

ORIGINAL RESEARCH

Comparative genomic analysis of *Thermus* provides insights into the evolutionary history of an incomplete denitrification pathway

Jian-Yu Jiao^{1,*}, Zheng-Han Lian^{1,*}, Meng-Meng Li^{1,*}, Nimaichand Salam¹, En-Min Zhou², Lan Liu¹, Hong Ming³, Guoxing Nie⁴, Wensheng Shu⁵, Guoping Zhao⁶, Brian P. Hedlund^{7,8,*}, and Wen-Jun Li^{1,9,*}

Edited by Meng Li, Shenzhen University, China; Received August 16, 2021; Accepted January 8, 2022; Published online April 29, 2022

Abstract

Biological denitrification is a crucial process in the nitrogen biogeochemical cycle, and *Thermus* has been reported to be a significant heterotrophic denitrifier in terrestrial geothermal environments. However, neither the denitrification potential nor the evolutionary history of denitrification genes in the genus *Thermus* or phylum *Deinococcota* is well understood. Here, we performed a comparative analysis of 23 *Thermus* genomes and identified denitrification genes in 15 *Thermus* strains. We confirmed that *Thermus* harbors an incomplete denitrification pathway as none of the strains contain the *nosZ* gene. Ancestral character state reconstructions and phylogenetic analyses showed that *narG*, *nirS*, and *norB* genes were acquired by the last common ancestor of *Thermales* and were inherited vertically. In contrast, *nirK* of *Thermales* was acquired via two distinct horizontal gene transfers from *Proteobacteria* to the genus *Caldithermus* and from an unknown donor to the common ancestor of all known *Thermus* species except *Thermus filiformis*. This study expands our understanding of the genomic potential for incomplete denitrification in *Thermus*, revealing a largely vertical evolutionary history of the denitrification pathway in the *Thermaceae*, and supporting the important role for *Thermus* as an important heterotrophic denitrifier in geothermal environments.

Keywords: comparative genomics; denitrification; evolutionary history; *Thermus*

Impact statement

For decades, *Thermus* was always considered to be aerobic. However, recent studies have suggested that the denitrification abilities of *Thermus* species may be widely underestimated. In the present study, we used comparative genomic analysis to investigate the evolutionary history of the denitrification pathway in *Thermus* and other members of the phylum *Deinococcota*. We revealed incomplete denitrification pathways to be common in *Thermus* and showed they are inherited mostly vertically, which further supports the importance of *Thermus* as a significant denitrifier in hydrothermal environments.

INTRODUCTION

Thermus species are members of the family *Thermaceae* and are generally Gram-strain-negative, high-G + C-content, nonmotile, rod-shaped, and obligately aerobic or facultatively anaerobic. The genus *Thermus* belongs to the phylum

Deinococcota, along with the genera *Calidithermus*, *Marinithermus*, *Meiothermus*, *Oceanithermus*, *Rhabdothermus*, *Vulcanithermus*, *Deinococcus*, *Deinobacterium*, and *Truepera*. To date, the genus *Thermus* comprises 19 validly

¹State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Life Sciences, Sun Yat-Sen University, Guangzhou, China. ²International Joint Research Center for Karstology, School of Earth Sciences, Yunnan University, Kunming, China. ³Synthetic Biology Engineering Laboratory of Henan Province, College of Life Sciences and Technology, Xinxiang Medical University, Xinxiang, China. ⁴College of Fisheries, Henan Normal University, Xinxiang, China. ⁵Institute of Ecological Science, School of Life Science, South China Normal University, Guangzhou, China. ⁶Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. ⁷School of Life Sciences, University of Nevada, Las Vegas, Nevada, USA. ⁸Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, Nevada, USA. ⁹State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China.

* Correspondence: Brian P. Hedlund, brian.hedlund@unlv.edu; Wen-Jun Li, liwenjun3@mail.sysu.edu.cn

[#]Jian-Yu Jiao, Zheng-Han Lian, and Meng-Meng Li contributed equally to this study.

DOI: 10.1002/mlf2.12009

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

published species names under the International Code of Nomenclature of Prokaryotes (ICNP)¹, and members of this genus are archetypal thermophilic bacteria that are readily isolated from terrestrial geothermal environments, especially circumneutral pH or alkaline hot springs². *Thermus* species are well known for their biotechnological applications, such as the production of thermostable enzymes (e.g., *Taq* DNA polymerase)³. Several species of *Thermus* were reported to have incomplete denitrification pathways with corresponding experimental evidence for the production of nitrite or nitrous oxide^{4,5}.

Denitrification is a crucial component of the nitrogen cycle in geothermal systems. Complete biological denitrification is a respiratory process that reduces nitrogenous oxides to dinitrogen and is normally catalyzed by nitrate reductase (NarGHI or NapAB), nitrite reductase (NirK or NirS), nitric oxide reductase (NorBC), and nitrous oxide reductase (NosZ) under oxygen-limiting or anaerobic conditions². Previous studies have shown that denitrification is highly active in hot springs and that members of the genus *Thermus* play a significant role as heterotrophic denitrifiers in hot spring environments^{4,6}. However, the distribution of denitrification genes within *Thermus* and the evolutionary history of the *Thermus* denitrification pathway have not been investigated.

Currently, only a few strains of the genus *Thermus* have been reported to be incomplete denitrifiers, and there is no consensus in the literature on the nature of denitrification in the genus *Thermus*. Furthermore, considering the high plasticity of *Thermus* genomes⁷, natural competence to transform DNA⁸, and abundance of insertion sequence (IS) elements and prophages on *Thermus* genomes, including the plasmid-borne nitrate conjugative element in *Thermus thermophilus* HB8 and NAR1^{5,9–11}, we hypothesize that denitrification genes might be obtained through various horizontal gene transfer (HGT) events. If so, the distribution and evolutionary origins of denitrification genes in various *Thermus* species might be highly variable and responsive to selective forces in individual geothermal systems, including the supply of oxidized nitrogen species.

To test our hypothesis, we performed a comparative genome analysis of *Thermus* species and analyzed three aspects of the evolutionary history of the *Thermus* denitrification pathway viz. (1) the presence of genes involved in denitrification pathway in a representative of all *Thermus* species; (2) phylogenetic analysis of major denitrification proteins; and (3) investigation of gene gain and loss events in *Thermus* genomes.

RESULTS AND DISCUSSION

Phylogenetic analysis shows monophyly of the genus *Thermus*

The genus *Thermus* presently harbors 19 species, most of which can be isolated from terrestrial geothermal environments. A phylogenetic dendrogram indicating the phylogenetic relationships within the phylum *Deinococcota* was generated based on a concatenated alignment of conserved

positions based on 120 bacterial marker genes, as implemented by Genome Taxonomy Database-Toolkit (GTDB-Tk)¹². The phylogenomic tree showed that the genus *Thermus* is monophyletic within the family *Thermaceae* with high bootstrap support (Figure 1A). Both average nucleotide identity (ANI) and average amino acid identity (AAI) values showed large genomic divergences among the *Thermus* genomes (Figure 1B,C and Tables S1–S2). However, among the 23 *Thermus* strains, *Thermus thermophilus* and *Thermus parvatiensis* formed a monophyletic clade, and both ANI and AAI values between these two species were higher than 95%, indicating they belong to the same species^{13,14}. Given the priority of the species *T. thermophilus*, we suggest that *T. parvatiensis* is a later heterotypic synonym of *T. thermophilus*. The independence of all other *Thermus* species was confirmed using a 95% ANI species boundary¹⁵. The evolutionary distances between *Thermus filiformis* ATCC 43280^T and the other *Thermus* strains were high, although *T. filiformis* ATCC 43280^T was monophyletic with other members of the genus *Thermus* with high bootstrap support (Figure 1 and Tables S1–S2).

