A neutral process of genome reduction in marine bacterioplankton

Xiaojun Wang[1,2^], Mei Xie[1^], Kaitlyn Elizabeth Yee Kei Ho[1a], Ying Sun[1b], Xiao Chu[1c], Shuangfei Zhang[3d], Victoria Ringel[4], Hui Wang[3], Xiao-Hua Zhang[5], Zongze Shao[6], Yanlin Zhao[7,8], Thorsten Brinkhoff[9], Jörn Petersen[4], Irene Wagner-Döbler[10], Haiwei Luo[1,11*]

[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]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China

[3]Biology Department, College of Science, Shantou University, 515063, Shantou, China

[4]German Culture Collection for Microorganisms and Cell Cultures (DSMZ), Inhoffenstraße 7B, 38124, Braunschweig, Germany

[5]Frontiers Science Center for Deep Ocean Multispheres and Earth System & College of Marine Life Sciences, Ocean University of China, Qingdao, China

[6]Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, 361005, Xiamen, China

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

[8]Key Laboratory of Marine Biotechnology of Fujian Province, Institute of Oceanology, Fujian Agriculture and Forestry University, Fuzhou, China

[9]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Carl von Ossietzky Str. 9-11, 26129, Oldenburg, Germany

[10]Braunschweig Center for Systems Biology (BRICS), Technische Universität Braunschweig, Rebenring 56, 38106, Brunswick, Germany

[11]Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[a] Current Affiliation: Max Planck Institute for Marine Microbiology, 28359, Bremen, Germany

[b] Current Affiliation: BGI Research, Qingdao, 266555, Qingdao, China

[c] Current Affiliation: Institute of Biomedical Research, Yunnan University, 650500, Kunming,

China

[d] Current Affiliation: School of Minerals Processing and Bioengineering, Central South University, 410083, Changsha, China

[^]These are co-first authors.

[*]**Corresponding author:**

Haiwei Luo
School of Life Sciences, The Chinese University of Hong Kong
Hong Kong SAR
Phone: (+852) 39436121
E-mail: haiweiluo@cuhk.edu.hk

Running title: A neutral process of marine bacterial genome reduction

Key words: Roseobacter, CHUG, genome reduction, genetic drift, mutation rate

Abstract

Marine bacterioplankton communities are dominated by cells equipped with small genomes. Streamlining selection has been accepted as the main force driving their genome reduction. Here, we report that a neutral evolutionary mechanism governs genome reduction in the Roseobacter group that represents 5-20% of the bacterioplankton cells in coastal waters. Using representative strains that fall into three genome size groups (2-3, 3-4, and 4-5 Mbp), we measured their genomic mutation rates (μ) through long-term mutation accumulation experiments followed by genome sequencing the resulting 437 mutant lines. We further calculated their effective population sizes (N_e) based on μ and the neutral genetic diversity of the studied species, the latter estimated based on multiple genome sequences of natural isolates collected from global oceans with their population structure considered. A surprising finding is that N_e scales positively with genome size, which is the opposite of the expectation from the streamlining selection theory. As the strength of random genetic drift is the inverse of N_e, this result instead suggests drift as the primary driver of genome reduction. Additionally, we report a negative scaling between μ and genome size, which is the first experimental evidence for the long-lasting hypothesis that mutation rate increases play a part in marine bacterial genome reduction. As μ scales inversely with N_e, genetic drift appears to be the ultimate cause of genome reduction in these Roseobacters. Our finding discounts, but is insufficient to reject, the streamlining theory because streamlining process is expected to be more effective in oligotrophic open ocean waters.

Introduction

Bacterioplankton in the surface ocean is dominated by genome-reduced lineages, which are characterized by an average genome size of less than 2 Mbp [1–3]. Among these, the most abundant genome-reduced lineages are the SAR11 clade (Pelagibacteriales) of Alphaproteobacteria, the SAR86 clade of Gammaproteobacteria, and the genus *Prochlorococcus* of Cyanobacteria [4]. Streamlining selection is believed to drive the genome reduction of these free-living bacteria [5,6]. According to this theory, selection favors smaller genome size in nutrient-limited surface oceans because i) removal of non-essential DNA provides a metabolic advantage for bacteria, and ii) a concomitant increased surface-to-volume ratio (correlated with a decreased genome size) promotes nutrient uptake through enhanced diffusive delivery of nutrients to cell surface [6–8]. The key concept of streamlining selection theory hinges on the effective population size (N_e), which describes how many individuals in a bacterial population are contributing to the observed neutral genetic diversity of the population [9]. The underlying hypothesis is that free-living marine bacteria with reduced genomes have larger N_e than their relatives carrying larger genomes, thereby increasing the efficiency of natural selection acting to eliminate superfluous genomic DNA in oligotrophic environments. By comparing the N_e of related lineages with very different genome sizes, we can start to assess the definitive role of selection to downsize the genomes of free-living bacterioplankton cells in surface oceans.

However, this hypothesis has never been tested experimentally. The N_e can be calculated if π_s (the neutral genetic diversity of a population) and μ (unbiased genomic mutation rate) are available according to the equation $\pi_s = 2 \times N_e \times \mu$ [10]. Here, π_s is approximated by the diversity at four-fold degenerate sites of a population whose members recombine more frequently than those involving other populations. As a gold standard, μ is measured using a

long-term mutation accumulation (MA) experiment, which requires the experimental populations to repeatedly go through single-colony bottlenecks to eliminate the effect of natural selection [11]. This is commonly achieved by growing bacterial cells to single colonies in many parallel lines and propagating them for hundreds of generations. By whole genome sequencing (WGS) the mutant lines and comparing the mutant genomes with the ancestor's genome, the unbiased genomic mutation rate and spectrum can be calculated. However, genome-reduced bacterioplankton lineages inhabiting surface oceans typically do not grow on solid media or do not form single colonies on solid media, thus posing a significant challenge for their genomic mutation rate measurements and N_e estimation. A genome-reduced member of *Prochlorococcus* in the high-light-adapted clade II (*Prochlorococcus marinus* AS9601, ~1.7 Mbp) is the only genome-reduced surface ocean bacterioplankton lineage that was subjected to MA/WGS procedure [12].

The Roseobacter group (Roseobacteraceae, Alphaproteobacteria) is globally abundant, comprising up to 20% of the bacterial cells in coastal waters and 5% in the open ocean [13–15]. Cultured members of the Roseobacter group exhibit a wide range of genome size spanning from 2.5-2.6 Mbp (CHUG, NAC11-7) to 6.5-8.1 Mbp (*Salipiger*, *Poseidonocella*) (with the 25%, 50%, and 75% percentile being 3.83, 4.36, and 4.69 Mbp, respectively, Fig. S1). Most cultured lineages thrive on nutrient-rich solid media, but they are generally not representative in the oligotrophic pelagic ocean environments [16]. Early community structure analyses based on 16S rRNA gene revealed that marine Roseobacter communities are dominated by a few uncultivated lineages [17], whose genomes are becoming increasingly available through sequencing novel cultured members and sequencing uncultured members by single-cell genomics and metagenomic binning. Classical examples are DC5-80-3 (also named RCA or *Planktomarina*) [18–

[21], CHAB-I-5 [15,22], ChesI-A/B (also named SAG-O19) [23], NAC11-7 [24], and more recently CHUG [25,26]. These lineages constitute a polyphyletic group in the phylogenomic tree but form a tight pelagic Roseobacter cluster (PRC) in a dendrogram clustered from the presence and absence of gene families, reflecting their convergent evolution towards shared genome content and reduced genome sizes [22,26].

