PNAS Nexus, 2024, 3, 1–13
https://doi.org/10.1093/pnasnexus/pgae057
Advance access publication 8 February 2024
Research Report

Gene inversion led to the emergence of brackish archaeal heterotrophs in the aftermath of the Cryogenian Snowball Earth

Lu Fan[a,b,1,*], Bu Xu[a,1], Songze Chen[a,b,1], Yang Liu[c,d,1], Fuyan Li[e], Wei Xie[f,g], Apoorva Prabhu[h], Dayu Zou[c,d], Ru Wan[i,j,k,l], Hongliang Li[k,l], Haodong Liu[a], Yuhang Liu[a], Shuh-Ji Kao[i,j], Jianfang Chen[k,l], Yuanqing Zhu[a,m], Christian Rinke[h], Meng Li[c,d], Maoyan Zhu[n,o,p] and Chuanlun Zhang[a,b,l,*]

[a]Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Department of Ocean Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen, Guangdong 518055, China
[b]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, Guangdong 511458, China
[c]Archaeal Biology Center, Institute for Advanced Study, Shenzhen University, Shenzhen, Guangdong 518060, China
[d]Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University, Shenzhen, Guangdong 518060, China
[e]Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawaii, Honolulu, HI 96822, USA
[f]School of Marine Sciences, Sun Yat-sen University, Zhuhai, Guangdong 519082, China
[g]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, Guangdong 519082, China
[h]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia
[i]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, Hainan 570228, China
[j]State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen, Fujian 361005, China
[k]Key Laboratory of Marine Ecosystem Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, Zhejiang 310012, China
[l]State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou, Zhejiang 310012, China
[m]Shanghai Sheshan National Geophysical Observatory, Shanghai Earthquake Agency, Shanghai 200062, China
[n]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[o]State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing, Jiangsu 210008, China
[p]Center for Excellence in Life and Paleoenvironment, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing, Jiangsu 210008, China
*To whom correspondence should be addressed: Email: fanl@sustech.edu.cn (L.F.); Email: zhangcl@sustech.edu.cn (C.Z.)
[1]L.F., B.X., S.C., and Y.L. contributed equally to this work.
Edited By: Christopher Dupont

Abstract

Land–ocean interactions greatly impact the evolution of coastal life on earth. However, the ancient geological forces and genetic mechanisms that shaped evolutionary adaptations and allowed microorganisms to inhabit coastal brackish waters remain largely unexplored. In this study, we infer the evolutionary trajectory of the ubiquitous heterotrophic archaea Poseidoniales (Marine Group II archaea) presently occurring across global aquatic habitats. Our results show that their brackish subgroups had a single origination, dated to over 600 million years ago, through the inversion of the magnesium transport gene *corA* that conferred osmotic-stress tolerance. The subsequent loss and gain of *corA* were followed by genome-wide adjustment, characterized by a general two-step mode of selection in microbial speciation. The coastal family of Poseidoniales showed a rapid increase in the evolutionary rate during and in the aftermath of the Cryogenian Snowball Earth (~700 million years ago), possibly in response to the enhanced phosphorus supply and the rise of algae. Our study highlights the close interplay between genetic changes and ecosystem evolution that boosted microbial diversification in the Neoproterozoic continental margins, where the Cambrian explosion of animals soon followed.

Keywords: brackish microorganisms, gene inversion, archaeal evolution, osmotic stress, Snowball Earth

Significance Statement

Global-scale land–ocean interactions stimulate ecosystem transformation and have had tremendous impacts on the evolution of animals and plants; however, they are poorly studied for microorganisms that evolved in deep time. One of the long-standing questions is the origination of brackish microbial communities, whose genetic repertoire might have recorded the dynamic environmental changes on continental margins. We demonstrate that the origination of the brackish subgroups of Marine Group II archaea likely happened after the Cryogenian glaciation. The complex rearrangement of the magnesium transporter gene *corA* in their genomes was demonstrated to be critical to this habitat transition. Our results provide a glimpse into the microbial evolution in Precambrian coastal oceans, where microorganisms might have paved the way for the arrival of animals.

Competing Interest: The authors declare no competing interests.
Received: August 18, 2023. **Accepted:** January 31, 2024

Introduction

Salinity is among the strongest environmental factors determining the community composition of both macro- and microorganisms (1, 2), which is so-called the "salinity divide" (3). Compared with distinct and characteristic biological populations in marine and freshwater ecosystems, biodiversity in brackish environments, such as estuaries and enclosed seas, is less understood, despite the importance of these environments in land–ocean interactions and their vulnerability to human activities. In the past decade, the existence of a unique, highly diverse and possibly globally distributed brackish community of bacteria and protists has been reported (4–6). Two recent studies inferred that the evolutionary transitions of bacteria and archaea among freshwater, brackish, and marine environments started in the early earth (7, 8); yet the geological forces and genetic mechanisms that drove the origination of brackish microorganisms are still largely unknown.

Brackish ecosystems are different from marine and freshwater ones in various environmental factors, such as salinity, nutrients, predators, and competitors (9). It is thus challenging to identify the key genetic changes and the primary selective force triggering the divergence between these ecosystems. Recent genome comparisons among saline and freshwater subgroups of bacteria and archaea have revealed functional differences in these two subgroups in osmotic regulation and in growth substrate specificity (10–16). Two fundamental questions remain, including (i) the primary vs. secondary genetic changes during the transition and (ii) whether the transition was caused by a sudden event or by gradual and cumulative effects (10, 17).

Land–ocean interactions may have strongly affected marine life evolution in geological history of the earth. A typical example is the Cambrian explosion of animals, which happened largely on continental margins (18). This drastic shift of global biota is attributed to interconnected global changes after the Cryogenian Snowball Earth and in the Ediacaran-Cambrian transition (18, 19) (~700–485.4 million years ago, Ma), with rapid sea level rise, orogeny, and enhanced land weathering (20, 21). Consequently, pulses of nutrients supplied to the ocean (22) boosted coastal primary productivity (23) and food web complexity (24). At the same time, the rise and variability of ocean oxygen levels (25, 26) might have accompanied a large increase in the abundance of planktonic algae (27). While numerous fossil records provide details on the radiation of multicellular animals and photoautotrophs during this period (28), evidence of a change in the evolution and ecology of bacterial and archaeal heterotrophs is lacking. Specifically, it is unclear whether the evolutionary rate of microorganisms accelerated in this period as that of animals (29) and whether changes in quantity and quality of organic matter altered the "microbial loop" (30) of the transforming food webs in the Ediacaran-Cambrian coastal waters. Moreover, the postglacial ocean is considered to be stably stratified with a surface layer of brackish and oxic water (31), but the biological impact of these expanded brackish environments on microbial evolution is currently unknown.

Marine Group II archaea (Ca. Poseidoniales) are among the most abundant planktonic archaea in global oceans, playing essential roles in marine carbon cycles (32, 33). Genomic analyses support the idea that they are heterotrophs living on algal-derived organic substrates. They are an important part of the "microbial loop" (32–37) and the microbial carbon pump processes (38) in the open ocean and coastal marine environments. Highly abundant Poseidoniales 16S rRNA genes were recently found at the estuary mixing zone with salinity below 15‰ (39), suggesting the possible existence of brackish Poseidoniales populations. In this study, an in-depth analysis of global brackish metagenomes allowed us to identify the evolutionary transition of Poseidoniales between marine and brackish habitats and to elucidate the important genetic changes that accompanied the origin of the brackish subclades. Furthermore, the evolutionary rates of Poseidoniaceae and Thalassarchaeaceae, the two families of Poseidoniales that adapt to nearshore and pelagic waters, respectively, were compared to reveal the possible geological and environmental driving forces during this marine–brackish shift.

Results and discussion

Global distribution of diverse and active brackish-specific Poseidoniaceae

Metagenome-assembled genomes (MAGs) of Poseidoniales were reconstructed from 282 metagenomes of global estuarine, enclosed sea, and coastal environments (Fig. S1 and Table S1) and phylogenetically analyzed together with those obtained in recent studies (32, 33, 37). This approach contributed 94 (20.7%) high-quality genomes to the updated nonredundant Poseidoniales genome dataset (455 MAGs, completeness median = 84.36%, contamination median = 1.09%) based on a cutoff of 99% average nucleotide identity (Table S2), which would contribute to a better understanding of the species diversity of Poseidoniales in low-salinity environments. Poseidoniales include two family-level subgroups, MGIIa (Poseidoniaceae) and MGIIb (Thalassarchaeaceae) (32). Abundance mapping of this nonredundant MAG dataset in global surface waters with distinct salinities shows that Poseidoniaceae are abundant in enclosed seas and estuaries, while Thalassarchaeaceae are detected exclusively in samples with salinities over 30‰ (Fig. 1A). This spatial distribution pattern is in line with the previous observation that Poseidoniaceae are adapted to more eutrophic and diverse coastal environments but Thalassarchaeaceae usually remain in the open ocean that is commonly oligotrophic (32–37). Notably, two distinct patterns of abundance distribution along the salinity gradient are detected in Poseidoniaceae: a "salt-preferred" pattern in which their abundance increased with the increase in salinity and a "brackish-specific" pattern in which they were enriched at salinities between 6.6 and 23‰ but depleted or absent at salinities beyond this range (Fig. 1A). Furthermore, we analyzed the abundance of Poseidoniales MAGs in the metatranscriptomes of 20 samples collected from three estuaries (Table S1). The results support that Poseidoniaceae of both patterns were in an active state (Fig. S2).

