# Comparative genomics of *Exiguobacterium* reveals what makes a cosmopolitan bacterium

De-Chao Zhang[a,b,c], Zhaolu Zhu[d,e,#], Xudong Li[d,e], Ziyu Guan[d], Jinshui Zheng[d,e,#]

[a]Institute of Oceanology, Chinese Academy of Sciences, 266071, Qingdao, China

[b]University of Chinese Academy of Sciences, 100049, Beijing, China

[c]Center for Ocean Mega-Science, Chinese Academy of Sciences, 7 Nanhai Road, 266071, Qingdao, China

[d]State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China

[e]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

[#]Corresponding author

Address correspondence to Jinshui Zheng (jszheng@mail.hzau.edu.cn); Zhaolu Zhu (zhaoluzhu@gmail.com)

**Running title:** Adaptation and evolution analysis of *Exiguobacterium*

**E-mail address of all authors:**

Dechao Zhang: zhangdechao@qdio.ac.cn

Zhaolu Zhu: zhaoluzhu@gmail.com

Xudong Li: xudongli_bioinfo@163.com

Ziyu Guan: ziyu_guan@163.com

Jinshui Zheng: jszheng@mail.hzau.edu.cn

## ABSTRACT

Although the adaptation strategies of bacteria to specific environmental conditions are widely reported, fewer studies have addressed how microbe with cosmopolitan distribution adapted to diverse habitats. *Exiguobacterium* is a versatile genus whose members have been commonly found in great variety of habitats. To understand the mechanism behind the universally of *Exiguobacterium*, we isolated 103 strains from diverse environments, and performed large-scale metabolic and adaptive ability tests. We found that the capacities of survival in a wide range of temperature, salinity and pH are common for most *Exiguobacterium* members. According to the core genome based phylogeny and ANI analysis, 26 putative species including 13 putative new ones were identified and two genetic groups were classified as Group I and II. Comparative genomic analysis revealed that *Exiguobacterium* members can not only utilize a variety of complex polysaccharides and proteins that are ubiquitous in both terrestrial and marine environments, but also have a number of chaperonins and transporters which could support them to survive in different extreme environments. In addition, we found that the species from Group I can be found in more diverse environments with larger genome size compared to those of Group II. Twenty-five transporter families involved in transport of organic or inorganic substrates and environments stresses resistance were predicted to be enriched in Group I strains. This study provided the comprehensive insight into general genetic basis of the cosmopolitan distribution of a bacteria genus and deciphered putative determinants behind the ecological difference of different groups belonging to the same genus.

**IMPORTANCE** The wide distribution characteristics make *Exiguobacterium* a valuable model for studying adaptive strategy of bacteria adapted to multiple habitats. In this study, we found that comprehensive capacity of diverse polysaccharides utilization and environmental stress resistance is the important basis for survival, and selective expansion of transporters is an evolution and adaptation strategy for extensive distribution. Our findings are significant for understanding the adaptation and evolution mechanisms of cosmopolitan bacteria and explaining the vital genomic

53 traits that facilitate niches adaptation.

54 **KEYWORDS** *Exiguobacterium*, cosmopolitan distribution, genomics, adaptation
55 strategies, polysaccharide utilization, transporters

56

## INTRODUCTION

Microbial community composition and diversity across landscape is nonrandom (1). Physical and chemical factors in the environment significantly influence the distribution patterns of microbes (2). There is a barrier between marine and non-marine habitats due to their strong physiochemical difference such as salinity, temperature, pH, dissolved oxygen and water chemistry (3). As a result, most marine microbes belong to different phylogenetic groups from their freshwater and terrestrial relatives, and there are rare transitions between these two niches (4, 5). It was frequently reported that these different bacteria members usually utilize different strategy for niche adaptations. Comparative genomics of ocean microbes revealed that many marine bacteria were identified with streamlined genome to reduce metabolic costs of maintaining nonessential genetic material to adapt to the nutrient-poor ocean environments (6, 7). And it is obviously different from free-living terrestrial bacteria, which usually have normal genome size with frequent horizontal genetic transfer events to facilitate the capacity of using diverse collection of nutrients and ability of resistance to complicated adverse environments (8, 9). The transition events of marine to non-marine or in reverse for bacteria require complex genome evolution (10, 11). Events of gaining and losing gene which involved in the abilities for transport, metabolism and assimilation different types of organic or inorganic nutrients play crucial roles during this progress (10, 12). However, this knowledge is mainly derived from the comparative genomics of a few bacteria with high abundance in either marine or non-marine microbiota (13, 14). The strategies of evolution and adaptation for microbes which have wide distribution in both marine and non-marine environments are not well studied yet.

Bacteria of the genus *Exiguobacterium* are Gram-positive facultative anaerobes which have been frequently isolated from various habitats including seawater, marine sediment, marine algae (15-17), soil (17), freshwater (18), plant rhizosphere (19), and some extreme environments such as salt lake (20), glacier and hot spring (21). Genomic analysis of these bacteria has provided some vital insights into their

86 psychrophilic and thermophilic adaptations and multiple toxic compound resistances
87 (22-24). However, comprehensive knowledge about these bacteria evolutionary
88 adaptation to marine and non-marine habitats remains undiscovered.

89 In this study, *Exiguobacterium* was used as a model to study how microbes with
90 cosmopolitan distribution adapted to both marine and terrestrial environments. We
91 leveraged different strategies to decipher the ecology and evolution of
92 *Exiguobacterium* spp.. We firstly mined the public database for 16S rRNA gene to
93 reveal the diversity and distribution of the genus. Then we isolated multiple strains
94 from marine and other habitats and tested their adaptive and metabolic features.
95 Furthermore, we sequenced 103 strains and performed large-scale phylogenomic and
96 comparative genomic analyses on a total 145 genomes representing strains isolated
97 from marine and non-marine habitats worldwide. Special attention was drawn on
98 genomic and metabolic characters responded to diverse environments.

99 **RESULTS AND DISCUSSION**

100 **Cosmopolitan distribution of *Exiguobacterium* spp. and their survival abilities in**
101 **a wide range of conditions.** To explore the diversity and distribution of members
102 belonging to *Exiguobacterium*, we focused on the bacterial strain with 16S rRNA
103 gene sequence having an identity above 95% to that of the reported species type strain
104 from this genus. A total of 2,582 *Exiguobacterium* 16S rRNA gene sequences with
105 unambiguous information of isolation source were collected from Genbank (Table S1).
106 We found that members of this genus were frequently isolated from territorial
107 environments (86.6%), including plant or rhizosphere (16.6%), animal skin or gut
108 (10.7%), freshwater or freshwater sediment (12.7%), contaminated water or soil (7%),
109 soil (6.8%), extreme environments (hot- or cold-associated or hypersaline
110 environments) (6%), air (4.8%) and other non-marine environments (22%) (Fig. 1A;
111 Table S1). The left 13.4% members were isolated from marine-associated
112 environments including sea water, algae and oceanic sediment. Combining the
113 location information, we found that *Exiguobacterium* can be isolated from all of the
114 continents and oceans. This result was accordance with the current notion that

*Exiguobacterium* was a cosmopolitan bacteria genus including many extremophiles surviving in both marine and non-marine environments worldwide (15).

To study the adaptation and evolution of *Exiguobacterium*, we performed large-scale phenotype tests and comparative genomics analysis. A total of 103 *Exiguobacterium* strains were collected here, including 87 isolates from marine-associated niches (marine sediment, seawater, algae, marine cold spring, hydrothermal vent, seamount, mangrove, marine fish and coral), 11 strains from territorial environments (soil, salt lake, coal mine and pig farm) and 5 type strains (Fig. 1B; Table S2). Among the 42 *Exiguobacterium* members with genome available up to the manuscript submitted, only 4 strains were isolated from marine habitats. We used more isolates from marine than territorial environments in this study.

Previous studies have reported that *Exiguobacterium* spp. could survive in a wide range of habitats including cold, hot, hypersaline and alkaline environments (16), while it is unclear if these features are shared by all the members or strain/species specific.

To assess the survival abilities of *Exiguobacterium* spp. to diverse conditions, we evaluated the growth potential of the collected 103 strains in culture medias with different of pH, temperatures and salinity, respectively. We found that the most *Exiguobacterium* members can survive and grow in a wide range of temperature, salinity and pH values, respectively. (Fig. 1C; Table S3). The pH test revealed that all the strains were alkali-resistant which could survive in the environments with pH value up to 11, and 24% strains showed growth at pH 5 (17%) and even 4 (7%). For salinity test, most of strains showed tolerance to saline, with 66% strains growing in environment with NaCl concentration above 10%. Temperature tests showed that the *Exiguobacterium* spp. have high and low temperature tolerance, with 91% strains growing at 4℃ and 61% surviving from 40 to 50℃. Moreover, no associations were found between the growth abilities and the source environments. These results suggest the extensive adaptability to survive in variable environments as a general feature of this genus.

**Phylogenetic analysis identified two genetic groups in *Exiguobacterium* genus.** We conducted comparative genomics analysis to investigate the strategies of evolution and adaptation to diverse environments. All genomes of the 103 collected *Exiguobacterium* strains were sequenced through Illumina Novaseq 6000 platform and assembled by SPAdes software (25). The 42 *Exiguobacterium* genomes available on GenBank were also included in the comparative analysis (Table S2). All the annotated proteins from 145 genomes were clustered into 8,728 groups, with 1,162 shared by all genomes classified as core gene families. The maximum-likelihood (ML) phylogenetic tree was constructed based on the core genome alignment (Fig. 2). Two genetic groups of *Exiguobacterium* were classified and well-supported by bootstrapping analysis, which is consistent with previous analysis based on 16S rRNA gene sequences (15). Using the threshold of ANI (average nucleotide identity) 95% to define different species (26), a total of 26 species can be classified including 13 putative new ones (N1 to N13) (Fig. 2; Table S4). There were 11 and 15 species belonging to Group I and II, respectively (Table S4).

We next annotated the tree by adding the isolated environment of each strain, and found that strains from different marine and terrestrial niches distributed around the whole tree, even isolates from the same species could found in different environments (Table S5). For example, members of *E. acetylicum* were found in seawater, ocean sediment, soil, rhizosphere, glacier and even animal gut. It is suggested that frequent transitions *Exiguobacterium* spp. among different niches of terrestrial and marine environments. This finding is different from those of some typical marine bacteria which usually contained different lineages adapting to marine and non-marine habits, respectively (27).