Unlike other PRC lineages, some CHUG isolates grow in single colonies and can be stably propagated on agar plates, thus rendering themselves appropriate for genomic mutation rate measurement through the classical MA/WGS procedure. Since other members of the Roseobacter group that carry larger and variable genomes and co-inhabit surface oceans are more readily available for μ determination, CHUG and its Roseobacter relatives commonly found in surface oceans create a unique opportunity to test the streamlining selection hypothesis.

Here, we report the μ of CHUG to be $(7.86 \pm 5.31) \times 10^{-10}$ base substitutions per nucleotide site per cell division, which, to date, represents the free-living bacterial species with the highest mutation rate measured by MA/WGS. Additionally, we report the N_e of CHUG to be 1.78×10^7, which, together with *Prochlorococcus* high light-adapted clade II (1.68×10^7) [12], represent the free-living bacterial species with the smallest N_e. We also determined μ and N_e of two other Roseobacter species, namely *Sulfitobacter pontiacus* and *Dinoroseobacter shibae*, and included the published *Ruegeria pomeroyi* data [20] for comparison. These four Roseobacter species are all commonly found in surface ocean habitats but have overall very different genome sizes (2.6, 3.5, 4.4, 4.6 Mbp). Remarkably, the assayed Roseobacter species that carry lower genome sizes have higher μ and smaller N_e. These scaling relationships imply that both genetic drift (i.e., the strength of drift is the inverse of N_e) and mutation rate increases play important roles in genome reduction of this bacterioplankton group. That N_e scales inversely with μ further implies that

139 genetic drift is the ultimate force that governs genome reduction. This is the first experimental
140 test of the prevailing streamlining selection theory and the finding is surprising because it
141 reverses the trend predicted by the streamlining theory in which streamlined marine
142 bacterioplankton lineages should have very large N_e such that genetic drift is negligible.
143
144 **Results**
145 *Genome sizes of the Roseobacter group members from surface oceans scale negatively with their*
146 *genomic mutation rates*
147 Here, we report the genomic mutation rate (μ) of three important members of the
148 Roseobacter group: CHUG, *Sulfitobacter pontiacus*, and *Dinoroseobacter shibae*. Along with μ
149 of *Ruegeria pomeroyi* we published earlier [20], we have four representative Roseobacter lineages
150 commonly found in surface ocean habitats and varying substantively in genome sizes (2.6, 3.5,
151 4.4 and 4.6 Mbp, respectively). For CHUG, a total of 200 mutation accumulation (MA) lines
152 were initiated from a single ancestral cell of strain HKCCA1288, 192 of which survived after 64
153 transfers with each line undergoing 1,472 cell divisions (corrected with death rate) and 180 of
154 which had over 50x coverage in WGS. Mutations were accumulated in 172 out of the 180 MA
155 lines, and a total of 596 base-pair substitution mutations (BPSs), 121 deletions and 29 insertions
156 were identified (Table S1).
157 In general, mutations are randomly distributed along the genomic regions of CHUG, though
158 clustered mutations were identified in two ribosomal RNA (rRNA) genes and 17 protein-coding
159 genes across multiple MA lines **(**Fig. 1A). In the former, 25 and 15 BPSs are clustered within
160 23S rRNA and 16S rRNA genes, respectively. Among these, 28 BPSs are contributed by a single
161 MA line and the remaining 12 are distributed across another 12 MA lines (Table S2). We

validated a randomly chosen BPS located in the 23S rRNA gene using PCR (Table S2). For the latter, 43 BPSs, three deletions, and 72 insertions fell into 17 genes across 72 MA lines, which represents a significant excess of mutations (bootstrap test; $p < 0.05$ for each gene, Fig. 1A). Most of the insertional mutations (66 of the 72) fall into three genes (I3V23_07740, I3V23_09110, and I3V23_11715), and most of them (60 of the 66) are frameshift mutations (Table S2), potentially leading to pseudogenization. Gene I3V23_07740 encodes the SLAC1 anion channel family protein linked to tellurite resistance, I3V23_09110 produces the DeoR/GlpR transcriptional regulator associated with fructose and glucose metabolism, and I3V23_11715 codes for the sarcosine oxidase subunit beta family protein (Table S2). Mutation clustering is a common phenomenon in mutation rate determination with the MA/WGS strategy [12,27–29], and it may be either caused by mutational hotspots or a result of positive selection as they likely increase fitness under experimental conditions. In the case of CHUG, frameshift mutations are enriched in the three genes, tentatively suggesting that deleting these genes likely confer benefits to the bacteria during the MA process. Thus, BPSs occurring in the above genes or intergenic region were excluded when calculating the genomic base-substitution mutation rate, and the remaining 505 BPSs translate to a μ of $(7.31 \pm 4.92) \times 10^{-10}$ (95% confidence interval [CI]: $6.69 \times 10^{-10} – 7.97 \times 10^{-10}$) base substitutions per site per cell division.

If the mutations clustered in the above genomic regions are indeed under selection, they may hitchhike mutations in linked genomic regions [30,31], thus inflating the mutation rate estimate. They may also have epistatic interactions with mutations in other genomic regions [32,33], likely delaying the fixation of other mutations and thus reducing the mutation rate estimate. To eliminate any such confounding factors in mutation rate estimation, we discarded the 83 MA lines with mutations accumulated in any of the 17 genes or the two ribosomal RNA genes with

185 excess mutations. The remaining 97 MA lines accumulated 270 BPSs, 19 deletions and 28
186 insertions (Table S1), translating to a μ of $(7.14 \pm 4.82) \times 10^{-10}$ (95% CI: $6.32 \times 10^{-10} – 8.05 \times$
187 10^{-10}), which is not significantly different from the mutation rate $(7.51 \pm 5.03) \times 10^{-10}$ (95% CI:
188 $6.60 \times 10^{-10} – 8.52 \times 10^{-10}$) derived from the remaining 83 MA lines mutated at aforementioned
189 genes but with these mutations excluded from the calculation (Wilcoxon–Mann–Whitney test, p
190 = 0.930).

191 Additional analyses were performed to evaluate the effect of selective pressure in the 97
192 MA lines where mutations are randomly distributed across the entire genomes. We found the
193 ratio of accumulated mutations in protein-coding sites to those in intergenic sites (233 vs 1) is
194 significantly larger than the ratio of the number of protein-coding sites to intergenic sites
195 (2,432,502 vs 214,229) (Fisher's exact test, $p < 0.001$). A similar pattern has also been reported
196 in MA experiments with other bacteria [34,35], and the increased mutation rate in coding regions
197 may be linked to transcription-induced mutations, considering that coding regions are transcribed
198 whereas most intergenic regions are not [36,37]. In terms of the protein-coding sites, the ratio of
199 accumulated nonsynonymous to synonymous mutations (171 vs 62) does not differ significantly
200 from that of nonsynonymous to synonymous sites (1,766,993 vs 665,509) (χ^2 test, $p = 0.85$),
201 supporting that selection did not play an important role during the MA process.

