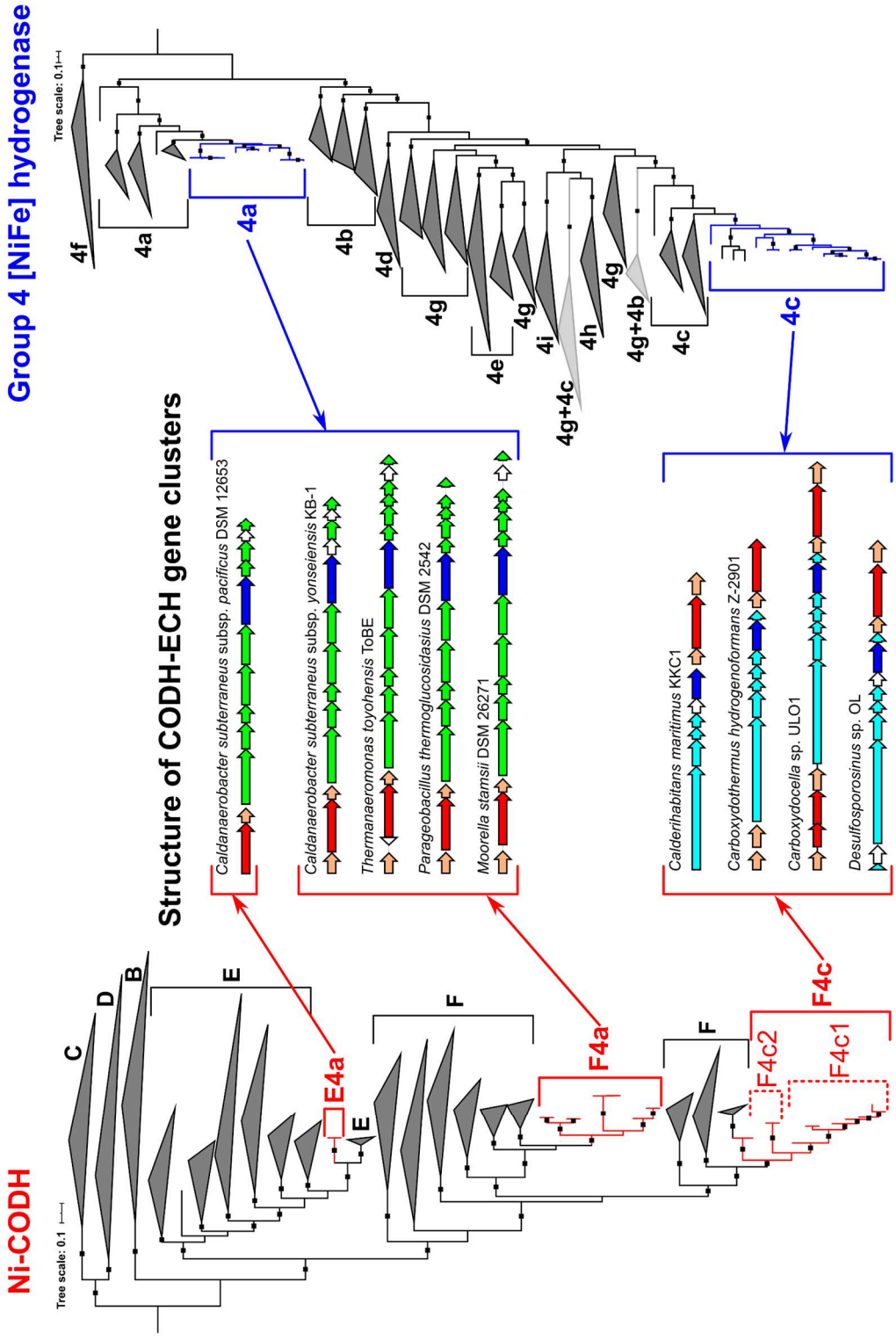


Fig. 4-2. Structural and phylogenetic classification of CODH–ECH gene clusters. Phylogenetic trees of CODH genes and Group 4 [NiFe] hydrogenase catalytic subunit genes are shown at left and right, respectively. Names of major clades are shown in black characters and branches in the trees according to the previous studies (Inoue *et al.*, 2019b; Søndergaard *et al.*, 2016; Techtmann *et al.*, 2012). Black squares indicate >0.8 support by bootstrap values. Structures of representative CODH–ECH gene clusters are shown at center. Red and blue colors indicate CODH and Group 4 [NiFe] hydrogenase catalytic subunit genes within the CODH–ECH gene clusters. Orenge arrows indicate CODH-related genes (*cooF* or *cooC*). Light green and lignt blue arrows indicate Group 4a and 4c [NiFe] hydrogenase related genes.

Table 4-3. Genomes harboring FirmiCODHech and associated ECH (Group 4 [NiFe] hydrogenase)

Genome		FirmiCODHech			Group 4 [NiFe] hydrogenase catalytic subunit		
Accession	**Taxonomy**	**Accession**	**Locus tag**	**Sub-clade**	**Accession**	**Locus tag**	**Class**
GCF_000156275.2	*Caldanaerobacter subterraneus* subsp. *pacificus* DSM 12653	WP_009610668.1	CDSM653_RS07355	E4a	WP_009610669.1	CDSM653_RS07320	4a
GCF_000806225.2	*Thermoanaerobacter* sp. YS13	WP_042833370.1	THYS13_RS04710	E4a	WP_042833365.1	THYS13_RS04650	4a
GCF_000166075.1	*Parageobacillus thermoglucosidasius* Y4.1MC1	WP_013400775.1	GY4MC1_RS09120	F4a	WP_013400781.1	GY4MC1_RS09155	4a
GCF_000178395.2	*Parageobacillus thermoglucosidasius* C56-YS93	WP_013400775.1	GEOTH_RS09250	F4a	WP_013876883.1	GEOTH_RS09285	4a
GCF_000258725.1	*Parageobacillus thermoglucosidasius* TNO-09.020	WP_003250289.1	GT20_RS08580	F4a	WP_003250304.1	GT20_RS08615	4a
GCF_000648295.1	*Parageobacillus thermoglucosidasius* NBRC 107763	WP_013400775.1	GT2_RS13950	F4a	WP_013400781.1	GT2_RS13985	4a
GCF_000966225.1	*Parageobacillus thermoglucosidasius* DSM 2542	WP_013400775.1	WH82_RS08365	F4a	WP_013400781.1	WH82_RS08330	4a
GCF_001295365.1	*Parageobacillus thermoglucosidasius* DSM 2542	WP_013400775.1	AOT13_RS13420	F4a	WP_013400781.1	AOT13_RS13385	4a
GCF_001587555.1	*Parageobacillus thermoglucosidasius* B4168	WP_003250289.1	B4168_RS01110	F4a	WP_003250304.1	B4168_RS01145	4a
GCF_001651535.1	*Parageobacillus thermoglucosidasius* GT23	WP_003250289.1	GT23_RS00370	F4a	WP_003250304.1	GT23_RS00335	4a
GCF_001700985.1	*Parageobacillus thermoglucosidasius* NCIMB 11955	WP_013400775.1	BCV53_RS13415	F4a	WP_013400781.1	BCV53_RS13380	4a
GCF_001902495.1	*Parageobacillus thermoglucosidasius* TM242	WP_013400775.1	BCV54_RS13805	F4a	WP_013400781.1	BCV54_RS13770	4a
GCF_003865195.1	*Parageobacillus thermoglucosidasius* TG4	WP_125009789.1	PTHTG4_RS08120	F4a	WP_125009791.1	PTHTG4_RS08155	4a
GCF_000007085.1	*Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4	WP_011025912.1	TTE_RS08175	F4a	WP_009610669.1	TTE_RS08140	4a
GCF_000473865.1	*Caldanaerobacter subterraneus* subsp. *yonseiensis* KB-1	WP_022587816.1	O163_RS22180	F4a	WP_022587809.1	O163_RS22145	4a
GCF_003116935.1	*Moorella* sp. Hama-1	WP_109207247.1	Hama1_RS10675	F4a	WP_109207254.1	Hama1_RS10710	4a
GCF_001373375.1	*Moorella glycerini* NMP	WP_054936715.1	BN1230_RS08005	F4a	WP_054936708.1	BN1230_RS07970	4a
GCF_002995805.1	*Moorella stamsii* DSM 26271	WP_054936715.1	MOST_RS16220	F4a	WP_054936708.1	MOST_RS16185	4a
GCF_001875325.1	*Moorella thermoacetica* DSM 21394	WP_075516371.1	MOTE_RS04415	F4a	WP_075516364.1	MOTE_RS04380	4a
GCF_900176005.1	*Thermanaeromonas toyohensis* ToBE	WP_084664092.1	B9A14_RS03610	F4a	WP_084664078.1	B9A14_RS03575	4a
GCF_000012865.1	*Carboxydothermus hydrogenoformans* Z-2901	WP_011344718.1	CHY_RS08505	F4c1	WP_011344721.1	CHY_RS08520	4c
GCF_001950325.1	*Carboxydothermus islandicus* SET	WP_075865517.1	ciss_RS06710	F4c1	WP_075865520.1	ciss_RS06725	4c
GCF_000214435.1	*Desulfotomaculum nigrificans* CO-1-SRB	WP_013809564.1	DESCA_RS01885	F4c1	WP_013809561.1	DESCA_RS01870	4c
GCF_002207765.1	*Calderihabitans maritimus* KKC1	WP_088553704.1	KKC1_RS06675	F4c1	WP_088553703.1	KKC1_RS06665	4c
GCF_000746025.1	Clostridiales bacterium DRI-13	WP_034425146.1	BR63_RS15745	F4c1	WP_034425152.1	BR63_RS15760	4c

Table 4-3. Continued.

GCF_002049255.1	*Carboxydocella* sp. ULO1	WP_078665971.1	ULO1_RS08880	F4c1	WP_078665968.1	ULO1_RS08895	4c
GCF_002049395.1	*Carboxydocella* sp. JDF658	WP_078665971.1	JDF658_RS09120	F4c1	WP_079907326.1	JDF658_RS09135	4c
GCF_003047205.1	*Carboxydocella thermautotrophica* 19	WP_078665971.1	CFE_RS00860	F4c1	WP_078665968.1	CFE_RS00845	4c
GCF_003054495.1	*Carboxydocella thermautotrophica* 41	WP_078665971.1	CTH_RS00865	F4c1	WP_078665968.1	CTH_RS00850	4c
GCF_900167165.1	*Carboxydocella sporoproducens* DSM 16521	WP_078665971.1	B5D20_RS09335	F4c1	WP_078665968.1	B5D20_RS09320	4c
GCF_000169155.1	*Thermosinus carboxydivorans* Nor1	WP_007288856.1	TCARDRAFT_RS04635	F4c1	WP_007288859.1	TCARDRAFT_RS04650	4c
GCF_001936615.1	*Desulfosporosinus* sp. OL	WP_075366373.1	DSOL_RS18930	F4c2	WP_075366375.1	DSOL_RS18945	4c
GCF_000092945.1	*Thermincola potens* JR	WP_013121776.1	THERJR_RS14890	F4c2	WP_013121779.1	THERJR_RS14905	4c
GCF_001263415.1	*Thermincola ferriacetica* Z-0001	WP_052217746.1	Tfer_RS07525	F4c2	WP_052217742.1	Tfer_RS07510	4c

