Zhang Xiao-Hua (Orcid ID: 0000-0002-7428-7775)
Luo Haiwei (Orcid ID: 0000-0001-8452-6066)

Ancestral Niche Separation and Evolutionary Rate Differentiation Between Sister Marine Flavobacteria Lineages

Chun-Xu Xue[†,1,2], Hao Zhang[†,2], He-Yu Lin[†,1], Ying Sun[2], Danli Luo[2], Yongjie Huang[2], Xiao-Hua Zhang[*,1], Haiwei Luo[*,2,3]

[1]College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

[2]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

[3]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518000, China

[†]These authors contributed equally to this study.

[*]**Corresponding author**:

Haiwei Luo
E-mail: hluo2006@gmail.com
The Chinese University of Hong Kong

Xiao-Hua Zhang
E-mail: xhzhang@ouc.edu.cn
Ocean University of China

Key words: Flavobacteria, evolutionary rate, genome reduction, mutation rate, glycoside hydrolase, CAZyme

Originality significance statement

Members of the flavobacteria group in *Bacteroidetes* are prevalent in marine

environments, representing up to 10-20% of the bacterial cells in some ocean regions.

Flavobacteria are generally considered as the specialists for polysaccharide

degradation because of the excess of genes encoding carbohydrate active enzymes,

and thus play an important role in marine carbon cycle. The present study employs a

combination of genomic analyses, physiological assays, ancestral reconstruction and

experimental evolution, and addresses the evolutionary processes and mechanisms

underlying niche differentiation of two related flavobacteria lineages.

Summary

Marine flavobacteria are specialists for polysaccharide degradation. They dominate in habitats enriched with polysaccharides, but are also prevalent in pelagic environments where polysaccharides are less available. These niches are likely occupied by distinct lineages, but evolutionary processes underlying their niche differentiation remain elusive. Here, genomic analyses and physiological assays indicate that the sister flavobacteria lineages *Leeuwenhoekiella* and *Nonlabens* likely explore polysaccharide-rich macroalgae and polysaccharide-poor pelagic niches, respectively. Phylogenomic analyses inferred that the niche separation likely occurred anciently and coincided with increased sequence evolutionary rate in *Nonlabens* compared to *Leeuwenhoekiella*. Further analyses ruled out the known mechanisms likely driving evolutionary rate acceleration, including reduced selection efficiency, decreased generation time and increased mutation rate. In particular, the mutation rates were determined using an unbiased experimental method, which measures the present-day populations and may not reflect ancestral populations. These data collectively lead to a new hypothesis that an ancestral and transient mutation rate increase resulted in evolutionary rate increase in *Nonlabens*. This hypothesis was supported by inferring that gains and losses of genes involved in SOS response, a mechanism known to drive transiently increased mutation rate, coincided with evolutionary rate acceleration. Our analyses highlight the evolutionary mechanisms underlying niche differentiation of flavobacteria lineages.

Introduction

Approximately one-half of the oceanic primary production is consumed by heterotrophic marine bacteria (Fuhrman and Azam, 1982; Fenchel, 1988). Among these, flavobacteria is one of the most effective groups because of its high abundance in global oceans and its specialized capability for polymeric organic matter utilization (Kirchman, 2002). Marine flavobacteria thrive in diverse ecological niches and are enriched in coastal waters and during phytoplankton blooms, in which they account for up to ~40% and ~70% of bacterial cells, respectively (Cottrell and Kirchman, 2000; Kirchman, 2002; Teeling et al., 2016). The unique role of flavobacteria in high molecular weight organic matter utilization largely results from an excess of genes encoding carbohydrate active enzymes (CAZymes) and peptidases carried in their genomes (Bauer et al., 2006; Fernández-Gomez et al., 2013). CAZymes often cluster with *susCD*-like transporter genes, and form polysaccharide utilization locus (PUL) for binding, degradation and uptake of different classes of polysaccharides (Kappelmann et al., 2019).