**Carbon and nitrogen source utilization for wide adaptation.** To explain the extensive distribution of *Exiguobacterium* strains in different niche types, we focused on the genes involved in different types of nutrition metabolism. Active enzymes involved in carbohydrate metabolism are defined as carbohydrate-active enzymes (CAZymes). A total of 7,864 genes belonging to 5 CAZyme superfamily were

identified from all the genomes, with 61.5%, 18.7%, 15.9%, 2.7% and 1.2% genes belonging to GH (glycoside hydrolase), CE (carbohydrate esterase), CBM (carbohydrate-binding module), PL (polysaccharide lyase) and AA (auxiliary activities), respectively (Fig. 3A; Table S6). Each genome sequence encoded 43 to 68 of these enzymes.

Many classes of enzyme for complex polysaccharide degradation were predicted for *Exiguobacterium* genomes (Fig. 3B; Table S6). The top 3 abundant classes are those associated with the degradation of starch, cellulose and chitin, and most of these CAZymes are potentially secreted (Fig. 3B). Family GH13 represents the main amylolytic enzymes family, the GH13_31 (α-glucosidase) and GH13_13 (pullulanase) are the top 2 frequent subfamilies, with 3.3 and 3 genes per genome, respectively. The most abundant family involved in cellulose degradation is GH1 (β-glucosidases), with each genome containing more than 4 genes. For chitin degradation, family CE4 (deacetylase) and CBM50 (chitin-binding) showed significantly higher abundance.

As the important storage polysaccharide, starch is produced by plants from both terrestrial and marine (28, 29). Cellulose is the most prevalent polysaccharide in nature, which make up the plant and algal cell walls (30, 31). As the second common polysaccharide after cellulose in nature, chitin is also widely distributed in terrestrial and marine ecosystems as a major structural component of crustacean shell, arthropod exoskeleton and the cell wall of diatom (32-34). The abundant CAZymes contained by members of *Exiguobacterium* giving the strong putative capacity for degradation of these polysaccharides ensure these bacteria to extensively gain the carbon source and survive in variable niches of both marine and non-marine environments.

Proteinaceous compounds are abundant forms of organic nitrogen in aquatic and soil (35). Extracellular microbial peptidases play an important role in both marine and terrestrial environments, as they directly link to organic nitrogen degeneration to contribute the global nitrogen cycling (35). In this study, a total of 3,912 putatively secreted peptidases were assigned to 20 families, including 43.7%, 38.1%, 15% and 3.2% belonging to metallo-, serine, cysteine and aspartic peptidase families,

202 respectively (Fig. 3C; Table S7). When normalized to genome size, the average
203 number of secreted peptidase coding genes was 9 genes per Mb, which is higher than
204 the overall level of bacteria (5.84 genes per Mb) (35). Among these peptidases, the
205 metallo peptidase M23 and serine peptidase S08 represent the top 2 ample peptidases
206 (Fig. 3C). M23 peptidases were reported to degenerate the bacterial extracellular
207 peptidoglycan, contributing to nutrition acquisition or defense against competitors (36,
208 37). Serine peptidases are often used as marker enzyme for proteolysis activity in soil,
209 and play important roles in the utilization of nitrogen sources in the environments (38).
210 The presence of abundant potentially secreted peptidases in *Exiguobacterium*
211 genomes could allow them to exploit different niches for nitrogen source uptake in
212 different environments.

213 To validate the potential ability of *Exiguobacterium* spp. to degrade and metabolize
214 complex carbohydrates and proteins, the amylase and protease activities of the 103
215 strains isolated in this study were tested on plates (Fig. S1). All of these strains
216 showed effective hydrolysis ability for starch, and approximately 70% of these strains
217 can degenerate proteins. Taking results from both the genomic analysis and activity
218 testing together, it provided strong evidence that most members of the
219 *Exiguobacterium* genus have the abilities to metabolic and utilize a wide range of
220 nutrition from marine and non-marine environments, which explains the genetic basis
221 for the cosmopolitan distribution of these bacteria.

222 **Genetic basis of keeping homeostasis in extreme environments.** As a versatile
223 genus, *Exiguobacterium* was found to survive in many extreme environments such as
224 cold, hot, saline and pollutant (16). Our phenotype test has proved that these adaptive
225 characteristics are shared by most members of this genus. We investigated the putative
226 genetic determinants behind these abilities from the whole genus.

227 Two strategies are used by bacteria to survive in cold environment, which are
228 utilization of unsaturated branched-chain fatty acids to maintain membrane fluidity
229 and expression of cold shock proteins (Csp) that stabilize the bacterial cytosol at low
230 temperatures (39, 40). From the genomic analysis, we found that all members of

*Exiguobacterium* genus could use the both strategies to cope with low temperature. Two types of fatty acid desaturase (FAD) involved in unsaturated branched-chain fatty acid production were identified in *Exiguobacterium* genomes (Fig. 4; Table S8). All genomes except those of AB2 and s126 encoded at least one FAD1 protein, while gene for FAD2 was mainly contained by the strains belonging to Group I. Three types of *csp* (*cspA*, *cspB* and *cspC*) were predicted from all the genomes (Fig. 4). Most genomes in Group I contained more than 2 *cspA* genes while those from Group II had only one. The *cspB* and *cspC* were harbored by members from Group I and II, respectively. It was reported that members can grow below 0°C mainly belonging to Group I (15). More *csp* and *fad* genes contained by member of Group I than II may contribute to this difference.

To support to survive in hot environments, all strains of *Exiguobacterium* spp. contain the shock gene cluster *grpE-dnaJ-dnaK* (Fig. 4; Table S8), which encodes chaperones that prevent aggregation and denaturation of proteins at high temperature (41). It was reported that proteins GroEL/GroES cooperating with DnaK/DnaJ to prevent protein misfolding in bacteria (42). Genes encoding GroEL and GroES were also discovered in all of the strains except for EHD646 (Fig. 4; Table S8). Additionally, other 3 types of heat shock protein (HSP20, HSP33 and HSP90) were also predicted. These HSPs are important chaperone for appropriate response to heat or oxidative stress because of their capacity of preventing irreversible protein denaturation (43-46). This suggests that members of this genus utilize multiple ways to cope with heat environments.

In bacteria, the $Na^+$:$H^+$ antiporters play crucial roles in the maintenance of intracellular pH homeostasis and dynamic balance of cellular $Na^+$. According to the Transporter Classification Database (TCDB), $Na^+$:$H^+$ antiporters mainly contain the large monovalent cation/proton antiporter (CPA) family such as CPA1, CPA2 and CPA3, and the NhaC $Na^+$:$H^+$ antiporter family (47, 48). Seven types of CPA1 and 1 from CPA2 of the $Na^+$:$H^+$ antiporters of were predicted from the genomes of *Exiguobacterium* spp. (Fig. 4; Table S8). The $Na^+$:$H^+$ antiporters from CPA1 and CPA2 family were found in partial *Exiguobacterium* genomes, with the former one

more frequently from ones of Group I and the last more common from those belonging to Group II. Compared to CPA1 and CPA2, CPA3 antiporters are more structurally complex with a multicomponent structure consisting of either seven or six members (49). This multicomponent $Na^+$:$H^+$ antiporter (Mrp) has been proved to provide $Na^+$/$H^+$ antiport activity and function in multiple compound resistance and pH homeostasis in *Bacillus subtilis* (49). In this study, Mrp antiporters were identified from all *Exiguobacterium* genomes (Fig. 4; Table S8). In addition, the antiporter from NhaC $Na^+$:$H^+$ antiporter (NhaC) family was identified in all *Exiguobacterium* genomes with copy number up to 6 (Fig. 4; Table S8). The presence of multiple types of $Na^+$:$H^+$ antiporter provides the basis for *Exiguobacterium* to maintain osmosis and pH balance in a variety of environments.

The combining results of the existence of diverse important proteins, including cold- and heat-shock protein, chaperonin, fatty acid desaturase and diverse $Na^+$:$H^+$ antiporters could explain the broad range of acceptable temperatures, pH and salinity for *Exiguobacterium* strains, and help to colonize in diverse habitats.

**Expansion of transporter families contributing to a wider adaptation.** By comparing the ecological difference between members of Group I and II, we found that strains from marine environments were more frequently assigned to Group II and species from Group I had a more diverse niche distribution. Based on the analysis of 16S rRNA gene sequences from GenBank, there were 54 and 292 respectively belonging to Group I and II among the 346 sequences from marine environments (Table S1). As for the strains isolated in this study, there were 25 and 62 respectively belonging to Group I and II among the 87 strains from marine environments. In Group I, 10 of the 11 species contained strains isolated not only from the marine environment, but also from various terrestrial environments such as soil, plant rhizosphere, fresh water, etc (Table S5). While in Group II, most species were mainly isolated from marine-associated environments (Table S5). More diverse niche distribution suggested that the members belonging to Group I have a stronger capacity for environments adaptation.

To understand the genetic background behind the ecological difference of these two groups, we performed comparative genomic analysis and found a similar increasing tendency of genome size and transporter number from species of Group I to Group II (Fig. 5A; Fig. 5B and Table S9). The average genome size and transporter number of Group I (3.12Mb and 648) was significant larger than those of Group II (2.90Mb and 610) ($p < 0.0001$, Wilcoxon test) (Fig. 5C; Fig. 5D). Moreover, the spearman correlation coefficients of transporter number with genome size and CDS number is 0.86 (Fig. 5E). The significant correlation ($p < 0.01$) with genome size and CDS number suggests that the expansion of transporter contributed to the difference of genome content in Group I and Group II strains.

Transporters are vital to all living organisms in the uptake of nutrients, secretion of metabolites, maintenance of ion concentration gradient across membranes and efflux of drug and toxins (50). In this study, 25 of the 247 identified transporter families were identified to have high degrees of correlation with both genome size and CDS number (correlation coefficients > 0.6, $p < 0.01$, Table 1). These 25 transporter gene families were significantly enriched in genome of Group I compared to Group II (Fig. 6; Table S9).

Seven of the 25 families are associated with the transport of diverse amino acids. Among them, cationic amino acid, polar amino acid, branched chain amino acid and basic amino acid are important components for nitrogen metabolism, protein synthesis, cell growth and energy production or conversion (51). Besides, 2 of the 25 families are involved in the transport of $Mg^{2+}$, including the cyclin M $Mg^{2+}$ exporter family and the CorA metal ion transporter family. $Mg^{2+}$ homeostasis is important in bacteria and has been reported to play a critical role in their thermotolerance (52, 53). Moreover, the inorganic phosphate transporter family was also expanded in Group I. Compared with the marine environment, the terrestrial environment is more diverse and has a variety of complex microenvironments due to the influence of climate or seasons, which lead to a diverse content of nutrient substrates (3, 54). The enhanced capability of Group I strains in important substrates transport meets the need of

cellular metabolism and functions, and provides important base to survive in more diverse terrestrial environments.

