Mechanisms Driving Genome Reduction of a Novel *Roseobacter* Lineage Showing Vitamin B_{12} Auxotrophy

Xiaoyuan Feng[1], Xiao Chu[1], Yang Qian[1], Michael W. Henson[2a], V. Celeste Lanclos[2], Fang Qin[3], Yanlin Zhao[3], J. Cameron Thrash[2], Haiwei Luo[1*]

[1]Simon F. S. Li Marine Science Laboratory, School of Life Sciences and State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[2]Department of Biological Sciences, University of Southern California, Los Angeles, CA USA

[3]Fujian Provincial Key Laboratory of Agroecological Processing and Safety Monitoring, College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

[a] Current Affiliation: Department of Geophysical Sciences, University of Chicago, Chicago, Illinois, USA

[*]Corresponding author:
Haiwei Luo
School of Life Sciences, The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Phone: (+852) 3943-6121
Fax: (+852) 2603-5646
E-mail: hluo2006@gmail.com

Keywords: *Roseobacter*, CHUG, genome reduction, vitamin B_{12} auxotrophy

Summary

Members of the marine *Roseobacter* group are key players in the global carbon and sulfur cycles. While over 300 species have been described, only 2% possess reduced genomes (mostly 3-3.5 Mbp) compared to an average roseobacter (>4 Mbp). These taxonomic minorities are phylogenetically diverse but form a Pelagic *Roseobacter* Cluster (PRC) at the genome content level. Here, we cultivated eight isolates constituting a novel *Roseobacter* lineage which we named 'CHUG'. Metagenomic and metatranscriptomic read recruitment analyses showed that CHUG members were globally distributed and active in marine environments. CHUG members possess some of the smallest genomes (~2.52 Mb) among all known roseobacters, but they do not exhibit canonical features of genome streamlining like higher coding density or fewer paralogues and pseudogenes compared to their sister lineages. While CHUG members are clustered with traditional PRC members at the genome content level, they show important differences. Unlike other PRC members, neither the relative abundances of CHUG members nor their gene expression levels are correlated with chlorophyll a concentration across the global samples. Moreover, CHUG members cannot synthesize vitamin B_{12}, a key metabolite made by most roseobacters but not by many phytoplankton species and thus thought to mediate the roseobacter-phytoplankton interactions. This combination of features is evidence for the hypothesis that CHUG members may have evolved a free-living lifestyle decoupled from phytoplankton. This ecological transition was accompanied by the loss of signature genes involved in roseobacter-phytoplankton symbiosis, suggesting that relaxation of purifying selection is likely an important driver of genome reduction in CHUG.

Introduction

The marine *Roseobacter* group is a subfamily-level lineage in the *Alphaproteobacteria* and plays an important role in global carbon and sulfur cycling (1, 2). It is highly abundant in the coastal environments, accounting for up to 20% of all bacterial cells (3–5). Over 300 species and 100 genera have been described (6), the vast majority of which harbor large and variable genomes and grow readily on nutrient-rich solid media which are not representative of the niches found in the oligotrophic oceans. Early culture-independent 16S rRNA gene surveys showed that the oceanic roseobacters are represented by a few uncultivated lineages (1, 7). Recently, novel cultivation techniques and single-cell genomics have made available (partial) genome sequences of several previously uncultivated lineages including NAC11-7 (8), DC5-80-3 (also called RCA) (9, 10) and CHAB-I-5 (11, 12). Although these lineages are spottily distributed throughout the *Roseobacter* phylogeny, they together form a pelagic *Roseobacter* cluster (PRC). The PRC members consistently harbor smaller genomes and show more similar genome content compared to other roseobacters (11). Learning their evolutionary histories is essential to understand how the genetic and metabolic diversity of the pelagic *Roseobacter* lineages has formed, which in turn helps appreciate their roles in oceanic carbon and sulfur cycles. However, most PRC members form orphan branches and lack closely related reference genomes, which hampers our further understanding of their evolutionary trajectories.

Here, we isolated eight closely related roseobacters from several ocean regions that consistently possessed some of the smallest genomes (~2.52 Mb) among all known roseobacters. They together formed a novel *Roseobacter* lineage which we named ‘CHUG’ (Clade Hidden and Underappreciated Globally) that was abundant and active in global oceans. Unlike other PRC members, the global distribution of CHUG members was uncorrelated with chlorophyll a (Chl-a) concentration and they cannot *de novo* synthesize vitamin B_{12}, which is

often the metabolite roseobacters supply to phytoplankton during their symbiosis (2, 13–15). Therefore, the reductive evolution of CHUG may also indicate a dissociation with phytoplankton, a feature so far unique to CHUG among pelagic roseobacters.

Materials and Methods

Detailed methods are described in Supplementary Text 1. Briefly, samples were collected from surface water of the South China Sea, the East China Sea and the northern Gulf of Mexico. Over 20 CHUG isolates were retrieved following different dilution cultivation procedures, and genomes of eight isolates from the three ocean regions were sequenced with Illumina platforms, assembled with SPAdes (16) and annotated with Prokka (17). Among these, the isolate HKCCA1288 was further sequenced with PacBio Sequel platform to obtain a complete and closed genome. The average nucleotide identity (ANI) between genomes was calculated using fastANI (18). The assembled genome size, gene number, coding density and GC content of each genome were obtained using CheckM (19), whereas the estimated genome size was adjusted as (*assembled genome size*)/(*completeness* + *contamination*) (20). Pseudogenes were predicted following our recent study (21), and other genomic features were summarized using custom scripts (https://github.com/luolab-cuhk/CHUG-genome-reduction-project). The phylogenetic ANOVA analyses were performed to compare the analyzed genomic traits while controlling for the evolutionary history of those traits using the 'phylANOVA' function of the 'phytools' R package (22).

The *TARA* Ocean metagenomic and metatranscriptomic sequencing data with size fractions up to 3 μm (prokaryote-enriched) (23, 24) and metagenomic sequencing data with size fraction of 5-20 μm (nanoplankton-enriched) (25) were mapped to all 79 roseobacters studied here using bowtie (26) and BLASTN (27). Only reads sharing >95% similarity

and >80% alignment to their best hit were kept for relative abundance and activity calculation. The correlation analysis was performed using the 'rcorr' function in the 'Hmisc' R package (28), and the significance level was adjusted using stringent Bonferroni correction for multiple comparisons.

The *Roseobacter* phylogeny was constructed based on 120 bacterial marker genes (29), and the reference *Roseobacter* genomes included in the phylogeny followed a previous study (30). The orthologous gene families were predicted with OrthoFinder (31), and a binary matrix of the presence and absence pattern of orthologous gene families were used to construct the genome content dendrogram. The gene copy number of each orthologous family was further used to estimate the ancestral genome content for CHUG, its sister group and the outgroup using BadiRate (32).

Results

The CHUG diversity

Eight strains constituting a novel lineage (Fig. 1A) within the *Roseobacter* group, which we named Clade Hidden and Underappreciated Globally (CHUG), were isolated from the coastal waters of the South China Sea, the East China Sea, and the northern Gulf of Mexico (Table S1). Their genomes shared ≥99.7% 16S rRNA gene identity and ≥93% whole-genome average nucleotide identity (ANI). The CHUG lineage further exhibited ≥98.2% 16S rRNA gene identity when sequences of a few uncultivated members were added (Fig. S1), which was comparable to other pelagic *Roseobacter* lineages (98% (10) for DC5-80-3 and 96% (33) or 98% (7) for CHAB-I-5). CHUG genomes had ≤96.5% 16S rRNA gene identity and ≤71% ANI compared to the sister group members (Fig. 1A). All CHUG isolates were sampled exclusively from pelagic environments, whereas members of their sister group and the outgroup inhabit highly diverse salty environments including pelagic ocean, saline lake, algal

127 culture and coastal sediment (Table S1).

128 We also constructed a dendrogram based on the presence/absence pattern of orthologous
129 gene families (Fig. 1B). Although not monophyletic in the phylogeny based on shared genes
130 (Fig. 1A), CHUG and seven other genomes from taxa previously sampled from pelagic
131 environments formed a coherent group called the Pelagic *Roseobacter* Cluster (PRC) (11).
132 One previously identified PRC member, *Roseobacter* sp. R2A57 (4.13 Mb), was not
133 affiliated with PRC in this study. To facilitate our analysis and discussion, we divided the 79
134 roseobacters used in the present study into five groups: CHUG (eight genomes), its sister
135 group (five genomes), the outgroup of CHUG and its sister group (six genomes), other PRC
136 members (seven genomes) and other reference roseobacters (53 genomes).

137

138 *Genomic features*

139 Among the eight CHUG genomes, one (HKCCA1288) was closed with 2.66 Mb and the
140 remaining draft genomes were nearly complete (≥98.5%) according to CheckM predictions
141 (Table S1). Among other roseobacter genomes under comparison, at least 17 genomes were
142 closed and the remaining ones were nearly complete (≥96.5%) (Table S1). Based on the
143 assembled genome sizes, CHUG members possessed much smaller genomes (2.52 ± 0.07 Mb,
144 Fig. 2A) than an average roseobacter (4.16 ± 0.68 Mb). Further, their genome sizes were
145 comparable to those of the NAC11-7 cluster represented by the strain HTCC2255 (estimated
146 complete size to be 2.34 Mb), which is a basal roseobacter with the smallest genome among
147 all known roseobacters (34). As in HTCC2255, no plasmids were found in the CHUG
148 genomes. However, the coding density of CHUG (91.7 ± 0.5%, Fig. 2B) showed no
149 significant difference with its sister group and the outgroup (90.7 ± 0.7% and 90.7 ± 0.6%,
150 respectively) based on the phylogenetic ANOVA analysis ($p > 0.05$, 'phylANOVA' in the
151 'phytools' R package; the same test used below unless stated otherwise), which performs a

simulation-based comparison while taking into account the influence of phylogeny on the trait evolution (22). CHUG genomes had a lower genomic GC content (55.4 ± 0.8%, Fig. 2C) compared to their sister group (63.5 ± 1.6%, $p < 0.05$), although no significant difference was identified compared to the outgroup (63.8 ± 2.6%). In terms of pseudogenes, the number (99 ± 24, Fig. 2D) and ratio (3.9 ± 0.9%, Fig. 2E) in CHUG members were not significantly different from those of the sister group (128 ± 51; 3.3 ± 1.1%) and outgroup (148 ± 37; 3.7 ± 0.9%). The seven other PRC members also had smaller genomes (3.26 ± 0.51 Mb, Fig. 2A) and a reduced GC content (49.6 ± 5.5%, Fig. 2C) compared to the 53 other reference roseobacters (genome size: 4.32 ± 0.64 Mb, GC content: 61.9 ± 4.1%; $p < 0.01$), but there was no significant difference between the two groups in terms of the coding density (90.4 ± 0.9% for seven PRC members versus 89.3 ± 1.5% for other roseobacters, Fig. 2B), or the number (108 ± 49 for seven PRC members versus 205 ± 134 for other roseobacters, Fig. 2D) and ratio of pseudogenes (3.2 ± 1.4% for seven PRC members versus 4.7 ± 2.4% for other roseobacters, Fig. 2E).

CHUG genomes showed increased use of carbon atoms per amino-acid-residue side chain (C-ARSC, 2.833 ± 0.005, Fig. 2F) compared to the sister group (2.799 ± 0.004, $p < 0.05$). However, no significant difference was found in CHUG members in the use of C-ARSC compared to the outgroup (2.803 ± 0.011), nor that of nitrogen atoms per amino-acid-residue side chain (N-ARSC, 0.345 ± 0.002, Fig. 2G) compared to the sister group (0.344 ± 0.008) or the outgroup (0.346 ± 0.006). Likewise, the seven other PRC genomes had significantly higher C-ARSC (2.879 ± 0.031, Fig. 2F) than the 53 other reference roseobacters (2.817 ± 0.026, $p < 0.01$), but there was no significant difference between their N-ARSC (0.336 ± 0.004 for seven PRC members versus 0.348 ± 0.009 for other roseobacters, Fig. 2G).

We further investigated the codon usage and amino acid usage patterns in these lineages.

