Phylogenetically and structurally diverse reductive dehalogenases link biogeochemical cycles in deep-sea cold seeps

Yingchun Han[1 #], Yongyi Peng[1, 2 #], Jiaxue Peng[1, 3], Lei Cao[4], Yangru Xu[5], Yi Yang[6], Minghuo Wu[7], Hao Zhou[7], Chen Zhang[8], Dongdong Zhang[5], Minxiao Wang[4], Chris Greening[9], Xiyang Dong[1, 10 *]

[1] Key Laboratory of Marine Genetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen 361005, China

[2] School of Marine Sciences, Sun Yat-Sen University, Zhuhai 519082, China

[3] College of Environmental Science and Engineering, Dalian Maritime University, Dalian 116026, China

[4] Deep Sea Research Center, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266000, China

[5] Institute of Marine Biology and Pharmacology, Ocean College, Zhejiang University, Zhoushan 316021, China

[6] Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China

[7] School of Ocean Science and Technology, Dalian University of Technology, Panjin 124221, China

[8] Department of Microbiology, Immunology and Molecular Genetics, University of Texas Health Science Center, San Antonio 78229, USA

[9] Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

[10] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China

[#] These authors contributed equally to this work.

* Correspondence can be addressed to Xiyang Dong (dongxiyang@tio.org.cn).

Abstract

Reductive dehalogenation is crucial for halogen cycling and environmental remediation, yet its ecological role is incompletely understood, especially in deep-sea environments. To address this gap, we investigated the diversity of reductive dehalogenases (RDases) and ecophysiology of organohalide reducers in deep-sea cold seeps, which are environments rich in halogenated compounds. Through genome-resolved metagenomic analysis 165 global cold seep sediment samples, we identified four types of RDases, namely prototypical respiratory, transmembrane respiratory, and cytosolic RDases, and one novel clade. These RDases are encoded by physiologically diverse microbes across four archaeal and 36 bacterial phyla, significantly broadening the known diversity of organohalide reducers. Halogen geochemistry, metatranscriptomic data, and metabolomic profiling confirm that organohalides occur at as high as 18 mg/g in these sediments and are actively reduced by microorganisms. This process is tightly linked to other biogeochemical cycles, including carbon, hydrogen, nitrogen, sulfur, and trace elements. RDases from cold seeps have diverse N-terminal structures across different gene groups, and *rdhA* genes in these environments are mostly functionally constrained and conserved. Altogether, these findings suggest that reductive dehalogenation is a central rather than supplemental process in deep-sea environments, mediated by numerous diverse microbes and novel enzymes.

Introduction

Globally situated along continental margins, deep-sea cold seeps are regions where fluids rich in methane and other hydrocarbons migrate from the deep subsurface to sediment-seawater interface[1, 2]. These fluids support a diverse array of microorganisms, including anaerobic methanotrophic archaea (ANME) and their syntrophic sulfate-reducing bacterial partners[3, 4]. Beyond hydrocarbons, cold seeps are also rich in various organic compounds like organohalides. Various organohalides, such as dichloroacetaldehyde and tetrachlorobenzene, are abundant in cold seep sediments in various locations, including the Eastern Gulf of Mexico and the Haima cold seep in the South China Sea[1, 3, 5]. These organohalides, given their high standard redox potential (E_o' = +0.24 to +0.58 V), are desirable electron acceptors for anaerobic respiration in these oxidant-limited environments[1, 6-11]. Organohalide-respiring microorganisms harness energy for growth by reducing these organohalides[6, 12, 13]. Notably some microorganisms within the *Chloroflexota* phylum, inhabiting deep-sea environments such as methane seeps, hydrothermal vents, and hadal trenches, can grow either obligately or facultatively on organohalides[14-17]. However, the diversity and roles of organohalide-reducing microorganisms in deep-sea cold seeps remain largely uncharted.

The process of reductive dehalogenation is mediated by reductive dehalogenases (RDases, also termed as RdhA, encoded by *rdhA*), which catalyze the cleavage of halides (like chloride or bromide) from the carbon backbone under anoxic conditions[18, 19]. For organohalide-respiring microorganisms, RDases are periplasmic membrane-associated proteins exported by the twin-arginine translocation (TAT) system. These enzymes are typically composed of two subunits: the periplasmic catalytic subunit RdhA usually harbors a cobalt cofactor that mediates halogen elimination and two tetranuclear [Fe-S] clusters for electron transfer, while the membrane-anchoring subunit RdhB typically contains two or three transmembrane helices (TMHs)[9, 19, 20]. The quinone-dependent RDases use respiratory quinols (e.g.

menaquinol), reduced by electron donors such as molecular hydrogen (H_2), whereas other enzymes are quinone-independent and appear to form respiratory supercomplexes with hydrogenases and potentially other primary dehydrogenases[19, 21, 22]. In addition to these prototypical RDases, two other classes of RDases are known. In the transmembrane respiratory RDases, the *rdhB* gene and TAT signal peptide are absent and instead the *rdhA* gene is directly connected to the membrane with one to three N-terminal TMHs[23-25]. In contrast, the cytosolic RDases (also called the catabolic RDases) are not connected to the respiratory chain and instead mediate the initial reductive dehalogenation of compounds so that they can be used as carbon and energy sources; these enzymes are typically NADPH-dependent and oxygen-tolerant, in contrast to their counterparts in organorespirers[19, 26]. However, classifying RDases remains challenging due to the limited number of well-characterized examples and their low sequence identity[9, 23, 26-28].

In this study, we analyzed a comprehensive dataset consisting of 165 metagenomic, 33 metatranscriptomic, and 55 metabolomic samples collected from deep-sea cold seeps globally. In addition to geochemical analyses, we focused on examining *rdhA* genes and organohalide reducers in these environments. We systematically analysed the phylogenetic diversity, ecological distribution, metabolic functions, genetic microdiversity, and structural evolution of various *rdhA* subfamilies and organohalide reducers. To do so, we relied on two major innovations. First, given the limited number of experimentally solved RDase structures[29-31], we employed AI-based protein structure modelling to better understand the structural and functional diversity within this superfamily. Second, we used state-of-the-art methods for quantifying genetic variations in microbial populations, including single-nucleotide variants (SNVs), to study the population dynamics of these enzymes across various environments[32-35]. Through developing a computational framework that integrates protein structure analysis with genetic variation, we were able to gain a more detailed understanding of the sequence-structure-function relationships of RDases, while uncovering a central role for organohalide reducers in shaping the biogeochemistry

and ecology of deep-sea sediments.

Results and Discussion

Geochemical evidence for reductive dehalogenation at cold seeps

Reductive dehalogenation can cause an increase in halogen profiles in cold seep sediments[36]. Our analysis of 1089 pore water samples revealed that (**Fig. 1a and Supplementary Fig. 1 and Supplementary Table 1**), on average, concentrations of dissolved chlorine (Cl^-) and bromine (Br^-) and the Br^-/Cl^- ratio in cold seep pore waters are slightly higher than those in overlying water and typical seawater[37, 38]. This trend is also observed in pore water iodide (I^-) concentrations in cold seep sediments, which are significantly higher compared to typical seawater[39]. Additionally, the solid-phase chlorine and bromine concentrations in 68 freeze-dried sediment samples showed a notable decrease with increasing sediment depth (**Fig. 1a and Supplementary Table 2**), potentially suggesting reductive dehalogenation decreases with sediment depth. However, other geological and biological processes may also contribute to these differences, as indicated by the varied dissolved halogen profiles found in 84 cold seep sediment cores (**Supplementary Figs. 2-7)**.

From 55 cold seep sediments, we found that the concentrations of total organic halogens (TOX) ranged between 6.7 mg/g and 17.6 mg/g, showing no significant variations across sediment depth but differed among the columns of cold seep sediments (**Fig. 1b and Supplementary Table 2**). Additionally, untargeted metabolomics analysis of these sediments identified 3,713 peaks, including 560 annotated metabolites. Among these metabolites (**Fig. 1c and Supplementary Table 3**), we found various halogenated compounds, some with a relative abundance reaching up to 10^3, such as chlorides (e.g., 2-chloro-L-phenylalanine, 2-chloro-5-methyl-cis-dienelactone), bromides (e.g., bromobenzene-3,4-oxide), fluorides (e.g., 5-fluorouridine triphosphate), and iodides (e.g., 5-iodo-2'-dUMP, 5-iodo-dCTP).

Cold seeps harbor a vast repertoire of reductive dehalogenase genes

Reductive dehalogenase (RdhA) sequences from cold seep metagenomes were identified and validated by analysing a non-redundant gene catalog, comprising 147 million genes[40], using a comprehensive workflow (**Supplementary Fig. 8**). The process yielded 3,993 validated RdhA sequences that fell into four phylogenetic groups **(Fig. 2**): (i) prototypical respiratory RDases (n = 1,466) associated with obligate and facultative organohalide-respiring microbes[19, 41, 42]; (ii) cytosolic RDases (n = 362) involved in liberating organohalides as energy and carbon sources[19, 26, 27]; (iii) transmembrane respiratory RDases (n = 1,963) characterized by a fusion of the *rdhA* gene with N-terminal TMHs[23, 26]; and (iv) a novel clade of RDases (n = 202). This clade, clustering between the cytosolic and prototypical clades, is characterized by the presence of both N-terminal TMHs (like transmembrane respiratory RDases) and a RdhB partner subunit (like prototypical respiratory RDases) **(Supplementary Table 4)**. These four groups could be further subclassified into six further subgroups, for example quinone-independent and quinone-dependent prototypical respiratory RDases, as depicted in **Fig. 2**.

Cold seep *rdhA* genes exhibited statistically significant differences across the four groups and six subgroups (P < 2.2e-16; **Fig. 3a and Supplementary Table 5**). Prototypical respiratory and transmembrane respiratory *rdhA* genes are more abundant than the cytosolic and novel clade *rdhA* genes (**Fig. 3a, inset**), suggesting that reductive dehalogenation in cold seeps is mostly associated with anaerobic respiration. Despite no notable differences in transcript abundance across groups (P > 0.05; **Fig. 3a**), the expression of *rdhA* subgroups (averaging 2.11 TPM; reaching 19.13 TPM for transmembrane respiratory *rdhA* genes) suggests microbial reductive dehalogenation occurs *in situ*. The distribution of *rdhA* genes appears to be influenced by sediment depth and cold seep types. A negative correlation between *rdhA* gene abundance and sediment depth (**Fig. 3b**) implies reduced dehalogenation activity in deeper sediments, potentially due to limited organohalides and nutrients, as observed in the Nankai

Trough subduction zone sediments[43]. In contrast, some *rdhA* subgroups display site-specific, depth-dependent trends indicative of niche specialization (**Supplementary Fig. 9a and Supplementary Table 6**). Significant differences in *rdhA* gene abundance were also observed among five cold seep types, especially in quinone-independent prototypical and cytosolic *rdhA* genes ($P < 0.05$; **Supplementary Fig. 10**).

Sulfate reduction and anaerobic methane oxidation, associated with *dsrA* and *mcrA* gene activities, respectively, are two important metabolic processes in cold seeps[44, 45]. The observed positive correlation between the abundances of *rdhA* and *dsrA* genes indicates that reductive dehalogenation and sulfate reduction processes likely occur simultaneously in cold seeps (**Fig. 3c and Supplementary Table 7**). Moreover, the fact these genes are present in similar abundances suggests reductive dehalogenation is potentially as crucial as sulfate reduction in these environments, underscoring its significance in the biogeochemical dynamics of cold seeps. Conversely, the negative correlation with *mcrA* gene abundance ($P < 0.001$; **Fig. 3d and Supplementary Fig. 11**) implies that anaerobic methane-oxidizing archaea may not engage in reductive dehalogenation[46, 47]. Additionally, the positive association of genes involved in osmotic stress protection and osmolyte transport with *rdhA* gene abundance (**Supplementary Figs 12-13 and Supplementary Table 8**) suggests that microorganisms deploy various salt tolerance mechanisms to maintain osmotic balance during reductive dehalogenation[48].