Comparative genome analysis of *Thermus* shows similar genomic features

The properties and statistics of the finished and draft *Thermus* genomes are summarized in Table 1. Twenty-three *Thermus* genomes with 20 different species names ranged in size from 2.02 to 2.56 Mbp, with G + C content between 64.81% and 69.50%. The number of protein-coding genes ranged from 2224 to 2749, while the number of predicted transfer RNA (tRNAs), encoding almost all 20 amino acids, ranged from 43 to 53. The estimated completeness of all the genomes based on conserved marker genes was more than 99% (except *Thermus amyloliquefaciens* YIM 77409^T at 97.81% and *T. parvatiensis* RL at 95.76%). The genome sizes of *Thermus* species were smaller than those of most members within the phylum *Deinococcota*, and the coding density of *Thermus* was higher compared with most other species (Table S3). The smaller genome sizes of *Thermus* reflect a well-documented pattern of genomic reduction among thermophiles¹⁶. Mobile genetic elements such as IS elements and genomic islands (GI) were also detected. Each *Thermus* genome contains a large number of IS elements (Table S4), and almost all the genomes (22/23) contain a GI (Table S5). However, no denitrification genes were detected close to any of the mobile genetic elements, suggesting that the denitrification genes might not be transferred through these two mobilomes. The detection of CRISPR-Cas system indicated their bacterial immune system for protection against alien DNA.

The genetic variability among *Thermus* genomes can be further determined from the distribution of the conserved (core) and species-specific (unique) genes. The pan-genome of the *Thermus* genomes comprised 6992 gene clusters including 1027 homologous gene clusters in the core genome (Figure 2A). Comparative genome analysis revealed that each

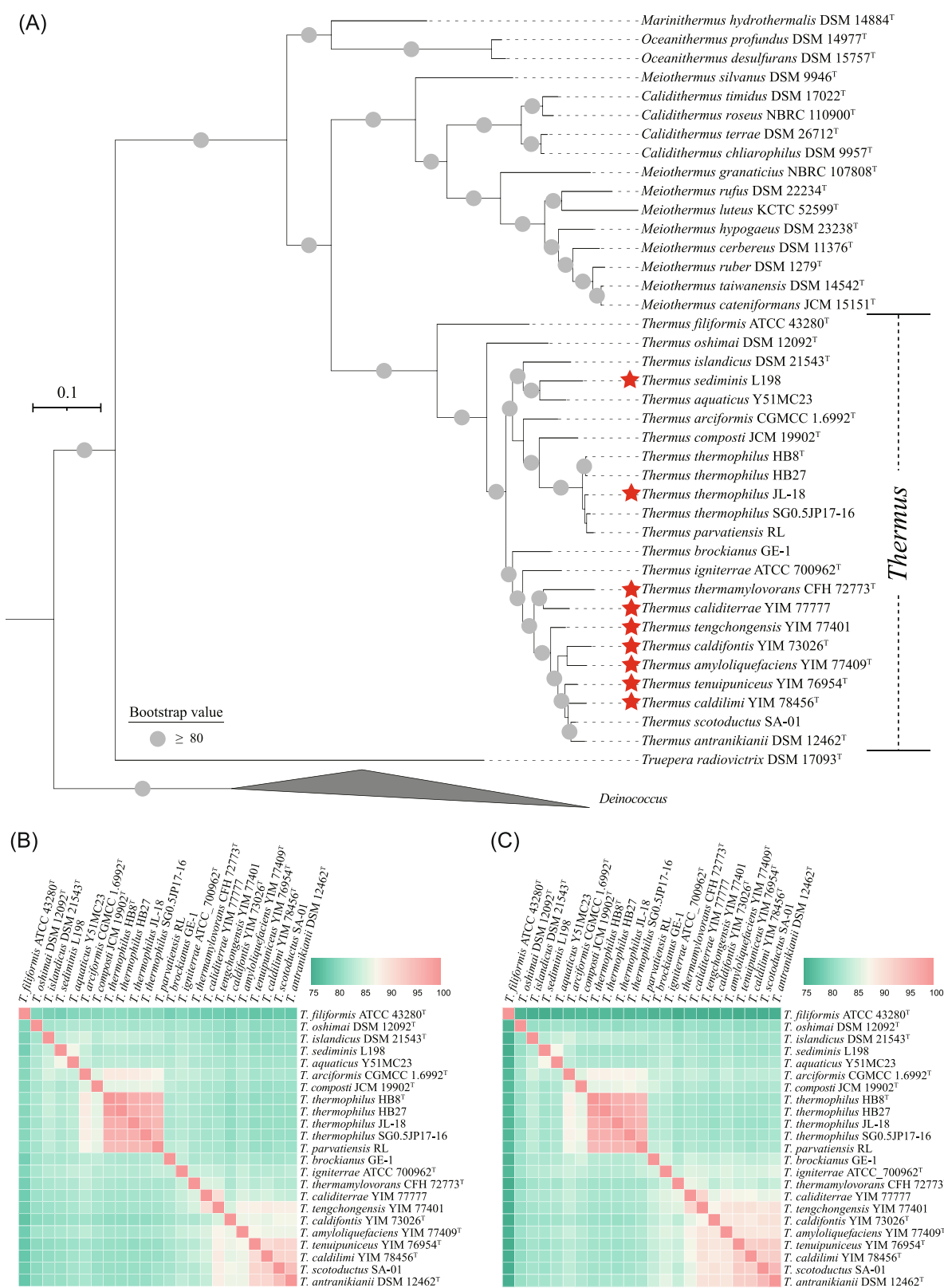


Figure 1. Phylogenetic inference of *Thermus* strains. (A) Phylogenetic placement of *Thermus* strains. Multiple sequence alignments of 120 bacterial marker genes were generated by GTDB-Tk and IQ-Tree was used to construct the phylogenomic tree. Red stars illustrate that the strains were isolated or the genomes were sequenced previously. (B) Average nucleotide identity (ANI) values shared among the *Thermus* species. (C) Average amino acid identity (AAI) values shared among the *Thermus* species. The ANI and AAI values are presented in Tables S1–S2, respectively.

Table 1. General features of the *Thermus* genomes.

Strain	Completeness (%)	Contamination (%)	Genome size (bp)	Scaffolds	<i>N</i> ₅₀ (scaffolds)	GC (%)	Coding density (%)	Predicted genes	Unique genes	tRNAs	rRNA copies	CRISPR
<i>Thermus amyloliquefaciens</i> YIM 77409 ^T	97.81	0	21,60,855	6	20,55,291	67.44	92.61	2315	129	48	2	5
<i>Thermus antranikianii</i> DSM 12462 ^T	100.00	0	21,65,150	33	2,18,058	64.81	94.46	2265	41	47	1	0
<i>Thermus aquaticus</i> Y51MC23	100.00	0	23,38,641	5	21,58,963	68.04	93.15	2501	145	53	2	10
<i>Thermus arciformis</i> CGMCC 1.6992 ^T	100.00	0	24,42,297	88	58,107	68.79	93.70	2643	105	47	2	12
<i>Thermus brockianus</i> GE-1	99.36	0	23,88,273	3	20,35,182	66.90	93.27	2530	132	47	2	9
<i>Thermus caldifontis</i> YIM 73026 ^T	100.00	0	21,63,786	41	1,93,267	64.85	94.27	2275	52	46	1	4
<i>Thermus caldlimi</i> YIM 78456 ^T	99.58	0	24,47,659	1	24,47,659	65.14	93.87	2639	155	47	2	0
<i>Thermus caliditerrae</i> YIM 77777	100.00	0.42	22,18,114	4	20,46,548	67.22	93.83	2285	76	50	2	5
<i>Thermus compositi</i> JCM 19902 ^T	100.00	0	20,80,570	64	82,383	67.70	94.64	2247	54	45	1	3
<i>Thermus filiformis</i> ATCC 43280 ^T	99.15	0.42	23,86,081	40	5,51,922	69.01	92.76	2410	188	47	2	15
<i>Thermus igniterrae</i> ATCC 700962 ^T	99.58	0.42	22,25,983	74	62,194	68.80	94.94	2331	52	43	2	10
<i>Thermus islandicus</i> DSM 21543 ^T	99.15	0	22,63,010	67	2,41,998	68.35	93.75	2412	145	47	2	2
<i>Thermus oshimai</i> DSM 12092 ^T	100.00	0	22,60,954	24	2,00,705	68.74	94.80	2355	93	48	1	15
<i>Thermus parvatiensis</i> RL	95.76	0	20,16,098	2	18,72,821	68.53	90.78	2463	189	47	2	2
<i>Thermus scotoductus</i> SA-01	100.00	0	23,55,186	2	23,46,803	64.89	94.25	2458	80	48	2	5
<i>Thermus sediminis</i> L198	99.15	0	21,60,271	4	20,28,157	68.21	92.46	2287	149	48	2	5
<i>Thermus tengchongensis</i> YIM 77401	100.00	2.54	25,62,314	5	22,24,342	66.40	92.57	2749	179	47	2	5
<i>Thermus tenuipunicus</i> YIM 76954 ^T	100.00	0	22,61,036	66	51,325	65.54	94.02	2375	76	47	1	5
<i>Thermus thermamylovorans</i> CFH 72773 ^T	99.58	0	22,50,808	54	1,71,757	69.50	94.05	2336	89	45	1	11
<i>Thermus thermophilus</i> HB27	99.58	0	21,27,482	2	18,94,877	69.41	93.86	2224	43	48	2	10
<i>T. thermophilus</i> HB8 ^T	99.58	0	21,16,056	3	18,49,742	69.50	93.80	2224	38	48	2	11
<i>T. thermophilus</i> JL-18	100.00	0	23,11,212	3	19,02,595	68.98	94.11	2459	51	49	2	6
<i>T. thermophilus</i> SG0.5JP17-16	100.00	0.42	23,03,227	2	18,63,201	68.62	93.81	2451	98	48	2	8

