CorA gene rearrangement triggered the salinity-driven speciation of Poseidoniales

Lu Fan[1,2*], Bu Xu[1*], Songze Chen[1,2*], Yang Liu[3,4*], Fuyan Li[5], Wei Xie[6], Apoorva Prabhu[7], Dayu Zou[3,4], Ru Wan[8,9,10,11], Hongliang Li[10,11], Haodong Liu[1], Yuhang Liu[1], Shuh-Ji Kao[8,9], Jianfang Chen[10,11], Yuanqing Zhu[1,12], Christian Rinke[7], Meng Li[3,4], Chuanlun Zhang[1,2,12#]

[1] Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Department of Ocean Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen, China. [2] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China. [3] Archaeal Biology Center, Institute for Advanced Study, Shenzhen University, Shenzhen, China. [4] Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University, Shenzhen, China. [5] Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawaii, Honolulu, Hawaii, USA. [6] School of Marine Sciences, Sun Yat-sen University & Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China. [7] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia. [8] State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China. [9] State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen, China. [10] Key Laboratory of Marine Ecosystem Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China. [11] State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou, China. [12] Shanghai Sheshan National Geophysical Observatory, Shanghai Earthquake Agency, Shanghai, China.

* These authors contributed equally to this work.

Corresponding author:

Chuanlu Zhang, zhangcl@sustech.edu.cn

running title: **Salinity-driven speciation by changing one gene**

ABSTRACT

The rise of microbial species is associated with multiple genetic changes and niche reconstruction [1,2]. While recombination, lateral gene transfer and point mutations can contribute to microbial speciation [3], acquisition of niche-specific genes was found to play an important role in initiating ecological specialization followed by genome-wide mutations [4]. The critical step at the very early microbial speciation between ecologically distinct habitats, such as land and ocean, however, is elusive. Here we show that the divergence of archaea Poseidoniales between brackish and marine waters was triggered by rearranging a magnesium transport gene *corA* in a global geological background. The *corA* gene was inserted within a highly conservative gene cluster and possibly function in concert with the other genes in this cluster in osmotic stress response. It then went through sporadic losses and gains that were coincident with the Pangea tectonic activities and sea-level rising. Notably, metabolic adjustment and proteome-wide amino acid substitution were found after the change of *corA*. Our results highlight salinity adaptation as the primary factor in microbial speciation at the interface between land and ocean. Such a process can start from simply changing one gene but may need coherent gene cluster rearrangement and work in tune with strong selective forces such as global landform changes.

Introduction

The origin of species is a fundamental question of evolution [5]. While the concept of microbial species is still in discussion, speciation has been considered essential in studying microbial pangenome, phylogenetics, biogeography, pathogen emergence, and ecological impacts of global changes [6]. Modeling the historical process of microbial speciation in natural environment is challenging because of the complex relationship between genetics and ecology of microorganisms [1]. Limited studies of microbial pangenomes in the last decade have proposed several modes of microbial speciation; one of them suggests selection drives speciation and is followed by genome-wide divergence [3]. This mode is supported by studies identifying niche-specific genes potentially critical in initial speciation of recently diverged clinical and marine bacteria [4,7–9].

A critical knowledge gap, however, is the trajectory of microbial speciation during evolutionary transitions between distinct habitats such as land and ocean in the deep history of the earth. Marine and freshwater ecosystems are different in various environmental factors such as salinity, nutrients, and competitors [10]; it is thus challenging to identify the key selective force triggering the speciation. Exploring new living substrates in a different habitat may be a strong motive [11], but considering the difficulty in microbial acclimation to different environmental salinities [12], such a benefit is likely insufficient in driving the transition. On the other hand, land-ocean transition of many microbial lineages is infrequent and ancient which is so called the 'salinity divide' [13,14]. The lack of reliable geological records leaves such speciation events in mists.

Another fundamental question is whether the genetic change in the initiation of the transition is caused by a sudden event (e.g. gene loss, gene gain or rearrangement) or a result of gradual and cumulative effects (e.g. genome-wide mutation, or combination of genes in adaptation to multiple niches). Recent genome comparisons among marine and non-marine subgroups of Pelagibacterales [15], Flavobacteria [11], *Rhodobacteraceae* [16], Actinobacteria [17], Hikarchaeia [18], *Synechococcus* [19] and Nitrososphaeria [20] have revealed common functional differences in osmotic regulation and substrate specificity that play a key role in microbial adaptation to habitats of different salinities. However, details of the gene-level transition across the salinity barrier are still obscure. Eiler et al. addressed this question by studying the freshwater group LD12 of Pelagibacterales and suggested a gradual tuning of metabolic pathways and transporters towards organic substrates in freshwater environments [15]. In contrast, Henson et al. proposed that the irreversible loss of two osmolyte transporters could be the critical

step in the formation of the LD12 clade [21]. These differing opinions likely resulted from the lack of genomes representative of the intermediate state during the marine-freshwater transition of Pelagibacterales. Therefore, studies analyzing larger genome sets are required to capture a more refined scale of the evolutionary trajectory are required to answer the question asked above.

Marine Group II Euryarchaea (now known as Poseidoniales) are among the most abundant archaeal plankton in global oceans with great ecological potential [22,23]. They are so far only found in marine waters. While Poseidoniales are still not cultured, genomic analyses support that they have a heterotrophic lifestyle by remineralizing organic matter such as algal-derived substrates [22–25]. Poseidoniales include two family-level subgroups, MGIIa (Poseidoniaceae) and MGIIb (Thalassarchaeaceae) [22]. Poseidoniaceae are found to be dominant in coastal areas where algal oligosaccharides would be more readily available, while most Thalassarchaeaceae are adapted to mesopelagic and oligotrophic waters where direct algal inputs are limited [26,27]. Recently, high abundance of Poseidoniales 16S rRNA genes was reported at the Pearl River estuary mixing zone with salinity below 15 practical salinity unit (PSU) [28], suggesting these populations of Poseidoniales might have evolved to adapt to brackish waters. However, detailed genetic evidence has not been provided because of the lack of genomic representatives of brackish-specific lineages of Poseidoniales.

The identification of brackish Poseidoniales

In this study, we sampled 128 global brackish metagenomes (**Fig. S1**, **Table S1**) to identify Poseidoniales lineages specifically adapted to brackish waters and to reconstruct their detailed evolutionary trajectory between salinity-distinct habitats. Poseidoniales metagenome-assembled genomes (MAGs) were reconstructed and phylogenetically analyzed together with those obtained in recent studies [22,23,27]. This approach contributed 94 (20.7%) novel genomes to the updated non-redundant Poseidoniales genome dataset (455 MAGs, completeness > 47%, completeness median = 70.76%, contamination < 10%, contamination median = 0.8%) based on a cutoff of 99% average nucleotide identity (ANI) (**Table S2**) and thus has filled a significant gap in species diversity of global low-salinity Poseidoniales. In the phylogenomic tree, genomes from global estuaries and enclosed seas form several clusters in the subclade of Poseidoniaceae (**Fig. S2**). Many of those clusters show remarkable evolutionary distance from adjacent oceanic relatives.