177 The CHUG genomes tended to comprise codons with more adenine/thymine (A/T) and less
178 guanine/cytosine (G/C) for 11 amino acids compared to the sister group and for 12 amino
179 acids compared to the outgroup, respectively ($p < 0.05$, Fig. S2 and Supplementary Text 2.1).
180 Furthermore, CHUG members possessed a higher fraction of isoleucine and lysine in their
181 proteomes but a lower fraction of glycine, proline, valine and tryptophan when compared to
182 the sister group or outgroup ($p < 0.05$, Fig. S3), which may be partially affected by the
183 differences of nitrogen (N) use in their corresponding codons (Supplementary Text 2.1).
184 Consistent with their genome size differences, CHUG genomes contained a significantly
185 smaller number of coding genes (2,486 ± 78, Fig. 2H) compared to the outgroup (3,939 ± 214,
186 $p < 0.01$) and the seven other PRC genomes (3,253 ± 545 genes, $p < 0.05$). The CHUG
187 genomes contained 2,215 ± 70 orthologous gene families (Fig. 2I) with 1.12 ± 0.01 gene copy
188 per family (Fig. 2J). By comparison, the outgroup genomes contained 3,259 ± 130 families (p
189 < 0.01) orthologous gene families with 1.20 ± 0.04 ($p > 0.05$) gene copy per family, while the
190 seven other PRC genomes possessed 2,678 ± 398 families ($p > 0.05$) orthologous gene
191 families with 1.21 ± 0.05 ($p < 0.01$) gene copy per family. No significant difference occurred
192 between CHUG and the sister group (3,865 ± 591 genes, 3,197 ± 345 gene families and 1.20
193 ± 0.05 copy per family). Additionally, while the number of genes and number of gene copies
194 per family of the seven other PRC genomes was not significantly different from those in the
195 53 other reference roseobacters (4,199 ± 644 genes and 1.25 ± 0.12 copy per family, Fig.
196 2H,J), the seven other PRC genomes had fewer orthologous families compared to the 53
197 other reference roseobacters (3,362 ± 362, $p < 0.01$, Fig. 2I).
198
199 *Global distribution and ecological drivers*
200 We used recruitment analysis with the global-scale *TARA* Ocean metagenomic and
201 metatranscriptomic datasets with size fractions up to 3 µm (prokaryote-enriched) (23, 24) to

quantify the global distribution of CHUG and other PRC members. The eight CHUG members recruited 0.0005% and 0.0008% of all metagenomic (Fig. 3A) and metatranscriptomic (Fig. 3B) reads, respectively. The CHUG members appeared to be less abundant and less active than other PRC representatives such as *Rhodobacterales* bacterium HTCC2255 (NAC11-7), *Rhodobacteraceae* bacterium SB2 (CHAB-I-5) and *Planktomarina temperata* RCA23 (RCA or DC5-80-3) (Welch's t-test, $p < 0.01$ for each). A similar pattern was also found using *TARA* Ocean metagenomic sequencing data with the size fraction of 5-20 µm (nanoplankton-enriched; Fig. 3C) (25). The CHUG members further represented 1.165% of the total reads from the nutrient perturbation experiments in mesocosm situated in the Red Sea (Fig. S4A) (35), and they also showed seasonality, as they recruited 0.007%, 0.032% and 1.623% of the total reads sampled at Kwangyang bay ocean (36) in February, May and August 2015, respectively (Fig. S4B).

Next, we sought to identify the ecological factors that may drive the global distribution and activity of the CHUG members, and to compare it to seven other PRC members using the *TARA* Ocean metagenomic (Fig. 3D) and metatranscriptomic (Fig. 3E) samples. The relative abundance and activity of CHUG members and the PRC member *Rhodobacteraceae* bacterium HIMB11 were not correlated with other PRC members, chlorophyll a (Chl-a) concentration, or the total carbon (Fig. 3D,E; Bonferroni corrected $p < 0.05$). On the other hand, the relative abundances of other PRC members were mutually positively correlated with each other, with Chl-a concentration, and with total carbon in both metagenomic and metatranscriptomic samples (Fig. 3D,E; Bonferroni corrected $p < 0.05$). In addition, the activity of CHUG genomes was positively correlated with nitrate and depth (Fig. 3E; Bonferroni corrected $p < 0.05$). From a gene-centric perspective, 58.6% ± 1.2% and 88.3% ± 12.7% genes from the eight CHUG genomes and seven other PRC members recruited *TARA* metatranscriptomic reads, respectively. Among the most expressed gene families (top 5%),

many were housekeeping genes involved in transcription, translation, cell cycle, respiration, the tricarboxylic acid cycle (TCA) cycle, and the biosynthesis of amino acids, chaperones, cell wall, and capsule (Fig. 4). Both CHUG and other PRC members also had highly expressed genes for light utilization (e.g. the photosynthesis gene cluster or proteorhodopsin) and nutrient (e.g. carbohydrates and amino acid) transport. Additional highly expressed genes among CHUG members included those involved in zinc transport, the cytochrome *cbb*$_3$-type oxidase, acetate transporters, and genes for mercury homeostasis, among which the latter two were exclusively found in CHUG members (Fig. 4). On the other hand, some highly expressed orthologous gene families specific to the seven other PRC members were related to phosphonate transport and taurine degradation.

Genome reduction and vitamin B_{12} auxotrophy

Since CHUG has a well-supported sister group and outgroup (Fig. 1A), we reconstructed the gene gain and loss events that were associated with the origin of the CHUG cluster (Fig. 5A). The last common ancestor (LCA) of the CHUG cluster was estimated to have 2,320 genes, 2,134 orthologous gene families (1.09 gene copy per family), and a genome size of 2.35 Mb. There were 172 families (185 genes) gained and 406 families (425 genes) lost on the ancestral branch leading to the LCA of CHUG, while 28 and 52 families (30 and 79 genes) underwent copy number increase and decrease, respectively. Compared to its sister group and the outgroup, CHUG members lost 412 Kb (9.8%) on the ancestral branch leading to its LCA (filled triangle in Fig. 5A).

We further compared the metabolic potential between CHUG (Fig. 5A), the reconstructed ancestors (Fig. 5B), seven other PRC genomes (Fig. 5C), and other reference roseobacters (Table S2). Since the CHUG genomes experienced net DNA and gene losses, we explored whether metabolic auxotrophies (i.e., inability to synthesize a compound

252 required for the growth) arose as a result of these losses. Among the sequenced CHUG
253 members, the genome of the strain HKCCA1288 was closed, which improved our auxotrophy
254 inference. CHUG genomes harbored the complete pathways for the synthesis of all 20 amino
255 acids, many of which, such as the synthesis of lysine (*dapD*) and methionine (*metH* and
256 *ahcY*), were under active expression in the wild (Fig. 4). They further encoded the key genes
257 for thiamine (vitamin B_1) synthesis (thiamine-phosphate pyrophosphorylase, *thiE*) and
258 pyridoxine (vitamin B_6) synthesis (pyridoxamine 5'-phosphate oxidase, *pdxH*). Nevertheless,
259 the key gene for biotin (vitamin B_7) synthesis (biotin synthase, *bioB*) was not found in CHUG
260 nor in the sister group and the outgroup, suggesting that the biotin auxotrophy in CHUG was
261 not part of their net gene losses.

262 Intriguingly, CHUG was auxotrophic for cobalamin (vitamin B_{12}) biosynthesis, which
263 can be synthesized by most roseobacters (2). This was validated using a growth assay, in
264 which CHUG strain HKCCA1288 did not grow in the defined medium lacking vitamin B_{12}
265 but grew well with the supplement of vitamin B_{12} (Fig. 6A). As a comparison, the model
266 roseobacter *Ruegeria pomeroyi* DSS-3, which is equipped with the *cobG* route, grew equally
267 well in the presence or absence of vitamin B_{12} (Fig. 6B). Mapping of the vitamin B_{12} *de novo*
268 synthesis to the phylogeny (Fig. 1A) indicates that the loss of this capability was most likely
269 associated with the genome reduction leading to the LCA of the CHUG lineage. On the other
270 hand, no genome content changes were inferred related to vitamin B_{12} synthesis by the
271 ancestral genome reconstruction (Fig. 5B). This controversy can be ascribed to the fact that
272 the *de novo* synthesis of cobinamide has two non-homologous pathways (i.e., aerobic and
273 anaerobic synthesis of cobinamide, the key precursor of vitamin B_{12}, via key genes *cobG* and
274 *cbiX*, respectively), and distinct pathways are maintained in the CHUG sister lineages (Fig.
275 1A). The ancestral genome reconstruction further inferred that the loss of vitamin B_{12} *de novo*
276 synthesis capability was compensated with the coincidental gain of a putative vitamin B_{12}

277 transporter (Fig. 5B), which was absent from all other PRC members capable of *de novo*
278 vitamin B_{12} synthesis (Fig. 5C). Taken together, the loss of *de novo* synthesis capability and
279 the gain of a putative transporter indicates that CHUG may have to acquire vitamin B_{12} or its
280 precursor from the environment.
281
282 *Metabolic potential for community interaction*
283 Besides the loss of genes for *de novo* vitamin B_{12} synthesis, the CHUG members have
284 also lost genes for chemotaxis (*cheAB*) and flagellar assembly (*fliC*). These genes were
285 essential to mediate roseobacter-phytoplankton interactions (37), but may become
286 dispensable when switching to a planktonic lifestyle (38). Consistent with this, the
287 quorum-sensing (QS) system (*luxR*), type IV secretion system (*virB*), and type VI secretion
288 system (*vasKF*) involved in organismal interactions were rarely found in the CHUG genomes
289 (Fig. 5A). CHUG members also lost the gene cluster encoding gene transfer agent (GTA),
290 which resembles small double-stranded DNA (dsDNA) bacteriophages that increase
291 horizontal gene transfer (HGT) and metabolic flexibility at high population density (39).
292
293 *Metabolic potential for nutrient acquisition*
294 Nitrogen (N) is a primarily limiting nutrient in surface oceans (40). Genes encoding the
295 nitrogen regulatory protein P-II (*glnBD*) were highly expressed in the wild CHUG
296 populations (Fig. 4). Genes encoding the high-affinity ammonium transporter (*amtB*) and
297 nitrogen regulation two-component system (*ntrBC* and *ntrXY*) were found in the CHUG
298 genomes. Genes encoding urease (*ureABC*) were also identified in CHUG members, though
299 the urea transport system (*urtABCDE*) was not in any CHUG genomes. It is possible that urea
300 was assimilated via passive diffusion across the cell membrane in CHUG as shown in other
301 bacteria (41), or that urea was taken up by another promiscuous transporter. The genes

encoding the transporter for nitrate/nitrite assimilation (*nrtABC*) were also missing in CHUG genomes. CHUG members retained the genes for the spermidine/putrescine transporter (*potABCD* and ABC.SP) (Table S2), and the latter was among the most highly expressed genes in the oceanic CHUG populations (Fig. 4). However, the CHUG members did not carry genes for other polyamine transport systems (*oocMQT* for octopine/nopaline and *potFGHI* for putrescine). The CHUG also retained and highly expressed *aapJMPQ* for the general L-amino acid transporter (Fig. 4), but lost genes encoding the polar amino acid transport system ABC.PA, which was prevalent in all other roseobacters studied here. CHUG further had a reduced number of genes (only one copy) encoding the branched-chain amino acid transport system (*livFGHKM*) compared to its sister group (at least three copies), the outgroup (at least three copies) and other PRC members (at least two copies; Table S2). Overall, fewer genes involved in the acquisition of amino acids were found in CHUG (Table S2), but they may remain efficient due to the high expression level of the retained genes.

Phosphorus (P) is often a co-limiting nutrient in surface oceans (40). To deal with P limitation, the CHUG members may be assisted by the essential regulatory and metabolic pathways known to be induced by P-limitation including the two-component regulatory system (*phoBR*), the high-affinity phosphate transporter (*pstABCS*) for phosphate acquisition and the C-P lyase (*phnGHIJKLM*) for phosphonate utilization. However, they have lost the *phoX* encoding an alkaline phosphatase for phosphodiester utilization (42) during the genome reduction process (Fig. 5A,B). A notable evolutionary innovation upon the emergence of the CHUG lineage was a gain of the gene encoding phospholipase C (*plcP*) (Fig. 5A,B), which was missing from all the seven other PRC members (Fig. 5C). The *plcP* is the key gene of the pathway for phospholipid substitution with non-phospholipids in response to P starvation, and was prevalently found in marine bacterioplankton (43).