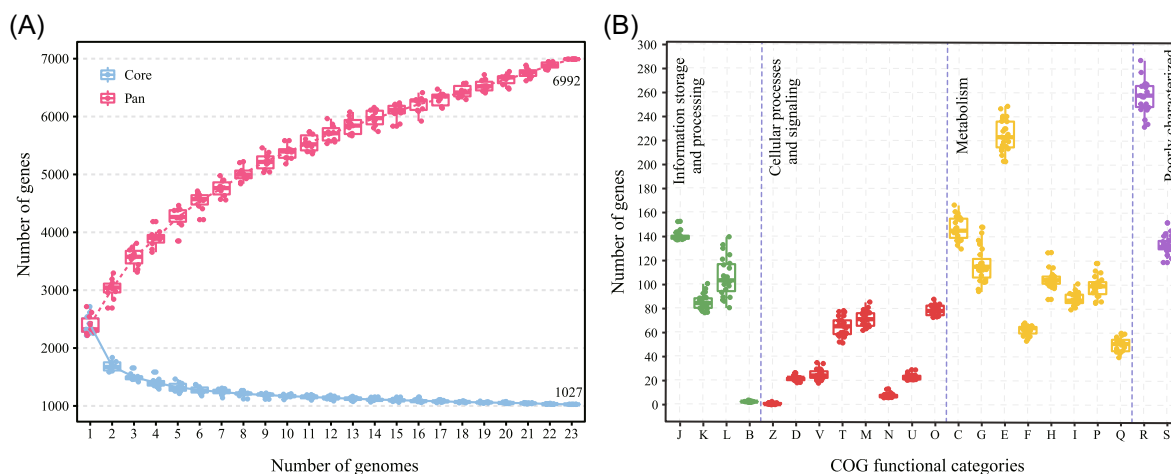


Figure 2. Pan- and core-genome evolution and functional annotations of *Thermus* genomes. (A) Pan- and core-genome evolution of *Thermus*. (B) Categorization of the function of each protein-coding gene was based on COG categories. J: Translation, ribosomal structure, and biogenesis; K: Transcription; L: Replication, recombination, and repair; B: Chromatin structure and dynamics; Z: Cytoskeleton; D: Cell-cycle control, cell division, chromosome partitioning; V: Defense mechanisms; T: Signal transduction mechanisms; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; U: Intracellular trafficking, secretion, and vesicular transport; O: Posttranslational modification, protein turnover, chaperones; C: Energy production and conversion; G: Carbohydrate transport and metabolism; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport, and catabolism; R: General function prediction only; S: Function unknown.

of the *Thermus* species contains a small number (38–189) of unique genes in their genomes (Table 1).

The properties and statistics of genes annotated into the cluster of orthologous groups (COGs) functional categories are shown in Figure 2B and Table S6. The most abundant COGs in the *Thermus* genomes are assigned to general function prediction only (COG category R), followed closely by CDSs dedicated to amino acid transport and metabolism (COG category E). These two categories are in higher proportions compared with other COG categories in all *Thermus* genomes. Notably, for an overall comparison among the genomes of *Thermus* strains, all the genomes contain a similar gene number in each COG category, and almost half of the gene clusters (1027) are conserved in all genomes, which suggests that *Thermus* genomes are more stable than generally considered.

***Thermus* incomplete denitrification pathways evolve through a combination of horizontal and vertical processes**

Nitrogen oxides can act as the terminal electron acceptors for respiration in the absence of oxygen. Previous studies have reported that the members of *Thermus*, such as *Thermus scotoductus* SE-1^T, *Thermus tenuipunctus* YIM 76954^T, and a variety of *T. thermophilus* and *Thermus oshimai* strains can grow anaerobically with nitrate as electron acceptors^{17,18}. In total, 15 of the examined genomes contained annotated denitrification genes (Figure 3), and the genes coding for denitrification enzymes were located close to genes for cytochrome c oxidase (Figures S1–S2), supporting their potential roles in respiration. Furthermore, the synteny *narGHI*, *nirS*, and *norBC* are conserved, suggesting that these genes evolve as

an evolutionary unit (Figures S1–S2). Fourteen of the *Thermus* genomes contain the genes coding for nitrate reductase. Additionally, *narGHI* was prevalent in other genera of *Thermaceae* but was absent from the genera *Truepera* and *Deinococcus*, which belong to the related families *Treuperaceae* and *Deinococcaceae*. Furthermore, phylogenetic analysis of NarG showed that the *Thermales* nitrate reductases were clustered in a single clade (Figure 4), indicating the *narG* genes, overall, were inherited vertically.

Nitrite reduction (NO_2^- to NO) is often a rate-limiting process in denitrification²¹. Although 14 genomes contain nitrite reductases, the enzymes for nitrite reduction are different in different *Thermus* species. *Thermus* species can catalyze nitrite reduction by two types of nitrite reductases, NirK or NirS. Five *Thermus* strains were shown to contain *nirK*, which encodes the Cu-containing nitrite reductase (Figure 3), whereas *nirS*, encoding the isofunctional tetra-heme cytochrome *cd1*-containing nitrite reductase, was present in 12 strains. This finding indicates that *nirS*-type nitrite reductases are more prevalent in *Thermus* species. A previous study reported the distinct roles of *nirK* and *nirS* in the genus *Thermus*: *nirS* was expressed higher in comparison to *nirK* under oxic and stable conditions. In contrast, *nirK* was expressed over a wider range of nitrite concentrations²¹. Three *Thermus* genomes, *T. oshimai* DSM 12092^T, *T. scotoductus* SA-01, and *Thermus antranikianii* DSM 12462^T, harbor two types of nitrite reductases, suggesting that they might be able to utilize nitrite more effectively over a wide concentration range and could use nitrite as an electron acceptor under oxic conditions²¹.