Three of the 25 transporter families are involved in transport of heavy metals ions, including the arsenite-antimonite efflux family, the iron/lead transporter family and peptide/opine/nickel uptake transporter family. These transporters have been shown to counteract the effects of toxic heavy metals (55-57). The members of major facilitator superfamily (MFS) are capable of transporting a wide range of substrates in response to ion gradients or function as drug:$H^+$ antiporter; and majority of bacterial drug efflux pumps classified within the MFS (58, 59). Four of the 25 families are belonging to MFS superfamily, among which 2 are associated with drug efflux. Moreover, 2 of the 25 families are involved in the formation of bacterial cell wall and biofilm, including the teichoic acid exporter family and the 4 TMS GlpM family. Teichoic acid is a major cell wall component of Gram-positive bacteria, and has been proved to play crucial roles in bacterial resistance to antimicrobial and survival under disadvantageous conditions (60, 61). The members of 4 TMS GlpM family is required for normal production of alginate (62). Alginates are important polymeric substances contributing to the formation and development of biofilm matrixes of numerous bacteria enhancing colonization and persistence under environmental stresses (63). Due to the emission or leach out from the industrial and agricultural fields, terrestrial and freshwater ecosystems are contaminated with heavy metals or pesticide severely (64, 65). The expansion of transporters involved in disadvantageous conditions resistance and efflux of drug and heavy metals may contribute a more diverse distribution in non-marine habitats of Group I strains.

Efficient transport of substances related to metabolism, cellular function or environment stresses resistance is crucial for bacterial survival in a variety of environments (66). In order to survive in more diverse environments, bacteria have to develop specific systems for their survival such as nutrient sensing and transport systems (67). Bacteria are often exposed to stress conditions in many stages of their life cycle; and the capacity of environmental stress resistance determines the

distribution of microorganisms (68, 69). Therefore, these expanded families that related to environment stresses resistance and transport of organic or inorganic substrates, may play crucial roles for *Exiguobacterium* spp. from Group I to survive in more diverse environments.

**Conclusions.** Numerous studies have investigated the evolution and adaptation mechanism of important pathogens (70). However, fewer studies have addressed how microbes with cosmopolitan distribution but relative low abundance to adapt to diverse habitats. The wide distribution characteristic makes the genus *Exiguobacterium* as a valuable model for studying the adaptive strategy of bacteria to multiple habitats. This study suggested that these bacteria with nomadic lifestyle and cosmopolitan distribution are usual generalists, which can utilize a variety of nutrients simultaneously and resist diverse environmental stresses.

Although strains of *Exiguobacterium* genus are generalists, the ecological difference between members of Group I and II was still discovered. We found the species from Group I was distributed in more diverse environments with larger genome size and the expansion of transporter families contributed to the difference of genome size. Most of the enriched transporter families are involved in environment stresses resistance and transport of organic or inorganic substrates, which may play vital role for Group I strains in adaptation to wider habitats.

## MATERIALS AND METHODS

**Analysis of *Exiguobacterium* 16S rRNA gene sequences.** We retrieved *Exiguobacterium* 16S rRNA gene sequences from GenBank. The information of isolation source of these sequences was collected. The habitats were classified into 13 types, including air, animal-associated environments, cold-associated environments, estuary-associated environments, contaminated water or soil, freshwater or freshwater sediment, hot-associated environments, hypersaline environments, marine-associated environments, plant, rhizosphere, soil and other inland environments.

**Bacterial isolation and culture.** A total of 103 strains were used for adaptive

375 experiments and genome sequencing, including 98 isolated from terrestrial and
376 marine environments worldwide by us, and 5 type strains obtained from DSMZ and
377 CGMCC (Table S2). Initially samples from marine and terrestrial environments were
378 macerated and mixed with sterile saline solution (0.8%) using a standard dilution
379 plating method on Marine agar 2216 (MA, Difco) and LB agar at 20 ℃, respectively.
380 All these 103strains can grow in sea salts free medium and were routinely cultivated
381 on LB agar and in liquid LB for subsequent genomic sequencing.

382 **Adaptive ability tests of pH, temperature and salinity**. To assess the range of
383 adaptation to pH, temperature and salinity, all of the 103 strains were measured for
384 assessing growth under the following conditions. The temperature range for growth
385 were measured at 4, 25, 30, 40, 45 and 55 °C (the lowest growth temperature was not
386 recorded for the strain that could not grow at 4°C) on LB and in liquid LB. The pH
387 range for growth was determined at 25 °C in liquid LB medium at pH 4.0-12.0 (with
388 intervals of 1 units) using the following buffers: citrate/$Na_2HPO_4$ buffer (pH 4.0-7.0),
389 Tris buffer (pH 7.5-9.0) and $NaHCO_3$/$Na_2CO_3$ buffer (pH 9.5-10.0); growth was
390 evaluated by measuring OD600 after 7 days of incubation. Growth with 0–15% (in 1 %
391 increments, w/v) NaCl was investigated after 14 days of cultivation at 25 °C in the
392 following medium: 0.1% peptone, 0.1% yeast extract, 0.03% KCl, 0.25%
393 $MgSO_4.7H_2O$, 0.05% $CaCl_2$.

394 **Degradation ability tests of complex carbohydrates and proteins.** Protease and
395 amylase activities were tested on LB agar supplemented with skim milk (2%, w/v)
396 and starch (0.4%, w/v). After 4–6 days at 25°C, a positive reaction was noticed when
397 transparent zones around the colonies were directly visible or detected after coloration
398 of the undegraded substrate.

399 **Genome sequencing, assembly and annotation**. Genomic DNA was extracted by
400 using a bacterial genomic DNA Mini kit (TaKaRa Bio) following the manufacturer's
401 protocol. Genomes of the 103 *Exiguobacterium* strains were sequenced using Illumina
402 NovaSeq 6000 platform. The raw reads of each genome were trimmed using
403 trimmomatic v0.36 (71) and *de novo* assembled using (25). The genomes of

*Exiguobacterium* deposited in GenBank were collected and filtered based on the criterion that genomes were at least 95% complete, with < 5% contamination based on CheckM analysis (72). The ANI values between two each genome pair was computed using the OrthoANI software (73). Gene predictions and annotations of all genomes were generated using Prokka (74).

**Phylogenetic tree construction.** Analysis of orthologous clusters was performed using the FastOrtho (http://enews.patricbrc.org/fastortho/), a faster reimplementation of OrthoMCL (75). In brief, an all-against-all BLAST was firstly performed with *E*-values < $1 \times 10^{-5}$. Then, ortholog groups were created with the MCL algorithm with an inflation value of 2, and the single-copy gene families were obtained using custom-made Python scripts. Protein sequences of each family were aligned by MUSCLE (76) and then trimmed by trimAL (77). All trimmed alignments were then concatenated into a new alignment by a local Python script. Single-copy core gene-based phylogenetic tree was constructed using RAxML (78) with 1000 bootstrap replicates, employing the LG+I+G+F model. iTOL was used for the phylogenetic tree visualization (79).

**Identification of carbohydrate-active enzymes and proteases.** For genes encoding carbohydrate-active enzymes and proteases, all of the annotation genes were searched against the CAZy database (www.cazy.org) (80) and peptidase database (MEROPS) (81) with *E*-values < $1 \times 10^{-5}$ by BLASTP. The potential secreted carbohydrate-active enzymes and peptidases were confirmed based on the identification of extracellular transport signals using SignalP (82). Genes related to carbohydrate-active enzymes and proteases were further classified into different groups according to the predictions.

**Identification of vital genes for environmental stresses resistance.** For each predicted protein by Prokka were annotated using BLASTP and Hmmscan against Clusters of Orthologous Groups (COG) database and PFAM database with *E*-values < $1 \times 10^{-5}$, respectively. Transporters were predicted by performing BLASTP with *E*-values < $1 \times 10^{-10}$ using *Exiguobacterium* protein sequences against all Transporter

Classification Database (TCDB) sequences (83).

**Data analyses.** Statistical analyses were performed using Wilcoxon test. Correlation analysis on the genome size, CDS number and transporter number was performed using chart.Correlation from the PerformanceAnalytics package in R (https://cran.r-project.org/web/packages/PerformanceAnalytics/index.html).

**Data availability.** The genomes supporting the results have been deposited at DDBJ/ENA/GenBank under the BioProjectID PRJNA644789 (accession numbers from JACSJK000000000 to JACSNI000000000) (Table S2).

## ACKNOWLEDGEMENTS

This study was supported by National Natural Science Foundation of China (31670002, 31970003 and 31770003).

We declare that we have no conflict of interest.

# REFERENCE


1. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Øvreås L, Reysenbach A-L, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4:102-112.
2. Bock C, Jensen M, Forster D, Marks S, Nuy J, Psenner R, Beisser D, Boenigk J. 2020. Factors shaping community patterns of protists and bacteria on a European scale. Environ Microbiol 22:2243-2260.
3. Valli S, Suvathi SS, Aysha OS, Nirmala P, Vinoth KP, Reena A. 2012. Antimicrobial potential of *Actinomycetes* species isolated from marine environment. Asian Pacific Journal of Tropical Biomedicine 2:469-473.
4. Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K. 2009. Infrequent marine–freshwater transitions in the microbial world. Trends Microbiol 17:414-422.
5. Logares R, Shalchian-Tabrizi K, Boltovskoy A, Rengefors K. 2007. Extensive dinoflagellate phylogenies indicate infrequent marine–freshwater transitions. Mol Phylogen Evol 45:887-903.
6. Getz EW, Tithi SS, Zhang LQ, Aylward FO. 2018. Parallel evolution of genome streamlining and cellular bioenergetics across the marine radiation of a bacterial phylum. mBio 9:e01089-18.
7. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. mBio 3:e00252-12.
8. Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett 10:1135-1142.
9. Eichorst SA, Trojan D, Roux S, Herbold C, Rattei T, Woebken D. 2018. Genomic insights into the *Acidobacteria* reveal strategies for their success in terrestrial environments. Environ Microbiol 20:1041-1063.
10. Zhang H, Yoshizawa S, Sun Y, Huang YJ, Chu X, González JM, Pinhassi J, Luo HW. 2019. Repeated evolutionary transitions of flavobacteria from marine to non-marine habitats. Environ Microbiol 21:648-666.
11. Cabello-Yeves PJ, Rodriguez-Valera F. 2019. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. Microbiome 7:117.
12. Tanaka M, Kumakura D, Mino S, Doi H, Ogura Y, Hayashi T, Yumoto I, Cai M, Zhou YG, Gomez-Gil B, Araki T, Sawabe T. 2020. Genomic characterization of closely related species in the Rumoiensis clade infers ecogenomic signatures to non-marine environments. Environ Microbiol 22:3205-3217.
13. Giovannoni SJ. 2017. SAR11 bacteria: The most abundant plankton in the oceans. Annual Review of Marine Science 9:231-255.
14. Santoro AE, Richter RA, Dupont CL. 2019. Planktonic marine archaea. Annual review of marine science 11:131-158.