To map the distribution of Poseidoniales in global coastal and pelagic surface waters with distinct salinities, we calculated their abundances in 267 metagenomes and 21 metatranscriptomes of two low-salinity enclosed seas, four major estuaries, four coastal regions, and eight pelagic regions of global oceans. In general, Poseidoniaceae are in high abundance in inland seas and estuaries, while Thalassarchaeaceae are only detected in most of the nearshore and all the pelagic samples. This spatial distribution pattern is in line with the previous observation that Poseidoniaceae are adapted to more eutrophic and diverse coastal environments [22,23]. Notably, two patterns of abundance distribution along the salinity gradient are detected in Poseidoniaceae: a 'salt-preferred' pattern in which their abundance increased with the increase of salinity and a 'brackish specific' pattern in which they were enriched in salinity between 6.6 to 23 PSU but depleted or absent in salinity beyond this range (**Fig. 1A**). The metatranscriptomic analysis supports that Poseidoniaceae of both patterns are in an active state (**Fig. S3**).

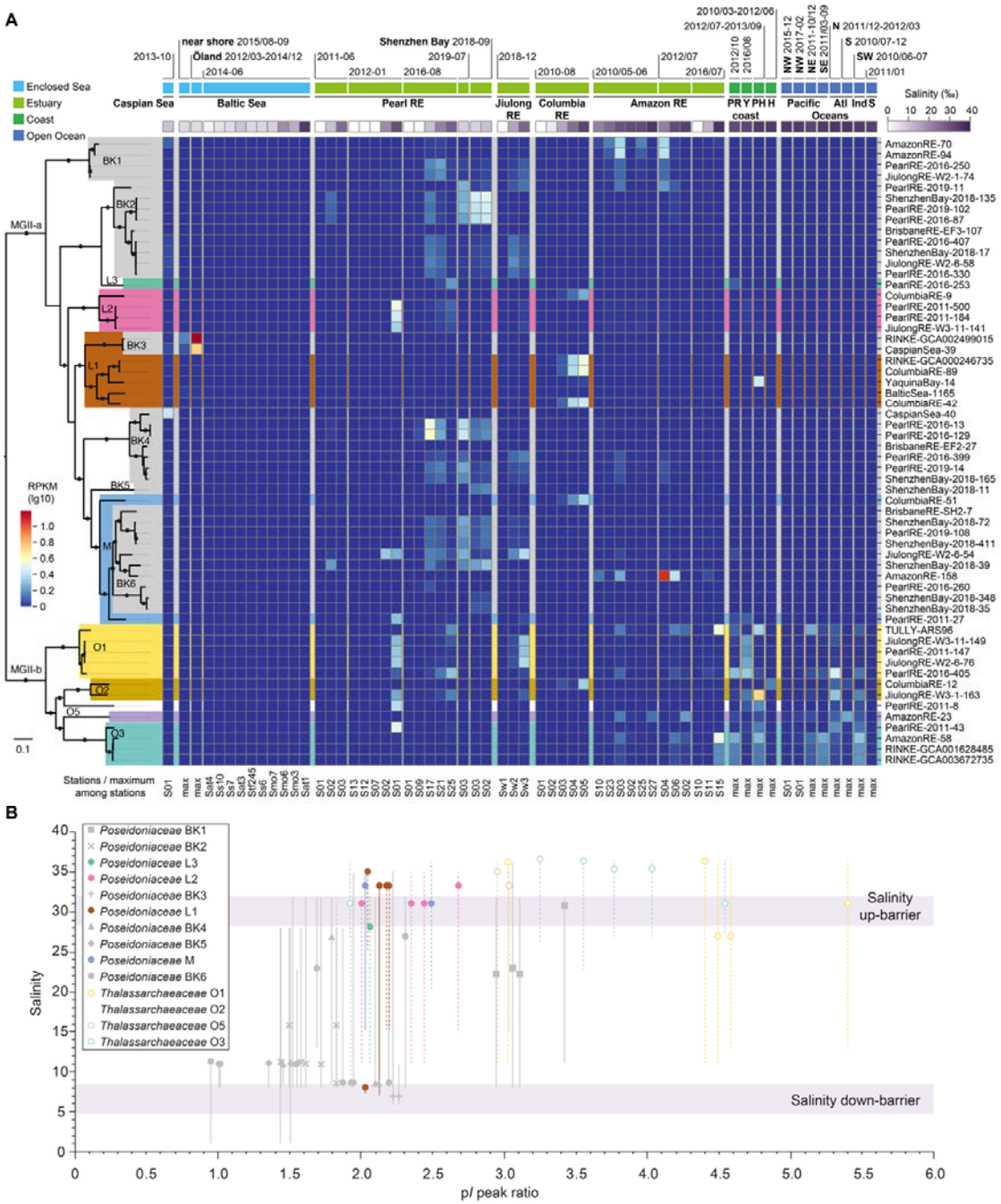

Fig. 1 | Global distribution and proteome adaptation of Poseidoniales. (**A**) Abundance pattern of Poseidoniales MAGs based on 246 metagenome samples from surface waters of global marine and brackish environments. MAGs with RPKM value above 0.4 in enclosed sea and estuarine samples are shown. In the maximum-likelihood tree at the left of the panel, solid dots on internal branches show branch supports of ultra-fast bootstrapping (1000) in IQTree > 90%. Shades on branches show Poseidoniales genera with the color code according to Rinke et al. 2019 [22], except for brackish clades which are in grey. Abbreviations of sampling areas

are: RE = river estuary, PR = Pearl River estuary, Y = Yangtze River estuary, PH = Port Hacking, H= Helgoland, Alt = Atlantic, Ind = Indian, S= Southern/South, N = North, NW = Northwest, NE = Northeast, SW = Southwest, SE= Southeast. Time ranges show the sampling period spans. 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 show Poseidoniales MAGs. Filled circles and solid lines belong to Poseidoniaceae MAGs, while empty circles and dashed lines belong to Thalassarchaeaceae MAGs. The position of each dot on the y-axis shows the optimal salinity of that MAG. The scale of each line on the y-axis shows the upper and down limits salinity of that MAG. The color code of dots and lines shows MGII genera according to Rinke et al. 2019 [22], except for brackish clades which are in grey.

Based on this observation, six monophyletic brackish-specific clades can be identified in the tree of Poseidoniaceae (**Fig. 1A and Fig. S2**). They either branch within or as close sisters to certain previously classified Poseidoniaceae genera [22]. Some of these brackish-specific lineages are found in different estuaries globally, while others are found to be enriched only in one estuary or one enclosed sea (**Fig. 1A**), implying that global dispersion and local adaptation may both have a function (see discussion in ref [29]). Repeated presence but interannual variability in abundance of these brackish specific lineages are observed in the Baltic Sea, the Pearl River estuary and the Amazon River estuary, suggesting they have specifically adapted to coastal brackish waters but factors other than salinity might impact their temporary abundances [27].