The phylogeny of the NirK from *Thermales* (Figure 5) revealed three major clades, representing the three genera *Thermus*, *Calidithermus*, and *Deinococcus*, which suggests a

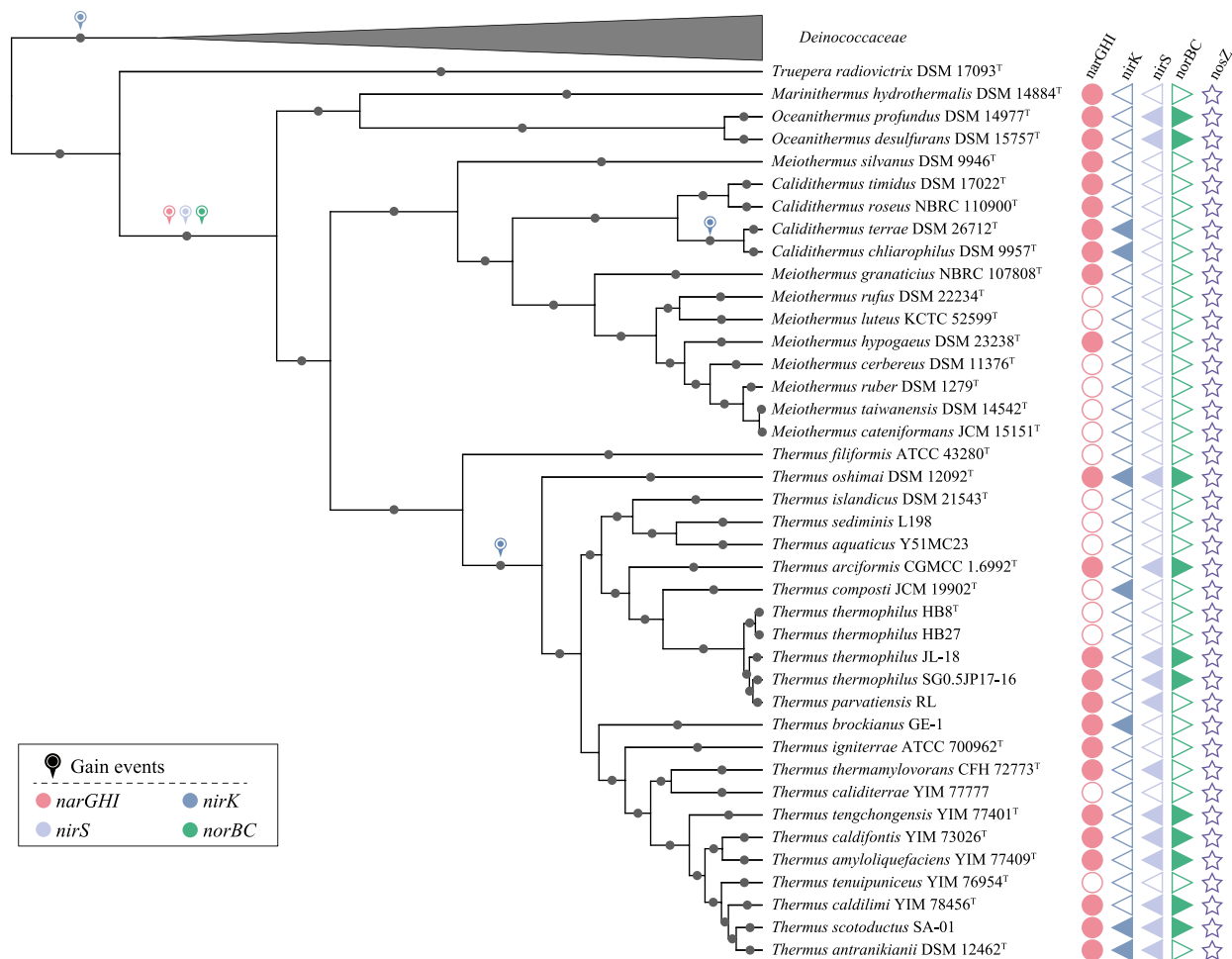


Figure 3. Gain events of denitrification genes in the phylum *Deinococcota*. The Bayesian tree was determined by MrBayes¹⁹, and gene gain events were calculated by using COUNT software²⁰. The black circles represent nodes with posterior possibilities higher than 0.9. The presence or absence of genes related to denitrification genes in the phylum *Deinococcota* is shown on the right of the Bayesian tree. The gain of denitrification genes in *Deinococcota* is marked.

more complex evolutionary history compared with other denitrification genes. The NirK of the genera *Calidithermus* and *Deinococcus* were clearly within a clade found in *Proteobacteria*, indicating potential HGT events from *Proteobacteria* to these two genera. However, the five *Thermus* NirK proteins were monophyletic, and their sister lineages contained multiple phyla (e.g., *Spirochaetes*, *Firmicutes*, *Acidobacteria*, *Crenarchaeota*, etc.), suggesting a different evolutionary origin. In contrast, NirS was only present in the order *Thermales* and formed a monophyletic clade (Figure 6). Furthermore, the ancestral character state reconstruction of NirS (Figure 3) showed that the last common ancestor of *Thermales* likely encoded NirS for nitrite reduction. Our results suggest that the *Thermales* acquired *nirS* before their divergence, and that *nirS* was subsequently inherited vertically with multiple gene losses; in contrast, *nirK* was acquired from different donors.

The genes *norBC*, coding for the nitric oxide reductase, were also detected in the order *Thermales*. In particular, nine *Thermus* genomes and two *Oceanithermus* genomes contain *norBC* (Figure 3). However, no *norBC* was detected

in the families *Deinococcaceae* and *Trueperaceae* (Figure 3). The phylogenetic analysis showed that the NorB of *Thermales* formed a single clade (Figure 7), which indicates that *norBC* was present in the common ancestor of the order and was inherited vertically but with multiple gene loss events.

No *nosZ* gene coding for nitrous-oxide reductase was detected in the *Thermus* genomes (Figure 3), which is consistent with the incomplete denitrification phenotype commonly reported in *Thermus* strains^{4,5}. However, one *nosZ* gene was detected in the genome of the type strain of *Deinococcus ficus* CC-FR2-10^T. The NosZ phylogenetic tree showed that it might have been horizontally transferred from *Firmicutes* (Figure S3). The nitrous-oxide reductase is the most sensitive enzyme in denitrification to oxygen^{22,23}, and the absence of *nosZ* in *Thermus* might reflect the niche of *Thermus* in relatively oxidized habitats within thermal environments. All known *Thermus* species are aerobic and tolerant of atmospheric concentrations of oxygen, despite the low solubility of oxygen at high temperatures. Furthermore, most geothermal springs are sourced with ammonium as the