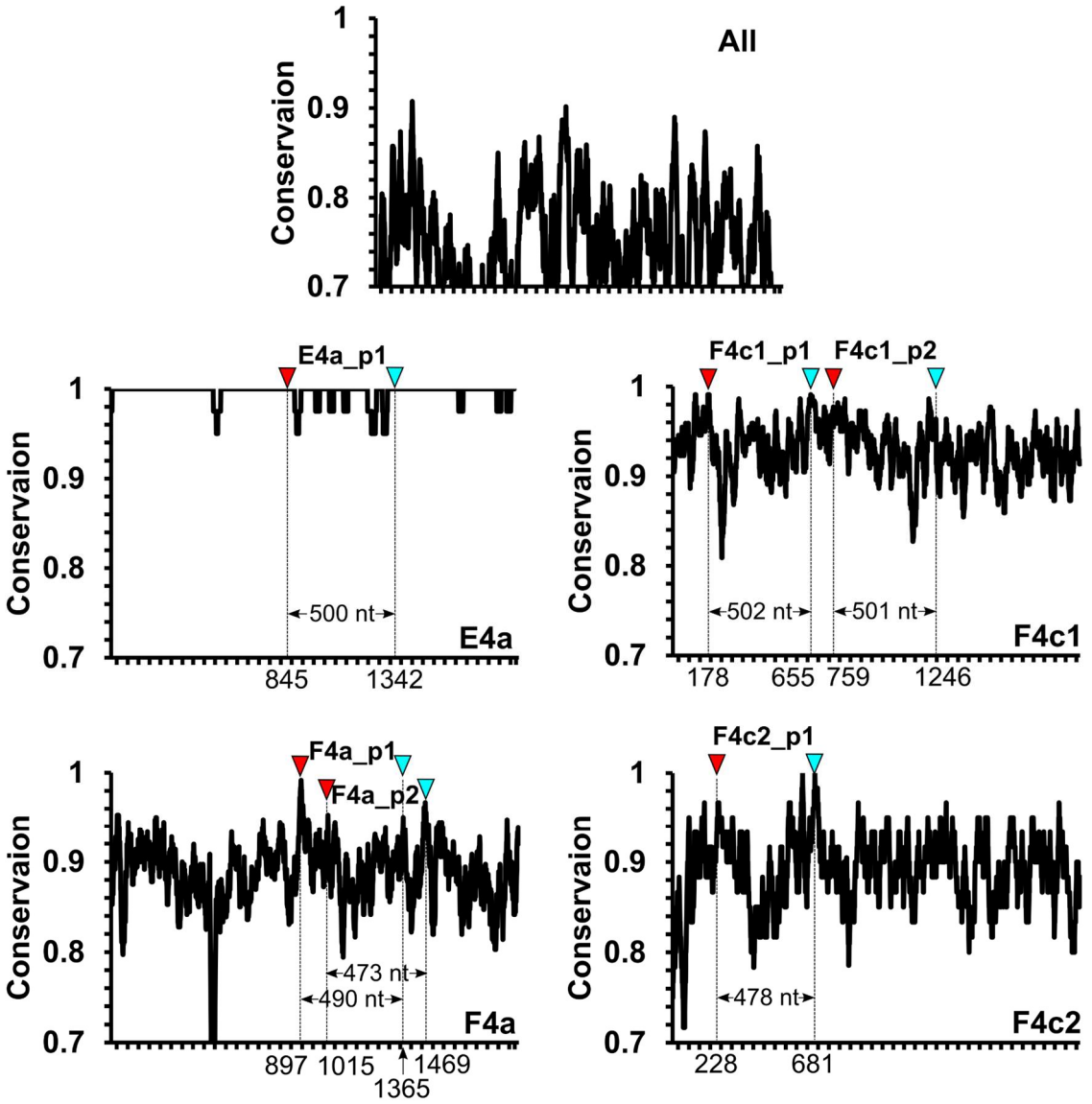


Fig. 4-3. Conservation of the FirmiCODHech nucleotide sequences and the primer regions. The average ratio of dominant base to all at each position in a 20 bases-length sliding window in each was calculated and shown in each multiple sequence alignments: All indicates the all 34 Firmi CODHech genes were used, and E4a, F4a, F4a1 and F4c2 indicates the FirmiCODHech genes of each subclade were used. Red and light blue arrow heads indicate the positions of forward and reverse primers of the each new primer set, respectively. The primer set names are shown between the arrow heads, and expected amplicon size are shown in the graph.

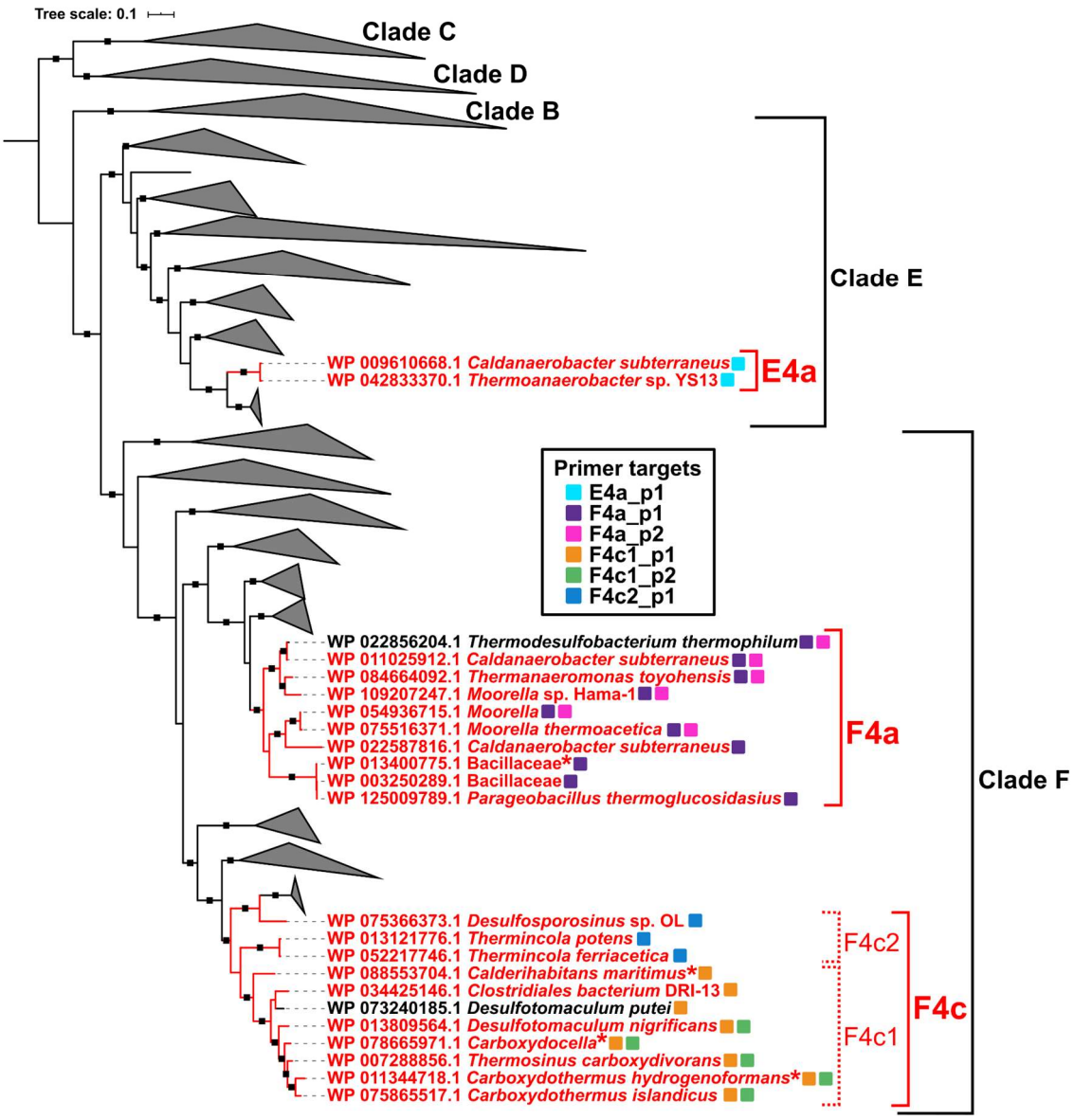


Fig. 4-4. Phylogeny of FirmiCODHech genes and targets of new primers. The tree was constructed using an alignment of 1,558 CODH proteins. Major clades B through F are indicated in the right. FirmiCODHech subclades are assigned and colored in red. Branches and leaves are indicated by different colors as follows: red, FirmiCODHechs; black, other CODHs. The targets of each new primer set are shown by squares placed on the right of leaves in different colors. Asterisks indicate FirmiCODHech genes added in the CODHech-mock community sample. The black squares on the branch indicate >0.8 support by bootstrap values.

Table 4-4. Primer sequences and efficiency of amplification for qPCR

Primer target	Fw primer sequence (5' -> 3')	Rv primer sequence (5' -> 3')	Amplification efficiency (%)	R^2
CHY_RS08505 (*Carboxydothermus hydrogenoformans* Z-2901)	GCTAATGACCCGTTAAAACC	GTTACAACCGGCAAATAAGG	93.7	0.999
ULO1_RS08880 (*Carboxydocella* sp. ULO1)	TAACAATCCTAAGGCGATCC	TGCCTCCTGAGTCATTAAAC	96.4	0.998
KKC1_RS06675 (*Calderihabitans maritimus* KKC1)	TGTCGCATTTGTTGGAAGGG	ATGGTATCGGCATCAGCACC	93.9	0.999
AOT13_RS13420 (*Parageobacillus thermoglucosidasius* DSM 2542)	TCCGCGGAGTCTGTCTCTTTG	CAACCAGTAGCCAGCAGCAG	97.7	0.999

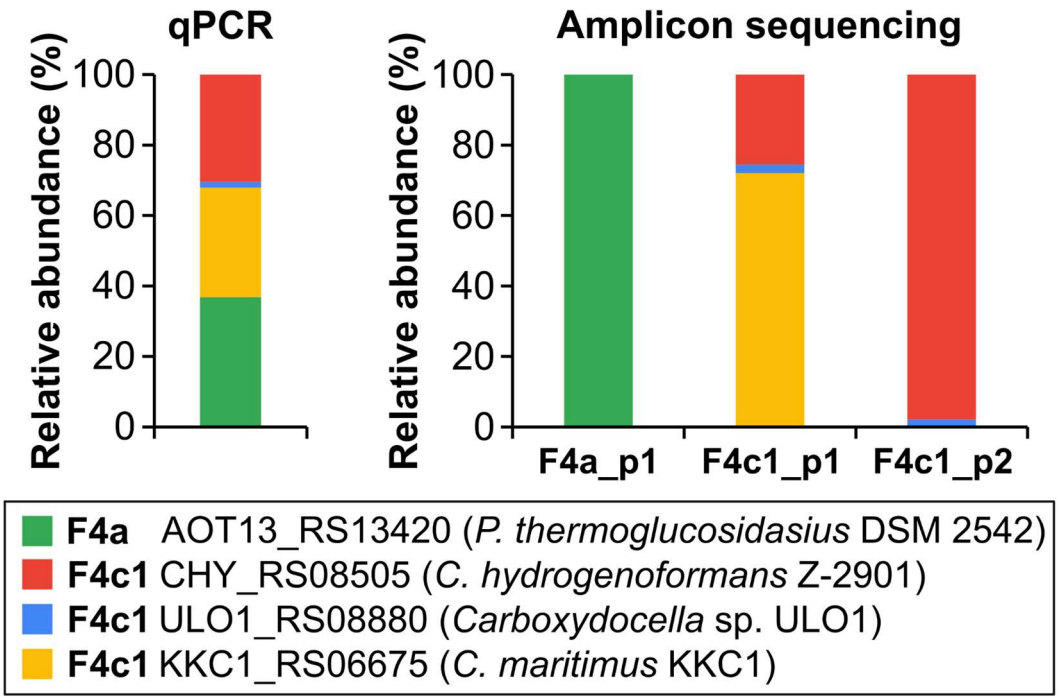


Fig. 4-5. Evaluation of specificity and quantitativity of the new primers using the CODHech-mock community sample. Composition of each FirmiCODHech genes estimated by qPCR is shown in left. Relative abundance of each CODH-OTU in amplicon sequencing by each new primer set is shown in right. Each FirmiCODHech genes are shown in different colors.

Table 4-5. Read statistics of amplicon sequencing in CODHech-mock community sample

Primer sets	Total reads	Merged QC reads	Denoised chimera-free reads	CODH reads	CODH OTUs
F4a_p1	122,285	88,442	42,364	42,354	43
F4c1_p1	65,790	57,720	30,129	30,129	3
F4c1_p2	64,059	54,321	28,610	28,610	2