Metabolic potential for energy acquisition

Members of the CHUG cluster maintained some energy conservation strategies that are commonly found in other roseobacters. One example was the acquisition of light energy. The complete photosynthesis gene cluster underlying the aerobic anoxygenic photosynthesis (AAnP) were identified in all CHUG members, five of the seven other PRC genomes, and 21 of the 64 non-PRC genomes (Table S2). Other light energy acquisition mechanisms including genes encoding proteorhodopsin and xanthorhodopsin were only found in the PRC member HTCC2255 and in the two *Octadecabacter* genomes, respectively. Two marker genes (*pufAB*) of the photosynthesis gene cluster were among the most highly expressed genes in oceanic CHUG and other PRC members, and the proteorhodopsin in *Rhodobacterales* bacterium HTCC2255 was also highly expressed (Fig. 4). In total, the potential for light utilization was found in 14 of the 15 PRC members, but in only 23 of 64 non-PRC roseobacters (Table S2). The association of the light acquisition trait with the PRC members was significant, which remains when the biased phylogenetic distribution of this trait was under control as shown by the binaryPGLMM analysis ($p < 0.05$) (44). This result indicates that light utilization may facilitate their survival under nutrient-depleted pelagic environments (45, 46). However, it is not clear why the reduced CHUG genomes employ the photosynthesis gene cluster rather than a rhodopsin system for light acquisition, considering that the photosynthesis gene cluster consists of about 40 genes (46) whereas a rhodopsin system requires only the rhodopsin gene and an associated chromophore retinal gene (47). In fact, the possibility of an evolutionary replacement of photosynthesis gene cluster with proteorhodopsin remains open, because proteorhodopsin and photosynthesis gene cluster occur in two closely related ecotypes of DC5-80-3 (also called RCA), respectively (48), suggesting that the replacement of one phototrophic type with the other could happen rapidly.

Another example for energy conservation is the oxidation of reduced inorganic

352 compounds. The CHUG carried genes for the oxidation of carbon monoxide (CO) and
353 sulfide/thiosulfate as energy sources. Most roseobacters encode type II carbon monoxide
354 dehydrogenase (*codh*), but only those with type I CODH may perform CO oxidation (49).
355 Four of the eight CHUG genomes possessed type I CODH (*coxL*) and thus may oxidize CO
356 (Fig. 5). This gene was further identified in three genomes from its sister group, three
357 genomes from the outgroup, three PRC genomes and 18 other reference genomes (Table S2).
358 All CHUG members possessed the sulfide:quinone oxidoreductase (*sqr*) for the oxidation of
359 sulfide to zero valence sulfur (S^o) (50), the persulfide dioxygenase (*pdo*) for the oxidation of
360 S^o to sulfite which could spontaneous react with S^o to generate thiosulfate (50), and the
361 complete *sox* pathway for the oxidation of thiosulfate to sulfate (51) (Fig. 5). The *sqr* and *pdo*
362 were also found in four other PRC genomes and 32 of the 64 non-PRC genomes, while the
363 *sox* pathway was found in all seven PRC genomes and 42 non-PRC genomes (Table S2).
364 Unlike the capability of light utilization, no uneven distribution was identified for *coxL*, *sox*,
365 *sqr* and *pdo* between PRC and non-PRC roseobacters (χ^2 test for *coxL* and binaryPGLMM
366 analysis for the remaining genes; $p > 0.05$).
367 CHUG cannot perform nitrate/nitrite reduction for energy conservation due to the lack of
368 genes involved in nitrate reduction to nitrite (nitrate reductase, periplasmic *napAB* or
369 membrane-bound *narGHI*), nitrite reduction to ammonium (nitrite reductase, *nirBD*) or nitrite
370 reduction to nitric oxide (NO-forming nitrite reductase, copper-containing *nirK* or
371 haem-containing *nirS*) (2). The *narGHI* and *nirBD* were identified in some genomes
372 affiliated with the sister group and the outgroup (Fig. 5A). These genes were also missing
373 from other PRC genomes, but were found in some reference *Roseobacter* genomes (Table
374 S2).
375

Other important metabolic pathways relevant to Roseobacter ecology

Among the major pathways for glycolysis, all CHUG members maintained the key gene encoding phosphogluconate dehydratase (*edd*) for the Entner-Doudoroff (ED) pathway, but none of them contained the key gene for phosphofructokinase (*pfk*) in the Embden-Meyerhof-Parnas (EMP) pathway (Fig. 5). Both pathways were prevalent in the sister group and the outgroup. Ancestral genome content reconstruction inferred that the EMP pathway was lost at the LCA of the CHUG lineage (Fig. 5B) as a result of genome reduction. Interestingly, the seven other PRC genomes held an identical pattern to CHUG, in which the ED pathway was universally preserved but the EMP pathway was missing. Although generating less ATP and NADH, the ED pathway can provide NADPH and may accompany increased resistance to oxidative stress compared with the EMP pathway (52). This likely confers an important benefit to these pelagic roseobacters inhabiting the surface ocean where reactive oxygen species (ROS) production is intensive (53). The catabolic product of the ED pathway, pyruvate, can be further degraded through the tricarboxylic acid cycle (TCA) cycle, the genes of which were highly expressed in environmental CHUG members (Fig. 4).

Many roseobacters can degrade aromatic compounds through the aerobic ring-cleaving pathways (54). All CHUG members harbored the protocatechuate ring cleavage pathway (protocatechuate 3,4-dioxygenase, *pcaGH*), which is one of the most common pathways for the degradation of monoaromatic compounds among roseobacters (55). However, they did not carry *paaABCDE* encoding ring-1,2-phenylacetyl-CoA epoxidase (key enzyme for the phenylacetate ring cleavage pathway) or *hmgA* encoding homogentisate 1,2-dioxygenase (key enzyme for the homogenisate ring cleavage pathway) (54). As these two pathways were inferred to be present in the LCA shared by CHUG and its sister group (filled circle in Fig. 5A), we hypothesize that their absence from CHUG resulted from genome reduction. All the three ring cleavage pathways were common in the seven other PRC genomes (Table S2).

Methylated compounds are important substrates for roseobacters (56). Briefly, the CHUG members possessed the metabolic potential to utilize dimethylsulfoniopropionate (DMSP) via both demethylation (DMSP demethylase, *dmdA*) and cleavage (*dddD* or *dddL*) pathway. However, genes encoding trimethylamine dehydrogenase (*tmd*) and trimethylamine monooxygenase (*tmm*) involved in trimethylamine N-oxide (TMAO) and trimethylamine (TMA) degradation, respectively, were not identified in the CHUG genomes, nor in most genomes affiliated with their sister group and the outgroup. However, these genes were identified in some other PRC members. Genes involved in taurine transport (*tauABC*) and degradation (*xsc*) were not found in CHUG members, but they were present, and the latter was highly expressed, in seven other PRC members (Fig. 4).

Discussion

The CHUG population dynamics are uncoupled from phytoplankton abundance

Though the novel lineage CHUG and the previously known Pelagic *Roseobacter* Cluster (PRC) members all reach high global abundance and activity, the ecological factors driving their global distribution are different. DC5-80-3 and NAC11-7 abundances were previously shown to be positively correlated with phytoplankton blooms (1, 4, 57–60) and their abundance and activity were both found to be significantly correlated with Chl-a abundance here (Fig. 3D,E). In the PRC lineage CHAB-I-5, a few members carry signature genes mediating organismal interactions (e.g., type VI secretion system and quorum sensing) (12), and thus may also explore microenvironments such as phytoplankton and organic particles. In fact, CHAB-I-5 abundance and activity was also positively correlated with Chl-a across the global ocean samples (Fig. 3D,E), though such a correlation was not found in a previous study with a more limited sampling effort (11). In the case of CHUG, no significant correlation with Chl-a was identified (Fig. 3D,E). Indeed, when the *TARA Ocean*

426 metagenomic sequencing reads at the nanoplankton-enriched size fraction (5-20 μm) were
427 recruited, CHUG members exhibited a lower relative abundance than the other PRC
428 representatives by approximately one order of magnitude (Fig. 3C). Together, these data
429 support our hypothesis that members of the CHUG lineage evolved a free-living lifestyle
430 decoupled from phytoplankton.

431 The possible contrasting roles of CHUG versus other pelagic roseobacters in relationship
432 to phytoplankton were further supported by the absence of the *de novo* vitamin B_{12} synthesis
433 in all CHUG members but its presence in all other PRC members. The auxotrophy for
434 vitamin B_{12} was also validated for HKCCA1288 - for which we generated a complete
435 genome sequence - in a growth assay (Fig. 6). The marine eukaryotic algae are predominantly
436 vitamin B_{12} auxotrophs (61), whereas most roseobacters have the potential to synthesize
437 vitamin B_{12} (2). This complementarity is one of the major mechanisms that facilitate
438 mutualistic interactions between roseobacters and phytoplankton (2, 13–15), which also helps
439 explain why roseobacters are often among the most abundant bacteria associated with marine
440 eukaryotic phytoplankton (62–64). The loss of vitamin B_{12} synthesis in CHUG is unusual
441 because members of the *Roseobacter* group are known as the dominant bacterial lineages
442 associated with marine phytoplankton groups (65) and their evolutionary history was likely
443 correlated with phytoplankton diversification (2, 66). They usually benefit from the fixed
444 carbon or other excretions released by phytoplankton and, in return, produce secondary
445 metabolites (e.g. vitamins, indole-3-acetic acid) to promote phytoplankton growth (15, 67,
446 68). These interactions likely occur in microzones immediately surrounding phytoplankton
447 cells, which may create gene flow barriers and facilitate population differentiation of
448 associated roseobacters (69). Therefore, the ecology and evolution of the *Roseobacter* group
449 in the pelagic ocean are generally shaped by marine phytoplankton, making the possible
450 separation from this ecological pattern in the CHUG lineage unique.

Potential evolutionary forces driving genome reduction of the CHUG roseobacters

The most abundant marine bacterioplankton, such as the *Pelagibacterales* (also called the SAR11 clade) in the Alphaproteobacteria and the *Prochlorococcus* in Cyanobacteria, are often equipped with very small genomes (70). The evolutionary mechanisms driving their genome reduction have been discussed extensively. Among these, positive selection for metabolic efficiency (i.e., 'genome streamlining') has been theorized as the dominant force driving their genome reduction (70, 71). Although CHUG members possessed smaller genomes and lower GC content compared to the sister group and the outgroup, they did not show other features of genome streamlining, such as higher coding density, fewer paralogues, or a lower proportion of pseudogenes (70, 72). Therefore, the genome reduction process of CHUG members did not meet the canonical definition of 'genome streamlining'.

Other important evidence against the genome streamlining explanation for CHUG genome reduction was from the genomic proxies for nutrient acquisition and saving strategies used by marine bacterioplankton. Among the selective factors that may drive bacterioplankton genome reduction in the pelagic ocean, N limitation is considered as the dominant factor (34, 70, 73, 74). Although the relative abundance of gene transcripts (but not the genes) in the wild CHUG populations was positively correlated with the nitrate concentration (Fig. 3E), which provides marginal evidence for a role of N limitation, other key evidence was missing. For example, we did not observe a reduced use of N in the amino acid sequences (approximated by N-ARSC) in CHUG compared to the sister group and the outgroup. Similar observation was used as evidence against the hypothesis that N limitation is a strong driver of genome streamlining in other marine bacterioplankton lineages (75, 76). A second potential ecological factor driving genome streamlining is P limitation (77), though this theory has been debated (78). Genome reduction likely leads to a sizable decrease in

476 cellular P requirement and thus may confer a competitive advantage in the P-limited marine
477 environments (79). Although a few important genes for P acquisition (*pst* for high-affinity
478 phosphate transporter and *phn* for C-P lyase) were retained during the CHUG genome
479 reduction and a gene encoding phospholipase C (*plcP*) responsible for cell membrane
480 phospholipid substitution for non-phosphorus lipids (43) was even acquired, the key P
481 scavenging gene encoding PhoX alkaline phosphatase was lost (Fig. 5). Therefore, available
482 evidence for either N or P limitation as a driver of CHUG genome reduction was
483 self-contradictory.

484 Because evidence for genome streamlining was weak in this lineage, we examined
485 neutral evolutionary forces as potential explanations for CHUG genome reduction. In fact,
486 neutral mechanisms have recently been considered to play an important role in driving
487 genome reduction of marine bacterioplankton lineages (80–82). Most of the prior studies
488 focused on *Prochlorococcus* (see references cited in the following paragraphs). While some
489 extended their discussions to *Pelagibacterales* (81, 83), knowledge on the evolutionary
490 mechanisms driving genome reduction of most other marine bacterioplankton lineages is
491 rather limited.