In the phylogenomic tree, Poseidoniaceae genomes from global estuaries and enclosed seas form several clusters (Fig. S3). Based on the above abundance analysis, six monophyletic brackish (BK)-specific clusters can be identified (labeled with a prefix "BK" in Figs. 1A and S3). They either branch within or as close sisters to previously classified Poseidoniaceae genus-level subgroups (32). Some of these brackish-specific lineages are found in different estuaries globally but enriched only in one estuary or one enclosed sea (Fig. 1A), implying that both global dispersion and local adaptation may play a role in their geographic distribution (8). Repeated presence but interannual variability in abundance of these brackish-specific lineages is observed in the Baltic Sea and the Pearl River estuary, suggesting they have specifically adapted to coastal brackish waters, but factors other than salinity might have impacted their temporary abundances (37).

Acidified proteome isoelectric point (p*I*) is recognized to be a strong indicator of microorganisms inhabiting saline environments (41, 42). We found that marine Poseidoniales have typical

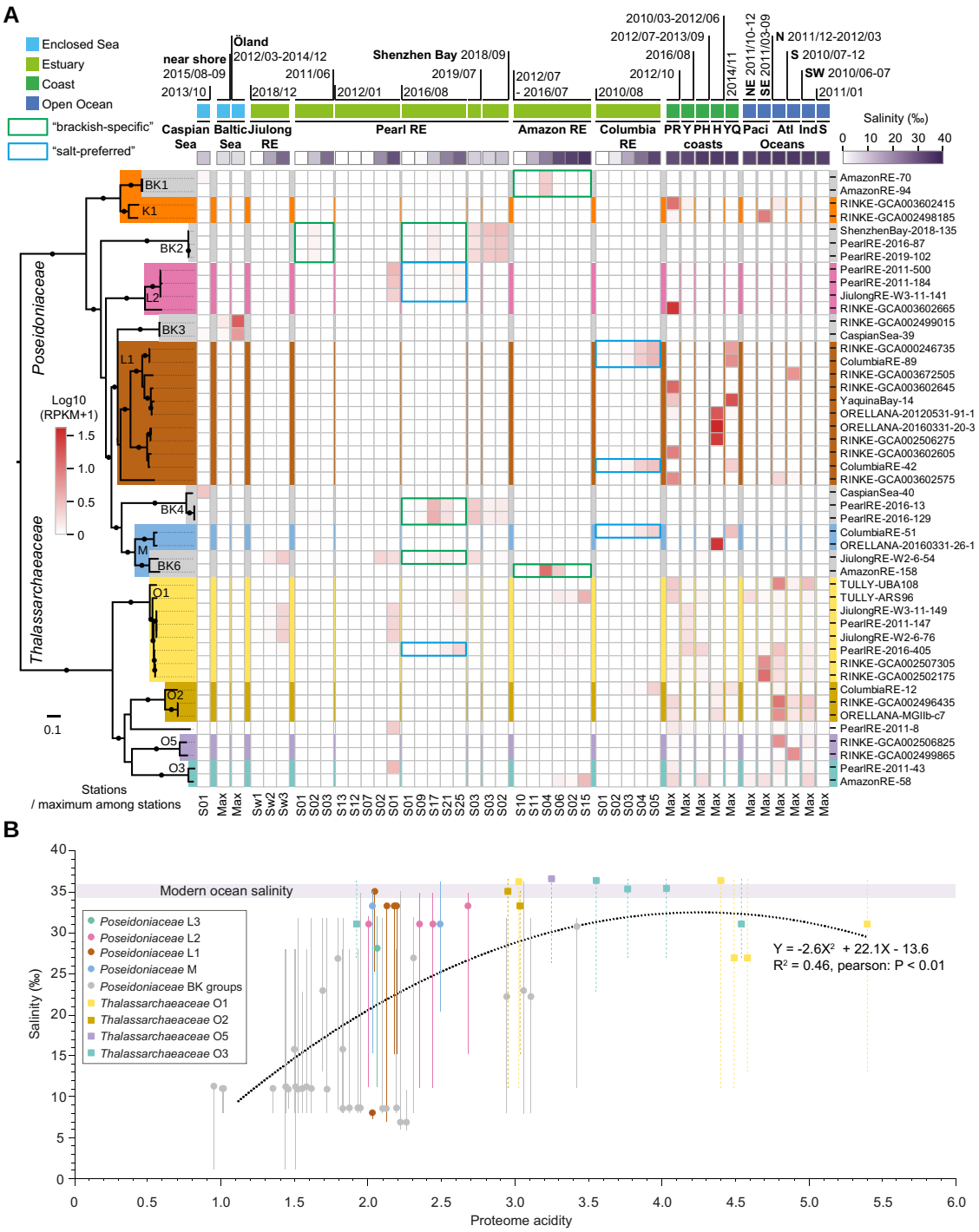

Fig. 1. Global distribution and proteome adaptation of brackish- and marine-specific Poseidoniales. A) Abundance pattern of Poseidoniales MAGs based on 246 metagenome samples from surface waters of global marine and brackish environments. MAGs with RPKM value >1 in enclosed sea and estuarine samples and >5 in coastal and oceanic samples are shown. The maximum-likelihood tree at the left of the panel is reconstructed based on 83 marker genes as the subset of the ar122 (40) (see Materials and methods). Solid dots on internal branches show branch supports of ultra-fast bootstrapping (1,000) in IQTree >95%. The letter codes and the shade colors of the genus-level subgroups are consistent with Rinke et al. (32), except for brackish clades (labeled with a prefix "BK") which are in gray. Abbreviations of sampling areas are: RE, river estuary; PR, Pearl River estuary; Y, Yangtze River estuary; PH, Port Hacking; H, Helgoland; YQ, Yaquina Bay; Alt, Atlantic; Ind, Indian; S, Southern/South; N, North; NE, Northeast; SW, Southwest; and SE, Southeast. Salinity shows the sample salinity or the average of the collection of samples. RPKM shows the abundance or the maximum abundance of each MAG in each sample or a collection of samples, respectively. B) Habitat salinity and proteome acidity of Poseidoniales MAGs found in estuaries and enclosed seas. Circles and solid lines belong to Poseidoniaceae MAGs, while squares and dashed lines belong to Thalassarchaeaceae MAGs. The position of each dot on the y-axis shows the optimal salinity of that organism. The scale of each vertical line on the y-axis shows the upper and lower limits of salinity for that organism. The different colors of circles, squares, and lines show Poseidoniales genus-level subgroups in consistent with those in Fig. 1 of Rinke et al. (32), except for brackish clades whose circles and lines are in gray. The regression of these dots is shown as the dashed black line.

acidified proteomes (Fig. S9 and supplementary material). To verify that the brackish-enriched Poseidoniaceae are specifically adapted to low-salinity habitats, we plot their optimum salinity values (i.e. the salinity of an environment in which a MAG had the highest abundance) detected in enclosed sea and estuarine samples according to the estimated acidity of their proteome p*I* patterns. Figure 1B clearly shows a polynomial correlation ($R^2 = 0.46$, $P < 0.01$, $n = 56$) between proteome acidity and salinity adaptation. Almost all Thalassarchaeaceae MAGs have acidity values over 2.9 (except one) and optimum salinity over 30‰ (except two). Most Poseidoniaceae with acidity above 2.0 were abundant at salinities from 20 to 36‰. In contrast, Poseidoniaceae MAGs of acidity below 2.0 belonging to brackish-specific clades have optimum salinities below 30‰ and down to 8‰. Some taxa are detectable even in river mouths at a salinity of around 1‰. This distinct distribution pattern of marine and brackish Poseidoniaceae subgroups is consistent with the suggestion of long-term divergent evolution (42).

The emergence of brackish Poseidoniaceae after the inversion of the *corA* gene in a stress-response gene cluster

Researchers have proposed various genes potentially contributing to the land–ocean divergence in microbial evolution, including those functioning in osmotic regulation, substrate preference, and adaptation to dynamic environments (3, 8). To identify genes potentially responsible for differentiating the marine and brackish Poseidoniales subgroups, we annotated Poseidoniaceae and Thalassarchaeaceae MAGs (Table S2) and conducted gene-centered comparison. Remarkably, the magnesium transporter gene *corA* (arCOG02265) is the only one that is present in >80% MAGs (30 in 33, or 90.9%) of the brackish clades and in <40% MAGs (18 in 369, or 11.0%) of the marine clades of Poseidoniaceae (Fig. S5). Student's t test shows that the proteome acidities of Poseidoniaceae encoding *corA* are significantly lower than those without this gene ($P = 0.016$, two-tailed; Fig. 2A) and MAG completeness does not significantly affect the presence/absence analysis of *corA* ($P = 0.266$, two-tailed; Fig. S6). CorA is one of the main import channels for magnesium ions (Mg^{2+}) in bacteria and archaea (43). Its potential bidirectional and concentration-regulated feature (44) suggests that it may act as a condition-dependent valve to maintain a stable intracellular magnesium concentration in hydrodynamic estuaries. This observation suggests that intracellular magnesium, which can potentially stabilize or modulate the structures of macromolecules such as DNAs, RNAs, and proteins, may be important in the adaptation of brackish Poseidoniaceae (supplementary material). It also highlights that the stress of salinity fluctuation in brackish environments instead of a steady and moderate salinity is potentially the primary barrier for brackish water adaptation of Poseidoniales.

