



Comparative Genomic Analysis Reveals Preserved Features in Organohalide-Respiring *Sulfurospirillum* Strains

Yi Yang,^a Torsten Schubert,^b Yan Lv,^{a,c} Xiuying Li,^a Jun Yan^a

^aKey Laboratory of Pollution Ecology and Environmental Engineering, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, Liaoning, China

^bResearch Group Anaerobic Microbiology, Institute of Microbiology, Friedrich Schiller University Jena, Jena, Germany

^cUniversity of Chinese Academy of Sciences, Beijing, China

ABSTRACT *Sulfurospirillum* species strains are frequently detected in various pristine and contaminated environments and participate in carbon, sulfur, nitrogen, and halogen elements cycling. Recently we obtained the complete genome sequences of two newly isolated *Sulfurospirillum* strains, ACS_{DCE} and ACS_{TCE}, capable of dechlorinating tetrachloroethene to *cis*-1,2-dichloroethene and trichloroethene under low-pH conditions, but a detailed analysis of these two genomes in reference to other *Sulfurospirillum* genomes for an improved understanding of *Sulfurospirillum* evolution and ecophysiology has not been accomplished. Here, we performed phylogenetic and pangenome analyses with 12 completed *Sulfurospirillum* genomes, including those of strain ACS_{TCE} and strain ACS_{DCE}, to unravel the evolutionary and metabolic potentials in the genus *Sulfurospirillum*. Based on 16S rRNA gene and whole-genome phylogenies, strains ACS_{TCE}, ACS_{DCE}, and JPD-1 could be clustered into a single species, proposed as “*Candidatus Sulfurospirillum acididehalogenans*.” TimeTree analysis suggested that the organohalide-respiring (OHR) *Sulfurospirillum* might acquire the ability to use chlorinated electron acceptors later than other energy conservation processes. Nevertheless, the ambiguity of the phylogenetic relations among *Sulfurospirillum* strains complicated the interpretation of acquisition and loss of metabolic traits. Interestingly, all OHR *Sulfurospirillum* genomes except the ones of *Sulfurospirillum multivorans* strains harbor a well-aligned and conserved region comprising the genetic components required for the organohalide respiration chain. Pangenome results further revealed that a total of 34,620 gene products, annotated from the 12 *Sulfurospirillum* genomes, can be classified into 4,118 homolog families and 2,075 singleton families. Various *Sulfurospirillum* species strains have conserved metabolisms as well as individual enzymes and biosynthesis capabilities. For instance, only the OHR *Sulfurospirillum* species strains possess the quinone-dependent pyruvate dehydrogenase (PoxB) gene, and only “*Ca. Sulfurospirillum acididehalogenans*” strains harbor urea transporter and urease genes. The plasmids found in strain ACS_{TCE} and strain ACS_{DCE} feature genes coding for type II toxin-antitoxin systems and transposases and are promising tools for the development of robust gene editing tools for *Sulfurospirillum*.

IMPORTANCE Organohalide-respiring bacteria (OHRB) play critical roles in the detoxification of chlorinated pollutants and bioremediation of subsurface environments (e.g., groundwater and sediment) impacted by anthropogenic chlorinated solvents. The majority of known OHRB cannot perform reductive dechlorination below neutral pH, hampering the applications of OHRB for remediating acidified groundwater due to fermentation and reductive dechlorination. Previously we isolated two *Sulfurospirillum* strains, ACS_{TCE} and ACS_{DCE}, capable of dechlorinating tetrachloroethene under acidic conditions (e.g., pH 5.5), and obtained the complete genomes of both strains. Notably, two plasmid sequences were identified in the genomes of strain ACS_{TCE} and strain ACS_{DCE} that may be conducive to unraveling the genetic modification mechanisms in the genus *Sulfurospirillum*. Our findings improve the current understanding of *Sulfurospirillum* species

Editor Timothy E. Mattes, University of Iowa

Copyright © 2022 Yang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yi Yang, yangyi@iae.ac.cn, or Jun Yan, junyan@iae.ac.cn.

The authors declare no conflict of interest.

Received 15 November 2021

Accepted 19 January 2022

Published 23 February 2022

strains regarding their biogeographic evolution, genome dynamics, and functional diversity. This study has applied values for the bioremediation of toxic and persistent organohalide pollutants in low-pH environments.

KEYWORDS *Sulfurospirillum*, comparative genomics, evolution, organohalide respiration

Members within the genus *Sulfurospirillum* are capable of versatile energy metabolism (e.g., nitrate reduction and organohalide respiration), enabling them to thrive in various pristine and contaminated environments (1–3). Consequently, *Sulfurospirillum* species strains are considered excellent candidates for biotechnological applications such as oil field souring control and bioremediation of sites impacted by chlorinated contaminants (4). For example, some *Sulfurospirillum* strains had demonstrated nitrate reduction to nitrite, which inhibits the growth of sulfate-reducing bacteria and consequently prevents the souring of oil reservoirs (5). A subset of *Sulfurospirillum* strains (e.g., *S. multivorans* strain DSM 12446, *Sulfurospirillum* sp. strain ACS_{TCE}, and *Sulfurospirillum* sp. strain ACS_{DCE}) can perform organohalide respiration (OHR) with tetrachloroethene (PCE) and trichloroethene (TCE) as electron acceptors, particularly at a pH as low as 5.5, suggesting that they play key roles for the natural attenuation of anthropogenic chlorinated solvents in acidic environments (2, 4, 6–8). In addition, some *Sulfurospirillum* strains are capable of synthesizing the unconventional norpseudo-type cobamide (i.e., vitamin B₁₂ derivative), which can be used to explore the impacts of cobamides on the activities of corrinoid-dependent enzymes and microbial community structures (9). In the contaminated environments, *Sulfurospirillum* strains were frequently found to coexist with obligate organohalide respiring bacteria (OHRB) phylotypes such as *Dehalococcoides mccartyi* and *Dehalobacter*, which might be explained by their ability to provide the required nutrients, such as cobamides and hydrogen, for these dechlorinators (4). The hypothesis that hydrogen produced by *Sulfurospirillum* via anaerobic oxidation of carboxylic acids enabling syntrophic growth with a hydrogenotrophic partner was later demonstrated in the coculture of *S. multivorans* strain DSM 12446 and *Methanococcus voltae* (10). Recently, the cross-feeding of hydrogen, acetate, and cobamides was reported during the cocultivation of *S. multivorans* strain DSM 12446 and *Dehalococcoides mccartyi* strain 195 (11). Based on these observations, *Sulfurospirillum* strains are considered suitable partner populations to explore OHR-related microbial ecology (9–11). Great efforts have been made to investigate the environmental distribution of the genus *Sulfurospirillum*; however, the physiological traits, ecological relevance, and evolutionary history of *Sulfurospirillum* strains are not well understood.

Due to the lack of efficient genetic manipulation systems and difficulties in studying OHR *Sulfurospirillum* using classical biochemical techniques, bioinformatic analysis based on metagenome-assembled genomes (MAGs) and complete genomes are alternative methods to investigate the physiology and metabolic potentials of geographically distributed *Sulfurospirillum* (12). For instance, a genome-wide comparison of *Sulfurospirillum* strains indicated that the non-OHR *Sulfurospirillum barnesii* strain SES-3, which was isolated from anoxic sediment contaminated with arsenate and selenate, uniquely harbors a rare hybrid gene cluster encoding polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS). Based on this genome-to-metabolite approach, the first anoxically biosynthesized NRPS-PKS-derived natural product, barnesin A, was identified by omics-guided isolation and total synthesis (13). Additionally, Goris et al. conducted a thorough comparative genomic analysis of *S. multivorans* strain DSM 12446 with closely related non-OHR *S. deleyianum* strain DSM 6946 and *S. barnesii* strain SES-3, which revealed the presence of an ~50-kbp region containing genes required for OHR and cobamide cofactor biosynthesis and horizontally acquired genes, enabling the catabolic flexibility in *S. multivorans* strain DSM 12446 (14). The genome of *S. cavolei* strain MES, assembled from the metagenomic sequences of an electrosynthetic microbiome, was compared with other 10 complete or draft *Sulfurospirillum* genomes featuring conserved and divergent

physiologies and metabolisms. This pangenome analysis revealed a total of 6,264 homolog families, including 1,082 homolog families shared among all 11 *Sulfurospirillum* genomes (i.e., core clusters), 1,991 homolog families shared among part of 11 genomes (i.e., accessory clusters), and 3,191 singleton families (i.e., unique clusters), indicating the commonalities in general functions as well as ecological pressures induced acquisition of unique gene sets (15). Combined with transcriptomics data, a recent *Sulfurospirillum* comparative genomics study on the region encoding OHR genetic components identified a two-component regulator that is responsible for PCE-induced gene expression in OHR *Sulfurospirillum* strains (16).

To date, at least 47 complete or draft *Sulfurospirillum* genomes are publicly available (www.ncbi.nlm.nih.gov/genome). Two novel *Sulfurospirillum* organisms, strain SL2-1 and strain SL2-2, which performed PCE-to-TCE and PCE-to-*cis*-1,2-dichloroethene (cDCE) dechlorination, respectively, were recently enriched from a PCE-dechlorinating consortium maintained for 10 years (8). The assembled genomes of strain SL2-1 and strain SL2-2 are highly identical, representing a new *Sulfurospirillum* species proposed as "*Candidatus Sulfurospirillum diekertiae*" (8). The increased numbers of *Sulfurospirillum* genomes require an updated genomic examination and offer the opportunity for comprehensive comparison of OHR and non-OHR *Sulfurospirillum*. Such efforts are promising to provide new insights into the physiological traits, metabolic potentials, and evolution characteristics in the genus *Sulfurospirillum*.