Acidified proteome isoelectric point (p*I*) was recognized to be a strong indicator of microorganisms inhabiting saline environments [12], which is a response to higher intracellular ion concentration [30]. The p*I* distribution of Poseidoniales proteomes shows typical acidification, suggesting they may import cations to balance osmotic pressure in seawater (**Fig. S4, SI**). Indeed, potassium transporter Trk is present in all Poseidoniales genomes (**Fig. S5, SI**). To verify that the brackish-enriched Poseidoniaceae are specifically adapted to low-salinity habitats because of evolution, we calculated the optimum salinity values of the MAGs (*i.e.*, the salinity of an environment in which a MAG has the highest abundance) detected in inland and estuarine samples and plotted them according to the estimated acidity of their proteome p*I* patterns. **Fig. 1B** clearly shows a correlation between proteome acidity and salinity adaptation -- most Poseidoniaceae with acidity above 2.0 enriched in

salinity from 20 to 35 PSU, with up-limit to over 36 PSU. All Thalassarchaeaceae MAGs have acidity values over 2.9 (except one) and optimum salinity over 30 PSU (except two). In contrast, Poseidoniaceae MAGs of acidity below 2.0 belong to brackish-specific clades and have optimum salinities below 30 and down to 8 PSU, which were detectable even in river mouth at a salinity around 1 PSU but never detected at a salinity above 32 PSU. This observation strongly suggests that the distinct distribution pattern of marine and brackish Poseidoniaceae subgroups results from long-term divergent evolution [12].

Evolution of key genes in salinity adaptation

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 [13]. To identify key genes potentially responsible for differentiating the marine and brackish Poseidoniaceae subgroups, we annotated genes of Poseidoniales MAGs and conducted gene-centered comparison between the group containing all brackish Poseidoniaceae and the group containing all marine Poseidoniaceae. Remarkably, two divalent cation transporters stood out as the only genes distinguishing these two groups -- the magnesium transporter CorA and the zinc/manganese ABC transport complex ZnuABC. *CorA* is the only gene present in almost all genomes of brackish clades while it was absent in nearly all other Poseidoniales genomes that contained the *znuABC* gene set (**Fig. 2C**, **Fig. S5**). This observation suggests that intracellular magnesium may be the key factor in low-salinity adaptation of brackish Poseidoniaceae, while zinc or manganese is required by marine subgroups of Poseidoniaceae. In addition, an additional extensive analysis of genes potentially involved in osmotic regulation (**SI**) suggests that three proteins/protein complexes are possibly responsible for the observation that Poseidoniaceae are more adapted to dynamic coastal waters while Thalassarchaeaceae generally restrict their habitat in pelagic zones with a more stable salinity [27] (**Fig. 2C, Fig. S5**).

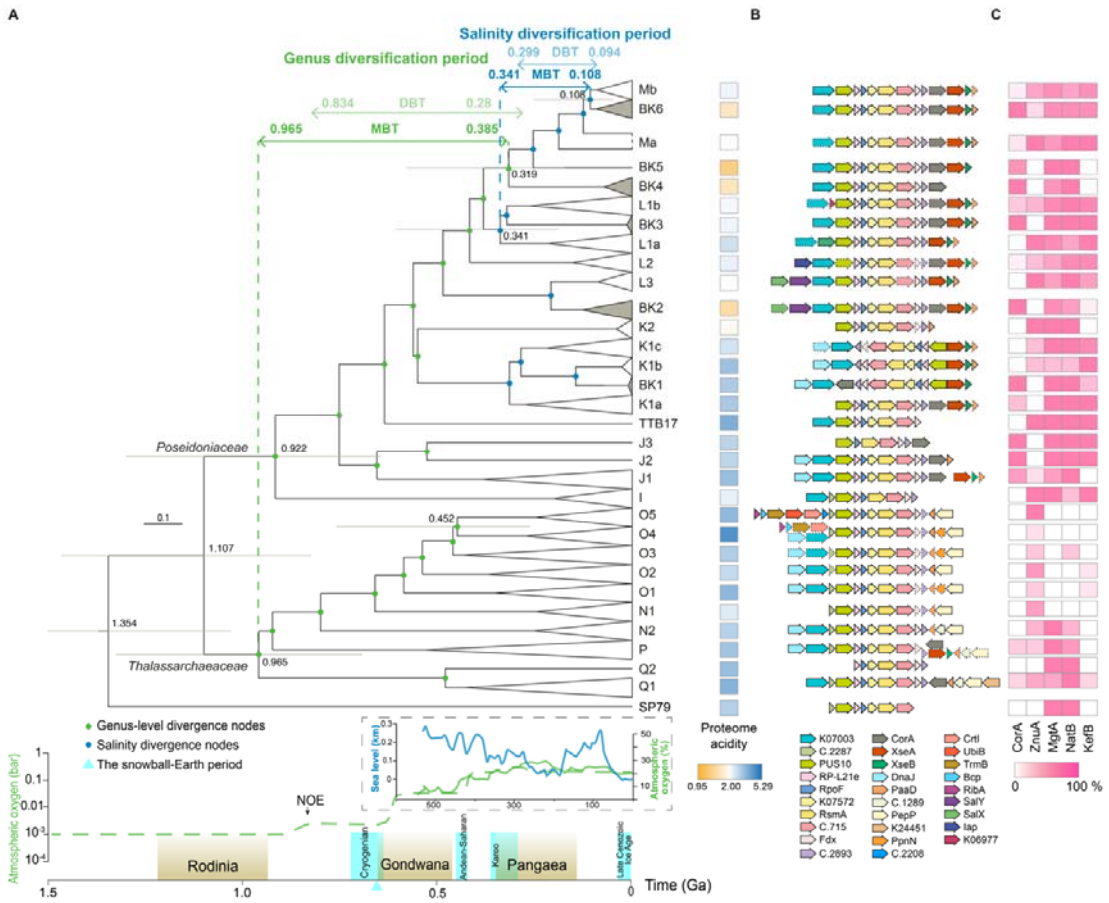

Fig. 2 | Proteome acidification, key gene changes and geological background in Poseidoniales evolution. **(A)** Geological timing of Poseidoniales evolution and major geological events. The tree is part of the tree in **Fig. S9A**. Bars on nodes show 95% confidence interval. Clades filled in grey are brackish subgroups. Proteome acidity levels show the median values of MAGs in each clade. Glaciation events shown here are Huronian (2.29–2.25 Ga), Sturtian (717–659 Ma), and Marinoan (645–635 Ma). MBT = methanogen basal tree, DBT= DPPAN basal tree. **(B)** Arrangement of the stress-response gene cluster in Poseidoniales genomes. Arrows with dashed edges suggest that the genes are present in part of the MAGs in each clade. Gene names are explained in **Fig. S5**. **(C)** The presence of representative ion transport genes. Color density shows the percentage of MAGs having the gene in each clade.