202 For the *Sulfitobacter* and *Dinoroseobacter*, MA experiments were conducted using two
203 strains, *S. pontiacus* EE-36 (=DSM 11700) and *D. shibae* DFL-12 (=DSM 16493^T), and the same
204 mutation calling method was implemented to calculate the spontaneous mutation rates. For EE-
205 36, an analysis of 58 MA lines with high-quality reads revealed 284 BPSs, 22 deletions and 14
206 insertions (Table S3&S4 and SI Results). Among the 51 MA lines without the potential effect of
207 the three mutation-enriched genes found here (Fig. 1C and SI Text 1.1), 243 BPSs, 11 insertions

and 18 deletions (Table S3) were kept, and the 243 BPSs translate to a μ of $(3.57 \pm 1.89) \times 10^{-10}$ (95% CI: $3.13 \times 10^{-10} – 4.03 \times 10^{-10}$). No selection was detected in the MA experiment of EE-36 (SI Text 1.1). For DFL-12, 149 MA lines with high quality reads yield 296 BPSs, 34 deletions and 82 insertions (Table S5&S6). After ignoring 87 MA lines potentially affected by the eight genes with an excess of mutations, the remaining 62 MA lines with 80 BPS lead to a μ of $(1.81 \pm 2.25) \times 10^{-10}$ (95% CI $1.44 \times 10^{-10} – 2.26 \times 10^{-10}$), along with seven insertions and nine deletions (Table S5). Significantly reduced BPSs were found in protein-coding genes and nonsynonymous sites than expected, suggesting a possible role of purifying selection in preventing the fixation of strong deleterious mutations during the MA process [38] of *D. shibae* DFL-12. For *R. pomeroyi* DSS-3 (=DSM 15171^T), the mutation rate was reported in our previous study [20], but its generation time was not corrected with its death rate. Here, we measured its death rate and updated its μ to be $(1.38 \pm 0.85) \times 10^{-10}$ (95% CI $1.17 \times 10^{-10} – 1.61 \times 10^{-10}$). In summary, the genomic mutation data of CHUG, *S. pontiacus* and *R. pomeroyi* are unbiased, whereas that of *D. shibae* may be slightly biased owing to a significant deficit in nonsynonymous mutations than expected (SI Results).

Among the four Roseobacter species (Fig. 2A), CHUG shows a significantly higher μ than the other three species (Wilcoxon–Mann–Whitney test, $p < 0.001$ in all three comparisons). Among the remaining species, *S. pontiacus* shows a significantly higher mutation rate than *D. shibae* and *R. pomeroyi* ($p < 0.001$ in both comparisons). No significant difference was found between *D. shibae* and *R. pomeroyi*. Our results show a negative correlation between genome size and μ among the Roseobacter lineages (dashed gray line in Fig. 2B [$r^2 = 0.992$, slope = −2.895, s.e.m. = 0.019, $p = 0.004$]) according to a generalized linear model (GLM) regression. This relationship is not caused by shared ancestry, as confirmed by phylogenetic generalized

231 least square (PGLS) regression analysis (solid blue line in Fig. 2B [r^2 = 0.991, slope = −2.894,
232 s.e.m. = 0.159, p = 0.003]). This conclusion remains robust by leaving out *D. shibae*.
233
234 *Genome sizes of the Roseobacter group members from surface oceans scale positively with*
235 *effective population sizes*
236 We calculated the N_e of CHUG according to the aforementioned equation $\pi_s = 2 \times N_e \times \mu$.
237 Here, π_s estimation requires the delineation of the population boundary. In many previous
238 studies involving bacterial N_e estimates, π_s was estimated based on a nominal bacterial species
239 [39,40]. However, N_e should be estimated based on neutral genetic diversity of a well-mixed
240 population, whereas a nominal bacterial species is commonly comprised of structured
241 populations [9,41]. The recently available tool PopCOGenT defines a bacterial population whose
242 members recombine more frequently than bacteria involving different populations [42], thereby
243 rendering itself a powerful method for bacterial N_e estimation. It was shown to outperform earlier
244 methods and was used to delineate population boundaries of many prokaryotic species for N_e
245 estimation [12]. We therefore used PopCOGenT to delineate populations for CHUG based on the
246 published genomes of 39 isolates we sampled mostly from brown algae ambient seawater [25,26] as
247 well as newly sequenced genomes of seven isolates from more diverse marine ecosystems such
248 as regular coastal seawater and coral ambient seawater (Fig. S2). This approach led to the
249 identification of two populations, one containing 21 non-redundant members (CHUG_MC0)
250 with a genomic median π_s of 0.043 translated to N_e of 2.74×10^7, and the other containing four
251 members (CHUG_MC1) with a median π_s of 0.013 and N_e of 8.27×10^6. This gave an average
252 N_e of $(1.78 \pm 1.35) \times 10^7$.

Using the same method, we estimated the mean N_e of the two populations of *S. pontiacus* (Fig. S3 and Table S9) and median N_e of five populations related to *R. pomeroyi* (Fig. S4 and Table S9) as $(4.01 \pm 3.73) \times 10^7$ and 1.51×10^8, respectively. There were only four *Dinoroseobacter* genomes publicly available, two (*D. shibae* DSM 112351 and *D. shibae* DFL-12 (= DSM 16493)) of which were derived from the same strain, and we extended the dataset by sequencing four additional strains. However, seven of them fall into a clonal complex (see Method) and are thus redundant. As a result, N_e of this species was estimated by using only two non-redundant genomes (*D. shibae* PD6 and *D. shibae* DFL-12), which gave 7.35×10^7 (Fig. S5) and remains to be updated in the future when more non-redundant strains become available.

The differences of N_e across the three genome size categories (2-3, 3-4, 4-5 Mbp) of these surface ocean Roseobacter lineages are statistically significant. Specifically, CHUG has a significantly lower N_e than the other three lineages (Wilcoxon–Mann–Whitney test, $p < 0.001$ in all cases), and *S. pontiacus* has a significantly lower N_e than *Ruegeria* sp. and *D. shibae* (Wilcoxon–Mann–Whitney test, $p < 0.001$ in all cases). Our results show that genome size correlates positively with N_e among these Roseobacter lineages (Fig. 2C). At first glance, this correlation is not significant according to the generalized linear model (GLM) regression analysis (dashed gray line in Fig. 2C [$r^2 = 0.890$, slope = 0.253, s.e.m. = 0.063, $p = 0.057$]. However, GLM is not appropriate in this case because there is a strong phylogenetic effect on the relationship between these two traits (indicated by the λ value of 1 in blue box of Fig. 2C). By controlling for the phylogenetic signal, the phylogenetic generalized least square (PGLS) supports significantly positive relationship (solid blue line in Fig. 2C [$r^2 = 0.871$, slope = 0.241, s.e.m. = 0.052, $p = 0.044$]).

Increased mutation rate correlates with increased power of genetic drift among the Roseobacters

According to the GLM regression, we found a significantly negative relationship between μ and N_e across the four Roseobacter lineages (dashed gray line in Fig. 2D [r^2 = 0.935, slope = −0.743, s.e.m. = 0.139, p = 0.033]). Although the observed relationship was impacted by phylogenetic effect (indicated by λ value at 1), it remains robust after controlling for shared ancestry, as shown by PGLS regression analysis (solid blue line in Fig. 2D [r^2 = 0.920, slope = −0.713, s.e.m. = 0.120, p = 0.027]).

The scaling relationships between genome size, mutation rate, and effective population size are robust across prokaryotes

Based on 31 prokaryotic species (mostly bacteria, mostly from non-marine habitats) with μ determined with the MA/WGS strategy (Table S10) and N_e calculated with the same approach as presented here, Chen et al. (2022) reported that both μ and the genome-wide mutation rate (U_P, a proxy for deleterious mutation load of a genome [11]) scales negatively with N_e. They also reported a negative scaling relationship between genome size and μ, but they found the negative correlation between genome size and N_e was not significant by both GLM and PGLS regression analyses [12]. By including the new data of the three Roseobacter lineages determined here, we confirmed the first three correlations (Fig. S6A&B&C). Intriguingly, we found a significant positive scaling relationship between genome size and N_e after controlling for the common ancestry (PGLS, solid blue line in Fig. S6D [r^2 = 0.251, slope = 0.147, s.e.m. = 0.048, p = 0.005]).