In this study, we performed comparative pangenome analysis on 12 complete *Sulfurospirillum* genomes, including those of two new *Sulfurospirillum* isolates, strain ACS_{DCE} and strain ACS_{TCEr} capable of dechlorinating PCE to TCE and cDCE, respectively, under low-pH conditions (e.g., pH 5.5). We found that a well-aligned and conserved region comprising the genetic components required for the organohalide respiration chain is present in all OHR *Sulfurospirillum* genomes except the ones in *S. multivorans* strains. Genomic differences between non-OHR and OHR *Sulfurospirillum* strains as well as variations among OHR *Sulfurospirillum* strains were observed and discussed. Findings of this study will advance our understanding of members of the genus *Sulfurospirillum* regarding their evolutionary traits, genome dynamics, and functional diversity.

RESULTS AND DISCUSSION

Proposition of "*Candidatus Sulfurospirillum acididehalogenans*" as a new *Sulfurospirillum* species. Multiple tools (e.g., JSpeciesWS, TYGS, and GTDB) and analyses (e.g., 16S rRNA genes and whole-genome sequences) were performed to reveal the phylogenetic placement of strain ACS_{DCE} and strain ACS_{TCE} in reference to other *Sulfurospirillum* species strains. Pairwise comparison of 16S rRNA gene sequences demonstrated that strains ACS_{DCEr}, ACS_{TCEr}, and JPD-1, which share 99.8 to 99.9% identities to each other, are clustered into a distinct subclade with 90.9% to 98.9% identities to other *Sulfurospirillum* species strains (Fig. 1A). Phylogenetic inference with complete genome sequences further demonstrated that strains ACS_{DCEr}, ACS_{TCEr}, and JPD-1 can be placed into a single cluster with "*Candidatus Sulfurospirillum diekertiae*" strain SL2-1 and strain SL2-2 as the closest relatives (Fig. 1B; see also Fig. S2 in the supplemental material). Pairwise comparison of genome sequences performed with TYGS found that strain ACS_{DCE} and strain ACS_{TCE} shared 99.3 to 99.7% dDDH (digital DNA-DNA hybridization) based on three different GBDP formulas (Table S2); by comparison, strain JPD-1 shared 69.3% and 84.7% dDDH with strain ACS_{DCE} and 69.3% and 83.4% dDDH with strain ACS_{TCE} (Table S2). ANIm, ANIb, and orthoANI analyses by JSpeciesWS and orthoANI demonstrated that the calculated ANI values for each pair of strains ACS_{TCEr}, ACS_{DCEr}, and JPD-1 were above the 95% threshold for species delineation (Tables S3 and S4, Fig. S3). Based on these results, we proposed to unify strains ACS_{DCEr}, ACS_{TCEr}, and JPD-1 into a new *Sulfurospirillum* species, designated "*Candidatus Sulfurospirillum acididehalogenans*."

OHR *Sulfurospirillum* species diverged from non-OHR *Sulfurospirillum* recently. Genetic processes (e.g., mutation, horizontal gene transfer) and extreme geological

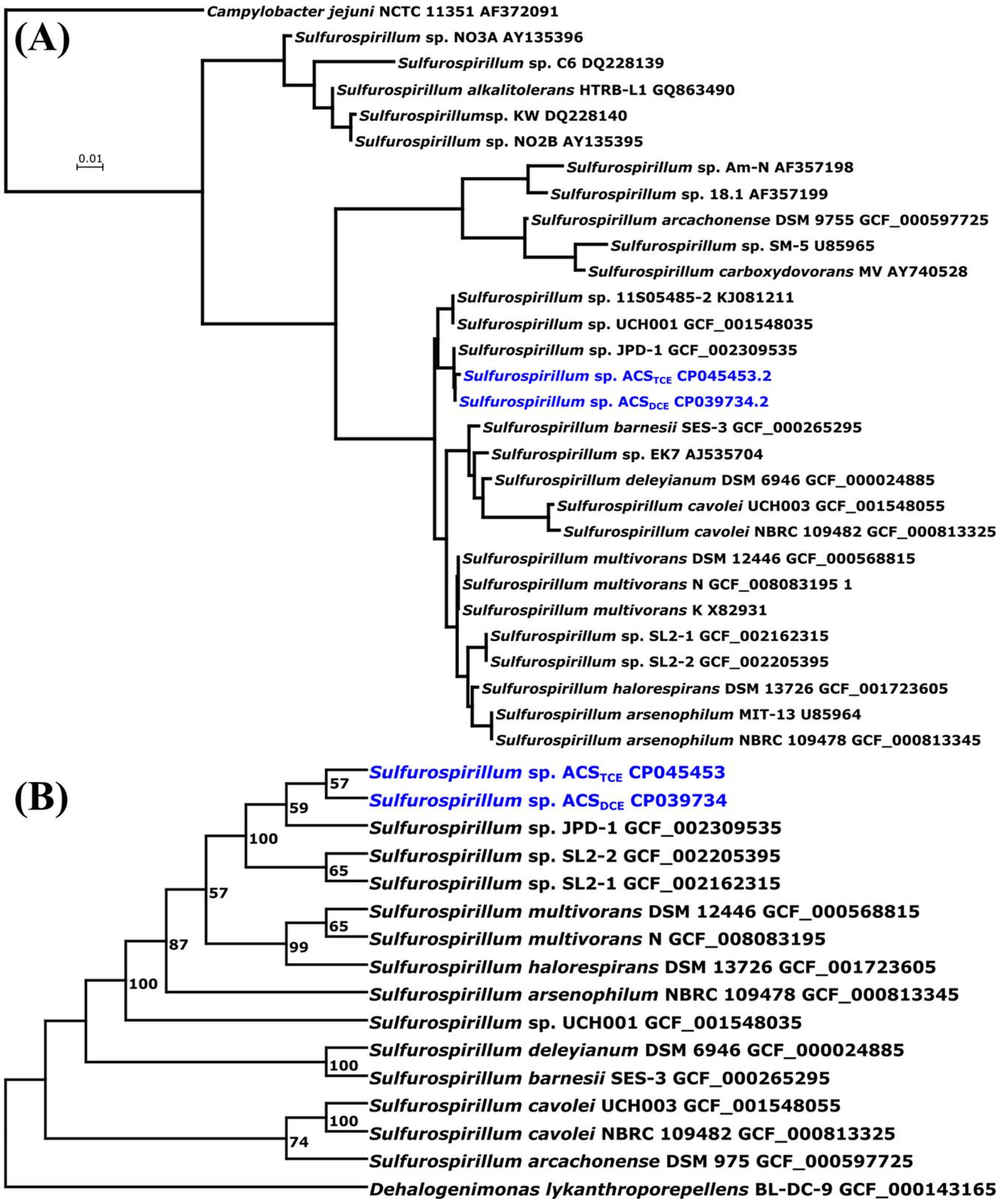


FIG 1 (A) 16S rRNA gene phylogenetic tree of *Sulfurospirillum* species strains. (B) The genome-based phylogenetic tree was inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches in panel B are GBDP pseudobootstrap support values of >60% from 100 replications, with an average branch support of 81.7%. Strain ACS_{TCE} and strain ACS_{DCE} are colored blue.

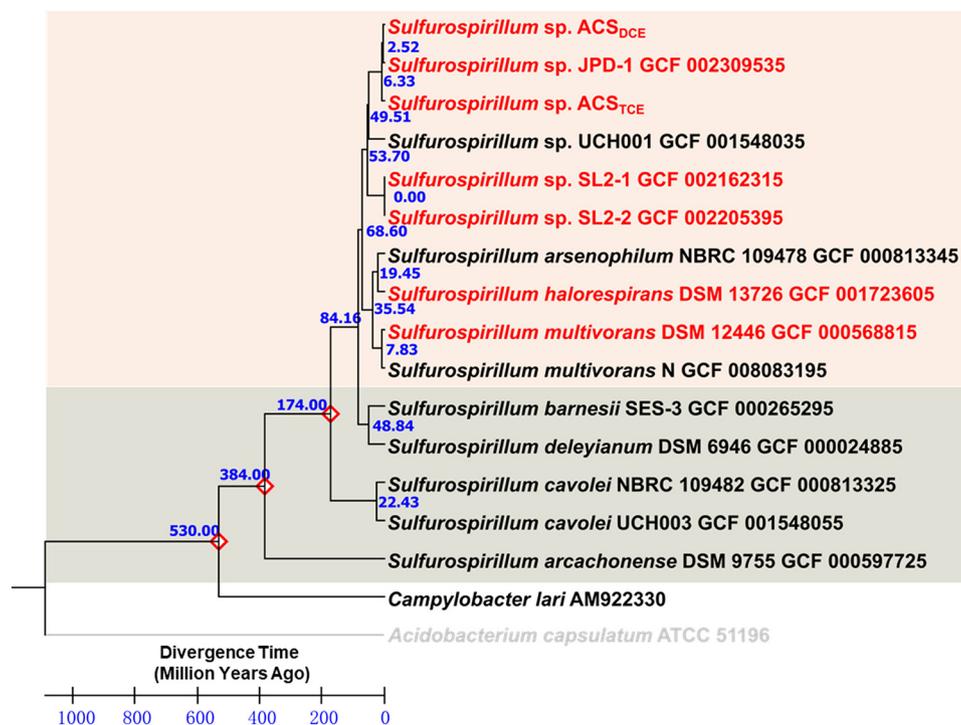


FIG 2 Phylogeny and molecular clock of selected 16 rRNA genes from *Sulfurospirillum* species strains and *Campylobacter lari* from *Proteobacteria*. *Acidobacterium capsulatum* ATCC 51196 was chosen as the outgroup. Branch lengths represent the divergence times (MYA) approximated by the RelTime method using Mega X. The OHR and non-OHR clusters of *Sulfurospirillum* species strains were colored in light orange and gray, respectively.