Despite their generally recognized role in polysaccharide and peptide degradation, a few studies have revealed niche specialization and adaptation among flavobacteria lineages. For example, although members of the genus *Formosa* have been largely isolated from the nutrient-enriched niches such as brown algae, a few were proposed to inhabit the pelagic ocean (Mann et al., 2013). We recently demonstrated three unidirectional transitions from marine to non-marine habitats during the evolutionary history of flavobacteria, which was likely facilitated by

repeated gains and losses of niche-specifying genes (Zhang et al., 2019). The present study investigates the potential niche separation between two well-established sister lineages, *Nonlabens* and *Leeuwenhoekiella*. While the two genera are distributed in overlapping marine niches such as seawater and invertebrates (Nedashkovskaya et al., 2005; Lau et al., 2006; Pinhassi et al., 2006; Park et al., 2012), they show distinct genomic features, suggestive of niche specialization. Members of *Leeuwenhoekiella* consistently have larger genome size (over 4 Mbp), versatile metabolic capabilities and gliding motility, whereas *Nonlabens* have reduced genomes (around 3 Mbp), decreased gliding motility, and use rhodopsins to acquire energy from sunlight in some members (Fernández-Gomez et al., 2013; Kwon et al., 2013).

A major benefit of using *Nonlabens* and *Leeuwenhoekiella* is that they allow for an in-depth analysis of evolutionary forces associated with their niche separation. As repeatedly shown in published phylogenomic trees in which these two lineages were included (Kumagai et al., 2018; Zhang et al., 2019), the genome-reduced *Nonlabens* evolved at an accelerated rate compared to *Leeuwenhoekiella*, manifested as extraordinarily long branches connected to the former members since the last common ancestor (LCA) shared by the two lineages. In fact, free-living marine bacterial lineages with reduced genome sizes commonly show accelerated rate of sequence evolution at genome-wide scale (Batut et al., 2014), with prominent examples being the genome-reduced *Prochlorococcus* lineages in Cyanobacteria, the SAR11 clade in Alphaproteobacteria, and the SAR86 clade in Gammaproteobacteria. A few evolutionary mechanisms have been invoked to explain the genome reduction

in free-living marine bacteria, which highlighted the role of positive selection (Giovannoni et al., 2005), increased mutation rate (Marais et al., 2008), and genetic drift (Luo et al., 2017). In theory, these evolutionary forces may potentially also drive the genome-wide accelerated evolutionary rate.

Among these mechanisms, genetic drift has been subject to rigorous tests. The underlying principle is that, because nonsynonymous substitutions leading to physicochemically dissimilar amino acid replacements (radical changes) are more likely to be deleterious than those leading to similar amino acid replacements (conservative changes) (Zuckerkandl and Pauling, 1965; Dayhoff et al., 1972), an excess of the former type across genomic regions is evidence for genetic drift as a result of effective population size (N_e) reduction (Zhang, 2000; Eyre-Walker et al., 2002; Hughes and Friedman, 2009; Wernegreen, 2011). Both sides of the causal relationship underlying this drift theory were shown to exist in several important free-living marine bacteria (Luo et al., 2017). For example, the radical nonsynonymous mutations were demonstrated to be under greater functional constraints than the conservative nonsynonymous mutations in all examined marine bacterial populations, indicating the former type of mutations are indeed more deleterious than the latter type (Luo et al., 2017). Further, genome reduction and genomic sequence evolutionary rate acceleration were shown to coincide with an accelerated accumulation of the more deleterious type of mutations (i.e., radical changes) in *Prochlorococcus* and in SAR86, suggesting that genetic drift, rather than natural selection, acted at the time that genome reduction and increased evolutionary rate

occurred in these lineages (Luo et al., 2017).