Organohalide reducers span across four archaeal and 36 bacterial phyla

From the cold seep genome catalog with 3,164 MAGs, we identified 586 *rdhA* genes that each fell into the four groups (**Supplementary Figs. 8 and 14**). These genes are distributed across 47 archaeal and 400 bacterial MAGs, encompassing four archaeal and 36 bacterial phyla, highlighting the wide distribution of organohalide reduction in these environments (**Fig. 4a, Supplementary Fig. 15 and Supplementary Table 9**). The most prevalent phyla harboring *rdhA* genes include *Chloroflexota* (n = 75),

Acidobacteriota (n = 56), *Bacteroidota* (n = 55), *Desulfobacterota* (n = 53), *Krumholzibacteriota* (n = 28), and *Asgardarchaeota* (n = 27). Most MAGs (79%, n = 355) encoded a single *rdhA* gene, whereas 92 MAGs across 15 phyla contained multiple *rdhA* genes **(Supplementary Fig. 15; Supplementary Table 9)**, in line with culture-based observations that highly versatile organohalide reducers produce multiple RDases[19] (**Supplementary Table 9**). Further examination revealed 36 mobile genetic elements (MGEs) associated with 22 *rdhA*-containing contigs (**Supplementary Table 10**), suggesting a role for MGEs in the horizontal transfer of *rdhA* genes and the resultant diversity of organohalide reducers. Additionally, the abundance of these organisms negatively correlates with sediment depth, in line with the decrease in *rdhA* gene abundance in deeper sediment layers (**Supplementary Fig. 16 and Fig. 3b**). To investigate potential substrates for organohalide reducers, we conducted molecular docking for RdhAs against 191 varied naturally occurring organohalides, including those commonly found in cold seeps **(Supplementary Table 11)**[3, 5, 14]. In total, we found 13,561 possible interactions between RdhAs and halohydrocarbons, with the binding energies ranged from -0.47 to -6.08 kcal/mol **(Supplementary Table 12)**.

The quinone-dependent respiratory *rdhA* genes were encoded by 136 microorganisms spanning 15 bacterial and four archaeal phyla, including most of the putative archaeal organohalide reducers (**Fig. 4a, Supplementary Fig. 15 and Supplementary Table 9**). These RdhAs show strong structural homology to the PceA dehalogenase (PDB id: 4UR3) from *Sulfurospirillum multivorans*[18], with TM-scores between 0.63 and 0.90 (**Supplementary Fig. 17**). In addition to be co-encoded with the membrane anchor subunit RdhB, this RDase is also frequently genomically associated with a partner uptake hydrogenase (Hup), the ferredoxin:NAD^+ oxidoreductase (Fnr), and transcriptional regulators (for example, GntR family, and NosR/NirI family) (**Fig. 4b and Supplementary Table 13**). Molecular docking analysis against 191 organohalides, including those abundant in cold seeps, demonstrated that *Thermoplasmata* RDase has a binding affinity of -4.1 kcal/mol for 2'-chloro-biphenyl-

2,3-diol and thus is a possible substrate (**Fig. 4c and Supplementary Tables 12**). Quinone-independent respiratory *rdhA* genes, typically found in obligate organohalide reducers *Dehalococcoides* and *Dehalogenimonas* within the phylum *Chloroflexota*[6, 19, 42], were also detected in *Acidobacteriota* and *Gemmatimonadota* **(Fig. 4a and Supplementary Fig. 15)**, broadening the known diversity of microorganisms harboring this gene type. Adjacent genes to *rdhA*, ferredoxin and NADH-coupled oxidoreductase were identified **(Supplementary Table 13)**, which are essential for the electron transport chain in organohalide respiration[49]. RDases in *Aminicenantaceae* (*Acidobacteriota*) and GWA2-58-10 (*Gemmatimonadota*) exhibit -4.0 to -4.8 kcal/mol binding affinities for chlorinated benzene and chlorinated paraffins **(Supplementary Tables 12)**, suggesting their suitability as substrates.

The transmembrane respiratory RDases were by far the most widespread enzymes, encoded by 287 microorganisms from 33 phyla, including multiple *Bacteroidota*, *Acidobacteriota*, *Desulfobacterota* and *Krumholzibacteriota* MAGs. This observation extends Atashgahi et al.'s findings that these enzymes are widespread despite being understudied[23]. Adjacent genes to *rdhA*, transcriptional regulators (e.g., GntR family), genes involved in electron transport (e.g., FAD dependent oxidoreductase) and MGEs (e.g., transposase) were identified **(Fig. 4b and Supplementary Table 13)**. We also observed transmembrane respiratory RDases form various phyla exhibit -3.2 to -6.0 kcal/mol binding affinities for chlorinated benzene and chlorinated alkane **(Supplementary Table 12)**, suggesting potential substrates. The novel clade was also widespread, encoded by 49 MAGs spanning 11 phyla, including the four abovementioned phyla and candidate phyla JAGLTZ01 and QNDG01 (**Fig. 4a, Supplementary Fig. 15 and Supplementary Table 9**). Despite being phylogenetically distant from the transmembrane respiratory RDases, this novel clade also likely has a respiratory function, given the *rdhA* gene is typically fused with between one to three N-terminal TMHs; however, a key difference is that 23 and 38 of the 50 genes in the novel clade are sometimes found alongside TAT systems and *rdhB* genes, respectively **(Supplementary Table 9)**. Novel RDases demonstrate binding

affinities ranging from -4.4 to -6.1 kcal/mol for chlorinated benzene **(Supplementary Table 12)**, indicating their suitability as substrates.

Cytosolic *rdhA* genes (n = 37) were mainly found in *Chloroflexota*, *Desulfobacterota*, and *Pseudomonadota*, but were also detected in a *Heimdallarchaeia* MAG **(Fig. 4a, Supplementary Fig. 15 and Supplementary Table 9)**. Consistent with a role in organic carbon degradation, these genes are often associated with bacterial transcriptional regulators IclR/MarR, beta oxidation genes, and galactose degradation genes **(Fig. 4b)**. In three MAGs, cytosolic *rdhA* genes were discovered on plasmids (**Supplementary Table 14**); this is reminiscent of BhbA, the plasmid-encoded first characterized enzyme from this family, which aerobically breaks down bromoxynil into 4-carboxy-2-hydroxymuconate-6-semialdehyde in *Comamonas* sp. 7D-2, which is then funneled into the tricarboxylic acid cycle[19, 50]. These cytosolic RdhAs have high TM-scores (0.92 to 0.95) (**Supplementary Fig. 18**) to the X-ray crystal structure of the 3-bromo-4-hydroxybenzoic acid reductase from *Nitratireductor pacificus* (NpRdhA; PDB id: 6ZY1)[27]. Two organohalide reducers from *Gammaproteobacteria*, encoding catabolic RDases, showed potential in reducing 1,5-dibromopentane, evidenced by binding energies of -2.44 and -3.11 kcal/mol (**Supplementary Table 12**). Further analysis revealed that these microorganisms contain genes responsible for the aerobic degradation of medium-chain alkanes (C5-C13, e.g. pentane; *alkB* and CYP153) **(Fig. 5 hydrocarbon degradation and Supplementary Table 15)**, suggesting a link between catabolic reductive dehalogenation and aerobic hydrocarbon degradation[19, 27, 50].

Cold seep halogen cycling is closely linked to central biogeochemical processes

We extensively annotated the metabolic capabilities of the RDase-encoding MAGs to gain insights into their potential ecophysiology and biogeochemical roles. Fermentation is the main process transforming organic carbon in cold seep environments[1, 51], which produces significant amounts of hydrogen (H_2) and various organic acids, both of which are key electron donors for organohalide reducers[41, 48, 52,

[53]. Consistently, 288 organohalide reducers were found to encode 495 hydrogenases. Of these, 337 are involved in H_2 oxidation[54], including group 1a, 1b, 1c, 1d, 1e, and 1f [NiFe]-hydrogenases that donate H_2-derived electrons to anaerobic respiratory chains **(Supplementary Fig. 19 and Supplementary Table 16)**. Additionally, genes responsible for formate (*fdhAB*, *fdoG*, *fdwB* and *fdoH*), acetate (*acs*, *acdA*, *ack* and *pta*), and lactate (*ldh*) oxidation were present in 226, 363, and 18 organohalide reducers, respectively (**Fig. 5, Supplementary Fig. 19 and Supplementary Table 17**). Most of these organohalide reducers involved in processing hydrogen and organic acids possess RdhAs that belong to prototypical and transmembrane respiratory RDases.

Organohalide reducers also participate in nitrogen, sulfur, and hydrocarbon cycles **(Fig. 5 nitrogen and sulfur metabolism)**. We identified 55 and 57 organohalide reducers with genes related to sulfate and thiosulfate reduction (reductive *dsrA* and *phsA*), respectively (**Supplementary Tables 17-18**). Additionally, 185 organohalide reducers harbor genes involved in the oxidation of sulfide, sulfur, and thiosulfate (*sqr*, *sdo* and *soxBYC*; **Supplementary Table 17**). 189 organohalide reducers harbor reductive genes associated with nitrogen metabolism (*narGH*, *nrfHA*, *nirBDKS* and *octR*), with a small proportion also capable of nitrogen fixation (n = 6) and nitrite oxidation (n = 5). Four organohalide reducers also encoded enzymes for the anaerobic degradation of hydrocarbons (**Fig. 5 hydrocarbon degradation, Supplementary Fig. 19 and Supplementary Table 15)**. Two of these, affiliated with *Desulfobacterota* and encoding quinone-dependent respiratory RDases, encode genes for the anaerobic degradation of n-alkanes (*assA*)[3, 55, 56]; they potentially remove chlorine from chlorinated paraffins, as suggested by their binding energy of -4.4 kcal/mol **(Supplementary Table 12)**. Altogether, the presence of these genes indicates that organohalide reducers are not only versatile in their metabolic functions but also well-equipped to adapt and thrive in various environmental conditions. Despite previous findings suggesting the coupling of AOM with reductive dehalogenation[7], *rdhA* genes were not found in the ANME group of microorganisms **(Supplementary Fig. 20 and**

Supplementary Table 9). Moreover, although 55 sulfate-reducing bacteria have the capacity for reductive dehalogenation, none seem to partner with ANME in this process (**Supplementary Table 18**). Considering the negative correlation between the gene abundance of *rdhA* and oxidative *mcrA* **(Fig. 3d)**, we hypothesize that these two processes are likely not coupled within cold seeps.

Cobalamin (vitamin B_{12}) acts as a cofactor for RDases, facilitating electron transfer[27, 42, 57]. Only 19% (n = 87) of organohalide reducers are capable of B_{12} biosynthesis through various pathways[58], including the aerobic (n = 2), anaerobic (n = 5), salvage remodeling (n = 12) and the Post-AdoCbi-P (n = 78) pathways **(Fig. 5, Supplementary Fig. 19 and Supplementary Table 19**). This finding is in line with existing knowledge that only a limited number of microbial genera in the ocean can synthesize B_{12} de novo[59, 60]. Among two capable of aerobic B12 biosynthesis, they encode cytosolic *rdhA* genes. The remaining 80% (n = 360) of organohalide reducers, which lack B_{12} biosynthesis genes, likely rely on B_{12} synthesized by other microorganisms or present in the environment for their reductive dehalogenation activities[61].

Cold seep RDases show diverse N-terminal structures across different gene groups

Consistent with the phylogenetic tree based on protein sequences, the structure-based phylogeny also supports the classification of cold seep RDases into four distinct groups (**Fig. 6 and Supplementary Table 20**): prototypical respiratory RDases, transmembrane respiratory RDases, cytosolic RDases, and the novel clade RDases. Structurally, bacterial and archaeal RDases from all four groups each contain corrinoid- and FeS-containing domains, each marked by a two-layered alpha-beta structure[29-31]; however, these enzymes greatly vary in their N-terminal structures in a manner consistent with their physiological roles. Our structural phylogeny revealed key aspects of RDase evolution that sequence analysis alone could not, suggesting at

328 least five key evolutionary events (**Fig. 6**). For the novel clade, the emergence of a
329 single N-terminal α-helix separated this group from other RDases (**Event 1 in Fig. 6**).
330 Most novel clade RDases (36 of 50) contain a single N-terminal transmembrane α-
331 helix (**Supplementary Fig. 21a**), likely anchoring these enzymes for a respiratory
332 function[62]. Half of these novel clade RDases (18 of 36) with a single TMH also
333 possess RdhB subunit, and the combination of the two forms a multi-transmembrane
334 helix predicted by AlphaFold Multimer (**Supplementary Fig. 22**). Other RDases
335 within this novel clade anchor to the membrane through the RdhB subunit
336 (**Supplementary Fig. 21b**).

