bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Evolutionary ecology of microbial populations inhabiting deep sea sediments associated with cold seeps

Xiyang Dong[1, 3 # *], Yongyi Peng[1, 2 #], Muhua Wang[2, 3], Laura Woods[4], Wenxue Wu[2, 3], Yong Wang[5], Xi Xiao[6], Jiwei Li[7], Kuntong Jia[2], Chris Greening[4], Zongze Shao[1, 3 *], Casey R.J. Hubert[8]

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

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

[5] Institute for Ocean Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

[6] Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou 510075, China

[7] Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

[8] Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada

[#] These authors contributed equally to this work.

* Correspondence can be addressed to Xiyang Dong (dongxiyang@tio.org.cn) and Zongze Shao (shaozz@163.com).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Deep sea cold seep sediments host abundant and diverse bacterial and archaeal populations that significantly influence biogeochemical cycles. While numerous studies have revealed the community structure and functional capabilities of cold seep microbiomes, little is known about their genetic heterogeneity within species. Here, we examined intraspecies diversity patterns of 39 abundant species identified in sediment layers down to 4.3 mbsf across six cold seep sites from around the world. These species were predicted to participate in methane oxidation and sulfate reduction, and based on their metabolic capabilities, grouped as aerobic methane-oxidizing bacteria (MOB), anaerobic methanotrophic archaea (ANME) and sulfate-reducing bacteria (SRB). These physiologically and phylogenetically diverse MOB, ANME and SRB display different degrees of intrapopulation sequence divergence and different evolutionary trajectories. Populations were in general characterized by low rates of homologous recombination and strong purifying selection with most of the nucleotide variation being synonymous. Functional genes related to methane (*pmoA* and *mcrA*) and sulfate (*dsrA*) metabolisms were found to be under strong purifying selection in the vast majority of species investigated, although examples of active positive selection were also observed. These genes differed in evolutionary trajectories across phylogenetic clades but are functionally conserved across cold seep sites. Intrapopulation diversification of MOB, ANME and SRB species as well as their *mcrA* and *dsrA* genes was observed to be depth-dependent and undergo divergent selection pressures throughout the sediment column. These results highlight the role of the interplay between ecological processes and the evolution of key bacteria and archaea in deep sea cold seep sediments and shed light on how microbial populations adapt in the subseafloor biosphere.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Cold seeps are widely distributed along continental margins across the globe and likely have an important environmental influence on the local biology, chemistry, and geology[1-4]. These seep fluids are enriched in hydrocarbon gases (mainly methane) and other energy rich hydrocarbon fluids, and markedly alter the sedimentary microbial community structure and function by promoting microbial growth, specialization, and adaptation[1]. Based on amplicon sequencing of genetic markers and genome-resolved metagenomics, numerous studies have revealed the extensive macrodiversity (i.e. the measure of population diversity within a community[5]) of archaeal and bacterial lineages and microbial metabolic versatility along different cold seep sites and sediment depths[1, 6-9]. Findings include the discovery of various lineages of archaeal anaerobic methanotrophs (ANME) and sulfate-reducing bacteria (SRB) as syntrophic aggregates to perform methane oxidation coupled to sulfate reduction at anoxic sediment layers[10, 11]. In the upper oxic sediment layers of cold seeps, methane is mostly consumed by aerobic methane-oxidizing bacteria (MOB) mainly from the order *Methylococcales* [9, 12]. Although macro-variations in microbial diversity and functions have been well-characterized, our knowledge of the microdiversity (i.e. the measure of genetic variation within a population[13-15]) of these metabolically and taxonomically diverse subseafloor microorganisms remains limited. By addressing crucial questions such as “what populations have high microdiversity levels”, “which genes are under selection” and “does homologous recombination occur”, intra-population microdiversity analyses can provide a more complete understanding of microbial ecology and evolution in the subseafloor biosphere[16, 17].

Traditional cultivation-based approaches have a fundamental role in studying genetic variation in microbial populations but are often not applicable to the study of the subseafloor biosphere as most microorganisms are extremely difficult to isolate[14, 18, 19]. With the emergence of bioinformatic tools for culture-free, high-resolution strain and subspecies analyses in complex environments[5, 15, 20-22], genome-resolved metagenomic analyses can now be conducted at large-scale to reveal fine-scale evolutionary mechanism dynamics and strain-level metabolic variation[13, 14, 16, 17, 23-25]. Pioneering studies have been conducted in which metagenomic data has been used to explore the roles of basic processes (natural selection, mutation, genetic drift, and recombination)

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in shaping microbial evolution of several typical subseafloor habitats. For instance, *in situ* work examining genomic variation of microbes inhabiting the upper two meters of anoxic subseafloor sediments in Aarhus Bay revealed that rates of genomic diversification and selection do not change with either sediment age or depth, likely due to energy limitation and reduced growth in this environment[26]. In contrast to non-vent sediments, population-specific differences in selection pressure were observed in both *Sulfurovum* and *Methanothermococcus* species between two distinct geochemically distinct hydrothermal vent fields, where energy availability and cell abundances are relatively high[27, 28]. A third study found that gene flow and recombination appeared to shape the evolution of microbial metapopulations that disperse frequently through the cold, oxic crustal fluids of the mid-Atlantic ridge[29]. In contrast, no direct studies of the evolutionary histories and selection pressures of cold seep sedimentary microorganisms have been conducted despite the importance of their roles on global biogeochemical cycles. Knowledge of the nucleotide variation of key functional genes related to methane metabolism in these organisms is similarly lacking.

Here we hypothesize that microbial evolution in cold seep sediments may differ from observations in highly productive deep-sea hydrothermal vents and energy-limited marine sediments, due to the impact of the continuous flow of hydrocarbon-rich fluids. To gain insights into evolutionary trajectories among microbial populations inhabiting cold seep sediments, we examined the metagenomic data of 68 cold seep sediment samples to track population microdiversity from metagenomic short-read alignments and performed microdiversity-aware genomic comparisons. Our study also revealed the depth- and site-dependent trends of microbial evolution in deep sea cold seep sediments through the analysis of inter-sample genomic variation of species. Microbial adaptation in a gradient environment is a highly dynamic and complex process involving the interaction of multiple evolutionary forces[22]. Thus, the exploration of adaptive fingerprints to uncover evolutionary mechanisms of specific taxa in cold seep sediments may hint at the factors impacting long-term evolution in deep subseafloor biosphere, as well as those processes that shape the evolution of genes involved in adaptation to specific environmental factors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results and Discussion

Depth profiles of species-level clusters in cold seep sediments

We assembled metagenomic data sequenced from 68 sediment samples obtained from six globally distributed cold seep sites, spanning various depth layers from 0 to 4.3 mbsf (**Supplementary Figure 1 and Supplementary Table 1**). After binning of metagenomic assemblies, 1261 prokaryotic species-level clusters (1041 bacterial and 220 archaeal; **Supplementary Table 2**) were recovered according to the suggested threshold of 95% average nucleotide identity (ANI) for delineating species[30-32]. These species clusters belonged to 85 phyla (70 bacterial and 15 archaeal; **Figure 1a**) based on the Genome Database Taxonomy (GTDB; version R06-202)[33-36], and were highly represented by bacterial phyla including Chloroflexota (n = 184), Proteobacteria (n = 125), Desulfobacterota (n = 101), Planctomycetota (n = 73) and Bacteroidota (n = 67). The top five archaeal phyla with the largest number of species-level clusters were Asgardarchaeota (n = 50), Thermoplasmatota (n = 44), Halobacteriota (n = 42), Thermoproteota (n = 35; mainly *Bathyarchaeia*) and Nanoarchaeota (n = 19). Most of these phyla lack an available cultured representative in the GTDB[35] and consist exclusively of MAGs and/or SAGs. Around 51% and 94% clusters could not be assigned to an existing genus and species, respectively (**Figure 1b**), confirming that the majority of the cold seep sediment species lack even uncultured representative genomes in the reference database.

At the phylum level, the total relative abundance of Halobacteriota was the highest across samples of all eight depth groups (**Supplementary Figure 2**), and ranged from 2.9%±3.7% at 0-0.05 mbsf (n = 11) to 20.7%±12.6% at 2.0-3.0 mbsf (n = 4) (**Supplementary Table 4**). At the species level (**Supplementary Figure 3**), clusters from ANME-1 (genus of QEXZ01) ranked the highest (up to 9.1%±9.1% at 0.3-4.5 mbsf), followed by ANME-2c groups (up to 2.0%±2.3% at 0-0.3 mbsf), indicating their different depth distributions[37]. Bacterial taxa belonging to the phyla Desulfobacterota and Caldatribacteriota are also very abundant, with the latter being especially prevalent in deeper sediments (**Supplementary Figure 2**). The most abundant bacterial species was assigned to Caldatribacteriota taxon SB_S5_bin2, with values of up to 2.2%±2.3% at 1.0-2.0 mbsf (**Supplementary Figure 4**). Three species belonging to Desulfobacterota, the potential bacterial partners of ANME-1 or ANME-

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

144 2[38], became relatively predominant (up to 3.5%±0.5%) at 2.0-4.5 mbsf. Overall,
145 sediment depth is one important factor shaping distributions of cold seep sediment
146 microbial species clusters[7, 39].