492 One potentially important neutral driver is random genetic drift due to a reduction of
493 effective population size (N_e). A previous study showed that the major genome reduction
494 event coincided with an accelerated rate of accumulating deleterious mutations in the early
495 evolution of *Prochlorococcus*, providing important evidence that genetic drift was likely the
496 primary mechanism of genome reduction in this lineage (81). Specifically, the power of
497 genetic drift (i.e., the inverse of N_e) of an ancestral lineage (e.g., the ancestral branch
498 underlying the ancient genomic events) can be approximated by the ratio of the radical
499 nonsynonymous nucleotide substitutions per radical nonsynonymous site (d_R) to the
500 conservative nonsynonymous nucleotide substitutions per conservative nonsynonymous site

(d_C) (81). Because the replacement by a physicochemically dissimilar amino acid (or radical change) is likely to be more deleterious than the replacement by a similar amino acid (or conservative change) (84, 85), the elevated d_R/d_C ratio is evidence for genetic drift acting to accumulate the deleterious type of mutations (i.e., the radical changes) in excess. In terms of the CHUG, the d_R/d_C ratio was not significantly elevated compared to its sister group (Fig. S5A) under two independent methods for biochemical classification of the 20 amino acids (Fig. S5B,C), suggesting that the deleterious type of mutations was not accumulated in excess at the ancestral branch leading to the LCA of the CHUG lineage (filled triangle in Fig. 5A). Since this ancestral branch corresponds to the time when major genome reduction occurred for CHUG, we can conclude that genetic drift was unlikely to be an important driver of CHUG genome reduction.

A second potentially important neutral driver of prokaryotic genome reduction is increased mutation rate, which was also proposed to explain *Prochlorococcus* genome reduction (86). Mathematical modeling predicts that not all auxiliary genes can be maintained by purifying selection when mutation rate is increased, and that an increase of 10 fold in mutation rate may lead to a 30% decrease in genome size (87). More recently, this hypothesis was supported with empirical data from comparative genomics analyses (82), though whether increased mutation rate is a truly important driver of prokaryotic genome reduction is debated (88). Given the potentially important role of increased mutation rate in driving prokaryotic genome reduction, determining the unbiased spontaneous mutation rate of the CHUG and the sister lineage using the mutation accumulation experiment followed by whole genome sequencing of the mutant lines becomes an urgent research need.

One more potentially important but rarely discussed neutral force leading to genome reduction is the loss of the genes that were important in the initial habitat but became dispensable after the bacteria switched to a new environment. This neutral loss mechanism,

termed relaxation of purifying selection, may also have contributed to genome reduction in *Prochlorococcus* (89). Importantly, the loss of dispensable genes under this mechanism is not related to the change of N_e but results instead from a change of habitat or lifestyle. Unlike other pelagic roseobacter members, CHUG members do not exhibit a correlative pattern between their global distributions and Chl-a (Fig. 3D,E), which can be used as a proxy for phytoplankton abundances (90). This is supported by evidence at the molecular and physiological level, in which the *de novo* synthesis of vitamin B_{12}, a fundamental metabolite roseobacters produce and supply to phytoplankton, was missing from all CHUG members but present in all other pelagic roseobacters (Fig. 1A, Fig. 6). Once the capability of *de novo* vitamin B_{12} synthesis was lost, the CHUG ancestor may have lost its ability to establish symbiosis with phytoplankton and subsequently undergone a major shift of its planktonic lifestyle, namely from phytoplankton-associated to free-living. Given that phytoplankton cell surfaces can be more densely populated compared to the bulk seawater (65), genes contributing to roseobacter-phytoplankton symbiosis (e.g., motility and chemotaxis), depending on population density and involved in interactions with other bacteria (e.g., quorum sensing, gene transfer agent), may have become dispensable during this transition (38). Indeed, the loss of these signature genes contributed to the genome reduction of CHUG (Fig. 5). We therefore propose that the relaxation of purifying selection may be one of the primary evolutionary forces leading to the major genome reduction of CHUG.

Data availability

Genomic sequences of the eight CHUG genomes are available at the NCBI GenBank database under the accession number PRJNA574877.

Code availability

The custom scripts used in this study are available in the online repository (https://github.com/luolab-cuhk/CHUG-genome-reduction-project).

Acknowledgments

This research was funded by the National Science Foundation of China (41776129), the Hong Kong Research Grants Council General Research Fund (14163917), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), and the Direct Grant of CUHK (4053257 & 3132809). The research was also supported by a Louisiana Board of Regents grant (LEQSF(2014-17)-RD-A-06) and a Simons Early Career Investigator in Marine Microbial Ecology and Evolution Award to JCT.

Conflict of Interest

The authors declare no competing interests concerning the submitted work.

References

1. Buchan A, González JM, Moran MA. Overview of the marine *Roseobacter* lineage. Appl Environ Microbiol 2005; 71(10):5665–77.

2. Luo H, Moran MA. Evolutionary ecology of the marine *Roseobacter* clade. Microbiol Mol Biol Rev 2014; 78(4):573–87.

3. Moran MA, Belas R, Schell MA, González JM, Sun F, Sun S et al. Ecological genomics of marine Roseobacters. Appl Environ Microbiol 2007; 73(14):4559–69.

4. Giebel H-A, Kalhoefer D, Lemke A, Thole S, Gahl-Janssen R, Simon M et al. Distribution of *Roseobacter* RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. ISME J 2011; 5(1):8–19.

5. Wemheuer B, Wemheuer F, Hollensteiner J, Meyer F-D, Voget S, Daniel R. The green

577 impact: bacterioplankton response toward a phytoplankton spring bloom in the southern
578 North Sea assessed by comparative metagenomic and metatranscriptomic approaches. Front
579 Microbiol 2015; 6:805.

580 6. Pujalte MJ, Lucena T, Ruvira MA, Arahal DR, Macián MC. The Family
581 *Rhodobacteraceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F,
582 editors. The Prokaryotes. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 439–512.

583 7. Buchan A, Hadden M, Suzuki MT. Development and application of quantitative-PCR tools
584 for subgroups of the *Roseobacter* clade. Appl Environ Microbiol 2009; 75(23):7542–7.

585 8. Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA. Comparing effective
586 population sizes of dominant marine alphaproteobacteria lineages. Environ Microbiol Rep
587 2014; 6(2):167–72.

588 9. Giebel H-A, Kalhoefer D, Gahl-Janssen R, Choo Y-J, Lee K, Cho J-C et al. *Planktomarina*
589 *temperata* gen. nov., sp. nov., belonging to the globally distributed RCA cluster of the marine
590 *Roseobacter* clade, isolated from the German Wadden Sea. Int J Syst Evol Microbiol 2013;
591 63(Pt 11):4207–17.

592 10. Voget S, Wemheuer B, Brinkhoff T, Vollmers J, Dietrich S, Giebel H-A et al. Adaptation
593 of an abundant *Roseobacter* RCA organism to pelagic systems revealed by genomic and
594 transcriptomic analyses. ISME J 2015; 9(2):371–84.

595 11. Billerbeck S, Wemheuer B, Voget S, Poehlein A, Giebel H-A, Brinkhoff T et al.
596 Biogeography and environmental genomics of the *Roseobacter*-affiliated pelagic CHAB-I-5
597 lineage. Nat Microbiol 2016; 1(7):16063.

598 12. Zhang Y, Sun Y, Jiao N, Stepanauskas R, Luo H. Ecological genomics of the
599 uncultivated marine *Roseobacter* lineage CHAB-I-5. Appl Environ Microbiol 2016;
600 82(7):2100–11.

601 13. Wagner-Döbler I, Ballhausen B, Berger M, Brinkhoff T, Buchholz I, Bunk B et al. The
602 complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a hitchhiker's
603 guide to life in the sea. ISME J 2010; 4(1):61–77.

604 14. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ et al. Cryptic carbon and
605 sulfur cycling between surface ocean plankton. Proc Natl Acad Sci U S A 2015;
606 112(2):453–7.

607 15. Cooper MB, Kazamia E, Helliwell KE, Kudahl UJ, Sayer A, Wheeler GL et al.

608 Cross-exchange of B-vitamins underpins a mutualistic interaction between *Ostreococcus*
609 *tauri* and *Dinoroseobacter shibae*. ISME J 2018.

610 16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al. SPAdes:
611 a new genome assembly algorithm and its applications to single-cell sequencing. J Comput
612 Biol 2012; 19(5):455–77.

613 17. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;
614 30(14):2068–9.

615 18. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput
616 ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun
617 2018; 9(1):5114.

618 19. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the
619 quality of microbial genomes recovered from isolates, single cells, and metagenomes.
620 Genome Res 2015; 25(7):1043–55.

621 20. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN et al.
622 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of
623 life. Nat Microbiol 2017; 2(11):1533–42.

624 21. Chu X, Li S, Wang S, Luo D, Luo H. Gene loss through pseudogenization contributes to
625 the ecological diversification of a generalist *Roseobacter* lineage. ISME J 2020.

626 22. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other
627 things). Methods in Ecology and Evolution 2012; 3(2):217–23.

628 23. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G et al. Ocean
629 plankton. Structure and function of the global ocean microbiome. Science 2015;
630 348(6237):1261359.

631 24. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M et al. Gene
632 expression changes and community turnover differentially shape the global ocean
633 metatranscriptome. Cell 2019; 179(5):1068-1083.e21.

634 25. Vargas C de, Audic S, Henry N, Decelle J, Mahé F, Logares R et al. Eukaryotic plankton
635 diversity in the sunlit ocean. Science 2015; 348(6237):1261605.

636 26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;
637 9(4):357–9.

638 27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search
639 tool. J Mol Biol 1990; 215(3):403–10.

640 28. Harrell Jr FE. Package 'Hmisc'. CRAN2018 2019; 2019:235–6.

641 29. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A et al. A
642 standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of
643 life. Nat Biotechnol 2018; 36(10):996–1004.

644 30. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M et al.
645 Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and
646 non-marine habitats. ISME J 2017; 11(6):1483–99.

647 31. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative
648 genomics. Genome Biol 2019; 20(1):238.

649 32. Librado P, Vieira FG, Rozas J. BadiRate: estimating family turnover rates by
650 likelihood-based methods. Bioinformatics 2012; 28(2):279–81.

651 33. Lekunberri I, Gasol JM, Acinas SG, Gómez-Consarnau L, Crespo BG, Casamayor EO et
652 al. The phylogenetic and ecological context of cultured and whole genome-sequenced
653 planktonic bacteria from the coastal NW Mediterranean Sea. Syst Appl Microbiol 2014;
654 37(3):216–28.

655 34. Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA. Evolutionary analysis of a
656 streamlined lineage of surface ocean *Roseobacters*. ISME J 2014; 8(7):1428–39.

657 35. Coello-Camba A, Diaz-Rua R, Duarte CM, Irigoien X, Pearman JK, Alam IS et al.
658 Picocyanobacteria community and cyanophage infection responses to nutrient enrichment in
659 a mesocosms experiment in oligotrophic waters. Front. Microbiol. 2020; 11.

660 36. Kim Y, Jeon J, Kwak MS, Kim GH, Koh I, Rho M. Photosynthetic functions of
661 *Synechococcus* in the ocean microbiomes of diverse salinity and seasons. PLoS ONE 2018;
662 13(1):e0190266.

663 37. Geng H, Belas R. Molecular mechanisms underlying *Roseobacter*phytoplankton
664 symbioses. Curr Opin Biotechnol 2010; 21(3):332–8.

665 38. Luo H, Moran MA. How do divergent ecological strategies emerge among marine
666 bacterioplankton lineages? Trends Microbiol 2015; 23(9):577–84.

667 39. Biers EJ, Wang K, Pennington C, Belas R, Chen F, Moran MA. Occurrence and

668 expression of gene transfer agent genes in marine bacterioplankton. Appl Environ Microbiol
669 2008; 74(10):2933–9.

670 40. Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW et al. Processes
671 and patterns of oceanic nutrient limitation. Nature Geosci 2013; 6(9):701–10.

672 41. Veaudor T, Cassier-Chauvat C, Chauvat F. Genomics of urea transport and catabolism in
673 Cyanobacteria: biotechnological implications. Front Microbiol 2019; 10:2052.

674 42. Luo H, Benner R, Long RA, Hu J. Subcellular localization of marine bacterial alkaline
675 phosphatases. Proc Natl Acad Sci U S A 2009; 106(50):21219–23.

676 43. Sebastián M, Smith AF, González JM, Fredricks HF, van Mooy B, Koblížek M et al.
677 Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus
678 deficiency. ISME J 2016; 10(4):968–78.

679 44. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
680 analyses in R. Bioinformatics 2019; 35(3):526–8.