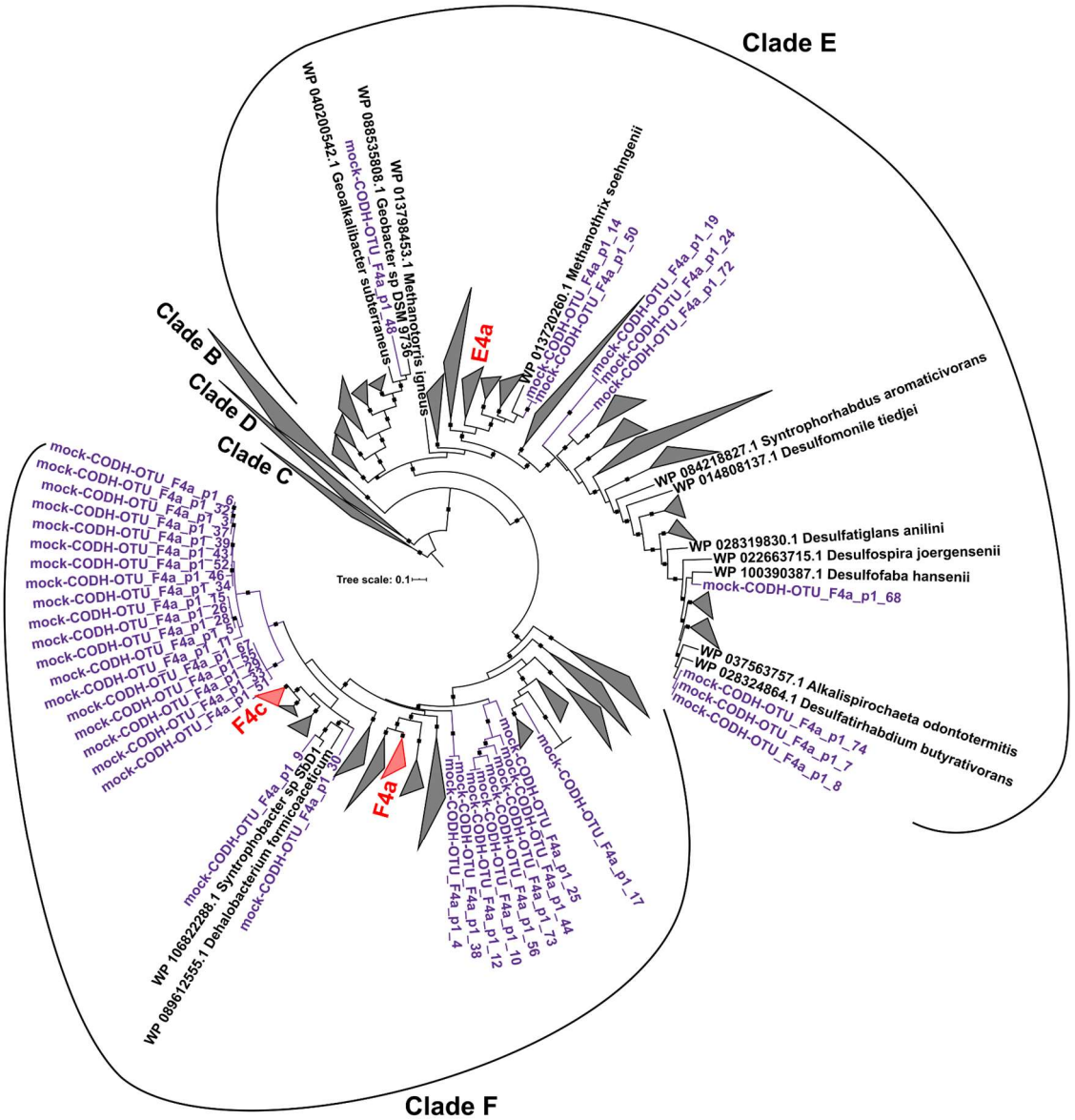


Fig. 4-6. Phylogeny of 'noisy CODH-OTUs' amplified by the F4a_p1 from the CODHech-mock community sample. The tree was constructed using an alignment of 1,558 CODH proteins and the CODH-OTUs. Major clades B through F are indicated. Branches and leaves are indicated by different colors as follows: red, FirmiCODHechs; black, other CODHs; purple, CODH-OTUs. The black squares on the branch indicate >0.8 support by bootstrap values.

CODHech genes amplified by new primers from environmental samples

To verify that the newly designed primers amplify CODHech genes from environments, I performed amplicon sequencing by using these primers with the sediment samples corrected from two hot springs of UN and JI, where thermophilic hydrogenogenic carboxydotrophs have been detected or isolated (Fukuyama *et al.*, 2017; Yoneda *et al.*, 2012, 2015) (Chapter 3). Of the all six primer sets tested, PCR amplification was observed for three (F4a_p1, F4c1_p1 and F4c1_p2) and two (F4a_p1 and F4c2_p1) in UN and JI, respectively. Sequencing of the PCR products produced >68,000 raw paired-end reads per primer per sample, leaving >39,000 chimera- and noise-free merged reads per primer per sample by the sequence processing (Table 4-6).

Table 4-6. Read statistics of amplicon sequencing in environmental samples

Samples	Primer sets	Total reads	Merged QC reads	Denoised chimera-free reads	CODH-OTUs		CODH-OTUs within FirmiCODHech subclades	
					# of reads	# of OTUs	# of reads	# of OTUs
UN	F4a_p1	81,415	68,345	39,136	37,080	53	0	0
UN	F4c1_p1	68,241	63,494	41,677	41,316	5	41,316	5
UN	F4c1_p2	73,395	67,320	40,855	36,766	3	36,766	3
JI	F4a_p1	112,727	89,115	46,348	41,502	26	0	0
JI	F4c2_p1	169,495	128,748	87,391	62,379	3	62,379	3

In the UN, the CODH reads amplified by the primer sets F4c1_p1 and F4c1_p2 were grouped into five and three CODH-OTUs within subclade F4c1, respectively (Tables 4-6, 4-7). Most of these CODH-OTUs, which accounted for ~99% of the read abundances, were phylogenetically related to the CODHech gene of *Desulfotomaculum nigrificans* CO-1-SRB by showing ≥98.2% pairwise identity (Table 4-7, Fig. 4-7). The CODH-OTUs, which were similar or identical to CODHech genes of *C. hydrogenoformans* Z-2901 and *Carboxydocella* sp. JDF658, respectively, were also

found, although they accounted for small part of the reads (Table 4-7, Fig. 4-7). In the JI, three CODH-OTUs were generated by F4c2_p1, all of which were similar to CODHech genes of *Thermincola* lineages (Table 4-7, Fig. 4-7). Meanwhile, 53 and 26 CODH-OTUs were also generated by the primer set F4a_p1 in the UN and the JI, respectively, but they were not classified into the F4a FirmiCODHech subclade (Table 4-6). As is the case in the CODHech-mock community sample, these CODH-OTUs were phylogenetically distinct CODHs forming clades within Clades E and F, which might be derived from unknown species (Fig. 4-8).

Table 4-7. Environmental CODH-OTUs which were classified into the FirmiCODHech subclades

CODH-OTUs*	Sub-clades	Closest CODHech genes			Relative abundance in each primer set (%)	
		Locus tag	Taxon	Identity (%)	UN	JI
F4c1_p1_1	F4c1	CHY_RS08505	*Carboxydothermus hydrogenoformans* Z-2901	**93.8**	0.017	0
F4c1_p1_3	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	98.2	85.599	0
F4c1_p1_4	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	98.7	13.975	0
F4c1_p1_6	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	99.1	0.327	0
F4c1_p1_7	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	99.3	0.082	0
F4c1_p2_2	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	99.1	85.566	0
F4c1_p2_6	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	99.8	13.624	0
F4c1_p2_19	F4c1	JDF658_RS09120	*Carboxydocella* sp. JDF658	100.0	0.811	0
F4c2_p1_1	F4c2	THERJR_RS14890	*Thermincola potens* JR	97.0	0	99.994
F4c2_p1_2	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	99.8	0	0.005
F4c2_p1_3	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	99.3	0	0.002

* CODH-OTU names were represented in the name of primer set which amplified it, and number identifier of the CODH-OTU, which were concatenated with "_" characters. CODH-OTUs which showed <94.5% identity with the closest CODHech genes were shown in bold.

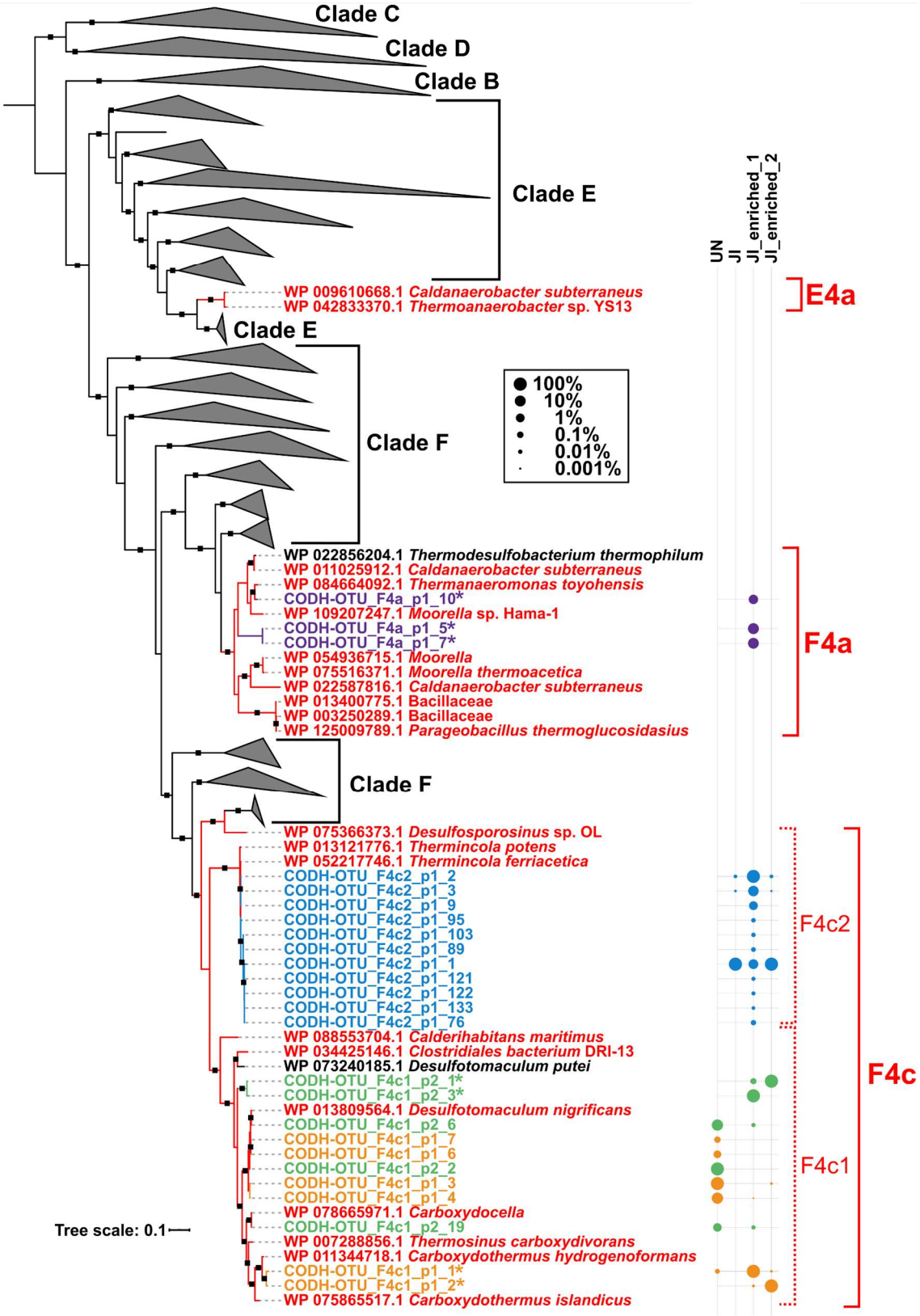


Fig. 4-7. Phylogeny and relative abundance of CODH-OTUs amplified by the new primer sets from the environmental and enriched samples. The tree was constructed using an alignment of 1,558 CODH proteins

and the CODH-OTUs. Major clades B through F are indicated in the right. FirmiCODHech subclades are assigned and colored in red. Branches and leaves are indicated by different colors as follows: red, FirmiCODHechs; black, other CODHs; other colors, CODH-OTUs amplified by each new primer set. The relative abundance of CODH-OTUs in each primer set are shown in bubble plot with different colors on the right of the tree. Asterisks indicate novel CODH-OTUs showing <94.5% pairwise sequence identity with the closest CODHech gene. The black squares on the branch indicate >0.8 support by bootstrap values.