147 **Selection of key functional taxa for microdiversity analysis**

148 Aerobic methane-oxidizing bacteria (MOB), anaerobic methanotrophic (ANME)
149 archaea, and sulfate-reducing bacteria (SRB) are three key groups of functional
150 microorganisms in cold seeps sediments[1, 6, 9]. To identify them, we screened the 1261
151 MAGs for the presence of three functional genes: the particulate methane
152 monooxygenase marker gene *pmoA* encoding for MOB, the methyl-coenzyme-M
153 reductase marker gene *mcrA* for ANME archaea, and the dissimilatory sulfite
154 reductase marker gene *dsrA* for SRB. To facilitate the microdiversity analyses of key
155 functional taxa, we kept only species-cluster representative MAGs with an estimated
156 quality score ≥50 (defined as the estimated completeness of a genome minus five
157 times its estimated contamination)[40] and at least 10× coverage[15, 22, 41, 42].

158 Three species belonging to *Gammaproteobacteria* were found to contain *pmoA* genes
159 and passed the required criteria for MAG quality (**Figure 2a and Supplementary**
160 **Table 3**). They were found exclusively in near-surface sediments (0-0.1 mbsf) from
161 the Haima cold seep in the South China Sea (**Supplementary Figure 5 and**
162 **Supplementary Table 5**), where oxygen is likely still available via penetration from
163 the water column. Functional annotation of 13 species genomes from
164 *Methanosarcinia* and *Syntropharchaeia* indicated the ability of these species to
165 perform anaerobic oxidation of methane (**Figure 2b and Supplementary Table 3**).
166 These 13 species represent four families: *Methanosarcinaceae*, HR1, ANME-2c and
167 ANME-1, and are widely distributed in 53 samples across sediment depths in the
168 range of 0.01-4.25 mbsf (**Supplementary Figure 5 and Supplementary Table 5**).
169 Among 117 *dsrA*-containing species genomes, 23 bacterial species from
170 Desulfobacterota perform sulfate reduction, likely coupled to anaerobic oxidation of
171 methane **(Figure 2c and Supplementary Table 3)**, and belonged to four clades:
172 "C00003060" (aka SEEP-SRB1c[43]), *Desulfobacterales*, *Desulfobulbia* and
173 *Desulfatiglandales*. They are widely distributed in 12 sediment columns and along
174 multiple sediment depths (0.01-4.25 mbsf) at various relative abundances
175 (**Supplementary Figure 5 and Supplementary Table 5**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genomic variations across different phylogenetic groups

At the genome level, we evaluated the 39 functional species for linkage disequilibrium (D'), pN/pS (the ratio of non-synonymous/synonymous mutations), r/m (the rate of recombination to mutation, gamma/mu; only for 10 species, **see Methods**), and number of single nucleotide variations for every thousand base pairs (SNVs/kbp). For populations from MOB, ANME and SRB, the evolutionary metrics varied greatly (**Figure 3a and Supplementary Table 5**), showing a wide distribution range in D' (~0.89 vs 0.72-0.99 vs 0.74-0.99), SNVs/kbp (5-41 vs 2-98 vs 2-89) and pN/pS (0.12-0.27 vs 0.11-0.25 vs 0.13-0.23). D' values indicate that MOB, ANME and SRB have not undergone high rates of homologous recombination similar to soil bacterial populations across a grassland meadow[17]. Population diversity in these groups as measured by SNVs/kbp is higher than that observed for soils in grassland meadows and subseafloor crustal fluids[17, 29]. These result indicate that deeply buried cold seep populations from MOB, ANME and SRB are under purifying selection, suggesting the possibility that these populations have reached an adaptive optimum for this stable environment, which is maintained by purging nonsynonymous mutations[44]. This result is in line with previous observations reported for bacterioplankton assemblages in sunlit freshwater and marine systems[16, 24, 45], and microbial populations in other deep sea biospheres[46, 47].

These data (**Figure 3a and Supplementary Table 5**) also indicate that microbial populations with different functional features in cold seep sediment habitats have diverse evolutionary modes, similar to observations in deep-sea hydrothermal vents with unique attributes[13, 27]. Accordingly, statistically significant differences were observed for almost all of the evolutionary metrics among the ANME and SRB lineages **(***P*<0.001, except 0.06 for D' of ANME groups; **Figures 3b and 3c)**. The SEEP-SRB1c group had the lowest SNVs/kbp ratio, highest pN/pS value and a low degree of recombination among the four SRB groups (**Figure 3b**). Fewer recombination events and lower nucleotide diversity are signals of selective sweep[24]. Thus, these data indicate that the SEEP-SRB1c group has undergone strong selection.

For HR1, *Methanosarcinaceae*, ANME-2c, *Desulfatiglandales* and ANME-1, negative correlations between D' and SNVs/kbp reflect a positive relationship between nucleotide diversity and homologous recombination **(Figure 4a and**

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 6). This is in agreement with the positive correlation (linear regression; R^2=0.34, P<0.001) found between r/m and SNVs/kbp for both ANME and SRB **(Supplementary Figure 6a and Supplementary Table 7)**. The ratio of nucleotide substitutions originating from homologous recombination to those originating from mutation (r/m ratio) can be used to measure the relative effect of homologous recombination on the genetic diversification of populations[48]. These results indicate that ANME and SRB populations could preserve high genome-wide diversity and prevent selective sweeps through increasing recombination rates to various environmental changes[14, 24]. In addition to the negative correlation between SNVs/kbp and D', negative correlations were also observed between SNVs/kbp and pN/pS for HR1 and *Methanosarcinaceae* **(Figure 4b, Supplementary Figure 7 and Supplementary Table 6)**. These correlations indicate that HR1 and *Methanosarcinaceae* populations are stabilized by frequent recombination while maintaining a high degree of intra-population diversity characterized by an accumulation of synonymous mutations, pointing to an ancient divergence of these two populations[45, 49]. In contrast, the *Desulfobulbia* populations had higher pN/pS values with more single nucleotide variants (linear regression; R^2=0.65, P=0.006; **Figure 4b and Supplementary Table 6)**. Additionally, they also had high SNVs/kbp and low degrees of within-species recombination **(Figure 3c and Supplementary Figure 8)**. These data suggest that *Desulfobulbia* populations may be in the process of subspecies establishment (i.e. speciation) or purging of non-synonymous mutations[14]. For ANME-2c and *Desulfobacterales* populations, the genome coverage (i.e. relative abundances of the populations) and SNVs/kbp fitted the linear regression model with a positive slope **(Figure 4c and Supplementary Table 6)**, indicating that population quantity may be constraining genomic microdiversification[42]. ANME-2c populations with higher abundances were found to show high single-nucleotide variations which were related to the high mutation rate or accumulation of mutations in the population (**Supplementary Figure 7**)[27, 50]. Relatively constant pN/pS ratios in those population further suggest that non-synonymous mutations in ANME-2c population might have been purged by purifying selection over a long period[45]. For the *Desulfobacterales* group, high-coverage populations were also reported to show relatively high degrees of recombination, and recombination did not bring changes in amino acids at the genome level despite the observed high nucleotide variations (**Supplementary Figure 8**)[22]. ANME-1 populations with higher abundances showed higher recombination

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rates (linear regression; R^2=0.12, P<0.001; **Figure 4d and Supplementary Table 6),** consistent with the positive linear regression relationship between r/m and coverage (linear regression; R^2=0.81, P<0.001; **Supplementary Figure 6b and Supplementary Table 7**). These findings are consistent with those of another study in which the relationship between population abundance and recombination rate was proposed to be the underlying mechanism responsible for the evolutionary success of the marine bacterium SAR11 in the near-surface epipelagic waters of the ocean[45].

Nucleotide variation of three key functional genes

For *pmoA*, *mcrA* and *dsrA* genes, SNVs/kbp ranged widely from 0.76 to 123 (45 on average, **Figure 5a and Supplementary Table 8)**. Based on the pN/pS values (0-1.43, 0.16 on average), these genes were under strong purifying selection. The evolutionary fitness of these three key functional genes was also consistent with that observed in previous studies for natural comammox *Nitrospira* populations[44] and *Thalassospira* bacterial populations isolated from million-year-old subseafloor sediments[19]. Our findings are in line with research showing that essential genes and enzymes catalyzing reactions that are difficult to bypass through alternative pathways are subject to higher purifying selection than nonessential ones[26, 44]. Even though the pN/pS values of the vast majority were well below 1 (indicating purifying selection), genes with pN/pS values above 1 and significantly higher than the genomic average were detected, which indicates that positive selection acted upon those genes. Although these microorganisms were deeply buried in subseafloor in which microbial evolution might operate differently from sunlit habitats[13], the distribution profile of pN/pS values are compatible with the neutral theory model[51] wherein most mutation events are neutral or deleterious[52] (**Figure 5b)**. Similarly, in microbial rare genes, evolution was found to proceed largely via neutral processes[53]. On the other hand, studies of the microbial inhabitants of wild bromeliads demonstrate patterns indicating the action of non-neutral processes[54].The similar evolutionary patterns appeared in both microbes and higher eukaryotes may be due to the conserved ancient mechanism.

For *dsrA* genes, SNVs/kbp and major allele frequency exhibited statistically significant differences among the four SRB groups (P≤0.005), while pN/pS values were similar among them **(Figure 5c)**. Similar evolutionary trends (relatively low SNVs/kbp and pN/pS values) were observed in *dsrA* genes from groups of

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Desulfobulbia, SEEP-SRB1c and *Desulfatiglandales*, **(Supplementary Figure 9)**. This indicates that *dsrA* genes in these groups were functionally stable following purifying selection. The *dsrA* gene in the *Desulfobacterales* group also had low pN/pS values (~0.12) but showed a broad range of SNVs/kbp. The *dsrA* gene of ETH-SRB1 SB_S7_bin23 populations with abnormally high pN/pS values were found with fewer synonymous mutations **(Supplementary Figure 10)**, indicating that *dsrA* genes in these populations were under strong positive selection and were likely further determined by random genetic drift[55].