Neighboring gene analysis integrated with phylogenetic interference analysis indicates that *corA* had a highly conserved evolutionary trajectory and may have played an essential role in stress response. Specifically, the tree of *corA* is highly congruent with the phylogenomic tree of Poseidoniales suggesting that after its gain in the common ancestor of Poseidoniales possibly from a bacterium (Fig. S7 and supplementary material), *corA* was generally passed vertically when Poseidoniales diversified (Figs. 2B and S8 and supplementary material). The absence of *corA* in some Poseidoniaceae and the majority of Thalassarchaeaceae are likely a result of sporadic loss (Figs. 2B and S9 and supplementary material). Moreover, in Poseidoniales, *corA* is exclusively found at the tail of a highly conserved gene cluster consisting of over 10 syntenous genes (Figs. 2C and S9). This gene cluster contains core gene sets involved in DNA repair (e.g. exonuclease VII), transcription (e.g. archaeal DNA-directed RNA polymerase RpoF), translational regulation (e.g. large subunit ribosomal protein L21e and ribosomal RNA small subunit methyltransferase A), and posttranslational modification (e.g. tRNA pseudouridine synthase 10) by modifying macromolecules such as DNAs, RNAs, and proteins (supplementary material). Adjacent and syntenous genes often form operons and transcribe simultaneously (49). By mapping available metatranscriptomic reads of estuarine samples (i.e. samples of the Pearl River estuary and the Columbia River estuary) to the brackish Poseidoniaceae genomes, we identify RNA reads mapped across the intergenic regions of *corA* and upstream three syntenous genes, suggesting that co-transcription could have possibly occurred (Fig. S10 and supplementary material). However, as little is known about the transcriptional mechanisms of Poseidoniales that are still uncultured, further evidence is needed to fully understand the transcriptional variation of *corA* in different Poseidoniales lineages.

Importantly, we find that the sporadically distributed brackish clades in the Poseidoniaceae evolutionary tree might have had a single origination, likely after the inversion of *corA* in the common ancestor of Poseidoniaceae. In the two basal genus-level subgroups P and Q1 of Thalassarchaeaceae, the coding direction of the only three *corA* copies is reversed to those of the syntenous genes in the stress-response gene cluster (Figs. 2B, C, and S9). As phylogenetic analysis suggests the copies of *corA* in Thalassarchaeaceae diverged earlier than those in Poseidoniaceae, they likely represent the ancestral versions (Fig. S7 and supplementary material). In this direction, despite the benefit that *corA* may have brought to these Thalassarchaeaceae lineages, the transcription of *corA* may have disrupted the integrity and transcription of the stress-response gene cluster, and the net effect might be deleterious. Consequently, it may be prone to negative selection and loss. In comparison, in the common ancestor of Poseidoniaceae, the *corA* gene was inverted and then lied in the same coding direction as the rest of the gene cluster, making co-transcription possible (Figs. 2B, C and S9 and supplementary material). However, after this inversion event, *corA* did not seem to immediately establish co-transcription with upstream syntenous genes in stress response, since a basal monophyletic clade of Poseidoniaceae containing genus-level subgroups J1, J2, and J3 partially contained *corA* yet showed no brackish-adaptational features. Instead, the emergence of brackish Poseidoniaceae likely only occurred after the branching out of J1, J2, and J3 (more likely after the branching out of TULLY-TrashBin17), but no later than the common ancestor of all the extant brackish lineages (indicated by the star in Fig. 2B), in which *corA* might have fully coupled its transcription to the upstream genes (supplementary material). This rearrangement provided a selective advantage for Poseidoniaceae in adaptation to brackish waters, and *corA* was thus preserved in further diversification. The following sporadic loss of *corA* gene in other Poseidoniaceae lineages, however, forced them to go back to the more saline water (Fig. 2B).

A proposal of two-step changes in habitat transition of Poseidoniaceae

Genetic mechanisms are elusive in the study of evolutionary transitions across the salinity barrier. Eiler et al. (10) suggested a gradual tuning of metabolic pathways and transporters of the marine clades of SAR11 bacteria toward organic substrates in freshwater environments. In contrast, Henson et al. (17) proposed that the irreversible loss of osmolyte transporters could be the critical step in the formation of freshwater SAR11 bacteria. We notice that

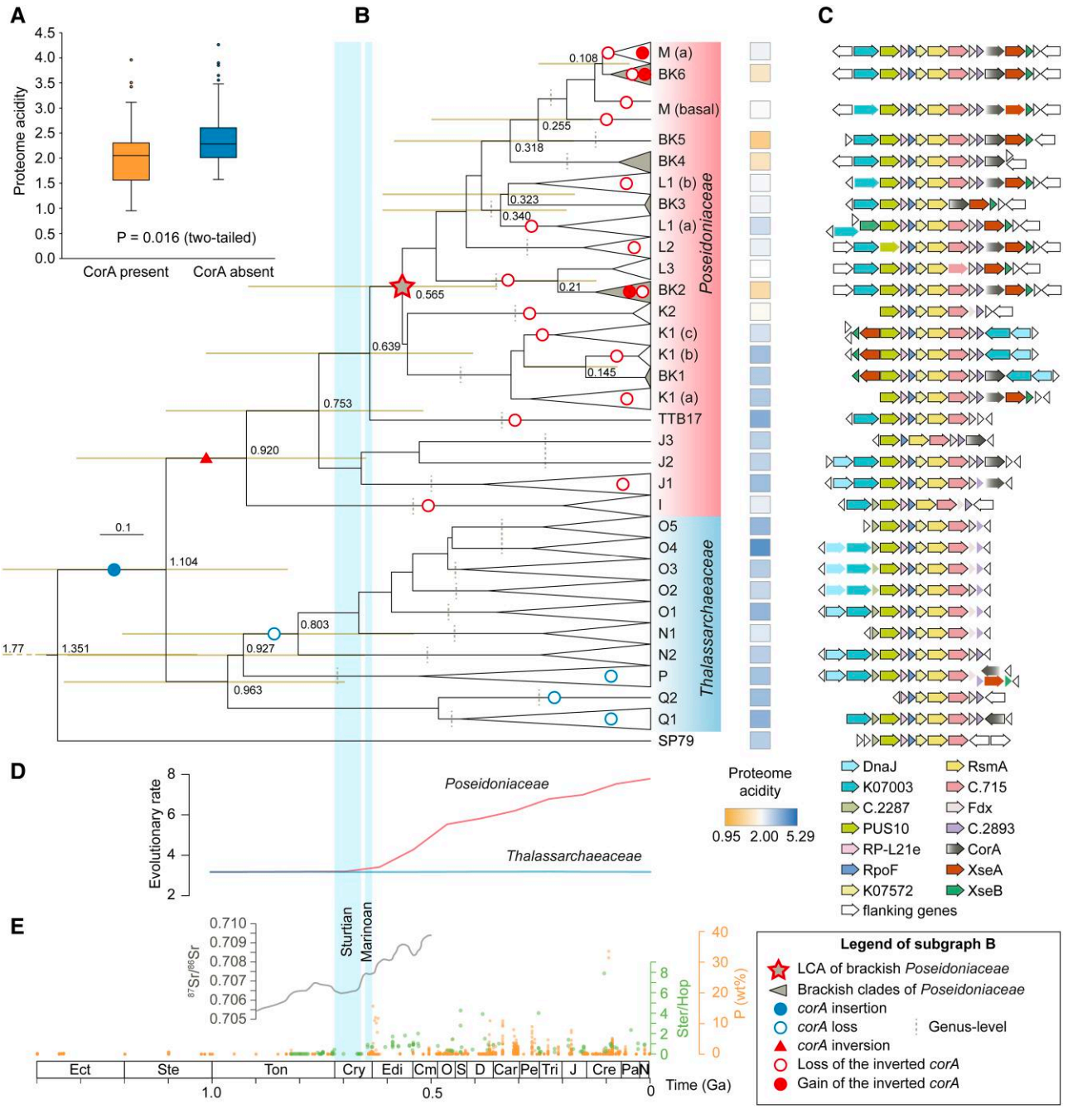

Fig. 2. Genetic changes, evolutionary rate, and geological background in the origination and evolution of brackish Poseidoniales. A) The distribution of proteome acidity in genomes of Poseidoniaceae with or without the *corA* gene. B) Molecular dating results and proteome acidity patterns of Poseidoniales subgroups with the change of the *corA* gene in evolution. The tree is part of the tree in Fig. S11A, which is reconstructed based on 39 of the 41 marker proteins described by Adam et al. (45). Values on the nodes show a 95% CI. The letter codes of the genus-level subgroups are consistent with Rinke et al. (32). Genus-level cutoff in the tree is derived from RED of GTDB taxonomic ranking (46). Small letters in brackets show subgenus-level subgroups. Proteome acidity levels at the right show the median values of MAGs in each clade. TTB17 = TULLY-TrashBin17. C) Arrangement of the stress-response gene cluster in Poseidoniales genomes. Arrows without edges suggest that the genes are present only in some of the MAGs in each clade. Potential functions of genes are explained in supplementary material. The completeness values of the representative MAGs in each genus/subgeneric subgroup range from 75.41 to 100% (median = 98.43%). Gene arrangement and synteny of each genus/subgeneric subgroup shown in Fig. 2C was summarized based on the observation in the MAGs in Fig. S9. The adjacency of these genes was based on evidence from a single contig in each MAG. Moreover, genes up- and down-streaming the gene cluster were also present in the same contig, therefore precluding gene absence caused by assembly problems. D) Evolutionary rate changes of Poseidoniaceae and Thalassarchaeaceae. E) Geological records of phosphorus deposit in shales from Reinhard et al. (22), the $^{87}Sr/^{86}Sr$ curve copied from Fig. 1 of Laakso et al. (47), and the relative contribution of eukaryotic and bacterial lipids to sedimentary organic matter is approximated by the sterane/hopane ratio (Ster/Hop) by Brocks et al (27). Glaciation events are based on Hoffman et al. (48). P, phosphorus; Ect, Ectasian; Ste, Stenian; Ton, Tonian; Cry, Cryogenian; Edi, Ediacaran; Cm, Cambrian; O, Ordovician; S, Silurian; D, Devonian; Car, Carboniferous; Pe, Permian; Tri, Triassic; J, Jurassic; Cre, Cretaceous; Pa, Paleogene; N, Neogene; Ga, billion years ago.

the evolutionary history of Poseidoniaceae transitioning between brackish and marine habitats is characterized by multiple salinity-based divergent events (Fig. 2B). To elucidate these complex evolutionary processes, we focused on the gene-gain and gene-loss events in association with the salinity and proteome acidification diversification of selected high-quality MAGs (completeness median = 93.63%, contamination median = 0, Table S3) belonging to the genus-level subgroups BK4, BK5, and M and the subgenus-level subgroup BK6 based on the ALE results (Table S4).