15. Vishnivetskaya TA, Kathariou S, Tiedje JM. 2009. The *Exiguobacterium* genus: biodiversity and biogeography. Extremophiles 13:541-555.
16. Kasana RC, Pandey CB. 2018. *Exiguobacterium*: an overview of a versatile genus with potential in industry and agriculture. Crit Rev Biotechnol 38:141-156.
17. Singh NK, Raichand R, Kaur I, Kaur C, Pareek S, Mayilraj S. 2013. *Exiguobacterium himgiriensis* sp. nov. a novel member of the genus *Exiguobacterium*, isolated from the Indian Himalayas. Antonie Van Leeuwenhoek 103:789-796.
18. Raichand R, Pareek S, Singh NK, Mayilraj S. 2012. *Exiguobacterium aquaticum* sp. nov., a member of the genus *Exiguobacterium*. Int J Syst Evol Microbiol 62:2150-2155.
19. Khandare RV, Rane NR, Waghmode TR, Govindwar SP. 2012. Bacterial assisted phytoremediation for enhanced degradation of highly sulfonated diazo reactive dye. Environmental Science and Pollution Research 19:1709-1718.
20. Jiang X, Xue Y, Wang L, Yu B, Ma Y. 2013. Genome sequence of a novel polymer-grade L-lactate-producing alkaliphile, *Exiguobacterium* sp. strain 8-11-1. Genome Announcements 1:e00616-13.
21. Vishnivetskaya TA, Lucas S, Copeland A, Lapidus A, Glavina del Rio T, Dalin E, Tice H, Bruce DC, Goodwin LA, Pitluck S, Saunders E, Brettin T, Detter C, Han C, Larimer F, Land ML, Hauser LJ, Kyrpides NC, Ovchinnikova G, Kathariou S, Ramaley RF, Rodrigues DF, Hendrix C, Richardson P, Tiedje JM. 2011. Complete genome sequence of the thermophilic bacterium *Exiguobacterium* sp. AT1b. J Bacteriol 193:2880.
22. Gutiérrez-Preciado A, Vargas-Chávez C, Reyes-Prieto M, Ordoñez OF, Santos-García D, Rosas-Pérez T, Valdivia-Anistro J, Rebollar EA, Saralegui A, Moya A, Merino E, Farías ME, Latorre A, Souza V. 2017. The genomic sequence of *Exiguobacterium chiriqhucha* str. N139 reveals a species that thrives in cold waters and extreme environmental conditions. PeerJ 5:e3162.
23. Dias LM, Folador ARC, Oliveira AM, Ramos RTJ, Silva A, Baraúna RA. 2018. Genomic architecture of the two cold-adapted genera *Exiguobacterium* and *Psychrobacter*: evidence of functional reduction in the *Exiguobacterium antarcticum* B7 Genome. Genome Biology and Evolution 10:731-741.
24. Castro-Severyn J, Pardo-Esté C, Mendez KN, Morales N, Marquez SL, Molina F, Remonsellez F, Castro-Nallar E, Saavedra CP. 2020. Genomic variation and arsenic tolerance emerged as niche specific adaptations by different *Exiguobacterium* strains isolated from the extreme Salar de Huasco environment in Chilean – Altiplano. Frontiers in Microbiology 11.
25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-477.
26. Kim M, Oh H-S, Park S-C, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346-351.
27. Logares R, Bråte J, Heinrich F, Shalchian-Tabrizi K, Bertilsson S. 2009. Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). Mol Biol Evol 27:347-357.
28. Collinson SR, Thielemans W. 2010. The catalytic oxidation of biomass to new materials

focusing on starch, cellulose and lignin. Coordination Chemistry Reviews 254:1854-1870.

29. Prabhu M, Chemodanov A, Gottlieb R, Kazir M, Nahor O, Gozin M, Israel A, Livney YD, Golberg A. 2019. Starch from the sea: the green macroalga *Ulva ohnoi* as a potential source for sustainable starch production in the marine biorefinery. Algal Research 37:215-227.

30. Domozych DS, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WGT. 2012. The cell walls of green algae: a journey through evolution and diversity. Frontiers in plant science 3:82-82.

31. Broxterman SE, Schols HA. 2018. Interactions between pectin and cellulose in primary plant cell walls. Carbohydrate Polymers 192:263-272.

32. Rinaudo M. 2006. Chitin and chitosan: Properties and applications. Progress in Polymer Science 31:603-632.

33. Nguyen STC, Freund HL, Kasanjian J, Berlemont R. 2018. Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. Appl Microbiol Biotechnol 102:1629-1637.

34. Elieh-Ali-Komi D, Hamblin MR. 2016. Chitin and chitosan: Production and application of versatile biomedical nanomaterials. International journal of advanced research 4:411-427.

35. Nguyen TTH, Myrold DD, Mueller RS. 2019. Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. Frontiers in microbiology 10:413-413.

36. Stohl EA, Chan YA, Hackett KT, Kohler PL, Dillard JP, Seifert HS. 2012. *Neisseria gonorrhoeae* virulence factor NG1686 is a bifunctional M23B family metallopeptidase that influences resistance to hydrogen peroxide and colony morphology. J Biol Chem 287:11222-11233.

37. Baba T, Schneewind O. 1996. Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*. The EMBO Journal 15:4789-4797.

38. Fuka MM, Engel M, Gattinger A, Bausenwein U, Sommer M, Munch JC, Schloter M. 2008. Factors influencing variability of proteolytic genes and activities in arable soils. Soil Biol Biochem 40:1646-1653.

39. Barria C, Malecki M, Arraiano C. 2013. Bacterial adaptation to cold. Microbiology 159:2437-2443.

40. De Maayer P, Anderson D, Cary C, Cowan DA. 2014. Some like it cold: understanding the survival strategies of psychrophiles. EMBO reports 15:508-517.

41. Feder ME, Hofmann GE. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu Rev Physiol 61:243-282.

42. Gragerov A, Nudler E, Komissarova N, Gaitanaris GA, Gottesman ME, Nikiforov V. 1992. Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli*. Proc Natl Acad Sci 89:10341.

43. Narberhaus F. 2002. α-Crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. Microbiol Mol Biol Rev 66:64-93.

44. Wholey WY, Jakob U. 2012. Hsp33 confers bleach resistance by protecting elongation factor Tu against oxidative degradation in Vibrio cholerae. Mol Microbiol 83:981-991.

45. Honoré FA, Méjean V, Genest O. 2017. Hsp90 is essential under heat stress in the bacterium *Shewanella oneidensis*. Cell Reports 19:680-687.

46. Segal N, Shapira M. 2015. HSP33 in eukaryotes – an evolutionary tale of a chaperone adapted

to photosynthetic organisms. The Plant Journal 82:850-860.

47. Krulwich TA, Hicks DB, Ito M. 2009. Cation/proton antiporter complements of bacteria: why so large and diverse? Mol Microbiol 74:257-260.

48. Lee C, Kang HJ, von Ballmoos C, Newstead S, Uzdavinys P, Dotson DL, Iwata S, Beckstein O, Cameron AD, Drew D. 2013. A two-domain elevator mechanism for sodium/proton antiport. Nature 501:573-577.

49. Kajiyama Y, Otagiri M, Sekiguchi J, Kosono S, Kudo T. 2007. Complex formation by the *mrpABCDEFG* gene products, which constitute a principal $Na^+/H^+$ antiporter in *Bacillus subtilis*. J Bacteriol 189:7511.

50. Busch W, Saier MH. 2002. The transporter classification (TC) system, 2002. Crit Rev Biochem Mol Biol 37:287-337.

51. Tesseraud S, Everaert N, Boussaid-Om Ezzine S, Collin A, Métayer-Coustard S, Berri C. 2011. Manipulating tissue metabolism by amino acids. World's Poultry Science Journal 67:243-252.

52. Prasad D, Verma N, Bakshi M, Narayan OP, Singh AK, Dua M, Johri AK. 2019. Functional characterization of a magnesium transporter of root endophytic fungus *Piriformospora indica*. Frontiers in Microbiology 9.

53. Yang Y, Huang S, Wang J, Jan G, Jeantet R, Chen XD. 2017. $Mg^{2+}$ improves the thermotolerance of probiotic *Lactobacillus rhamnosus* GG, *Lactobacillus casei* Zhang and *Lactobacillus plantarum* P-8. Lett Appl Microbiol 64:283-288.

54. DeMarco J, Mack MC, Bret-Harte MS. 2011. The effects of snow, soil microenvironment, and soil organic matter quality on N availability in three Alaskan Arctic plant communities. Ecosystems 14:804-817.

55. Große C, Scherer J, Koch D, Otto M, Taudte N, Grass G. 2006. A new ferrous iron-uptake transporter, EfeU (YcdN), from *Escherichia coli*. Mol Microbiol 62:120-131.

56. Zhao FJ. 2016. A novel pathway of arsenate detoxification. Mol Microbiol 100:928-930.

57. Cavazza C. 2011. Extracytoplasmic Nickel‐binding proteins. Encyclopedia of Inorganic and Bioinorganic Chemistry:1-11.

58. dos Santos SC, Teixeira MC, Dias PJ, Sá-Correia I. 2014. MFS transporters required for multidrug/multixenobiotic (MD/MX) resistance in the model yeast: understanding their physiological function through post-genomic approaches. Frontiers in Physiology 5.

59. Chen LH, Tsai HC, Yu PL, Chung KR. 2017. A major facilitator superfamily transporter-mediated resistance to oxidative stress and fungicides requires Yap1, Skn7, and MAP kinases in the citrus fungal pathogen Alternaria alternata. PLoS One 12.

60. Chen L, Hou WT, Fan T, Liu BH, Pan T, Li YH, Jiang YL, Wen W, Chen ZP, Sun LF, Zhou CZ, Chen YX. 2020. Cryo-electron microscopy structure and transport mechanism of a wall teichoic acid abc transporter. mBio 11:e02749-19.

61. Swoboda JG, Campbell J, Meredith TC, Walker S. 2010. Wall teichoic acid function, biosynthesis, and inhibition. ChemBioChem 11:35-45.