681 45. Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M et al. Genomic and
682 functional adaptation in surface ocean planktonic prokaryotes. Nature 2010; 468(7320):60–6.

683 46. Brinkmann H, Göker M, Koblížek M, Wagner-Döbler I, Petersen J. Horizontal operon
684 transfer, plasmids, and the evolution of photosynthesis in *Rhodobacteraceae*. ISME J 2018;
685 12(8):1994–2010.

686 47. Pinhassi J, DeLong EF, Béjà O, González JM, Pedrós-Alió C. Marine Bacterial and
687 Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. Microbiol
688 Mol Biol Rev 2016; 80(4):929–54.

689 48. Sun Y, Zhang Y, Hollibaugh JT, Luo H. Ecotype diversification of an abundant
690 *Roseobacter* lineage. Environ Microbiol 2017; 19(4):1625–38.

691 49. Cunliffe M. Correlating carbon monoxide oxidation with cox genes in the abundant
692 Marine *Roseobacter* Clade. ISME J 2011; 5(4):685–91.

693 50. Xin Y, Liu H, Cui F, Liu H, Xun L. Recombinant *Escherichia coli* with sulfide:quinone
694 oxidoreductase and persulfide dioxygenase rapidly oxidises sulfide to sulfite and thiosulfate
695 via a new pathway. Environ Microbiol 2016; 18(12):5123–36.

696 51. Friedrich CG, Quentmeier A, Bardischewsky F, Rother D, Kraft R, Kostka S et al. Novel
697 genes coding for lithotrophic sulfur oxidation of *Paracoccus pantotrophus* GB17. J. Bacteriol.

2000; 182(17):4677–87.

52. Klingner A, Bartsch A, Dogs M, Wagner-Döbler I, Jahn D, Simon M et al. Large-Scale 13C flux profiling reveals conservation of the Entner-Doudoroff pathway as a glycolytic strategy among marine bacteria that use glucose. Appl Environ Microbiol 2015; 81(7):2408–22.

53. Diaz JM, Hansel CM, Voelker BM, Mendes CM, Andeer PF, Zhang T. Widespread production of extracellular superoxide by heterotrophic bacteria. Science 2013; 340(6137):1223–6.

54. Gulvik CA, Buchan A. Simultaneous catabolism of plant-derived aromatic compounds results in enhanced growth for members of the *Roseobacter* lineage. Appl Environ Microbiol 2013; 79(12):3716–23.

55. Alejandro-Marín CM, Bosch R, Nogales B. Comparative genomics of the protocatechuate branch of the β-ketoadipate pathway in the *Roseobacter* lineage. Marine Genomics 2014; 17:25–33.

56. Chen Y. Comparative genomics of methylated amine utilization by marine *Roseobacter* clade bacteria and development of functional gene markers (tmm, gmaS). Environ Microbiol 2012; 14(9):2308–22.

57. Wagner-Döbler I, Biebl H. Environmental biology of the marine *Roseobacter* lineage. Annu Rev Microbiol 2006; 60:255–80.

58. West NJ, Obernosterer I, Zemb O, Lebaron P. Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. Environ Microbiol 2008; 10(3):738–56.

59. Rich VI, Pham VD, Eppley J, Shi Y, DeLong EF. Time-series analyses of Monterey Bay coastal microbial picoplankton using a ‘genome proxy’ microarray. Environ Microbiol 2011; 13(1):116–34.

60. Landa M, Blain S, Christaki U, Monchy S, Obernosterer I. Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. ISME J 2016; 10(1):39–50.

61. Helliwell KE. The roles of B vitamins in phytoplankton nutrition: new perspectives and prospects. New Phytol 2017; 216(1):62–8.

728 62. González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C et al.
729 Bacterial community structure associated with a dimethylsulfoniopropionate-producing North
730 Atlantic algal bloom. Appl Environ Microbiol 2000; 66(10):4237–46.

731 63. Amin SA, Parker MS, Armbrust EV. Interactions between diatoms and bacteria.
732 Microbiol Mol Biol Rev 2012; 76(3):667–84.

733 64. Li S, Chen M, Chen Y, Tong J, Wang L, Xu Y et al. Epibiotic bacterial community
734 composition in red-tide dinoflagellate *Akashiwo sanguinea* culture under various growth
735 conditions. FEMS Microbiol Ecol 2019; 95(5).

736 65. Seymour JR, Amin SA, Raina J-B, Stocker R. Zooming in on the phycosphere: the
737 ecological interface for phytoplankton-bacteria relationships. Nat Microbiol 2017; 2:17065.

738 66. Luo H, Csuros M, Hughes AL, Moran MA. Evolution of divergent life history strategies
739 in marine alphaproteobacteria. MBio 2013; 4(4).

740 67. Durham BP, Dearth SP, Sharma S, Amin SA, Smith CB, Campagna SR et al. Recognition
741 cascade and metabolite transfer in a marine bacteria-phytoplankton model system. Environ
742 Microbiol 2017:3500–13.

743 68. Shibl AA, Isaac A, Ochsenkühn MA, Cárdenas A, Fei C, Behringer G et al. Diatom
744 modulation of select bacteria through use of two unique secondary metabolites. Proc Natl
745 Acad Sci U S A 2020; 117(44):27445–55.

746 69. Wang X, Zhang Y, Ren M, Xia T, Chu X, Liu C et al. Cryptic speciation of a pelagic
747 *Roseobacter* population varying at a few thousand nucleotide sites. ISME J 2020.

748 70. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for
749 microbial ecology. ISME J 2014; 8(8):1553–65.

750 71. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D et al. Genome
751 streamlining in a cosmopolitan oceanic bacterium. Science 2005; 309(5738):1242–5.

752 72. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM et al.
753 Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the
754 surface ocean. Proc Natl Acad Sci U S A 2013; 110(28):11463–8.

755 73. Luo H, Thompson LR, Stingl U, Hughes AL. Selection maintains low genomic GC
756 content in marine SAR11 lineages. Mol Biol Evol 2015; 32(10):2738–48.

757 74. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM et al.

758 Environmental drivers of a microbial genomic transition zone in the ocean's interior. Nat
759 Microbiol 2017; 2(10):1367–73.

760 75. Grzymski JJ, Dussaq AM. The significance of nitrogen cost minimization in proteomes of
761 marine microorganisms. ISME J 2012; 6(1):71–80.

762 76. Lee MD, Ahlgren NA, Kling JD, Walworth NG, Rocap G, Saito MA et al. Marine
763 *Synechococcus* isolates representing globally abundant genomic lineages demonstrate a
764 unique evolutionary path of genome reduction without a decrease in GC content. Environ
765 Microbiol 2019; 21(5):1677–86.

766 77. Hessen DO, Jeyasingh PD, Neiman M, Weider LJ. Genome streamlining and the
767 elemental costs of growth. Trends Ecol Evol (Amst) 2010; 25(2):75–80.

768 78. Vieira-Silva S, Touchon M, Rocha EPC. No evidence for elemental-based streamlining of
769 prokaryotic genomes. Trends Ecol Evol (Amst) 2010; 25(6):319-20; author reply 320-1.

770 79. Thingstad T, Rassoulzadegan F. Conceptual models for the biogeochemical role of the
771 photic zone microbial food web, with particular reference to the Mediterranean Sea. Progress
772 in Oceanography 1999; 44(1-3):271–86.

773 80. Batut B, Knibbe C, Marais G, Daubin V. Reductive genome evolution at both ends of the
774 bacterial population size spectrum. Nat Rev Microbiol 2014; 12(12):841–50.

775 81. Luo H, Huang Y, Stepanauskas R, Tang J. Excess of non-conservative amino acid
776 changes in marine bacterioplankton lineages with reduced genomes. Nat Microbiol 2017;
777 2:17091.

778 82. Bourguignon T, Kinjo Y, Villa-Martín P, Coleman NV, Tang Q, Arab DA et al. Increased
779 mutation rate is linked to genome reduction in prokaryotes. Curr Biol 2020;
780 30(19):3848-3855.e4.

781 83. Viklund J, Ettema TJG, Andersson SGE. Independent genome reduction and
782 phylogenetic reclassification of the oceanic SAR11 clade. Mol Biol Evol 2012;
783 29(2):599–615.

784 84. Zuckerkandl E, Pauling L, Bryson V, Vogel HJ. Evolving genes and proteins. Science
785 1965:68–71.

786 85. Dayhoff MO. Atlas of protein sequence and structure. National Biomedical Research
787 Foundation; 1972.

788 86. Dufresne A, Garczarek L, Partensky F. Accelerated evolution associated with genome
789 reduction in a free-living prokaryote. Genome Biol 2005; 6(2):R14.

790 87. Marais GAB, Calteau A, Tenaillon O. Mutation rate and genome reduction in
791 endosymbiotic and free-living bacteria. Genetica 2008; 134(2):205–10.

792 88. Gu J, Wang X, Ma X, Sun Y, Xiao X, Luo H. Unexpectedly high mutation rate of a
793 deep-sea hyperthermophilic anaerobic archaeon. ISME J 2021; In press.

794 89. Luo H, Friedman R, Tang J, Hughes AL. Genome reduction by deletion of paralogs in the
795 marine cyanobacterium *Prochlorococcus*. Mol Biol Evol 2011; 28(10):2751–60.

796 90. Roesler C, Uitz J, Claustre H, Boss E, Xing X, Organelli E et al. Recommendations for
797 obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global
798 analysis of WET Labs ECO sensors. Limnol. Oceanogr. Methods 2017; 15(6):572–85.

799 91. Nguyen L-T, Schmidt HA, Haeseler A von, Minh BQ. IQ-TREE: a fast and effective
800 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015;
801 32(1):268–74.

802

803 **Figure legend**

804 **Fig. 1. (A)** Maximum likelihood phylogenomic tree showing the position of CHUG in
805 the *Roseobacter* group. The phylogeny was inferred using IQ-TREE (91) based on a
806 concatenation of 45,904 amino acid sites over 120 conserved bacterial proteins (29). Solid
807 circles in the phylogeny indicate nodes with bootstrap values >95%. The potential of aerobic
808 (key gene *cobG*, red) and anaerobic (key gene *cbiX*, green) cobinamide synthesis (the first
809 stage of Vitamin B_{12} synthesis) is labeled at the tips. Subclades of the *Roseobacter* group are
810 marked according to a recent study (30). **(B)** Dendrogram of the same *Roseobacter* genomes
811 based on the presence/absence pattern of orthologous gene families.

812 **Fig. 2.** Genomic feature comparisons between CHUG, their sister group, the outgroup,
813 seven other PRC members, and other reference roseobacters. The significance level in
814 genomic features between CHUG and the other four groups are shown in red, while that
815 between seven other PRC members and the remaining groups are shown in blue. The markers
816 * and ** denote $p < 0.05$ and $p < 0.01$ (phylANOVA analysis (22)), respectively.
817 Abbreviations: C-ARSC, carbon atoms per amino-acid-residue side chain; N-ARSC, nitrogen
818 atoms per amino-acid-residue side chain.