Integration of neighboring gene analysis and phylogenetic analysis of *corA* and *znuA* suggests their distinct evolutionary trajectories. The tree of *corA* was highly congruent to the phylogenomic tree of Poseidoniales and adjacent Marine Group III (MGIII) archaea suggesting that *corA* was generally passed vertically when

Poseidoniales diversified (**Fig. S6**). The absence of *corA* in some Poseidoniaceae and the majority of Thalassarchaeaceae was likely a result of sporadic loss (**Fig. S5, SI**). In Poseidoniales, *corA* was exclusively found at the tail of a highly conservative gene cluster consisting of over ten syntropic genes (**Fig. 2B, Fig. S5)**. This gene cluster contained core gene sets involved in DNA repair, transcription, translational regulation, and post-translational modification by modulating macromolecules such as DNAs, RNAs and proteins (**SI**). Notably, in the three genomes of basal Thalassarchaeaceae, *corA* was in opposite coding direction to the rest of the gene cluster while in Poseidoniaceae it was always in the same coding direction. Such an arrangement suggests an inversion event in the ancestor of either Poseidoniaceae or Thalassarchaeaceae (**SI**). As adjacent and syntropic genes often form operons and transcribe simultaneously [31], genes in this cluster are possibly regulated in concert with each other in stress response. In contrast, the evolution of *znuA* in Poseidoniales is often mediated by lateral gene transfer (LGT) (**Fig. S6**) including 39 gain- and 35 loss events of *znuA*, respectively, as suggested by amalgamated likelihood estimation (ALE) (**Fig. S5**). Moreover, the *znuABC* gene set is not associated with any specific genes or gene clusters (**Fig. S5**).

Both magnesium and zinc are essential elements in cellular functions. While they may be involved in various ways in cells' response to environmental stresses such as drastic salinity fluctuation at estuaries, their common and fundamental role is likely to stabilize or modulate structures of macromolecules such as DNAs, RNAs, and proteins (**SI**). At least 133 of the 159 (83.6%) Poseidoniales MAGs in **Fig. S5** encode either *corA*, *znuABC*, or both, suggesting that the import of divalent cations may be crucial to the survival of Poseidoniales. On the other hand, the mutually exclusive distribution of *corA* and *znuABC* in Poseidoniales suggests these two ion transporters may be functionally redundant (**Fig. S5, SI**). However, genetic conservation and genomic location of *corA* suggest that the regulatory coupling of magnesium import with other essential stress-response functions is the strongest determinant of brackish Poseidoniaceae. In marine Poseidoniaceae whose *corA* is lost from the stress-response gene cluster, *znuABC* is then obtained as compensation for divalent cation transporter.

Gene gain and loss during habitat shift

The sporadic loss of the habitat-determinant gene *corA* in Poseidoniaceae genera provides a unique opportunity to reconstruct the evolutionary history of Poseidoniaceae transitions between brackish and marine habitats at

227 species to strain level. There were at least ten salinity-based divergent events in the evolutionary history of
228 Poseidoniaceae (**Fig. 2A**). As an example, we tracked the gene gain and loss events in association with the
229 salinity and proteome acidification diversification in subgroups BK4, BK5 and M (MAG completeness
230 minimum = 55.96%, completeness median = 80.28%, contamination < 5%) based on the ALE results (**Table**
231 **S4**). The *corA* gene was replaced by a new copy in the common ancestor of BK4 and then vertically transferred
232 in this subgroup. In BK5, M and BK6, *corA* is vertically transferred followed by sporadic losses in marine
233 species of the M subgroup. Remarkably, immediately before or after the loss of *corA*, z*nuABC* was gained and
234 maintained for at least one copy (**Fig. 3**). One additional gain event of *corA* and subsequent loss of *znuABC* is
235 found in a branch of M-b. This observation again supports the hypothesis that these two ion transporters may
236 partially be functionally redundant.
237
238 Accompanying and especially following the change of *corA*, various and massive gains and losses of habitat-
239 specification genes possibly mediated by LGT were observed in Poseidoniaceae genomes during their gradual
240 diversification to marine or brackish environments. For example, as the less acidified proteome of brackish
241 Poseidoniaceae comprises more basic amino acids than that of their marine counterparts (**Fig. 1B**), the
242 concentration of free basic amino acids in the cytoplasmic pool needs to be lowered by either increasing export
243 or stopping biosynthesis to maintain charge balance. Indeed, the acquisition of lysine/arginine efflux and the
244 loss of lysine biosynthesis happened in the ancestors of BK4 and BK6, respectively (**Fig. 3**). Moreover,
245 microorganisms living in brackish and marine environments face great physiochemical differences in nutrient
246 availability, substrate composition and stress types. Accordingly, we found large-scale loss of genes involved in
247 phosphorus and low-abundant metal uptake, peptide degradation (peptidases/proteases), and organic matter
248 utilization in BK4 and BK6 lineages, reflecting a relative eutrophic brackish environment where extra substrate
249 transport machinery is unnecessary. At the same time, some peptidases and biosynthetic enzymes required for
250 brackish water adaptation were obtained in brackish taxa. For example, the acquisition of fructose/tagatose
251 bisphosphate aldolase by genome ShenzhenBay-2018-35 and 2-keto-4-pentenoate hydratase in the catechol
252 meta-cleavage pathway by genome PearlRE-2016-260 possibly reflect the adaptation of brackish
253 Poseidoniaceae to consuming terrestrial substrates (**Table S4**).
254
255 In addition, frequent loss and gain of different types of ribosomal proteins were found. Variations in the content
256 and number of ribosomal proteins could contribute significantly to differences in maintaining large RNA

structure and thus ribosomal performance [32]. The turnover of ribosomal proteins suggests a dynamic population of ribosomes with heterogeneous protein composition and potentially diverse functions in response to habitat shift [33].

All these results support a model in salinity transition within a microbial genus that the transition is initiated with the sudden change of a key gene involved in osmotic stress adaptation and followed by gradual tuning in metabolism and proteome acidification. This model is in line with the 'salinity divide' paradigm that salinity is the primary factor in spatial isolation of aquatic microbes [34].