62. Schweizer HP, Po C, Bacic MK. 1995. Identification of *Pseudomonas aeruginosa glpM*, whose gene product is required for efficient alginate biosynthesis from various carbon sources. J Bacteriol 177:4801.

63. Moradali MF, Rehm BHA. 2019. The role of alginate in bacterial biofilm formation, p 517-537. *In* Cohen E, Merzendorfer H (ed), Extracellular sugar-based biopolymers matrices doi:10.1007/978-3-030-12919-4_13. Springer International Publishing, Cham.

64. Xing W, Zheng Y, Scheckel KG, Luo Y, Li L. 2019. Spatial distribution of smelter emission heavy metals on farmland soil. Environ Monit Assess 191:115.
65. Haddaway NR, Brown C, Eggers S, Josefsson J, Kronvang B, Randall N, Uusi-Kämppä J. 2016. The multifunctional roles of vegetated strips around and within agricultural fields. A systematic map protocol. Environmental Evidence 5:18.
66. Lorca GL, Barabote RD, Zlotopolski V, Tran C, Winnen B, Hvorup RN, Stonestrom AJ, Nguyen E, Huang L-W, Kim DS, Saier MH. 2007. Transport capabilities of eleven gram-positive bacteria: comparative genomic analyses. Biochim Biophys Acta 1768:1342-1366.
67. Samyn DR, Persson BL. 2016. Inorganic phosphate and sulfate transport in *S. cerevisiae*, p 253-269. *In* Ramos J, Sychrová H, Kschischo M (ed), Yeast membrane transport doi:10.1007/978-3-319-25304-6_10. Springer International Publishing, Cham.
68. Chu H, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. 2010. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environ Microbiol 12:2998-3006.
69. Jock S, Langlotz C, Geider K. 2005. Survival and possible spread of *Erwinia amylovora* and related plant-pathogenic bacteria exposed to environmental stress conditions. J Phytopathol 153:87-93.
70. Sperandio V. 2018. Pathogens' adaptation to the human host. Proc Natl Acad Sci 115:9342.
71. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.
72. 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.
73. Lee I, Ouk Kim Y, Park S-C, Chun J. 2016. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100-1103.
74. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.
75. Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178-2189.
76. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.
77. 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.
78. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
79. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256-W259.
80. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2008. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233-D238.
81. Rawlings ND, Waller M, Barrett AJ, Bateman A. 2013. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 42:D503-D509.
82. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal

665 peptides from transmembrane regions. Nat Methods 8:785-786.
666 83. Saier MH, Jr., Reddy VS, Tamang DG, Västermark Å. 2013. The transporter classification
667 database. Nucleic Acids Res 42:D251-D258.

668 **Table 1** Correlation analysis of transporter family with both genome size and CDS number.
669

| Genome content | TC number | Family | Correlation coefficients | p-value |
|---|---|---|---|---|
| Genome size | 2.A.1 | **The Major Facilitator (MFS) Superfamily** | 0.73 | 0 |
| CDS number | 2.A.1 | | 0.66 | 0 |
| Genome size | 2.A.1.3 | The Drug:$H^+$ Antiporter-2 (14 Spanner) (DHA2) Family | 0.73 | 0 |
| CDS number | 2.A.1.3 | | 0.66 | 0 |
| Genome size | 2.A.1.76 | The Uncharacterized Major Facilitator 24 Family | 0.69 | 0 |
| CDS number | 2.A.1.76 | | 0.61 | 4.44E-16 |
| Genome size | 2.A.1.32 | The Putative Aromatic Compound/Drug Exporter (ACDE) Family | 0.68 | 0 |
| CDS number | 2.A.1.32 | | 0.60 | 8.88E-16 |
| Genome size | 2.A.1.6 | The Metabolite:$H^+$ Symporter (MHS) Family | 0.66 | 0 |
| CDS number | 2.A.1.6 | | 0.67 | 0 |
| Genome size | **2.A.3** | **The Amino Acid-Polyamine-Organocation (APC) Superfamily** | 0.67 | 0 |
| CDS number | **2.A.3** | | 0.61 | 2.22E-16 |
| Genome size | 2.A.3.1 | The Amino Acid Transporter (AAT) Family | 0.71 | 0 |
| CDS number | 2.A.3.1 | | 0.66 | 0 |
| Genome size | 2.A.3.3 | The Cationic Amino Acid Transporter (CAT) Family | 0.68 | 0 |
| CDS number | 2.A.3.3 | | 0.60 | 8.88E-16 |
| Genome size | **3.A.1** | **The ATP-binding Cassette (ABC) Superfamily** | 0.75 | 0 |
| CDS number | **3.A.1** | | 0.69 | 0 |
| Genome size | 3.A.1.5 | The Peptide/Opine/Nickel Uptake Transporter (PepT) Family | 0.72 | 0 |
| CDS number | 3.A.1.5 | | 0.68 | 0 |
| Genome size | 3.A.1.3 | The Polar Amino Acid Uptake Transporter (PAAT) Family | 0.73 | 0 |
| CDS number | 3.A.1.3 | | 0.66 | 0 |

| | | | | |
|---|---|---|---|---|
| Genome size | 3.A.1.155 | The Phage Infection Protein (PIP) Family | 0.72 | 0 |
| CDS number | 3.A.1.155 | | 0.63 | 0 |
| Genome size | 3.A.1.104 | The Teichoic Acid Exporter (TAE) Family | 0.68 | 0 |
| CDS number | 3.A.1.104 | | 0.60 | 1.33E-15 |
| Genome size | 1.A.112 | The Cyclin M $Mg^{2+}$ Exporter (CNNM) Family | 0.72 | 0 |
| CDS number | 1.A.112 | | 0.70 | 0 |
| Genome size | 1.A.8 | The Major Intrinsic Protein (MIP) Family | 0.70 | 0 |
| CDS number | 1.A.8 | | 0.61 | 2.22E-16 |
| Genome size | 1.A.35 | The CorA Metal Ion Transporter (MIT) Family | 0.68 | 0 |
| CDS number | 1.A.35 | | 0.60 | 8.88E-16 |
| Genome size | 1.I.3 | The Bacterial (Planctomycetes) Nuclear Pore-like Complex (B-NPC) Family | 0.63 | 0 |
| CDS number | 1.I.3 | | 0.62 | 2.22E-16 |
| Genome size | 2.A.23 | The Dicarboxylate/Amino Acid:Cation ($Na^+$ or $H^+$) Symporter (DAACS) Family | 0.71 | 0 |
| CDS number | 2.A.23 | | 0.65 | 0 |
| Genome size | 2.A.26 | The Branched Chain Amino Acid: Cation Symporter (LIVCS) Family | 0.61 | 4.44E-16 |
| CDS number | 2.A.26 | | 0.61 | 2.22E-16 |
| Genome size | 2.A.45 | The Arsenite-Antimonite (ArsB) Efflux Family | 0.60 | 1.11E-15 |
| CDS number | 2.A.45 | | 0.62 | 0 |
| Genome size | 2.A.118 | The Basic Amino Acid Antiporter (ArcD) Family | 0.68 | 0 |
| CDS number | 2.A.118 | | 0.60 | 8.88E-16 |
| Genome size | 2.A.20 | The Inorganic Phosphate Transporter (PiT) Family | 0.68 | 0 |
| CDS number | 2.A.20 | | 0.60 | 8.88E-16 |
| Genome size | 2.A.108 | The Iron/Lead Transporter (ILT) Family | 0.69 | 0 |
| CDS number | 2.A.108 | | 0.61 | 4.44E-16 |
| Genome size | 2.A.76 | The Resistance to Homoserine/Threonine (RhtB) Family | 0.61 | 8.88E-16 |
| CDS number | 2.A.76 | | 0.61 | 4.44E-16 |

| | | | | |
|---|---|---|---|---|
| Genome size | 3.A.6 | The Type III (Virulence-related) Secretory Pathway (IIISP) Family | 0.62 | 0 |
| CDS number | 3.A.6 | | 0.62 | 0 |
| Genome size | 5.B.1 | The gp91$^{phox}$ Phagocyte NADPH Oxidase-associated Cytochrome $b_{558}$ (Phox) Family | 0.73 | 0 |
| CDS number | 5.B.1 | | 0.67 | 0 |
| Genome size | 9.B.34 | The Kinase/Phosphatase/Cyclic-GMP Synthase/Cyclic di-GMP Hydrolase (KPSH) Family | 0.75 | 0 |
| CDS number | 9.B.34 | | 0.76 | 0 |
| Genome size | 9.B.72 | The 4 TMS GlpM (GlpM) Family | 0.63 | 0 |
| CDS number | 9.B.72 | | 0.63 | 0 |

670

**Fig. 1 Cosmopolitan distribution of *Exiguobacterium* strains**. (A) Relative abundance of 16S rRNA gene sequences among 13 types of habitats. (B) The isolation sites of 98 *Exiguobacterium* strains in this study (The pink dots represent the sampling location). (C) Temperature, pH and salinity tolerance test of 103 *Exiguobacterium* strains. In the temperature experiment, the intermediate temperature between 4℃ and 25℃ was not tested. The lowest growth temperature was not recorded for the strain that could not grow at 4℃.

**Fig. 2 Phylogenetic analysis of *Exiguobacterium*.** The tree was built using IQ-tree based on the concatenated amino acid sequence alignments of 1,162 core genes. Bootstrap support values were calculated from 1000 replicates. 'T' represents type strain: NIO-1109 for *E. enclense*, HHS31 for *E. indicum*, DSM20416 for *E. acetylicum*, 255-15 for *E. sibiricum*, s145 for *E. artemiae*, 7-3 for *E. sibiricum*, DSM14481 for *E. undae*, JCM12280 for *E. oxidotolerans*, DSM6208 for *E. aurantiacum*, s122 for *E. himgiriensis*, 12-1 for *E. alkaliphilum*, s149 for *E. mexicanum*, DSM16307 for *E. marinum*, s124 for *E. aestuarii* and s121 for *E. profundum*. N1 to N13 represents putative new species.

**Fig. 3 Carbon and nitrogen source utilization.** (A) Number of carbohydrate-active enzymes (CAZymes) encoded in *Exiguobacterium* genomes. (B) Number of plant polysaccharides degradation enzymes encoded in *Exiguobacterium* genomes. (C) Number of extracellular peptidases encoded in *Exiguobacterium* genomes. Brackets: Total number of genomes in each *Exiguobacterium* species. Asterisk: CAZyme with potential secretion signal.

693 **Fig. 4 Vital genes detected across *Exiguobacterium* genomes for keeping**
694 **homeostasis in extreme environments.** The heatmap represents the vital gene
695 number with distribution across the 145 genomes. The maximum-likelihood tree was
696 constructed by RAxML as described in methods.