The hypothesis that increased mutation rate as a result of repair gene losses may play a role in accelerated sequence evolution in free-living marine bacteria is straightforward and was used to explain the accelerated sequence evolution in *Prochlorococcus* (Dufresne et al., 2005). The hypothesis that increased mutation rate may drive genome reduction of marine bacterioplankton lineages was supported by a mathematical model (Marais et al., 2008). However, direct experimental measures of unbiased mutation rates in free-living marine bacterial lineages showing accelerated evolutionary rate and possessing reduced genomes have not been available. With increased mutation rate, not all auxiliary genes can be maintained by purifying selection (Marais et al., 2008); some of these are pseudogenized through disabling mutations and subsequently lost by mutations that are commonly biased towards deletion over insertion (Kuo and Ochman, 2009). Because genome reduction in marine bacterioplankton lineages are often associated with loss of important repair genes (Partensky and Garczarek, 2010; Viklund et al., 2011), elevation of mutation rate is potentially an important mechanism driving genome reduction in marine bacterioplankton lineages (Marais et al., 2008). In addition to repair gene loss, increased mutation rates could also be triggered by error-prone DNA replication under SOS conditions (Maor-Shoshani et al., 2000). However, the relationship between genome reduction and the bacterial SOS response is yet to be explored.

The mutation accumulation (MA) experiment followed by whole-genome sequencing (WGS) of the mutant lines is becoming a standard approach for

spontaneous mutation rate determination, and the MA/WGS data of over 20 model

bacteria have become available (Long et al., 2018). In the MA procedure, the

laboratory populations arising from the same ancestor cell independently go through

repeated single-cell bottlenecks over hundreds to several thousand cell divisions. This

procedure minimizes the ability of natural selection to eliminate deleterious mutations

except the lethal ones and thus is considered an approximately unbiased method for

mutation rate determination (Lynch et al., 2016). Currently the common practice of

creating single-cell bottlenecks is by propagating the population on solid medium.

Unfortunately, most marine bacterioplankton lineages with reduced genomes either

remain uncultivated (e.g., SAR86), do not grow on solid media (e.g., SAR11), or are

practically infeasible for long-term propagation on solid medium (e.g.,

Prochlorococcus), and thus are not suitable for the MA/WGS procedure. Members of

Nonlabens and *Leeuwenhoekiella* grow readily on solid medium and thus are

amenable to the MA/WGS procedure, providing an unprecedented opportunity to

examine the various drivers including mutation rate effect on elevated evolutionary

rate and genome reduction in free-living marine bacteria.

Results and Discussion

Niche specialization of present-day populations in Nonlabens and Leeuwenhoekiella

We performed detailed experimental and bioinformatics analyses on a present-

day representative member SY33080 in the *Nonlabens* lineage and compared it to

strain ZYFB001 in the sister lineage *Leeuwenhoekiella*. Both of them grow on starch,

Tween 20, Tween 40, and Tween 80, but neither use gelatin, casein, cellulose, alginate, and chitin. The API 20NE experiment result shows only *L.* ZYFB001 has β-glucosidase and β-galactosidase activities. Furthermore, the GN2 Biolog analysis showed that *L.* ZYFB001 utilizes 34 more carbon compounds than *N.* SY33080 (Table 1), among which 24 are carbohydrates or organic acids and only the remaining three are amino acids, consistent with the more abundant glycoside hydrolase (GH) genes (103 vs. 26) and more diverse glycoside hydrolase gene families (35 vs. 17) in *L.* ZYFB001 than in *N.* SY33080 (Fig. 1A).