819 **Fig. 3.** The global distribution of CHUG and its ecological correlation with
820 environmental factors. **(A, B & C)** The relative abundance of CHUG and other PRC
821 members in the bacterial communities based on recruitment analysis using the metagenomic
822 *TARA* Ocean sequencing samples with size fractions up to 3 μm (A), and metatranscriptomic
823 sequencing samples with size fractions up to 3 μm (B), and metagenomic sequencing samples
824 with size fraction of 5-20 μm (C). **(D & E)** Correlation analysis between the relative
825 abundance of CHUG and other PRC members and environmental parameters measured in the
826 *TARA* Ocean metagenomic (D) and metatranscriptomic (E) samples. The *p* value is adjusted
827 using stringent Bonferroni correction. Nonsignificant correlations are indicated by crosses for

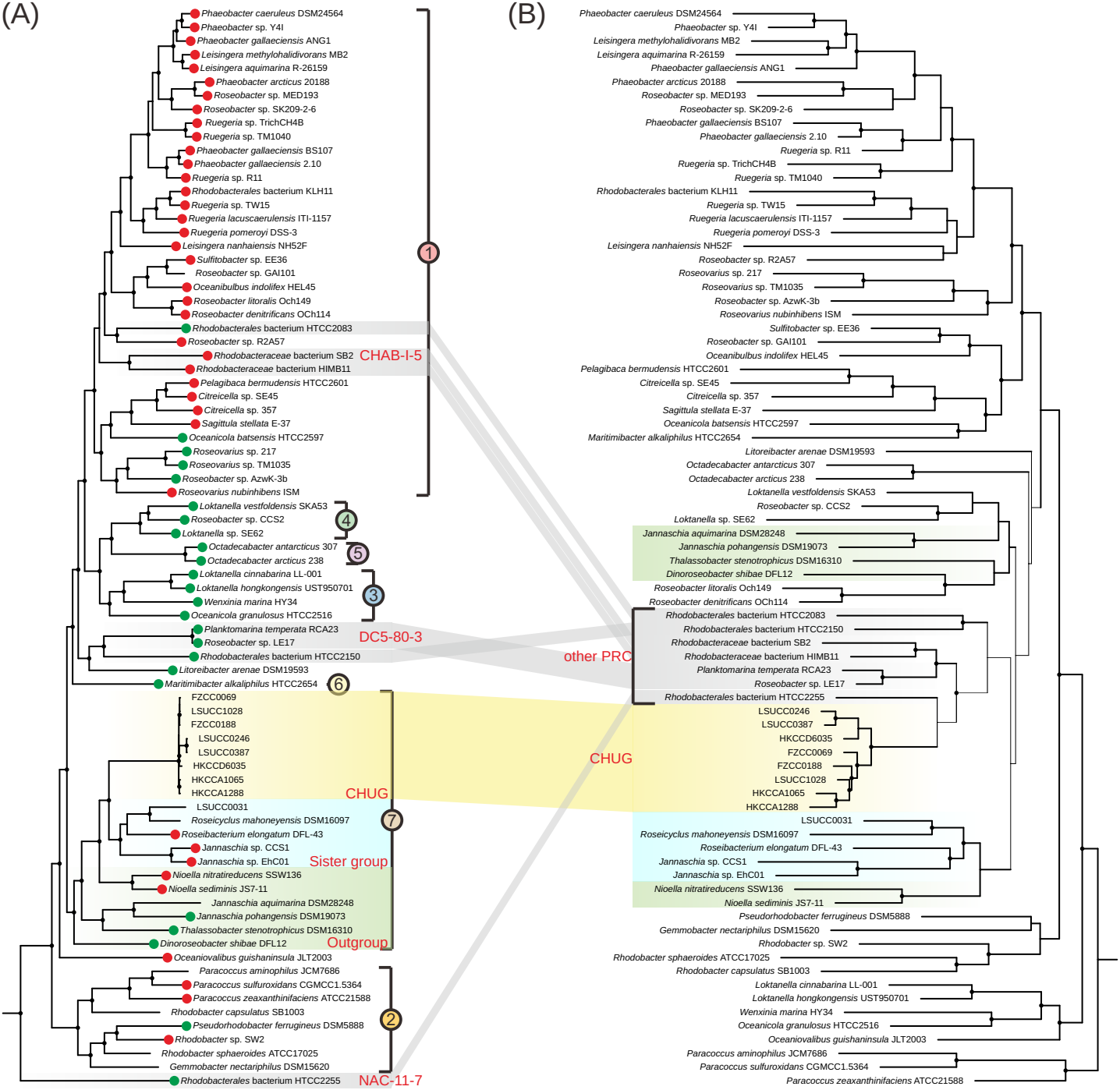
828 $p > 0.05$ after adjusting. Abbreviations: AO, Arctic Ocean; NAO, North Atlantic Ocean; SAO,
829 South Atlantic Ocean; IO, Indian Ocean; MS, Mediterranean Sea; NPO, North Pacific Ocean;
830 SPO, South Pacific Ocean; RS, Red Sea; SO, Southern Ocean; fCDOM, fluorescence,
831 colored dissolved organic matter.

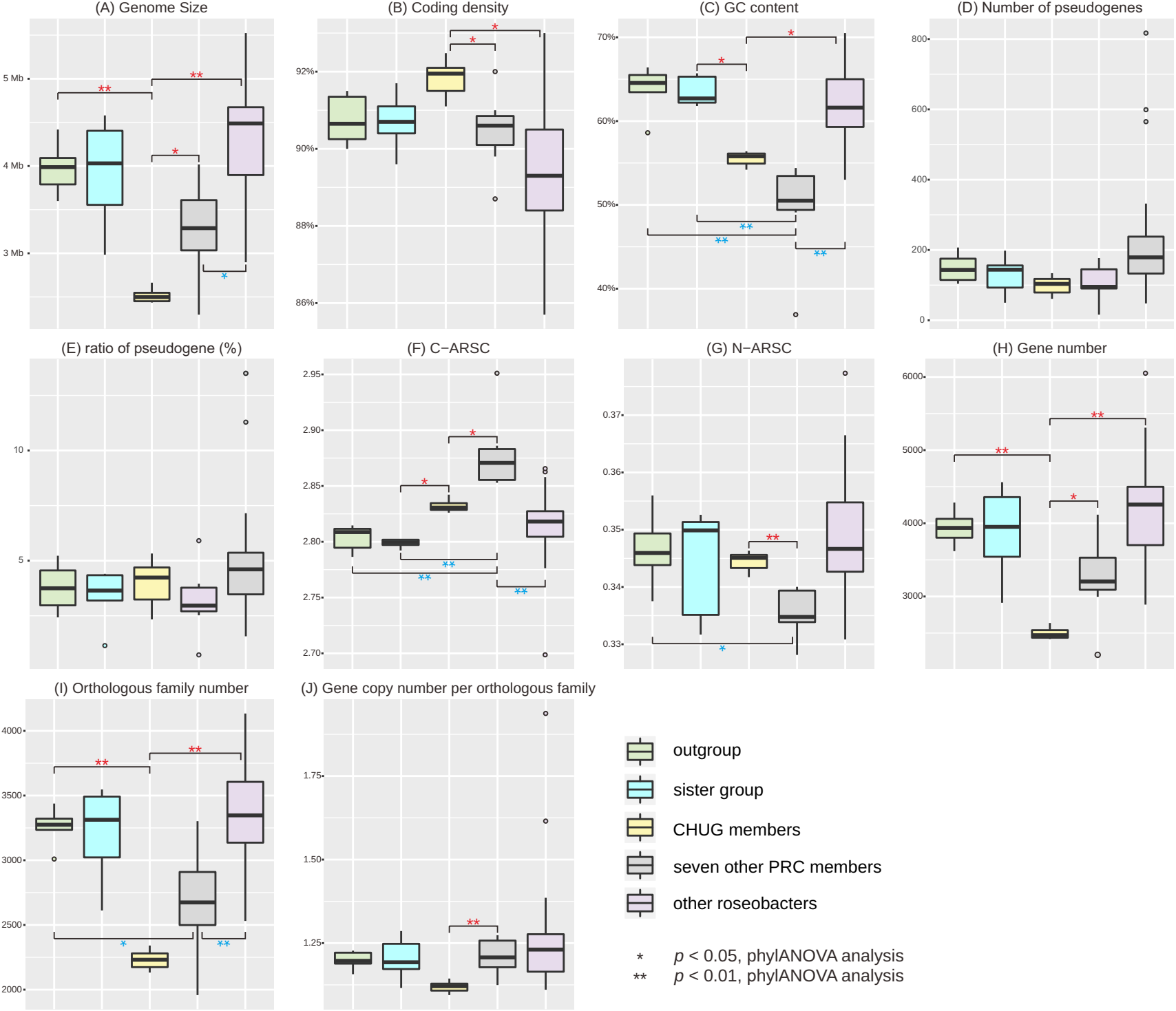
832 **Fig. 4.** The average expression level of gene families in CHUG and seven other PRC
833 members. Gene families with their gene expression level at top 5% found exclusively in
834 CHUG members, exclusively in seven other PRC members, and shared by CHUG and other
835 PRC members are shown in magenta, orange, and red dots, respectively. The remaining gene
836 families are shown in gray dots. Gene families specific to CHUG and seven other PRC
837 members are shown in the upper and right panel, respectively. Abbreviations: RPKM, Reads
838 Per Kilobase per Million mapped reads; *aapJ*, general L-amino acid transport system;
839 ABC.MS, multiple sugar transport system; ABC.PA, polar amino acid transport system;
840 ABC.SP, spermidine/putrescine transport system; *acnB*, aconitate hydratase 2; *actP*, acetate
841 permease; *ahcY*, adenosylhomocysteinase; ccoO, cytochrome cbb_3-type oxidase; *dapD*,
842 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase; *glnB*, nitrogen regulatory
843 protein P-II; *gmd*, GDPmannose 4,6-dehydratase; *icd*, isocitrate dehydrogenase; *ilvC*,
844 ketol-acid reductoisomerase; *katG*, catalase-peroxidase; *leuD*,
845 3-isopropylmalate/(R)-2-methylmalate dehydratase; *merA*, mercuric reductase; *merC*, *merP*
846 and *merT*, mercuric ion transport system; *metH*, 5-methyltetrahydrofolate--homocysteine
847 methyltransferase; *phnA*, phosphonoacetate hydrolase; *phnD*, phosphonate transport system;
848 *potD*, spermidine/putrescine transport system; *pufA* and *pufB*, light-harvesting complex 1;
849 *rbsB*, ribose transport system; *rhaS*, rhamnose transport system; *ureJ*, urease; *wza*,
850 polysaccharide biosynthesis/export protein; *xsc*, sulfoacetaldehyde acetyltransferase; *yaeT*,
851 Outer membrane protein assembly factor; *zntA*, lead, cadmium, zinc and mercury transporting
852 ATPase; *zupT*, zinc transporter.