For *mcrA*, the three evolutionary metrics (SNVs/kbp, pN/pS and major alleles frequency) were significantly different (P<0.001) among the four ANME groups **(Figure 5d and Supplementary Figure 11)**, indicating differences in evolutionary trends of *mcrA* genes across these groups. The *mcrA* gene from HMR20_21 in the ANME-1 group were found to have high pN/pS values with low synonymous mutation rates, indicating positive selection or relaxed purifying selection **(Supplementary Figure 10)**. For *mcrA* genes from ANME-2c, SNVs/kbp (linear regression; R^2=0.47, P<0.001) and pN/pS values (linear regression; R^2=0.17, P=0.003) positively correlated with gene coverage **(Supplementary Figure 12)**. This suggests that mutations were maintained for *mcrA* genes during the clonal expansion of ANME-2c populations.

Major allele frequency (0.79 on average) appeared to show a direct relationship with SNVs/kbp for *mcrA* and *dsrA* genes (**Supplementary Figures 9 and 11**). High SNVs/kbp corresponded to high major allele frequencies (~0.8), while the distributions of major allele frequency were mostly scattered (0.56-0.99) at low values of SNVs/kbp. Specific major alleles were fixed in most *mcrA* and *dsrA* genes (major allele frequency: 0.70-0.98) with varying degrees of nucleotide diversity (SNVs/kbp: 1-123). For *mcrA* and *dsrA* genes with relatively low major allele frequency (0.56-0.70) and low pN/pS values (0-0.26), their genetic heterogeneity was preserved and strongly selected in populations from HR1, SEEP-SRB1c, *Desulfobulbia* and *Desulfobacterales* which might help diversification of functional genes under different environmental conditions[56].

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Depth- and site-dependent trends of microdiversity

To determine whether microbial microdiversity was depth-dependent in deep-sea cold seep sediments, we assessed the relationship between evolutionary metrics and sediment depth. At genome level, a negative correlation was observed between SNVs/kbp and depth while pN/pS and D' showed a positive relationship with sediment depth for ANME and SRB populations **(Figures 6a-c and Supplementary Table 9)**. This suggests that as depth below the sea floor increases along the sediment column, microbial populations exhibit less microdiversity and levels of homologous recombination, as well as more relaxed purifying selection. This depth-dependent trend differs from that found for non-seep subseafloor sediments in which buried microbial populations show uniformly low genetic heterogeneity across sediment depths[26]; but similar to results from a pelagic freshwater system on the surface of the Earth where a lower mutation rate was observed in the deeper water layer[42]. This is likely related to the higher energy supply and larger population sizes in cold seep sediments compared to non-seep subseafloor habitats[29]. Depth-dependent microdiversity patterns were also observed for *mcrA* and *dsrA* genes. In general, SNVs/kbp and pN/pS are negatively correlated with sediment depth while major allele frequency correlated positively with depth but the degree to which it did so was population-specific **(Figure 6d-f and Supplementary Table 9)**. This suggests that *mcrA* and *dsrA* genes had lower degrees of microdiversity and were subject to higher levels of purifying selection when ANME and SRB were buried deeper. Additionally, based on observations at genome and gene levels, ANME and SRB populations likely undergo distinct selection pressures arising from sediment depths **(Figure 6)**.

Cold seep microbiomes are reported to be locally selected and diversified (macrodiversity) by unique benthic biogeochemical conditions and environmental gradients such as for methane and sulfate concentrations[6, 7, 9, 39]. Evolutionary metrics (D', SNVs/kbp, pN/pS) at the genome level showed significant differences (P<0.001) among different cold seep sites **(Supplementary Figure 13a)**, indicating that different physicochemical conditions of various cold seep systems also influenced the intra-population diversity and evolutionary processes in microbial populations. We also observed that there were differences (P=0.006) in evolutionary metrics (SNVs/kbp and major alleles frequency) for the three functional genes among different cold seep

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sites, **(Supplementary Figure 13b),** indicative of site-dependent microdiversity for functional genes. However, no clear difference (*P*=0.64) was observed among them with regard to pN/pS **(Supplementary Table 10)**. This suggests that those three key genes from ANME and SRB are mostly functionally conserved across different cold seep sites.

Conclusion

By analyzing a suite of population genetics parameters through metagenomic read mapping, here we show that microbial evolutionary processes of microbes in deep-sea cold seep sediments are much more complex than previously thought, and are governed by factors that differ from those observed in energy-limited marine sediments, hydrothermal vents and low biomass subseafloor fluids[26, 27, 29]. The 39 abundant MOB, ANME and SRB species in cold seep sediments had diverse evolutionary modes with different degrees of nucleotide variation and varying degrees of homologous recombination, demonstrating that selection pressure exerted by seep fluids enriched in methane may operate differently on different species-level populations. The investigated species from MOB, ANME and SRB in general showed low homologous recombination and strong purifying selection, the latter process being especially strong in functional genes related to methane (*pmoA* and *mcrA*) and sulfate (*dsrA*) metabolisms. Evolutionary metrics of these genes differed across species identities but were functionally conserved across various cold seep sites, supporting that the importance of relatively stable cold seep environmental conditions (i.e. continuous supply of methane and sulfate) in affecting evolutionary processes of genes essential in energy metabolism. We further found that sediment depths can not only shape the community structure of cold seep microbes in sediment depth layers from 0 to 4.3 mbsf, but also be one of the driving factors of microdiversity patterns for microbial genomes and genes. The depth-dependent trends of microdiversity might be the mixing consequences of redox condition changes and age of the sediment[12, 26]. Together, this study improves our understanding of principles that drive evolution of slow-growing deep-sea microbes in one of the unique subseafloor biosphere. However, these conclusions were only based on abundant populations from MOB, ANME and SRB, genomic microdiversification of other species at lower abundance levels can still play important functional roles was not evaluated due to

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

368 relatively shallow sequencing[5]. Deeper sequencing is therefore needed to depict a full
369 picture of microdiversity within microbial populations in deep sea sediments[57].

370

371 **Methods**

372 **Metagenomes for deep-sea cold seep sediments**

373 Metagenomic data sets were compiled from 68 sediment samples (0 to 4.3 mbsf)
374 collected from six globally distributed cold seep sites (**Supplementary Figure 1**).
375 These sites are as follows: Eastern Gulf of Mexico; Northwestern Gulf of Mexico;
376 Scotian Basin; Haiyang4, Site F, and Haima cold seeps in the South China Sea
377 (**Supplementary Table 1**). For samples from Northwestern Gulf of Mexico,
378 metagenomic data sets along with metadata were downloaded from NCBI Sequencing
379 Read Archive (SRA) and NCBI BioSample databases[58]. Samples of SY5, SY6 and
380 S11 were obtained from sediments of Haima cold seep areas[59] and raw sequencing
381 data were deposited in NCBI SRA (PRJNA739036 and PRJNA738468). For other
382 samples, sample collection and DNA sequencing were detailed previously[7, 8, 60-62]. The
383 68 sediment samples were catalogued into eight groups according to depth below the
384 seafloor (**Supplementary Figure 2**): 0-0.05 mbsf (n = 11); 0.05-0.1 mbsf (n = 14);
385 0.1-0.2 mbsf (n = 16); 0.2-0.3 mbsf (n = 8); 0.3-1.0 mbsf (n = 6);1.0-2.0 mbsf (n = 4);
386 2.0-3.0 mbsf (n = 4); 3.0-4.5 mbsf (n = 5).

387 **Metagenome assembly and binning**

388 Paired-end raw reads were quality-controlled by trimming primers and adaptors and
389 filtering out artifacts and low-quality reads using the Read_QC module within the
390 metaWRAP pipeline (v1.3.2; –skip-bmtagger)[63]. For each cold seep site, filtered reads
391 from each metagenome were individually assembled and co-assembled using
392 MEGAHIT (v1.1.3; default parameters) based on succinct *de Bruijn* graphs[64]. Contigs
393 less than 1000 bp were removed. For each assembly, contigs were binned using the
394 binning module (parameters: –maxbin2 –concoct –metabat2) and consolidated into a
395 final bin set using the Bin_refinement module (parameters: –c 50 –x 10) within
396 metaWRAP. The quality of the obtained MAGs was estimated by the lineage-specific

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

397 workflow of CheckM (v1.0.12)[65]. MAGs estimated to be at least 50% complete and
398 with less than 10% contamination were retained.

399 **Species-level clustering and taxonomic assignment**

400 Species-level clustering and representative species identification was performed using
401 dRep (v3.2.2)[31] with an average nucleotide identity (ANI) cutoff value of 95%. The
402 taxonomic classifications of representative MAGs were assigned based on the
403 Genome Database Taxonomy GTDB (release 06-RS202)[35] via the classify workflow
404 of GTDB-Tk (v1.5.1)[66]. To calculate the relative abundance of each MAG, CoverM
405 was used in genome mode (v0.6.0; parameters: –min-read-percent-identity 0.95 –min-
406 read-aligned-percent 0.75 –trim-min 0.10 –trim-max 0.90;
407 https://github.com/wwood/CoverM).

408 **Functional annotations and phylogenetic analysis**

409 METABOLIC-G, an implementation of METABOLIC (v4.0)[67], was used to predict
410 metabolic and biogeochemical functional trait profiles of MAGs. For phylogenetic
411 analysis of functional genes related to methane and sulfate metabolism, amino acid
412 sequences were aligned using the MUSCLE algorithm[68] included in the software
413 package MEGA X[69]. All positions with less than 95% site coverage were excluded.
414 The maximum-likelihood phylogenetic tree was constructed in MEGA X using the
415 Jones Taylor Thornton matrix-based model, bootstrapped with 50 replicates. The
416 output trees were visualized and beautified in the Interactive Tree Of Life (iTOL;
417 v6)[70].