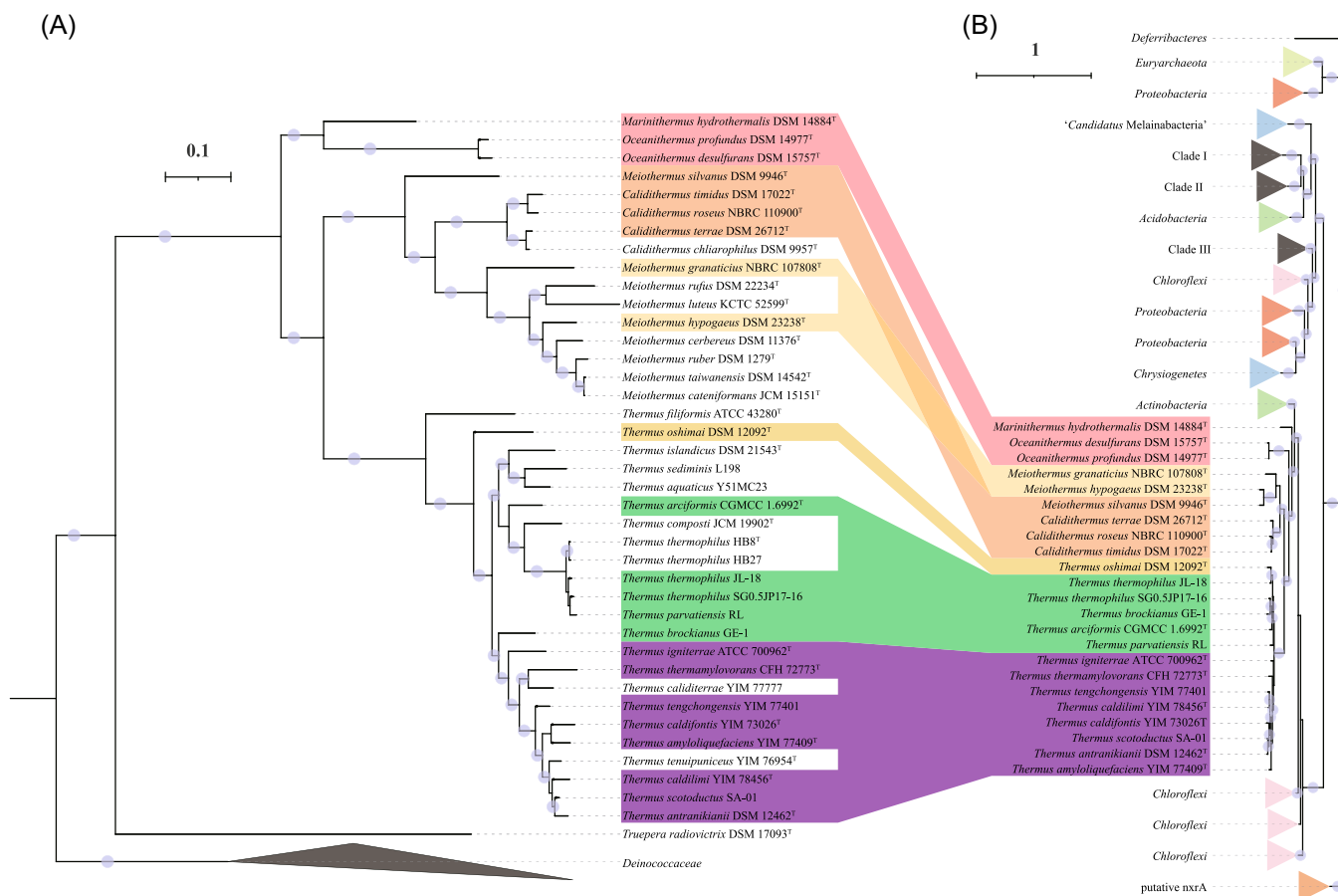


Figure 4. Consistency between the phylogenomic tree of *Deinococcota* and the phylogenetic tree of NarG. (A) The phylogenomic tree of *Deinococcota* was constructed as in Figure 1A. (B) The NarG sequences listed in Table S7 were aligned using MUSCLE with 100 iterations. The NarG tree was constructed using IQ-Tree with the parameters (-alrt 1000 -bb 1000 -nt AUTO). The best-fit model (LG + R9) was determined by ModelFinder.

dominant source of dissolved organic nitrogen²⁴, and aerobic chemolithotrophic ammonia oxidation is thought to be the main process delivering oxidized nitrogen as a source for denitrifiers⁶; the frequent coupling of nitrification and denitrification might select for oxygen tolerance and therefore select against nitrous-oxide reductase, consistent with high nitrous oxide fluxes measured in some geothermal systems⁴.

In total, the phylogenetic analyses of *Thermus* denitrification proteins were not always consistent with phylogenomic trees (Figures 1A and 4–7), suggesting some elements of nonvertical evolution. In addition to the complex evolution of NirK, the denitrification enzymes (NarG, NirS, and NorB) from *T. oshimai* DSM 12092^T consistently branched in the middle of the clade of the *Thermus* enzymes. This position contrasts with the phylogenomic analyses in which *T. oshimai* DSM 12092^T is deep branching within the genus *Thermus* (Figure 1A). The reason may be due to a recent HGT event leading to the gain of the denitrification gene cluster in *T. oshimai* DSM 12092^T from a *Thermus* donor, as suggested by a previous study²⁵, or due to the low phylogenetic resolution of single genes at the highest and lowest taxonomic ranks²⁶.

The emerging view of the evolution of denitrification genes in *Thermales* is one in which the last common ancestor of the order gained an incomplete denitrification gene cluster containing *narGHI*, *nirS*, and *norBC*, which was generally vertically inherited, but with multiple gene losses, particularly for *nirS* and *norBC*. A completely separate evolution is inferred for NirK. Although the genera *Thermus*, *Caldithermus*, and *Deinococcus* (*Deinococcales*) contain *nirK*, phylogenetic analysis of NirK showed that these proteins are not closely related among the genera and were thus gained through independent HGT events. Further back, we infer that the last common ancestor of the *Deinococcota* did not contain genes for complete denitrification, and the *Thermales* acquired these genes early in their evolutionary history, allowing them to adapt to low-oxygen or anoxic conditions where nitrification or other processes provide oxidized nitrogen species to support denitrification.

In summary, we first analyzed the distribution and evolution of denitrification genes in *Deinococcota* genomes and provided insights into the evolutionary history of the *Thermus* denitrification pathway. Similar to other studies, our results show that the *Thermus* genomes contain an incomplete denitrification pathway; more than half of the investigated *Thermus* genomes contain one or more denitrification genes,

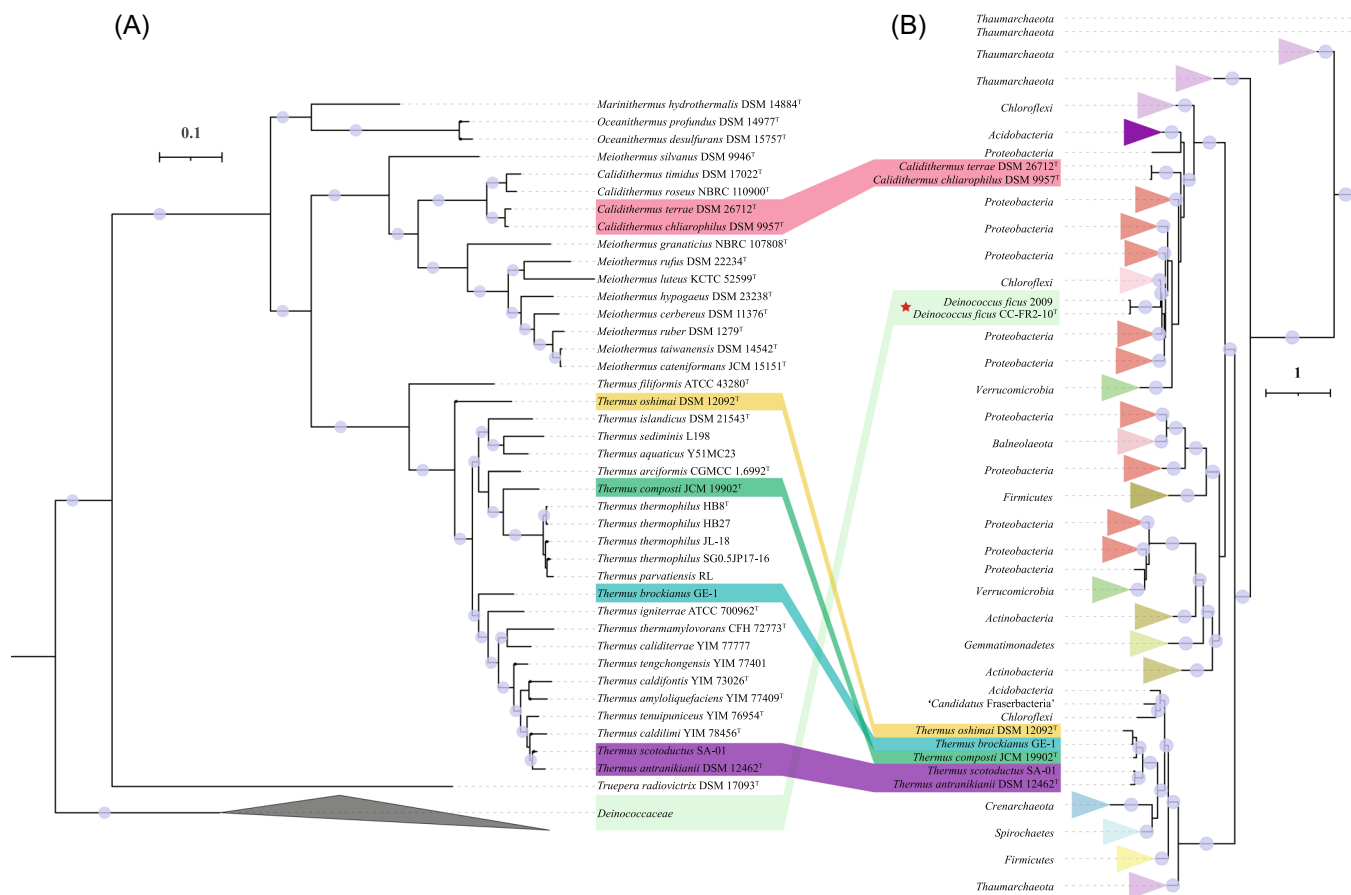


Figure 5. Discrepancy between the phylogenomic tree of *Deinococcota* and the phylogenetic tree of NirK. (A) The phylogenomic tree of *Deinococcota* was constructed as in Figure 1A. (B) The NirK protein sequences listed in Table S8 were aligned using MUSCLE with 100 iterations. The NirK tree was constructed using IQ-Tree with the parameters (-alrt 1000 -bb 1000 -nt AUTO). The best-fit model (WAG + F + R10) was determined by ModelFinder. Red star illustrates the strains from the family Deinococcaceae.

and none of them contain the *nosZ*. The phylogenetic analysis revealed distinct evolutionary histories of *narGHI*, *nirS*, and *norBC* versus *nirK*, and demonstrated that the last common ancestor of the *Thermales* acquired these genes early in their evolutionary history, and these genes were largely inherited vertically. Our results support the incomplete denitrification pathway in *Thermus* and suggest that this pathway allowed it to adapt to the low-oxygen or anoxic conditions that are common in geothermal environments, which also expand our knowledge of the evolution of the denitrification pathway in *Thermus* and support the importance of *Thermus* as a significant heterotrophic denitrifier in thermal environments.