events (e.g., great oxygenation event, neoproterozoic oxygenation event) can affect the long-term evolution of microorganisms; however, our understanding of microbial evolution is limited by the lack of geological and biological evidence (e.g., fossils, genetics) through the geological time scale. We applied the TimeTree analysis with the RelTime method to evaluate the evolution time frame of several newly sequenced *Sulfurospirillum* species strains by following the well-established protocols (17, 18). The TimeTree analysis suggested that the ancestor of the new species "*Candidatus Sulfurospirillum acididehalogenans*" emerge approximately between the Eocene and Miocene (e.g., between 49.41 and 6.33 million years ago [MYA]) (Fig. 2), which was much later than the estimated appearance time (e.g., the Neoproterozoic era, 1,000 to 541 MYA) of obligate organohalide respiring *Dehalococcoidia* (18). The TimeTree analysis is based on 16S rRNA gene sequences and intended to infer the divergence times of strains within the genus *Sulfurospirillum*. Such an approach can only provide a rough estimate; however, this effort reflected our interests in unresolved questions, including (i) how does the organohalide respiration metabolism evolve, (ii) when does this energy conservation emerge in geoscale time, and (iii) when were the organohalide-respiratory genes transferred into microorganisms of different genera (e.g., *Sulfurospirillum*). To the best of our knowledge, no solid evidence is available to predict or support when an insertion event (e.g., the organohalide respiration region) in the *Sulfurospirillum* genome occurs, since dating the hypothetical insertion event is still difficult. Nevertheless, the dechlorinating *Sulfurospirillum* species strains could not obtain organohalide respiration genes before their own existence; therefore, the insertion event or organohalide respiration genes horizontally transferred occurred roughly at the same time or later than the divergence time of *Sulfurospirillum* species strains.

Furthermore, a previous genomic comparison study based on the genome sequence of *Sulfurospirillum multivorans* discovered a 50-kbp organohalide-respiring region (14), which was also identified in the genome of another species, *Sulfurospirillum halorespirans*

(6), indicating the ability to respire halogenated organic compounds was horizontally acquired by some ancestors of the genus *Sulfurospirillum* recently. Ancestors of the OHR *Sulfurospirillum* strains may diverge from the non-OHR *Sulfurospirillum* strains via horizontal acquisition of the genetic components required for organohalide respiration chain and *de novo* cobamide biosynthesis, and OHR *Sulfurospirillum* strains were subsequently distributed to different niches on the planet Earth, which could be inferred by the observation that different OHR *Sulfurospirillum* strains with conserved arrangements of the gene cluster responsible for OHR were isolated from geographically distinct origins. Nonetheless, the non-OHR *Sulfurospirillum* sp. strain UCH001, which was isolated from chlorinated ethene-contaminated groundwater in Japan but could not dechlorinate chlorinated ethenes (19), was closely related to the OHR *Sulfurospirillum* strains. Such an inconsistency may be due to the loss of gene clusters responsible for organohalide respiration and requires further investigation.

Shared features in the chromosomes of strain ACS_{DCE} and strain ACS_{TCE}. The chromosome of strain ACS_{TCE} has 2,998 features, including 2,853 coding genes, 3 CRISPR (clustered regularly interspaced short palindromic repeats) arrays, 26 CRISPR repeats, 23 CRISPR spacers, and 50 noncoding RNAs. Only 1,338 coding sequences could be assigned with a SEED (<https://pubseed.theseed.org>) annotation ontology across 910 distinct SEED functions. By comparison, the chromosome of strain ACS_{DCE} harbors 2,993 features, including 2,852 coding genes, 3 CRISPR arrays, 26 CRISPR repeats, 23 CRISPR spacers, and 55 noncoding RNAs. Only a total of 1,351 coding sequences could be assigned with a SEED annotation ontology across 911 distinct SEED functions. Functioning-based metabolic reconstruction comparison between strain ACS_{DCE} and strain ACS_{TCE} suggested that both strains share a total of 956 functioning roles defined within the SEED subsystems (e.g., biotin biosynthesis, cobamide synthesis, coenzyme A biosynthesis, and heme and siroheme biosynthesis) (Table S5). BLAST of the annotated coding sequences showed that the homologs of 70 genes in the chromosome of strain ACS_{DCE} were not found in the chromosome of strain ACS_{TCE}, including those genes encoding hydroxylamine reductase, arginyl-tRNA protein transferase, DNA helicase restriction/modification system component YeeBC, and 50 hypothetical genes with unknown functions (Table S6). Likewise, 23 coding genes in the chromosome of strain ACS_{TCE} were not matched in the chromosome of strain ACS_{DCE}, such as genes encoding cytochrome *c*₅₅₂ precursor, anaerobic C₄-dicarboxylate transporter DcuA, heterodisulfide reductase subunit B-like protein/putative succinate dehydrogenase subunit, and other 10 hypothetical proteins (Table S7). Genes encoding tetrathionate reductase, nitrate reductase, formate dehydrogenase, and [NiFe] hydrogenase were present in both chromosomes of strains ACS_{TCE} and strain ACS_{DCE}. However, no genes related to nitrite ammonification were identified, suggesting that they can reduce nitrate as an electron acceptor and produce nitrite only. In pure cultures, strain ACS_{TCE} and strain ACS_{DCE} could transform up to 60% of an initial 5 mM nitrate to nitrite (Fig. S4 and S5). Whether the accumulation of nitrite from nitrate transformation inhibits the activities of *Sulfurospirillum* and whether strains ACS_{DCE} and ACS_{TCE} can transform nitrite to ammonia require further investigation.

Potential roles of plasmids in strain ACS_{DCE} and strain ACS_{TCE}. Except for strain ACS_{TCE} and strain ACS_{DCE}, plasmids are not found in *Sulfurospirillum* species strains, presumably due to the absence of plasmid or difficulty in faithful assembly and identification of plasmid solely using Illumina short reads (20). In contrast, circular plasmids with sizes of 38,046 bp and 39,868 bp were assembled from the genome sequencing data of strain ACS_{TCE} and strain ACS_{DCE}, respectively, using PacBio long reads and Illumina short reads (21, 22). Plasmid sequence alignment by MAUVE demonstrated that the synteny of the two plasmid sequences was not conserved, and only two regions with a total size of 9.5 kbp could be mapped between the two plasmid sequences. Plasmids of strain ACS_{TCE} and strain ACS_{DCE} carried 94 and 57 coding sequences, of which only 19 and 20 encoded hypothetical proteins, respectively. By comparison, the eggNOG-mapper pipeline found that 37 coding sequences of strain ACS_{TCE} plasmid and 32 of strain ACS_{DCE} plasmid (Table S8 and S9) matched with orthologous groups in the

eggNOG database. While most of the orthologous genes related to the coding sequences in the plasmids of strain ACS_{TCE} and strain ACS_{DCE} were found in Proteobacteria, a few coding sequences (e.g., *paraA* and *mcrB*) present in these two plasmids were closely related to those found in *Clostridia* of Firmicutes and Bacteroidetes (Tables S8 and S9). Therefore, we hypothesized that these two plasmids originate from Proteobacteria with additional coding sequences horizontally acquired from other phyla (e.g., Firmicutes). Two prophage-related regions with lengths of 20,367 bp and 26,549 bp were found in the plasmids of strain ACS_{TCE} and strain ACS_{DCE}, respectively. In addition, the plasmid and chromosomal sequences of strain ACS_{TCE} shared an identical repeat region with an approximate size of 1 kbp and was in strain ACS_{DCE}, indicating the exchange of genetic materials between the plasmid and the chromosome in both strains. Sequences encoding components (e.g., a toxic protein and its cognate antitoxin protein) of various type II toxin-antitoxin (TA) systems (e.g., HicA-HicB, YafQ-DinJ, and RelE-RelB) are both present in the plasmids of the strain ACS_{TCE} and strain ACS_{DCE}. Type II TA systems have been proposed to play roles in genome stabilization, abortive phage infection, stress modulation, and antibiotic persistence (23); however, how such a system is related to the survival of strain ACS_{TCE} and strain ACS_{DCE} is unknown. Four and three genes coding for transposases were annotated from the plasmids of strain ACS_{TCE} and strain ACS_{DCE}, respectively. A phylogenetic analysis on transposases annotated from *Sulfurospirillum* genomes indicated that five transposases from the plasmid sequences were clustered within the IS30/IS982 family. One IS21 family transposase on the plasmid of strain ACS_{DCE} was clustered with the transposases annotated on the chromosomes of strain ACS_{DCE} and strain ACS_{TCE} (Fig. S6). Restriction sites for restriction endonucleases such as BamHI, BglII, EcoRI, PvuI, and Sall were identified in both plasmids, providing retrospective *in silico* evidence of the movement of genomic sequences between the plasmid and chromosome in strain ACS_{TCE} and strain ACS_{DCE}. Overall, these results improve our understanding of the genomic characteristics of *Sulfurospirillum* species strains and are promising for the development of molecular tools for editing *Sulfurospirillum* genomes.