418 **Calculation of evolutionary metrics**

419 Filtered reads from each sample were mapped to all species-cluster representative
420 MAGs concatenated together using Bowtie2 (v2.2.5; default parameters)[71]. Population
421 statistics and nucleotide metrics including linkage disequlibrium (D'), nucleotide
422 diversity (SNVs/kbp), nonsynonymous to synonymous mutation ratio (pN/pS) and
423 major allele frequency were calculated from these mappings using the profile module
424 of the inStrain program (v1.5.4; –database mode; default parameters)[15] at genome and
425 gene levels. Genetic annotation of MAGs was performed with Prodigal (v2.6.3; –p

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

meta)[72] for the gene module of inStrain.

Inferring recombination rates

Filtered reads were mapped to each of 39 selected representative MAGs using Bowtie2 (v2.2.5; –sensitive-local mode)[71]. The mcorr package (https://github.com/kussell-lab/mcorr)[73] was used to calculate the rate of recombination to mutation (gamma/mu) for each population. MAGs in which (1) normally distributed residuals for the model fit and (2) the bootstrapping mean was within 2X of the final estimate for gamma/mu[17] were retained, resulting in a set of 10 genomes for inferring recombination rates.

Statistical analyses

Statistical analysis was carried out in R (v4.0.0). Shapiro-Wilk and Bartlett's tests were used to assess the normality and variance homogeneity of the data. The Kruskal-Wallis rank sum test with Chi-square correction was used for comparison of evolutionary metrics in genomes and genes among different groups. Pearson's product-moment correlation was performed to assess the relationship between various evolutionary metrics (D', pN/pS, r/m, coverage and SNVs/kbp for genomes; pN/pS, major allele frequency, coverage and SNVs/kbp for genes) and their relationship with sediment depth. Linear regression was used to fit the data and predict the linear correlation between the two indexes mentioned above on population. These metrics were used to test the evolutionary processes in the cold seep sediment populations and the effect of sediment depth on them.

Data availability

MAGs, files for the phylogenetic trees and other related information have been uploaded to figshare (DOI: 10.6084/m9.figshare.17195003).

References

1. Joye, S. B., The Geology and Biogeochemistry of Hydrocarbon Seeps. *Annu Rev Earth Pl Sc* **2020,** *48*, (1), 205-231.
2. Suess, E., Marine cold seeps and their manifestations: geological control, biogeochemical criteria and environmental conditions. *Int J Earth Sci* **2014,** *103*, (7), 1889-1916.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Levin, L. A.; Baco, A. R.; Bowden, D. A.; Colaco, A.; Cordes, E. E.; Cunha, M. R.; Demopoulos, A. W. J.; Gobin, J.; Grupe, B. M.; Le, J.; Metaxas, A.; Netburn, A. N.; Rouse, G. W.; Thurber, A. R.; Tunnicliffe, V.; Van Dover, C. L.; Vanreusel, A.; Watling, L., Hydrothermal Vents and Methane Seeps: Rethinking the Sphere of Influence. *Front Mar Sci* **2016,** *3*, (72).

4. Chakraborty, A.; Ruff, S. E.; Dong, X.; Ellefson, E. D.; Li, C.; Brooks, J. M.; McBee, J.; Bernard, B. B.; Hubert, C. R. J., Hydrocarbon seepage in the deep seabed links subsurface and seafloor biospheres. *Proc Natl Acad Sci U S A* **2020,** *117*, (20), 11029-11037.

5. Gregory, A. C.; Gerhardt, K.; Zhong, Z. P.; Bolduc, B.; Temperton, B.; Konstantinidis, K. T.; Sullivan, M. B., MetaPop: a pipeline for macro- and microdiversity analyses and visualization of microbial and viral metagenome-derived populations. *Microbiome* **2022,** *10*, (1), 49.

6. Li, Z.; Pan, D.; Wei, G.; Pi, W.; Zhang, C.; Wang, J. H.; Peng, Y.; Zhang, L.; Wang, Y.; Hubert, C. R. J.; Dong, X., Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *Isme J* **2021,** *15*, (8), 2366-2378.

7. Dong, X.; Rattray, J. E.; Campbell, D. C.; Webb, J.; Chakraborty, A.; Adebayo, O.; Matthews, S.; Li, C.; Fowler, M.; Morrison, N. M.; MacDonald, A.; Groves, R. A.; Lewis, I. A.; Wang, S. H.; Mayumi, D.; Greening, C.; Hubert, C. R. J., Thermogenic hydrocarbon biodegradation by diverse depth-stratified microbial populations at a Scotian Basin cold seep. *Nat Commun* **2020,** *11*, (1), 5825.

8. Dong, X.; Greening, C.; Rattray, J. E.; Chakraborty, A.; Chuvochina, M.; Mayumi, D.; Dolfing, J.; Li, C.; Brooks, J. M.; Bernard, B. B.; Groves, R. A.; Lewis, I. A.; Hubert, C. R. J., Metabolic potential of uncultured bacteria and archaea associated with petroleum seepage in deep-sea sediments. *Nat Commun* **2019,** *10*, (1), 1816.

9. Ruff, S. E.; Biddle, J. F.; Teske, A. P.; Knittel, K.; Boetius, A.; Ramette, A., Global dispersion and local diversification of the methane seep microbiome. *Proc Natl Acad Sci U S A* **2015,** *112*, (13), 4015-20.

10. Vigneron, A.; Alsop, E. B.; Cruaud, P.; Philibert, G.; King, B.; Baksmaty, L.; Lavallee, D.; Lomans, B. P.; Eloe-Fadrosh, E.; Kyrpides, N. C.; Head, I. M.; Tsesmetzis, N., Contrasting Pathways for Anaerobic Methane Oxidation in Gulf of Mexico Cold Seep Sediments. *mSystems* **2019,** *4*, (1), e00091-18.

11. Boetius, A.; Ravenschlag, K.; Schubert, C. J.; Rickert, D.; Widdel, F.; Gieseke, A.; Amann, R.; Jorgensen, B. B.; Witte, U.; Pfannkuche, O., A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **2000,** *407*, (6804), 623-6.

12. Emil Ruff, S., Microbial Communities and Metabolisms at Hydrocarbon Seeps. In *Marine Hydrocarbon Seeps*, Teske, A.; Carvalho, V., Eds. Springer International Publishing: Cham, 2020; pp 1-19.

13. Anderson, R. E., Tracking Microbial Evolution in the Subseafloor Biosphere. *mSystems* **2021,** *6*, (4), e0073121.

14. Van Rossum, T.; Ferretti, P.; Maistrenko, O. M.; Bork, P., Diversity within species: interpreting strains in microbiomes. *Nat Rev Microbiol* **2020,** *18*, (9), 491-506.

15. Olm, M. R.; Crits-Christoph, A.; Bouma-Gregson, K.; Firek, B. A.; Morowitz, M. J.; Banfield, J. F., inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat Biotechnol* **2021,** *39*, (6), 727-736.

16. Delmont, T. O.; Kiefl, E.; Kilinc, O.; Esen, O. C.; Uysal, I.; Rappe, M. S.; Giovannoni, S.; Eren, A. M., Single-amino acid variants reveal evolutionary processes that shape the

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

biogeography of a global SAR11 subclade. *Elife* **2019,** *8*, e46497.