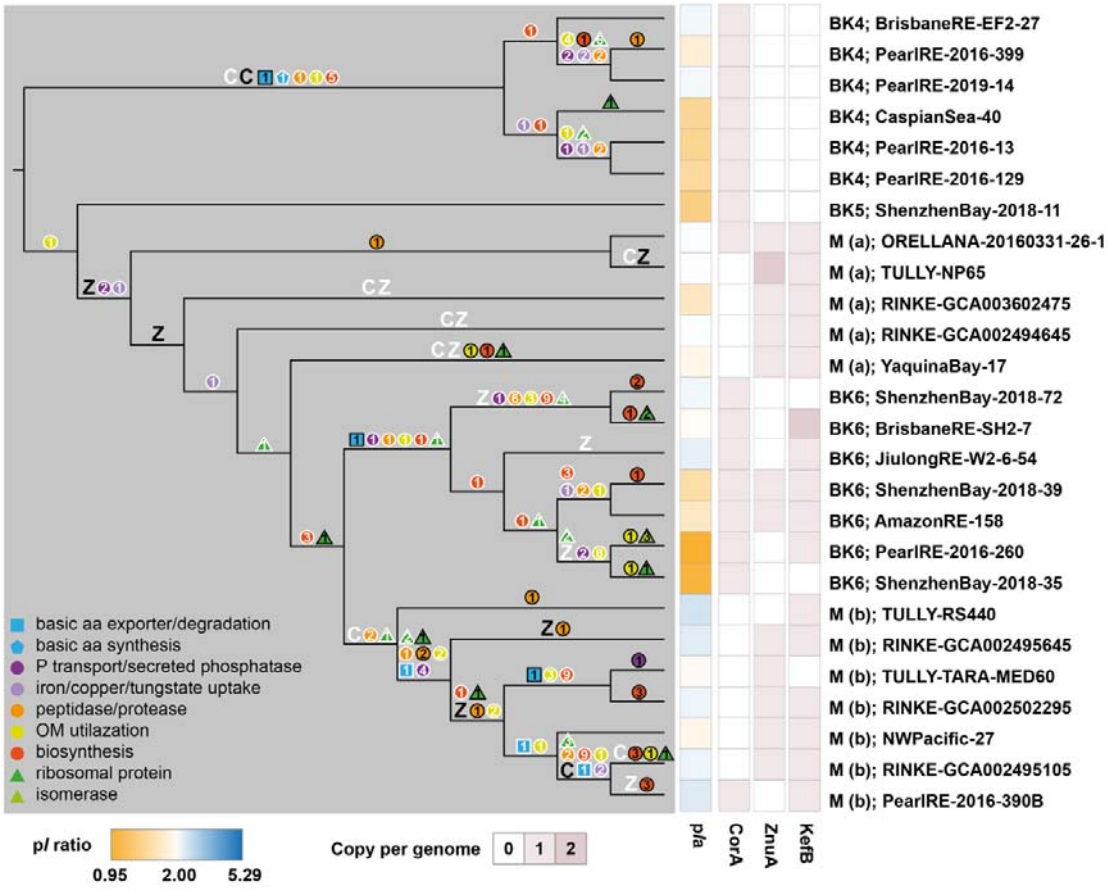

Fig. 3 | Essential gene gain and loss events during the diversification of the BK4-BK5-M-BK6 monophyletic clade of Poseidoniaceae. The cladogram of the BK4-BK5-M-BK6 clade of **Fig. S7A** is shown. Gene gain/loss events between adjacent internal nodes or between adjacent internal nodes and terminal taxa are illustrated on relevant branches. Capital C and Z mean the *corA* gene and the *znuABC* gene set, respectively.

Shapes with black edges or letters in black show gain events, while that with white edges or letters in white show loss events. Numbers inside the shapes indicate the number of genes. p/a, proteome acidity.

Dating the salinity divergence in deep time

Finally, we establish a geological time scheme in the evolution of the brackish and the marine Poseidoniales to study the possible correlation between salinity adaptation and geo-environment changes. Our recent approach has found a strong correlation between the diversification of Nitrososphaeria and geological events such as the Great Oxidation Event [35]. Building on this analytic framework, we calculated the evolutionary timepoint of Poseidoniales diversification in the phylogenetic tree of Archaea by conducting a molecular clock analysis. The archaeal root and three oxidation constraints were used for time calibration (**SI**). We then projected global geological events to the time scale of Poseidoniales diversification (~1.5 Gyr to now) and found that the divergence of Poseidoniaceae and Thalassarchaeaceae happened at about 0.969 Gyr (**Fig. 2A**).

Notably, two identifiable periods existed in the evolutionary history of Poseidoniales, which was supported by applying different rooting strategies and modeling methods (**SI**). The first period was from 0.834 to 0.28 Gyr, when the formation of all the Poseidoniales genus-level subgroups happened. This period coincides with the increase of atmospheric oxygen from that of the boring billion years to the current level [36]. As Poseidoniales are aerobic heterotrophs, an increase in ocean oxygen level may enhance their oxidative capacity to access more complex organic substrates and/or facilitate their exploration to previously anoxic marine habitats that are rich in various organic matter. Both processes promote niche diversification. The second period was from 0.284 to 0.094 Gyr when the divergence of the brackish and the marine subgroups of Poseidoniaceae happened. This period was aligned to the Pangea tectonic period and the large-scale change of sea level [37]. A similar observation was made for amphipods diversification due to habitat shift during the closing of Tethys [38] and the breakup of Gondwana [39]. Our finding provides strong evidence that, like animals, landform and sea level changes, which create drastic changes in the salinity of coastal aquatic habitats, can be a strong driving force for the cross-salinity evolution of Poseidoniales.

Conclusion

In this study, we track the evolution trajectory of Poseidoniales subgroups transited between marine and brackish habitats by focusing on identifying the primary selective force and the key genetic change that initiated the transitions. We discover that the speciation was triggered by the insertion, inversion, possibly regulatory coupling, and subsequent loss of a key gene *corA* in a highly conservative stress-response gene cluster. The regulatory coupling of *corA* with this gene cluster potentially contributes to osmotic adaptation of Poseidoniales in brackish waters. The sudden losses and gains of *corA* were then followed by metabolic acclimation and diversification mediated by LGT. The establishment of a geological time frame in genome evolution of Poseidoniales demonstrates that this salinity-based speciation is possibly selected by strong coastal hydrodynamics in the background of global tectonic activities and sea level changes.

Based on these discoveries of Poseidoniales evolution, we propose a model with a hierarchical structure of selection in the early process of microbial speciation (Fig. 4). In this model, the sudden change of primary and qualitative niche trait (e.g. salinity adaptation in this study) precedes the gradual changes in secondary and accumulative traits (e.g. proteome acidity change and multidimensional metabolism adjustment). We also highlight the essential roles of strong selective force (e.g. rapid environmental salinity change) and gene co-regulation (e.g. in a stress-response gene cluster) to fix the primary niche trait and facilitate further speciation. This model is potentially applicable to elucidate long-stand puzzles in other ancient microbial speciation events between ecologically distinct habitats such as the emergence of pathogens [8], endosymbionts [40], and extremophiles [41].