Pangenome analysis revealed conserved and differed features in OHR *Sulfurospirillum*. A total of 12 complete *Sulfurospirillum* genomes were selected for genomic comparison analysis to unravel the core functions and core protein families using the OrthoMCL tool, including seven experimentally verified OHR *Sulfurospirillum* genomes (i.e., strains ACS_{DCE}, ACS_{TCE}, JPD-1, SL2-1, SL2-2, *S. multivorans* DSM 12446, and *S. halorespirans* DSM 13726) and four non-OHR *Sulfurospirillum* genomes (i.e., strains UCH001, UCH003, *S. deleyianum* DSM 6946, and *S. barnesii* SES-3). *S. multivorans* strain N was demonstrated to be incapable of dechlorinating chlorinated ethenes despite its genome harbors two homologous reductive dehalogenase genes. For pangenome analysis, we grouped strain N with other OHR *Sulfurospirillum* species strains, since strain N may have lost its OHR capability recently by a transposition event (16). Generally, all examined *Sulfurospirillum* species strains shared a variety of conserved sequence regions (Fig. 3). Furthermore, whole-sequence alignment of all OHR *Sulfurospirillum* species strains by Mauve demonstrated that all of them except *S. multivorans* strain DSM 12446 and *S. multivorans* strain N have a larger well-aligned and conserved regions (i.e., 330 kbp) containing a previously proposed 50-kb OHR region (Fig. S7). A 148-kbp block that was not found in other OHR *Sulfurospirillum* genomes was inserted into the end of the OHR region only in the genomes of *S. multivorans* strain DSM 12446 and *S. multivorans* strain N. The OHR regions of eight OHR strains were estimated to be between 63 kbp (e.g., *S. halorespirans*) and 74 kbp (e.g., *S. multivorans*). A total of 34,620 coding sequences (i.e., 32,545 homolog gene sequences and 2,075 singleton gene sequences), annotated from 12 *Sulfurospirillum* genomes, was classified into 6,193 families (i.e., 4,118 homolog families and 2,075 singleton families). The average numbers of gene sequences, genes in homologs, genes in singletons, and homolog families for the total of 12 *Sulfurospirillum* genomes were 2,885, 2,712, 173, and 2,625, respectively. By comparison, the average numbers of gene sequences, genes in homologs, and homolog families, but not genes in singletons, found in the eight OHR *Sulfurospirillum* genomes were larger than those of the four non-OHR *Sulfurospirillum*

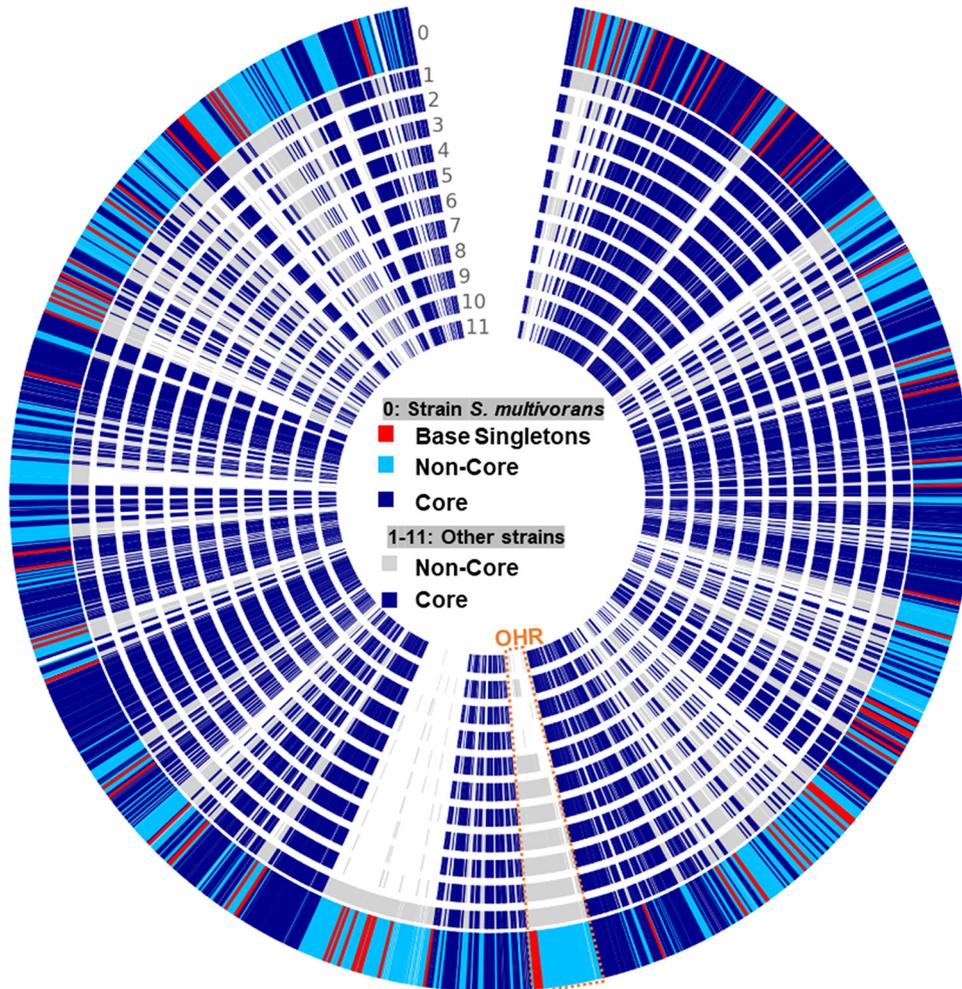


FIG 3 Circle view of pangenome with *S. multivorans* strain DSM 12446 as the base genome. The order of the genes in the rest of the pangenome is aligned to the position of its ortholog in the base genome. Genomes 1 to 11 are the following: 1, strain N; 2, *S. haloespirans*; 3, strain SL2-1; 4, strain SL2-2; 5, strain JPD-1; 6, strain ACS_{DCE}; 7, strain ACS_{TCE}; 8, strain UCH003; 9, strain UCH001; 10, strain SES-3; 11, *S. deleyianum*. The OHR colored in orange stands for the organohalide respiration region.

genomes, probably because all OHR *Sulfurospirillum* species strains have larger genome sizes than those of non-OHR *Sulfurospirillum* species strains by horizontally acquiring the genomic regions responsible for OHR (Table S1). The shared homolog families for each pair of *Sulfurospirillum* genomes were summarized in Table 1.

A total of 63 homolog families was only found in the genomes of OHR *Sulfurospirillum* species strains, including homologous genes encoding reductive dehalogenase, quinone-dependent pyruvate dehydrogenase (i.e., PoxB), NADPH-dependent FMN reductase, short-chain dehydrogenases/reductase, transcriptional regulator AraC family, Ser-tRNA(Ala) deacylase/Gly-tRNA(Ala) deacylase, phosphinothricin N-acetyltransferase, proteins associated with cobamide transport and biosynthesis (e.g., vitamin B₁₂-ABC transporter permease component BtuC, B₁₂-binding component BtuF, uroporphyrinogen-III methyltransferase/uroporphyrinogen-III synthase, and cobalt-precorrin-6A reductase), and a cluster related to propanoate metabolism (e.g., acetyl-coenzyme A synthetase, methylisocitrate lyase, 2-methylcitrate synthase, and 2-methylcitrate dehydratase). The previously reported 50-kbp gene region (e.g., 54 to 61 coding sequences starting from the gene representing carboxymuconolactone decarboxylase family protein) containing reductive dehalogenase genes and the gene cluster coding for (nor)cobamide biosynthesis from uroporphyrinogen III (14, 24) were conserved in all eight OHR *Sulfurospirillum* genomes (Fig. 4) but not present in

TABLE 1 Pairwise comparisons of 12 *Sulfurospirillum* genomes for the shared homolog families

Homolog family	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
G1, <i>S. deleyianum</i>	2,085	1,847	1,849	1,747	1,776	1,781	1,794	1,794	1,773	1,746	1,854	1,845
G2, strain N	1,847	3,146	2,050	2,089	2,133	2,248	2,309	2,308	2,227	2,187	2,534	3,140
G3, strain SES-3	1,849	2,050	2,255	1,812	1,847	1,862	1,906	1,903	1,827	1,813	2,001	2,050
G4, strain UCH001	1,747	2,089	1,812	2,307	2,030	2,062	2,104	2,102	2,070	2,052	2,113	2,088
G5, strain UCH003	1,776	2,133	1,847	2,030	2,389	2,096	2,109	2,107	2,074	2,057	2,154	2,133
G6, strain JPD-1	1,781	2,248	1,862	2,062	2,096	2,632	2,485	2,484	2,427	2,376	2,303	2,243
G7, strain SL2-1	1,794	2,309	1,906	2,104	2,109	2,485	2,797	2,779	2,394	2,351	2,347	2,306
G8, strain SL2-2	1,794	2,308	1,903	2,102	2,107	2,484	2,779	2,791	2,391	2,351	2,343	2,305
G9, strain ACS _{DCE}	1,773	2,227	1,827	2,070	2,074	2,427	2,394	2,391	2,628	2,565	2,253	2,224
G10, strain ACS _{TCE}	1,746	2,187	1,813	2,052	2,057	2,376	2,351	2,351	2,565	2,578	2,214	2,187
G11, <i>S. halorespirans</i>	1,854	2,534	2,001	2,113	2,154	2,303	2,347	2,343	2,253	2,214	2,750	2,531
G12, <i>S. multivorans</i>	1,845	3,140	2,050	2,088	2,133	2,243	2,306	2,305	2,224	2,187	2,531	3,141

the non-OHR *Sulfurospirillum* genomes. The only exception is that *S. barnesii* strain SES-3 of the non-OHR *Sulfurospirillum* group has several genes encoding *de novo* cobamide biosynthesis (Fig. 3) (14). All eight OHR *Sulfurospirillum* species strains and the non-OHR *S. cavolei* strain UCH003 possess the complete gene set for nitrogen fixation. All examined *Sulfurospirillum* species strains except *S. barnesii* strain SES-3 and *S. deleyianum* strain DSM 6946 possess the complete gene set for tetrathionate reductase.