17. Crits-Christoph, A.; Olm, M. R.; Diamond, S.; Bouma-Gregson, K.; Banfield, J. F., Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. *Isme J* **2020,** *14*, (7), 1834-1846.
18. Zhao, S.; Lieberman, T. D.; Poyet, M.; Kauffman, K. M.; Gibbons, S. M.; Groussin, M.; Xavier, R. J.; Alm, E. J., Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell Host Microbe* **2019,** *25*, (5), 656-667 e8.
19. Orsi, W. D.; Magritsch, T.; Vargas, S.; Coskun, O. K.; Vuillemin, A.; Hohna, S.; Worheide, G.; D'Hondt, S.; Shapiro, B. J.; Carini, P., Genome Evolution in Bacteria Isolated from Million-Year-Old Subseafloor Sediment. *mBio* **2021,** *12*, (4), e0115021.
20. Costea, P. I.; Munch, R.; Coelho, L. P.; Paoli, L.; Sunagawa, S.; Bork, P., metaSNV: A tool for metagenomic strain level analysis. *PLoS One* **2017,** *12*, (7), e0182392.
21. Eren, A. M.; Esen, O. C.; Quince, C.; Vineis, J. H.; Morrison, H. G.; Sogin, M. L.; Delmont, T. O., Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **2015,** *3*, e1319.
22. Sjoqvist, C.; Delgado, L. F.; Alneberg, J.; Andersson, A. F., Ecologically coherent population structure of uncultivated bacterioplankton. *Isme J* **2021,** *15*, (10), 3034-3049.
23. Kiefl, E.; Esen, O. C.; Miller, S. E.; Kroll, K. L.; Willis, A. D.; Rappé, M. S.; Pan, T.; Eren, A. M., Structure-informed microbial population genetics elucidate selective pressures that shape protein evolution. *bioRxiv* **2022**, 2022.03.02.482602.
24. Bendall, M. L.; Stevens, S. L.; Chan, L. K.; Malfatti, S.; Schwientek, P.; Tremblay, J.; Schackwitz, W.; Martin, J.; Pati, A.; Bushnell, B.; Froula, J.; Kang, D.; Tringe, S. G.; Bertilsson, S.; Moran, M. A.; Shade, A.; Newton, R. J.; McMahon, K. D.; Malmstrom, R. R., Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *Isme J* **2016,** *10*, (7), 1589-601.
25. Schloissnig, S.; Arumugam, M.; Sunagawa, S.; Mitreva, M.; Tap, J.; Zhu, A.; Waller, A.; Mende, D. R.; Kultima, J. R.; Martin, J.; Kota, K.; Sunyaev, S. R.; Weinstock, G. M.; Bork, P., Genomic variation landscape of the human gut microbiome. *Nature* **2013,** *493*, (7430), 45-50.
26. Starnawski, P.; Bataillon, T.; Ettema, T. J.; Jochum, L. M.; Schreiber, L.; Chen, X.; Lever, M. A.; Polz, M. F.; Jorgensen, B. B.; Schramm, A.; Kjeldsen, K. U., Microbial community assembly and evolution in subseafloor sediment. *Proc Natl Acad Sci U S A* **2017,** *114*, (11), 2940-2945.
27. Anderson, R. E.; Reveillaud, J.; Reddington, E.; Delmont, T. O.; Eren, A. M.; McDermott, J. M.; Seewald, J. S.; Huber, J. A., Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat Commun* **2017,** *8*, (1), 1114.
28. Hoffert, M.; Anderson, R. E.; Reveillaud, J.; Murphy, L. G.; Stepanauskas, R.; Huber, J. A., Genomic Variation Influences *Methanothermococcus* Fitness in Marine Hydrothermal Systems. *Front Microbiol* **2021,** *12*, 714920.
29. Anderson, R. E.; Graham, E. D.; Huber, J. A.; Tully, B. J., Microbial populations are shaped by dispersal and recombination in a low biomass subseafloor habitat. *bioRxiv* **2021**, 2021.02.03.429647.
30. Jain, C.; Rodriguez, R. L.; Phillippy, A. M.; Konstantinidis, K. T.; Aluru, S., High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **2018,** *9*, (1), 5114.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31. Olm, M. R.; Brown, C. T.; Brooks, B.; Banfield, J. F., dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *Isme J* **2017,** *11*, (12), 2864-2868.
32. Olm, M. R.; Crits-Christoph, A.; Diamond, S.; Lavy, A.; Matheus Carnevali, P. B.; Banfield, J. F., Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems* **2020,** *5*, (1), e00731-19.
33. Rinke, C.; Chuvochina, M.; Mussig, A. J.; Chaumeil, P. A.; Davin, A. A.; Waite, D. W.; Whitman, W. B.; Parks, D. H.; Hugenholtz, P., A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat Microbiol* **2021,** *6*, (7), 946-959.
34. Parks, D. H.; Chuvochina, M.; Waite, D. W.; Rinke, C.; Skarshewski, A.; Chaumeil, P. A.; Hugenholtz, P., A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **2018,** *36*, (10), 996-1004.
35. Parks, D. H.; Chuvochina, M.; Rinke, C.; Mussig, A. J.; Chaumeil, P. A.; Hugenholtz, P., GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **2022,** *50*, (D1), D785-D794.
36. Parks, D. H.; Chuvochina, M.; Chaumeil, P. A.; Rinke, C.; Mussig, A. J.; Hugenholtz, P., A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **2020,** *38*, (9), 1079-1086.
37. Chadwick, G. L.; Skennerton, C. T.; Laso-Perez, R.; Leu, A. O.; Speth, D. R.; Yu, H.; Morgan-Lang, C.; Hatzenpichler, R.; Goudeau, D.; Malmstrom, R.; Brazelton, W. J.; Woyke, T.; Hallam, S. J.; Tyson, G. W.; Wegener, G.; Boetius, A.; Orphan, V. J., Comparative genomics reveals electron transfer and syntrophic mechanisms differentiating methanotrophic and methanogenic archaea. *PLoS Biol* **2022,** *20*, (1), e3001508.
38. Skennerton, C. T.; Chourey, K.; Iyer, R.; Hettich, R. L.; Tyson, G. W.; Orphan, V. J., Methane-Fueled Syntrophy through Extracellular Electron Transfer: Uncovering the Genomic Traits Conserved within Diverse Bacterial Partners of Anaerobic Methanotrophic Archaea. *mBio* **2017,** *8*, (4), e00530-17.
39. Hoshino, T.; Doi, H.; Uramoto, G. I.; Wormer, L.; Adhikari, R. R.; Xiao, N.; Morono, Y.; D'Hondt, S.; Hinrichs, K. U.; Inagaki, F., Global diversity of microbial communities in marine sediment. *Proc Natl Acad Sci U S A* **2020,** *117*, (44), 27587-27597.
40. Parks, D. H.; Rinke, C.; Chuvochina, M.; Chaumeil, P. A.; Woodcroft, B. J.; Evans, P. N.; Hugenholtz, P.; Tyson, G. W., Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2017,** *2*, (11), 1533-1542.
41. Bouma-Gregson, K.; Crits-Christoph, A.; Olm, M. R.; Power, M. E.; Banfield, J. F., *Microcoleus* (Cyanobacteria) form watershed-wide populations without strong gradients in population structure. *Mol Ecol* **2022,** *31*, (1), 86-103.
42. Okazaki, Y.; Nakano, S.-i.; Toyoda, A.; Tamaki, H., Long-read-resolved, ecosystem-wide exploration of nucleotide and structural microdiversity of lake bacterioplankton genomes. *bioRxiv* **2022**, 2022.03.23.485478.
43. Zhao, R.; Biddle, J. F., Helarchaeota and co-occurring sulfate-reducing bacteria in subseafloor sediments from the Costa Rica Margin. *ISME Communications* **2021,** *1*, (1), 25.
44. Palomo, A.; Dechesne, A.; Cordero, O. X.; Smets, B. F., Evolutionary Ecology of Natural Comammox Nitrospira Populations. *mSystems* **2022,** *7*, (1), e0113921.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45. Lopez-Perez, M.; Haro-Moreno, J. M.; Coutinho, F. H.; Martinez-Garcia, M.; Rodriguez-Valera, F., The Evolutionary Success of the Marine Bacterium SAR11 Analyzed through a Metagenomic Perspective. *mSystems* **2020,** *5*, (5), e00605-20.
46. Biddle, J. F.; Sylvan, J. B.; Brazelton, W. J.; Tully, B. J.; Edwards, K. J.; Moyer, C. L.; Heidelberg, J. F.; Nelson, W. C., Prospects for the study of evolution in the deep biosphere. *Front Microbiol* **2011,** *2*, 285.
47. Shoemaker, W. R.; Jones, S. E.; Muscarella, M. E.; Behringer, M. G.; Lehmkuhl, B. K.; Lennon, J. T., Microbial population dynamics and evolutionary outcomes under extreme energy limitation. *Proc Natl Acad Sci U S A* **2021,** *118*, (33), e2101691118.
48. Vos, M.; Didelot, X., A comparison of homologous recombination rates in bacteria and archaea. *Isme J* **2009,** *3*, (2), 199-208.
49. Dixit, P. D.; Pang, T. Y.; Maslov, S., Recombination-Driven Genome Evolution and Stability of Bacterial Species. *Genetics* **2017,** *207*, (1), 281-295.
50. Koskella, B.; Hall, L. J.; Metcalf, C. J. E., The microbiome beyond the horizon of ecological and evolutionary theory. *Nat Ecol Evol* **2017,** *1*, (11), 1606-1615.
51. Ohta, T., The Nearly Neutral Theory of Molecular Evolution. *Annu Rev Ecol Evol Syst* **1992,** *23*, (1), 263-286.
52. Nei, M.; Suzuki, Y.; Nozawa, M., The neutral theory of molecular evolution in the genomic era. *Annu Rev Genomics Hum Genet* **2010,** *11*, (1), 265-89.
53. Coelho, L. P.; Alves, R.; Del Rio, A. R.; Myers, P. N.; Cantalapiedra, C. P.; Giner-Lamia, J.; Schmidt, T. S.; Mende, D. R.; Orakov, A.; Letunic, I.; Hildebrand, F.; Van Rossum, T.; Forslund, S. K.; Khedkar, S.; Maistrenko, O. M.; Pan, S.; Jia, L.; Ferretti, P.; Sunagawa, S.; Zhao, X. M.; Nielsen, H. B.; Huerta-Cepas, J.; Bork, P., Towards the biogeography of prokaryotic genes. *Nature* **2022,** *601*, (7892), 252-256.
54. Louca, S.; Jacques, S. M. S.; Pires, A. P. F.; Leal, J. S.; Srivastava, D. S.; Parfrey, L. W.; Farjalla, V. F.; Doebeli, M., High taxonomic variability despite stable functional structure across microbial communities. *Nat Ecol Evol* **2016,** *1*, (1), 15.
55. Rocha, E. P. C., Neutral Theory, Microbial Practice: Challenges in Bacterial Population Genetics. *Mol Biol Evol* **2018,** *35*, (6), 1338-1347.
56. Orsi, W. D., Ecology and evolution of seafloor and subseafloor microbial communities. *Nat Rev Microbiol* **2018,** *16*, (11), 671-683.
57. Merrill, B. D.; Carter, M. M.; Olm, M. R.; Dahan, D.; Tripathi, S.; Spencer, S. P.; Yu, B.; Jain, S.; Neff, N.; Jha, A. R.; Sonnenburg, E. D.; Sonnenburg, J. L., Ultra-deep Sequencing of Hadza Hunter-Gatherers Recovers Vanishing Microbes. *bioRxiv* **2022**, 2022.03.30.486478.
58. 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* **2020,** *10*, (1), 5772.
59. Niu, M.; Fan, X.; Zhuang, G.; Liang, Q.; Wang, F., Methane-metabolizing microbial communities in sediments of the Haima cold seep area, northwest slope of the South China Sea. *FEMS Microbiol Ecol* **2017,** *93*, (9).
60. Lu, R.; Gao, Z.-M.; Li, W.-L.; Wei, Z.-F.; Wei, T.-S.; Huang, J.-M.; Li, M.; Tao, J.; Wang, H.-B.; Wang, Y., Asgard archaea in the haima cold seep: Spatial distribution and genomic insights. *Deep-Sea Res Pt I* **2021,** *170*, 103489.
61. Li, W. L.; Dong, X.; Lu, R.; Zhou, Y. L.; Zheng, P. F.; Feng, D.; Wang, Y., Microbial ecology

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of sulfur cycling near the sulfate-methane transition of deep-sea cold seep sediments. *Environ Microbiol* **2021,** *23*, (11), 6844-6858.