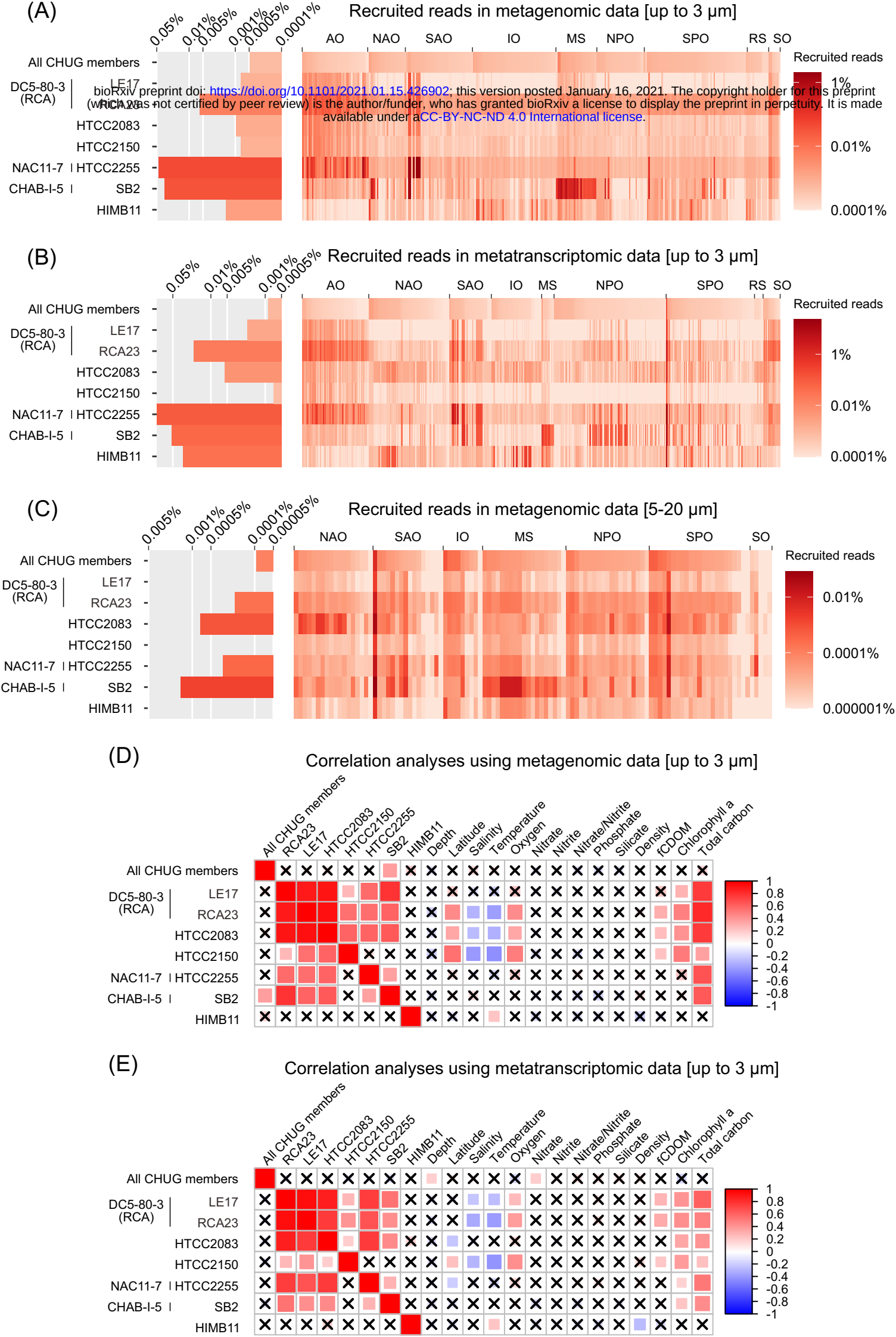
853 **Fig. 5.** The phyletic pattern of select genes. The solid and open circles in the right panel
854 represent the presence/absence of the genes, respectively. **(A)** The phyletic pattern in the
855 CHUG, its sister group and its outgroup. The phylogenomic tree shown in the left panel is
856 pruned from the full phylogenomic tree shown in Fig. 1A, and branch length is ignored for
857 better visualization. The ancestral genome reconstruction was performed with BadiRate (32).
858 Each ancestral and leaf node is associated with three numbers, representing the total number
859 of orthologous gene families at this node, and the number of orthologous gene families
860 gained and lost on the branch leading to this node. The LCA of CHUG, the LCA shared by
861 CHUG and its sister group, and the LCA shared by CHUG, its sister group and the outgroup
862 are marked with a filled triangle, a filled circle, and a filled star, respectively. **(B)** The
863 estimated phyletic pattern of the above-mentioned three LCAs. **(C)** The gene presence and
864 absence pattern in the CHUG and other seven PRC genomes. The dendrogram in the left
865 panel is pruned from that shown in Fig. 1B. Abbreviations: *thiE*, thiamine-phosphate
866 pyrophosphorylase; *pdxH*, pyridoxamine 5'-phosphate oxidase; *bioB*, biotin synthase; *cobG*,
867 precorrin-3B synthase; *cbiX*, sirohydrochlorin cobaltochelatase; *cobV*,
868 adenosylcobinamide-GDP ribazoletransferase; *btuB*, vitamin B12 transporter; *amtB*,
869 ammonium transport system; nitrogen regulatory protein P-II (*glnBD*); *ntrBC*, nitrogen
870 regulation two-component system; *ntrXY*, nitrogen regulation two-component system;
871 *ureABC*, urease; *urtABCDE*, urea transport system; *nrtABC*, nitrate/nitrite transport system;
872 *phoBR*, two-component phosphate regulatory system; *pstABCS*, phosphate transport system
873 (high affinity); *phnGHIJKLM*, carbon-phosphorus (C-P) lyase; *phoX*, alkaline phosphatase;
874 *plcP*, phospholipase C; PGC, photosynthesis gene cluster; *coxL*, carbon monoxide
875 dehydrogenase (type I forming); *sqr*, sulfide quinone oxidoreductase; *pdo*, persulfide
876 dioxygenase; *sox*, thiosulfate oxidizing SOX complex; *napAB*, nitrate reductase (periplasmic);
877 *narGHI*, nitrate reductase (membrane-bound); *nirBD*, nitrite reductase; *nirK*,

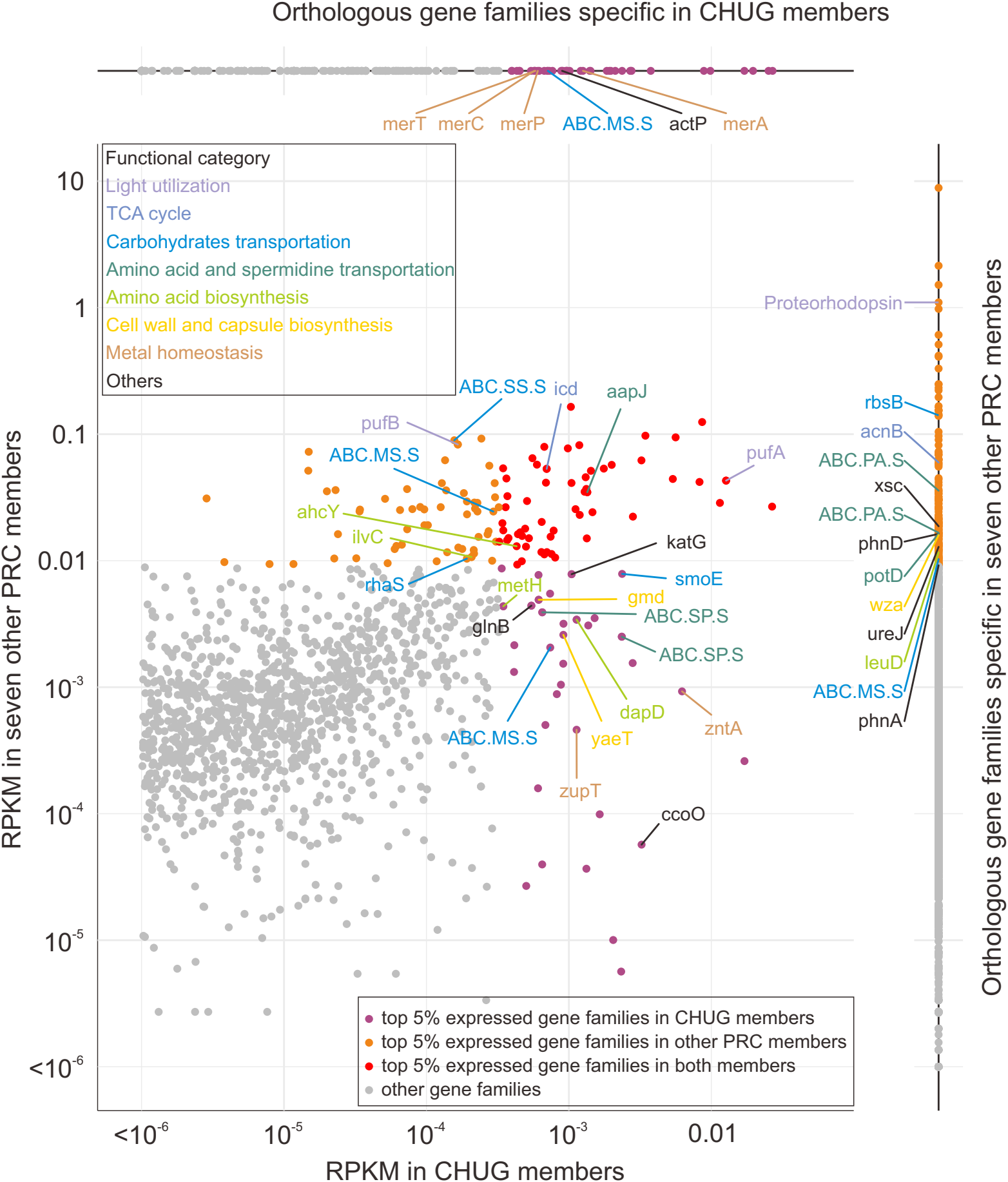
878 copper-containing NO-forming nitrite reductase; *nirS*, haem-containing NO-forming nitrite
879 reductase; *pfk*, phosphofructokinase; *edd*, phosphogluconate dehydratase; *pcaGH*,
880 protocatechuate 3,4-dioxygenase; *paaABCDE*, ring-1,2-phenylacetyl-CoA epoxidase; *hmgA*,
881 homogentisate 1,2-dioxygenase; *cheAB*, chemotaxis family protein; *fliC*, flagellin; *luxR*,
882 quorum-sensing system regulator; *virB*, type IV secretion system protein; vas*KF*, type VI
883 secretion system protein; GTA, gene transfer agent; *dmdA*, DMSP demethylase; *dddD*,
884 DMSP acyl-CoA transferase; *dddL*, dimethylpropiothetin dethiomethylase; *tmd*,
885 trimethylamine dehydrogenase; *tmm*, trimethylamine monooxygenase; *tauABC*, taurine
886 transport system; *xsc*, sulfoacetaldehyde acetyltransferase.

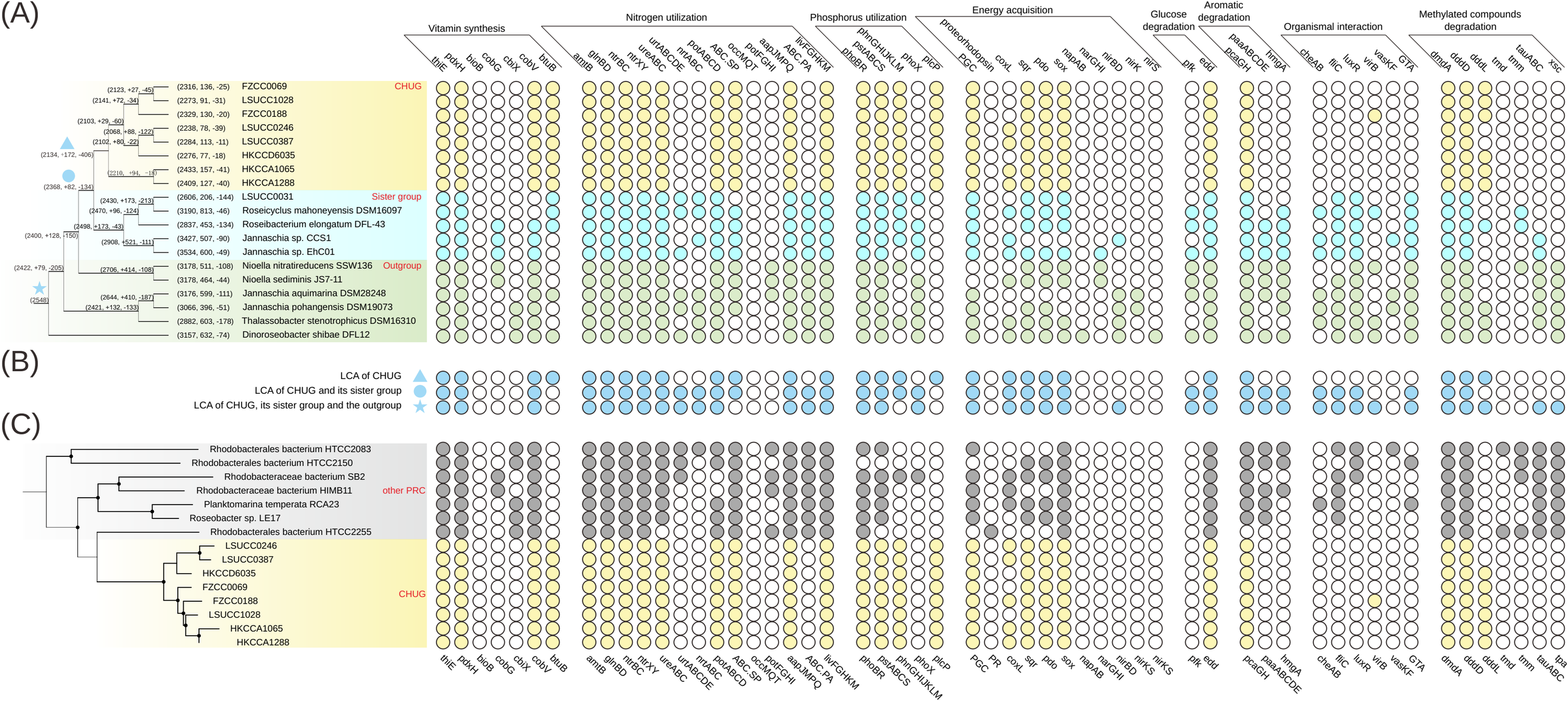
887 **Fig. 6.** Growth assay of (A) CHUG strain HKCCA1288 and (B) model roseobacter
888 *Ruegeria pomeroyi* DSS-3. Strains cultured on defined marine ammonium mineral salts
889 (MAMS) medium with and without vitamin B_{12} were plotted in red and blue, respectively.
890 Three triplicates were performed for each condition and error bars denote standard deviation.
891