337 Of the prototypical respiratory RDases, both the quinone-dependent and quinone-
338 independent enzymes share structural similarities. They each contain a domain
339 analogous to the quinone-binding domain found in quinone-dependent enzymes **(Fig.**
340 **6)**, such as the $DhPceA_2B_2$ from the tetrachloroethene-dechlorinating strain
341 *Desulfitobacterium hafniense* TCE1[29]. However, this domain in quinone-independent
342 RDases does not exhibit quinone-binding affinity, in contrast to their quinone-
343 dependent counterparts which show a significant binding affinity (-6.39 kcal/mol) and
344 possess shorter substrate transport tunnels (**Event 2 in Fig. 6 and Supplementary**
345 **Fig. 23**). Additionally, among the eleven RDases from *Heimdallarchaeia*, five
346 quinone-dependent RDases have complete N-terminal 4Fe-4S ferredoxin domains,
347 with three others having incomplete ones (**Event 3 in Fig. 6, Supplementary Fig. 24**
348 **and Supplementary Table 9**). The presence of an additional 4Fe-4S domain in these
349 archaeal RDases could potentially enhance the efficiency of the dehalogenation
350 process by using multiple electron pathways from the quinone pool to the active sites[8,
351 29].

352 Cytosolic RDases are variable in length, ranging from 300 to 1000 amino acids, and
353 they cluster into three distinct groups correlated with their respective phyla (**Event 4**
354 **in Fig. 6 and Supplementary Table 9**). Eight cold seep cytosolic RDases,
355 predominantly found in the *Pseudomonadota* phylum (with more than 700 amino

acids), function as monomers due to the presence of an additional vestigial cobalamin-binding core domain[26]. Half of these cytosolic RDases are characterized by a unique C-terminal NADP-linked ferredoxin reductase domain[26], with an alpha-beta three-layered sandwich structure, allowing electrons derived from NAD(P)H to drive reductive dehalogenation (**Supplementary Table 20**). According to the presence or absence of C-terminal reductase domain, these cytosolic RDases are termed as self-sufficient and truncated cytosolic RDases[26]. Another clade of cytosolic RDases (labelled as simplified cytosolic RDases) is typically present in *Desulfobacterota* and *Chloroflexota*, with an average length of 395 amino acids (**Supplementary Fig. 25**). Simplified cytosolic RDases possess a shortened N terminus compared to both self-sufficient and truncated cytosolic RDases, as they lack the vestigial cobalamin-binding domain, similar to prototypical respiratory RDases. The considerable variation in the length of cytosolic RdhA proteins indicates significant genetic diversity, which may be indicative of their functional adaptability among various microbial taxa.

In the transmembrane respiratory RdhA proteins, the presence of an integral membrane domain is a common feature (**Event 5 in Fig. 6**), yet not all members of this group exhibit N-terminal TMHs. Specifically, certain RdhAs in the two earliest clades within this group lack transmembrane α-helices (**Fig. 6**). Nevertheless, approximately half of these proteins compensate for this absence by utilizing transmembrane domains located in adjacent genes for membrane anchoring (**Supplementary Table 9**). The number of α-helices within these transmembrane domains varies, with some having one to three α-helices (**Supplementary Fig. 26**), which likely impacts how these proteins embed in the membrane and potentially in turn their substrate specificity and catalytic efficiency.

Cold seep *rdhA* genes are mostly functionally constrained and conserved

To investigate the evolutionary ecology of *rdhA* genes in cold seeps, we analyzed their genetic diversity across various sediment sites. Nucleotide diversity across *rdhA* genes

was consistently low, with an average of 0.013 (**Fig. 7a and Supplementary Table 21**). Most *rdhA* genes were under strong purifying selection, as evidenced by low pN/pS values (average 0.21 with 91.4% of ratios less than 0.4), suggesting reductive dehalogenation provides an adaptive advantage *in situ*. However, some quinone-dependent prototypical and cytosolic *rdhA* genes showed evidence of positive selection (**Supplementary Table 21**), potentially due to genetic drift or beneficial mutations[63]. Further 3D structural predictions for *rdhA* genes with high pN/pS values (>1.5)[64] demonstrated that amino acid changes related to non-synonymous mutations did not significantly alter overall structure or catalytic domains of the proteins (**Supplementary Fig. 27**). Additionally, the variation in nucleotide diversity and pN/pS ratios was statistically significant across different *rdhA* types (P = 2.2e-13 and P < 2.2e-16, respectively; **Fig. 7a and Supplementary Table 21**), indicating unique evolutionary trajectories for each *rdhA* type. Nucleotide diversity also varied significantly among different types of cold seeps (P = 0.0006), with gas hydrates, oil and gas seeps, and methane seep samples exhibiting higher nucleotide diversity (**Supplementary Fig. 28a**). Nevertheless, the pN/pS ratios were consistent across the five cold seep types (P = 0.25; **Supplementary Fig. 28b**). Overall, these findings are in line with previous observations of key functional genes in microbial populations within cold seeps[32, 65], highlighting the strong functional constraints and conservation of essential metabolic *rdhA* genes across various cold seep environments[66].

To further elucidate the association of nucleotide polymorphisms and protein structures of RDases in cold seeps, genetic variation and biophysical characteristics of RDases were examined. This included single-codon variant (SCV), synonymous (s) and nonsynonymous (ns) polymorphism rates of each codon ($pS^{(site)}$ and $pN^{(site)}$), as well as evaluating relative solvent accessibility (RSA) to indicate the exposure or burial of a site, and distance to ligand (DTL) to represent the proximity to the nearest active site at each codon position. A total of 50,531 SCVs (1,684 per metagenome on average) were found, and $pS^{(site)}$ exceeded $pN^{(site)}$ by a ratio of 6:1 on average of all SCVs, exhibiting particularly high enrichment of synonymous polymorphism within

the *rdhA* genes. Additionally, $pN^{(site)}$ and $pS^{(site)}$ values varied significantly from site to site (78.33% ± 0.13% versus 79.75% ± 0.11% of total variance, ANOVA), hinting varying selective pressures at different loci. The patterns of nucleotide polymorphism are largely determined by the structural and functional constraints of proteins[67]. The gene-wide distributions for $pN^{(site)}$ and $pS^{(site)}$ relative to RSA and DTL highlighted that $pN^{(site)}$ exhibited a distinct inclination for sites with higher RSA and DTL compared to $pS^{(site)}$ (**Fig 7b)**. This result suggests that buried sites (with lower RSA) and functional constraint sites (with lower DTL) are likely to purify nonsynonymous mutation for preserving protein structural stability and function, while revealing comparatively greater tolerance towards synonymous changes. This finding was also supported by the Pearson correlations of $\log_{10}(pN^{(site)})$ and $\log_{10}(pS^{(site)})$ to RSA and DTL for each gene-sample pair (**Fig. 7b)**, which showed consistently positive correlations for ns-polymorphism sites with an average Pearson coefficient of r_{RSA} = 0.22 and r_{DTL} = 0.08, whereas the correlations for s-polymorphism sites tended to cluster around 0, averaging r_{RSA} = 0.01 and r_{DTL} = 0.02. Together, by integrating genetic variants and structural features of RdhA proteins, we found that most polymorphism of *rdhA* genes yielded a strong purifying selection in order to preserve protein stability and metabolic activity of RDases.

Conclusions

In this study, we provide multiple levels of evidence that reductive dehalogenation is a key process supporting the ecology and biogeochemistry of deep-sea cold seeps. The concentrations of organohalide substrates reaching as high as 18 mg/g, widespread distribution of both prototypical and transmembrane respiratory reductive dehalogenases, and high abundance, expression, and conservation of RDase genes all suggest organohalides are a central rather than supplementary electron acceptor in these environments. Physiologically and phylogenetically diverse bacteria and archaea from some 40 phyla encode RDases, including key players in hydrogen, sulfur, carbon, and nitrogen cycling, with organohalide metabolism likely both directly and indirectly

influencing these cycles. It is probable that cold seep microbes have different organohalide substrate preferences, given the chemical diversity of the organohalides available to them and the vast structural diversity of the RDases they encode, which may contribute to metabolic niche partitioning in these competitive environments. In this regard, our analysis also revealed new phylogenetic and structural diversity in the RDases, as well as further insights into the complex evolutionary pathways of these enzymes. Notably, we discovered a novel clade of RDases that contains TMHs, TAT signal peptides, and *rdhB* genes, integrating features of both transmembrane and prototypical respiratory features. Finally, our research affirms the deep-sea environment as a repository of novel enzymes, which can be used both directly to remediate organohalide pollutants but also more broadly to advance sequence-structure-function relationships in this space.

Materials and Methods

Geochemical and metabolomics analyses

For the analysis of dissolved and solid-phase concentrations of Cl^- and Br^-, samples were measured using a Dionex Ion Chromatograph (Thermo Fisher, USA). The set included 1089 pore water samples and six overlying water samples from 63 sediment cores with lengths of 28, 32, and 40 cm, and 21 longer cores (ranging from 0-320 cm and 0-400 cmbsf) from the Qiongdongnan, Shenhu, Haima, and Site F cold seeps in the South China Sea (**Supplementary Fig. 1**). Additionally, 68 freeze-dried sediment samples were collected from the Qiongdongnan and Shenhu cold seeps.

For total organic halogens, 55 sediment samples from the Qiongdongnan and Shenhu cold seeps were analyzed using a multi X® 2500 AOX/TOX analyzer (Analytik Jena, Germany). The untargeted metabolomics analyses of these 55 sediment samples involved processing each sample in a solution of methanol, acetonitrile, and water in a 2:2:1 ratio, with a 20 mg/L internal standard. The samples were vortexed for 30

seconds, ground with steel beads for 10 minutes at 45 Hz, and ultrasonicated for 10 minutes in an ice bath. High-resolution LC-MS/MS analysis of all samples was conducted using a Waters Acquity I-Class PLUS ultra-high performance liquid tandem with a Waters Xevo G2-XS QTOF high-resolution mass spectrometer. Both primary and secondary mass spectrometry data were collected in MSe mode, controlled by the acquisition software (MassLynx V4.2, Waters). The raw data were then processed using Progenesis QI software, based on the METLIN database and Biomark's self-built library for identification. The relative abundance of metabolites was quantified based on their normalized peak areas, i.e. the percentage of peak area for each metabolite in the total peak area. The chemical structures of halogenated compounds were visualized using online platform MolView[68] (https://molview.org/).

Metagenomic and metatranscriptomic datasets from cold seep sediments

Metagenomes were compiled from 165 deep-sea sediment samples, with sediment depths ranging from 0 to 68.55 mbsf and water depths varying from 860 to 3005 meters. These samples were collected from 16 geographically diverse cold seep sites from around the world (**Supplementary Fig. 1**)[1, 3, 40, 51, 69-80], encompassing various types of cold seeps such as oil and gas seeps, methane seeps, gas hydrates, asphalt volcanoes, and mud volcanoes. The sites include: Eastern North Pacific (ENP), Santa Monica Mounds (SMM), Western Gulf of Mexico (WGM), Eastern Gulf of Mexico (EGM), Northwestern Gulf of Mexico (NGM), Scotian Basin (SB), Haakon Mosby mud volcano (HM), Mediterranean Sea (MS), Laptev Sea (LS), Jiaolong cold seep (JL), Shenhu area (SH), Haiyang4 (HY4), Qiongdongnan Basin (QDN), Xisha Trough (XST), Haima seep (HM1, HM3, HM5, HM_SQ, S11, SY5, and SY6) and site F cold seep (RS, SF, FR, and SF_SQ). Additionally, 33 metatranscriptomic data were obtained from our previous publications[51, 77, 79, 81], including those from South China Sea cold seeps in Jiaolong, Haima, Qiongdongnan Basin, and the Shenhu area. The detailed information of the metagenomic and metatranscriptomic datasets used in this study were described in our previous cold seep gene catalog publication[82].

Metagenomic and metatranscriptomic data processing

The workflow for metagenomic analyses is described in detail (e.g. software and parameters) in **Supplementary Table 22** [40]. Briefly, metagenomic reads were quality controlled and assembled into contigs. The protein-coding sequences predicted from contigs were clustered at 95% amino acid identity to generate a non-redundant gene catalog (n = 147,289,169). MAGs were derived from assembled contigs of over 1,000 bp using various binning software tools, including MetaBAT2[83], MaxBin2[84], CONCOCT[85], SemiBin[86], VAMB[87] and Rosella (https://github.com/rhysnewell/rosella). Produced MAGs were refined, quality controlled, and clustered at 95% average nucleotide identity, resulting in 3,114 species-level representative MAGs. The taxonomy of each MAG was initially assigned using GTDB-TK v2.1.1 with reference to GTDB R207 [88, 89] and then validated using a maximum-likelihood phylogenomic tree. The phylogenomic tree for MAGs was inferred based on concatenation of 43 conserved single-copy genes that were used as phylogenetic markers in CheckM v1.2.1[90], constructed using IQ-TREE v2.2.0.3[91] with best-fit model and 1000 ultrafast bootstraps.