The eight OHR *Sulfurospirillum* genomes contain a total of 4,113 gene clusters (Fig. 5). Genomes of “*Candidatus Sulfurospirillum acididehalogenans*” strains ACS_{TCE}, ACS_{DCE}, and JPD-1 shared 59 clusters consisting of 179 coding sequences (Fig. 5). Notably,

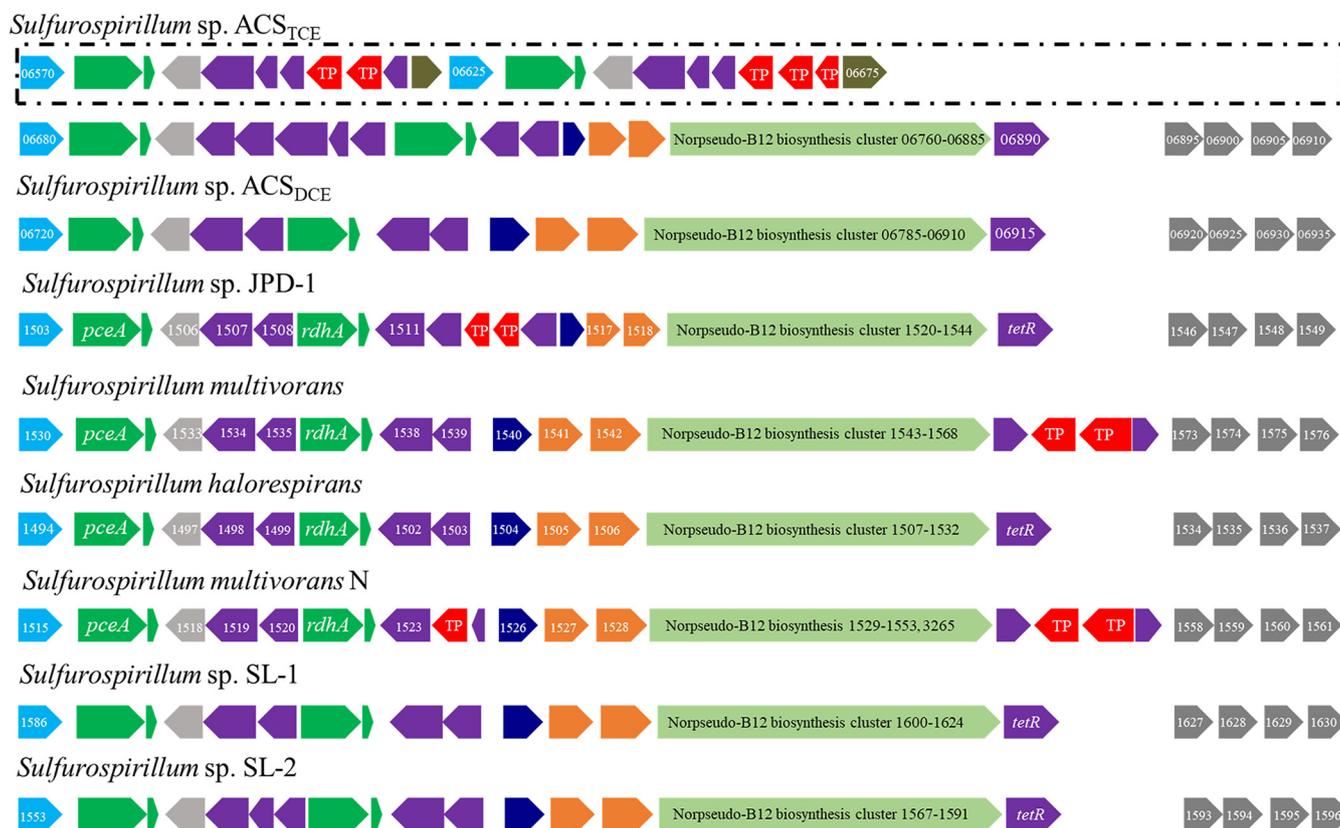


FIG 4 Illustration of the preserved OHR-related gene regions in the OHR *Sulfurospirillum* species strains. These regions encode carboxymuconolactone decarboxylase family protein (light blue), followed by the reductive dehalogenase with associated putative membrane anchor protein (green), iron-sulfur cluster assembly scaffold protein (light gray), sensor histidine kinase and response regulator transcription factor (purple), DUF4405 domain-containing protein (dark blue), components of a putative quinol dehydrogenase (orange), norcobamide biosynthesis clusters (light green), and transposase (red). The OHR region of strain ACS_{TCE} possesses two additional reductive dehalogenase gene clusters presented in the dashed box.

"*Candidatus Sulfurospirillum acididehalogenans*" strains ACS_{DCE}, ACS_{TCE}, and JPD-1 have a 14-genes cluster encoding urea transporter UrtABCDEFGF, urease, two-component sensor histidine kinase, and a hybrid sensor histidine kinase/response regulator. The presence of urease is hypothesized to allow strains ACS_{DCE}, ACS_{TCE}, and JPD-1 to tolerate low-pH conditions by yielding ammonia to neutralize protons (25); however, *S. multivorans* also can grow and dechlorinate PCE under acidic conditions (i.e., pH 5.5) and does not possess the urease gene cluster. "*Candidatus Sulfurospirillum acididehalogenans*" and *S. multivorans* probably could possess additional acid tolerance mechanisms, such as involvement of F₀F₁-ATPase in pH homeostasis, amino acid-dependent decarboxylase/antiporter systems, and deiminase and deaminase systems. Genes encoding F₀F₁-ATP synthase and various decarboxylase (e.g., arginine decarboxylase, aspartate decarboxylase) and agmatine deiminase family proteins were present in the genomes of "*Candidatus Sulfurospirillum acididehalogenans*" and *S. multivorans*. Comparatively, a total of 246 coding sequences from all OHR *Sulfurospirillum* species strains except "*Candidatus Sulfurospirillum acididehalogenans*" strains ACS_{DCE}, ACS_{TCE}, and JPD-1 were classified into 49 clusters, including genes for the L-proline glycine betaine ABC transport system (i.e., ProVWX), arsenite oxidase, 2-oxoglutarate/malate translocator, ferric iron ABC transporter, respiratory arsenate reductase Mo binding, and FeS subunits ArrAB. Strain ACS_{TCE} and strain ACS_{DCE} shared 152 orthologous gene clusters containing 310 coding sequences, most of which code for CRISPR-associated proteins, transposases, and mobile element proteins. Two adjacent genes annotated as carbon monoxide dehydrogenases, CooS, and carbon monoxide dehydrogenase accessory protein, CooC, were present in the genomes of strain ACS_{DCE} and strain ACS_{TCE} but not in other *Sulfurospirillum* genomes. By comparison, all the OHR *Sulfurospirillum* species strains, except strain ACS_{DCE} and strain ACS_{TCE}, harbor genes encoding polysulfide reductase NrfD, potassium-transporting ATPase, nitrous oxide reductase NosZ, and nitrous oxide reductase maturation protein NosL.

Differences between PCE-to-cDCE and PCE-to-TCE dechlorinating reductive dehalogenases. All examined OHR *Sulfurospirillum* genomes possess two copies of reductive dehalogenase homologous (RdhA) genes, except that four RdhA copies were found in the genome of strain ACS_{TCE} (Fig. 4). Each of these RdhA genes is adjacent to a gene with a size of 150 to 225 bp encoding reductive dehalogenase membrane anchor protein RdhB. Phylogenetic analysis indicated that all 18 RdhAs encoded by the genomes of known *Sulfurospirillum* species strains can be grouped into two distinct clusters, with several RdhAs belonging to the genera *Desulfovibrio*, *Desulfomonile*, and *Desulforhopalus* as the closest relatives (Fig. 6). Seven RdhAs in cluster I (colored in blue in Fig. 6) were 100% identical to each other, but the RdhA from strain ACS_{TCE} shares 99.8% amino acid similarity with the other seven RdhAs in cluster I. The substrate spectrum of the eight putative RdhAs in cluster I remains elusive. By comparison, the three identical RdhAs (ACS_{TCE}_3, ACS_{TCE}_17, and ACS_{TCE}_2843) of strain ACS_{TCE} and the other five putative RdhAs (Shal_1516, JPD-1_1501, SL2-2_1608, ACS_{DCE}_8 and Smul_N_1565) in cluster II (colored red in Fig. 6) were grouped with the characterized PceA (Smul_1557) of *S. multivorans* strain DSM 12446, indicating that they can dechlorinate PCE. The putative RdhAs of Shal_1516, JPD-1_1501, SL2-2_1608, ACS_{DCE}_8, and Smul_N_1565 shared 92.1%, 94.0%, 96.2%, 97.2%, and 99.8% similarities, respectively, with PceA of *S. multivorans*. The three identical RdhAs of strain ACS_{TCE} (colored red in Fig. 6) share 97.8% similarity with the RdhA of strain SL2-1 (SL2-1_1591), and these four RdhAs were predicted to be responsible for dechlorinating PCE to TCE based on the fact that strain ACS_{TCE} and strain SL2-1 could only dechlorinate PCE to TCE. Only four critical residue differences (Ser²⁷⁹ versus Ala²⁷⁹, Gly²⁸⁶ versus Cys²⁸⁶, Ser³¹² versus Cys³¹², and Pro³²⁰ versus Ala³²⁰) were identified among the four putative PCE-to-TCE dechlorinating RdhAs (SL2-1_1591, ACS_{TCE}_3, ACS_{TCE}_17, and ACS_{TCE}_2843) and the other six putative PCE-to-cDCE dechlorinating RdhAs (Shal_1516, JPD-1_1501, SL2-2_1608, ACS_{DCE}_8, Smul_1557, and N_1565). The architecture of the active site of PCE-to-TCE and PCE-to-cDCE dechlorinating RdhAs of *Sulfurospirillum* appears to be similar (26). Based on the structure of PceA in *S. multivorans* (27), one has to assume that the residues identified here are not directly involved in the formation of the enzyme's