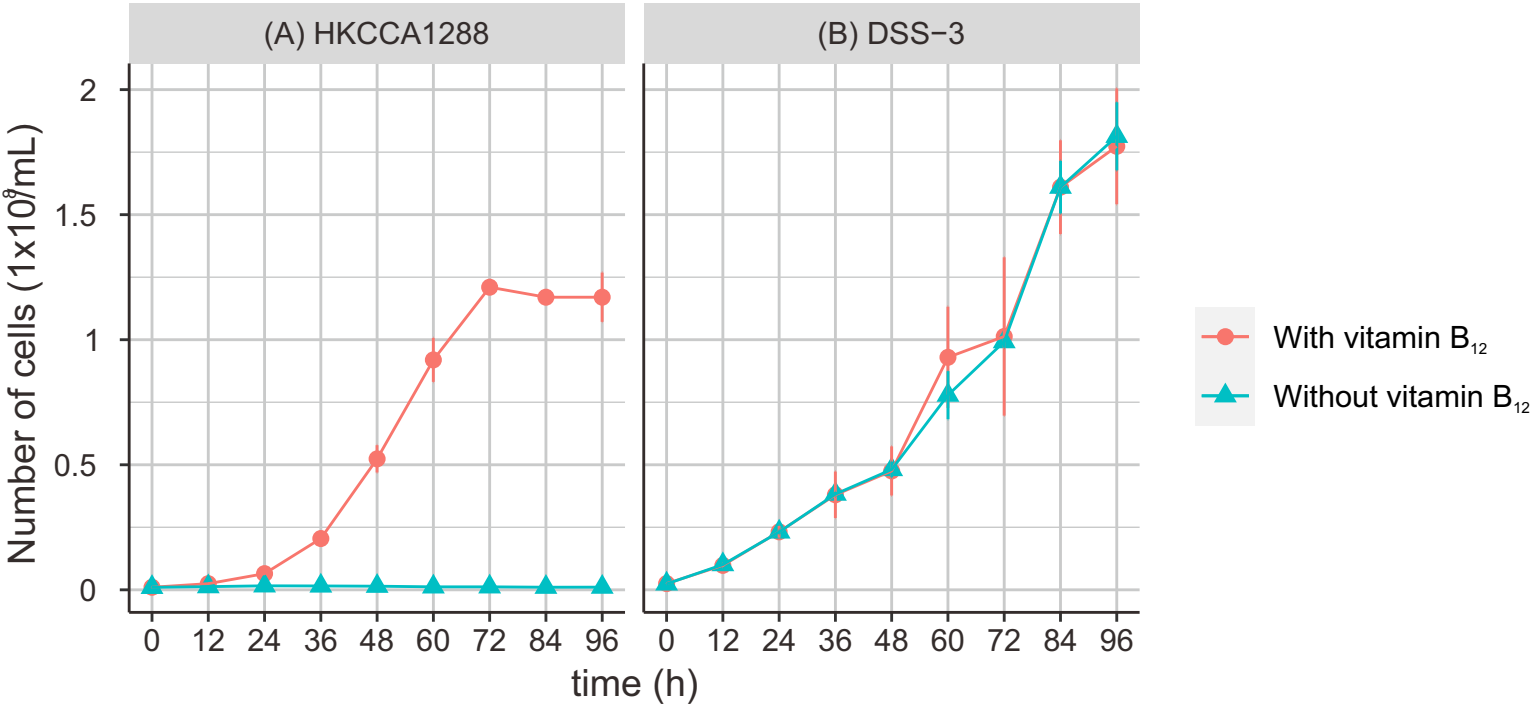

(A)
(B)
Phaeobacter caeruleus DSM24564
Phaeobacter sp. Y4I
Phaeobacter gallaeciensis ANG1
Leisingera methylohalidivorans MB2
Leisingera aquimarina R-26159
Phaeobacter arcticus 20188
Roseobacter sp. MED193
Roseobacter sp. SK209-2-6
Ruegeria sp. TrichCH4B
Ruegeria sp. TM1040
Phaeobacter gallaeciensis BS107
Phaeobacter gallaeciensis 2.10
Ruegeria sp. R11
Rhodobacterales bacterium KLH11
Ruegeria sp. TW15
Ruegeria lacuscaerulensis ITI-1157
Ruegeria pomeroyi DSS-3
Leisingera nanhaiensis NH52F
Sulfitobacter sp. EE36
Roseobacter sp. GAI101
Oceanibulbus indolifex HEL45
Roseobacter litoralis Och149
Roseobacter denitrificans OCh114
Rhodobacterales bacterium HTCC2083
Roseobacter sp. R2A57
Rhodobacteraceae bacterium SB2
Rhodobacteraceae bacterium HIMB11
Pelagibaca bermudensis HTCC2601
Citreicella sp. SE45
Citreicella sp. 357
Sagittula stellata E-37
Oceanicola batsensis HTCC2597
Roseovarius sp. 217
Roseovarius sp. TM1035
Roseobacter sp. AzwK-3b
Roseovarius nubinhibens ISM
Loktanella vestfoldensis SKA53
Roseobacter sp. CCS2
Loktanella sp. SE62
Octadecabacter antarcticus 307
Octadecabacter arcticus 238
Loktanella cinnabarina LL-001
Loktanella hongkongensis UST950701
Wenxinia marina HY34
Oceanicola granulosus HTCC2516
Planktomarina temperata RCA23
Roseobacter sp. LE17
Rhodobacterales bacterium HTCC2150
Litoreibacter arenae DSM19593
Maritimibacter alkaliphilus HTCC2654
FZCC0069
LSUCC1028
FZCC0188
LSUCC0246
LSUCC0387
HKCCD6035
HKCCA1065
HKCCA1288
LSUCC0031
Roseicyclus mahoneyensis DSM16097
Roseibacterium elongatum DFL-43
Jannaschia sp. CCS1
Jannaschia sp. EhC01
Nioella nitratireducens SSW136
Nioella sediminis JS7-11
Jannaschia aquimarina DSM28248
Jannaschia pohangensis DSM19073
Thalassobacter stenotrophicus DSM16310
Dinoroseobacter shibae DFL12
Oceaniovalibus guishaninsula JLT2003
Paracoccus aminophilus JCM7686
Paracoccus sulfuroxidans CGMCC1.5364
Paracoccus zeaxanthinifaciens ATCC21588
Rhodobacter capsulatus SB1003
Pseudorhodobacter ferrugineus DSM5888
Rhodobacter sp. SW2
Rhodobacter sphaeroides ATCC17025
Gemmobacter nectariphilus DSM15620
Rhodobacterales bacterium HTCC2255
Sister group
Outgroup
CHAB-I-5
DC5-80-3
CHUG
NAC-11-7
other PRC
1
2
3
4
5
6
7
Pseudorhodobacter ferrugineus DSM5888
Gemmobacter nectariphilus DSM15620
Rhodobacter sp. SW2
Rhodobacter sphaeroides ATCC17025
Rhodobacter capsulatus SB1003
Loktanella cinnabarina LL-001
Loktanella hongkongensis UST950701
Wenxinia marina HY34
Oceanicola granulosus HTCC2516
Oceaniovalibus guishaninsula JLT2003
Paracoccus aminophilus JCM7686
Paracoccus sulfuroxidans CGMCC1.5364
Paracoccus zeaxanthinifaciens ATCC21588
de novo VB12 synthesis (cobG pathway)
de novo VB12 synthesis (cbiX pathway)
Tree scale: 0.1
Tree scale: 0.01


(A) Genome Size
5 Mb
4 Mb
3 Mb
(B) Coding density
92%
90%
88%
86%
(C) GC content
70%
60%
50%
40%
(D) Number of pseudogenes
800
600
400
200
0
(E) ratio of pseudogene (%)
10
5
(F) C−ARSC
2.95
2.90
2.85
2.80
2.75
2.70
(G) N−ARSC
0.37
0.36
0.35
0.34
0.33
(H) Gene number
6000
5000
4000
3000
(I) Orthologous family number
4000
3500
3000
2500
2000
(J) Gene copy number per orthologous family
1.75
1.50
1.25
outgroup
sister group
CHUG members
seven other PRC members
other roseobacters
* p < 0.05, phylANOVA analysis
** p < 0.01, phylANOVA analysis




Orthologous gene families specific in CHUG members
merT
merC
merP
ABC.MS.S
actP
merA
Functional category
Light utilization
TCA cycle
Carbohydrates transportation
Amino acid and spermidine transportation
Amino acid biosynthesis
Cell wall and capsule biosynthesis
Metal homeostasis
Others
RPKM in seven other PRC members
10
1
0.1
0.01
10-3
10-4
10-5
<10-6
ABC.SS.S
icd
aapJ
pufB
ABC.MS.S
pufA
ahcY
ilvC
katG
rhaS
metH
smoE
gmd
glnB
ABC.SP.S
ABC.SP.S
dapD
zntA
ABC.MS.S
yaeT
zupT
ccoO
Proteorhodopsin
rbsB
acnB
ABC.PA.S
xsc
ABC.PA.S
phnD
potD
wza
ureJ
leuD
ABC.MS.S
phnA
Orthologous gene families specific in seven other PRC members
top 5% expressed gene families in CHUG members
top 5% expressed gene families in other PRC members
top 5% expressed gene families in both members
other gene families
<10-6
10-5
10-4
10-3
0.01
RPKM in CHUG members


(A)
Vitamin synthesis
Nitrogen utilization
Phosphorus utilization
Energy acquisition
Glucose degradation
Aromatic degradation
Organismal interaction
Methylated compounds degradation
thiE pdxH bioB cobG cbiX cobV btuB
amtB glnBD ntrBC ntrXY ureABC urtABCDE nrtABC potABCD ABC-SP occMQT potFGHI aapJMPQ ABC-PA livFGHKM
phoBR pstABCS phnGHIJKLM phoX plcP
PGC proteorhodopsin coxL sqr pdo sox napAB narGHI nirBD nirK nirS
pfk edd
pcaGH paaABCDE hmgA
cheAB fliC luxR virB vasKF GTA
dmdA dddD dddL tmd tmm tauABC xsc
(2123, +27, -45) (2316, 136, -25) FZCC0069 CHUG
(2141, +72, -34) (2273, 91, -31) LSUCC1028
(2329, 130, -20) FZCC0188
(2103, +29, -60)
(2068, +88, -122) (2238, 78, -39) LSUCC0246
(2102, +80, -22) (2284, 113, -11) LSUCC0387
(2134, +172, -406) (2276, 77, -18) HKCCD6035
(2210, +94, -18) (2433, 157, -41) HKCCA1065
(2368, +82, -134) (2409, 127, -40) HKCCA1288
(2430, +173, -213) (2606, 206, -144) LSUCC0031 Sister group
(2470, +96, -124) (3190, 813, -46) Roseicyclus mahoneyensis DSM16097
(2498, +173, -43) (2837, 453, -134) Roseibacterium elongatum DFL-43
(2400, +128, -150) (3427, 507, -90) Jannaschia sp. CCS1
(2908, +521, -111) (3534, 600, -49) Jannaschia sp. EhC01
(2422, +79, -205) (3178, 511, -108) Nioella nitratireducens SSW136 Outgroup
(2706, +414, -108) (3178, 464, -44) Nioella sediminis JS7-11
(2548) (3176, 599, -111) Jannaschia aquimarina DSM28248
(2644, +410, -187) (3066, 396, -51) Jannaschia pohangensis DSM19073
(2421, +132, -133) (2882, 603, -178) Thalassobacter stenotrophicus DSM16310
(3157, 632, -74) Dinoroseobacter shibae DFL12
(B)
LCA of CHUG
LCA of CHUG and its sister group
LCA of CHUG, its sister group and the outgroup
(C)
Rhodobacterales bacterium HTCC2083
Rhodobacterales bacterium HTCC2150
Rhodobacteraceae bacterium SB2
Rhodobacteraceae bacterium HIMB11 other PRC
Planktomarina temperata RCA23
Roseobacter sp. LE17
Rhodobacterales bacterium HTCC2255
LSUCC0246
LSUCC0387
HKCCD6035
FZCC0069
FZCC0188 CHUG
LSUCC1028
HKCCA1065
HKCCA1288
thiE pdxH bioB cobG cbiX cobV btuB
amtB glnBD ntrBC ntrXY ureABC urtABCDE nrtABC potABCD ABC-SP occMQT potFGHI aapJMPQ ABC-PA livFGHKM
phoBR pstABCS phnGHIJKLM phoX plcP
PGC PR coxL sqr pdo sox napAB narGHI nirBD nirKS nirKS
pfk edd
pcaGH paaABCDE hmgA
cheAB fliC luxR virB vasKF GTA
dmdA dddD dddL tmd tmm tauABC tpa


(A) HKCCA1288
(B) DSS−3
Number of cells (1x10⁸/mL)
2
1.5
1
0.5
0
0 12 24 36 48 60 72 84 96
time (h)
With vitamin B12
Without vitamin B12