62. 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* **2020,** *2*, (4), 431-441.

63. Uritskiy, G. V.; DiRuggiero, J.; Taylor, J., MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **2018,** *6*, (1), 158.

64. Li, D.; Luo, R.; Liu, C. M.; Leung, C. M.; Ting, H. F.; Sadakane, K.; Yamashita, H.; Lam, T. W., MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **2016,** *102*, 3-11.

65. 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* **2015,** *25*, (7), 1043-55.

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

67. Zhou, Z.; Tran, P. Q.; Breister, A. M.; Liu, Y.; Kieft, K.; Cowley, E. S.; Karaoz, U.; Anantharaman, K., METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **2022,** *10*, (1), 33.

68. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **2004,** *32*, (5), 1792-7.

69. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K., MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **2018,** *35*, (6), 1547-1549.

70. Letunic, I.; Bork, P., Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **2021,** *49*, (W1), W293-W296.

71. Langmead, B.; Salzberg, S. L., Fast gapped-read alignment with Bowtie 2. *Nat Methods* **2012,** *9*, (4), 357-9.

72. Hyatt, D.; Chen, G. L.; Locascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J., Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **2010,** *11*, (1), 119.

73. Lin, M.; Kussell, E., Inferring bacterial recombination rates from large-scale sequencing datasets. *Nat Methods* **2019,** *16*, (2), 199-204.

Acknowledgements

The work was supported by the Scientific Research Foundation of Third Institute of Oceanography, MNR (No. 2022025), the National Natural Science Foundation of China (No. 41906076), the Science and Technology Projects in Guangzhou (No. 202102020970), and Guangdong Basic and Applied Basic Research Foundation (No. 20201910240000691).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contributions

XD, ZS and CRJH designed this study. XD and YP performed analysis. XD, YP, MW, LW, WW, KJ and CG interpreted the data. YW, XX, JL and CRJH contributed to data collection. XD, YP and LW wrote the paper, with input from other authors.

Conflict of interest

The authors declare no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legends

Figure 1. Classification of species-level MAGs recovered from global cold seep sediments. (a) Sankey based on assigned GTDB taxonomy showing recovered archaeal and bacterial MAGs at different phylogenetic levels. Numbers indicate the number of MAGs recovered for the lineage. (b) Total MAGs unclassified by GTDB-Tk at each taxonomic level. MAGs were dereplicated at species level (i.e., 95% ANI). Detailed statistics for 1261 MAGs are provided in **Supplementary Table 2**.

Figure 2. Maximum-likelihood phylogenetic trees of three key functional genes in cold seep sediments. Phylogenetic trees are based on alignments of amino-acid sequences of (a) PmoA, (b) McrA and (c) DsrA protein sequences. The sequences from the same taxonomic groups are highlighted in the same colors. Black dots indicate bootstrap values 50-100%. Scale bars indicate the average number of substitutions per site. Detailed annotations of species-cluster representative MAGs are provided in **Supplementary Table 3**.

Figure 3. Genome-wide evolutionary metrics of three key functional microbial groups in cold seep sediments. (a) Relationships between SNV density (SNVs/kbp), linkage disequilibrium (D'), the ratio of nonsynonymous to synonymous polymorphisms (pN/pS ratio), and genome coverage at genome level. Each dot represents one species-level microbial population. (b)-(c) Box plots showing comparison of SNV density, D' and pN/pS of sulfate-reducing bacteria and anaerobic methanotrophic archaea across different taxonomic groups. P-values of differences across different taxonomic groups were calculated using Kruskal-Wallis rank sum test. Source data is available in **Supplementary Table 5**.

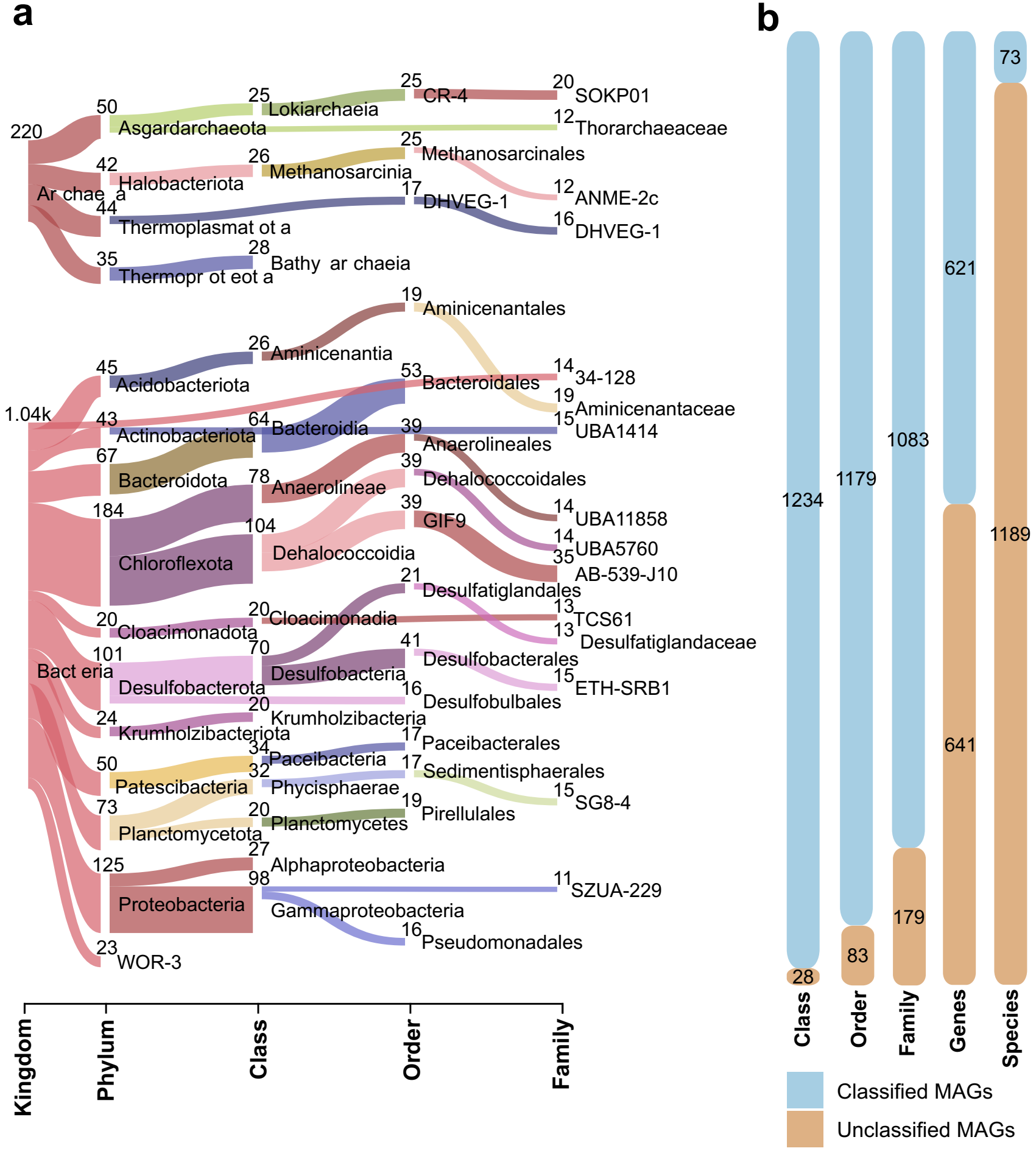
Figure 4. Genome-wide comparison of evolutionary metrics for microbial populations in cold seep sediments. (a)-(b) D' and pN/pS ratio in relation to SNV density. (c)-(d) SNV density and D' in relation to genome coverage. Each dot represents one species-level microbial population. Linear regressions and R^2 values