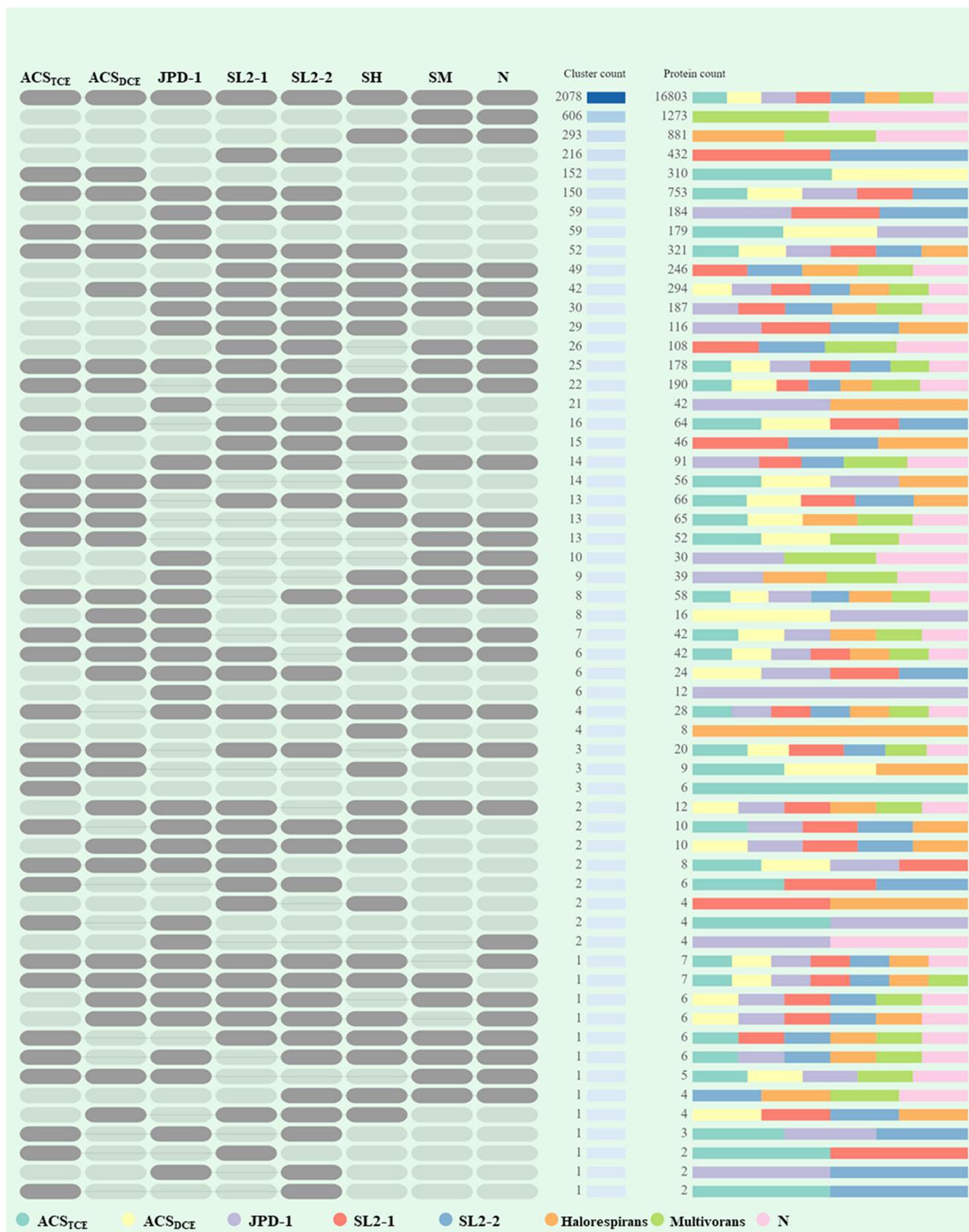


FIG 5 Orthologous gene clusters across eight OHR *Sulfurospirillum* strains. The filled and gray blocks indicate the presence and absence of orthologous gene clusters in each genome. SH, SM, and N stand for *S. halorespirans*, *S. multivorans*, and strain N, respectively.

genetic contents of *Sulfurospirillum* will assist us in understanding these versatile microorganisms regarding organohalide respiration.

Despite physiological and evolutionary differences among the eight OHR *Sulfurospirillum* species strains, the second copies of RdhAs (e.g., cluster I) are highly conserved (i.e., >99% amino acid sequence identity) in all eight OHR *Sulfurospirillum* strains (14, 28). Generally, reductive dehalogenases are known to be phylogenetically diverse (29), and we did not expect to observe that the second copies of RdhAs with unknown functions were highly conserved in OHR *Sulfurospirillum* species strains distributed in geographically distinct locations. Despite the detection of the second copy of RdhA in the transcriptomic and proteomic studies on *S. multivorans* grown on PCE (14, 24), the function of the second copy of RdhA remains unclear. Hypothetically, the two genes encoding RdhAs (i.e., *pceA* and the second copy of the *rdhA* gene) were acquired horizontally before the speciation of the genus *Sulfurospirillum*, and more evolutionary pressures could be forced on the functional *pceA* gene than the second copy of the *rdhA* gene.

Genome-inferred metabolic capacities and electron transport mechanism in *Sulfurospirillum*. In addition to OHR, *Sulfurospirillum* species strains can conserve energy using a variety of electron acceptors (e.g., fumarate, dimethyl sulfoxide [DMSO], thiosulfate, arsenate, selenate, and nitrate) and electron donors (e.g., hydrogen, formate, hydrogen sulfide, lactate, and pyruvate) (4). For example, *S. halospirans* strain PCE-M2 possesses a gene cluster encoding the SoxCDXYZAB proteins for thiosulfate oxidation. By comparison, other *Sulfurospirillum* species strains, including strain ACS_{DCE} and strain ACS_{TCEr}, have only two copies of the *phsA* gene encoding thiosulfate reductase. *Sulfurospirillum* species strains have been implicated in arsenate reduction with the potential of mobilizing arsenic in underground aquifers. While a single copy of the arsenate reductase gene *arsC* is present in the genomes of strains ACS_{DCEr}, ACS_{TCEr}, and JPD-1, other OHR *Sulfurospirillum* species strains harbor multiple copies of the *arsC* gene as well as arsenite oxidase genes *aioAB* (14). The ability to oxidize arsenite to arsenate at the presence of azurin protein as an electron acceptor was not observed in pure culture study, demonstrating that *aioAB* genes may not be functioning.

The electron transport chain of OHR can be generally classified into two categories: quinone dependent, represented by *Sulfurospirillum*, and quinone independent, represented by *Dehalococcoides* (12, 26, 30). OHR *Sulfurospirillum* species strains are quinone dependent and likely express the NapGH-like quinol dehydrogenases to transfer electrons from the menaquinone pool to PceA, similar to the function of NapGH quinol dehydrogenase in nitrate reduction (12, 24, 26). One of the possible electron sources for replenishing the electron pool is through oxidizing pyruvate via two different enzymes: pyruvate:ferredoxin/ferredoxin oxidoreductase (PFOR) and ubiquinone-dependent pyruvate dehydrogenase (PoxB) (14). The PFOR protein catalyzes the conversion of pyruvate to acetyl-CoA and carbon dioxide with simultaneous transfer of two electrons to ferredoxin or flavodoxin, while the PoxB enzyme could convert pyruvate into acetate and carbon dioxide accompanied by transferring the generated electrons directly to the menaquinone pool (31). The gene encoding PFOR has been found in all 12 *Sulfurospirillum* genomes; by comparison, the *poxB* genes are only present in the eight OHR *Sulfurospirillum* genomes, suggesting that *poxB* is probably related to the electron transfer chain of OHR. The inferred *poxB* products of OHR *Sulfurospirillum* species strains share 86.9% to 100% amino acid similarities to each other and are closely related to the ones identified in *Malaciobacter marinus* and gammaproteobacterial microorganisms (Fig. S8). The coexistence of both genes in a bacterial genome is uncommon, and the advantages for OHR *Sulfurospirillum* species strains possessing the *pfoR* and *poxB* genes simultaneously are not clear.