Gene abundances in the non-redundant gene catalog across 165 metagenomes were quantified using Salmon (v.1.9.0)[92] in mapping-based mode (parameters: --validateMappings -meta) and read counts were normalized to GPM (genes per million). The relative abundance of each MAG was calculated using CoverM in genome mode (v0.6.1; https://github.com/wwood/CoverM; parameters: -min-read-percent-identity 0.95 -min-read-aligned-percent 0.75 -trim-min 0.10 -trim-max 0.90 -m relative_abundance) by mapping quality-controlled reads from the 165 metagenomes to all MAGs.

Regarding metatranscriptomes, raw reads were quality filtered (--skip-bmtagger) using Read_QC module within the metaWRAP (v1.3.2) pipeline[93]. To remove ribosomal RNAs from quality-controlled reads, SortMeRNA (v2.1)[94] was used with

default settings. Transcript abundances for *rdhA* genes were determined by mapping clean reads from 33 metatranscriptomes to the non-redundant gene catalog using Salmon (v.1.9.0; parameters: -validateMappings -meta)[92]. The transcript abundances of *rdhA* gene were calculated as TPM (transcripts per million).

Functional annotations and phylogenetic analysis

To identify *rdhA* genes, we used reference *rdhA* sequences (n = 1,040) from the Reductive Dehalogenase Database (https://rdasedb.biozone.utoronto.ca/)[28]. We searched for potential *rdhA* sequences in the non-redundant gene catalog against reference *rdhA* genes in RDaseDB using DIAMOND blastp (v2.0.8)[95], with >30% percentage identity and >50% query coverage as the cut off[41, 52]. Furthermore, we extracted a gene catalog encoding *rdhA* genes from the InterPro database using Pfam id "PF13486" (n = 4,163) and NCBI's Protein Family Models using NCBI HMM accession "TIGR02486" (n = 1,235). The hmmsearch tool in HMMER v3.3.2 was applied (E-value < 1E-10) using the amino acid sequences of non-redundant gene catalog to the reference gene catalog. To identify potential *rdhA* sequences in MAGs, Prodigal (v2.6.3; parameter: -meta)[96] was utilized to predict protein-coding sequences of all MAGs. These predicted sequences were also searched against the *rdhA* reference sequences using DIAMOND blastp (v2.0.15.153)[95] and HMMER v3.2.1, with the same parameters as mentioned above **(Supplementary Fig. 8)**. For each gene which passed the required criteria for *rdhA* gene identification by blastp or hmmsearch were merged and then the retrieved results were manually checked according to gene length (> 300 aa) and two Fe-S conserved motifs (CXXCXXCXXXCP, CXXCXXXCP). In brief, pairwise alignment of identified *rdhA* amino acid sequences for conserved active site analysis was performed and visualized using MAFFT v7.471 (-auto option)[97] and Jalview[98]. Phylogenetic trees were further constructed to validate the phylogenetic clades of RdhA, *rdhA* amino acid sequences and reference sequences were aligned using MUSCLE (v3.8.1551, default settings)[99] and trimmed using TrimAL (v1.4.1)[100] with default options. Maximum-likelihood

trees were constructed using IQ-TREE (v2.2.0.3)[91] with the "-m MFP -B 1000" options. All trees were visualized by using iTOL (v6)[101].

The MAGs were also annotated using DRAM (v1.3.5; parameter: --min_contig_size 1000)[102] against KEGG, Pfam, MEROPS and dbCAN databases. To annotate genes involved in hydrocarbon degradation, CANT-HYD database[103] with the HMMs of 16 hydrocarbon-degrading genes was searched against with parameters "-cut_nc" using HMMER v3.2.1. The hydrogenases were annotated using DIAMOND blastp (v2.0.15.153; options: --id 50 --query-cover 80 --evalue 1E-20) against local protein databases (https://doi.org/10.26180/c.5230745)[104], further confirmed and classified using the HydDB tool[54]. The VB_{12}Path was employed to annotate genes involved the VB_{12} synthesis pathway[58].

Identification of mobile genetic elements

Classification of *rdhA*-containing contigs as belonging to chromosomes, plasmids or viruses was performed using Genomad v.1.5.0 with default parameters[105]. Integrons, integrative conjugative element (ICEs), IS elements and transposons were identified using HMM searches of the proteins against the 68 marker HMM profiles by default, which are available on proMGE (http://promge.embl.de/)[106].

Protein modeling and molecular docking

Deep learning approach AlphaFold2 has yielded remarkable progress in predicting protein structures. The three-dimensional structures of 586 *rdhA* sequences were generated with AlphaFold (v2.0; full_dbs)[107]. The pLDDT values, a measure for confidence of the AlphaFold structure prediction, range from 65 to 97, 91 on average. The protein complexes of novel clade RDases were predicted using AlphaFold (v2.0; model_preset = multimer)[107]. The paired alignment of structures relied on 3Di and amino acid based alignment via Foldseek v8.ef4e960[108]. Structure-based RdhA tree was further built using Foldtree (https://github.com/DessimozLab/fold_tree) based on

a local structural alphabet[109] and visualized using iTOL (v6)[101]. Structural similarity and homology relationships of RdhAs with 1,007,623 protein domains in the ECOD database[110] were investigated using Foldseek easy-search module[108] (--tmscore-threshold 0.3 -e 0.001). TMHMM v2.0 was employed to predict transmembrane topology of RdhA proteins (https://services.healthtech.dtu.dk/services/TMHMM-2.0/).

Ledock (https://www.lephar.com/) was used to predict the binding poses of 191 halohydrocarbons on different RDases (RMSD: 1.0, Number of binding poses: 20, size: 10). Prediction of ligand channels was performed on the Caver v3.0 (probe_radius 0.9, shell_radius 3, shell_depth 4)[111]. All of the structures were visualized and exported as images using PyMOL (http://www.pymol.org).

Nucleotide diversity and pN/pS ratio analyses

All metagenomic filtered reads from each sample were mapped to an indexed database of the *rdhA*-containing genomes using Bowtie2 (v2.3.5.1; default parameters)[112]. The nucleotide diversity and pN/pS ratio of *rdhA* genes were calculated from these mappings using the profile module of the inStrain program (v1.6.2; default parameters)[35] at the gene level. To perform gene level profiling, genes were predicted by the software Prodigal (v2.6.3; settings: -p meta)[96] for each MAG carrying *rdhA* genes. A total of 475 *rdhA* genes were retained for microdiversity analyses, which satisfied threshold criteria of having a breadth of 50% and 5× coverage.

Structure-based polymorphism analyses

The variants across the RdhA protein structures were explored using the microbial population genetics framework implemented in anvi'o (v7.1)[113]. First, population statistics of *rdhA* genes, including coverage, single nucleotide variants (SNVs) and single codon variants (SCVs), were calculated from the mappings using the profile module of the anvi'o program (v7.1; default parameters). Only SNV positions

mapping with greater than 10× coverage and the average coverage of *rdhA* genes exceeded 2× in each metagenome were retained (reducing our metagenomic samples size from 165 to 30). Additionally, *rdhA* genes that are present in a minimum of 20 out of 30 metagenomes were included for further analysis. The remaining 200 *rdhA* genes with high-quality structure (pLDDT > 80) were imported for structure-based polymorphism analyses. The SCV data of each site on the high-quality structure (pLDDT > 80) was integrated with 'anvi-display-structure', which filtered for variants that had at least 0.05 departure from consensus. The producing results including synonymous (s) and nonsynonymous (ns) polymorphism rates of each codon ($pS^{(site)}$ and $pN^{(site)}$) and relative solvent accessibility (RSA). Ligand-binding residues of each gene were predicted with per-residue binding frequencies greater than 0.5 using InteracDome. Distance-to-ligand (DTL) was calculated using append_dist_to_lig.py (https://merenlab.org/data/anvio-structure/chapter-III/). The relationship between SCVs, relative solvent accessibility (RSA), and distance to ligand (DTL) in reductive dehalogenases were explored.

Statistical analyses

Statistical analyses were carried out in R v4.2.3. The normality and variance homogeneity of the data were evaluated using Shapiro-Wilk Test and Levene's test, respectively. To compare gene abundance and evolutionary metrics among different *rdhA* groups, the Kruskal-Wallis rank-sum test was employed. For paired comparisons within *rdhA* groups, the Wilcoxon test was utilized. Pearson's product-moment correlation and linear regression were performed to assess the relationship between gene abundance, evolutionary metrics and their relationships with sediment depth. ANOVA analysis was performed to determine the portion of variability in the polymorphism data and the Pearson coefficients for each gene-sample pair were calculated using the R script available at https://merenlab.org/data/anvio-structure/chapter-IV/.

Data availability

The non-redundant gene and MAGs catalogs derived form 165 cold seep metagenomes can be found in figshare (https://doi.org/10.6084/m9.figshare.22568107). The *rdhA*-carrying MAGs and phylogenetic trees of RdhA based on amino acid sequences and protein structures are available at figshare (https://doi.org/10.6084/m9.figshare.23499363).

Code Availability

The present study did not generate codes, and mentioned tools used for the data analysis were applied with default parameters unless specified otherwise.

Reference

1. Dong, X. et al. Metabolic potential of uncultured bacteria and archaea associated with petroleum seepage in deep-sea sediments. *Nat. Commun.* **10**, 1816 (2019).
2. Joye, S.B. The geology and biogeochemistry of hydrocarbon seeps. *Annu. Rev. Earth Pl. Sci.* **48**, 205-231 (2020).
3. Dong, X. et al. Thermogenic hydrocarbon biodegradation by diverse depth-stratified microbial populations at a Scotian Basin cold seep. *Nat. Commun.* **11**, 5825 (2020).
4. Kleindienst, S. et al. Diverse sulfate-reducing bacteria of the Desulfosarcina/Desulfococcus clade are the key alkane degraders at marine seeps. *ISME J* **8**, 2029-2044 (2014).
5. Lyu, L. et al. Bioaccumulation of emerging persistent organic pollutants in the deep-sea cold seep ecosystems: Evidence from chlorinated paraffin. *J Hazard Mater* **445**, 130472 (2022).
6. Zhang, C., Atashgahi, S., Bosma, T.N.P., Peng, P. & Smidt, H. Organohalide respiration potential in marine sediments from Aarhus Bay. *FEMS Microbiol Ecol* **98** (2022).
7. Yang, Y. et al. Roles of Organohalide-Respiring Dehalococcoidia in Carbon Cycling. *mSystems* **5** (2020).
8. Spang, A. et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol* **4**, 1138-1148 (2019).
9. Schubert, T., Adrian, L., Sawers, R.G. & Diekert, G. Organohalide respiratory chains: composition, topology and key enzymes. *FEMS Microbiol Ecol* **94** (2018).
10. Leri, A.C. et al. A marine sink for chlorine in natural organic matter. *Nat Geosci* **8**, 620-624 (2015).
11. Futagami, T., Morono, Y., Terada, T., Kaksonen, A.H. & Inagaki, F. Distribution of dehalogenation activity in subseafloor sediments of the Nankai Trough subduction zone. *Philos T R Soc B* **368** (2013).
12. Dolfing, J. in Organohalide-Respiring Bacteria. (eds. L. Adrian & F.E. Löffler) 31-48

(Springer Berlin Heidelberg, Berlin, Heidelberg; 2016).