A broad two-step genomic change was identified during the marine–brackish inter-transition (Fig. 3). The first step was the gain or loss of the *corA* gene. Specifically, the gene was replaced by a new copy in the common ancestor of BK4 and then vertically transferred into this subgroup. In the monophyletic clade containing BK5, M, and BK6, *corA* was vertically transferred, followed by sporadic losses in marine species of the genus M (Fig. 3).

In the second step, accompanying and especially following the change of *corA*, massive gains or losses of habitat-specific genes possibly mediated by lateral gene transfer are observed in Poseidoniaceae genomes during their gradual diversification to marine or brackish environments (Fig. 3). For example, as the less-acidified proteomes of brackish Poseidoniaceae comprise more basic amino acids than those of their marine counterparts (Fig. 1B), the concentration of free basic amino acids in the cytoplasmic pool needs to be lowered by either increasing export or stopping biosynthesis to maintain charge balance. Indeed, the acquisition of lysine/arginine efflux happened in the ancestors of BK4 and BK6, respectively, and the loss of lysine biosynthesis happened in the ancestor of BK4 (Fig. 3). Moreover, microorganisms living in brackish and marine environments face great physiochemical differences in nutrient availability, substrate composition, and stress types. Accordingly, we find large-scale loss of genes involved in phosphorus and trace element uptake, peptide degradation (peptidases/proteases), and organic-matter utilization in the BK4 and BK6 lineages. At the same time, some peptidases and biosynthetic enzymes required for brackish water adaptation were obtained in brackish taxa. For example, the acquisition of fructose/tagatose bisphosphate aldolase by organism ShenzhenBay-2018-35 and 2-keto-4-pentenoate hydratase in the catechol meta-cleavage pathway by organism PearlRE-2016-260 possibly reflects the adaptation of brackish Poseidoniaceae to consume terrestrial substrates (Fig. 3 and Table S4) (50).

This observation is consistent with the fundamental rule that salinity is the strongest environmental barrier in the microbial habitat transition between land and ocean and highlights the possibility that multiple stages may be required for microorganisms to complete the physiological adaptation from one saline environment to the next. Specifically, the sudden change of primary and qualitative niche traits (e.g. adaptation to salinity fluctuation) precedes the gradual changes in secondary and accumulative traits (e.g. proteome acidity change and multidimensional metabolism adjustment) in Poseidoniaceae genomes. This evolutionary model supports the hypothesis that gene-specific selection drives speciation followed by genome-wide divergence (51), which may shed light on the debate raised by Eiler et al. (10) and Henson et al. (17) and may also explain the previous discovery of "temporal fragmentation of speciation" (52).

Habitat divergence and evolutionary rate changes of Poseidoniales potentially affected by the enhanced land weathering and the rise of algae after the Snowball Earth

Temporal connection between genetic changes and major geological events can provide insights into the potential environmental driving force of microbial evolution in the absence of a fossil record. Despite the frequent discussion on the uncertainty of methodology that may impair dating accuracy, molecular clock analysis aided by biomarker records has unveiled important events in life evolution before the emergence of multicellular organisms (7, 53–58). We here establish a geological time scheme in the evolutionary diversification of Poseidoniales to infer the possible geological and environmental driving force in the origination of brackish lineages (Fig. 2B). To ensure the robustness of molecular dating analysis on Poseidoniales, we here compare the results from alternative phylogenetic tree topologies (i.e. "methanogen-basal" vs. "DPANN-basal," supplementary material), select a modeling method based on the justification in our previous study (54), and build our conclusions on multiple interconnected lines of genetic and geological evidence (Fig. 2B–E).

Our dating analysis shows that the origination of brackish Poseidoniaceae was temporally linked to the emergence of eukaryotic algae in the Ediacaran Period. Based on the "methanogen-basal" tree topology, the emergence of Poseidoniales and the division of Poseidoniaceae and Thalassarchaeaceae might have happened around 1.104 billion years ago (Ga) with a 95% CI from 1.476 to 0.826 Ga (Figs. 2B and S11A). The divergence of the crown groups of both Poseidoniaceae and Thalassarchaeaceae might have happened before 0.9 Ga. The insertion of *corA* into the common ancestor of Poseidoniales was in the period from 1.351 (95% CI = 1.77 to 1.031) Ga to 1.104 (95% CI = 1.476 to 0.826) Ga and its inversion happened between 1.104 (95% CI = 1.476 to 0.826) Ga and 0.920 (95% CI = 1.308 to 0.647) Ga (Figs. 2B and S11A). Notably, the common ancestor of extant brackish Poseidoniaceae emerged after 639 (95% CI = 1,012 to 403) but no later than 564 (95% CI = 916 to 348) Ma (611 and 541 Ma according to the "DPANN-basal" tree, Figs. 2B and S11), covering the Marinoan Glaciation (~650 to 632.3 Ma) to Ediacaran (635 to 541 Ma) period. This period is also highlighted by the rise of eukaryotic algae (27), which might have enhanced the production of organic substrates for the growth of Poseidoniaceae (34). After that the divergence of brackish and marine lineages of Poseidoniaceae happened multiple times in an algal-dominant ocean until a most recent major divide between the stem groups of subclades BK6 and M (a) around 108 (95% CI = 254 to 46) Ma (Fig. 2B).

The evolutionary rates of Poseidoniaceae and Thalassarchaeaceae were remarkably different in the past 700 Ma, probably reflecting different levels of influence from the land. The rapid increase in evolutionary rate of Poseidoniaceae started generally with the Cryogenian glaciation and accelerated in the Ediacaran-Cambrian period (Figs. 2D and S12). After that, the evolutionary rate increased in a constant pattern until the present. This tendency is consistent with the increase of continental weathering (20, 21), coastal sedimentary phosphorite (22), and the rise of algae (27) (Fig. 2E). It is also consistent with the origination and diversification of brackish species Poseidoniaceae after the Snowball Earth and throughout the Phanerozoic (Fig. 2B), implying a potential impact of land nutrient supply and consequently coastal algal bloom on the evolution of nearshore-inhabiting Poseidoniaceae. In contrast, the evolutionary rate of Thalassarchaeaceae stayed at a low level in the past 1 billion years (Figs. 2D and S12) in line with their preferred environments that are oligotrophic and distant from the shore.

Genetic and geological integration supporting the origination of brackish Poseidoniaceae

The emergence of novel life forms in earth's history was likely triggered by a mixture of geological and ecological changes, while genetic innovation made it possible. However, collecting mutually corroborated lines of evidence to prove such ancient processes is a formidable task, especially for microorganisms that lack fossil