Summary. In this study, we performed phylogenetic, pangenomic, and evolutionary analyses on 12 complete *Sulfurospirillum* genomes, including two newly sequenced ones of strain ACS_{DCE} and strain ACS_{TCE} and proposed a new species, "*Candidatus Sulfurospirillum acididehalogenans*," represented by strains ACS_{TCEr}, ACS_{DCEr}, and JPD-1. The relatively preserved region identified in all examined OHR *Sulfurospirillum* species strains, but

not *S. multivorans* strains, suggested that they have a common ancestor that acquired the OHR capability recently and an additional insertion event(s) occurred in *S. multivorans* strains. Comparison between the PCE-to-TCE RdhAs and PCE-to-cDCE RdhAs identified the differences of four amino acid residues, but how these residue differences affect the substrate specificity of reductive dehalogenases remains unclear. The versatile metabolisms in *Sulfurospirillum* species strains ensure their great potentials in biotechnological applications, including the cleanup of soil and groundwater contaminated with a combination of chemicals (e.g., mixtures of arsenate, chlorinated ethenes, nitrate, and selenate). The ability to grow under unfavorable conditions (e.g., low pH) further emphasizes their roles and functions in specialized environmental settings. Nevertheless, some predictions on the metabolisms and capabilities of *Sulfurospirillum* were solely based on *in silico* genomic analysis, and experimental evidence is warranted in future studies. The plasmids harbored by strain ACS_{TCE} and strain ACS_{DCE} represent a promising tool for developing a robust gene editing tool that can advance the understanding of physiology and metabolism in the genus *Sulfurospirillum*.

MATERIALS AND METHODS

Sequence data set and phylogenetic analysis. As of January 2022, 47 *Sulfurospirillum* assemblies are publicly available (<https://www.ncbi.nlm.nih.gov/assembly/?term=Sulfurospirillum>). CheckM (version 1.0.18) (32) assessment showed that the average completeness and contamination of 35 assemblies were 85.5% and 0.9%, respectively. Since the incompleteness and contamination of metagenome-assembled draft genomes (33) may affect the analysis of core, auxiliary, and singleton families, we only focused on the completed *Sulfurospirillum* genomes in this study. Among them, 12 completed *Sulfurospirillum* genomes, including the ones of strain ACS_{TCE} and strain ACS_{DCE}, were selected for the following analyses (see Table S1 in the supplemental material and supplemental tables at <https://doi.org/10.6084/m9.figshare.17014352.v1>). Genome sequences of strain ACS_{TCE} and strain ACS_{DCE} were recently deposited and are publicly available in the NCBI genome database (21, 22). Phylogenetic analysis of *Sulfurospirillum* 16S rRNA gene sequences, which were retrieved from the Ribosomal Database Project (RDP) database, release 11 update 5 (34), was performed using PhyML with the general time-reversible (GTR) substitution model (35). Whole-genome-based taxonomy was analyzed using the Type (Strain) Genome Server (TYGS) (36) and Genome Taxonomy Database (GTDB), as described previously (37, 38). Briefly, RNAmmer v1.2 was applied to extract *Sulfurospirillum* 16S rRNA gene sequences (39), which were compared with those of available 10,997 type strains in the TYGS database for finding additional closely related type strains. Genomes were selected for pairwise comparisons using GBDP (Genome BLAST Distance Phylogeny) and accurate intergenomic distances (40), and the results were used to construct phylogenetic trees with a balanced minimum evolution via FASTME 2.1.4 (41). Species boundary was defined as 70% DNA-DNA hybridization (DDH). Average nucleotide identity (ANI) (e.g., ANIm [42], ANIb [43]) was calculated using JSpeciesWS (44) and orthoANI (45) to evaluate if two or more genomes can be classified into the same species. The ANI threshold for species boundary is defined as 95%.

Molecular clock analysis. MEGA X (46) for molecular evolutionary genetics analysis was applied to predict the divergence times using the RelTime method (47, 48) and the Tamura-Nei model (49). The 16S rRNA gene sequences were aligned with MUSCLE using the unweighted pair group method using average linkages algorithm and then analyzed for phylogeny reconstruction with the minimum evolution method or neighbor-joining method using Mega X. The nucleotide sequence alignment and phylogenetic tree were used as the input for RelTime-ML. The divergence times shown in Fig. S1 for several *Sulfurospirillum* species were predicted by TimeTree (50, 51) and were applied to set the divergence time calibration constraints by following the published approach (18, 52).

Whole-genome comparison and pangenome analysis. *Sulfurospirillum* genomes were reannotated using the RAST (Rapid Annotation using Subsystem Technology) tool with default parameters (53, 54) and eggNOG-mapper (55–57) to ensure annotation conformity with formats and consistency across all genomes. Annotated coding sequences were verified by BLAST search (58) against NCBI nonredundant protein sequences and by UniProt database search (e.g., UniProtKB reference proteomes plus Swiss-Prot) (59). All 12 completed *Sulfurospirillum* genomes were used for pangenome analysis, which was constructed and performed by OrthoMCL (60) with default parameters using the KBase platform (61). The protein sequences of the key functions were retrieved and analyzed using KEGG (<https://www.kegg.jp>) (62, 63) and BRENDA (www.brenda-enzymes.org) (64) to identify the key pathways and modules among various *Sulfurospirillum* species strains.

Analysis of functional sequences. Protein sequences for building the phylogenetic inference tree were retrieved from the UniProt database (www.uniprot.com) (59). All protein sequences for building the phylogenetic trees of reductive dehalogenases are listed in the Text S1 (see also Data Set 1 at <https://doi.org/10.6084/m9.figshare.17014145.v1>) (65). The phylogenetic trees for reductive dehalogenases and transposases were built with Geneious software version 11.1.5 using the MUSCLE and FastTree or PhyML with default settings (Biomatters Inc., Newark, NJ, USA).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.02 MB.

FIG S1, TIF file, 0.7 MB.

FIG S2, TIF file, 0.3 MB.

FIG S3, TIF file, 0.3 MB.

FIG S4, TIF file, 0.2 MB.

FIG S5, TIF file, 0.2 MB.

FIG S6, TIF file, 0.5 MB.

FIG S7, TIF file, 0.9 MB.

FIG S8, TIF file, 0.9 MB.

ACKNOWLEDGMENTS

We thank all the researchers for sharing the publicly available genomic data and the open-source bioinformatics tools. Special thanks to Frank Löffler's for support and guidance.

This work was supported by the National Key Research and Development Program of China (2019YFC1804400), National Natural Science Foundation of China (grant no. 41907287, 41977295, and 41907220), Liaoning Revitalization Talents Program XLYC1807139, and Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS), grant no. ZDBS-LY-DQC038.

We declare no competing financial interest.