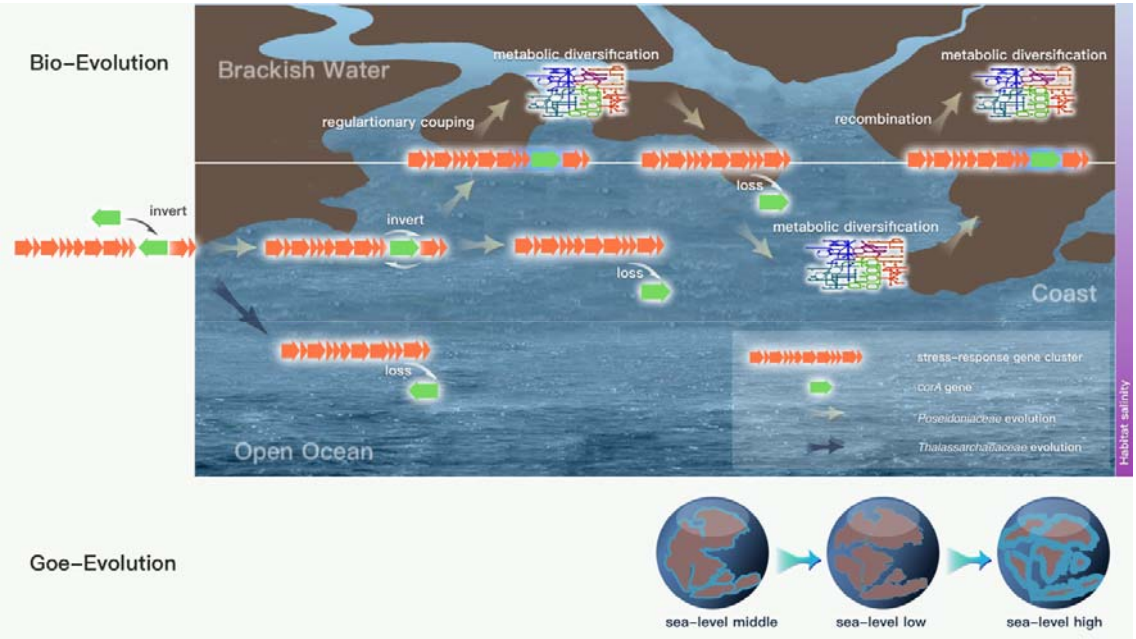

Fig. 4 | A schematic picture shows the process of genetic changes during the speciation of Poseidoniales between salinity distinct habitats and the potential selective force. The three globes roughly show the movement of land plates during the Pangaea tectonic period. The light blue edges of the continents conceptually show the merged coastal land as a result of high sea-level.

Methods

Sampling metagenomes. Microplanktons were collected at the Pearl River estuary in 2011, 2012, 2016 [42], 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: OEP001542), 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□ in lab till further processes. 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 [43], the Baltic Sea [29,44–46], the Columbia River estuary [47], the Amazon River estuary [48], the Yaquina Bay estuary [49], the Helgoland coast[27], the Port Hacking coast[22], and the Tara Oceans Project [50] 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) [51].

Generation of the global non-redundant Poseidoniales genome dataset. To obtain potential brackish Poseidoniales genomes, we used IDBA-UD (v. 1.1.3) [52] 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 [53], Maxbin2 [54], and CONCOCT [55] methods. Bins with completeness >50% and contamination <10% as evaluated by CheckM (v. 1.0.5) [56] were kept and those classified as Poseidoniales by GTDB-tk (v. 1.3.0, release 95) [57] were used for downstream analysis. Previously published marine Poseidoniales genomes generated by Rinke et al. [22], Tully [23], and Orellana et al. [27] were downloaded from their online deposits. The combination of downloaded genomes with those generated in this study results in 835 Poseidoniales metagenome-assembled genomes (MAGs) (**Table S2**). Potential contaminant contigs in each MAG were further removed by manual check aided by acdc (v. 1.2.1) [58]. A non-redundant MAG dataset was generated by using dRep (v. 2.6.2) [59] and setting a cutoff of 99% average nucleotide identity. This dataset contains 455 Poseidoniales MAGs. Quality check and taxonomic classification of these MAGs were conducted by using CheckM and GTDB-tk, respectively. Genes and proteins of the MAGs were predicted by using Prodigal (v. 2.6.3) [60].

Phylogenomics of the non-redundant Poseidoniales MAGs. We used hmmsearch (v. 3.1b2; -E 1E-5) [61] to search for the 122 archaeal single-copy marker proteins [56] in the 455 non-redundant Poseidoniales MAGs based on hidden Markov models (HMMs) in Pfam [62] and TIGRfam [63] databases. MGIII euryarchaeal and other archaeal genomes were used as the outgroup (**Table S3**). Marker proteins present in ≥60% taxa were retained and aligned, respectively, by using MUSCLE (v. 3.8.1551; --maxiters 16) [64]. The alignment matrixes were denoised by using trimAl (v1.2rev59; -automated1) [65] and then concatenated. Missing data were filled with gaps. A maximum-likelihood tree was reconstructed by using FastTree (v. 2.1.10; -gamma -lg) [66] and visualized in the Interactive Tree of Life (iTOL, v.5.1.1) [67].

MAG abundance calculation. To profile the distribution of Poseidoniales in global marine surface water, the non-redundant 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, 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) [68], 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 before mapping using Bedtools (v. 2.27.1). Read mapping was conducted by using Bowtie2 (v. 2.3.5) [69] and followed by sorting and format convert to BAM files by using SAMtools (v. 1.9) [70]. The BAM files were filtered by using BamM (v. 1.7.3) (https://github.com/minillinim/BamM;) with thresholds of 99% identity and 75% coverage. Finally, bbmap (http://jgi.doe.gov/data-and-tools/bb-tools/) was used to calculate read counts for each contig and the RPKM (Reads Per Kbp of each genome per Mbp of each metagenomic sample) value was calculated for each MAG in each sample, respectively.

Proteome acidity estimation. The isoelectric points (p*I*) of proteins of MAGs were calculated by using Pepstats of the EMBOSS package [71]. p*I* frequency distribution of a proteome was calculated as previously described [12]. Proteome acidity in this study is defined as the ratio of the frequency of the acidic peak (p*I* 4.5) to the frequency of the semi-acidic peak (p*I* 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.

Functional annotation and comparison of MAGs. Protein sequences of MAGs were annotated based on the KEGG database by using kofamscan [72], and the COG [73], arCOG[74], Pfam [75] and Tigrfam databases [63] by using BLASTp [76] (E-value < 10^{-3}, bit score > 50, similarity > 50%, and coverage > 70%), respectively. Genes

specifically enriched in brackish Poseidoniales were defined as those present in > 85% brackish MAGs but < 5% in marine MAGs. Genes specifically enriched in marine Poseidoniales were defined *vice versa*.

Tree of Poseidoniales and other archaea. To build the tree for the dataset containing 188 taxa (**Fig. S7**), 39 of the 41 marker proteins described by Adam et al., 2017 [77] were used. The other two proteins were excluded because they are absent in most of the Poseidoniales MAGs in this dataset. Detection of the marker proteins in each MAG was based on functional annotation. The marker proteins of each MAG were identified according to the genome functional annotations. A multisequence alignment of concatenated marker proteins was constructed by using MUSCLE and automatically trimmed by using trimAl. Removal of compositional heterogenous sites was conducted by applying a χ^2-score-based approach [78]. Maximum likelihood trees were reconstructed with the LG+C60+F model implemented in IQ-TREE (v. 2.0.3) [79] and then visualized in iTOL. Maximum-likelihood trees of the dataset containing 230 taxa (**Fig. S8**) were reconstructed in the same manner.

Amalgamated likelihood estimation (ALE) analysis. Functional genes in the 231-taxa dataset were aligned by using MAFFT L-INS-I [80] and denoised by using trimAl (automated1). The ML tree was constructed by using IQ-TREE with the parameters "-seqtype AA -m LG+PMSF+G -B 1000 --bnni". The ALEml_undated algorithm of the ALE package [81] 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.