By normalizing the genome size, we found a greater CAZyme gene density (38.27 Mbp^{-1} vs. 25.96 Mbp^{-1}) but a similar peptidase gene density (30.37 Mbp^{-1} vs. 30.88 Mbp^{-1}) in *L.* ZYFB001 compared to *N.* SY33080 (Fig. 1B), giving rise to a starkly elevated peptidase:CAZyme ratio in *N.* SY33080 (1.19) compared to that in *L.* ZYFB001 (0.79). The trend remains consistent when taking into account the remaining genomes of each lineage: the CAZyme gene density in *Leeuwenhoekiella* lineage is significantly higher than that in the *Nonlabens* lineage (Wilcoxon-Mann-Whitney test, *p*<0.001), whereas the peptidase gene density is not significantly different (Wilcoxon-Mann-Whitney test, *p*>0.05). This leads to a significantly elevated peptidase:CAZyme ratio in the *Nonlabens* lineage compared to the *Leeuwenhoekiella* lineage (Wilcoxon-Mann-Whitney test, *p*<0.001). Since marine *Bacteroidetes* taking a primarily planktonic lifestyle tend to have higher peptidase:CAZyme ratios compared to those adopting a primarily algae-associated lifestyle (Xing et al., 2015; Bennke et al., 2016), our results strongly suggest that the

extant members of the *Nonlabens* lineage and the *Leeuwenhoekiella* lineage diversified ecologically towards taking planktonic and algae-associated lifestyles, respectively.

Bacteroidetes utilize complex polysaccharides often through PUL, which consists of *susC*-like TonB-dependent receptors (TBDRs), paired *susD*-like genes, CAZymes, and auxiliary genes such as sulfatase and regulatory genes (Bennke et al., 2016). Among these functional units, SusD binds to the degraded polysaccharides and bring them to TBDR for transportation into the cell (Sonnenburg et al., 2010). *L*. ZYFB001 possesses 59 TBDR genes, 34 of which are co-located with *susD*-like genes. By contrast, *N*. SY33080 only has 22 TBDR genes, two of which form partnership with *susD*-like genes. We identified 15 functional PULs potentially involved in the utilization of a diverse array of polysaccharides in *L*. ZYFB001 but only one PUL in *N*. SY33080 (Table S1).

In summary, *N*. SY33080 possesses a reduced genome, features low catabolic potential for polysaccharides, paralleling the planktonic lifestyle of several model marine flavobacteria strains such as *Polaribacter* sp. MED152, *Polaribacter* sp. Hel1 33 49, and *Dokdonia donghaensis* MED134 (González et al., 2008; González et al., 2011; Xing et al., 2015). In contrast, *L*. ZYFB001 possesses a larger genome and features highly diverse and abundant genes involved in polysaccharide utilization, likened to the algae-associated paradigm of other model marine flavobacteria strains such as *Zobellia galactanivorans* Dsij[T], *Gramella forsetii* KT0803, *Formosa agariphila* KMM 3901, and *Polaribacter* sp. Hel1 85 (Mann et al., 2013; Kabisch et

al., 2014; Xing et al., 2015; Barbeyron et al., 2016).

Niche separation of Leeuwenhoekiella and Nonlabens occurred anciently and accompanied by genome expansion and reduction, respectively

Among the three sister lineages, *Nonlabens* branched off prior to the diversification of *Leeuwenhoekiella* from *Dokdonia* (Fig. 2, Fig. S1). Ancestral genome reconstruction inferred that the LCA (N33) of *Nonlabens*, *Dokdonia* and *Leeuwenhoekiella* carried a genome with 2,216 gene families. It underwent distinct evolutionary processes toward the LCA of *Nonlabens* (N19) and toward the LCA of *Leeuwenhoekiella* (N31). The former reduced the genome size by 3.5%, whereas the latter experienced a stepwise genome expansion via the ancestral node of N32 (LCA of *Dokdonia* and *Leeuwenhoekiella*) by a net increase of 12.5% (Fig. 2, Table S2). Further diversification within *Nonlabens* was not associated with a clear trend of genome reduction, though an overall trend of expansion was reconstructed along the evolution of *Leeuwenhoekiella* (Table S2). Therefore, the dichotomy of lifestyle between extant members of *Nonlabens* and *Leeuwenhoekiella* is a result of two independent processes including early genome reduction in the former lineage and persistent genome expansion in the latter lineage.