13. Wang, S. et al. Electron transport chains in organohalide-respiring bacteria and bioremediation implications. *Biotechnol Adv* **36**, 1194-1206 (2018).
14. Liu, R. et al. Novel Chloroflexi genomes from the deepest ocean reveal metabolic strategies for the adaptation to deep-sea habitats. *Microbiome* **10** (2022).
15. Fincker, M. et al. Metabolic strategies of marine subseafloor Chloroflexi inferred from genome reconstructions. *Environmental Microbiology* **22**, 3188-3204 (2020).
16. Vuillemin, A., Kerrigan, Z., D'Hondt, S. & Orsi, W.D. Exploring the abundance, metabolic potential and gene expression of subseafloor Chloroflexi in million-year-old oxic and anoxic abyssal clay. *FEMS Microbiology Ecology* **96** (2020).
17. Liu, Y., Wang, L., Liu, R. & Fang, J. Biogeochemical cycling of halogenated organic compounds in the ocean: Current progress and future directions. *Deep Sea Research Part I: Oceanographic Research Papers*, 104237 (2024).
18. Bommer, M. et al. Structural basis for organohalide respiration. *Science* **346**, 455-458 (2014).
19. Fincker, M. & Spormann, A.M. Biochemistry of Catabolic Reductive Dehalogenation. *Annu Rev Biochem* **86**, 357-386 (2017).
20. Turkowsky, D. et al. An integrative overview of genomic, transcriptomic and proteomic analyses in organohalide respiration research. *FEMS Microbiol Ecol* **94**, fiy013 (2018).
21. Kublik, A. et al. Identification of a multi-protein reductive dehalogenase complex in *Dehalococcoides mccartyi* strain CBDB1 suggests a protein-dependent respiratory electron transport chain obviating quinone involvement. *Environmental Microbiology* **18**, 3044-3056 (2016).
22. Hartwig, S. et al. A H_2-oxidizing, 1,2,3-trichlorobenzene-reducing multienzyme complex isolated from the obligately organohalide-respiring bacterium *Dehalococcoides mccartyi* strain CBDB1. *Environmental Microbiology Reports* **9**, 618-625 (2017).
23. Atashgahi, S. Discovered by genomics: putative reductive dehalogenases with N-terminus transmembrane helixes. *FEMS Microbiol Ecol* **95** (2019).
24. Liu, J. & Haggblom, M.M. Genome-Guided Identification of Organohalide-Respiring Deltaproteobacteria from the Marine Environment. *Mbio* **9** (2018).
25. Jochum, L.M. et al. Single-Cell Genomics Reveals a Diverse Metabolic Potential of Uncultivated Desulfatiglans-Related Deltaproteobacteria Widely Distributed in Marine Sediment. *Front Microbiol* **9**, 2038 (2018).
26. Fisher, K. et al. Efficient NADPH-dependent dehalogenation afforded by a self-sufficient reductive dehalogenase. *J Biol Chem* **299**, 105086 (2023).
27. Payne, K.A. et al. Reductive dehalogenase structure suggests a mechanism for B12-dependent dehalogenation. *Nature* **517**, 513-516 (2015).
28. Molenda, O. et al. 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 Proc Imp* **22**, 663-678 (2020).
29. Cimmino, L. et al. Structure of a membrane-bound menaquinol:organohalide oxidoreductase. *Nat Commun* **14**, 7038 (2023).
30. Bommer, M. et al. Structural basis for organohalide respiration. *Science* **346**, 455-458 (2014).
31. Halliwell, T., Fisher, K., Payne, K.A.P., Rigby, S.E.J. & Leys, D. Catabolic Reductive Dehalogenase Substrate Complex Structures Underpin Rational Repurposing of Substrate

Scope. *Microorganisms* **8** (2020).

32. Dong, X. et al. Evolutionary ecology of microbial populations inhabiting deep sea sediments associated with cold seeps. *Nat Commun* **14**, 1127 (2023).
33. Anderson, R.E. et al. Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat Commun* **8**, 1114 (2017).
34. Sjoqvist, C., Delgado, L.F., Alneberg, J. & Andersson, A.F. Ecologically coherent population structure of uncultivated bacterioplankton. *ISME J* **15**, 3034-3049 (2021).
35. Olm, M.R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat Biotechnol* **39**, 727-736 (2021).
36. Berg, R.D. & Solomon, E.A. Geochemical constraints on the distribution and rates of debromination in the deep subseafloor biosphere. *Geochimica et Cosmochimica Acta* **174**, 30-41 (2016).
37. Ji, F. et al. Geochemistry of hydrothermal vent fluids and its implications for subsurface processes at the active Longqi hydrothermal field, Southwest Indian Ridge. *Deep Sea Research Part I: Oceanographic Research Papers* **122**, 41-47 (2017).
38. Millero, F.J., Feistel, R., Wright, D.G. & McDougall, T.J. The composition of Standard Seawater and the definition of the Reference-Composition Salinity Scale. *Deep Sea Research Part I: Oceanographic Research Papers* **55**, 50-72 (2008).
39. Hu, C.-Y. et al. Biogeochemical cycles at the sulfate-methane transition zone (SMTZ) and geochemical characteristics of the pore fluids offshore southwestern Taiwan. *Journal of Asian Earth Sciences* **149**, 172-183 (2017).
40. Han, Y. et al. A comprehensive genomic catalog from global cold seeps. *Scientific Data* **10** (2023).
41. Peng, P. et al. Organohalide-respiring Desulfoluna species isolated from marine environments. *ISME J* **14**, 815-827 (2020).
42. Cimmino, L. et al. Structure of a membrane-bound menaquinol:organohalide oxidoreductase. *Nature Communications* **14**, 7038 (2023).
43. Atashgahi, S. et al. Geochemical and microbial community determinants of reductive dechlorination at a site biostimulated with glycerol. *Environ Microbiol* **19**, 968-981 (2017).
44. Müller, A.L., Kjeldsen, K.U., Rattei, T., Pester, M. & Loy, A. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *The ISME Journal* **9**, 1152-1165 (2015).
45. Zhang, C. et al. Unexpected genetic and microbial diversity for arsenic cycling in deep sea cold seep sediments. *NPJ Biofilms Microbi* **9**, 13 (2023).
46. Wang, Y., Wegener, G., Hou, J., Wang, F. & Xiao, X. Expanding anaerobic alkane metabolism in the domain of Archaea. *Nature Microbiology* **4**, 595-602 (2019).
47. Wang, Y., Wegener, G., Ruff, S.E. & Wang, F. Methyl/alkyl-coenzyme M reductase-based anaerobic alkane oxidation in archaea. *Environ Microbiol* **23**, 530-541 (2021).
48. Xu, G., Zhao, X., Zhao, S., Rogers, M.J. & He, J. Salinity determines performance, functional populations, and microbial ecology in consortia attenuating organohalide pollutants. *The ISME Journal* **17**, 660-670 (2023).
49. Chen, Z., Tom, N.P.B., Siavash, A. & Hauke, S. Genome-resolved transcriptomics reveals novel organohalide-respiring bacteria from Aarhus Bay sediments. *bioRxiv*, 2023.2004.2017.537210 (2023).

50. Chen, K. et al. Molecular characterization of the enzymes involved in the degradation of a brominated aromatic herbicide. *Molecular Microbiology* **89**, 1121-1139 (2013).
51. Zhang, C. et al. The majority of microorganisms in gas hydrate-bearing subseafloor sediments ferment macromolecules. *Microbiome* **11**, 37-37 (2023).
52. Deng, Z. et al. Marine Dehalogenator and Its Chaperones: Microbial Duties and Responses in 2,4,6-Trichlorophenol Dechlorination. *Environ Sci Technol* **57**, 11300-11312 (2023).
53. Lipson David, A., Raab Theodore, K., Pérez Castro, S. & Powell, A. Organohalide-Respiring Bacteria at the Heart of Anaerobic Metabolism in Arctic Wet Tundra Soils. *Applied and Environmental Microbiology* **87**, e01643-01620 (2021).
54. Sondergaard, D., Pedersen, C.N.S. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* **6** (2016).
55. Khot, V. et al. CANT-HYD: A Curated Database of Phylogeny-Derived Hidden Markov Models for Annotation of Marker Genes Involved in Hydrocarbon Degradation. *Frontiers in Microbiology* **12** (2022).
56. Zhang, C. et al. Marine sediments harbor diverse archaea and bacteria with the potential for anaerobic hydrocarbon degradation via fumarate addition. *FEMS Microbiol Ecol* **97**, fiab045 (2021).
57. Kunze, C. et al. Cobamide-mediated enzymatic reductive dehalogenation via long-range electron transfer. *Nat Commun* **8**, 15858 (2017).
58. Zhou, J. et al. VB_{12}Path for Accurate Metagenomic Profiling of Microbially Driven Cobalamin Synthesis Pathways. *mSystems* **6**, 10.1128/msystems.00497-00421 (2021).
59. Zhou, J. et al. The diversity and ecological significance of microbial traits potentially involved in B12 biosynthesis in the global ocean. *mLife* **2**, 416-427 (2023).
60. Wang, J., Zhu, Y.-G., Tiedje, J.M. & Ge, Y. Global biogeography and ecological implications of cobamide-producing prokaryotes. *The ISME Journal*, wrae009 (2024).
61. Liu, J., Adrian, L. & Häggblom Max, M. Transcriptomic and Proteomic Responses of the Organohalide-Respiring Bacterium Desulfoluna spongiiphila to Growth with 2,6-Dibromophenol as the Electron Acceptor. *Applied and Environmental Microbiology* **86**, e02146-02119 (2020).
62. Gnidehou, S. et al. Iodotyrosine dehalogenase 1 (DEHAL1) is a transmembrane protein involved in the recycling of iodide close to the thyroglobulin iodination site. *The FASEB Journal* **18**, 1574-1576 (2004).
63. Rocha, E.P.C. Neutral Theory, Microbial Practice: Challenges in Bacterial Population Genetics. *Mol Biol Evol* **35**, 1338-1347 (2018).
64. McBride, J.M. et al. AlphaFold2 Can Predict Single-Mutation Effects. *Physical Review Letters* **131**, 218401 (2023).
65. Dong, X. et al. Phylogenetically and catabolically diverse diazotrophs reside in deep-sea cold seep sediments. *Nat Commun* **13**, 4885 (2022).
66. Starnawski, P. et al. Microbial community assembly and evolution in subseafloor sediment. *PNAS* **114**, 2940-2945 (2017).
67. Kiefl, E. et al. Structure-informed microbial population genetics elucidate selective pressures that shape protein evolution. *Science Advances* **9**, eabq4632 (2023).
68. Smith, T.J. MOLView: A program for analyzing and displaying atomic structures on the Macintosh personal computer. *Journal of Molecular Graphics* **13**, 122-125 (1995).

69. Glass, J.B. et al. Microbial metabolism and adaptations in Atribacteria-dominated methane hydrate sediments. *Environ. Microbiol.* **23**, 4646-4660 (2021).
70. Yu, H. et al. Sulfate differentially stimulates but is not respired by diverse anaerobic methanotrophic archaea. *ISME J.* **16**, 168-177 (2022).
71. Laso-Perez, R. et al. Anaerobic degradation of non-methane alkanes by "*Candidatus* Methanoliparia" in hydrocarbon seeps of the Gulf of Mexico. *mBio* **10**, e01814-01819 (2019).
72. Zhao, R., Summers, Z.M., Christman, G.D., Yoshimura, K.M. & Biddle, J.F. Metagenomic views of microbial dynamics influenced by hydrocarbon seepage in sediments of the Gulf of Mexico. *Sci. Rep* **10**, 5772 (2020).
73. Ruff, S.E. et al. In situ development of a methanotrophic microbiome in deep-sea sediments. *ISME J.* **13**, 197-213 (2019).
74. Li, L. et al. Bacteria and archaea synergistically convert glycine betaine to biogenic methane in the Formosa cold seep of the South China Sea. *mSystems* **6**, e0070321 (2021).
75. Zhang, H. et al. Metagenome sequencing and 768 microbial genomes from cold seep in South China Sea. *Sci. Data* **9**, 480 (2022).
76. Lu, R. et al. Asgard archaea in the haima cold seep: Spatial distribution and genomic insights. *Deep-Sea Res. Pt. I* **170**, 103489 (2021).
77. Xiao, X. et al. Metal-driven anaerobic oxidation of methane as an important methane sink in methanic cold seep sediments. *Microbiol. Spectr.* **11**, e05337-05322 (2023).
78. Li, J. et al. Deep sea cold seep is an atmospheric Hg sink and MeHg source. *Research Square* https://doi.org/10.21203/rs.3.rs-2323106/v1**.** (2022).
79. Dong, X. et al. Phylogenetically and catabolically diverse diazotrophs reside in deep-sea cold seep sediments. *Nat. Commun.* **13**, 4885 (2022).
80. Li, W.-L., Wu, Y.-Z., Zhou, G.-w., Huang, H. & Wang, Y. Metabolic diversification of anaerobic methanotrophic archaea in a deep-sea cold seep. *Mar. Life Sci. Tech.* **2**, 431-441 (2020).
81. Li, W.L. et al. Microbial ecology of sulfur cycling near the sulfate-methane transition of deep-sea cold seep sediments. *Environ. Microbiol.* **23**, 6844-6858 (2021).
82. Han, Y. et al. A comprehensive genomic catalog from global cold seeps. *Sci Data* **10**, 596 (2023).
83. Kang, D.D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *Peerj* **7** (2019).
84. Wu, Y.W., Simmons, B.A. & Singer, S.W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).
85. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**, 1144-1146 (2014).
86. Pan, S., Zhu, C., Zhao, X.-M. & Coelho, L.P. A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments. *Nat. Commun.* **13**, 2326 (2022).
87. Nissen, J.N. et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat. Biotechnol.* **39**, 555-560 (2021).
88. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P. & Parks, D.H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927 (2019).
89. Parks, D.H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat.*

Biotechnol. **38**, 1079–1086 (2020).

90. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055 (2015).

91. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: A Fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).

92. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417-419 (2017).

93. Uritskiy, G.V., DiRuggiero, J. & Taylor, J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).

94. Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217 (2012).

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

96. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

97. Katoh, K. & Standley, D.M. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**, 1933-1942 (2016).

98. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

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

100. Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

101. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. **49**, W293-W296 (2021).

102. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**, 8883-8900 (2020).

103. Khot, V. et al. CANT-HYD: A Curated Database of Phylogeny-Derived Hidden Markov Models for Annotation of Marker Genes Involved in Hydrocarbon Degradation. *Frontiers in microbiology* **12**, 764058 (2021).

104. Bay, S.K. et al. Trace gas oxidizers are widespread and active members of soil microbial communities. *Nat Microbiol* **6**, 246-256 (2021).

105. Camargo, A.P. et al. Identification of mobile genetic elements with geNomad. *Nat Biotechnol* (2023).

106. Khedkar, S. et al. Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Research* **50**, 3155-3168 (2022).

107. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

108. van Kempen, M. et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* (2023).

109. Moi, D. et al. Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses. *bioRxiv*, 2023.2009.2019.558401 (2023).
110. Cheng, H. et al. ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol* **10**, e1003926 (2014).
111. Chovancova, E. et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **8**, e1002708 (2012).
112. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
113. Kiefl, E. et al. Structure-informed microbial population genetics elucidate selective pressures that shape protein evolution. *Sci Adv* **9**, eabq4632 (2023).

Acknowledgements

The work was supported by the National Science Foundation of China (No. 92351304 and No. 32300100), the Natural Science Foundation of Fujian Province (No. 2023J06042), Scientific Research Foundation of Third Institute of Oceanography, MNR (No. 2022025 and No. 2023022), State Key Laboratory of Marine Geology, Tongji University (No. MGK202303), and China Postdoctoral Science Foundation (2023M734096). We thank Qiuyun Jiang, Jing Liao, Chengpeng Li, Zhaochao Deng, Fabai Wu and Xinyue Liu for geochemical analyses and helpful discussions.

Author contributions

X.D. and Y.H. designed this study. Y.H. and Y.P. performed the omics analysis. M.W. and J.P. contributed to metabolomic data. L.C. and M.W. provided extensive geochemical data for both pore and overlying waters. D.Z. and Y.X. measured total organic halogen of cold seep sediments. Y.Y., H.Z., C.Z., and C.G. are actively engaged in discussions and data interpretations. Y.H., Y.P., C.G., and X.D. wrote the paper, with input from other authors.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. Halogen profiles of cold seep porewater and sediments. (a) Concentration profiles of dissolved and solid chlorine (Cl^-) and bromine (Br^-) in 1086 porewater samples and 68 sediment samples collected from Qiongdongnan, Shenhu, Haima, and Site F cold seeps. Blue lines and red lines represent the average concentration of Cl^- and Br^- in the typical seawater and overlying water of cold seeps, respectively. (b) Concentration profiles of total organic halogen in sediments collected from Qiongdongnan and Shenhu cold seeps. (c) The chemical structures of halogenated compounds identified in sediments from Qiongdongnan and Shenhu cold seeps. Detailed data on halogen profiles of cold seep porewater and sediments can be found in **Supplementary Table 1-3**.

Figure 2. Maximum-likelihood phylogenetic tree of reductive dehalogenase (RdhA) identified from the cold seep non-redundant gene catalog. Branches are color-coded to represent different RdhA subgroups. Epoxyqueuosine reductase (QueG) was used as the outgroup. Scale bar indicates the mean number of substitutions per site. Bootstrap values over 50% were shown next to the nodes. The n values denote the count of RdhA sequences within each subgroup.

Figure 3. Relative abundance patterns of *rdhA* genes in cold seep sediments. (a) Relative abundance and expression levels of 3,993 *rdhA* genes across different sediment samples, measured as GPM for metagenomes and TPM for transcripts. Insert plots show the *rdhA* gene abundance and expression levels of the four groups. Gene groups and subgroups are in different colors. *P* values of differences across different types of *rdhA* genes were computed through Kruskal-Wallis rank-sum tests. Relationships between *rdhA* abundance and (b) sediment depths (mbsf), (c) reductive *dsrA* abundance, and (d) oxidative *mcrA* abundance. Each point represents the average gene abundance for a sample, accompanied by linear regression lines and R^2 values specific to each gene type in corresponding colors. Detailed statistics for significance

tests and linear regressions are provided in **Supplementary Tables 5-7**.

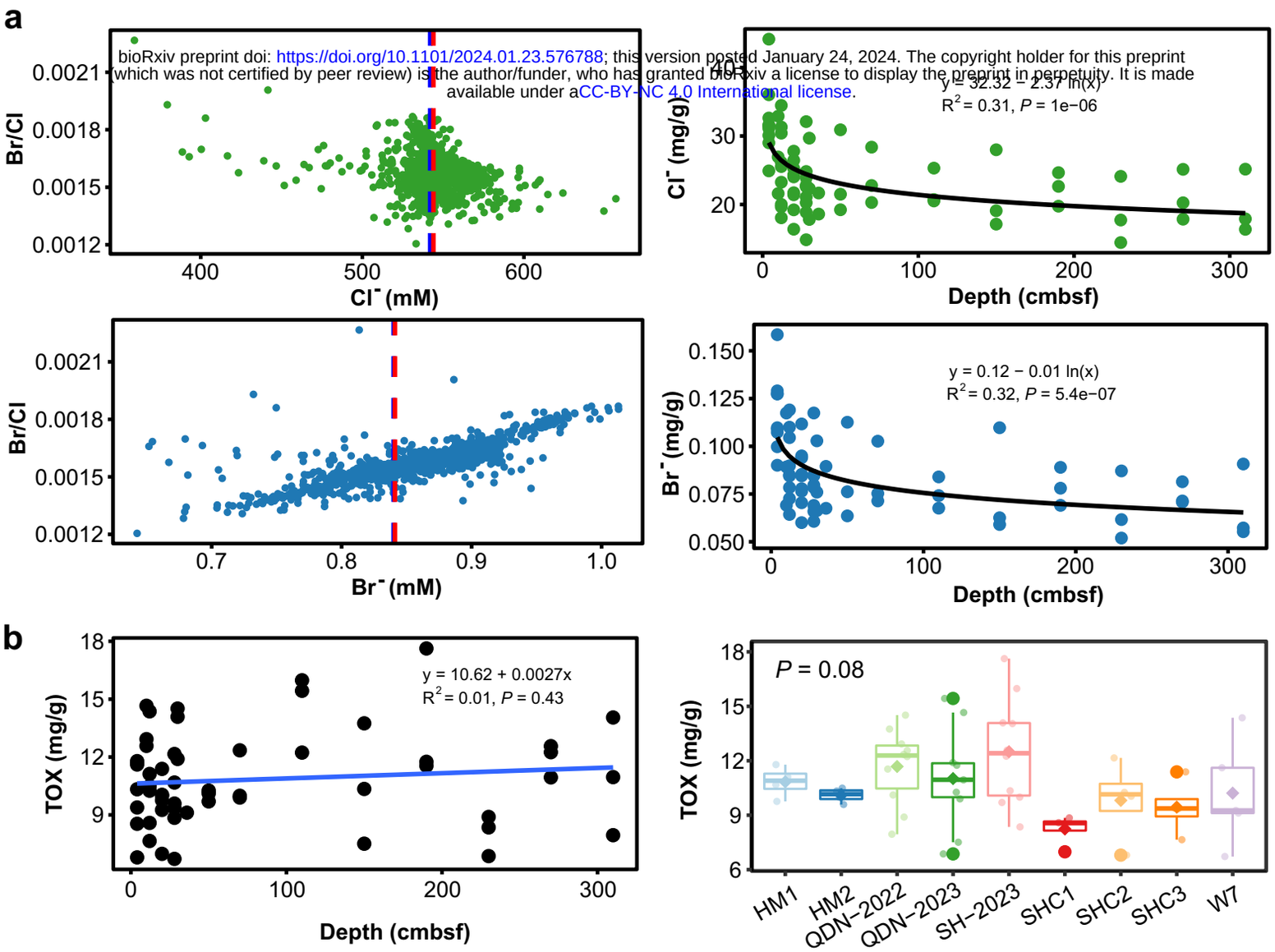
Figure 4. Taxonomic classification, genomic context and potential substrate docking of cold seep organohalide-reducing microorganisms. (a) A Sankey diagram showing the taxonomy of archaeal and bacterial MAGs according to GTDB, categorizing organohalide-reducers across taxonomic levels. (b) Genome synteny of 14 contigs showcasing five types of RDases (highlighted in red) and accompanying genes: regulatory (green), transport-related (orange), functional (blue), mobile genetic elements (MGEs, yellow), other annotated genes (purple), and unknown (grey). (c) Overlay of the active sites of cold seep quinone-dependent respiratory and cytosolic RdhA (lightpink) against known RdhA crystal structures (PDB id: 8Q4H and PDB id: 6ZY1, in green). Substrates 2'-Chloro-biphenyl-2,3-diol and 1,5-dibromopentane, predicted through molecular docking with binding affinities of -4.12 and -3.11 kcal/mol respectively, are shown in magenta. Details for taxonomic classification, molecular docking, and *rdhA*-containing contig annotations are provided in **Supplementary Tables 9 and 11-13**.

Figure 5. Reductive dehalogenation is closely linked to carbon, nitrogen, sulfur, and trace element metabolic processes. RDases of different colors represent different types, with orange indicating prototypical respiratory RDases, purple representing transmembrane respiratory RDases, green indicating cytosolic RDases, and cyan signifying the novel clade of RDases. Dark orange denotes potential electron transfer units, while orange-yellow represents potential electron donors and electron transfer pathways. Different background colors signify distinct metabolic pathways: light blue indicates hydrocarbon degradation, dark blue represents sulfur metabolism, light red represents nitrogen metabolism and olive green denotes B_{12} biosynthesis. Detailed annotations for different genes are provided in **Supplementary Tables 14-19.**

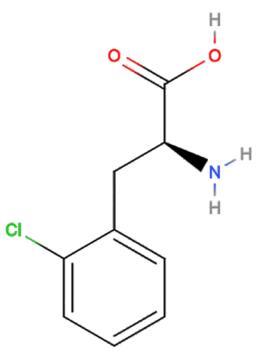
Figure 6. Structure-based phylogeny of cold seep reductive dehalogenases and reference proteins. Each group of reductive dehalogenases (RDases) is labeled by a unique color: Novel clade RDases, turquoise; Quinone-independent respiratory

RDases, gold; Quinone-dependent respiratory RDases, orange; Cytosolic RDases, green; Transmembrane respiratory RdhAs, purple. The three RDases that have been validated experimentally are highlighted with stars. Surrounding the tree, the AlphaFold-predicted 3D structures of representative RDases are displayed, each labeled with their taxonomic phylum and corresponding pLDDT scores. The indication of the five main inferred evolutionary events of the tree marked at the origin of each fold/architecture, numbered 1 through 5.