Discussion

The accepted genome streamlining theory attributes the process of genome reduction to selection for efficient use of limited nutrients [5]. For this reductive process to be visible to natural selection, the lineages undergoing genome streamlining must have greater N_e than their co-occurring sister lineages that carry larger genomes [6]. To study streamlining process in surface ocean bacterioplankton cells, N_e is the key parameter and bacterioplankton lineages found in surface oceans should be compared. The marine Roseobacter group provides a unique opportunity to test it. On one hand, group members that co-exist in surface oceans span a wide range of genome sizes. On the other hand, many Roseobacter populations are primarily associated with nutrient-rich marine habitats such as coral holobionts [43] and benthic environments [44]. These species are not appropriate targets to study the streamlining process because their growths are less limited by nutrients and their genome sizes are additionally shaped by characteristics of their habitats (e.g., host immune responses for host-associated Roseobacters) that are very different from those of surface oceans. In the case of the four Roseobacter lineages studied here, CHUG is featured by decoupling itself from marine eukaryotic phytoplankton groups [26], whereas *Dinoroseobacter shibae* is found to be primarily associated with marine phytoplankton [45]. Commonly found in between the two oceanic niches is *Ruegeria pomeroyi* [46]. Less is known for *Sulfitobacter pontiacus*, though other *Sulfitobacter* species are commonly associated with phytoplankton [47,48]. Although the difference in their ecological strategies (phytoplankton-associated versus free-living) are relevant to the difference in their genome sizes, these lineages remain valuable for testing the streamlining theory because they are all commonly found in surface ocean habitats and are generally subjected to nutrient limitation. Simulations

with an agent-based model showed that carbon limitation at geological timescales is the primary force that shapes the high genomic G+C content of *R. pomeroyi* DSS-3 [49].

A surprising result from the present study is that the N_e of the surface ocean Roseobacter lineages scales positively with their genome sizes (Fig. 2C). This is unexpected because it reverses the trend predicted by the streamlining selection theory. As the strength of genetic drift is the inverse of N_e, this finding means that genetic drift becomes increasingly powerful in bacterioplankton species that carry increasingly small genomes. For selection to eliminate a deleterious mutation and promote a beneficial mutation, the mutation needs to be sufficiently deleterious or beneficial (i.e., the absolute value of selection coefficient s, $|s|$, is large enough) to overcome the power of genetic drift ($1/N_e$), or the condition $|s| > 1/N_e$ needs to be fulfilled. For a population with a decreased N_e, more deleterious mutants are expected to be fixed and more beneficial variants will be lost by chance, rendering natural selection less effective. Therefore, our finding suggests that natural selection becomes less effective for the surface ocean Roseobacter populations with decreasing genome sizes.

For genetic drift to play a role in bacterial genome reduction, it needs to work with certain mutational processes. Most mutations are deleterious and purged primarily by purifying selection and recombination [50,51]. In a small population with limited opportunities for recombination, deleterious mutations are accumulated by random and irreversible loss of genotypes that are depleted with deleterious mutations and by random fixation (i.e., 100% in frequency) of genotypes that are loaded with abundant deleterious mutations. This process is known as "Muller's ratchet" [52]. A direct effect of Muller's ratchet in bacteria is genome reduction due to accumulation of irreversible disabling mutations that lead to pseudogenization and gene loss. Because obligate endosymbiotic bacteria are featured by very small population sizes with

343 extremely low recombination rates, their reductive evolution is generally driven by Muller's
344 ratchet [53,54].

345 As marine bacterioplankton lineages have much higher recombination rates than obligate
346 endosymbionts, their evolution was thought to be unlikely to be driven by Muller's ratchet [55].
347 Although this is true in the overall well-connected oceans, recombination may be restricted in a
348 frozen ocean of the globe in some unusual geological periods of the Earth. A prominent example
349 is that in the global icehouse climate conditions during the Neoproterozoic Snowball Earth that
350 comprises Sturtian (approx. 717–659 Ma) and Marinoan (approx. 645–635 Ma) glaciations,
351 *Prochlorococcus* cells were likely restricted to a few biotic refugia such as cryoconite holes and
352 sea-ice brine channels [56]. Such highly patchy and disconnected habitats likely created little
353 opportunity for recombination, which was evidenced by extremely few gene acquisitions at that
354 stage [57]. Additionally, *Prochlorococcus* at that time experienced severe population bottlenecks,
355 bringing its N_e down to 10^4-10^5 or even lower according to simulations by an agent-based model
356 [57]. The very small N_e and the very rare recombination support an escalated role of Muller's
357 ratchet as a driver of the *Prochlorococcus* evolution at that time. As the *Prochlorococcus* major
358 genome reduction event, where ~30% of the genomic DNA was eliminated, also occurred at that
359 time, it is natural to come up with a new theory that Muller's ratchet likely acted as a main
360 mechanism of the historical genome reduction in *Prochlorococcus* [57].

361 Unlike *Prochlorococcus* that are found exclusively between 40° N and 40° S and thus
362 vulnerable to global icehouse climate, oceanic members of the Roseobacter group are globally
363 distributed and generally enriched in ocean regions at middle and high latitudes [14,15], thereby
364 more resistant to glaciation events and less likely to undergo severe population bottlenecks

365 during historical glaciation events. For this reason, genome reduction in the oceanic members of
366 the Roseobacter group is less likely to be driven by Muller's ratchet.

367 Alternatively, genetic drift may lead to genome reduction through mutation rate increases.
368 This theory further builds on two population genetic processes: increased strength of genetic drift
369 leads to increased mutation rate, and mutation rate increase leads to genome reduction. If
370 mutation rate exceeds selection coefficient of a gene, purifying selection may not be able to keep
371 the gene [58,59]. With increasing mutation rates, more non-essential genes are lost and genome
372 reduction occurs. This explains why mutation rate increase leads to genome reduction. In fact, a
373 potential role of mutation rate increases in marine bacterial genome reduction has been
374 repeatedly hypothesized for *Prochlorococcus* and SAR11 [8,59–61]. The present study is the first that
375 experimentally validated this long-lasting hypothesis, though genome-reduced Roseobacters,
376 rather than *Prochlorococcus* and SAR11, were used in the analysis. A theoretical model predicts
377 that bacterial genome size is reduced by 30% in response to mutation rate increase by 10 times [59].
378 It seems that our empirical data from natural Roseobacter lineages is broadly consistent with this
379 prediction.