REFERENCES

- Lacroix E, Brovelli A, Barry DA, Holliger C. 2014. Use of silicate minerals for pH control during reductive dechlorination of chloroethenes in batch cultures of different microbial consortia. *Appl Environ Microbiol* 80: 3858–3867. <https://doi.org/10.1128/AEM.00493-14>.
- Yang Y, Capiro NL, Marcet TF, Yan J, Pennell KD, Löffler FE. 2017. Organohalide respiration with chlorinated ethenes under low pH conditions. *Environ Sci Technol* 51:8579–8588. <https://doi.org/10.1021/acs.est.7b01510>.
- van der Stel AX, Wösten MMSM. 2019. Regulation of respiratory pathways in Campylobacterota: a review. *Front Microbiol* 10:1719. <https://doi.org/10.3389/fmicb.2019.01719>.
- Goris T, Diekert G. 2016. The genus *Sulfurospirillum*, p 209–234. In Adrian L, Löffler FE (ed), *Organohalide-respiring bacteria*. Springer, Berlin, Germany.
- Hubert C, Voordouw G. 2007. Oil field souring control by nitrate-reducing *Sulfurospirillum* spp. that outcompete sulfate-reducing bacteria for organic electron donors. *Appl Environ Microbiol* 73:2644–2652. <https://doi.org/10.1128/AEM.02332-06>.
- Goris T, Schenz B, Zimmermann J, Lemos M, Hackermüller J, Schubert T, Diekert G. 2017. The complete genome of the tetrachloroethene-respiring Epsilonproteobacterium *Sulfurospirillum halorespirans*. *J Biotechnol* 255: 33–36. <https://doi.org/10.1016/j.jbiotec.2017.06.1197>.
- Pietari JM. 2002. Characterization of tetrachloroethene dechlorinating cultures and isolation of a novel tetrachloroethene to cis-1, 2-dichloroethene halorespiring bacterium. PhD thesis. University of Washington, Seattle, WA.
- Buttet GF, Murray AM, Goris T, Burion M, Jin B, Rolle M, Holliger C, Maillard J. 2018. Coexistence of two distinct *Sulfurospirillum* populations respiring tetrachloroethene-genomic and kinetic considerations. *FEMS Microbiol Ecol* 94:fy018. <https://doi.org/10.1093/femsec/fy018>.
- Schubert T. 2017. The organohalide-respiring bacterium *Sulfurospirillum multivorans*: a natural source for unusual cobamides. *World J Microbiol Biotechnol* 33:93. <https://doi.org/10.1007/s11274-017-2258-x>.
- Kruse S, Goris T, Westermann M, Adrian L, Diekert G. 2018. Hydrogen production by *Sulfurospirillum* species enables syntrophic interactions of Epsilonproteobacteria. *Nat Commun* 9:4872. <https://doi.org/10.1038/s41467-018-07342-3>.
- Kruse S, Turkowsky D, Birkigt J, Matturro B, Franke S, Jehmlich N, von Bergen M, Westermann M, Rossetti S, Nijenhuis I, Adrian L, Diekert G, Goris T. 2021. Interspecies metabolite transfer and aggregate formation in a co-culture of *Dehalococcoides* and *Sulfurospirillum* dehalogenating tetrachloroethene to ethene. *ISME J* 15:1794–1809. <https://doi.org/10.1038/s41396-020-00887-6>.
- Turkowsky D, Jehmlich N, Diekert G, Adrian L, von Bergen M, Goris T. 2018. An integrative overview of genomic, transcriptomic and proteomic analyses in organohalide respiration research. *FEMS Microbiol Ecol* 94: fy013. <https://doi.org/10.1093/femsec/fy013>.
- Rischer M, Raguz L, Guo H, Keiff F, Diekert G, Goris T, Beemelmans C. 2018. Biosynthesis, synthesis, and activities of barnesin A, a NRPS-PKS hybrid produced by an anaerobic Epsilonproteobacterium. *ACS Chem Biol* 13:1990–1995. <https://doi.org/10.1021/acschembio.8b00445>.
- Goris T, Schubert T, Gadkari J, Wubet T, Tarkka M, Buscot F, Adrian L, Diekert G. 2014. Insights into organohalide respiration and the versatile catabolism of *Sulfurospirillum multivorans* gained from comparative genomics and physiological studies. *Environ Microbiol* 16:3562–3580. <https://doi.org/10.1111/1462-2920.12589>.
- Ross DE, Marshall CW, May HD, Norman RS. 2016. Comparative genomic analysis of *Sulfurospirillum cavolei* MES reconstructed from the metagenome of an electrosynthetic microbiome. *PLoS One* 11:e0151214. <https://doi.org/10.1371/journal.pone.0151214>.
- Esken J, Goris T, Gadkari J, Bischler T, Forstner KU, Sharma CM, Diekert G, Schubert T. 2020. Tetrachloroethene respiration in *Sulfurospirillum* species is regulated by a two-component system as unraveled by comparative genomics, transcriptomics, and regulator binding studies. *Microbiologyopen* 9:e1138. <https://doi.org/10.1002/mbo3.1138>.
- Mello B. 2018. Estimating timetrees with MEGA and the TimeTree resource. *Mol Biol Evol* 35:2334–2342. <https://doi.org/10.1093/molbev/msy133>.
- Yang Y, Zhang Y, Capiro NL, Yan J. 2020. Genomic characteristics distinguish geographically distributed *Dehalococcoides*. *Front Microbiol* 11: 546063. <https://doi.org/10.3389/fmicb.2020.546063>.
- Miura T, Uchino Y, Tsuchikane K, Ohtsubo Y, Ohji S, Hosoyama A, Ito M, Takahata Y, Yamazoe A, Suzuki K, Fujita N. 2015. Complete genome sequences of *Sulfurospirillum* strains UCH001 and UCH003 isolated from groundwater in Japan. *Genome Announc* 3:e00236-15. <https://doi.org/10.1128/genomeA.00236-15>.
- Gutierrez-Barranquero JA, Cazorla FM, de Vicente A, Sundin GW. 2017. Complete sequence and comparative genomic analysis of eight native *Pseudomonas syringae* plasmids belonging to the pPT23A family. *BMC Genomics* 18:365. <https://doi.org/10.1186/s12864-017-3763-x>.
- Yang Y, Huo L, Li X, Yan J, Löffler FE. 2021. Complete genome sequence of *Sulfurospirillum* sp. strain ACS_{DCE}, an anaerobic bacterium that respire tetrachloroethene under acidic pH conditions. *Microbiol Resour Announc* 10:e01360-20. <https://doi.org/10.1128/MRA.01360-20>.
- Huo L, Yang Y, Lv Y, Li X, Löffler FE, Yan J. 2020. Complete genome sequence of *Sulfurospirillum* strain ACS_{TCE1}, a tetrachloroethene-respiring

- anaerobe isolated from contaminated soil. *Microbiol Resour Announc* 9: e00941-20. <https://doi.org/10.1128/MRA.00941-20>.
23. Fraikin N, Goormaghtigh F, Van Melderen L. 2020. Type II toxin-antitoxin systems: evolution and revolutions. *J Bacteriol* 202:e00763-19. <https://doi.org/10.1128/JB.00763-19>.
 24. Goris T, Schifffmann CL, Gadkari J, Schubert T, Seifert J, Jehmlich N, von Bergen M, Diekert G. 2015. Proteomics of the organohalide-respiring *Epsilonproteobacterium Sulfurospirillum multivorans* adapted to tetrachloroethene and other energy substrates. *Sci Rep* 5:13794. <https://doi.org/10.1038/srep13794>.
 25. Lund P, Tramonti A, De Biase D. 2014. Coping with low pH: molecular strategies in neutralophilic bacteria. *FEMS Microbiol Rev* 38:1091–1125. <https://doi.org/10.1111/1574-6976.12076>.
 26. Schubert T, Adrian L, Sawers RG, Diekert G. 2018. Organohalide respiratory chains: composition, topology and key enzymes. *FEMS Microbiol Ecol* 94:fiy035. <https://doi.org/10.1093/femsec/fiy035>.
 27. Bommer M, Kunze C, Fessler J, Schubert T, Diekert G, Dobbek H. 2014. Structural basis for organohalide respiration. *Science* 346:455–458. <https://doi.org/10.1126/science.1258118>.
 28. Buttet GF, Holliger C, Maillard J. 2013. Functional genotyping of *Sulfurospirillum* spp. in mixed cultures allowed the identification of a new tetrachloroethene reductive dehalogenase. *Appl Environ Microbiol* 79:6941–6947. <https://doi.org/10.1128/AEM.02312-13>.
 29. Hug LA, Maphosa F, Leys D, Löffler FE, Smidt H, Edwards EA, Adrian L. 2013. Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. *Philos Trans R Soc Lond B Biol Sci* 368:20120322. <https://doi.org/10.1098/rstb.2012.0322>.
 30. Wang S, Qiu L, Liu X, Xu G, Siegert M, Lu Q, Juneau P, Yu L, Liang D, He Z, Qiu R. 2018. Electron transport chains in organohalide-respiring bacteria and bioremediation implications. *Biotechnol Adv* 36:1194–1206. <https://doi.org/10.1016/j.biotechadv.2018.03.018>.
 31. Tittmann K. 2009. Reaction mechanisms of thiamin diphosphate enzymes: redox reactions. *FEBS J* 276:2454–2468. <https://doi.org/10.1111/j.1742-4658.2009.06966.x>.
 32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
 33. Shaiber A, Eren AM. 2019. Composite metagenome-assembled genomes reduce the quality of public genome repositories. *mBio* 10:e00725-19. <https://doi.org/10.1128/mBio.00725-19>.
 34. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42:D633–D642. <https://doi.org/10.1093/nar/gkt1244>.
 35. Guindon S, Dufayard JF, 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. <https://doi.org/10.1093/sysbio/syq010>.
 36. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
 37. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38:1079–1086. <https://doi.org/10.1038/s41587-020-0501-8>.
 38. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
 39. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
 40. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. <https://doi.org/10.1186/1471-2105-14-60>.
 41. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* 32:2798–2800. <https://doi.org/10.1093/molbev/msv150>.
 42. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
 43. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91. <https://doi.org/10.1099/ijs.0.64483-0>.
 44. Richter M, Rossello-Mora R, Oliver Glockner F, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32:929–931. <https://doi.org/10.1093/bioinformatics/btv681>.
 45. Lee I, Ouk Kim Y, Park SC, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100–1103. <https://doi.org/10.1099/ijsem.0.000760>.
 46. Kumar S, Stecher G, Li M, Niyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>.
 47. Tamura K, Battistuzzi FU, Billings-Ross P, Murillo O, Filipiński A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A* 109:19333–19338. <https://doi.org/10.1073/pnas.1213199109>.
 48. Tamura K, Tao Q, Kumar S. 2018. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Mol Biol Evol* 35:1770–1782. <https://doi.org/10.1093/molbev/msy044>.
 49. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>.
 50. Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timeliness, timetrees, and divergence times. *Mol Biol Evol* 34:1812–1819. <https://doi.org/10.1093/molbev/msx116>.
 51. Marin J, Battistuzzi FU, Brown AC, Hedges SB. 2017. The timetree of prokaryotes: new insights into their evolution and speciation. *Mol Biol Evol* 34:437–446. <https://doi.org/10.1093/molbev/msw245>.
 52. McDonald BR, Currie CR. 2017. Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *mBio* 8:e00644-17. <https://doi.org/10.1128/mBio.00644-17>.
 53. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
 54. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
 55. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
 56. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *bioRxiv* <https://doi.org/10.1101/2021.06.03.446934>.
 57. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34:2115–2122. <https://doi.org/10.1093/molbev/msx148>.
 58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 59. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.
 60. Li L, Stoecckert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>.
 61. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia JM, Chia JM, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018.

- KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
62. Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* 29:28–35. <https://doi.org/10.1002/pro.3711>.
 63. Kanehisa M. 2017. Enzyme annotation and metabolic reconstruction using KEGG. *Methods Mol Biol* 1611:135–145. https://doi.org/10.1007/978-1-4939-7015-5_11.
 64. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. 2019. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* 47:D542–D549. <https://doi.org/10.1093/nar/gky1048>.
 65. Molenda O, Puentes JL, Cao X, Nesbo CL, Tang S, Morson N, Patron J, Lomheim L, Wishart DS, Edwards EA. 2020. Insights into origins and function of the unexplored majority of the reductive dehalogenase gene family as a result of genome assembly and ortholog group classification. *Environ Sci Process Impacts* 22:663–678. <https://doi.org/10.1039/c9em00605b>.