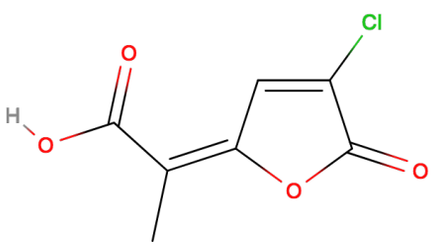
Figure 7. **Microdiversity of cold seep *rdhA* genes and their corresponding protein structural features.** (a) Comparison of nucleotide diversity and pN/pS for *rdhA* genes across gene groups. (b) Gene-wide distributions and Pearson correlations for $pN^{(site)}$ (red) and $pS^{(site)}$ (blue) relative to relative solvent accessibility (RSA) and distance to ligand (DTL). The pN(site) and pS(site) distribution were generated based on RSA and DTL values of each site from 200 predicted RdhA structures. The Pearson correlations of $\log_{10}(pN^{(site)})$ and $\log_{10}(pS^{(site)})$ to RSA and DTL for each gene-sample pair were calculated using linear models, and the mean values were represented as dashed lines. P values of differences across different types of *rdhA* genes were computed through Kruskal-Wallis rank-sum tests. The boxplot features include: center lines, medians; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range from the 25th and 75th percentiles; points, outliers. The n values denote the number of independent results utilized for statistical derivation. Detailed data on the microdiversity of *rdhA* genes can be found in **Source data 1-2.**



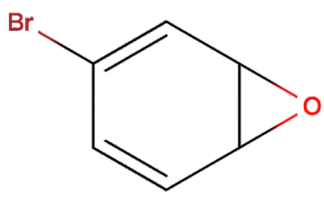
c



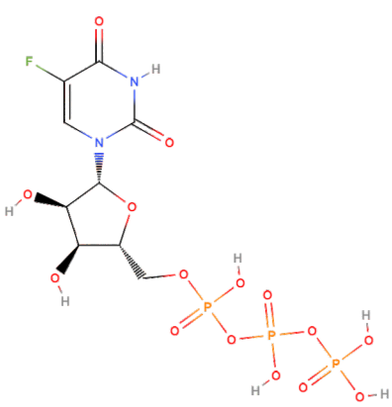
2-Chloro-L-phenylalanine



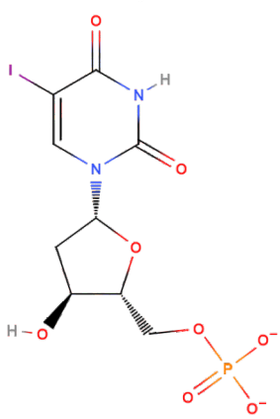
2-Chloro-5-methyl-cis-dienelactone



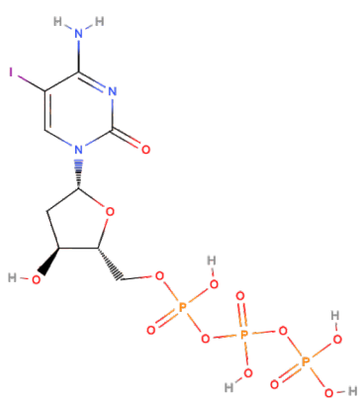
Bromobenzene-3,4-oxide



5-Fluorouridine-5'-triphosphate



5-iodo-2'-dUMP



5-iodo-dCTP

Transmembrane respiratory group1 RdhA
n = 1,480

Transmembrane respiratory group2 RdhA
n = 483

QueG

Novel clade RdhA
n = 202

Cytosolic RdhA
n = 362

Quinone-dependent
prototypical respiratory RdhA

Tree scale: 1

bootstrap

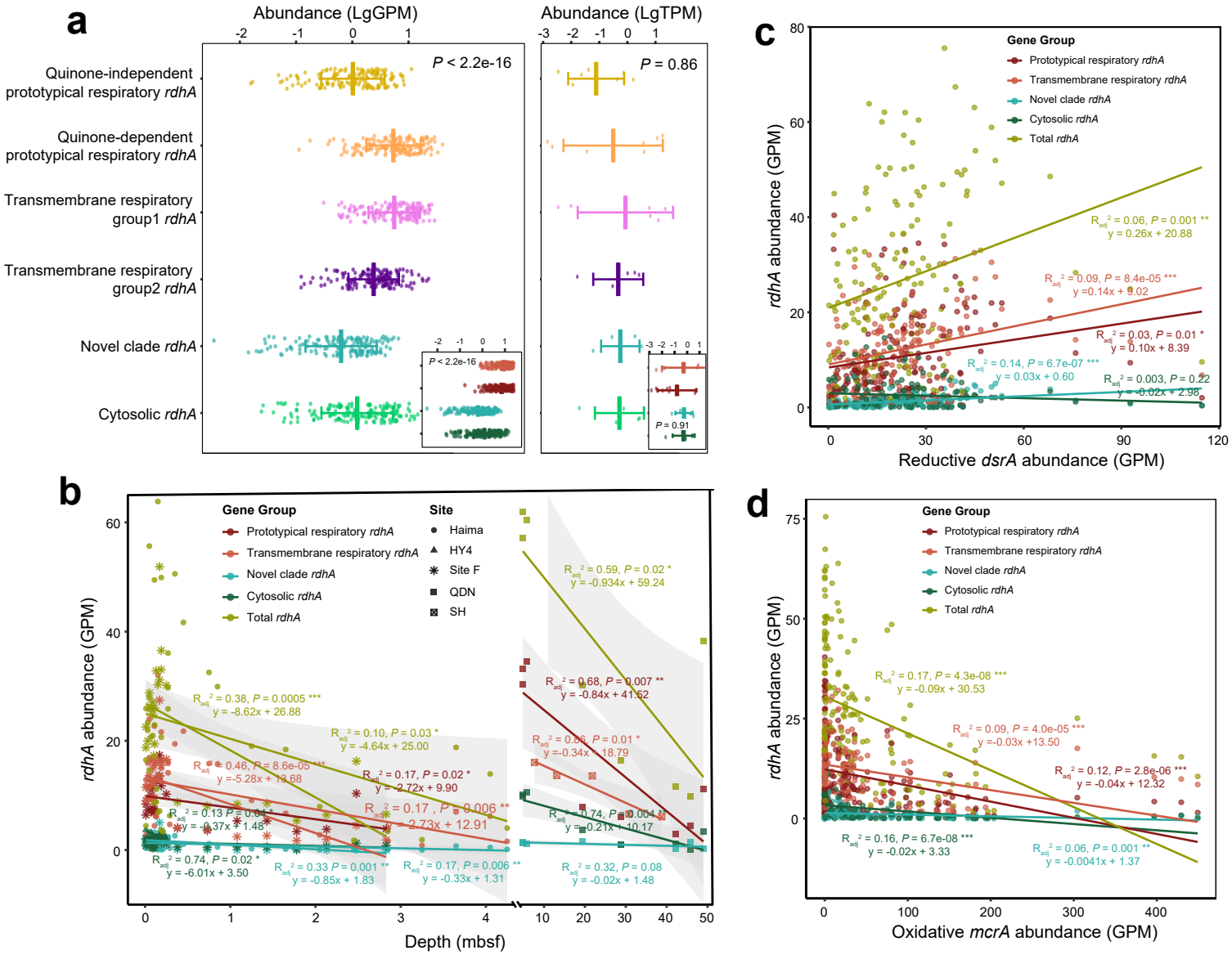
- 50
- 63
- 75
- 88
- 100

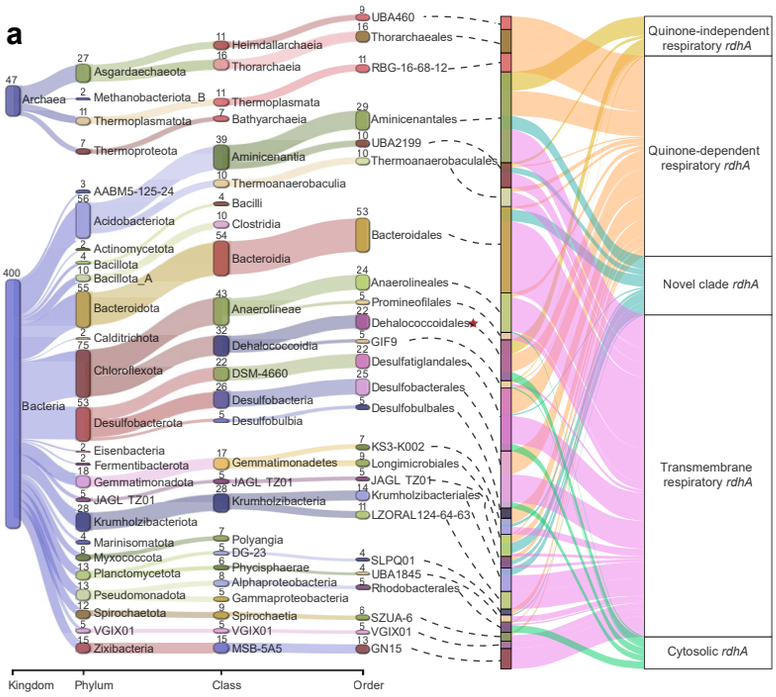
Quinone-independent
prototypical respiratory RdhA
n = 273

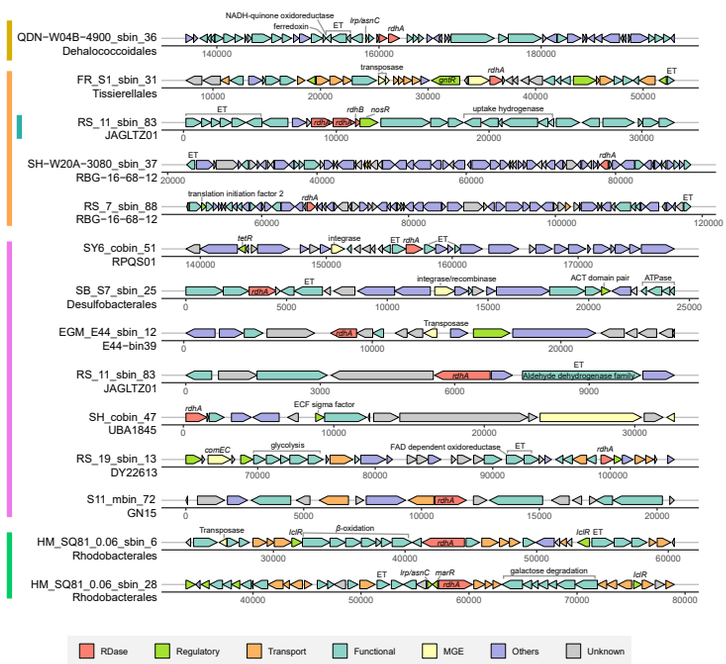
Quinone-dependent prototypical respiratory RdhA
n = 1,193

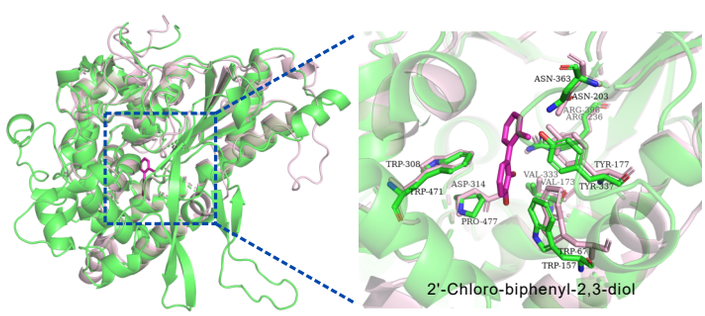
Gene Group

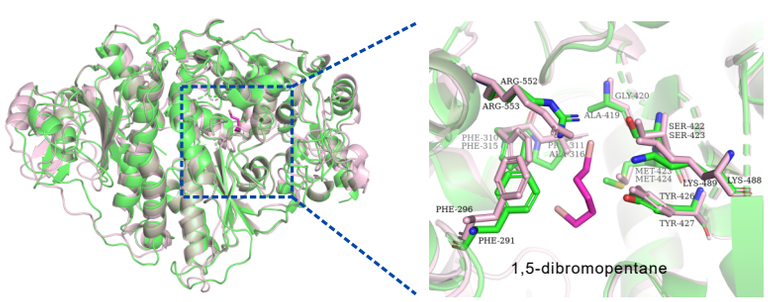
- Prototypical respiratory RdhA
- Transmembrane respiratory RdhA
- Novel clade RdhA
- Cytosolic RdhA