Among the 223 gene families lost at the LCA (N19) of *Nonlabens* (Table S3), 173 were consistently annotated by at least two databases. Many of them are involved in the degradation of carbohydrates derived from marine macroalgae (Fig. 2). For example, GH3 family was lost, which consists of β-glucosidase and 1,4-β-xylosidase for degrading β-glucan in brown algae (Kabisch et al., 2014) and β-xylan in green

macroalgae (Domozych et al., 2012), respectively. A second example is the loss of GH32 family with endo- and exo-type of levanase and invertase activities targeting fructan and sucrose, respectively (Versluys et al., 2018). While these carbohydrates are largely found in terrestrial plants including vegetables, fruits and crops (Huber, 1989; Muir et al., 2007; van Arkel et al., 2013), fructan and sucrose were also found in certain species of green macroalgae (e.g. *Cladophorales*) and brown algae (e.g. *Laminaria*), respectively (Nwachukwu and Chukwu, 2012; Versluys et al., 2018). A third example is the loss of trehalase, which enables the utilization of trehalose, the common algal disaccharide accumulated in red algae (Stewart, 1974). A fourth example is the loss of multiple genes responsible for the utilization of xylose, the major component of xylan found in the cell wall of green macroalgae and red algae (Turvey and Williams, 1970; Huizing and Rietema, 1975). This includes the loss of xylose transporter, xylose isomerases catalyzing the interconversion between D-xylose and D-xylulose, and D-xylulose kinase phosphorylating D-xylulose for further degradation. A fifth example is the loss of genes encoding α-amylase and extracellular starch binding protein for the utilization of starch, which is accumulated as granule in red algae (Viola et al., 2001). Furthermore, utilization of polysaccharides is often facilitated by the coordinated activities of genes clustered in PUL, which consists of CAZymes and two types of PUL signature genes, the *susC*-like TonB-dependent receptors and the *susD*-like outer membrane proteins (Bennke et al., 2016). The losses of several copies of both *susC* and *susD* genes further reduce the potential of *Nonlabens* LCA to utilize macroalgae derived polysaccharides.

On the other hand, 145 gene families were predicted to be gained at the LCA of *Nonlabens* (Table S4), of which only 61 were annotated by two or more databases. Among these, a few are involved in the interaction with microalgae. The genes gained at N19 include GH17 and GH30, which may have β-1,3-glucosidase and β-1,6-glucanase activities, respectively (Fig. S2A & B). They are often adjacently located within PUL for the degradation of laminarin or chrysolaminarin, with GH30 firstly cleaving β-1,6-glucose side chains followed by cleaving the remaining 1,3 bonds via GH17 for further utilization (Kappelmann et al., 2019). Laminarin and chrysolaminarin are used as the storage polysaccharides in brown algae and diatoms, respectively (Kappelmann et al., 2019). Although the backbones of laminarin and chrysolaminarin have the same chemical composition, the former is capped at the reducing end by a D-mannitol (Read et al., 1996), which needs to be removed through the activity of mannitol dehydrogenase as one step of laminarin degradation (Kappelmann et al., 2019). The absence of the mannitol dehydrogenase genes in the LCA of *Nonlabens* suggests that these two glycoside hydrolase genes (GH17 and GH30) were likely involved in the utilization of chrysolaminarin, which makes up to 80% of dry weight in diatom (Gügi et al., 2015). Another example of gain at N19 is the two β-mannosidase gene families in GH2 (Fig. S2C), which produces β-D-mannose residues from hydrolyzing β-D-mannosides. Oligomannosides are the dominant unit linked to membrane-bound proteins through *N*-glycosidic bond in microalgae (Mathieu-Rivet et al., 2013). The gained genes also include GH25, which exclusively encodes a lysozyme for peptidoglycan degradation. There is evidence that

GH25 breaks down the cell wall of blue-green algae, allowing the bacteria such as *Cytophaga* (another major lineage of *Bacteroidetes*) to utilize the algal protoplasmic nutrients (Cole, 1982).