380 However, the population genetic process giving rise to increased mutation rate in bacteria
381 has been more disputable. On one hand, the "mutator theory" [62–64] favors natural selection as the
382 primary force to boost mutation rate; it posits that a high mutation rate may provide transient
383 advantages for prokaryotes in a changing environment as it increases the opportunity to gain
384 beneficial mutations. On the other hand, the "drift-barrier" model posits that natural selection
385 favors low mutation rates and acts to increase replication fidelity to a threshold beyond which
386 genetic drift can overcome the effects of natural selection and thus fitness advantages start to
387 decrease. The drift-barrier model favors a primary role of genetic drift in increasing mutation

388 rates and predicts an inverse relationship between N_e and μ [11,40]. It gained a great success in
389 explaining why eukaryotes (generally having small N_e) have higher mutation rates than
390 prokaryotes (generally having large N_e) [11,40]. From within bacteria, the model was originally
391 skeptical when genome-reduced marine bacterioplankton lineages were concerned. These marine
392 bacteria, especially *Prochlorococcus* and SAR11, were believed to have unusually large N_e [6,65]
393 and assumed to have high μ (because of losses of important repair systems) [8,59–61]. Not until
394 recently when the N_e and μ of a genome-reduced *Prochlorococcus* strain became available, the
395 drift-barrier model received stronger support from within bacteria [12]. The inclusion of the three
396 Roseobacter species determined here, especially the CHUG, continues to support the negative
397 scaling relationship between N_e and μ (Fig. S6C), thus strengthening the drift-barrier theory. A
398 gross relationship across deeply-branching bacterial lineages means that genetic drift is a
399 universal rule dictating the evolution of mutation rates in bacteria. However, it does not mean
400 that this rule is necessarily manifested by an analysis of bacterial species from a certain habitat
401 because the pattern could be shadowed by selective pressure imposed by the particular
402 environment. It is therefore remarkable to observe the inverse scaling relationship between N_e
403 and μ holds when only the four Roseobacter species were compared (Fig. 2D), supporting the
404 idea that genetic drift drives mutation rate increases in the studied Roseobacters.
405 Our study shows that genetic drift and mutation rate increases are the two primary
406 population genetic mechanisms that mediate genome reduction in surface ocean Roseobacter
407 lineages. We also show that mutation rate increase itself is driven by genetic drift. Taken together,
408 we are able to conclude that genetic drift is the ultimate mechanism for genome reduction in the
409 studied marine bacteria. Our new result contributes to the ongoing discussion on the population
410 genetic mechanisms giving rise to the small genomes in marine bacterial cells that dominate the

411 marine bacterioplankton communities. It provides an alternative view to the prevailing
412 streamlining selection theory, but the presented evidence from the studied surface ocean
413 Roseobacter lineages, several of which are commonly associated with phytoplankton where
414 nutrients are more available than in the bulk seawater, is not sufficient to reject the streamlining
415 selection theory. The genome streamlining process is deemed to be more efficient in
416 bacterioplankton cells that are found in highly oligotrophic and stratified oceans such as the
417 ocean gyres where nutrients are extremely depleted, and thus the streamlining selection theory is
418 believed to work best in explaining the genome reduction process in those bacterioplankton
419 lineages. Future work should be focused on comparing lineages that dominate the
420 bacterioplankton communities in oligotrophic oceans.
421

422 **Data and code availability**

423 All the datasets generated, analyzed and presented in the current study are available in the
424 Supplementary Information. Raw reads of the surviving lines of three Roseobacter lineages are
425 available at the NCBI SRA under accession no. PRJNA1067981. The closed genome of *S.*
426 *pontiacus* EE-36 are available at the NCBI GenBank under accession no. SAMN39584998. All
427 the scripts are deposited at https://github.com/Xiaojun928/Genome_reduction_bacterioplankton.
428

429 **Author contribution**

430 H.L. conceived and designed the study. X.W. performed the bioinformatics. H.L. and X.W. wrote
431 the paper. M.X. and K.H. performed the CHUG MA experiment, M.X. measured the death rates
432 of all MA experiments, validated the mutations of CHUG, and constructed DNA library for
433 genome sequencing. Y.S. and X.C. performed the *S. pontiacus* EE-36 MA experiments. Y.S.

434 assembled the complete genome of *S. pontiacus* EE-36. S.Z. and H.W. conducted the *D. shibae*
435 DFL-12 MA experiments. X.C. and M.X. cultured new CHUG isolates. V.R., T.B., J.P. and I.W.D
436 contributed the four new *D. shibae* strains. Z.S. and X-H.Z. contributed five and five *S.*
437 *pontiacus* strains, respectively. M.X, Y.S. T.B., J.P. and I.W.D edited the manuscript. I.W.D
438 provided substantial comments that significantly improved this manuscript.
439

440 **Acknowledgements**
441 We thank Danli Luo for the isolation of two CHUG members from ambient water of corals, Ho
442 Ching Leung for helping prepare the CHUG MA experiments, and Yang Qian for his assistance
443 in the isolation of some CHUG strains. This work is supported by the Hong Kong Research
444 Grants Council General Research Fund (14110820) and the Hong Kong Research Grants Council
445 Area of Excellence Scheme (AoE/M-403/16). X.W. was supported by the China Postdoctoral
446 Science Foundation (2022M722220). T.B, J.P and I.W.D were supported by the Deutsche
447 Forschungsgemeinschaft within the Collaborative Research Centre TRR51 Roseobacter.
448

449 **Competing interests**
450 The authors declare no competing interests in relation to this work.
451
452

453 **Figure legends**
454 **Fig. 1. Distribution of genomic mutations in three Roseobacter species determined in this**
455 **study by mutation accumulation (MA) experiments followed by whole genome sequencing**
456 **of 437 MA lines in total. (A)** Base-substitution mutations and insertion/deletion mutations

457 across the whole genome of CHUG HKCCA1288 determined by 180 MA lines. The height of
458 each bar represents the number of base substitutions (black), insertions (red) and deletions
459 (yellow) across all MA lines within each protein-coding gene. Black diamonds and red triangles
460 denote base substitutions and insertions that occurred on the remaining genomic regions
461 (intergenic regions and non-protein-coding genes), respectively; both diamonds and triangles are
462 shown with transparence, thus genomic regions with more mutations show deeper color than
463 those with less mutations. The genomic position of insertion/deletion mutation refers to the
464 position of the first mutated site. The locus tag of the 17 genes with statistical enrichment of
465 mutations is shown in purple arrows. **(B)** Base-substitution mutations and insertion/deletion
466 mutations across the whole genome of *Sulfitobacter pontiacus* EE-36 determined by 58 MA lines.
467 Different types of mutations are illustrated with the same symbols as those in (A). The locus tag
468 of the three genes with statistical enrichment of mutations is shown in purple arrows.
469 Chromosome and plasmid are separated with the vertical dash line. **(C)** Base-substitution
470 mutations and insertion/deletion mutations across the whole genome of *Dinoroseobacter shibae*
471 DFL-12 determined by 149 MA lines. Different types of mutations are illustrated with the same
472 symbols as those in (A). The locus tag of the eight genes with statistical enrichment of mutations
473 is shown in purple arrows. Chromosome, chromids, and plasmids are separated with the vertical
474 dash line.
475
476 **Fig. 2. Phylogenomic tree of the Roseobacter group and scaling relationships between**
477 **genome size, genomic mutation rate (base-substitution mutation rate per cell division per**
478 **nucleotide site; μ) and effective population size (N_e) for the studied Roseobacter species.** All
479 trait values were logarithmically transformed. **(A)** The maximum likelihood phylogenomic tree

480 of representative Roseobacter genomes based on 120 conserved bacterial genes, with the four
481 studied Roseobacter species marked in red. **(B-D)** Scaling relationships involving μ, N_e, and
482 genome size across four phylogenetically diverse lineages within the marine Roseobacter group,
483 with all trait values logarithmically transformed. The dashed grey lines and solid blue lines
484 represent the generalized linear model (GLM) and phylogenetic generalized least square (PGLS)
485 regression, respectively.