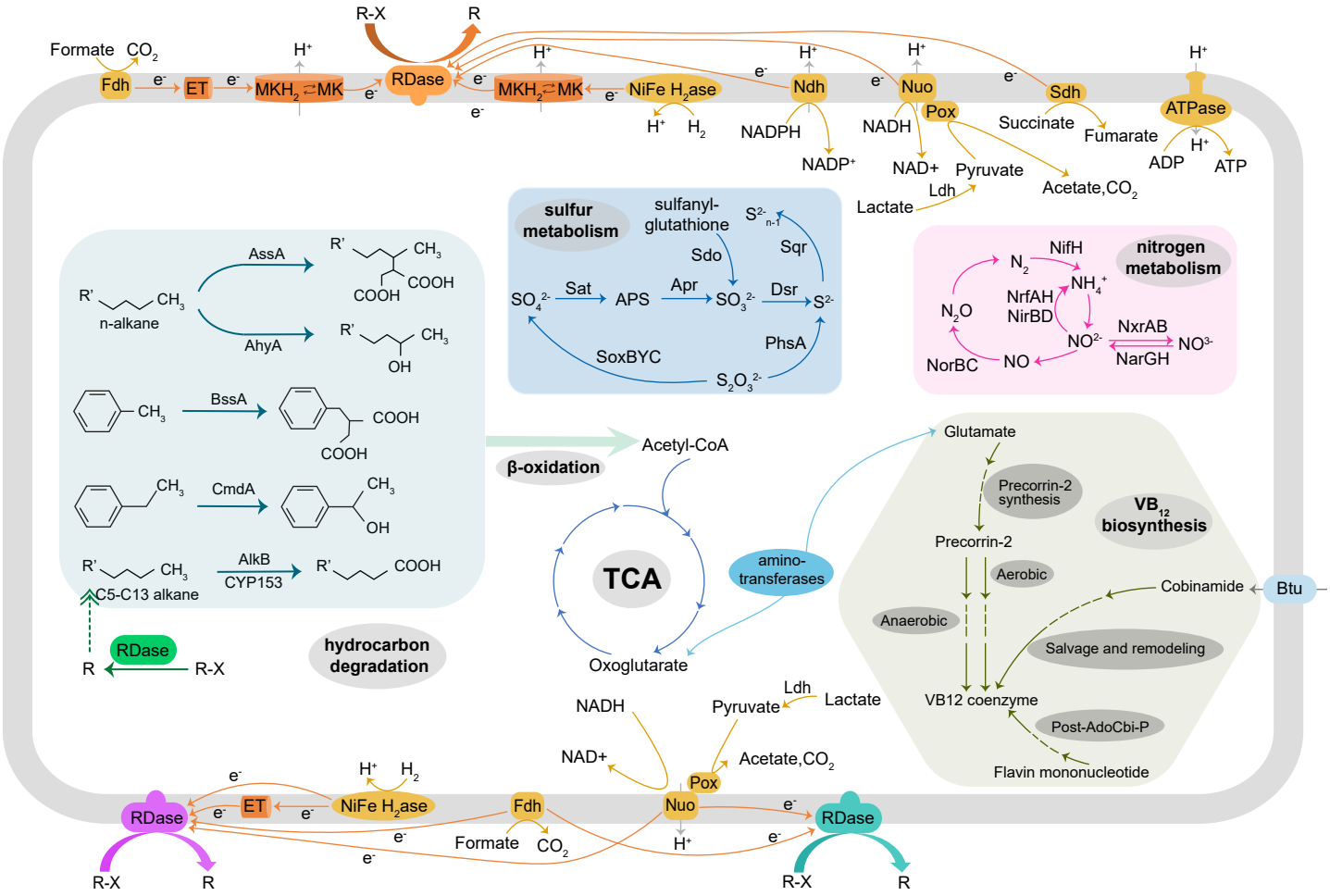


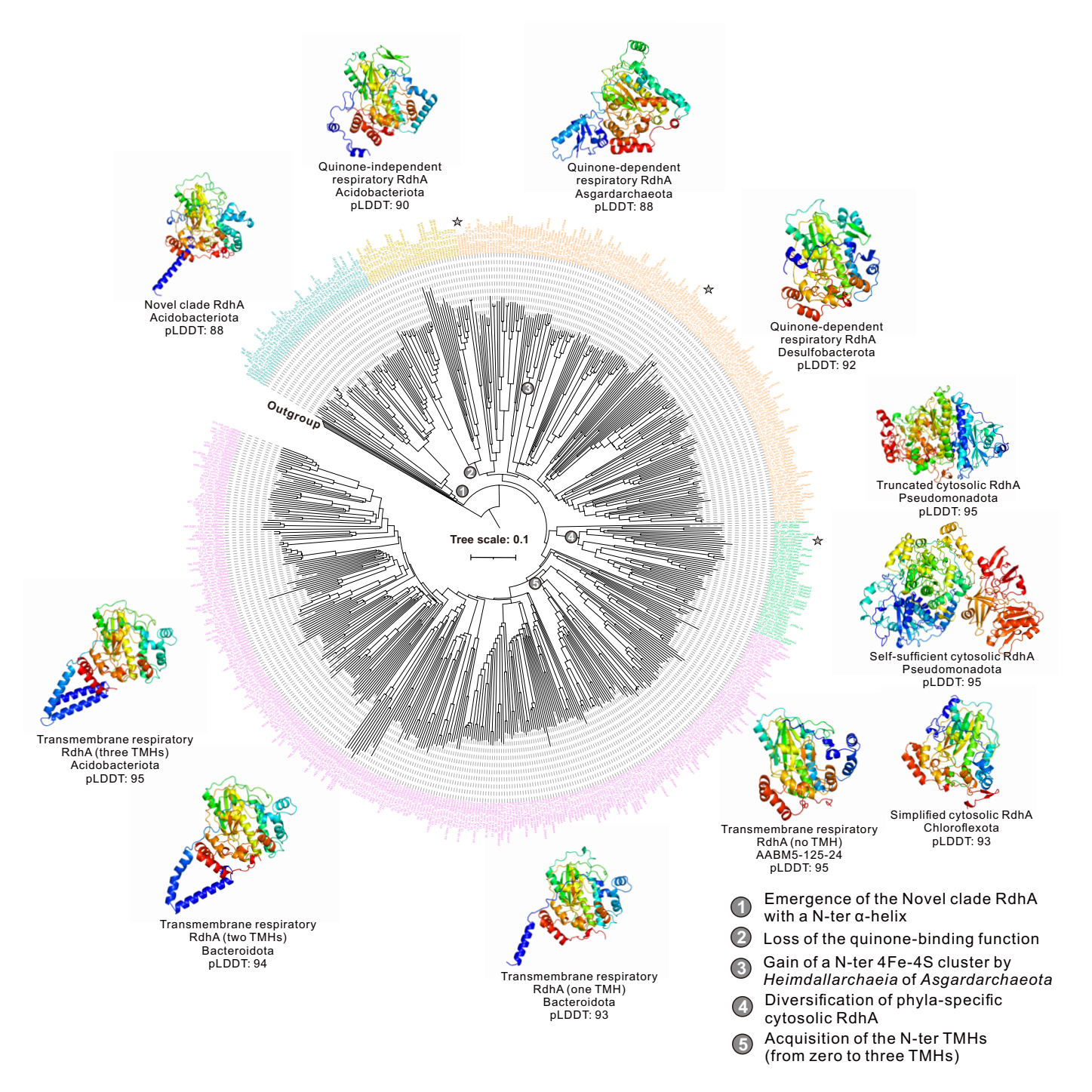


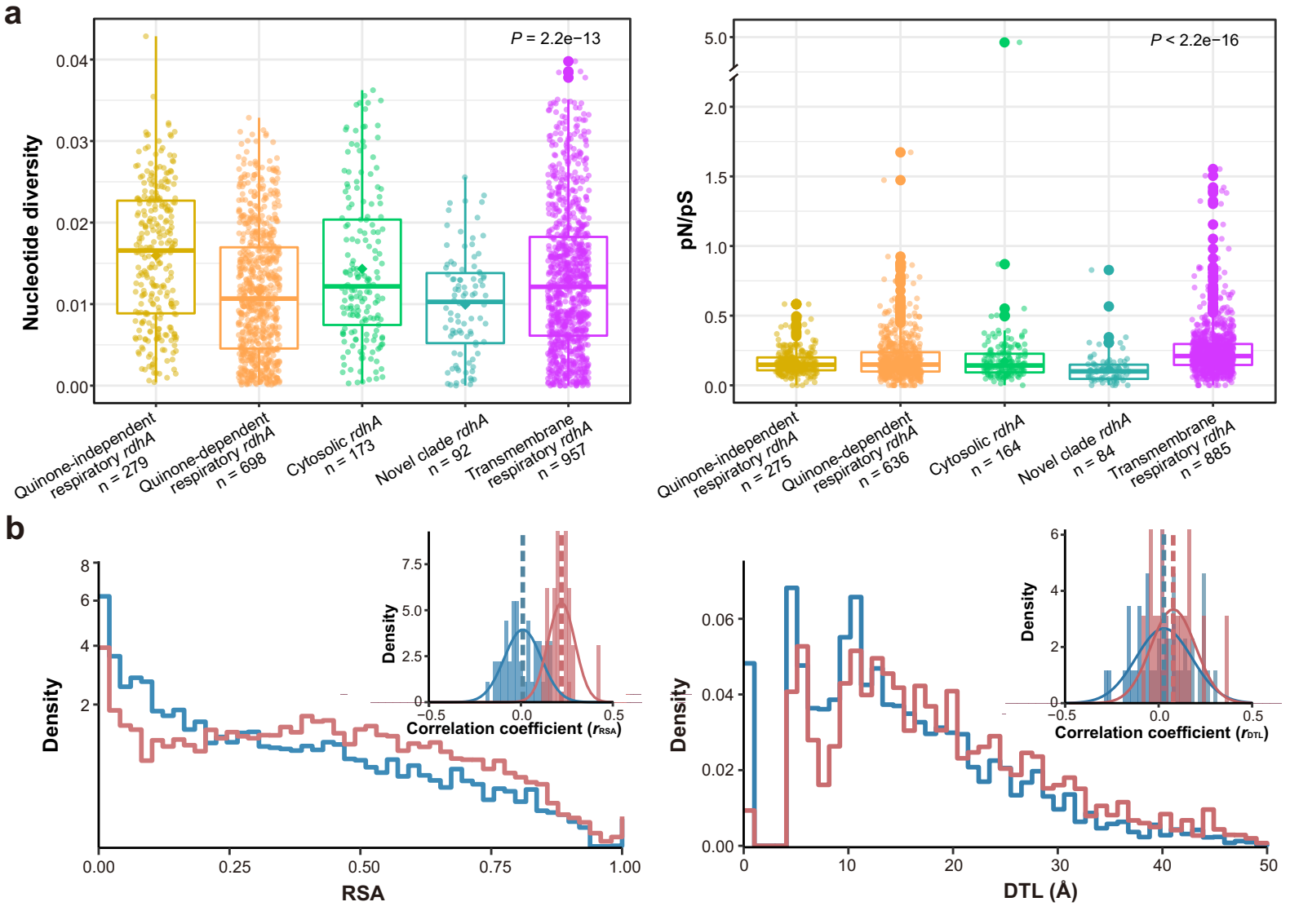









Quinone-independent
respiratory RdhA
Acidobacteriota
pLDDT: 90
Quinone-dependent
respiratory RdhA
Asgardarchaeota
pLDDT: 88
Novel clade RdhA
Acidobacteriota
pLDDT: 88
Quinone-dependent
respiratory RdhA
Desulfobacterota
pLDDT: 92
Outgroup
Tree scale: 0.1
1
2
3
4
5
Truncated cytosolic RdhA
Pseudomonadota
pLDDT: 95
Self-sufficient cytosolic RdhA
Pseudomonadota
pLDDT: 95
Transmembrane respiratory
RdhA (three TMHs)
Acidobacteriota
pLDDT: 95
Transmembrane respiratory
RdhA (no TMH)
AABM5-125-24
pLDDT: 95
Simplified cytosolic RdhA
Chloroflexota
pLDDT: 93
Transmembrane respiratory
RdhA (two TMHs)
Bacteroidota
pLDDT: 94
Transmembrane respiratory
RdhA (one TMH)
Bacteroidota
pLDDT: 93
1 Emergence of the Novel clade RdhA with a N-ter α-helix
2 Loss of the quinone-binding function
3 Gain of a N-ter 4Fe-4S cluster by *Heimdallarchaeia* of *Asgardarchaeota*
4 Diversification of phyla-specific cytosolic RdhA
5 Acquisition of the N-ter TMHs (from zero to three TMHs)


a
Nucleotide diversity
P = 2.2e−13
0.00
0.01
0.02
0.03
0.04
Quinone-independent respiratory rdhA n = 279
Quinone-dependent respiratory rdhA n = 698
Cytosolic rdhA n = 173
Novel clade rdhA n = 92
Transmembrane respiratory rdhA n = 957
pN/pS
P < 2.2e−16
0.0
0.5
1.0
1.5
2.0
5.0
Quinone-independent respiratory rdhA n = 275
Quinone-dependent respiratory rdhA n = 636
Cytosolic rdhA n = 164
Novel clade rdhA n = 84
Transmembrane respiratory rdhA n = 885
b
Density
RSA
0.00
0.25
0.50
0.75
1.00
Correlation coefficient (rRSA)
−0.5
0.0
0.5
DTL (Å)
0
10
20
30
40
50
Correlation coefficient (rDTL)