Gene gain/loss event analysis for the BK4-BK5-M-BK6 monophyletic clade of Poseidoniaceae. The event number of a gene (KO or arCOG entry) in a terminal taxon (MAGs) is 1 if the gene is present and is 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. [82] Gene gain/loss events between adjacent internal nodes or between adjacent internal nodes and terminal taxa are defined as the following: 1) A loss event is defined if the event number of the older node (an internal node) is greater than 0.8 and is eight times greater than that of the younger node (an internal node or a terminal taxon); and 2) A gain event is defined if the event number of the younger node (an internal node or a terminal taxon) is greater than 0.8 and is eight times greater than that of the older node (an internal node).

Molecular clock analysis. 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% confidence interval [83]. The root of Archaea (4.38-3.46 Ga) [84–86] and three constraints (i.e. the roots of Thermoproteales, Sulfolobales and Thermoplasma) related to the Great Oxygenation Event (2.32 Ga) [87,88] were used for calibration as introduced in our previous study [35].

Reference of geological events. The estimation of geological atmospheric oxygen level in Fig. 2A is based on Catling and Zahnle 2020 [36] and Campbell and Allen 2008 [89]. Glaciation events are based on Tang and Chen 2013 [90]. Tectonic active period is based on Nance et al. 2014 [91]. Sea level changes are based on Marcilly et al. 2022 [37].

REFERENCES AND NOTES

1. Shapiro, B. J. & Polz, M. F. Microbial Speciation. *Cold Spring Harb. Perspect. Biol.* **7**, a018143 (2015).
2. Baquero, F., Coque, T. M., Galán, J. C. & Martinez, J. L. The Origin of Niches and Species in the Bacterial World. *Front. Microbiol.* **12**, 657986 (2021).
3. Shapiro, B. J., Leducq, J.-B. & Mallet, J. What Is Speciation? *PLOS Genet.* **12**, e1005860 (2016).
4. Shapiro, B. J. *et al.* Population Genomics of Early Events in the Ecological Differentiation of Bacteria. *Science* **336**, 48–51 (2012).
5. Schemske, D. W. Adaptation and *The Origin of Species*. *Am. Nat.* **176**, S4–S25 (2010).
6. Bobay, L.-M. The Prokaryotic Species Concept and Challenges. in *The Pangenome: Diversity, Dynamics and Evolution of Genomes* (Springer, 2020).
7. Luo, C. *et al.* Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci U A* **108**, 7200–5 (2011).
8. Boritsch, E. C. *et al.* pks5-recombination-mediated surface remodelling in Mycobacterium tuberculosis emergence. *Nat. Microbiol.* **1**, 15019 (2016).
9. Chiner-Oms, Á. *et al.* Genomic determinants of speciation and spread of the *Mycobacterium tuberculosis* complex. *Sci. Adv.* **5**, eaaw3307 (2019).

10. Walsh, D. A., Lafontaine, J. & Grossart, H.-P.-. P. On the eco-evolutionary relationships of fresh and salt water bacteria and the role of gene transfer in their adaptation. in *Lateral gene transfer in evolution* 55–77 (Springer, 2013).
11. Zhang, H. *et al*. Repeated evolutionary transitions of flavobacteria from marine to non-marine habitats. *Env. Microbiol* **21**, 648–666 (2019).
12. Cabello-Yeves, P. J. & Rodriguez-Valera, F. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome* **7**, 117 (2019).
13. Logares, R. *et al*. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* **17**, 414–22 (2009).
14. Paver, S. F., Muratore, D., Newton, R. J. & Coleman, M. L. Reevaluating the salty divide: Phylogenetic specificity of transitions between marine and freshwater systems. *mSystems* **3**, e00232-18 (2018).
15. Eiler, A. *et al*. Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *ISME J* **10**, 1902–14 (2016).
16. Simon, M. *et al*. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J* **11**, 1483–1499 (2017).
17. Penn, K. & Jensen, P. R. Comparative genomics reveals evidence of marine adaptation in Salinispora species. *BMC Genomics* **13**, 86 (2012).
18. Martijn, J. *et al*. Hikarchaeia demonstrate an intermediate stage in the methanogen-to-halophile transition. *Nat Commun* **11**, 5490 (2020).
19. Cabello-Yeves, P. J. *et al*. Novel Synechococcus genomes reconstructed from freshwater reservoirs. *Front Microbiol* **8**, 1151 (2017).
20. Ren, M. & Wang, J. Phylogenetic divergence and adaptation of Nitrososphaeria across lake depths and freshwater ecosystems. *ISME J* (2022) doi:10.1038/s41396-022-01199-7.
21. Henson, M. W., Lanclos, V. C., Faircloth, B. C. & Thrash, J. C. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J* **12**, 1846–1860 (2018).
22. Rinke, C. *et al*. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME J* **13**, 663–675 (2019).
23. Tully, B. J. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat Commun* **10**, 271 (2019).

24. Li, M. *et al*. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* **6**, 8933 (2015).
25. Zhang, C. L., Xie, W., Martin-Cuadrado, A.-B. B. & Rodriguez-Valera, F. Marine Group II Archaea, potentially important players in the global ocean carbon cycle. *Front Microbiol* **6**, 1108 (2015).
26. Chen, S. *et al*. Interactions Between Marine Group II Archaea and Phytoplankton Revealed by Population Correlations in the Northern Coast of South China Sea. *Front Microbiol* **12**, 785532 (2021).
27. Orellana, L. H. *et al*. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. *ISME J* **13**, 3024–3036 (2019).
28. Xie, W. *et al*. Localized high abundance of Marine Group II archaea in the subtropical Pearl River Estuary: implications for their niche adaptation. *Env. Microbiol* **20**, 734–754 (2017).
29. Hugerth, L. W. *et al*. Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol* **16**, 279 (2015).
30. Fagerbakke, K. M., Norland, S. & Heldal, M. The inorganic ion content of native aquatic bacteria. *Can J Microbiol* **45**, 304–11 (1999).
31. Gao, L. *et al*. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
32. Sharma, M. R. *et al*. Structure of the Mammalian Mitochondrial Ribosome Reveals an Expanded Functional Role for Its Component Proteins. *Cell* **115**, 97–108 (2003).
33. Ferretti, M. B. & Karbstein, K. Does functional specialization of ribosomes really exist? *RNA* **25**, 521–538 (2019).
34. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc Natl Acad Sci U A* **104**, 11436–40 (2007).
35. Yang, Y. *et al*. The evolution pathway of ammonia-oxidizing archaea shaped by major geological events. *Mol Biol Evol* **38**, 3637–3648 (2021).
36. Catling, D. C. & Zahnle, K. J. The Archean atmosphere. *Sci. Adv.* **6**, eaax1420 (2020).
37. Marcilly, C. M., Torsvik, T. H. & Conrad, C. P. Global Phanerozoic sea levels from paleogeographic flooding maps. *Gondwana Res.* **110**, 128–142 (2022).
38. Hou, Z., Sket, B., Fiser, C. & Li, S. Eocene habitat shift from saline to freshwater promoted Tethyan amphipod diversification. *Proc Natl Acad Sci U A* **108**, 14533–8 (2011).