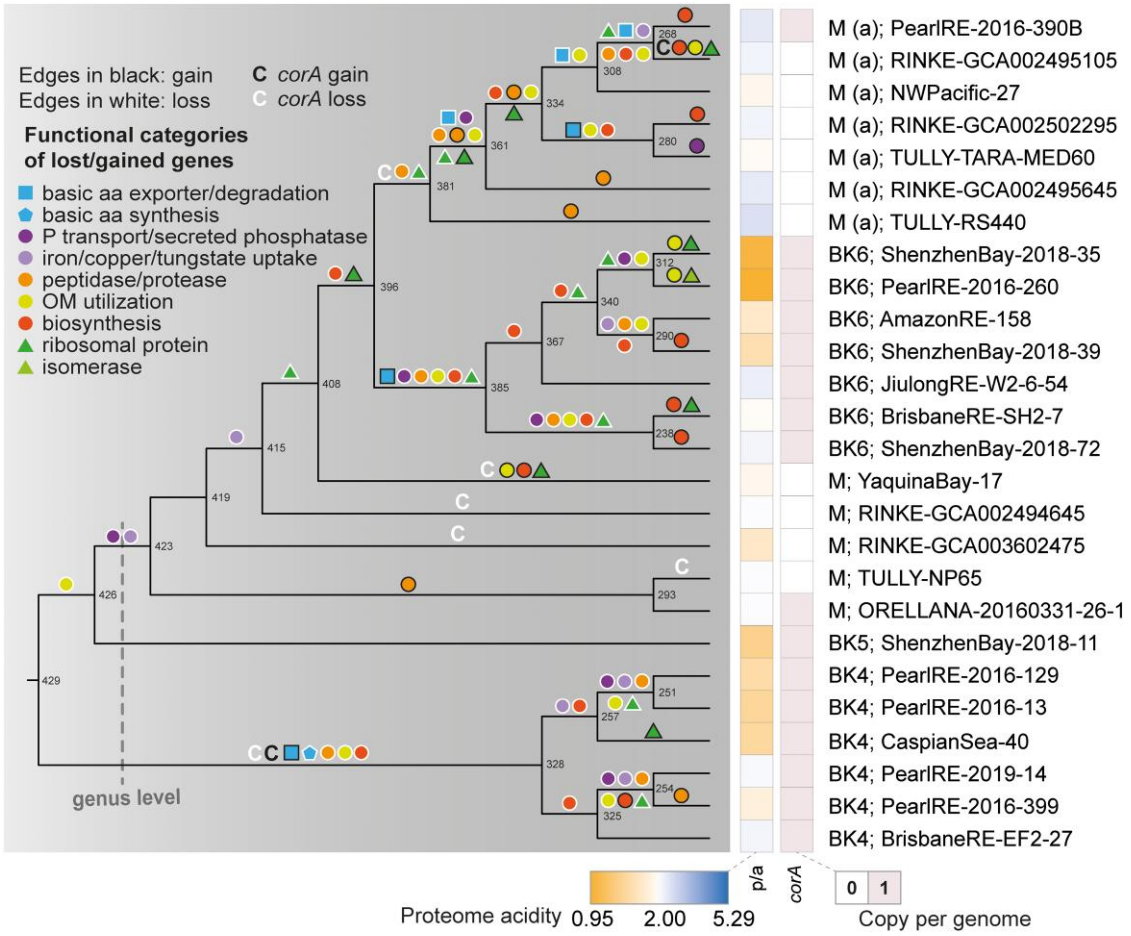

Fig. 3. Essential gene-gain and gene-loss events in the marine–brackish divergence of the BK4-BK5-M-BK6 monophyletic clade of Poseidoniaceae. The cladogram of the BK4-BK5-M-BK6 subclade of the tree in Fig. S13A is shown. The letter codes of the genus-level subgroups are consistent with Rinke et al. (32). Genus-level cutoff in the tree is derived from RED of GTDB taxonomic ranking (46). Small letters in brackets show subgenus-level subgroups. Node numbers are shown in consistence with Table S4. Gene-gain and gene-loss events between adjacent internal nodes or between adjacent internal nodes and terminal taxa are illustrated on relevant branches. Capital C means the *corA* gene. MAG completeness values range from 66.5 to 100% (median = 93.63%), and the contamination values range from 0 to 4.35% (median = 0).

records. Here, based on the detailed genetic features of brackish vs. marine lineages and the trend of global environmental changes on continental margins in history, we reconstruct the possible trajectory in the origination of brackish Poseidoniaceae (Fig. 4). After the divergence of the stem groups of Poseidoniaceae and Thalassarchaeaceae in about 1.1 Ga, the *corA* gene inserted at the end of an archaeal stress-response gene cluster in earlier times was inverted, providing a genetic predisposition for brackish adaptation. Several hundred million years later, on the postglacial Ediacaran coasts, a rapid increase in phosphorus from land supply and organic-matter release might have led to the rise of algae. The nutrient-stimulated algal boom should be most significant in coastal waters, and the flourishing of algal-derived organic matter then fed the emerging animal species (59) in a transforming food web on the continental shelf. These algal-derived organic matters also accelerated the evolution and diversification of coast-inhabiting Poseidoniaceae, possibly along with other heterotrophic bacteria and micro- and macropredators, and resulted in a highly competitive community on the Ediacaran-Cambrian coasts. At the same time, enhanced nutrient supply might have stimulated algal blooms in the expanded brackish water bodies during the postglacial transgression (31). This food-rich but predator-poor environment formed a remarkable niche gap. Consequently, Poseidoniaceae populations were continually selected when (i) the *corA* gene formed co-transcription with upstream stress-response genes to tolerate the stress of salinity fluctuation, (ii) the proteome was less acidified to meet lower average salinity, and (iii) the metabolism was better adapted to more nearshore environments. This gradual tuning in gene regulation, proteome amino acid composition, and substrate metabolism finally led to the origination of the brackish Poseidoniaceae. Later, when land weathering and phosphorus supply kept increasing in Phanerozoic, some brackish lineages continued exploring upstream habitats with salinity going down to 1‰ and even developed capabilities to utilize land-plant derived organic compounds, which increased rapidly from Silurian to Carboniferous (60). On the other hand, some lineages lost the *corA* gene and returned to marine habitats. In some rare cases, they might further gain the *corA* genes, possibly through recombination and shuffle between marine and brackish niches in expanded shallow marine environments (61).

Conclusion

The origin of global brackish microorganisms is a long-standing question. In this study, we apply detailed genome comparison and evolutionary analysis to investigate the key and detailed genetic changes causing the origination of the brackish lineages of the

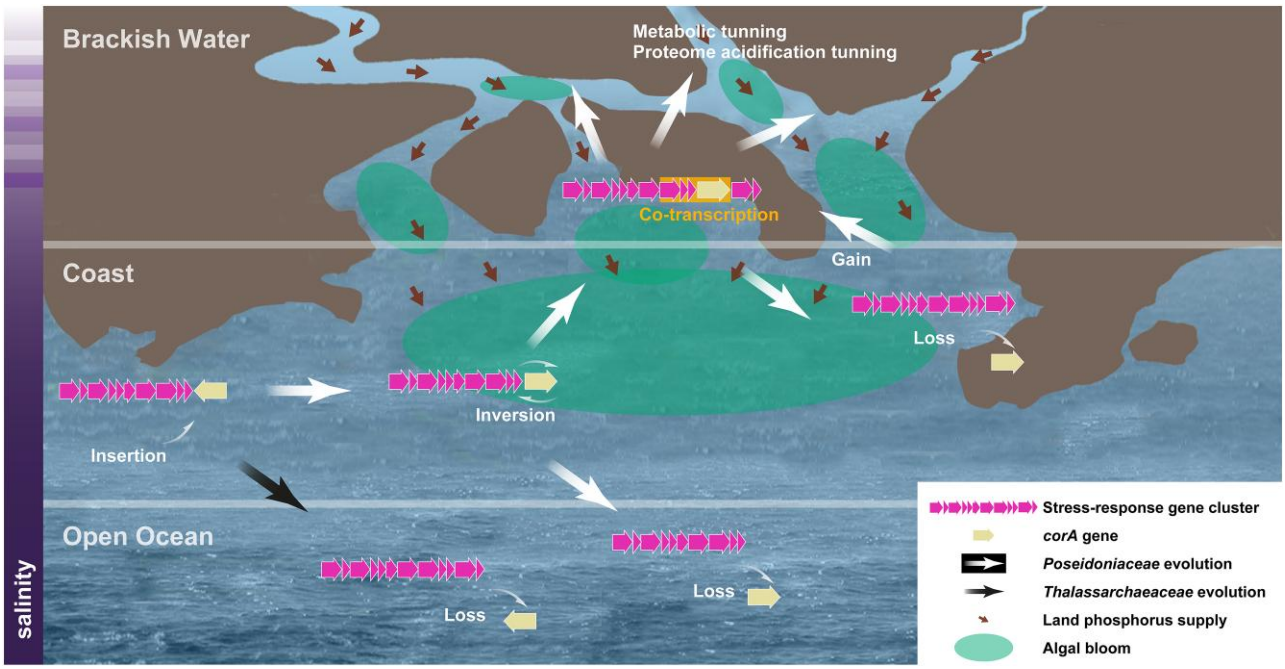

Fig. 4. A schematic diagram shows the process of genetic changes during the transition of Poseidoniales between salinity distinct habitats and environmental features in the origination of brackish Poseidoniaceae in the Ediacaran-Cambrian coasts.

most abundant archaeal heteroplankton Poseidoniales. We find that the divergent and globally distributed brackish lineages of Poseidoniaceae have a single origination. Importantly, we identify the *corA* gene as the key gene for adaptation in brackish environments, which may regulate intracellular magnesium concentration to stabilize macromolecules. The regulation of this gene was likely coupled to a highly conserved archaeal stress-response gene cluster after an ancient inversion event, allowing concerted responses to multiple stresses, including fluctuating salinities in brackish waters. Importantly, we find that the losses and gains of *corA* were followed by metabolic acclimation and diversification mediated by lateral gene transfer of substrate-specific genes. Molecular dating and geological inference suggest brackish Poseidoniaceae in the Ediacaran-Cambrian possibly flourished due to the rise of algal productivity in coastal and even brackish waters because of enhanced phosphorus supply in response to land weathering (62).

Our results highlight the coordination of genome rearrangement and global environmental changes driven by major geological events in the emergence of novel planktonic archaea. On the other hand, the emergence of these brackish microbes filled the new niche gaps in the Ediacaran-Cambrian coasts and contributed to the recycling of organic matter produced by algae through the microbial loop; thus, brackish Poseidoniales may have been an important player in the microbial carbon pump at that time (38, 63). The accelerated evolutionary rate of coastal Poseidoniales since the Ediacaran period indicates that evolution of both the macroheterotrophs (the animals) (29) and microheterotrophs (Poseidoniaceae and other heterotrophic archaea or bacteria) has been stimulated since the late-Proterozoic. Therefore, these heterotrophic archaea and bacteria could have enhanced the biomineralization of organic matter in seawater (61) after the Snowball Earth.