References

1. Mende, D. R. *et al.* Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol* **2**, 1367–1373 (2017).
2. Pachiadaki, M. G. *et al.* Charting the Complexity of the Marine Microbiome through Single-Cell Genomics. *Cell* **179**, 1623-1635.e11 (2019).
3. Yooseph, S. *et al.* Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**, 60–66 (2010).
4. Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science (New York, N.Y.)* **348**, 1261359 (2015).
5. Giovannoni, S. J. *et al.* Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science* **309**, 1242–1245 (2005).
6. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).
7. Kiørboe, T. Random walk and diffusion. in *A Mechanistic Approach to Plankton Ecology* 10–34 (Princeton University Press, 2008).
8. Partensky, F. & Garczarek, L. Prochlorococcus: Advantages and Limits of Minimalism. *Annual Review of Marine Science* **2**, 305–331 (2010).
9. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
10. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, 1983). doi:10.1017/CBO9780511623486.
11. Lynch, M. *et al.* Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016).
12. Chen, Z. *et al.* Prochlorococcus have low global mutation rate and small effective population size. *Nat Ecol Evol* **6**, 183–194 (2022).
13. Moran, M. A. *et al.* Ecological Genomics of Marine *Roseobacters*. *Appl. Environ. Microbiol.* **73**,

4559–4569 (2007).

14. Selje, N., Simon, M. & Brinkhoff, T. A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature* **427**, 445–448 (2004).
15. Zhang, Y., Sun, Y., Jiao, N., Stepanauskas, R. & Luo, H. Ecological Genomics of the Uncultivated Marine Roseobacter Lineage CHAB-I-5. *Appl Environ Microbiol* **82**, 2100–2111 (2016).
16. Swan, B. K. *et al.* Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proceedings of the National Academy of Sciences* **110**, 11463–11468 (2013).
17. Buchan, A., González, J. M. & Moran, M. A. Overview of the marine *Roseobacter* lineage. *Appl. Environ. Microbiol.* **71**, 5665–5677 (2005).
18. Giebel, H.-A. *et al.* Planktomarina temperata gen. nov., sp. nov., belonging to the globally distributed RCA cluster of the marine Roseobacter clade, isolated from the German Wadden Sea. *International Journal of Systematic and Evolutionary Microbiology* **63**, 4207–4217 (2013).
19. Liu, Y. *et al.* Metagenome-assembled genomes reveal greatly expanded taxonomic and functional diversification of the abundant marine Roseobacter RCA cluster. *Microbiome* **11**, 265 (2023).
20. Sun, Y. *et al.* Spontaneous mutations of a model heterotrophic marine bacterium. *ISME J* **11**, 1713–1718 (2017).
21. Voget, S. *et al.* Adaptation of an abundant Roseobacter RCA organism to pelagic systems revealed by genomic and transcriptomic analyses. *The ISME Journal* **9**, 371–384 (2015).
22. Billerbeck, S. *et al.* Biogeography and environmental genomics of the Roseobacter-affiliated pelagic CHAB-I-5 lineage. *Nature Microbiology* **1**, 16063 (2016).
23. Luo, H., Swan, B. K., Stepanauskas, R., Hughes, A. L. & Moran, M. A. Evolutionary analysis of a streamlined lineage of surface ocean Roseobacters. *The ISME journal* **8**, 1428–39 (2014).
24. Luo, H., Swan, B. K., Stepanauskas, R., Hughes, A. L. & Moran, M. A. Comparing effective population sizes of dominant marine alphaproteobacteria lineages. *Environ Microbiol. Rep.* **6**, 167–172 (2014).
25. Chu, X. *et al.* Coastal Transient Niches Shape the Microdiversity Pattern of a Bacterioplankton

538 Population with Reduced Genomes. *mBio* **0**, e00571-22 (2022).

539 26. Feng, X. *et al.* Mechanisms driving genome reduction of a novel Roseobacter lineage. *ISME J* (2021)
540 doi:10.1038/s41396-021-01036-3.

541 27. Farlow, A. *et al.* The spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*.
542 *Genetics* **201**, 737–744 (2015).

543 28. Gu, J. *et al.* Unexpectedly high mutation rate of a deep-sea hyperthermophilic anaerobic archaeon.
544 *ISME J* (2021) doi:10.1038/s41396-020-00888-5.

545 29. Konrad, A. *et al.* Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in Caenorhabditis
546 elegans Spontaneous Mutation Accumulation Lines of Differing Population Size. *Molecular Biology*
547 *and Evolution* **34**, 1319–1334 (2017).

548 30. Shaver, A. C. *et al.* Fitness Evolution and the Rise of Mutator Alleles in Experimental Escherichia
549 coli Populations. *Genetics* **162**, 557–566 (2002).

550 31. Ho, W.-C. *et al.* Evolutionary Dynamics of Asexual Hypermutators Adapting to a Novel Environment.
551 *Genome Biology and Evolution* **13**, evab257 (2021).

552 32. Lyons, D. M., Zou, Z., Xu, H. & Zhang, J. Idiosyncratic epistasis creates universals in mutational
553 effects and evolutionary trajectories. *Nat Ecol Evol* **4**, 1685–1693 (2020).

554 33. Halligan, D. L. & Keightley, P. D. Spontaneous Mutation Accumulation Studies in Evolutionary
555 Genetics. *Annual Review of Ecology, Evolution, and Systematics* **40**, 151–172 (2009).

556 34. Dillon, M. M. & Cooper, V. S. The fitness effects of spontaneous mutations nearly unseen by
557 selection in a bacterium with multiple chromosomes. *Genetics* **204**, 1225–1238 (2016).

558 35. Long, H. *et al.* Mutation Rate, Spectrum, Topology, and Context-Dependency in the DNA Mismatch
559 Repair-Deficient Pseudomonas fluorescens ATCC948. *Genome Biol Evol* **7**, 262–271 (2014).

560 36. Beletskii, A. & Bhagwat, A. S. Transcription-induced mutations: Increase in C to T mutations in the
561 nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **93**,
562 13919–13924 (1996).

563 37. Datta, A. & Jinks-Robertson, S. Association of Increased Spontaneous Mutation Rates with High

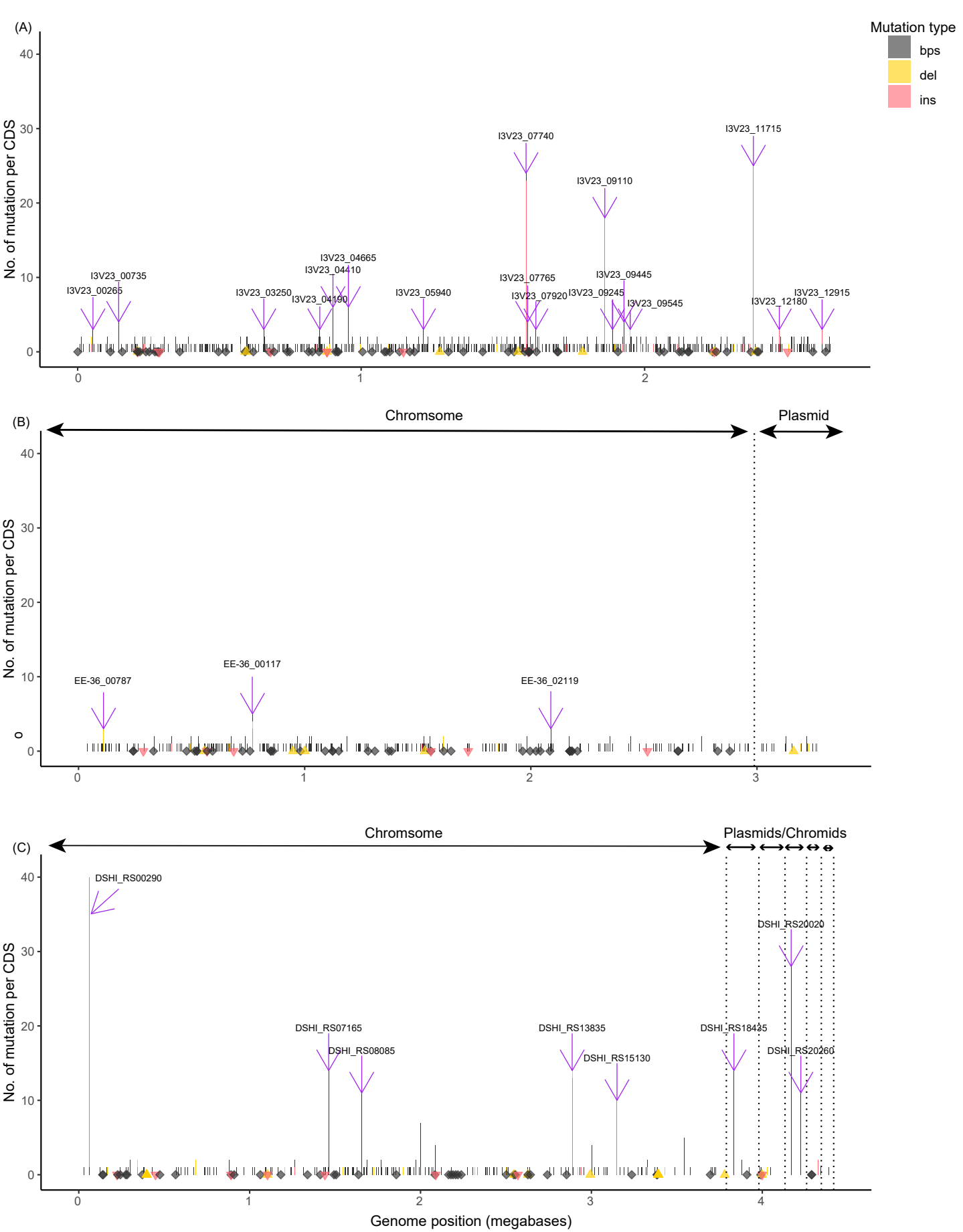
Levels of Transcription in Yeast. *Science* **268**, 1616–1619 (1995).