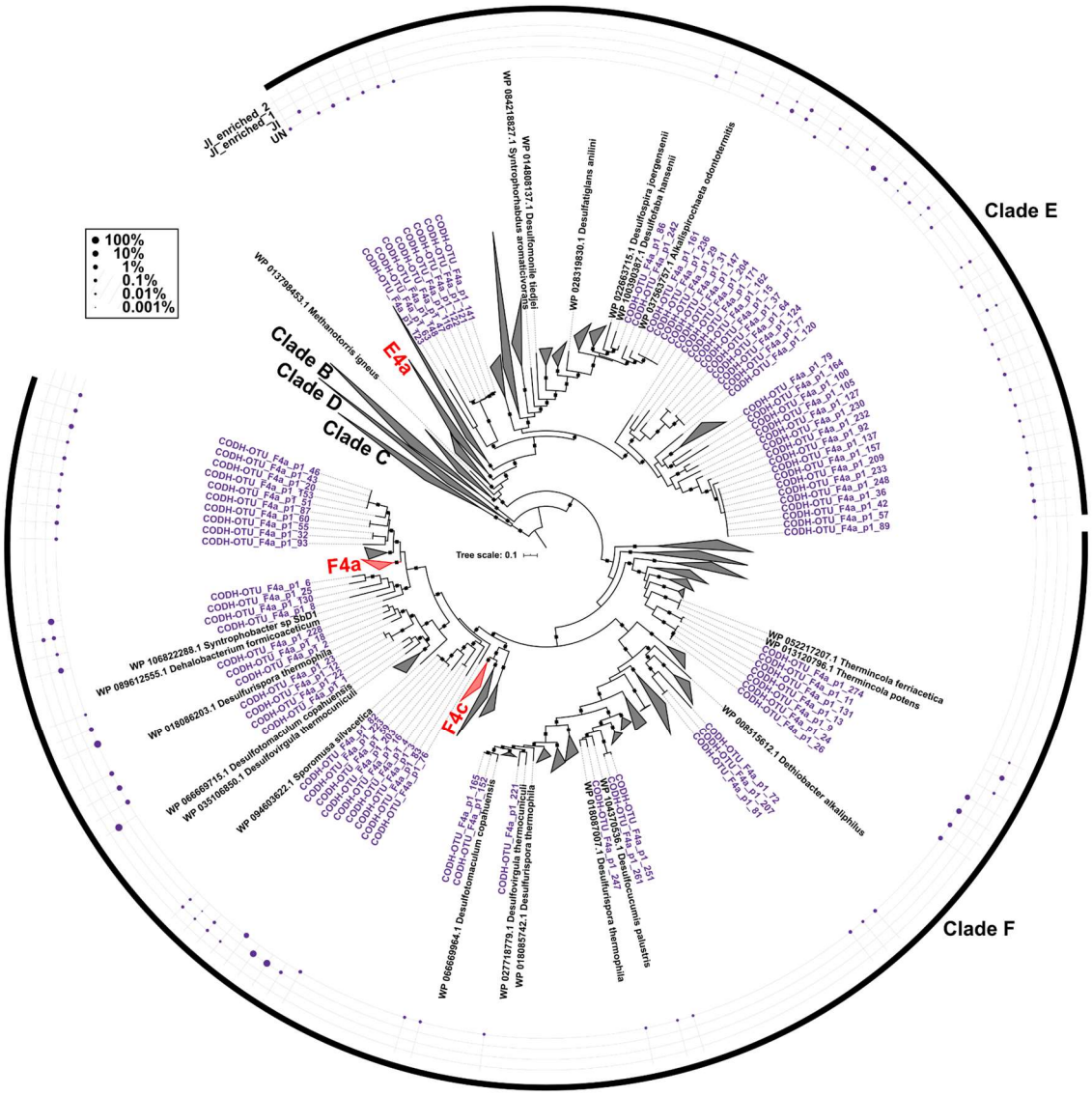


Fig. 4-8. Phylogeny of 'noisy CODH-OTUs' amplified by the F4a_p1 from the environmental and enrichment samples. The tree was constructed using an alignment of 1,558 CODH proteins and the CODH-OTUs. Major clades B–F are indicated. Branches and leaves are indicated by different colors as follows: red, FirmiCODHechs; black, other CODHs; purple, CODH-OTUs. The relative abundance of CODH-OTUs is shown in bubble plot. The black squares on the branch indicate >0.8 support by bootstrap values.

CODHech genes amplified by new primer sets from enrichment samples

To verify that the CODH-OTUs which were found by the new primers were CODHech genes derived from thermophilic hydrogenogenic carboxydotrophs, I tried to enrich these microorganisms in the sediment sample of JI by incubation under 10% CO. After five days, H_2 productions of 3.6 and 1.7% were observed in the head spaces of both two replicates (JI_enriched_1 and JI_enriched_2, respectively), suggesting that endogenous thermophilic hydrogenogenic carboxydotrophs had actively grown during enrichment. By PCR amplification using the new primers, products with expected sizes were obtained by four and three primer sets in JI_enriched_1 and 2, respectively. Sequencing of the PCR products produced >71,000 raw paired-end reads per primer per sample, and >28,000 chimera and noise-free merged reads per primer per sample were left (Table 4-8). While only three CODH-OTUs of F4c2 FirmiCODHech subclade were observed in the original JI sample (Table 4-6), I found 22 CODH-OTUs within F4a, F4c1 and F4c2 subclades in the enriched JI samples (Tables 4-8, 4-9).

Table 4-8. Read statistics of amplicon sequencing in enrichment samples

Samples	Primer sets	Total reads	Merged QC reads	Denoised chimera-free reads	CODH-OTUs		CODH-OTUs within FirmiCODHech subclades	
					# of reads	# of OTUs	# of reads	# of OTUs
JI_enriched_1	F4a_p1	128,964	108,172	67,696	66,676	21	13,462	3
JI_enriched_1	F4c1_p1	221,536	200,466	138,881	138,881	3	138,881	3
JI_enriched_1	F4c1_p2	92,499	54,465	28,819	11,550	4	11,550	4
JI_enriched_1	F4c2_p1	218,325	207,194	158,881	158,800	12	158,368	11
JI_enriched_2	F4c1_p1	111,569	100,723	63,905	63,905	3	63,905	3
JI_enriched_2	F4c1_p2	71,126	66,414	42,240	42,214	1	42,214	1
JI_enriched_2	F4c2_p1	123,192	116,928	96,397	96,190	4	96,104	3

In addition to the three CODH-OTUs of *Thermincola* lineages, which were observed in the original JI (Table 4-7), 12 other CODH-OTUs, which were

phylogenetically related to and showed ≥94.5% pairwise sequence identities with the FirmiCODHech genes of the known thermophilic hydrogenogenic carboxydotrophs including *D. nigrificans* CO-1-SRB and *Carboxydocella* sp. JDF658 (Byrne-Bailey *et al.*, 2010; Fukuyama *et al.*, 2017; Parshina *et al.*, 2005b; Sokolova *et al.*, 2002; Zavarzina *et al.*, 2007), were found in the enriched JI samples (Table 4-9, Fig. 4-7). The three *Thermincola* CODH-OTUs, which were observed in the original JI, remained abundant (~99% in relative abundance per primer set per sample) in the enriched JI samples, but the newly detected 12 CODH-OTUs accounted for small fraction of the reads obtained (0.001–1% in relative abundance per primer set) (Table 4-9, Fig. 4-7).

On the other hand, I found phylogenetically novel four CODH-OTUs (F4a_p1_5, F4a_p1_7, F4c1_p2_1 and F4c1_p2_3) in the enriched JI samples, which formed distinct branches from other known FirmiCODHech genes within F4a and F4c subclades (Fig. 4-7). These CODH-OTUs showed only ~89% pairwise nucleotide identity with the closest CODHech gene (Table 4-9). Furthermore, I found one (F4a_p1_10) and two (F4c1_p1_1 and F4c1_p1_2) CODH-OTUs which were phylogenetically related to FirmiCODHech genes of *Thermanaeromonas toyohensis* ToBE and *C. hydrogenoformans* Z-2901, respectively, with >0.8 support by bootstrap (Fig. 4-7). These CODH-OTUs, however, showed <94.5% pairwise sequence identities with the closest FirmiCODHech genes (Table 4-9), which is below the tentative threshold for classifying species in this study (Fig. 4-1) and implied that these might be derived from uncultured species. These novel CODH-OTUs were abundant in the enriched JI samples accounting for ~100% of the reads obtained by each primer set (Table 4-9). Note that the CODH-OTUs within the F4a subclade accounted for ~20% of the reads which were generated by the F4a_p1 from the JI_enriched_1, and the other reads were grouped into 18 noisy CODH-OTUs as in the

original JI sample (Table 4-8; Fig. 4-8).

Table 4-9. CODH-OTUs in enrichment samples, which were classified into the FirmiCODHech subclades

CODH-OTUs*	Sub-clades	Closest CODHech genes			Relative abundance in each primer set (%)	
		Locus tag	Taxon	Identity (%)	JI_enriched_1	JI_enriched_2
F4a_p1_5	F4a	Hama1_RS10675	*Moorella* sp. Hama 1	**85.7**	9.854	0
F4a_p1_7	F4a	Hama1_RS10675	*Moorella* sp. Hama 1	**85.7**	8.355	0
F4a_p1_10	F4a	B9A14_RS03610	*Thermanaeromonas toyohensis* ToBE	**92.9**	1.981	0
F4c1_p1_1	F4c1	CHY_RS08505	*Carboxydothermus hydrogenoformans* Z-2901	**93.8**	99.997	0.002
F4c1_p1_2	F4c1	CHY_RS08505	*Carboxydothermus hydrogenoformans* Z-2901	**93.4**	0.002	99.997
F4c1_p1_3	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	98.2	0	0.002
F4c1_p1_4	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	98.7	0.001	0
F4c1_p2_1	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	**89.2**	0.061	100
F4c1_p2_3	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	**88.5**	99.922	0
F4c1_p2_6	F4c1	DESCA_RS01885	*Desulfotomaculum nigrificans* CO-1-SRB	99.8	0.009	0
F4c1_p2_19	F4c1	JDF658_RS09120	*Carboxydocella* sp. JDF658	100.0	0.009	0
F4c2_p1_1	F4c2	THERJR_RS14890	*Thermincola potens* JR	97.0	2.903	99.904
F4c2_p1_2	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	99.8	89.416	0.005
F4c2_p1_3	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	99.3	6.300	0.001
F4c2_p1_9	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	99.1	1.025	0
F4c2_p1_76	F4c2	THERJR_RS14890	*Thermincola potens* JR	98.1	0.027	0
F4c2_p1_89	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	98.4	0.015	0
F4c2_p1_95	F4c2	Tfer_RS07525	*Thermincola ferriacetica* Z-0001	98.4	0.012	0
F4c2_p1_103	F4c2	THERJR_RS14890	*Thermincola potens* JR	98.4	0.011	0
F4c2_p1_121	F4c2	THERJR_RS14890	*Thermincola potens* JR	97.7	0.008	0
F4c2_p1_122	F4c2	THERJR_RS14890	*Thermincola potens* JR	98.1	0.008	0
F4c2_p1_133	F4c2	THERJR_RS14890	*Thermincola potens* JR	97.9	0.005	0

* CODH-OTU names were represented in the name of primer set which amplified it, and number identifier of the CODH-OTU, which were concatenated with "_" characters. CODH-OTUs which showed <94.5% identity with the closest CODHech genes were shown in bold.

Discussion

In this study, I provided new primers which can amplify CODHech genes of the phylum Firmicutes and enable to evaluate diversity of hydrogenogenic carboxydotrophs in environments by culture-independent way. Compared to the previous CODH-targeted primers which amplify ~1,300 bp region of broad sequence diversity of CODH genes (Matson *et al.*, 2011), the primers were designed to amplify FirmiCODHech genes and be applicable to deep-sequencing platform with 300 bp paired-end reads. Amplicon sequencing with these primers in combination with CO-enrichment successfully identified rare thermophilic hydrogenogenic carboxydotrophs. To our knowledge, this study identified the CODHech genes including unknown lineages in environments by CODH-targeted primers for the first time.

Phylogenetic analysis of CODH genes prior to design of primers revealed that the 34 FirmiCODHech genes were found in three different subclades within Clades E and F. The six new primer sets were designed to amplify each subclade, and each primer set was considered to specifically amplify targeting-CODHech genes with reflecting original abundance as shown in amplicon sequencing using the CODHech-mock community (Fig. 4-5). Four of the six new primer sets efficiently amplified the multiple CODH-OTUs within the FirmiCODHech subclades from the two hot spring sediment samples (Fig. 4-7), which have never been detected by previous CODH-targeted primers (Hoshino & Inagaki, 2017; Matson *et al.*, 2011). Since the primer sets F4c1_p1 and F4c1_p2, whose targets overlap each other, these primers might amplify separate regions of the same CODH genes (Fig. 4-3). If this redundancy in the CODH-OTUs of F4c1_p1 and F4c1_p2 is removed, it is considered that the new primers detected at least six lineages of CODH-OTUs within the FirmiCODHech subclades from the two hot spring sediment samples.

The increase in the CODH-OTUs related to the FirmiCODHech genes of known hydrogenogenic carboxydotrophs in the CO-enriched samples with hydrogen producing activity verified that the new primers amplify CODHech genes of endogenous thermophilic hydrogenogenic carboxydotrophs. Meanwhile, the new primers also detected the phylogenetically distinct lineages of CODH-OTUs within the subclades F4a and F4c which might be enriched on the course of hydrogen production under CO gas. These newly detected CODH-OTUs conserved the His and Cys residues in the C-cluster of CODH genes, which is a catalytic metal cluster comprising Ni, Fe and S (Dobbek *et al.*, 2001; Inoue *et al.*, 2019a), suggesting that these conserved the function of CODHs. These results correctively indicated that these CODH-OTUs detected might be derived from unknown thermophilic hydrogenogenic carboxydotrophs. The occurrence of the CODH-OTUs only in enriched samples also suggested that natural abundance of these thermophilic hydrogenogenic carboxydotrophs in the JI sample before enrichment was smaller than detection limit. Taking into account the wide distribution of these microorganisms as shown in Chapter 3, it is predicted that unexplored rare thermophilic hydrogenogenic carboxydotrophs exist in wide variety of environments.

The new primers I provided here would pave the way for exploration of thermophilic hydrogenogenic carboxydotrophs, which have mainly relied on isolation and cultivation. CODH amplicon sequencing in combination with CO-enrichment is one effective option with the least cultivation bias for evaluation of rare hydrogenogenic carboxydotrophs including unknown species. However, we should take care when we evaluate phylogenetically novel CODH-OTUs, because CODH genes are often horizontaly transferred (Techtmann *et al.*, 2012). Also, it should be noted that the primer set F4a_p1 yielded noisy CODH-OTUs, although it can work as PCR amplification

primer for F4a FirmiCODHech genes. This primer should be revised by redesigning more specific one using larger CODH database including the phylogenetically novel CODH genes. Albeit these properties, the new primers are still worth using for evaluation of the diversity of rare hydrogenogenic carboxydotrophs with providing high-throughput deep-sequencing screening. Further studies using these primers in vast variety of environments including lake or marine sediments, soils and composts will reveal the diversity and distribution of thermophilic hydrogenogenic carboxydotrophs including unknown species.