We are, however, aware of the caveats that are inherited with molecular clock calculation of geological ages of ancient microorganisms, which are improving in precision given the growing database of high-quality genomes and rapid development in methodology (55, 64, 65). We expect that our prediction of the genetic changes in habitat specificity of Poseidoniales can be further validated with in vivo experiments when pure or enriched cultures are available (34). In addition, modeling Ediacaran coastal hydrodynamic and biogeochemical cycles with sedimental records of biomarkers for specific brackish microorganisms may provide further support for our findings.

Materials and methods

Sample and dataset collection

Microplanktons were collected at the Pearl River estuary in 2011, 2012, 2016 (66), 2018 (in the adjacent Shenzhen Bay), and 2019 (National Omics Data Encyclopedia, NODE, https://www.biosino.org/node/, project: OEP001662), the Jiulong River estuary in 2018 (NODE project: OEP000961), the Yangtze River estuary in 2016 (NODE project: OEP001524), the Brisbane River estuary in 2020 (NCBI project: PRJNA872317), and the Northwest Pacific in 2015 and 2017 (NODE project: OEP001662; Table S1).

Immediately after collection, surface water was first filtered through 2.7-μm pore-size glass fiber filters (Shanghai Mosutech, Shanghai, China) to remove large particles, and the filtrates were then filtered through 0.22-μm pore-size membrane filters (Pellicon cartridge, Millipore Corp., Billerica, MA, USA) to collect microbial cells. Filters were then frozen in liquid nitrogen and stored at −80 °C in the laboratory untill further processing. DNA was extracted by using the FastDNA SPIN kit for soil (MP Biomedicals, Solon, OH, USA) following the manufacturers' instructions. Metagenome sequencing was conducted on an Illumina HiSeq 2500 platform at Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). Raw reads of the published metagenomes of the Caspian Sea (67), the Baltic Sea (5, 68–70), the Columbia River estuary (71), the Amazon River estuary (72), the Yaquina Bay (73), the Helgoland coast (37), the Port Hacking coast (32), and the *Tara* Oceans Project (74) were downloaded

from public databases (Table S1). Clean reads of the above metagenomes were generated by using the reads_qc module of MetaWRAP (v. 1.2.1) (75).

Generation of the global nonredundant Poseidoniales genome dataset

To obtain potential brackish Poseidoniales genomes, we used IDBA-UD (v. 1.1.3) (76) to assembly clean reads of the metagenomes of the Pearl River estuary, Shenzhen Bay, the Brisbane River estuary, the Jiulong River estuary, the Yangtze River estuary, the Columbia River estuary, the Amazon River estuary, Caspian Sea, and Baltic Sea (Table S1). Contigs longer than 2 kb were used for binning by using the binning module of MetaWRAP recruiting metaBAT2 (77), Maxbin2 (78), and CONCOCT (79) methods. Bins with completeness >50% and contamination <10% as evaluated by CheckM (v. 1.0.5) (40) and those classified as Poseidoniales by GTDB-Tk (v. 1.3.0, release 95) (80) were used for further selection and analyses. Previously published Poseidoniales genomes generated by Rinke et al. (32), Tully (33), and Orellana et al. (37) were downloaded from their online deposits. The combination of downloaded genomes with those generated in this study results in 835 Poseidoniales MAGs (Table S2). Potential contaminant contigs in each MAG were further removed by manual check aided by using acdc (v. 1.2.1) (81). A nonredundant MAG dataset was generated by using dRep (v. 2.6.2) (82) with a cutoff of 99% average nucleotide identity. This dataset contains 455 Poseidoniales MAGs. The quality check and taxonomic classification of these MAGs were conducted using CheckM and GTDB-Tk, respectively. Specifically, a manually selected subset containing 148 MGII-specific single-copy marker genes (present in at least 328 of the 455 Poseidoniales MAGs) were selected for MAG completeness assessment when CheckM was used (Table S5). Genes and proteins of the MAGs were predicted using Prodigal (v. 2.6.3) (83).

Poseidoniales abundance and activity calculation

To profile the distribution of Poseidoniales in global marine surface water, the nonredundant MAGs were mapped by clean reads of metagenomes and metatranscriptomes obtained from surface samples of the Pearl River estuary, Shenzhen Bay, the Jiulong River estuary, the Yangtze River estuary, the Columbia River estuary, the Amazon River estuary, Caspian Sea, Baltic Sea, the Helgoland region, the Port Hacking offshore region, the Yaquina Bay, Northwest Pacific, and the Tara oceans project (Table S1). To minimize potential unspecific mapping, rRNA and tRNA genes in the MAGs were identified by using Metaxa (v. 2.2) (84), and low-complexity regions were predicted by using DustMasker (v. 1.0.0) (https://github.com/ncbi/ncbi-cxx-toolkit-conan). These regions of the MAGs were masked by using Bedtools (v. 2.27.1) before mapping. Read mapping was conducted by using Bowtie2 (v. 2.3.5) (85) and followed by sorting and format convert to BAM files by using SAMtools (v. 1.9) (86). The BAM files were filtered by using BamM (v. 1.7.3) (https://github.com/minillinim/BamM); with thresholds of 99% identity and 75% read coverage. Finally, bbmap (http://jgi.doe.gov/data-and-tools/bb-tools/) was used to calculate read counts for each contig and the reads per kb of each genome per millions of reads (RPKM) was calculated for each MAG in each sample, respectively.

Proteome acidity estimation

The pIs of proteins in MAGs were calculated by using Pepstats from the EMBOSS package (87). pI frequency distribution of a proteome was calculated, as previously described (42). Proteome acidity in this study is defined as the ratio of the frequency of the acidic peak (pI 4.5) to the frequency of the semi-acidic peak (pI 6.25), as shown in Fig. S4.

Habitat salinity range analysis

Habitat salinity was investigated by calculating the abundance (RPKM) of Poseidoniales MAGs in metagenomes from diverse salinities (Table S1). A MAG is considered present in a metagenome if its RPKM value is above 0.01. The up-limit habitat salinity of a Poseidoniales taxon is set as the highest salinity where it is present, and the down-limit is set as the lowest salinity where it is present. Its optimum habitat salinity is set as the salinity, where it has the highest RPKM value.

Phylogenomic analysis of the nonredundant Poseidoniales MAGs

For the phylogenetic tree in Fig. S3, we used hmmsearch (v. 3.1b2; -E 1E-5) (88) to search for the 122 archaeal single-copy marker proteins (40) in the 455 nonredundant Poseidoniales MAGs based on hidden Markov models in Pfam (89) and TIGRfam (90) databases. MGIII euryarchaeal and other archaeal genomes were used as the outgroup (Table S3). Eighty-three Marker proteins present in ≥60% taxa were retained and aligned, respectively, by using MUSCLE (v. 3.8.1551; –maxiters 16; Table S6) (91). The alignment matrixes were denoised by using trimAl (v1.2rev59; -automated1) (92) and then concatenated. Missing data were filled with gaps. A maximum-likelihood tree was reconstructed by using FastTree (v. 2.1.10; -gamma -lg) (93) and visualized in the Interactive Tree of Life (iTOL, v.5.1.1) (94). Assignment of the MAGs to genus-level subgroups was conducted based on GTDB-Tk classification (80) and named according to Rinke et al. (32).

The above selected 83 marker genes were used to reconstruct the phylogenetic tree in Fig. 1A. For the phylogenetic trees in Fig. S13 containing 188 taxa and the ones in Fig. S14 containing 231 taxa, and the one in Fig. S8, 39 of the 41 marker proteins described by Adam et al. (45) were used. The other two proteins were excluded because they are absent in most of the Poseidoniales MAGs in this dataset. The marker proteins of each MAG were identified according to the genome functional annotations. Alignment and trimming were conducted as above. Removal of compositional heterogeneous sites was conducted by applying a χ^2-score-based approach (95). Maximum-likelihood trees were reconstructed with the LG + C60 + F model implemented in IQ-Tree (v. 2.0.3) (96) and then visualized in iTOL. Ultra-fast bootstrapping supports for branches were calculated after 1,000 iterations.

For the tree of CorA in Fig. S7, protein sequences annotated as "ALR," "CorA," "LPE10," "MRS2," "NMR," or "ZntB" were obtained from the NCBI NR database, and a nonredundant dataset was generated by using CDHIT (v4.8.1) (97) with 70% identity cutoff. A tree was inferenced by using FastTree. Taxa on the branches next to Poseidoniales and MGIII were selected for tree reconstruction. For this tree and trees of CorA and ZnuA in Fig. S8, protein sequences were aligned by using MAFFT L-INS-I (98) and denoised by using trimAl (automated1). The trees were constructed by using IQ-Tree with the parameters "-seqtype AA -m MFP -B 1000 –bnni."

Functional annotation and comparison of MAGs

Protein sequences of MAGs were annotated based on the KEGG database by using kofamscan (99), and the COG (100), arCOG (101), Pfam (102), and Tigrfam databases (90) by using BLASTp

(103) (E-value $<10^{-3}$, bit score >50, similarity >50%, and coverage >70%), respectively. The accuracy of annotation was further verified by clustering all the protein sequences of MAGs based on 95% similarity before annotation. Genes specifically enriched in brackish Poseidoniales were defined by plotting and visualizing the presence percentage of all the annotated genes (arCOG and KEGG) in the brackish clades and the normal marine clades of Poseidoniaceae (Fig. S5).

To test whether MAG completeness has an impact on the presence/absence result of *corA* in Poseidoniaceae, the *F* test was conducted in Microsoft Excel for Mac (v. 16.69), and a *P*-value of 0.002 was obtained, suggesting the variances of the two groups were not significantly different. The Student's *t* test was then conducted by setting a two-tailed distribution and type of two-sample equal variance.