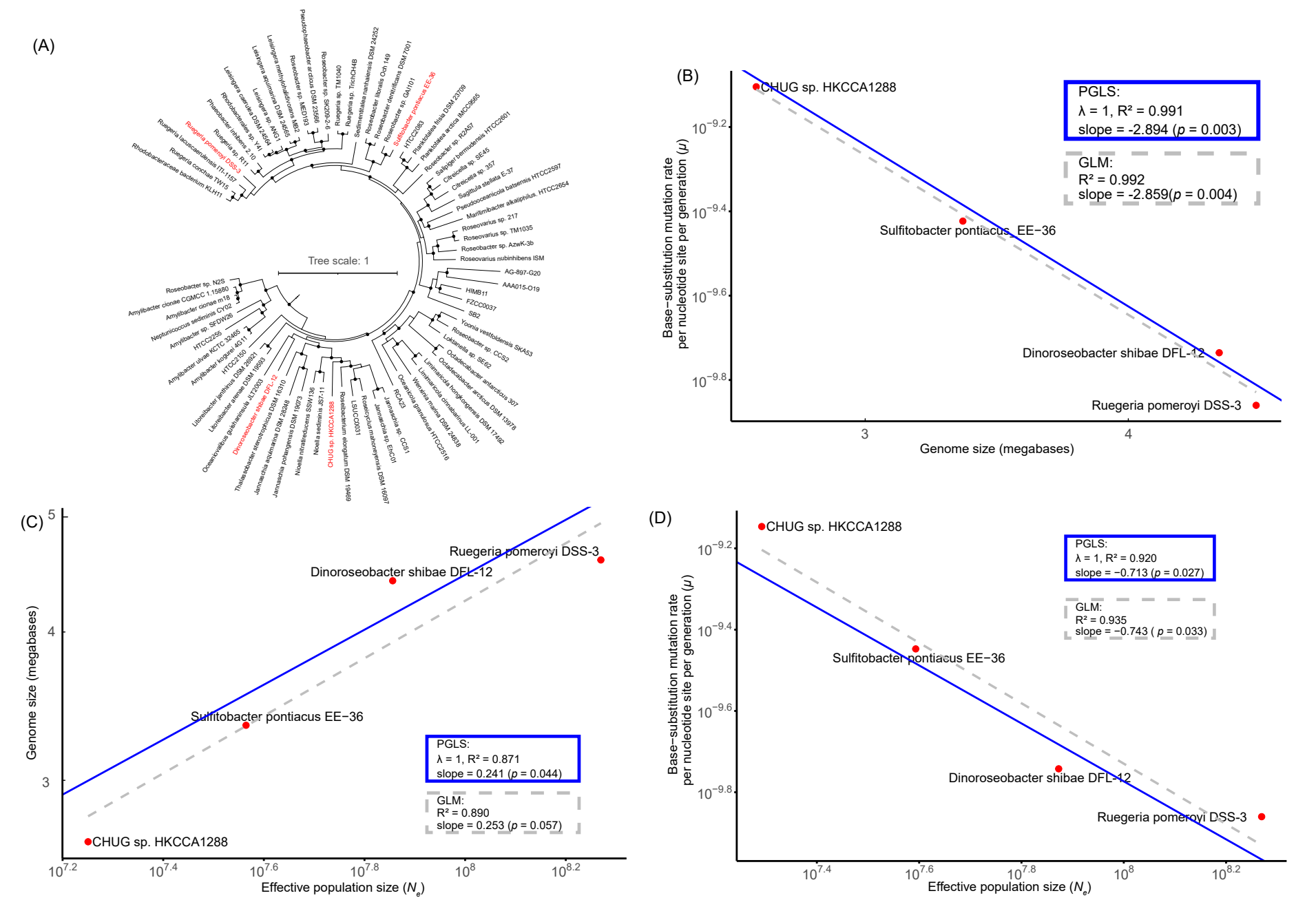
38. Keightley, P. D. *et al.* Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Res.* **19**, 1195–1201 (2009).
39. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
40. Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci. USA* **109**, 18488–18492 (2012).
41. Price, M. N. & Arkin, A. P. Weakly Deleterious Mutations and Low Rates of Recombination Limit the Impact of Natural Selection on Bacterial Genomes. *mBio* **6**, (2015).
42. Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J. & Polz, M. F. A reverse ecology approach based on a biological definition of microbial populations. *Cell* **178**, 820–834 (2019).
43. Huggett, M. J. & Apprill, A. Coral microbiome database: Integration of sequences reveals high diversity and relatedness of coral-associated microbes. *Environ Microbiol Rep* **11**, 372–385 (2019).
44. Lenk, S. *et al. Roseobacter* clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *ISME J* **6**, 2178–2187 (2012).
45. Wagner-Döbler, I. *et al.* The complete genome sequence of the algal symbiont *Dinoroseobacter shibae* : a hitchhiker's guide to life in the sea. *ISME J* **4**, 61–77 (2010).
46. Fu, H., Uchimiya, M., Gore, J. & Moran, M. A. Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proceedings of the National Academy of Sciences* **117**, 3656–3662 (2020).
47. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101 (2015).
48. Shibl, A. A. *et al.* Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proceedings of the National Academy of Sciences* **117**, 27445–27455 (2020).
49. Hellweger, F. L., Huang, Y. & Luo, H. Carbon limitation drives GC content evolution of a marine bacterium in an individual-based genome-scale model. *ISME J.* **12**, 1180–1187 (2018).
50. Hughes, A. L. Near-Neutrality: the Leading Edge of the Neutral Theory of Molecular Evolution. *Ann N Y Acad Sci* **1133**, 162–179 (2008).