697 **Fig. 5 Comparison and Spearman's correlation analysis of genome size with**
698 **transporter.** (A) Trends in genome size of *Exiguobacterium* species (the species with
699 5 strains were selected to show the trends). (B) Trends in transporter number of
700 *Exiguobacterium* species (the species with 5 strains were selected to show the trends).
701 (C) Comparison of genome size between Group I and Group II (the black '****'
702 represents significantly different, $p < 0.0001$, Wilcoxon tests). (D) Comparison of
703 transporter number between Group I and Group II. (E) Spearman's correlation
704 analysis of genome size, CDS number and transporter number (the genome size, CDS
705 number and transporter number is shown in the central diagonal; the scatterplots are
706 depicted with a fitted red line and on the corresponding side for each pairing is the
707 Spearman rank correlation coefficient $r_s$, the red asterisks '***' represents
708 significance levels, $p < 0.01$).

709 **Fig. 6 Comparison of 25 transporter family between Group I and Group II.** All
710 pairwise comparisons were significantly different (Wilcoxon test).

711 **Fig. S1 Validation of the ability of *Exiguobacterium* spp. to degrade and**
712 **metabolize complex carbohydrates and peptides.**

713 **Table S1 Information of *Exiguobacterium* 16S rRNA gene sequences in**

714 **Genebank.**

715 **Table S2 Genome features and isolation source of 145 *Exiguobacterium* strains.**

716 **Table S3 Temperature, pH and salinity tolerance test of 103 *Exiguobacterium***
717 **strains.**

718 **Table S4 Average nucleotide identity (ANI) analysis of 145 *Exiguobacterium***
719 **strains.**

720 **Table S5 Niches distribution of 145 *Exiguobacterium* strains basing on**
721 **phylogenetic.**

722 **Table S6 Gene number of carbohydrate-active enzymes (CAZymes) detected in**
723 **each *Exiguobacterium* genome.**

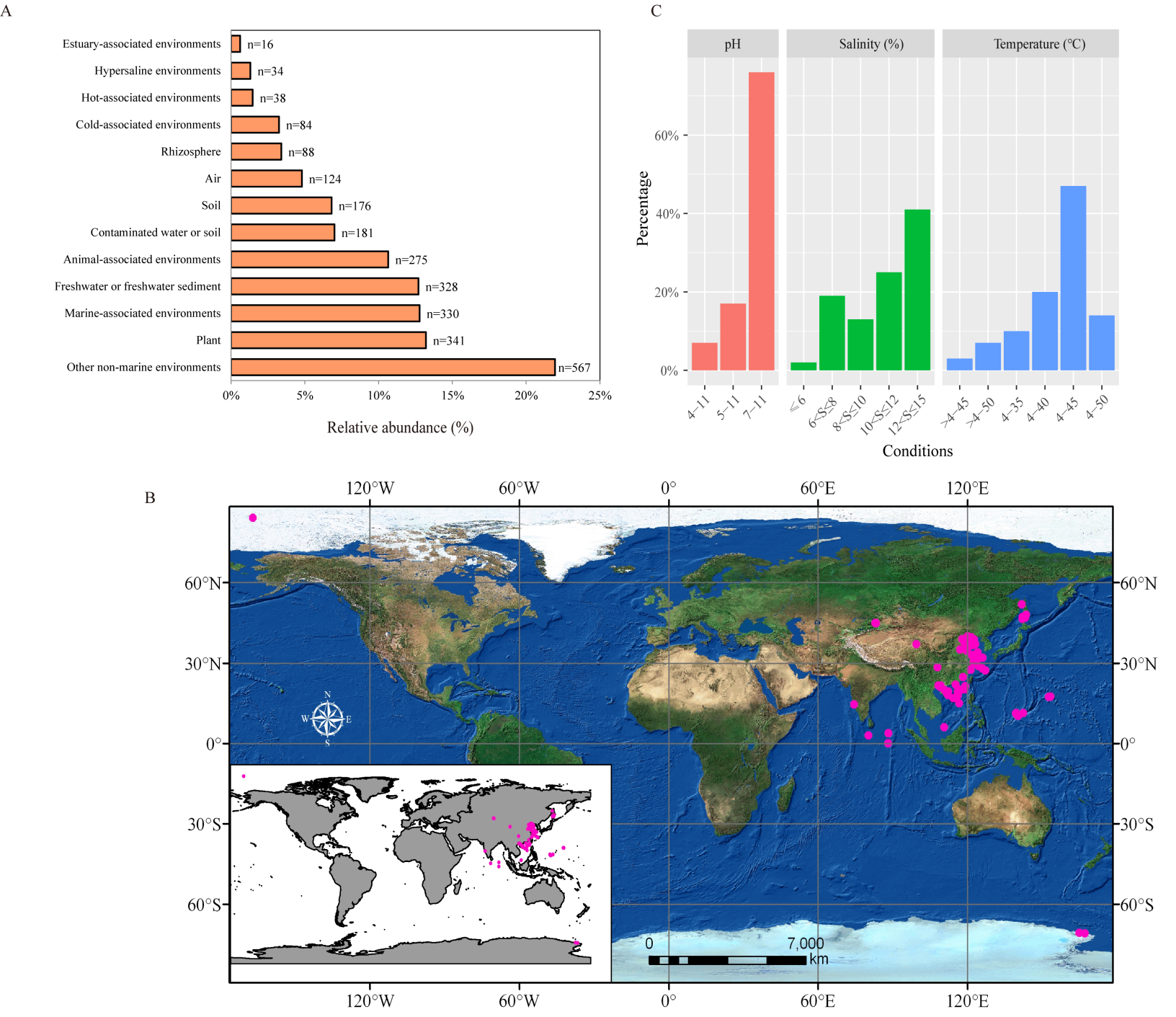
724 **Table S7 Gene number of extracellular peptidases detected in each**
725 ***Exiguobacterium* genome using the MEROPS peptidase database.**

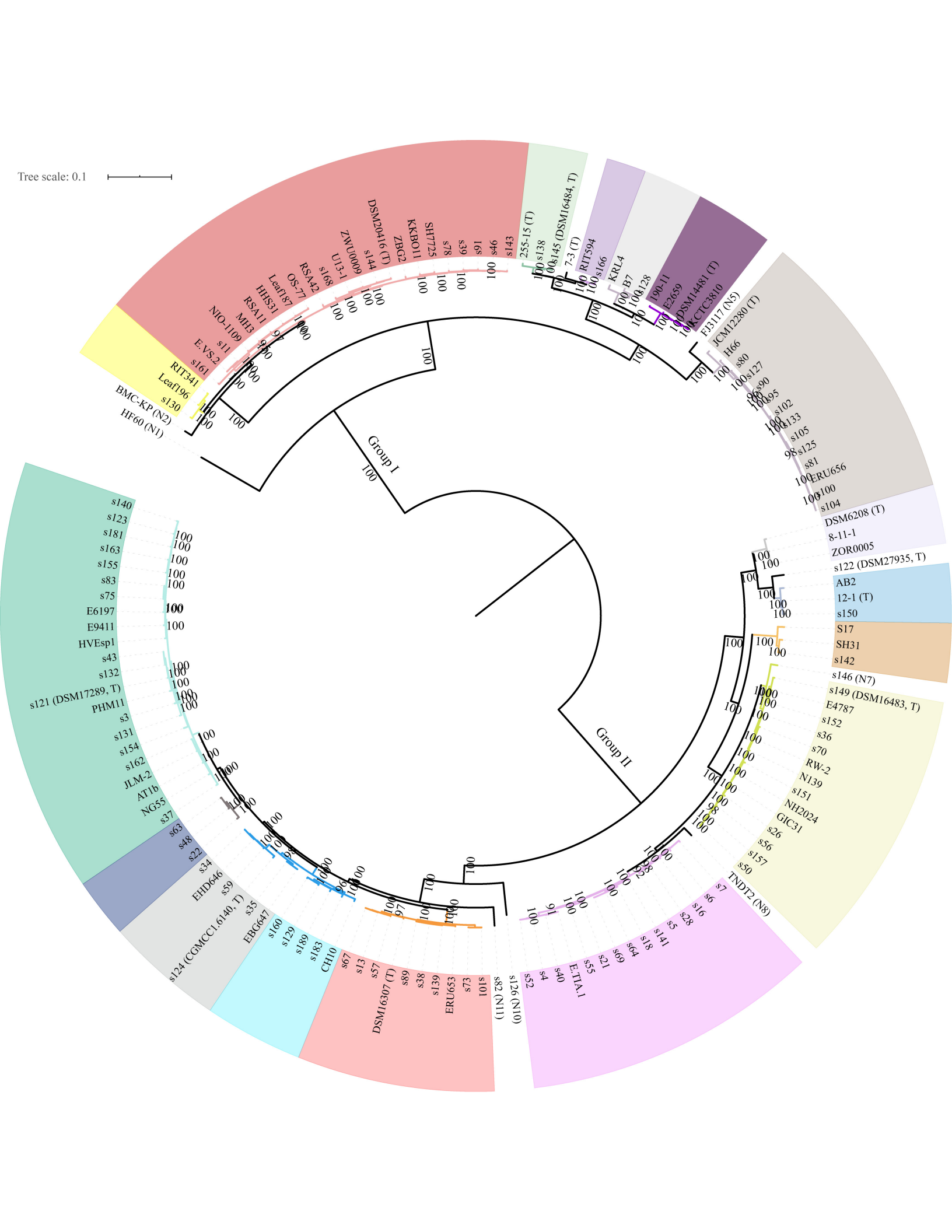
726 **Table S8 Genetic basis for *Exiguobacterium* strains adaptation to diverse**
727 **habitats.**