To compare proteome acidity values of Poseidoniaceae with and without *corA*, the *F* test was conducted in Microsoft Excel for Mac (v. 16.69), and a *P*-value of 0.002 was obtained, suggesting the variances of the two groups were significantly different. The Student's *t* test was then conducted by setting a two-tailed distribution and type of two-sample unequal variance.

Amalgamated likelihood estimation analysis

The ALEml_undated algorithm of the ALE package (104) was used to reconcile the functional gene tree against the phylogenomic tree to infer the numbers of duplication, loss, transfer (within the sampled genome set), and origination (including both transfer from other phyla outside the species tree or de novo gene formation) on each branch of the Thermoplasmatota species tree. The results were visualized in iTOL.

Co-transcription analysis

We analyzed the transcriptional levels of *corA* and the syntenous genes of Poseidoniales by visualizing the mapping profile of the metatranscriptomic reads on the contigs of MAGs. Because Poseidoniales are in relatively low abundance in these samples, to obtain enough read mapping depth for analysis, we pooled the BAM files of all the metatranscriptomic samples from the Pearl River estuary, the Columbia River estuary, and the Amazon River estuary generated from the above metatranscriptomic mapping analysis (Fig. S2). Read distributions on the open reading frame annotated contigs were visualized in Geneious (version 2022.2 created by Biomatters. Available from https://www.geneious.com).

Gene-gain and gene-loss analyses for the BK4-BK5-M-BK6 monophyletic clade of Poseidoniaceae

Selected MAGs with relatively high quality (completeness values range from 66.5 to 100%, median = 93.63%; contamination values range from 0 to 4.35%, median = 0) were processed for analysis. The event number of a gene (KO or arCOG entry) in a terminal taxon (MAG) is set to 1 if the gene is present and 0 if absent. The event number of a gene in an internal node is defined as the DTLO event numbers calculated by applying the branchwise_numbers_of_events.py script described by Sheridan et al. (105). Gene-gain and gene-loss events between adjacent internal nodes or between adjacent internal nodes and terminal taxa are defined as the following: (i) a loss event is defined if the event number of the older node (an internal node) is >0.8 and is eight times greater than that of the younger node (an internal node or a terminal taxon); and (ii) a gain event is defined if the event number of the younger node (an internal node or a terminal taxon) is >0.8 and eight times greater than that of the older node (an internal node).

Molecular clock analysis and speciation rate calculation

Node divergence time of the 231-taxa maximum-likelihood trees was estimated by using RelTime in MEGA X (v10.1.5) with the LG + G model and with 95% CI (106). The root of Archaea (4.38–3.46 Ga) (107–109), three constraints related to the Great Oxygenation Event (i.e. the roots of Thermoproteales, Sulfolobales, and Thermoplasma, <2.32 Ga) and one related to the gene transfer between cyanobacteria and methanogens (the latest common ancestor of Methanosarcinales and Methanomicrobiales) (110, 111) were used for calibration (Table S7). The evolutionary rate of each branch in the time trees was estimated by BAMM (v2.5.0) (112).

Acknowledgments

The authors thank Zongjun Yin (Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences), Weiqi Yao (Southern University of Science and Technology), and Jie Li (Institute of Microbiology, Chinese Academy of Sciences) for their valuable suggestions in geological records and archaeal transcription, and Glynn Gorick (Cambridge, UK) for his contribution to the artwork of Fig. 4. Computation in this study was supported by the Centre for Computational Science and Engineering at the Southern University of Science and Technology.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This study was supported by the National Natural Science Foundation of China (nos. 92351301, 42376113, and 91951120), the Guangdong Basic and Applied Basic Research Foundation (no. 2021B1515120080), the Open Project of Key Laboratory of Environmental Biotechnology, CAS (no. KF2021006), the Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Southern University of Science and Technology (no. ZDSYS201802081843490), the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (no. K19313901), and the Project of Educational Commission of Guangdong Province of China (no. 2020KTSCX123).

Author Contributions

L.F. and C.Z. conceived this study. B.X., S.C., F.L., W.X., A.P., D.Z., R.W., H. Li, H. Liu, Y.L., S.-J.K., J.C., Y.Z., C.R., and M.L. collected the samples and extracted DNA. L.F., B.X., S.C., Y.L., F.L., A.P., D.Z., R.W., and H. Li analyzed the metagenome data, produced the genomes, and conducted all other analyses. L.F., B.X., M.Z., and C.Z. interpreted the results and drafted the manuscript. All authors contributed to the final version of the manuscript.

Data Availability

Metagenome sequences generated in this work have been deposited at National Omics Data Encyclopedia (NODE, https://www.biosino.org/node/) in projects OEP001662, OEP000961, OEP001662, and OEP001524, and in NCBI BioProject PRJNA1024631 with the SRA numbers SRR26337533, SRR26337544, SRR26337555,

SRR26337577, and SRR26337578. Download of the MAGs analyzed in this study is available at https://doi.org/10.6084/m9.figshare.23988984

References

1 Lee CE, Bell MA. 1999. Causes and consequences of recent freshwater invasions by saltwater animals. *Trends Ecol Evol*. 14:284–288.
2 Lozupone CA, Knight R. 2007. Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A*. 104:11436–11440.
3 Logares R, *et al.* 2009. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol*. 17:414–422.
4 Herlemann DP, *et al.* 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J*. 5: 1571–1579.
5 Hugerth LW, *et al.* 2015. Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol*. 16:279.
6 Filker S, Kühner S, Heckwolf M, Dierking J, Stoeck T. 2019. A fundamental difference between macrobiota and microbial eukaryotes: protistan plankton has a species maximum in the freshwater-marine transition zone of the Baltic Sea. *Environ Microbiol*. 21:603–617.
7 Ngugi DK, *et al.* 2023. Postglacial adaptations enabled colonization and quasi-clonal dispersal of ammonia-oxidizing archaea in modern European large lakes. *Sci Adv*. 9:eadc9392.
8 Jurdzinski KT, *et al.* 2023. Large-scale phylogenomics of aquatic bacteria reveal molecular mechanisms for adaptation to salinity. *Sci Adv*. 9:eadg2059.
9 Walsh D, Lafontaine J, Grossart HP. 2013. On the eco-evolutionary relationships of fresh and salt water bacteria and the role of gene transfer in their adaptation. In: Gophna U, editor. *Lateral gene transfer in evolution*. 1st ed. New York (NY): Springer. p. 55–77.
10 Eiler A, *et al.* 2016. Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *ISME J*. 10: 1902–1914.
11 Zhang H, *et al.* 2019. Repeated evolutionary transitions of flavobacteria from marine to non-marine habitats. *Env Microbiol*. 21: 648–666.
12 Simon M, *et al.* 2017. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J*. 11:1483–1499.
13 Penn K, Jensen PR. 2012. Comparative genomics reveals evidence of marine adaptation in *Salinispora* species. *BMC Genomics*. 13:86.
14 Cabello-Yeves PJ, *et al.* 2017. Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front Microbiol*. 8:1151.
15 Martijn J, *et al.* 2020. Hikarchaeia demonstrate an intermediate stage in the methanogen-to-halophile transition. *Nat Commun*. 11:5490.
16 Ren M, Wang J. 2022. Phylogenetic divergence and adaptation of Nitrososphaeria across lake depths and freshwater ecosystems. *ISME J*. 16:1491–1501.
17 Henson MW, Lanclos VC, Faircloth BC, Thrash JC. 2018. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J*. 12:1846–1860.
18 Wood R, *et al.* 2019. Integrated records of environmental change and evolution challenge the Cambrian explosion. *Nat Ecol Evol*. 3:528–538.
19 Smith MP, Harper DAT. 2013. Causes of the Cambrian explosion. *Science*. 341:1355–1356.
20 Halverson GP, Wade BP, Hurtgen MT, Barovich KM. 2010. Neoproterozoic chemostratigraphy. *Precambrian Res*. 182: 337–350.
21 Sharoni S, Halevy I. 2023. Rates of seafloor and continental weathering govern Phanerozoic marine phosphate levels. *Nat Geosci*. 16:75–81.
22 Reinhard CT, *et al.* 2017. Evolution of the global phosphorus cycle. *Nature*. 541:386–389.
23 Xiang L, Schoepfer SD, Zhang H, Cao C, Shen S. 2018. Evolution of primary producers and productivity across the Ediacaran-Cambrian transition. *Precambrian Res*. 313:68–77.
24 Butterfield NJ. 2009. Macroevolutionary turnover through the Ediacaran transition: ecological and biogeochemical implications. *Geol Soc Lond Spec Publ*. 326:55–66.
25 Chen X, *et al.* 2015. Rise to modern levels of ocean oxygenation coincided with the Cambrian radiation of animals. *Nat Commun*. 6:7142.
26 Zhang F, *et al.* 2019. Global marine redox changes drove the rise and fall of the Ediacara biota. *Geobiology*. 17:594–610.
27 Brocks JJ, *et al.* 2017. The rise of algae in Cryogenian oceans and the emergence of animals. *Nature*. 548:578–581.
28 Erwin DH, Valentine JW. 2013. *The Cambrian explosion: the construction of animal biodiversity*. Greenwood Village (CO): Roberts and Company.
29 Erwin DH, *et al.* 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science*. 334:1091–1097.
30 Pehr K, *et al.* 2018. Ediacara biota flourished in oligotrophic and bacterially dominated marine environments across Baltica. *Nat Commun*. 9:1807.
31 Hurtgen MT, Halverson GP, Arthur MA, Hoffman PF. 2006. Sulfur cycling in the aftermath of a 635-Ma snowball glaciation: evidence for a syn-glacial sulfidic deep ocean. *Earth Planet Sci Lett*. 245:551–570.
32 Rinke C, *et al.* 2019. A phylogenomic and ecological analysis of the globally abundant marine group II archaea (*Ca.* Poseidoniales ord. nov.). *ISME J*. 13:663–675.
33 Tully BJ. 2019. Metabolic diversity within the globally abundant marine group II *Euryarchaea* offers insight into ecological patterns. *Nat Commun*. 10:271.
34 Zhang CL, Xie W, Martin-Cuadrado A-BB, Rodriguez-Valera F. 2015. Marine group II archaea, potentially important players in the global ocean carbon cycle. *Front Microbiol*. 6:1108.
35 Damashek J, *et al.* 2021. Transcriptional activity differentiates families of marine group II *Euryarchaeota* in the coastal ocean. *ISME Commun*. 1:5.
36 Pereira O, *et al.* 2021. Seasonality of archaeal proteorhodopsin and associated marine group IIb ecotypes (Ca. Poseidoniales) in the North Western Mediterranean Sea. *ISME J*. 15:1302–1316.
37 Orellana LH, *et al.* 2019. Niche differentiation among annually recurrent coastal marine group II *Euryarchaeota*. *ISME J*. 13: 3024–3036.
38 Zhang C, *et al.* 2018. Evolving paradigms in biological carbon cycling in the ocean. *Natl Sci Rev*. 5:481–499.
39 Xie W, *et al.* 2017. Localized high abundance of marine group II archaea in the subtropical pearl river estuary: implications for their niche adaptation. *Env Microbiol*. 20:734–754.
40 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 25:1043–1055.
41 Fagerbakke KM, Norland S, Heldal M. 1999. The inorganic ion content of native aquatic bacteria. *Can J Microbiol*. 45:304–311.
42 Cabello-Yeves PJ, Rodriguez-Valera F. 2019. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome*. 7:117.