MATERIALS AND METHODS

Genome sequencing, assembly, and annotation

The genomes of *T. tenuipunctus* YIM 76954^T, *T. sediminis* L198, *T. caliditerrae* YIM 77777, *T. tengchongensis* YIM 77401, *Thermus caldifontis* YIM 73026^T, *T. amyloliquefaciens* YIM 77409^T, *T. thermophilus* JL-18, *T. caldilimi* YIM 78456^T, and *T. thermamylovorans* CFH 72773^T were sequenced and

assembled as reported earlier^{2,17,27–32}. Other reference genomes of *Thermus* species were downloaded from the NCBI database (Table S3). The habitat of these strains was present in Table S3, and the completeness and contamination of each genome were calculated using CheckM³³. Putative protein-coding sequences (CDSs) of each genome were predicted using Prodigal³⁴ with the “-p single” parameter, and the CDSs were annotated against eggNOG, KEGG, and NCBI-nr databases using DIAMOND³⁵ by applying *E*-values < 1e-5. The predicted CDSs were also uploaded to KEGG Automatic Annotation Server (KAAS)³⁶ with “bidirectional best hit” and “for prokaryotes” parameters. The tRNAs and rRNAs were identified by tRNAscan-SE version 2.0.2³⁷ and RNAmmer version 1.2³⁸, respectively. Detection of the CRISPR-Cas system in the genomes of *Thermus* was done using the CRISPRCas-Finder tool (<https://crisprcas.i2bc.paris-saclay.fr>)³⁹. The ViroBLAST (<http://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php>)⁴⁰ was used to identify the targets of the spacers, while the nonsimilar spacers were analyzed using BLAST against the NT database. The ISSaga software (http://issaga.biotoul.fr/issaga_index.php)⁴¹ was used to determine the IS elements, and the IslandPath-DIMOB⁴² was used to

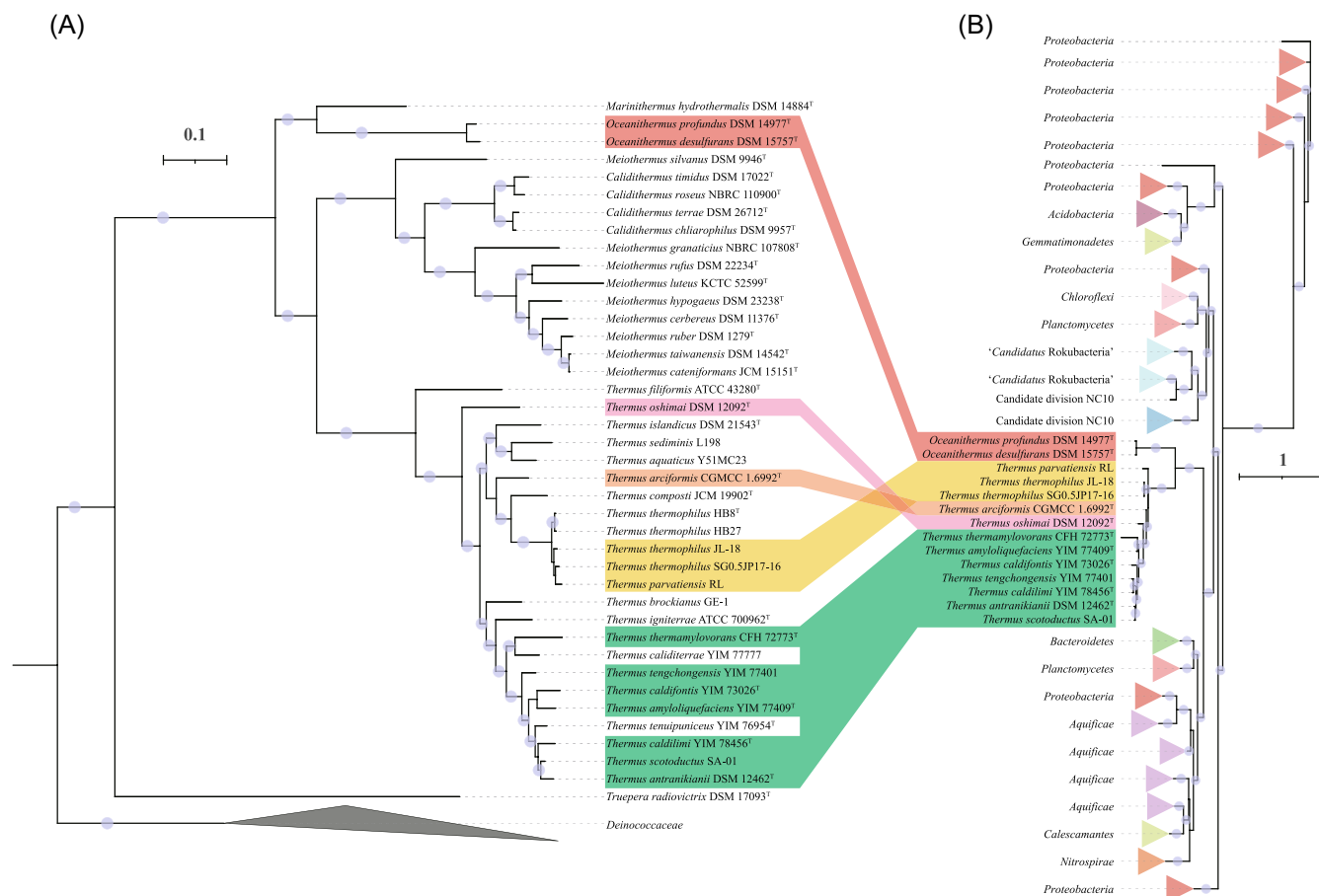


Figure 6. Consistency between the phylogenomic tree of *Deinococcota* and the phylogenetic tree of NirS. (A) The phylogenomic tree of *Deinococcota* was constructed as in Figure 1A. (B) The NirS protein sequences listed in Table S9 were aligned using MUSCLE with 100 iterations. The NirS tree was constructed using IQ-Tree with the parameters (-alrt 1000 -bb 1000 -nt AUTO). The best-fit model (LG + F + R10) was determined by ModelFinder.

predict GI, to provide the evidence of horizontal transfer regions.

Phylogenetic analysis

A reference genomic dataset of the genus *Thermus* was established by downloading type strain genomes from NCBI (Table S3). The genomes with estimated completeness >95% and contamination <5% were kept for phylogenetic analysis. The phylogeny of *Thermus* was generated as mentioned in our previous study^{43,44}. Briefly, GTDB-Tk software¹² was used to generate the multiple sequence alignments (MSAs) with 120 bacterial marker genes, and IQ-Tree⁴⁵ was used for calculating a maximum-likelihood phylogeny for MSAs with parameters (-alrt 1000 -bb 1000 -nt AUTO). The best-fit model (LG + F + R10) determined by ModelFinder⁴⁶ was chosen according to Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (Corrected AIC), and Bayesian Information Criterion (BIC).