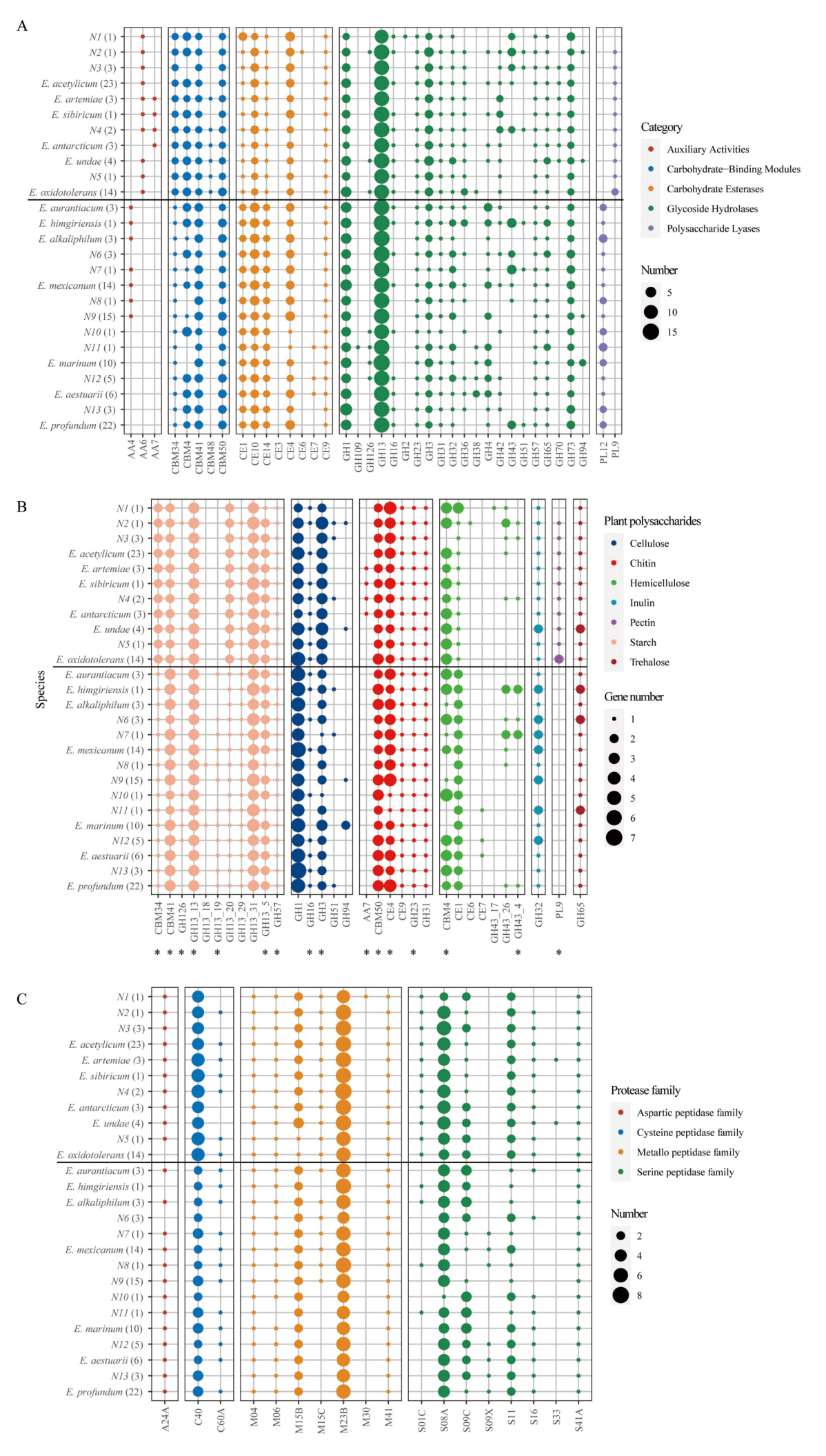
728 **Table S9 Transporter analysis of 145 *Exiguobacterium* genomes.**