Chapter 5
Integration and outlook

The microbial hydrogenogenic carbon monoxide (CO) oxidizing activity achieved by hydrogenogenic carboxydotrophs with a membrane associated CO-oxidizing and H_2-producing machinery, the carbon monoxide dehydrogenase (CODH)/energy converting hydrogenase (ECH) complex, is considered as an important metabolic process in microbial community as well as biotechnological application. However, previous isolation-based studies biased in hydrothermal environments, which revealed ~20 isolates, have resulted in the limited information about CO metabolism and ecology of thermophilic hydrogenogenic carboxydotrophs.

First, to reveal novel CO metabolism of thermophilic hydrogenogenic carboxydotrophs, I performed *de novo* sequencing and feature analysis of the *Calderihabitans maritimus* KKC1 genome, which is the first isolate from marine sediment and the phylogenetically novel bacterium (Chapter 2). This analysis revealed that *C. maritimus* KKC1 harbored six CODH gene clusters, which is the highest number encoded in a single genome, including three known CODH gene clusters, CODH–acetyl-CoA synthase (Wood-Ljungdahl pathway), CODH–ECH (energy conservation) and CooF–CODH–flavin adenine dinucleotide-nicotinamide adenine dinucleotide oxidoreductase (reducing power production), while the other three had novel genomic contexts. One of the three novel CODH genes was associated with 2-oxoglutarate:ferredoxin oxidoreductase, which is a CO_2-fixing enzyme in reverse tricarboxylic acid (RTCA) cycle, and expected to comprise novel CO metabolism where

CO is incorporated via RTCA cycle. From these results, it is predicted that unexplored hydrogenogenic carboxydotrophs might possess the novel CO metabolism and the exploration of these microbes in environments is important for understanding the diversity of CO metabolisms.

To explore more thermophilic hydrogenogenic carboxydotrophs in wide variety of environments, culture-independent method, which is high-throughput and avoids cultivation bias, is desired. But design of CODH-targeted universal primers was difficult due to the sequence diversity. Therefore, in Chapter 3, I constructed a reference database of these microorganisms with revealing distribution of CODH–ECH gene clusters upon 16S rRNA phylogeny to enable exploration of thermophilic hydrogenogenic carboxydotrophic isolates by microbial community analysis. This analysis identified 71 genomes of bacteria and archaea including the 46 overlooked potential hydrogenogenic carboxydotrophs, whose hydrogenogenic CO oxidizing activities have never been reported, from ~140,000 prokaryotic genomes, and expanded the estimation for diversity of these microorganisms to four phyla, 26 genera, and 43 species. By microbial community analysis using this reference data, I showed that potential thermophilic hydrogenogenic carboxydotrophs in phylum Firmicutes widely distributed in hydrothermal environments with small relative abundance.

While microbial community analysis can easily evaluate the 71 hydrogenogenic carboxydotrophic isolates, we cannot define hydrogenogenic carboxydotroph by the 16S rRNA gene sequence. In Chapter 4, I designed new primers for PCR amplification of CODH genes of CODH–ECH gene clusters (CODHech genes) to directly evaluate the diversity of thermophilic hydrogenogenic carboxydotrophs including unknown species by culture-independent way. My primer design strategy dividing target CODHech genes

of Firmicutes members based on phylogenetic subclades provided the six new primer sets, four of which effectively amplified the CODHech genes. Amplicon sequencing with these primers in two hot spring sediments in combination with CO-enrichment successfully identified at least six lineages of CODHech genes which might be derived from rare thermophilic hydrogenogenic carboxydotrophs possibly including novel species. To our knowledge, this is the first time that the CODHech genes are identified by CODH-targeted primers in environments.

These studies expanded our knowledge on metabolic and phylogenetic diversity and distribution of thermophilic hydrogenogenic carboxydotrophs. When correctively considering these results, it is predicted that thermophilic hydrogenogenic carboxydotrophs including still unknown species with novel CO metabolism are widely distributed in hydrothermal environments. In the future work, application of the culture-independent techniques, which I developed and validated in hydrothermal environments, to vast variety of environments including lake or marine sediments, soils and composts will pave the way for revealing diversity and distribution of these microorganisms.

Acknowledgements

First of all, I would like to express my sincere gratitude to Professor emeritus Yoshihiko Sako and Professor Takashi Yoshida who directed, discussed and communicated with me and gave me critical advice. I am grateful to Professor Shigeki Sawayama and Professor Tatsuya Sugawara for thoughtful words and reviewing on my thesis.

I would like to express my deep appreciate to Professor Hiroyuki Ogata, Professor Susumu Goto, and Dr. Yosuke Nishimura who gave me helpful advices for bioinformatics. I am also grateful to Professor Hisato Yasuda for all his help in correcting sediment samples. I would like to express my great thank to Dr. Yasuko Yoneda for isolation of novel thermophilic hydrogenogenic carboxydotrophs and giving me advices. I would like to appreciate Mr. Koichiro Nakano for all his help in genomic sequencing.

I thank my colleagues in the Laboratory of Marine Microbiology for support and encouragements. Especially, I would like to offer my special thanks to Dr. Yuto Fukuyama, Dr. Masao Inoue, Mr. Kenta Mise, Mr. Tatsuki Oguro and Mr. Issei Nakamoto as co-workers for valuable discussion and advices. I would like to appreciate to Dr. Ayumi Tanimura, Mr. Eitaro Ikeda, Mr. Yusuke Ogami, Mr. Shin Fujiwara and Mr. Taiki Hino who helped me in sample correction and discussed with me. I would like to express my thank to Dr. Takashi Daifuku, Mr. Kyosuke Takao, Ms. Maho Masuda, Ms. Mika Matsumoto, Mr. Hiroyasu Watai, Dr. Daichi Morimoto, Mr. Florian Prodinger, Dr. Shigeko Kimura, Dr. Sigitas Sulcius, Mr. Kento Tominaga and Mr. Tatsuhiro Isozaki for spending a lot of time with me and giving me a lot of encouragement. I also express my special thank to Ms. Ritsuko Mizuno for all her help and great encouragement.

My doctoral works were supported by Grants-in-Aid for Scientific Research (A) 25252038, and (S) 16H06381 from The Ministry of Education, Culture, Sports, Science and Technology (MEXT) and by a Grant-in-Aid from the Japan Society for the Promotion of Science (JSPS) Fellows (16J11269).

Finally, I would like to thank my family for giving me warm support throughout my student life.

References

Adam, P. S., Borrel, G. & Gribaldo, S. (2018). Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci U S A* **115**, E1166–E1173.

Adam, P., Hecht, S., Eisenreich, W., Kaiser, J., Gräwert, T., Arigoni, D., Bacher, A. & Rohdich, F. (2002). Biosynthesis of terpenes: Studies on 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase. *Proc Natl Acad Sci U S A* **99**, 12108–12113.

Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L. & Olmedo, G. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* **11**, 332.

Alex, A. & Antunes, A. (2015). Whole genome sequencing of the symbiont *Pseudovibrio* sp. from the intertidal marine sponge *Polymastia penicillus* revealed a gene repertoire for host-switching permissive lifestyle. *Genome Biol Evol* **7**, 3022–3032.

Alves, J. I., van Gelder, A. H., Alves, M. M., Sousa, D. Z. & Plugge, C. M. (2013). *Moorella stamsii* sp. nov., a new anaerobic thermophilic hydrogenogenic carboxydotroph isolated from digester sludge. *Int J Syst Evol Microbiol* **63**, 4072–4076.

Aragão, D., Mitchell, E. P., Frazão, C. F., Carrondo, M. A. & Lindley, P. F. (2008). Structural and functional relationships in the hybrid cluster protein family: structure of the anaerobically purified hybrid cluster protein from *Desulfovibrio vulgaris* at 1.35Å resolution. *Acta Crystallogr Sect D* **64**, 665–674.

Bae, S. S., Kim, Y. J., Yang, S. H., Lim, J. K., Jeon, J. H., Lee, H. S., Kang, S. G., Kim, S.-J. & Lee, J.-H. (2006). *Thermococcus onnurineus* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent area at the PACMANUS field. *J Microbiol Biotechnol* **16**, 1826–1831.

Balk, M., van Gelder, T., Weelink, S. A. & Stams, A. J. M. (2008). (Per)chlorate Reduction by the Thermophilic Bacterium *Moorella perchloratireducens* sp. nov., Isolated from Underground Gas Storage. *Appl Environ Microbiol* **74**, 403–409.

Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* **93**, 9188–9193.

Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci U S A* **91**, 1609–1613.

Berendsen, E. M., Wells-Bennik, M. H. J., Krawczyk, A. O., de Jong, A., van Heel, A., Holsappel, S., Eijlander, R. T. & Kuipers, O. P. (2016). Draft Genome Sequences of Seven Thermophilic Spore-Forming Bacteria Isolated from Foods That Produce Highly Heat-Resistant Spores, Comprising *Geobacillus* spp., *Caldibacillus debilis*, and *Anoxybacillus flavithermus*. *Microbiol Resour Announc* **4**, e00105-16.

Besemer, J., Lomsadze, A. & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* **29**, 2607–2618.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

Bonam, D. & Ludden, P. W. (1987). Purification and characterization of carbon monoxide dehydrogenase, a nickel, zinc, iron-sulfur protein, from *Rhodospirillum rubrum*. *J Biol Chem* **262**, 2980–2987.

Bott, T. L. & Brock, T. D. (1969). Bacterial growth rates above 90°C in Yellowstone hot springs. *Science* **164**, 1411–1412.

Brady, A. L., Sharp, C. E., Grasby, S. E. & Dunfield, P. F. (2015). Anaerobic carboxydotrophic bacteria in geothermal springs identified using stable isotope probing. *Front Microbiol* **6**, 897.

Brock, T. D. (1967). Life at high temperatures. *Science* **158**, 1012–1019.

Brock, T. D. & Darland, G. K. (1970). Limits of microbial existence: temperature and pH. *Science* **169**, 1316–1318.

Brumm, P. J., Land, M. L. & Mead, D. A. (2015). Complete genome sequence of *Geobacillus thermoglucosidasius* C56-YS93, a novel biomass degrader isolated from obsidian hot spring in Yellowstone National Park. *Stand Genomic Sci* **10**, 73.

Buchfink, B., Xie, C. & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60.

Byrne-Bailey, K. G., Wrighton, K. C., Melnyk, R. A., Agbo, P., Hazen, T. C. & Coates, J. D. (2010). Complete genome sequence of the electricity-producing '*Thermincola potens*' strain JR. *J Bacteriol* **192**, 4078–4079.

Canganella, F., Jones, W. J., Gambacorta, A. & Antranikian, G. (1998). *Thermococcus guaymasensis* sp. nov. and *Thermococcus aggregans* sp. nov., two novel thermophilic archaea isolated from the Guaymas Basin hydrothermal vent site. *Int J Syst Evol Microbiol* **48**, 1181–1185.

Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. (2009). trimAl: a tool

for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973.

Carvalho, S. M., Marques, J., Romão, C. C. & Saraiva, L. M. (2019). Metabolomics of *Escherichia coli* Treated with the Antimicrobial Carbon Monoxide-Releasing Molecule CORM-3 Reveals Tricarboxylic Acid Cycle as Major Target. *Antimicrob Agents Chemother* **63**, e00643-19.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, btz848.

Chen, I.-M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R. & other authors. **(2019).** IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* **47**, D666–D677.

Conrad, R. (1996). Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O, and NO). *Microbiol Mol Biol Rev* **60**, 609–640.

Conte, L., Szopa, S., Séférian, R. & Bopp, L. (2019). The oceanic cycle of carbon monoxide and its emissions to the atmosphere. *Biogeosciences* **16**, 881–902.

Dashekvicz, M. P. & Uffen, R. L. (1979). Identification of a Carbon Monoxide-Metabolizing Bacterium as a Strain of *Rhodopseudomonas gelatinosa* (Molisch) van Niel. *Int J Syst Bacteriol* **29**, 145–148.

Davidge, K. S., Sanguinetti, G., Yee, C. H., Cox, A. G., McLeod, C. W., Monk, C. E., Mann, B. E., Motterlini, R. & Poole, R. K. (2009). Carbon monoxide-releasing antibacterial molecules target respiration and global transcriptional regulators. *J Biol Chem* **284**, 4516–4524.

Davidova, M. N., Tarasova, N. B., Mukhitova, F. K. & Karpilova, I. U. (1994). Carbon monoxide in metabolism of anaerobic bacteria. *Can J Microbiol* **40**, 417–425.

Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679.

Diender, M., Stams, A. J. M. & Sousa, D. Z. (2015). Pathways and Bioenergetics of Anaerobic Carbon Monoxide Fermentation. *Front Microbiol* **6**, 1275.

Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R. & Meyer, O. (2001). Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* **293**, 1281–1285.

Drake, H. L. & Daniel, S. L. (2004). Physiology of the thermophilic acetogen *Moorella thermoacetica*. *Res Microbiol* **155**, 869–883.

Drake, H. L., Küsel, K. & Matthies, C. (2002). Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie Van Leeuwenhoek* **81**, 203–213.

Dufraigne, C., Fertil, B., Lespinats, S., Giron, A. & Deschavanne, P. (2005). Detection and characterization of horizontal transfers in prokaryotes using genomic signature. *Nucleic Acids Res* **33**, e6.

Dürre, P. & Eikmanns, B. J. (2015). C1-carbon sources for chemical and fuel production by microbial gas fermentation. *Curr Opin Biotechnol* **35**, 63–72.

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

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS

amplicon sequencing. *bioRxiv* 081257.

Ensign, S. A. & Ludden, P. W. (1991). Characterization of the CO oxidation/H_2 evolution system of *Rhodospirillum rubrum*. Role of a 22-kDa iron-sulfur protein in mediating electron transfer between carbon monoxide dehydrogenase and hydrogenase. *J Biol Chem* **266**, 18395–18403.

Esquivel-Elizondo, S., Maldonado, J. & Krajmalnik-Brown, R. (2018). Anaerobic carbon monoxide metabolism by *Pleomorphomonas carboxyditropha* sp. nov., a new mesophilic hydrogenogenic carboxydotroph. *FEMS Microbiol Ecol* **94**, fiy056.

Fardeau, M.-L., Salinas, M. B., L'Haridon, S., Jeanthon, C., Verhé, F., Cayol, J.-L., Patel, B. K. C., Garcia, J.-L. & Ollivier, B. (2004). Isolation from oil reservoirs of novel thermophilic anaerobes phylogenetically related to *Thermoanaerobacter subterraneus*: reassignment of *T. subterraneus*, *Thermoanaerobacter yonseiensis*, *Thermoanaerobacter tengcongensis* and *Carboxydibrachium pacificum* to *Caldanaerobacter subterraneus* gen. nov., sp. nov., comb. nov. as four novel subspecies. *Int J Syst Evol Microbiol* **54**, 467–474.

Ferry, J. G. (1999). Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol Rev* **23**, 13–38.

Fontaine, F. E., Peterson, W. H., McCoy, E., Johnson, M. J. & Ritter, G. J. (1942). A New Type of Glucose Fermentation by *Clostridium thermoaceticum* n. sp. *J Bacteriol* **43**, 701–715.

Fox, J. D., He, Y., Shelver, D., Roberts, G. P. & Ludden, P. W. (1996a). Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. *J Bacteriol* **178**, 6200–6208.

Fox, J. D., Kerby, R. L., Roberts, G. P. & Ludden, P. W. (1996b). Characterization of the CO-induced, CO-tolerant hydrogenase from *Rhodospirillum rubrum* and the gene encoding the large subunit of the enzyme. *J Bacteriol* **178**, 1515–1524.

Fraenkel, D. G. & Vinopal, R. T. (1973). Carbohydrate metabolism in bacteria. *Annu Rev Microbiol* **27**, 69–100.

Fuchs, G. (2011). Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? *Annu Rev Microbiol* **65**, 631–658.

Fukuda, E. & Wakagi, T. (2002). Substrate recognition by 2-oxoacid:ferredoxin oxidoreductase from *Sulfolobus* sp. strsain 7. *Biochim Biophys Acta - Protein Struct Mol Enzymol* **1597**, 74–80.

Fukuyama, Y., Oguro, T., Omae, K., Yoneda, Y., Yoshida, T. & Sako, Y. (2017). Draft genome sequences of two hydrogenogenic carboxydotrophic bacteria, *Carboxydocella* sp. strains JDF658 and ULO1, isolated from two distinct volcanic fronts in Japan. *Genome Announc* **5**, e00242-17.

Fukuyama, Y., Omae, K., Yoneda, Y., Yoshida, T. & Sako, Y. (2018). Insight into energy conservation via alternative carbon monoxide metabolism in *Carboxydothermus pertinax* revealed by comparative genome analysis. *Appl Environ Microbiol* **84**, e00458-18.

Fukuyama, Y., Tanimura, A., Inoue, M., Omae, K., Yoshida, T. & Sako, Y. (2019). Draft Genome Sequences of Two Thermophilic *Moorella* sp. Strains, Isolated from an Acidic Hot Spring in Japan. *Microbiol Resour Announc* **8**, e00663-19.

Furdui, C. & Ragsdale, S. W. (2000). The Role of Pyruvate Ferredoxin Oxidoreductase in Pyruvate Synthesis during Autotrophic Growth by the Wood-Ljungdahl Pathway. *J Biol Chem* **275**, 28494–28499.

Geelhoed, J. S., Henstra, A. M. & Stams, A. J. M. (2016). Carboxydotrophic growth of *Geobacter sulfurreducens*. *Appl Microbiol Biotechnol* **100**, 997–1007.

Gencic, S., Duin, E. C. & Grahame, D. A. (2010). Tight Coupling of Partial Reactions in the Acetyl-CoA Decarbonylase/Synthase (ACDS) Multienzyme Complex from *Methanosarcina thermophila*: acetyl C-C bond fragmentation at the A cluster promoted by protein conformational changes. *J Biol Chem* **285**, 15450–15463.

Grépinet, O., Boumart, Z., Virlogeux-Payant, I., Loux, V., Chiapello, H., Gendrault, A., Gibrat, J.-F., Chemaly, M. & Velge, P. (2012). Genome Sequence of the Persistent *Salmonella enterica* subsp. enterica Serotype Senftenberg Strain SS209. *J Bacteriol* **194**, 2385–2386.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321.

Haouari, O., Fardeau, M.-L., Casalot, L., Tholozan, J.-L., Hamdi, M. & Ollivier, B. (2006). Isolation of sulfate-reducing bacteria from Tunisian marine sediments and description of *Desulfovibrio bizertensis* sp. nov. *Int J Syst Evol Microbiol* **56**, 2909–2913.

Harada, J., Yamada, T., Giri, S., Hamada, M., Nobu, M. K., Narihiro, T., Tsuji, H. & Daimon, H. (2018). Draft Genome Sequence of *Moorella* sp. Strain Hama-1, a Novel Acetogenic Bacterium Isolated from a Thermophilic Digestion Reactor. *Microbiol Resour Announc* **6**, e00517-18.

He, Y., Shelver, D., Kerby, R. L. & Roberts, G. P. (1996). Characterization of a CO-responsive Transcriptional Activator from *Rhodospirillum rubrum*. *J Biol Chem* **271**, 120–123.

Hensley, S. A., Jung, J.-H., Park, C.-S. & Holden, J. F. (2014). *Thermococcus paralvinellae* sp. nov. and *Thermococcus cleftensis* sp. nov. of hyperthermophilic heterotrophs from deep-sea hydrothermal vents. *Int J Syst Evol Microbiol* **64**, 3655–3659.

Hille, R., Dingwall, S. & Wilcoxen, J. (2015). The aerobic CO dehydrogenase from *Oligotropha carboxidovorans*. *J Biol Inorg Chem* **20**, 243–251.

Hoshino, T. & Inagaki, F. (2017). Distribution of anaerobic carbon monoxide dehydrogenase genes in deep subseafloor sediments. *Lett Appl Microbiol* **64**, 355–363.

Huang, H., Wang, S., Moll, J. & Thauer, R. K. (2012). Electron Bifurcation Involved in the Energy Metabolism of the Acetogenic Bacterium *Moorella thermoacetica* Growing on Glucose or H_2 plus CO_2. *J Bacteriol* **194**, 3689–3699.

Hubert, C., Loy, A., Nickel, M., Arnosti, C., Baranyi, C., Brüchert, V., Ferdelman, T., Finster, K., Christensen, F. M. & other authors. (2009). A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. *Science* **325**, 1541–1544.

Inagaki, S., Masuda, C., Akaishi, T., Nakajima, H., Yoshioka, S., Ohta, T., Pal, B., Kitagawa, T. & Aono, S. (2005). Spectroscopic and Redox Properties of a CooA Homologue from *Carboxydothermus hydrogenoformans*. *J Biol Chem* **280**, 3269–3274.

Inoue, M., Nakamoto, I., Omae, K., Oguro, T., Ogata, H., Yoshida, T. & Sako, Y. (2019a). Structural and Phylogenetic Diversity of Anaerobic Carbon-Monoxide Dehydrogenases. *Front Microbiol* **9**, 3353.

Inoue, M., Tanimura, A., Ogami, Y., Hino, T., Okunishi, S., Maeda, H., Yoshida, T. & Sako, Y. (2019b). Draft Genome Sequence of *Parageobacillus*

thermoglucosidasius Strain TG4, a Hydrogenogenic Carboxydotrophic Bacterium Isolated from a Marine Sediment. *Microbiol Resour Announc* **8**, e01666-18.

Inoue, T., Takao, K., Yoshida, T., Wada, K., Daifuku, T., Yoneda, Y., Fukuyama, K. & Sako, Y. (2013). Cysteine 295 indirectly affects Ni coordination of carbon monoxide dehydrogenase-II C-cluster. *Biochem Biophys Res Commun* **441**, 13–17.

Itoh, T., Suzuki, K. & Nakase, T. (2002). *Vulcanisaeta distributa* gen. nov., sp. nov., and *Vulcanisaeta souniana* sp. nov., novel hyperthermophilic, rod-shaped crenarchaeotes isolated from hot springs in Japan. *Int J Syst Evol Microbiol* **52**, 1097–1104.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114.

Ji, S., Zhao, R., Li, Z., Li, B., Shi, X. & Zhang, X.-H. (2013). *Ferrimonas sediminum* sp. nov., isolated from coastal sediment of an amphioxus breeding zone. *Int J Syst Evol Microbiol* **63**, 977–981.

Jiang, B., Henstra, A.-M., Paulo, P. L., Balk, M., van Doesburg, W. & Stams, A. J. M. (2009). Atypical one-carbon metabolism of an acetogenic and hydrogenogenic *Moorella thermoacetica* strain. *Arch Microbiol* **191**, 123–131.

Jung, G. Y., Kim, J. R., Jung, H. O., Park, J.-Y. & Park, S. (1999). A new chemoheterotrophic bacterium catalyzing water-gas shift reaction. *Biotechnol Lett* **21**, 869–873.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res* **47**, D590–D595.

Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.

Kerby, R. L., Hong, S. S., Ensign, S. A., Coppoc, L. J., Ludden, P. W. & Roberts, G. P. (1992). Genetic and physiological characterization of the *Rhodospirillum rubrum* carbon monoxide dehydrogenase system. *J Bacteriol* **174**, 5284–5294.

Kerby, R. L., Ludden, P. W. & Roberts, G. P. (1995). Carbon monoxide-dependent growth of *Rhodospirillum rubrum*. *J Bacteriol* **177**, 2241–2244.

Khalil, M. A. K. & Rasmussen, R. A. (1990). The global cycle of carbon monoxide: Trends and mass balance. *Chemosphere* **20**, 227–242.

Khelaifia, S., Fardeau, M.-L., Pradel, N., Aussignargues, C., Garel, M., Tamburini, C., Cayol, J.-L., Gaudron, S., Gaill, F. & Ollivier, B. (2011). *Desulfovibrio piezophilus* sp. nov., a piezophilic, sulfate-reducing bacterium isolated from wood falls in the Mediterranean Sea. *Int J Syst Evol Microbiol* **61**, 2706–2711.

Kim, B. C., Grote, R., Lee, D. W., Antranikian, G. & Pyun, Y. R. (2001). *Thermoanaerobacter yonseiensis* sp. nov., a novel extremely thermophilic, xylose-utilizing bacterium that grows at up to 85 degrees C. *Int J Syst Evol Microbiol* **51**, 1539–1548.

Kim, M.-S., Bae, S. S., Kim, Y. J., Kim, T. W., Lim, J. K., Lee, S. H., Choi, A. R., Jeon, J. H., Lee, J.-H. & other authors. (2013). CO-Dependent H_2 Production by Genetically Engineered *Thermococcus onnurineus* NA1. *Appl Environ Microbiol* **79**, 2048–2053.

Kim, M., Oh, H.-S., Park, S.-C. & Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* **64**, 346–351.

King, G. M. & Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat Rev Microbiol* **5**, 107–118.

Kochetkova, T. V, Rusanov, I. I., Pimenov, N. V, Kolganova, T. V, Lebedinsky, A. V, Bonch-Osmolovskaya, E. A. & Sokolova, T. G. (2011). Anaerobic transformation of carbon monoxide by microbial communities of Kamchatka hot springs. *Extremophiles* **15**, 319–325.

Kozhevnikova, D. A., Taranov, E. A., Lebedinsky, A. V, Bonch-Osmolovskaya, E. A. & Sokolova, T. G. (2016). Hydrogenogenic and sulfidogenic growth of *Thermococcus* archaea on carbon monoxide and formate. *Microbiology* **85**, 400–410.