For phylogeny based on genes of the denitrification pathway, datasets were derived from the NCBI reference genomes. The genomes were annotated as the methods above. The NarG protein sequences with length >1000 aa and <1300 aa were retained for this study. Reference NirK

protein datasets with sequences longer than 400 aa were kept for further analysis. Similarly, full-length NirS (>400 aa), NorB with the length of 450–850 aa, and NosZ with the average length of 662 aa were kept. Finally, representative sequences for each dataset (Tables S7–S11) were aligned using MUSCLE⁴⁷ with 100 iterations. Phylogenetic trees were constructed using IQ-Tree⁴⁵ with the parameters mentioned above. All trees were visualized and annotated using iTOL⁴⁸.

Gene content comparison

Pan- and core-genome analysis was performed using the Roary pipeline⁴⁹ at 40% minimum sequence identity. The ANI among the genomes of genus *Thermus* was calculated by using the pyANI⁵⁰ with BLAST method. Orthologous proteins were identified based on reciprocal BLAST best hits of predicted amino acid sequences (*E*-value < 1e-5), and AAI of each pair of genomes was calculated as the mean identity of all orthologous proteins. All plots were generated with R⁵¹ with the ggplot2 package⁵². A Bayesian tree based on MSAs was constructed with the parameters (ngen = 3000000 Nruns = 2 Nchains = 4 diagnfreq = 1000 relburnin = yes burnfrac = 0.25 samplefreq = 100 printfreq = 100)⁴³ by using MrBayes software¹⁹. The evolutionary history of the

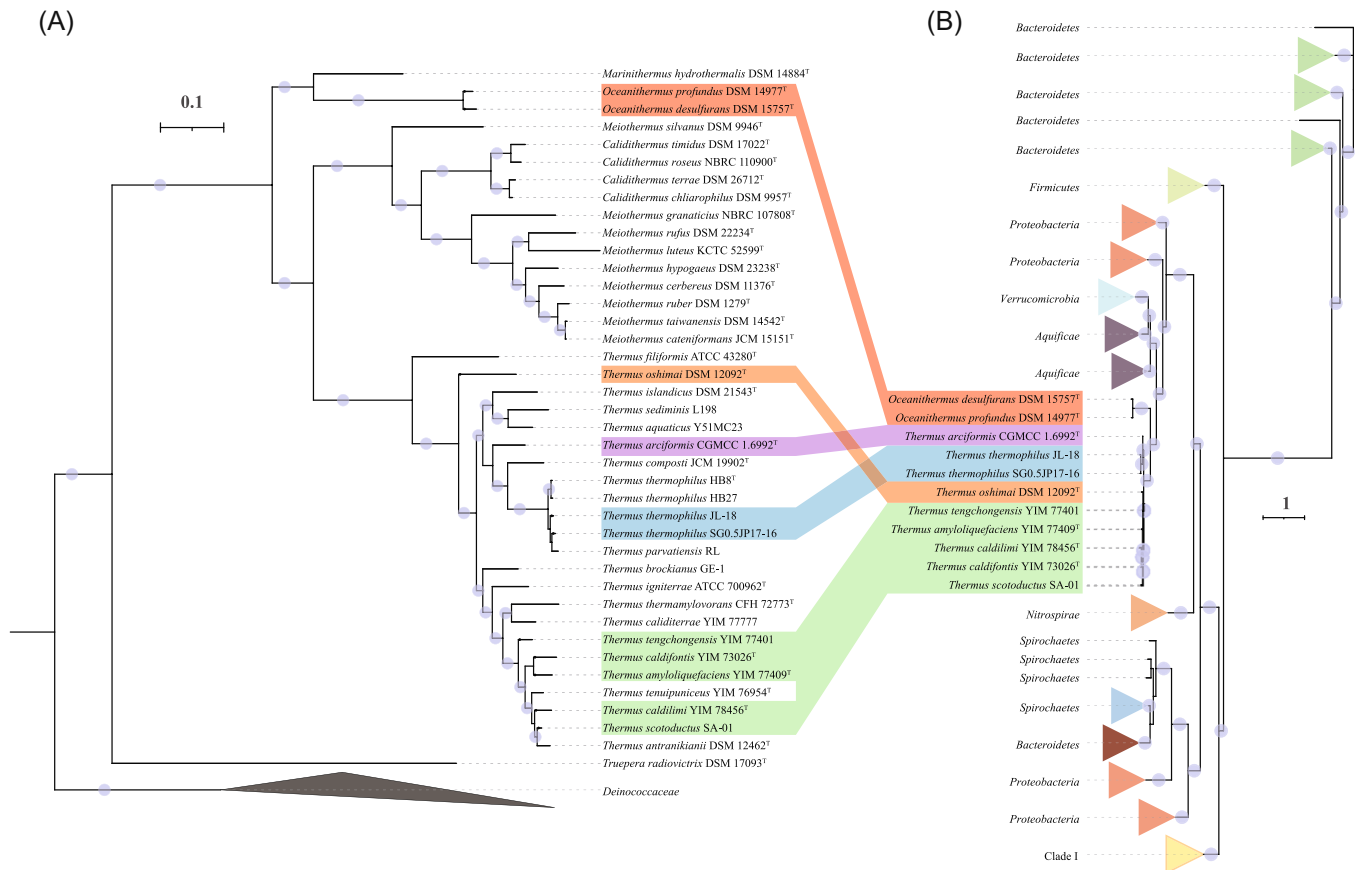


Figure 7. Consistency between the phylogenomic tree of *Deinococcota* and the phylogenetic tree of NorB. (A) The phylogenomic tree of *Deinococcota* was constructed as in Figure 1A. (B) The NorB protein sequences listed in Table S10 were aligned using MUSCLE with 100 iterations. The NorB tree was constructed using IQ-Tree with the parameters (-alrt 1000 -bb 1000 -nt AUTO). The best-fit model (LG + R10) was determined by ModelFinder.

phylum *Deinococcota* was inferred by COUNT²⁰ as described previously^{43,53}.

ACKNOWLEDGMENTS

We thank Xue-Qin Jia for collecting the data about the isolate information of *Deinococcota* strains. We are grateful to the staff from the Yunnan Tengchong Volcano and Spa Tourist Attraction Development Corporation for their assistance. This study was supported by funding from the National Natural Science Foundation of China (Nos. 91951205, 92051108, 31850410475, and 31970122), the National Science and Technology Fundamental Resources Investigation Program of China (2021FY100900), and the U.S. National Science Foundation (DEB 1557042 and DEB 1841658).

AUTHOR CONTRIBUTIONS

Wen-Jun Li and Brian P. Hedlund jointly conceived the study. Jian-Yu Jiao conceptualized the research goals under the supervision of Wen-Jun Li and Brian P. Hedlund, Jian-Yu Jiao, Zheng-Han Lian, Meng-Meng Li, En-Min Zhou, Lan Liu, and Hong Ming performed the bioinformatics analyses. Jian-Yu Jiao and Zheng-Han Lian prepared the figures and tables. Jian-Yu Jiao, Zheng-Han Lian, Meng-

Meng Li, Nimaichand Salam, Brian P. Hedlund, and Wen-Jun Li wrote the manuscript. All authors revised and approved the final manuscript.

ETHICS STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

DATA AVAILABILITY

All the genomes described in this article have been deposited in NCBI, and the accession numbers were provided in Table S3. The *Thermus* information can also be found in eLMSG (an eLibrary of Microbial Systematics and Genomics, <https://www.biosino.org/elmsg/index>) under accession numbers MSG069325, MSG066408, MSG019978, MSG066894, MSG067219, MSG066285, MSG068958, MSG065983, MSG068209, MSG065453, MSG067283, MSG007907. The R codes have been deposited in Github (<https://github.com/lianzhh-pub/Rcode-mLife2021>).

SUPPORTING INFORMATION

Additional Supporting Information for this article can be found online at <https://doi.org/10.1002/mlf2.12009>.