39. Cannizzaro, A. G. & Berg, D. J. Gone with Gondwana: Amphipod diversification in freshwaters followed the breakup of the supercontinent. *Mol Phylogenet Evol* **171**, 107464 (2022).
40. Suzuki, T. A. & Ley, R. E. The role of the microbiota in human genetic adaptation. *Science* **370**, eaaz6827 (2020).
41. Shu, W.-S. S. & Huang, L.-N. N. Microbial diversity in extreme environments. *Nat Rev Microbiol* (2021) doi:10.1038/s41579-021-00648-y.
42. Xu, B. *et al*. A holistic genome dataset of bacteria, archaea and viruses of the Pearl River estuary. *Sci Data* **9**, 49 (2022).
43. Mehrshad, M., Amoozegar, M. A., Ghai, R., Shahzadeh Fazeli, S. A. & Rodriguez-Valera, F. Genome reconstruction from metagenomic data sets reveals novel microbes in the brackish waters of the Caspian Sea. *Appl Env. Microbiol* **82**, 1599–612 (2016).
44. Larsson, J. *et al*. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* **8**, 1892–903 (2014).
45. Alneberg, J. *et al*. BARM and BalticMicrobeDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. *Sci Data* **5**, 180146 (2018).
46. Alneberg, J. *et al*. Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol* **3**, 119 (2020).
47. Fortunato, C. S. & Crump, B. C. Microbial gene abundance and expression patterns across a river to ocean salinity gradient. *PLoS One* **10**, e0140578 (2015).
48. Satinsky, B. M. *et al*. The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome* **2**, 17 (2014).
49. Kieft, B. *et al*. Microbial community structure-function relationships in Yaquina Bay estuary reveal spatially distinct carbon and nitrogen cycling capacities. *Front Microbiol* **9**, 1282 (2018).
50. Pesant, S. *et al*. Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* **2**, 150023 (2015).
51. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
52. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–8 (2012).

53. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
54. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
55. Alneberg, J. *et al*. Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**, 1144–6 (2014).
56. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–55 (2015).
57. Chaumeil, P.-A. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* (2019) doi:10.1093/bioinformatics/btz848.
58. Lux, M. *et al.* acdc - Automated Contamination Detection and Confidence estimation for single-cell genome data. *BMC Bioinformatics* **17**, 543 (2016).
59. 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* **11**, 2864–2868 (2017).
60. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
61. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29-37 (2011).
62. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
63. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31**, 371–373 (2003).
64. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–7 (2004).
65. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–3 (2009).
66. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

67. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
68. Bengtsson-Palme, J. *et al.* metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* (2015) doi:10.1111/1755-0998.12399.
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–9 (2012).
70. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
71. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–7 (2000).
72. Aramaki, T. *et al.* KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* (2019) doi:10.1093/bioinformatics/btz859.
73. Galperin, M. Y. *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* **49**, D274–D281 (2021).
74. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life Basel* **5**, 818–40 (2015).
75. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
76. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
77. Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**, 2407–2425 (2017).
78. Viklund, J., Ettema, T. J. G. & Andersson, S. G. E. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29**, 599–615 (2012).
79. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
80. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–66 (2002).
81. Szöllõsi, G. J., Rosikiewicz, W., Boussau, B., Tannier, E. & Daubin, V. Efficient exploration of the space of reconciled gene trees. *Syst Biol* **62**, 901–12 (2013).
82. Sheridan, P. O. *et al.* Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. *Nat Commun* **11**, 5494 (2020).

608 83. Tamura, K., Tao, Q. & Kumar, S. Theoretical foundation of the RelTime method for estimating divergence
609 times from variable evolutionary rates. *Mol Biol Evol* **35**, 1770–1782 (2018).
610 84. Ueno, Y. *et al.* Geological sulfur isotopes indicate elevated OCS in the Archean atmosphere, solving faint
611 young sun paradox. *Proc Natl Acad Sci U A* **106**, 14784–9 (2009).
612 85. Valley, J. W. *et al.* Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography. *Nat.*
613 *Geosci.* **7**, 219–223 (2014).
614 86. Wolfe, J. M. & Fournier, G. P. Reply to 'Molecular clocks provide little information to date methanogenic
615 Archaea'. *Nat. Ecol. Evol.* **2**, 1678–1678 (2018).
616 87. Blank, C. E. Phylogenomic dating–a method of constraining the age of microbial taxa that lack a
617 conventional fossil record. *Astrobiology* **9**, 173–91 (2009).
618 88. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
619 89. Campbell, I. H. & Allen, C. M. Formation of supercontinents linked to increases in atmospheric oxygen.
620 *Nat. Geosci.* **1**, 554–558 (2008).
621 90. Tang, H. & Chen, Y. Global glaciations and atmospheric change at ca. 2.3 Ga. *Geosci. Front.* **4**, 583–596
622 (2013).
623 91. Nance, R. D., Murphy, J. B. & Santosh, M. The supercontinent cycle: A retrospective essay. *Gondwana*
624 *Res.* **25**, 4–29 (2014).
625

626 **Acknowledgements**

627 This study was supported by National Natural Science Foundation of China (Nos. 91851210, 91951120,
628 42141003), the State Key R&D Project of China Grant (No. 2018YFA0605800), the Guangdong Basic and
629 Applied Basic Research Foundation (No. 2021B1515120080), the Shenzhen Key Laboratory of Marine Archaea
630 Geo-Omics, Southern University of Science and Technology (ZDSYS201802081843490), the Southern Marine
631 Science and Engineering Guangdong Laboratory (Guangzhou) (No. K19313901), and the Project of Educational
632 Commission of Guangdong Province of China (No. 2020KTSCX123). Computation in this study was supported
633 by the Centre for Computational Science and Engineering at the Southern University of Science and
634 Technology.
635

636 **Author contributions**

Lu Fan and Chuanlun Zhang conceived this study. Bu Xu, Songze Chen, Fuyan Li, Wei Xie, Apoorva Prabhu, Dayu Zou, Ru Wan, Hongliang Li, Haodong Liu, Yuhang Liu, Shuji Gao, Jianfang Chen, Christian Rinke, and Meng Li collected the samples and extracted DNA. Lu Fan, Bu Xu, Songze Chen, Yang Liu, Fuyan Li, Apoorva Prabhu, Dayu Zou, Ru Wan, and Hongliang Li analyzed the metagenome data, produced the genomes, and conducted all other analyses. Lu Fan, Bu Xu, and Chuanlun Zhang interpreted the results and drafted the manuscript. All authors contributed to the final version of the manuscript. Lu Fan, Bu Xu, Songze Chen, and Yang Liu contributed equally to this work.

Competing interest declaration

The authors declare no competing interests.

Additional Information

Supplementary Information is available for this paper.