Ladapo, J. & Whitman, W. B. (1990). Method for isolation of auxotrophs in the methanogenic archaebacteria: role of the acetyl-CoA pathway of autotrophic CO_2 fixation in *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A* **87**, 5598–5602.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T. & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100–3108.

Lee, H. S., Kang, S. G., Bae, S. S., Lim, J. K., Cho, Y., Kim, Y. J., Jeon, J. H., Cha, S.-S., Kwon, K. K. & other authors. **(2008).** The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydotrophic metabolism. *J Bacteriol* **190**, 7491–7499.

Letunic, I. & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**, W242–W245.

Lill, R. & Mühlenhoff, U. (2005). Iron–sulfur-protein biogenesis in eukaryotes. *Trends*

Biochem Sci **30**, 133–141.

Lim, J. K., Kang, S. G., Lebedinsky, A. V, Lee, J.-H. & Lee, H. S. (2010). Identification of a novel class of membrane-bound [NiFe]-hydrogenases in *Thermococcus onnurineus* NA1 by *in silico* analysis. *Appl Environ Microbiol* **76**, 6286–6289.

Lynch, M. D. J. & Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13**, 217–229.

Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Thiessen, P. A., Geer, L. Y. & Bryant, S. H. (2002). CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* **30**, 281–283.

Marteinsson, V. T., Birrien, J.-L., Reysenbach, A.-L., Vernet, M., Marie, D., Gambacorta, A., Messner, P., Sleytr, U. B. & Prieur, D. (1999). *Thermococcus barophilus* sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **49**, 351–359.

Matson, E. G., Gora, K. G. & Leadbetter, J. R. (2011). Anaerobic Carbon Monoxide Dehydrogenase Diversity in the Homoacetogenic Hindgut Microbial Communities of Lower Termites and the Wood Roach. *PLoS One* **6**, e19316.

McMurdie, P. J. & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217.

Meyer, F. M. & Stülke, J. (2013). Malate metabolism in *Bacillus subtilis*: distinct roles for three classes of malate-oxidizing enzymes. *FEMS Microbiol Lett* **339**, 17–22.

Mino, K. & Ishikawa, K. (2003). Characterization of a Novel Thermostable O-Acetylserine Sulfhydrylase from *Aeropyrum pernix* K1. *J Bacteriol* **185**, 2277–

2284.

Mohr, T., Aliyu, H., Küchlin, R., Polliack, S., Zwick, M., Neumann, A., Cowan, D. & de Maayer, P. (2018). CO-dependent hydrogen production by the facultative anaerobe *Parageobacillus thermoglucosidasius*. *Microb Cell Fact* **17**, 108.

Moreno-Hagelsieb, G. & Janga, S. C. (2008). Operons and the effect of genome redundancy in deciphering functional relationships using phylogenetic profiles. *Proteins* **70**, 344–352.

Mori, K., Hanada, S., Maruyama, A. & Marumo, K. (2002). *Thermanaeromonas toyohensis* gen. nov., sp. nov., a novel thermophilic anaerobe isolated from a subterranean vein in the Toyoha Mines. *Int J Syst Evol Microbiol* **52**, 1675–1680.

Mörsdorf, G., Frunzke, K., Gadkari, D. & Meyer, O. (1992). Microbial growth on carbon monoxide. *Biodegradation* **3**, 61–82.

Müller, A. L., de Rezende, J. R., Hubert, C. R. J., Kjeldsen, K. U., Lagkouvardos, I., Berry, D., Jørgensen, B. B. & Loy, A. (2014). Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J* **8**, 1153–1165.

Munk, A. C., Copeland, A., Lucas, S., Lapidus, A., Del Rio, T. G., Barry, K., Detter, J. C., Hammon, N., Israni, S. & other authors. (2011). Complete genome sequence of *Rhodospirillum rubrum* type strain ($S1^T$). *Stand Genomic Sci* **4**, 293–302.

Nakagawa, T., Iino, T., Suzuki, K. & Harayama, S. (2006). *Ferrimonas futtsuensis* sp. nov. and *Ferrimonas kyonanensis* sp. nov., selenate-reducing bacteria belonging to the *Gammaproteobacteria* isolated from Tokyo Bay. *Int J Syst Evol Microbiol* **56**, 2639–2645.

Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. (2004). Biased biological

functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**, 760–766.

Nevin, K. P., Holmes, D. E., Woodard, T. L., Hinlein, E. S., Ostendorf, D. W. & Lovley, D. R. (2005). *Geobacter bemidjiensis* sp. nov. and *Geobacter psychrophilus* sp. nov., two novel Fe(III)-reducing subsurface isolates. *Int J Syst Evol Microbiol* **55**, 1667–1674.

Nobre, L. S., Seixas, J. D., Romão, C. C. & Saraiva, L. M. (2007). Antimicrobial Action of Carbon Monoxide-Releasing Compounds. *Antimicrob Agents Chemother* **51**, 4303–4307.

Novikov, A. A., Sokolova, T. G., Lebedinsky, A. V, Kolganova, T. V & Bonch-Osmolovskaya, E. A. (2011). *Carboxydothermus islandicus* sp. nov., a thermophilic, hydrogenogenic, carboxydotrophic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol* **61**, 2532–2537.

Oda, Y., Larimer, F. W., Chain, P. S. G., Malfatti, S., Shin, M. V., Vergez, L. M., Hauser, L., Land, M. L., Braatsch, S. & other authors. **(2008).** Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. *Proc Natl Acad Sci U S A* **105**, 18543–18548.

Oelgeschläger, E. & Rother, M. (2008). Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch Microbiol* **190**, 257–269.

Oh, Y.-K., Kim, H.-J., Park, S., Kim, M.-S. & Ryu, D. D. Y. (2008). Metabolic-flux analysis of hydrogen production pathway in *Citrobacter amalonaticus* Y19. *Int J Hydrogen Energy* **33**, 1471–1482.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P. & other authors. **(2017).** vegan:

Community Ecology Package. R package.

Oshima, T. & Imahori, K. (1974). Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *Int J Syst Bacteriol* **24**, 102–112.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996–1004.

Parshina, S. N., Kijlstra, S., Henstra, A. M., Sipma, J., Plugge, C. M. & Stams, A. J. M. (2005a). Carbon monoxide conversion by thermophilic sulfate-reducing bacteria in pure culture and in co-culture with *Carboxydothermus hydrogenoformans*. *Appl Microbiol Biotechnol* **68**, 390–396.

Parshina, S. N., Sipma, J., Nakashimada, Y., Henstra, A. M., Smidt, H., Lysenko, A. M., Lens, P. N. L., Lettinga, G. & Stams, A. J. M. (2005b). *Desulfotomaculum carboxydivorans* sp. nov., a novel sulfate-reducing bacterium capable of growth at 100% CO. *Int J Syst Evol Microbiol* **55**, 2159–2165.

Peng, T., Pan, S., Christopher, L. P., Sparling, R. & Levin, D. B. (2016). Growth and metabolic profiling of the novel thermophilic bacterium *Thermoanaerobacter* sp. strain YS13. *Can J Microbiol* **62**, 762–771.

Peretó, J. (2012). Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks. *Chem Soc Rev* **41**, 5394–5403.

Pierce, E., Xie, G., Barabote, R. D., Saunders, E., Han, C. S., Detter, J. C., Richardson, P., Brettin, T. S., Das, A. & other authors. **(2008).** The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*).

Environ Microbiol **10**, 2550–2573.

van der Ploeg, J. R., Weiss, M. A., Saller, E., Nashimoto, H., Saito, N., Kertesz, M. A. & Leisinger, T. (1996). Identification of sulfate starvation-regulated genes in *Escherichia coli*: a gene cluster involved in the utilization of taurine as a sulfur source. *J Bacteriol* **178**, 5438–5446.

Price, M. N., Dehal, P. S. & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490.

Pruesse, E., Peplies, J. & Glöckner, F. O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596.

R Core Team. **(2016).** R: A language and environment for statistical computing.

Ragsdale, S. W. (1997). The Eastern and Western branches of the Wood/Ljungdahl pathway: how the East and West were won. *Biofactors* **6**, 3–11.

Ragsdale, S. W. (2004). Life with carbon monoxide. *Crit Rev Biochem Mol Biol* **39**, 165–195.

Reysenbach, A. L., Wickham, G. S. & Pace, N. R. (1994). Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl Environ Microbiol* **60**, 2113–2119.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–277.

van Rijsbergen, C. J. (1979). *Information Retrieval*, 2nd edn. Butterworth: Boston.

Robb, F. T. & Techtmann, S. M. (2018). Life on the fringe: microbial adaptation to growth on carbon monoxide. *F1000Research* **7**.

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584.

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat Prod Rep* **16**, 565–574.

Rose, J. J., Wang, L., Xu, Q., McTiernan, C. F., Shiva, S., Tejero, J. & Gladwin, M. T. (2017). Carbon Monoxide Poisoning: Pathogenesis, Management, and Future Directions of Therapy. *Am J Respir Crit Care Med* **195**, 596–606.

Sant'Anna, F. H., Lebedinsky, A. V, Sokolova, T. G., Robb, F. T. & Gonzalez, J. M. (2015). Analysis of three genomes within the thermophilic bacterial species *Caldanaerobacter subterraneus* with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. *BMC Genomics* **16**, 757.

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K. & other authors. **(2019).** Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **47**, D23–D28.

Schade, G. W. & Crutzen, P. J. (1999). CO emissions from degrading plant matter (II). Estimate of a global source strength. *Tellus B Chem Phys Meteorol* **51**, 909–918.

Schattner, P., Brooks, A. N. & Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33**, W686–W689.

Schauder, R., Preuß, A., Jetten, M. & Fuchs, G. (1988). Oxidative and reductive acetyl CoA/carbon monoxide dehydrogenase pathway in *Desulfobacterium autotrophicum*. *Arch Microbiol* **151**, 84–89.

Schoelmerich, M. C. & Müller, V. (2019). Energy conservation by a hydrogenase-dependent chemiosmotic mechanism in an ancient metabolic pathway. *Proc Natl Acad Sci U S A* **116**, 6329–6334.

Schuchmann, K. & Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* **12**, 809–821.

Seemann, M., Tse Sum Bui, B., Wolff, M., Miginiac-Maslow, M. & Rohmer, M. (2006). Isoprenoid biosynthesis in plant chloroplasts via the MEP pathway: Direct thylakoid/ferredoxin-dependent photoreduction of GcpE/IspG. *FEBS Lett* **580**, 1547–1552.

Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B. & Dunfield, P. F. (2014). Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J* **8**, 1166–1174.

Shelobolina, E. S., Nevin, K. P., Blakeney-Hayward, J. D., Johnsen, C. V, Plaia, T. W., Krader, P., Woodard, T., Holmes, D. E., VanPraagh, C. G. & Lovley, D. R. (2007). *Geobacter pickeringii* sp. nov., *Geobacter argillaceus* sp. nov. and *Pelosinus fermentans* gen. nov., sp. nov., isolated from subsurface kaolin lenses. *Int J Syst Evol Microbiol* **57**, 126–135.

Shiba, H., Kawasumi, T., Igarashi, Y., Kodama, T. & Minoda, Y. (1982). The Deficient Carbohydrate Metabolic Pathways and the Incomplete Tricarboxylic Acid Cycle in an Obligately Autotrophic Hydrogen-oxidizing Bacterium. *Agric*

Biol Chem **46**, 2341–2345.

Shiba, H., Kawasumi, T., Igarashi, Y., Kodama, T. & Minoda, Y. (1985). The CO_2 assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Arch Microbiol* **141**, 198–203.

Singer, S. W., Hirst, M. B. & Ludden, P. W. (2006). CO-dependent H_2 evolution by *Rhodospirillum rubrum*: Role of CODH:CooF complex. *Biochim Biophys Acta - Bioenerg* **1757**, 1582–1591.

Slepova, T. V., Sokolova, T. G., Lysenko, A. M., Tourova, T. P., Kolganova, T. V., Kamzolkina, O. V., Karpov, G. A. & Bonch-Osmolovskaya, E. A. (2006). *Carboxydocella sporoproducens* sp. nov., a novel anaerobic CO-utilizing/H_2-producing thermophilic bacterium from a Kamchatka hot spring. *Int J Syst Evol Microbiol* **56**, 797–800.

Slobodkin, A., Reysenbach, A.-L., Mayer, F. & Wiegel, J. (1997). Isolation and characterization of the homoacetogenic thermophilic bacterium *Moorella glycerini* sp. nov. *Int J Syst Bacteriol* **47**, 969–974.

Slobodkin, A. I., Sokolova, T. G., Lysenko, A. M. & Wiegel, J. (2006). Reclassification of *Thermoterrabacterium ferrireducens* as *Carboxydothermus ferrireducens* comb. nov., and emended description of the genus *Carboxydothermus*. *Int J Syst Evol Microbiol* **56**, 2349–2351.