ORCID

Jian-Yu Jiao  <http://orcid.org/0000-0002-6900-2296>

Zheng-Han Lian  <http://orcid.org/0000-0002-0562-5849>

Meng-Meng Li  <http://orcid.org/0000-0003-1686-8122>

En-Min Zhou  <http://orcid.org/0000-0002-8253-9055>

Lan Liu  <http://orcid.org/0000-0001-8139-7246>

Brian P. Hedlund  <http://orcid.org/0000-0001-8530-0448>

Wen-Jun Li  <http://orcid.org/0000-0002-1233-736X>

REFERENCES

- Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol*. 2020;70:5607–12.
- Zhou EM, Murugapiran SK, Mefferd CC, Liu L, Xian WD, Yin YR, et al. High-quality draft genome sequence of the *Thermus amyloliquefaciens* type strain YIM 77409^T with an incomplete denitrification pathway. *Stand Genomic Sci*. 2016;11:1–9.
- Bruins ME, Janssen AE, Boom RM. Thermozyes and their applications. *Appl Biochem Biotechnol*. 2001;90:155–86.
- Hedlund BP, McDonald AI, Lam J, Dodsworth JA, Brown JR, Hungate BA. Potential role of *Thermus thermophilus* and *T. oshimai* in high rates of nitrous oxide (N₂O) production in ~80°C hot springs in the US Great Basin. *Geobiology*. 2011;9:471–80.
- Ramírez-Arcos S, Fernández-Herrero LA, Berenguer J. A thermophilic nitrate reductase is responsible for the strain specific anaerobic growth of *Thermus thermophilus* HB8. *Biochim Biophys Acta Gene Struct Expression*. 1998;1396:215–27.
- Dodsworth JA, Hungate BA, Hedlund BP. Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US Great Basin hot springs with abundant ammonia-oxidizing archaea. *Environ Microbiol*. 2011;13:2371–86.
- Brüggemann H, Chen CY. Comparative genomics of *Thermus thermophilus*: plasticity of the megaplasmid and its contribution to a thermophilic lifestyle. *J Biotechnol*. 2006;124:654–61.
- Lorenz MG, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev*. 1994;58:563–602.
- Kumwenda B, Litthauer D, Reva O. Analysis of genomic rearrangements, horizontal gene transfer and role of plasmids in the evolution of industrial important *Thermus* species. *BMC Genomics*. 2014;15:1–13.
- Tripathi C, Mishra H, Khurana H, Dwivedi V, Negi RK, Lal R. Complete genome analysis of *Thermus parvatiensis* and comparative genomics of *Thermus* spp. provide insights into genetic variability and evolution of natural competence as strategic survival attributes. *Front Microbiol*. 2017;8:1410.
- Hidaka Y, Hasegawa M, Nakahara T, Hoshino T. The entire population of *Thermus thermophilus* cells is always competent at any growth phase. *Biosci Biotechnol Biochem*. 1994;58:1338–9.
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2020;36:1925–7.
- Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, daCosta MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol*. 2018;68:461–6.
- Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *ISME J*. 2017;11:2399–406.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9:1–8.
- Sabath N, Ferrada E, Barve A, Wagner A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol Evol*. 2013;5:966–77.
- Zhou EM, Xian WD, Jiao JY, Liu L, Li MM, Ding YP, et al. Physiological and genomic properties of *Thermus tenuipunicus* sp. nov., a novel slight reddish color member isolated from a terrestrial geothermal spring. *Syst Appl Microbiol*. 2018;41:611–18.
- Kieft TL, Fredrickson JK, Onstott TC, Gorby YA, Kostandarithes HM, Bailey TJ, et al. Dissimilatory reduction of Fe (III) and other electron acceptors by a *Thermus* isolate. *Appl Environ Microbiol*. 1999;65:1214–21.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
- Csűös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*. 2010;26:1910–12.
- Liu RR, Tian Y, Zhou EM, Xiong MJ, Xiao M, Li WJ. Distinct expression of the two NO-forming nitrite reductases in *Thermus antranikianii* DSM 12462^T improved environmental adaptability. *Microb Ecol*. 2020;80:614–26.
- Dalsgaard T, Stewart FJ, Thamdrup B, De Brabandere L, Revsbech NP, Ulloa O, et al. Oxygen at nanomolar levels reversibly suppresses process rates and gene expression in anammox and denitrification in the oxygen minimum zone off northern Chile. *mBio*. 2014;5:e01966–14.
- Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, et al. A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. *PLoS One*. 2013;8:e53350.
- Holloway JM, Nordstrom DK, Böhlke JK, McCleskey RB, Ball JW. Ammonium in thermal waters of Yellowstone National Park: processes affecting speciation and isotope fractionation. *Geochim Cosmochim Acta*. 2011;75:4611–36.
- Ramírez-Arcos S, Fernández-Herrero LA, Marín I, Berenguer J. Anaerobic growth, a property horizontally transferred by an Hfr-like mechanism among extreme thermophiles. *J Bacteriol*. 1998;180:3137–43.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018;36:996–1004.
- Ming H, Zhao ZL, Ji WL, Ding CL, Cheng LJ, Niu MM, et al. *Thermus thermamylovorans* sp. nov., isolated from a hot spring. *Int J Syst Evol Microbiol*. 2020;70:1729–37.
- Zhou EM, Xian WD, Mefferd CC, Thomas SC, Adegboruwa AL, Williams N, et al. *Thermus sediminis* sp. nov., a thiosulfate-oxidizing and arsenate-reducing organism isolated from Little Hot Creek in the Long Valley Caldera, California. *Extremophiles*. 2018;22:983–91.
- Mefferd CC, Zhou EM, Yu TT, Ming H, Murugapiran SK, Huntemann M, et al. High-quality draft genomes from *Thermus caliditerrae* YIM 77777 and *T. tengchongensis* YIM 77401, isolates from Tengchong, China. *Genome Announc*. 2016;4:e00312–16.
- Khan IU, Habib N, Hussain F, Xian WD, Amin A, Zhou EM, et al. *Thermus caldifontis* sp. nov., a thermophilic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol*. 2017;67:2868–72.
- Murugapiran SK, Huntemann M, Wei CL, Han J, Detter JC, Han CS, et al. Whole genome sequencing of *Thermus oshimai* JL-2 and *Thermus thermophilus* JL-18, incomplete denitrifiers from the United States Great Basin. *Genome Announc*. 2013;1:e00106–12.
- Li MM, Xian WD, Zhang XT, Yin YR, Zhou EM, Ding YP, et al. *Thermus calidilimi* sp. nov., a thermophilic bacterium isolated from a geothermal area. *Antonie Van Leeuwenhoek*. 2019;112:1767–74.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55.
- Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform*. 2010;11:119.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35:182–5.

- 37 Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25:955–64.
- 38 Lagesen K, Hallin P, Rødland EA, Stærfeldt H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35:3100–08.
- 39 Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 2007;35:52–7.
- 40 Deng WJ, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics.* 2007;23:2334–6.
- 41 Varani AM, Siguier P, Goubeyre E, Charneau V, Chandler M. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* 2011;12:1–9.
- 42 Bertelli C, Brinkman FSL. Improved genomic island predictions with IslandPath-DIMOB. *Bioinformatics.* 2018;34:2161–7.
- 43 Jiao JY, Fu L, Hua ZS, Liu L, Salam N, Liu PF, et al. Insight into the function and evolution of the Wood–Ljungdahl pathway in *Actinobacteria*. *ISME J.* 2021;1–14.
- 44 Jiao JY, Liu L, Hua ZS, Fang BZ, Zhou EM, Salam N, et al. Microbial dark matter coming to light: challenges and opportunities. *Natl Sci Rev.* 2021;8:nwaa280.
- 45 Nguyen LT, Schmidt HA, Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
- 46 Kalyaanamoorthy S, Minh BQ, Wong T, Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–9.
- 47 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
- 48 Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44:242–5.
- 49 Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31:3691–3.
- 50 Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2016;8:12–24.
- 51 Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2013.
- 52 Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis. *Meas Interdiscip Res Perspect.* 2019;17:160–7.
- 53 Hua ZS, Wang YL, Evans PN, Qu YN, Goh KM, Rao YZ, et al. Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea. *Nat Commun.* 2019;10:1–11.

How to cite this article: Jiao J-Y, Lian Z-H, Li M-M, Salam N, Zhou E-M, Liu L, et al. Comparative genomic analysis of *Thermus* provides insights into the evolutionary history of an incomplete denitrification pathway. *mLife*. 2022;1:198–209. <https://doi.org/10.1002/mlf2.12009>