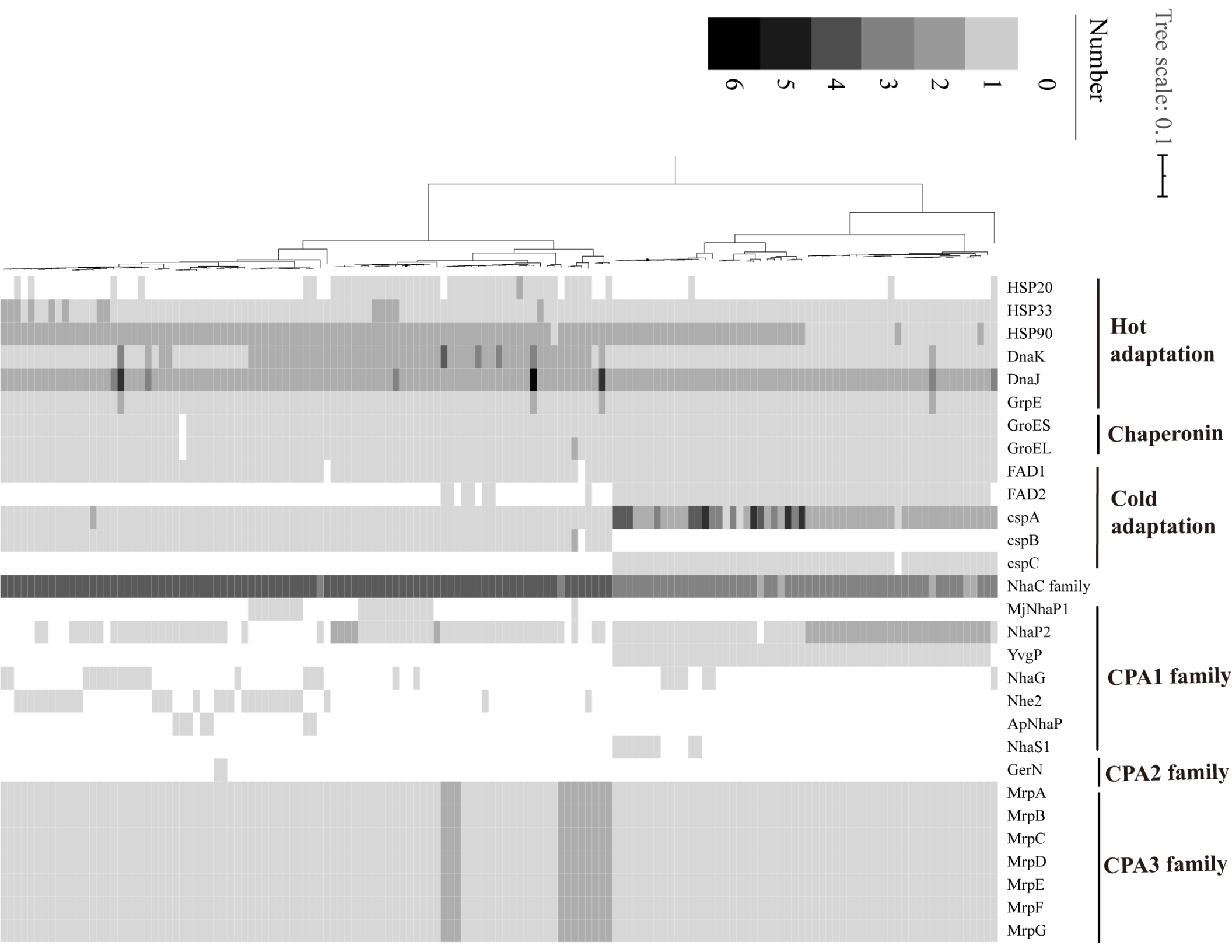
729

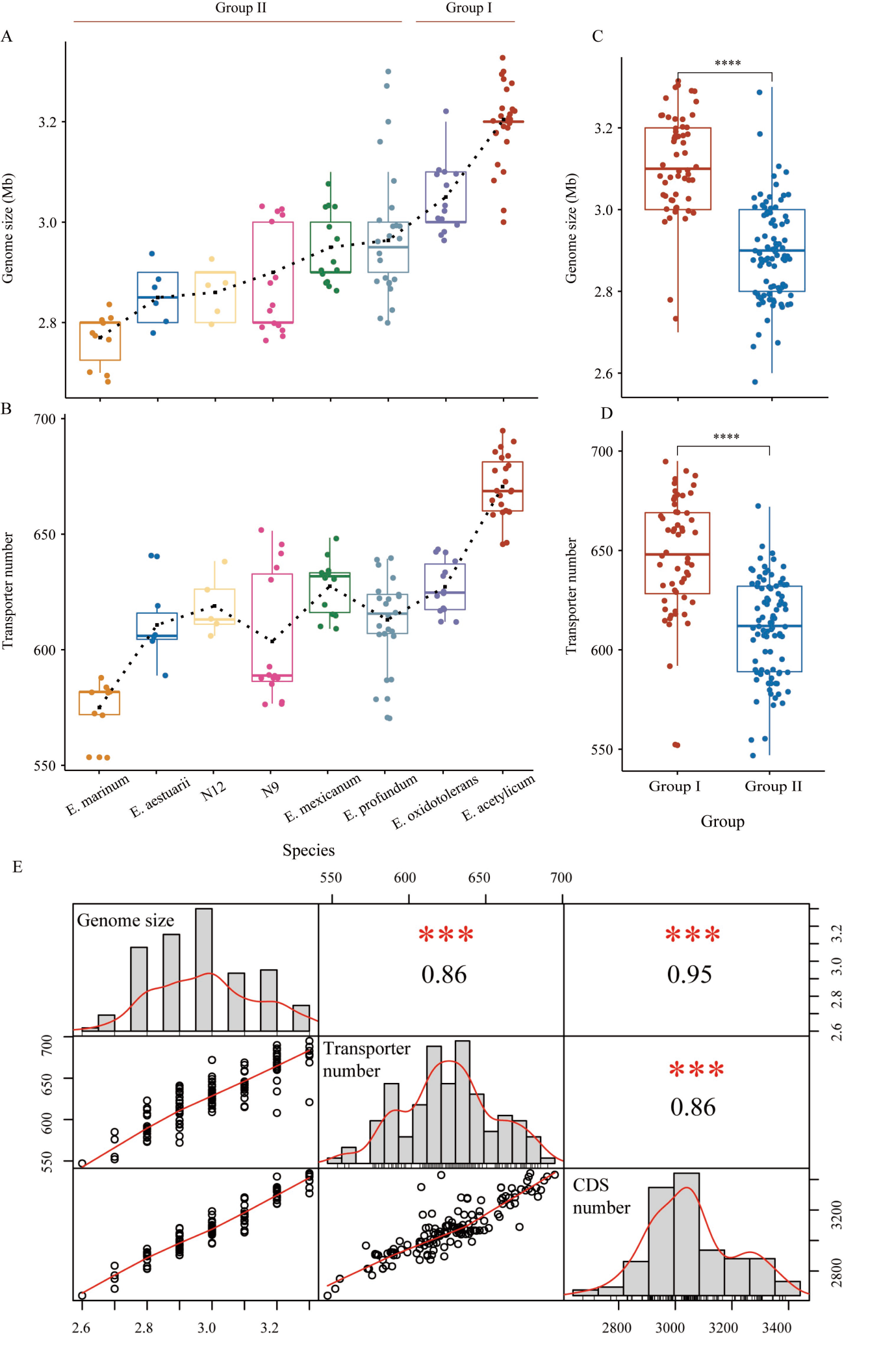
730

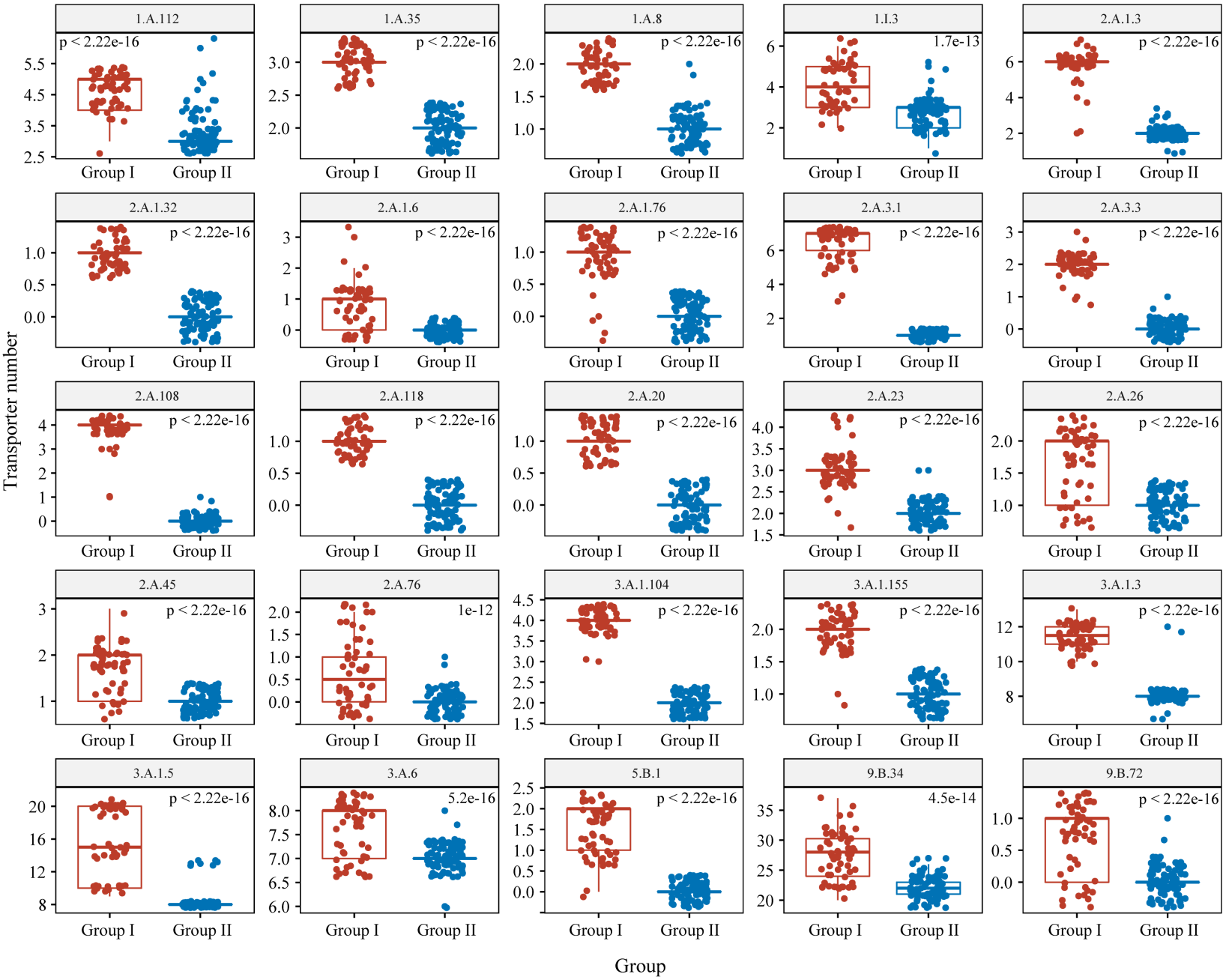

A
Estuary-associated environments
n=16
Hypersaline environments
n=34
Hot-associated environments
n=38
Cold-associated environments
n=84
Rhizosphere
n=88
Air
n=124
Soil
n=176
Contaminated water or soil
n=181
Animal-associated environments
n=275
Freshwater or freshwater sediment
n=328
Marine-associated environments
n=330
Plant
n=341
Other non-marine environments
n=567
0%
5%
10%
15%
20%
25%
Relative abundance (%)
C
pH
Salinity (%)
Temperature (°C)
Percentage
60%
40%
20%
0%
4–11
5–11
7–11
≤6
6<S≤8
8<S≤10
10<S≤12
12<S≤15
>4–45
>4–50
4–35
4–40
4–45
4–50
Conditions
B
120°W
60°W
0°
60°E
120°E
60°N
30°N
0°
30°S
60°S
N
W
E
S
0
7,000
km


Tree scale: 0.1
Group I
Group II
HF60 (N1)
BMC-KP (N2)
Leaf196
RIT341
s130
s161
E.VS.2
s11
NIO-1109
MH3
RSA11
HHS31
Leaf187
OS-77
RSA42
s168
U13-1
ZWU0009
s144
DSM20416 (T)
ZBG2
KKBO11
SH7725
s78
s39
s91
s46
s143
255-15 (T)
s138
s145 (DSM16484, T)
7-3 (T)
RIT594
s166
KRL4
B7
s128
190-11
E2659
DSM14481 (T)
KCTC3810
FJ3117 (N5)
JCM12280 (T)
H66
s80
s127
s90
s95
s102
s133
s105
s125
s81
ERU656
s104
DSM6208 (T)
8-11-1
ZOR0005
s122 (DSM27935, T)
AB2
12-1 (T)
s150
S17
SH31
s142
s146 (N7)
s149 (DSM16483, T)
E4787
s152
s36
s70
RW-2
N139
s151
NH2024
GIC31
s26
s56
s157
s50
TNDT2 (N8)
s7
s6
s16
s28
s5
s141
s18
s64
s69
s21
s55
E.TIA.1
s40
s4
s52
s126 (N10)
s82 (N11)
101
s73
ERU653
s139
s38
s89
DSM16307 (T)
s57
s13
s67
CH10
s183
s189
s129
s160
EBG647
s35
s124 (CGMCC1.6140, T)
s59
EHD646
s34
s22
s48
s63
s37
NG55
AT1b
JLM-2
s162
s154
s131
s3
PHM11
s121 (DSM17289, T)
s132
s43
HVEsp1
E9411
E6197
s75
s83
s155
s163
s181
s123
s140


A
N1 (1)
N2 (1)
N3 (3)
E. acetylicum (23)
E. artemiae (3)
E. sibiricum (1)
N4 (2)
E. antarcticum (3)
E. undae (4)
N5 (1)
E. oxidotolerans (14)
E. aurantiacum (3)
E. himgiriensis (1)
E. alkaliphilum (3)
N6 (3)
N7 (1)
E. mexicanum (14)
N8 (1)
N9 (15)
N10 (1)
N11 (1)
E. marinum (10)
N12 (5)
E. aestuarii (6)
N13 (3)
E. profundum (22)
AA4 AA6 AA7
CBM34 CBM4 CBM41 CBM48 CBM50
CE1 CE10 CE14 CE3 CE4 CE6 CE7 CE9
GH1 GH109 GH126 GH13 GH16 GH2 GH23 GH3 GH31 GH32 GH36 GH38 GH4 GH42 GH43 GH51 GH57 GH65 GH70 GH73 GH94
PL12 PL9
Category
Auxiliary Activities
Carbohydrate−Binding Modules
Carbohydrate Esterases
Glycoside Hydrolases
Polysaccharide Lyases
Number
5
10
15
B
Species
CBM34 CBM41 GH126 GH13_13 GH13_18 GH13_19 GH13_20 GH13_29 GH13_31 GH13_5 GH57
GH1 GH16 GH3 GH51 GH94
AA7 CBM50 CE4 CE9 GH23 GH31
CBM4 CE1 CE6 CE7 GH43_17 GH43_26 GH43_4
GH32
PL9
GH65
Plant polysaccharides
Cellulose
Chitin
Hemicellulose
Inulin
Pectin
Starch
Trehalose
Gene number
1
2
3
4
5
6
7
C
A24A C40 C60A
M04 M06 M15B M15C M23B M30 M41
S01C S08A S09C S09X S11 S16 S33 S41A
Protease family
Aspartic peptidase family
Cysteine peptidase family
Metallo peptidase family
Serine peptidase family
Number
2
4
6
8


Number
6
5
4
3
2
1
0
Tree scale: 0.1
HSP20
HSP33
HSP90
DnaK
DnaJ
GrpE
GroES
GroEL
FAD1
FAD2
cspA
cspB
cspC
NhaC family
MjNhaP1
NhaP2
YvgP
NhaG
Nhe2
ApNhaP
NhaS1
GerN
MrpA
MrpB
MrpC
MrpD
MrpE
MrpF
MrpG
Hot adaptation
Chaperonin
Cold adaptation
CPA1 family
CPA2 family
CPA3 family


Group II
Group I
A
Genome size (Mb)
3.2
3.0
2.8
B
Transporter number
700
650
600
550
E. marinum
E. aestuarii
N12
N9
E. mexicanum
E. profundum
E. oxidotolerans
E. acetylicum
Species
C
****
Genome size (Mb)
3.2
3.0
2.8
2.6
D
****
Transporter number
700
650
600
550
Group I
Group II
Group
E
550
600
650
700
Genome size
***
0.86
***
0.95
Transporter number
***
0.86
CDS number
700
650
600
550
3.2
3.0
2.8
2.6
3200
2800
2.6
2.8
3.0
3.2
2800
3000
3200
3400


1.A.112
p < 2.22e-16
1.A.35
p < 2.22e-16
1.A.8
p < 2.22e-16
1.I.3
1.7e-13
2.A.1.3
p < 2.22e-16
2.A.1.32
p < 2.22e-16
2.A.1.6
p < 2.22e-16
2.A.1.76
p < 2.22e-16
2.A.3.1
p < 2.22e-16
2.A.3.3
p < 2.22e-16
2.A.108
p < 2.22e-16
2.A.118
p < 2.22e-16
2.A.20
p < 2.22e-16
2.A.23
p < 2.22e-16
2.A.26
p < 2.22e-16
2.A.45
p < 2.22e-16
2.A.76
1e-12
3.A.1.104
p < 2.22e-16
3.A.1.155
p < 2.22e-16
3.A.1.3
p < 2.22e-16
3.A.1.5
p < 2.22e-16
3.A.6
5.2e-16
5.B.1
p < 2.22e-16
9.B.34
4.5e-14
9.B.72
p < 2.22e-16
Group I
Group II
Transporter number
Group