43 Maguire ME. 2006. The structure of CorA: a Mg(2+)-selective channel. *Curr Opin Struct Biol*. 16:432–438.
44 Payandeh J, Pfoh R, Pai EF. 2013. The structure and regulation of magnesium selective ion channels. *Biochim Biophys Acta*. 1828: 2778–2792.
45 Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of archaea: new perspectives on their diversity, evolution and ecology. *ISME J*. 11:2407–2425.
46 Rinke C, *et al.* 2021. A standardized archaeal taxonomy for the genome taxonomy database. *Nat Microbiol*. 6:946–959.
47 Laakso TA, Sperling EA, Johnston DT, Knoll AH. 2020. Ediacaran reorganization of the marine phosphorus cycle. *Proc Natl Acad Sci U S A*. 117:11961–11967.
48 Hoffman PF, Kaufman AJ, Halverson GP, Schrag DP. 1998. A neoproterozoic snowball earth. *Science*. 281:1342–1346.
49 Gao L, *et al.* 2020. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science*. 369:1077–1084.
50 Bugg TD, Ahmad M, Hardiman EM, Singh R. 2011. The emerging role for bacteria in lignin degradation and bio-product formation. *Curr Opin Biotechnol*. 22:394–400.
51 Shapiro BJ, Leducq J-B, Mallet J. 2016. What is speciation? *PLoS Genet*. 12:e1005860.
52 Retchless AC, Lawrence JG. 2007. Temporal fragmentation of speciation in bacteria. *Science*. 317:1093–1096.
53 Chen S-C, *et al.* 2020. The great oxidation event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc Natl Acad Sci U S A*. 117:10414–10421.
54 Yang Y, *et al.* 2021. The evolution pathway of ammonia-oxidizing archaea shaped by major geological events. *Mol Biol Evol*. 38:3637–3648.
55 Ren M, *et al.* 2019. Phylogenomics suggests oxygen availability as a driving force in *Thaumarchaeota* evolution. *ISME J*. 13: 2150–2161.
56 Betts HC, *et al.* 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol*. 2:1556–1562.
57 Boden JS, Konhauser KO, Robbins LJ, Sánchez-Baracaldo P. 2021. Timing the evolution of antioxidant enzymes in cyanobacteria. *Nat Commun*. 12:4742.
58 Blank CE. 2009. Not so old archaea—the antiquity of biogeochemical processes in the archaeal domain of life. *Geobiology*. 7:495–514.
59 Knoll AH. 2017. Food for early animal evolution. *Nature*. 548: 528–530.
60 Laakso TA, Strauss JV, Peterson KJ. 2020. Herbivory and its effect on phanerozoic oxygen concentrations. *Geology*. 48: 410–414.
61 Peters SE, Gaines RR. 2012. Formation of the 'great unconformity' as a trigger for the Cambrian explosion. *Nature*. 484:363–366.
62 Dodd MS, *et al.* 2023. Uncovering the Ediacaran phosphorus cycle. *Nature*. 618:974–980.
63 Jiao N, *et al.* 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat Rev Microbiol*. 8:593–599.
64 Szöllõsi GJ, *et al.* 2022. Relative time constraints improve molecular dating. *Syst Biol*. 71:797–809.
65 Wolfe JM, Fournier GP. 2018. Horizontal gene transfer constrains the timing of methanogen evolution. *Nat Ecol Evol*. 2: 897–903.
66 Xu B, *et al.* 2022. A holistic genome dataset of bacteria, archaea and viruses of the Pearl River estuary. *Sci Data*. 9:49.
67 Mehrshad M, Amoozegar MA, Ghai R, Shahzadeh Fazeli SA, Rodriguez-Valera F. 2016. Genome reconstruction from metagenomic data sets reveals novel microbes in the brackish waters of the Caspian Sea. *Appl Env Microbiol*. 82:1599–1612.
68 Larsson J, *et al.* 2014. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J*. 8:1892–1903.
69 Alneberg J, *et al.* 2018. BARM and BalticMicrobeDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. *Sci Data*. 5:180146.
70 Alneberg J, *et al.* 2020. Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol*. 3:119.
71 Fortunato CS, Crump BC. 2015. Microbial gene abundance and expression patterns across a river to ocean salinity gradient. *PLoS One*. 10:e0140578.
72 Satinsky BM, *et al.* 2014. The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome*. 2:17.
73 Kieft B, *et al.* 2018. Microbial community structure-function relationships in Yaquina Bay estuary reveal spatially distinct carbon and nitrogen cycling capacities. *Front Microbiol*. 9:1282.
74 Pesant S, *et al.* 2015. Open science resources for the discovery and analysis of Tara oceans data. *Sci Data*. 2:150023.
75 Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 6:158.
76 Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 28:1420–1428.
77 Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. 3:e1165.
78 Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 32:605–607.
79 Alneberg J, *et al.* 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 11:1144–1146.
80 Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*. 36:1925–1927.
81 Lux M, *et al.* 2016. , acdc—automated contamination detection and confidence estimation for single-cell genome data. *BMC Bioinformatics*. 17:543.
82 Olm MR, Brown CT, Brooks B, Banfield JF. 2017. Drep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 11:2864–2868.
83 Hyatt D, *et al.* 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 11: 119.
84 Bengtsson-Palme J, *et al.* 2015. METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour*. 15:1403–1414.
85 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9:357–359.
86 Li H, *et al.* 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078–2079.
87 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 16:276–277.
88 Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39: W29–W37.
89 Finn RD, *et al.* 2014. Pfam: the protein families database. *Nucleic Acids Res*. 42:D222–D230.

90 Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res*. 31:371–373.

91 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32: 1792–1797.

92 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25:1972–1973.

93 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 5: e9490.

94 Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 49:W293–W296.

95 Viklund J, Ettema TJG, Andersson SGE. 2012. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol*. 29:599–615.

96 Minh BQ, *et al*. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 37:1530–1534.

97 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 28:3150–3152.

98 Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30:3059–3066.

99 Aramaki T, *et al*. 2020. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 36:2251–2252.

100 Galperin MY, *et al*. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res*. 49:D274–D281.

101 Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life Basel*. 5:818–840.

102 El-Gebali S, *et al*. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res*. 47:D427–D432.

103 Camacho C, *et al*. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.

104 Szöllősi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013. Efficient exploration of the space of reconciled gene trees. *Syst Biol*. 62:901–912.

105 Sheridan PO, *et al*. 2020. Gene duplication drives genome expansion in a major lineage of *Thaumarchaeota*. *Nat Commun*. 11:5494.

106 Tamura K, Tao Q, Kumar S. 2018. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Mol Biol Evol*. 35:1770–1782.

107 Ueno Y, *et al*. 2009. Geological sulfur isotopes indicate elevated OCS in the Archean atmosphere, solving faint young sun paradox. *Proc Natl Acad Sci U S A*. 106:14784–14789.

108 Valley JW, *et al*. 2014. Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography. *Nat Geosci*. 7:219–223.

109 Wolfe JM, Fournier GP. 2018. Reply to 'molecular clocks provide little information to date methanogenic Archaea'. *Nat Ecol Evol*. 2:1678.

110 Blank CE. 2009. Phylogenomic dating—a method of constraining the age of microbial taxa that lack a conventional fossil record. *Astrobiology*. 9:173–191.

111 Bekker A, *et al*. 2004. Dating the rise of atmospheric oxygen. *Nature*. 427:117–120.

112 Rabosky DL, *et al*. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol Evol*. 5:701–707.