Soboh, B., Linder, D. & Hedderich, R. (2002). Purification and catalytic properties of a CO-oxidizing:H_2-evolving enzyme complex from *Carboxydothermus hydrogenoformans*. *Eur J Biochem* **269**, 5712–5721.

Sokolova, T. & Lebedinsky, A. (2013). CO-Oxidizing Anaerobic Thermophilic

Prokaryotes, pp. 203–231. *In* Satyanarayana, T., Littlechild, J. & Kawarabayasi, Y. (ed), *Thermophilic Microbes Environ Ind Biotechnol*, Springer.

Sokolova, T. G., González, J. M., Kostrikina, N. A., Chernyh, N. A., Tourova, T. P., Kato, C., Bonch-Osmolovskaya, E. A. & Robb, F. T. (2001). *Carboxydobrachium pacificum* gen. nov., sp. nov., a new anaerobic, thermophilic, CO-utilizing marine bacterium from Okinawa Trough. *Int J Syst Evol Microbiol* **51**, 141–149.

Sokolova, T. G., González, J. M., Kostrikina, N. A., Chernyh, N. A., Slepova, T. V., Bonch-Osmolovskaya, E. A. & Robb, F. T. (2004a). *Thermosinus carboxydivorans* gen. nov., sp. nov., a new anaerobic, thermophilic, carbon-monoxide-oxidizing, hydrogenogenic bacterium from a hot pool of Yellowstone National Park. *Int J Syst Evol Microbiol* **54**, 2353–2359.

Sokolova, T. G., Kostrikina, N. A., Chernyh, N. A., Tourova, T. P., Kolganova, T. V & Bonch-Osmolovskaya, E. A. (2002). *Carboxydocella thermautotrophica* gen. nov., sp. nov., a novel anaerobic, CO-utilizing thermophile from a Kamchatkan hot spring. *Int J Syst Evol Microbiol* **52**, 1961–1967.

Sokolova, T. G., Jeanthon, C., Kostrikina, N. A., Chernyh, N. A., Lebedinsky, A. V, Stackebrandt, E. & Bonch-Osmolovskaya, E. A. (2004b). The first evidence of anaerobic CO oxidation coupled with H_2 production by a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Extremophiles* **8**, 317–323.

Sokolova, T. G., Henstra, A.-M., Sipma, J., Parshina, S. N., Stams, A. J. M. & Lebedinsky, A. V. (2009). Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. *FEMS Microbiol Ecol* **68**, 131–141.

Søndergaard, D., Pedersen, C. N. S. & Greening, C. (2016). HydDB: A web tool for

hydrogenase classification and analysis. *Sci Rep* **6**, 34212.

Spormann, A. M. & Thauer, R. K. (1988). Anaerobic acetate oxidation to CO_2 by *Desulfotomaculum acetoxidans*. *Arch Microbiol* **150**, 374–380.

Srinivas, T. N. R., Vijaya Bhaskar, Y., Bhumika, V. & Anil Kumar, P. (2013). *Photobacterium marinum* sp. nov., a marine bacterium isolated from a sediment sample from Palk Bay, India. *Syst Appl Microbiol* **36**, 160–165.

Stupperich, E., Hammel, K. E., Fuchs, G. & Thauer, R. K. (1983). Carbon monoxide fixation into the carboxyl group of acetyl coenzyme A during autotrophic growth of *Methanobacterium*. *FEBS Lett* **152**, 21–23.

Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M. & Abe, S. (1983). *Bacillus thermoglucosidasius* sp. nov., a New Species of Obligately Thermophilic Bacilli. *Syst Appl Microbiol* **4**, 487–495.

Svetlichny, V. A., Sokolova, T. G., Gerhardt, M., Ringpfeil, M., Kostrikina, N. A. & Zavarzin, G. A. (1991). *Carboxydothermus hydrogenoformans* gen. nov., sp. nov., a CO-utilizing thermophilic anaerobic bacterium from hydrothermal environments of Kunashir Island. *Syst Appl Microbiol* **14**, 254–260.

Svetlitchnyi, V., Dobbek, H., Meyer-Klaucke, W., Meins, T., Thiele, B., Römer, P., Huber, R. & Meyer, O. (2004). A functional Ni-Ni-[4Fe-4S] cluster in the monomeric acetyl-CoA synthase from *Carboxydothermus hydrogenoformans*. *Proc Natl Acad Sci U S A* **101**, 446–451.

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. & Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PLoS One* **9**, e105592.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011).

MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* **28**, 2731–2739.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V, Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**, 22–28.

Tavares, A. F. N., Teixeira, M., Romão, C. C., Seixas, J. D., Nobre, L. S. & Saraiva, L. M. (2011). Reactive Oxygen Species Mediate Bactericidal Killing Elicited by Carbon Monoxide-releasing Molecules. *J Biol Chem* **286**, 26708–26717.

Techtmann, S. M., Lebedinsky, A. V., Colman, A. S., Sokolova, T. G., Woyke, T., Goodwin, L. & Robb, F. T. (2012). Evidence for Horizontal Gene Transfer of Anaerobic Carbon Monoxide Dehydrogenases. *Front Microbiol* **3**, 132.

Techtmann, S. M., Colman, A. S. & Robb, F. T. (2009). 'That which does not kill us only makes us stronger': the role of carbon monoxide in thermophilic microbial consortia. *Environ Microbiol* **11**, 1027–1037.

Thauer, R. K., Jungermann, K. & Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol Rev* **41**, 100–180.

Toshchakov, S. V, Lebedinsky, A. V, Sokolova, T. G., Zavarzina, D. G., Korzhenkov, A. A., Teplyuk, A. V, Chistyakova, N. I., Rusakov, V. S., Bonch-Osmolovskaya, E. A. & other authors. (2018). Genomic insights into energy metabolism of *Carboxydocella thermautotrophica* coupling hydrogenogenic CO oxidation with the reduction of Fe(III) minerals. *Front Microbiol* **9**, 1759.

Vignais, P. M. & Billoud, B. (2007). Occurrence, Classification, and Biological

Function of Hydrogenases: An Overview. *Chem Rev* **107**, 4206–4272.

Wakim, B. T. & Uffen, R. L. (1983). Membrane association of the carbon monoxide oxidation system in *Rhodopseudomonas gelatinosa*. *J Bacteriol* **153**, 571–573.

Wareham, L. K., Southam, H. M. & Poole, R. K. (2018). Do nitric oxide, carbon monoxide and hydrogen sulfide really qualify as 'gasotransmitters' in bacteria? *Biochem Soc Trans* **46**, 1107–1118.

Wrighton, K. C., Agbo, P., Warnecke, F., Weber, K. A., Brodie, E. L., DeSantis, T. Z., Hugenholtz, P., Andersen, G. L. & Coates, J. D. (2008). A novel ecological role of the Firmicutes identified in thermophilic microbial fuel cells. *ISME J* **2**, 1146–1156.

Wu, L. & Wang, R. (2005). Carbon Monoxide: Endogenous Production, Physiological Functions, and Pharmacological Applications. *Pharmacol Rev* **57**, 585–630.

Wu, M., Ren, Q., Durkin, S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., Madupu, R., Sullivan, S. A., Kolonay, J. F. & other authors. (2005). Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet* **1**, e65.

Xavier, J. C., Preiner, M. & Martin, W. F. (2018). Something special about CO-dependent CO_2 fixation. *FEBS J* **285**, 4181–4195.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134.

Yoneda, Y., Yoshida, T., Kawaichi, S., Daifuku, T., Takabe, K. & Sako, Y. (2012). *Carboxydothermus pertinax* sp. nov., a thermophilic, hydrogenogenic, Fe(III)-reducing, sulfur-reducing carboxydotrophic bacterium from an acidic hot spring.

Int J Syst Evol Microbiol **62**, 1692–1697.

Yoneda, Y., Yoshida, T., Yasuda, H., Imada, C. & Sako, Y. (2013a). A thermophilic, hydrogenogenic and carboxydotrophic bacterium, *Calderihabitans maritimus* gen. nov., sp. nov., from a marine sediment core of an undersea caldera. *Int J Syst Evol Microbiol* **63**, 3602–3608.

Yoneda, Y., Yoshida, T., Daifuku, T., Kitamura, T., Inoue, T., Kano, S. & Sako, Y. (2013b). Quantitative detection of carboxydotrophic bacteria *Carboxydothermus* in a hot aquatic environment. *Fundam Appl Limnol* **182**, 161–170.

Yoneda, Y., Kano, S. I., Yoshida, T., Ikeda, E., Fukuyama, Y., Omae, K., Kimura-Sakai, S., Daifuku, T., Watanabe, T. & Sako, Y. (2015). Detection of anaerobic carbon monoxide-oxidizing thermophiles in hydrothermal environments. *FEMS Microbiol Ecol* **91**, fiv093.

Zavarzina, D. G., Sokolova, T. G., Tourova, T. P., Chernyh, N. A., Kostrikina, N. A. & Bonch-Osmolovskaya, E. A. (2007). *Thermincola ferriacetica* sp. nov., a new anaerobic, thermophilic, facultatively chemolithoautotrophic bacterium capable of dissimilatory Fe(III) reduction. *Extremophiles* **11**, 1–7.

Zeigler, D. R. (2014). The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology* **160**, 1–11.

Zerbino, D. R. & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829.

Zhao, Y., Caspers, M. P., Abee, T., Siezen, R. J. & Kort, R. (2012). Complete Genome Sequence of *Geobacillus thermoglucosidans* TNO-09.020, a Thermophilic Sporeformer Associated with a Dairy-Processing Environment. *J Bacteriol* **194**, 4118.

Publication list

1. **Omae, K., Yoneda, Y., Fukuyama, Y., Yoshida, T., & Sako, Y. (2017).** Genomic analysis of *Calderihabitans maritimus* KKC1, a thermophilic hydrogenogenic carboxydotrophic bacterium isolated from marine sediment. *Appl Environ Microbiol* **83**, e00832-17.
2. **Omae, K., Fukuyama, Y., Yasuda, H., Mise, K., Yoshida, T., & Sako, Y. (2019).** Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs revealed by microbial community analysis in sediments from multiple hydrothermal environments in Japan. *Arch Microbiol* **201**, 969–982.
3. **Omae, K., Oguro, T., Inoue, M., Fukuyama, Y., Yoshida, T., & Sako, Y.** Diversity analysis of thermophilic hydrogenogenic carboxydotrophs by carbon monoxide dehydrogenase amplicon sequencing using new primers. (In preparation for submission).

Other publications

1. **Yoneda, Y., Kano, S. I., Yoshida, T., Ikeda, E., Fukuyama, Y., Omae, K., Kimura-Sakai, S., Daifuku, T., Watanabe, T., & Sako, Y. (2015).** Detection of anaerobic carbon monoxide-oxidizing thermophiles in hydrothermal environments. *FEMS Microbiol Ecol* **91**, fiv093.
2. **Nishimura, Y., Watai, H., Honda, T., Mihara, T., Omae, K., Roux, S., Blanc-Mathieu, R., Yamamoto, K., Hingamp, P., Sako, Y., Sullivan, M. B., Goto, S.,**

Ogata, H., & Yoshida, T. (2017). Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* **2**, e00359-16.

3. **Fukuyama, Y., Omae, K., Yoneda, Y., Yoshida, T., & Sako, Y. (2017).** Draft genome sequences of *Carboxydothermus pertinax* and *C. islandicus*, hydrogenogenic carboxydotrophic bacteria. *Genome Announc* **5**, e01648-16.
4. **Fukuyama, Y., Oguro, T., Omae, K., Yoneda, Y., Yoshida, T., & Sako, Y. (2017).** Draft genome sequences of two hydrogenogenic carboxydotrophic bacteria, *Carboxydocella* sp. strains JDF658 and ULO1, isolated from two distinct volcanic fronts in Japan. *Genome Announc* **5**, e00242-17.
5. **Fukuyama Y., Omae K., Yoneda Y., Yoshida T., & Sako Y. (2018).** Insight into energy conservation via alternative carbon monoxide metabolism in *Carboxydothermus pertinax* revealed by comparative genome analysis. *Appl Environ Microbiol* **84**, e00458-18.
6. **Inoue, M., Nakamoto, I., Omae, K., Oguro, T., Ogata, H., Yoshida, T., & Sako, Y. (2019).** Structural and phylogenetic diversity of anaerobic carbon-monoxide dehydrogenases. *Front Microbiol* **9**, 3353.
7. **Fukuyama, Y., Omae, K., Yoshida, T., & Sako, Y. (2019).** Transcriptome analysis of a thermophilic and hydrogenogenic carboxydotroph *Carboxydothermus pertinax*. *Extremophiles* **23**, 389–398.
8. **Fukuyama, Y., Tanimura, A., Inoue, M., Omae, K., Yoshida, T., & Sako, Y. (2019).** Draft genome sequences of two thermophilic bacteria *Moorella* spp., isolated from an acidic hot spring in Japan. *Microbiol Resour Announc* **8**, e00663-19.