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

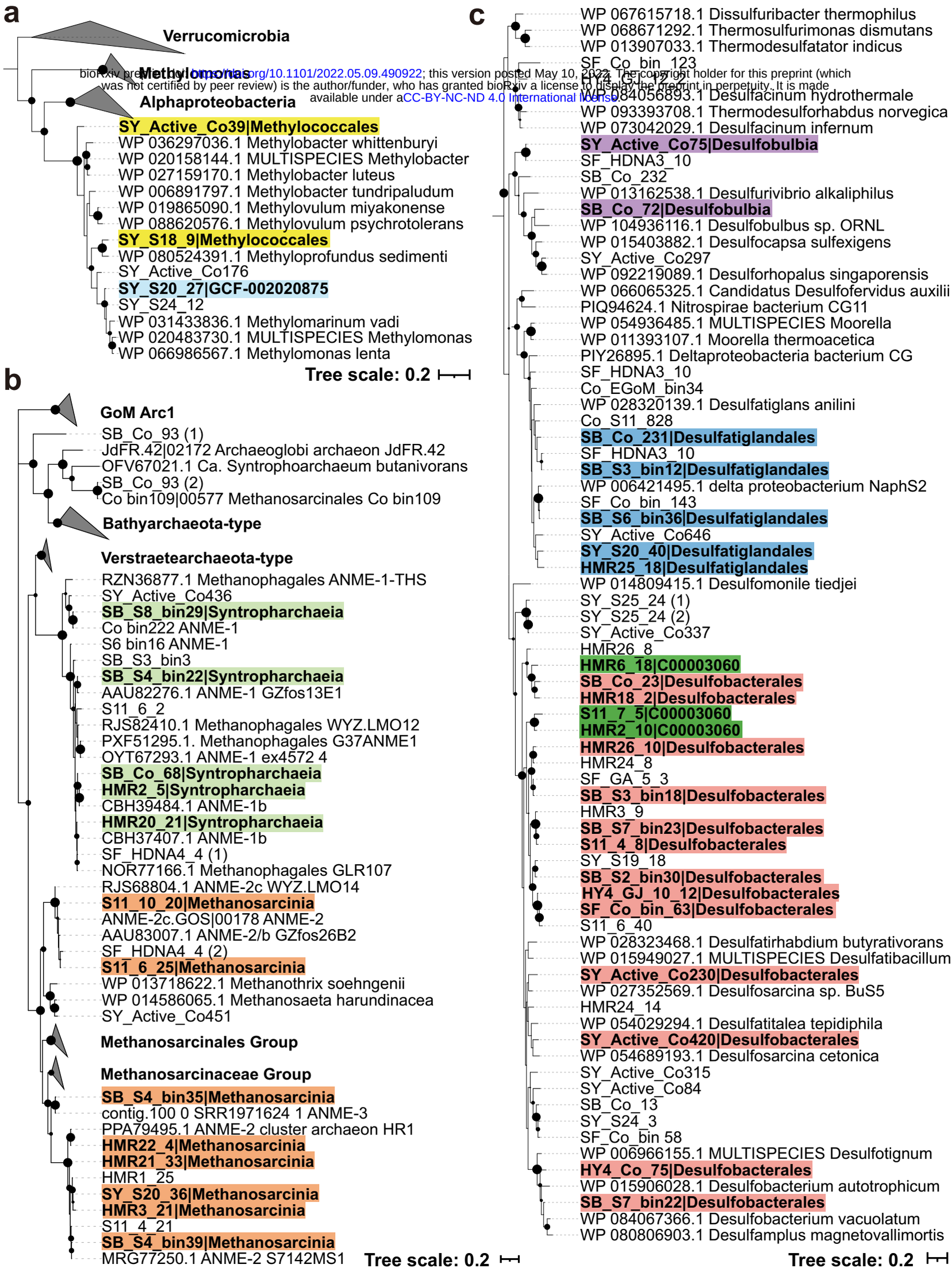
703 are indicated for different taxonomic groups. Detailed statistics for linear regressions
704 are provided in **Supplementary Table 6**.

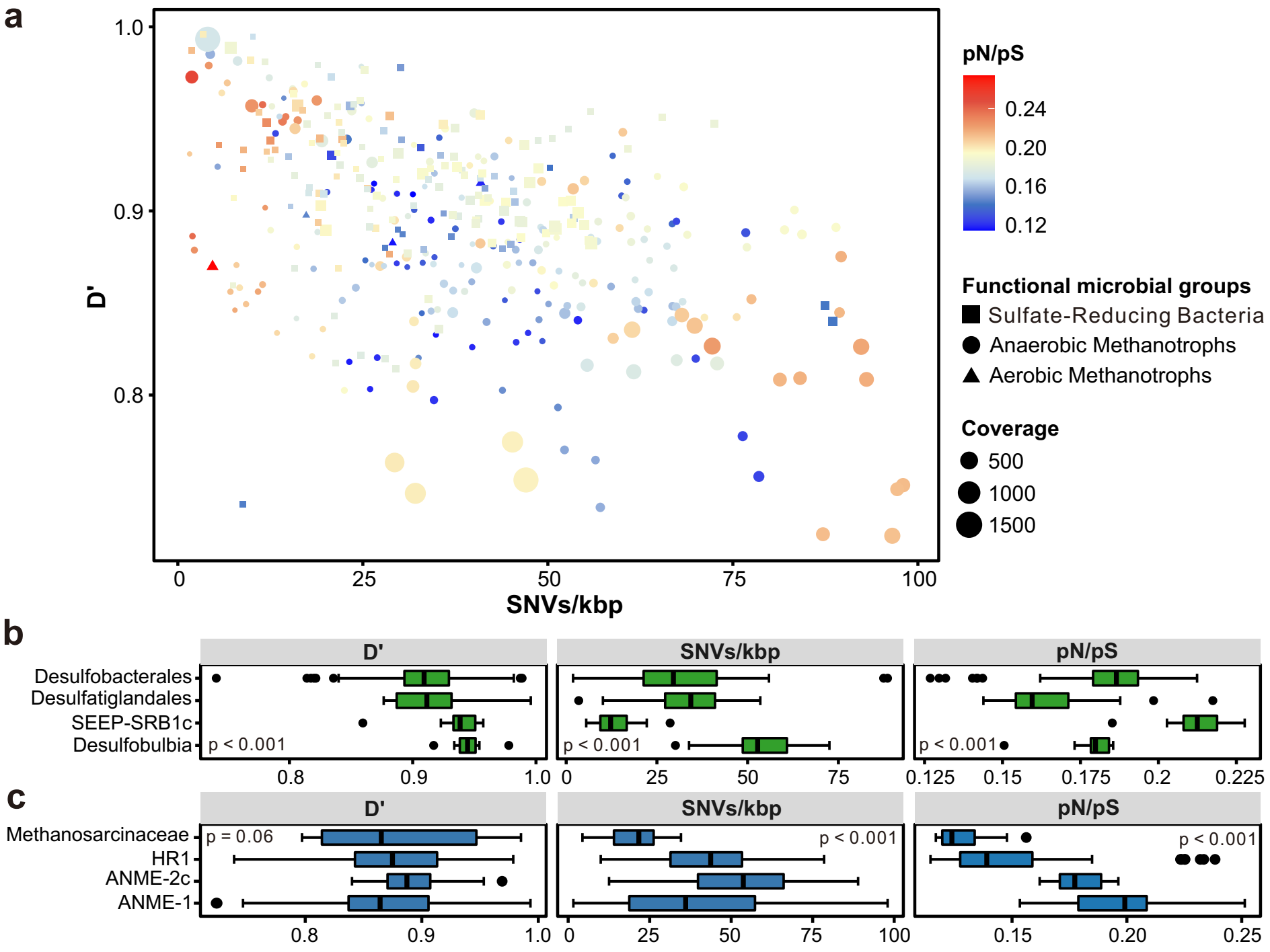
705 **Figure 5. Gene-specific evolutionary metrics of three key functional microbial**
706 **groups in cold seep sediments.** (a) Relationships between SNV density, pN/pS and
707 gene coverage at gene level. Each dot represents one species-level microbial
708 population. (b) Frequency histogram of pN/pS. (c)-(d) Box plots showing comparison
709 of SNV density, pN/pS and major allele frequency of *dsrA* and *mcrA* genes across
710 different taxonomic groups. P-values of differences across different taxonomic groups
711 were calculated using Kruskal-Wallis rank sum test. Source data is available in
712 **Supplementary Table 8**.

713 **Figure 6. Relationships between evolutionary metrics and cold seep sediment**
714 **depths (mbsf) at the whole-genome and gene levels.** (a)-(c) Comparison of
715 SNVs/kbp, pN/pS ratio and D' against sediment depths at the whole-genome level.
716 (d)-(f) Comparison of SNVs/kbp, pN/pS ratio and major allele frequency against
717 sediment depths at the gene level. Each dot represents one species-level microbial
718 population. Linear regressions and R^2 values are indicated for different taxonomic
719 groups. Detailed statistics for linear regressions are provided in **Supplementary**
720 **Table 9**.