590 51. Nei, M. Selectionism and Neutralism in Molecular Evolution. *Molecular Biology and Evolution* **22**,
591 2318–2342 (2005).
592 52. Muller, H. J. The relation of recombination to mutational advance. *Mutation Research/Fundamental*
593 *and Molecular Mechanisms of Mutagenesis* **1**, 2–9 (1964).
594 53. Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of*
595 *the National Academy of Sciences* **93**, 2873–2878 (1996).
596 54. van Ham, R. C. H. J. *et al.* Reductive genome evolution in Buchnera aphidicola. *Proceedings of the*
597 *National Academy of Sciences* **100**, 581–586 (2003).
598 55. Batut, B., Knibbe, C., Marais, G. & Daubin, V. Reductive genome evolution at both ends of the
599 bacterial population size spectrum. *Nat. Rev. Microbiol.* **12**, 841–850 (2014).
600 56. Zhang, H., Sun, Y., Zeng, Q., Crowe, S. A. & Luo, H. Snowball Earth, population bottleneck and
601 *Prochlorococcus* evolution. *Proc Royal Soc B* **288**, 20211956 (2021).
602 57. Zhang, H., Hellweger, F. L. & Luo. Genome reduction occurred in early Prochlorococcus with an
603 unusually low effective population size. *The ISME Journal* **in press**, (2024).
604 58. Biebricher, C. K. & Eigen, M. The error threshold. *Virus Res* **107**, 117–127 (2005).
605 59. Marais, G. A. B., Calteau, A. & Tenaillon, O. Mutation rate and genome reduction in endosymbiotic
606 and free-living bacteria. *Genetica* **134**, 205–210 (2008).
607 60. Dufresne, A., Garczarek, L. & Partensky, F. Accelerated evolution associated with genome reduction
608 in a free-living prokaryote. *Genome Biol* **6**, R14 (2005).
609 61. Viklund, J., Ettema, T. J. G. & Andersson, S. G. E. Independent Genome Reduction and Phylogenetic
610 Reclassification of the Oceanic SAR11 Clade. *Molecular Biology and Evolution* **29**, 599–615 (2012).
611 62. Taddei, F. *et al.* Role of mutator alleles in adaptive evolution. *Nature* **387**, 700–702 (1997).
612 63. Tenaillon, O., Toupance, B., Nagard, H. L., Taddei, F. & Godelle, B. Mutators, population size,
613 adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**, 485–493
614 (1999).
615 64. Giraud, A., Radman, M., Matic, I. & Taddei, F. The rise and fall of mutator bacteria. *Curr Opin*

Microbiol **4**, 582–585 (2001).

65. Kashtan, N. *et al*. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**, 416–420 (2014).


(A)
Mutation type
bps
del
ins
No. of mutation per CDS
I3V23_00735
I3V23_00265
I3V23_03250
I3V23_04665
I3V23_04410
I3V23_04190
I3V23_05940
I3V23_07740
I3V23_07765
I3V23_07920
I3V23_09110
I3V23_09245
I3V23_09445
I3V23_09545
I3V23_11715
I3V23_12180
I3V23_12915
(B)
Chromsome
Plasmid
No. of mutation per CDS
EE-36_00787
EE-36_00117
EE-36_02119
(C)
Chromsome
Plasmids/Chromids
No. of mutation per CDS
DSHI_RS00290
DSHI_RS07165
DSHI_RS08085
DSHI_RS13835
DSHI_RS15130
DSHI_RS18435
DSHI_RS20020
DSHI_RS20260
Genome position (megabases)


(A)
Tree scale: 1
Pseudophaeobacter arcticus DSM 23566
Roseobacter sp. SK209-2-6
Roseobacter sp. TM1040
Ruegeria sp. TrichCH4B
Sediminitalea nanhaiensis DSM 24252
Roseobacter litoralis Och 149
Roseobacter denitrificans DSM 7001
Sulfitobacter pontiacus EE-36
Roseobacter sp. GAI101
HTCC2083
Planktotalea frisia DSM 23709
Planktotalea arctica IMCC9605
Roseobacter sp. RZA57
Salipiger bermudensis HTCC2601
Citreicella sp. SE45
Citreicella sp. 357
Sagittula stellata E-37
Pseudooceanicola batsensis HTCC2597
Marinibacter alkaliphilus HTCC2654
Roseovarius sp. 217
Roseovarius sp. TM1035
Roseobacter sp. AzwK-3b
Roseovarius nubinhibens ISM
AG-897-G20
AAA015-O19
HIMB11
FZCC0037
SB2
Yoonia vestfoldensis SKA53
Roseobacter sp. CCS2
Lokiarchaeota sp. SE62
Octadecabacter antarcticus 307
Octadecabacter arcticus DSM 13978
Limimaricola hongkongensis DSM 17492
Limimaricola cinnabarinus LL-001
Wenxinzhinanlia marina DSM 24838
Oceanicola granulosus HTCC2516
RCA23
Jannaschia sp. CCS1
Jannaschia sp. EhC01
Roseivivax halotolerans DSM 16097
LSUCC0031
Roseibacterium elongatum DSM 19469
CHUG sp. HKCCA1288
Nisaea sediminis JS7-11
Nisaea nitritireducens SSW136
Nisaea denitrificans DSM 18693
Jannaschia pohangensis DSM 18673
Jannaschia aquimarina DSM 28248
Thalassobacter stenotrophicus DSM 16310
Dinoroseobacter shibae DFL-12
Litoreibacter arenae DSM 19593
Litoreibacter janthinus DSM 26921
Oceanicaulis gilvimarinus JCM 32003
Amylibacter kogurei NBRC 101654
Amylibacter ulvae KCTC 32465
HTCC2160
Amylibacter sp. HTCC2255
Amylibacter sp. SFDW26
Neptunicoccus sediminis CY02
Amylibacter cionae m19
Amylibacter clonae CGMCC 1.15880
Roseobacter sp. N2S
Ruegeria pomeroyi DSS-3
Ruegeria lacuscaerulensis ITI-1157
Ruegeria conchae TW15
Rhodobacteraceae bacterium KLH11
Phaeobacter inhibens 2.10
Ruegeria sp. R11
Rhodobacterales sp. Y41
Leisingera caerulea DSM 24564
Leisingera sp. ANG1
Leisingera aquimarina DSM 24565
Roseobacter sp. MED193
Leisingera methylohalidivorans M662
(B)
CHUG sp. HKCCA1288
Sulfitobacter pontiacus EE−36
Dinoroseobacter shibae DFL-12
Ruegeria pomeroyi DSS-3
PGLS:
λ = 1, R² = 0.991
slope = -2.894 (p = 0.003)
GLM:
R² = 0.992
slope = -2.859(p = 0.004)
Base−substitution mutation rate per nucleotide site per generation (μ)
Genome size (megabases)
(C)
Ruegeria pomeroyi DSS-3
Dinoroseobacter shibae DFL-12
Sulfitobacter pontiacus EE−36
CHUG sp. HKCCA1288
PGLS:
λ = 1, R² = 0.871
slope = 0.241 (p = 0.044)
GLM:
R² = 0.890
slope = 0.253 (p = 0.057)
Genome size (megabases)
Effective population size (Ne)
(D)
CHUG sp. HKCCA1288
Sulfitobacter pontiacus EE−36
Dinoroseobacter shibae DFL-12
Ruegeria pomeroyi DSS-3
PGLS:
λ = 1, R² = 0.920
slope = −0.713 (p = 0.027)
GLM:
R² = 0.935
slope = −0.743 (p = 0.033)
Base−substitution mutation rate per nucleotide site per generation (μ)
Effective population size (Ne)