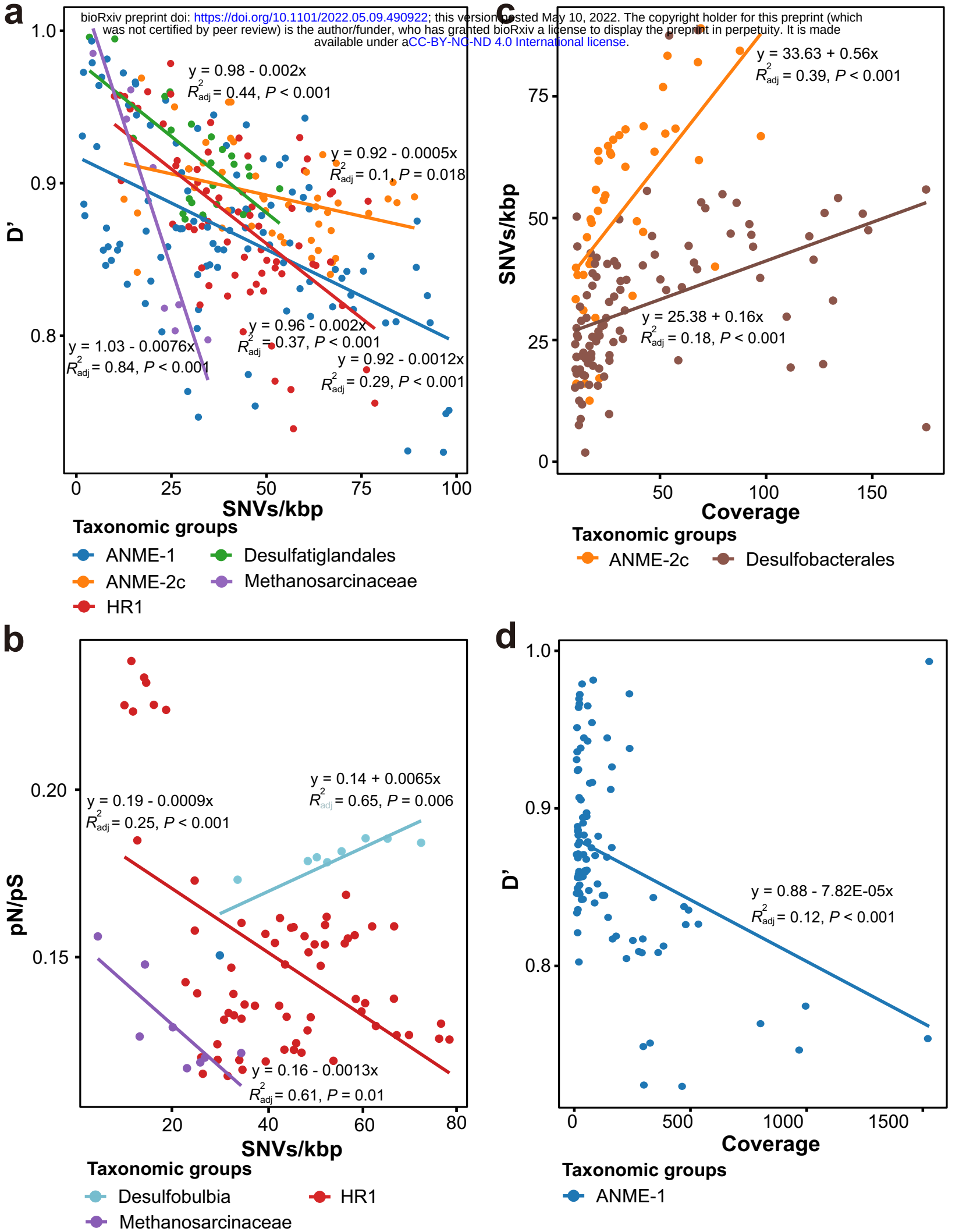

a
220
Ar chae a
50
Asgardarchaeota
42
Halobacteriota
44
Thermoplasmat ot a
35
Thermopr ot eot a
25
Lokiarchaeia
26
Methanosarcinia
28
Bathy ar chaeia
25
CR-4
12
Thorarchaeaceae
25
Methanosarcinales
17
DHVEG-1
20
SOKP01
12
ANME-2c
16
DHVEG-1
1.04k
45
Acidobacteriota
43
Actinobacteriota
67
Bacteroidota
184
Chloroflexota
20
Cloacimonadota
101
Desulfobacterota
24
Krumholzibacteriota
50
Patescibacteria
73
Planctomycetota
125
Proteobacteria
23
WOR-3
Bact eria
19
Aminicenantales
26
Aminicenantia
64
Bacteroidia
78
Anaerolineae
104
Dehalococcoidia
20
Cloacimonadia
70
Desulfobacteria
20
Krumholzibacteria
34
Paceibacteria
32
Phycisphaerae
20
Planctomycetes
27
Alphaproteobacteria
98
Gammaproteobacteria
53
Bacteroidales
39
Anaerolineales
39
Dehalococcoidales
39
GIF9
21
Desulfatiglandales
41
Desulfobacterales
16
Desulfobulbales
17
Paceibacterales
17
Sedimentisphaerales
19
Pirellulales
14
34-128
19
Aminicenantaceae
15
UBA1414
14
UBA11858
14
UBA5760
35
AB-539-J10
13
TCS61
13
Desulfatiglandaceae
15
ETH-SRB1
15
SG8-4
11
SZUA-229
16
Pseudomonadales
Kingdom
Phylum
Class
Order
Family
b
1234
1179
1083
621
641
179
83
28
73
1189
Class
Order
Family
Genes
Species
Classified MAGs
Unclassified MAGs

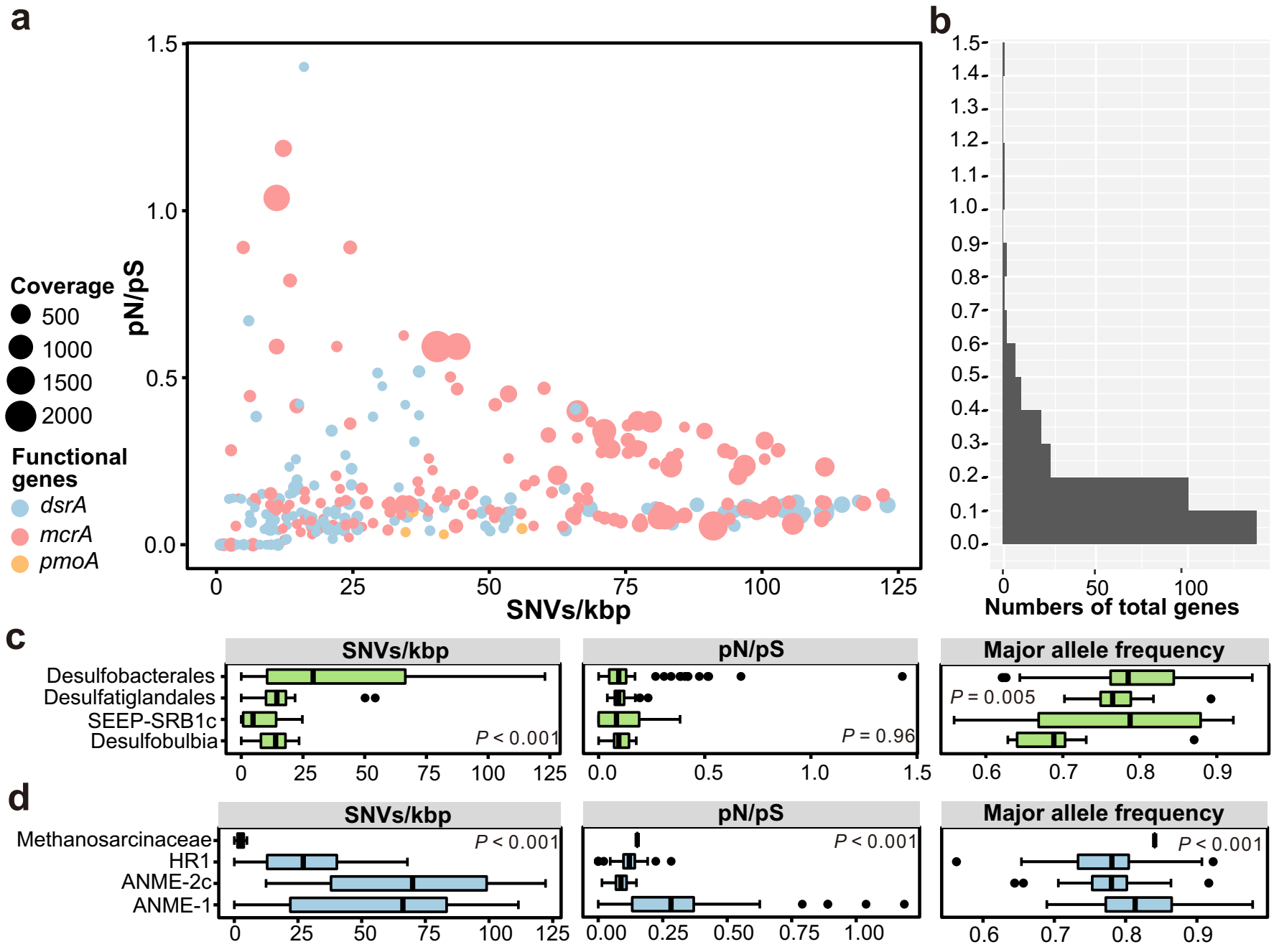
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




a
1.0
0.9
0.8
D'
0
25
50
75
100
SNVs/kbp
pN/pS
0.24
0.20
0.16
0.12
Functional microbial groups
Sulfate-Reducing Bacteria
Anaerobic Methanotrophs
Aerobic Methanotrophs
Coverage
500
1000
1500
b
D'
SNVs/kbp
pN/pS
Desulfobacterales
Desulfatiglandales
SEEP-SRB1c
Desulfobulbia
p < 0.001
p < 0.001
p < 0.001
0.8
0.9
1.0
0
25
50
75
0.125
0.15
0.175
0.2
0.225
c
D'
SNVs/kbp
pN/pS
Methanosarcinaceae
HR1
ANME-2c
ANME-1
p = 0.06
p < 0.001
p < 0.001
0.8
0.9
1.0
0
25
50
75
100
0.15
0.20
0.25

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.09.490922; this version posted May 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




a
b
c
d
Coverage
500
1000
1500
2000
Functional genes
dsrA
mcrA
pmoA
pN/pS
SNVs/kbp
Numbers of total genes
SNVs/kbp
pN/pS
Major allele frequency
Desulfobacterales
Desulfatiglandales
SEEP-SRB1c
Desulfobulbia
P < 0.001
P = 0.96
P = 0.005
Methanosarcinaceae
HR1
ANME-2c
ANME-1
P < 0.001
P < 0.001
P < 0.001

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.04.490644 posted May 04, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.