Other important gains at N19 include the gene encoding a carotenoid biosynthesis protein (PF04240 in Pfam) (Fig. 2). Carotenoids are likely among the earliest light-absorbing compounds and are well known for their photoprotective function against UV radiation (Pinhassi et al., 2016). Consequently, the gain of carotenoid synthesis gene may facilitate niche transition to the pelagic water where cells are more exposed to UV light. Another interesting prediction was that *crtI* for β-carotene synthesis experienced an expansion of copy number at N19 via horizontal gene transfer (Fig. S3). β-carotene is one type of carotenoid, which is used for retinal synthesis, an essential component of rhodopsin (Pinhassi et al., 2016). In oligotrophic pelagic environments where labile organic matter is commonly depleted, rhodopsin-containing flavobacteria utilize light as one source of energy to power cellular activities (Yoshizawa et al., 2014). While three types of rhodopsins, including proton-pumping proteorhodopsin (PR), sodium-pumping rhodopsin (NaR) and chloride-pumping rhodopsin (ClR) (Béja et al., 2000; Inoue et al., 2013; Yoshizawa et al., 2014), have been identified in *Nonlabens*, none was predicted to be gained at the LCA (N19) of the *Nonlabens* lineage (Fig. S4). Therefore, acquisition of an additional copy of *crtI* could be a case of pre-adaptation in early *Nonlabens* and facilitate the transition to rhodopsin-based photoheterotrophy in the later stage of their evolution.

Selection efficiency was not reduced despite the genome-wide accelerated sequence

evolutionary rate in Nonlabens compared to that in Leeuwenhoekiella

An important observation is that genome reduction in the *Nonlabens* lineage is associated with acceleration of amino acid sequence evolution across genomic regions. This is supported by the severely violated molecular clock between these two lineages (likelihood ratio test, $p<0.001$), and is shown by the extraordinarily longer branches connecting the LCA (N33) of the *Nonlabens*, *Dokdonia* and *Leeuwenhoekiella* lineages to the extant *Nonlabens* compared the branches connecting this LCA to the extant *Leeuwenhoekiella* (Fig. 2). A higher rate of sequence evolution at genome-wide scale often results from genetic drift, or increased mutation rate, or a combination of these factors (Graur et al., 2016). As genome reduction of *Nonlabens* largely occurred on the long ancestral branch leading to the LCA (N19) of the *Nonlabens* lineage, the power of genetic drift was tested on this branch.

Given this genome reduction event having occurred very anciently, the relative rate of radical versus conservative nonsynonymous substitutions (the d_R/d_C ratio) across the genomic regions can be a proxy for N_e (Zhang, 2000; Eyre-Walker et al., 2002; Hughes and Friedman, 2009; Wernegreen, 2011; Luo et al., 2017). The potential change of N_e associated with this genome reduction event can be approximated by comparing the d_R/d_C on the branch leading to the LCA of the *Nonlabens* lineage to that of its sister branch giving rise to the LCA (N32) of the *Dokdonia* and *Leeuwenhoekiella* lineages. Using the outgroup genomes (Fig. S1) as the reference, pairwise comparisons across 492 single-copy orthologous genes were performed between the target clade (*Nonlabens*) and the reference genomes, as well as between

the control clade (*Leeuwenhoekiella* and *Dokdonia*) and the reference genomes,

respectively. These d_R/d_C analyses led to opposite interpretations when different

classification schemes of the 20 amino acids were used (Fig. S5A, Table S5). While

the genome-wide d_R/d_C ratio was consistently greater on the ancestral branch leading

to the N19 than that on the branch leading to N32 based on volume and polarity, the

d_R/d_C ratio was instead consistently inflated on the ancestral branch leading to N32

when amino acids were classified according to charge. As genomes of *Dokdonia* are

smaller and more biased toward low GC content compared to *Leeuwenhoekiella*,

which might affect the result, the above analysis was also done after the *Dokdonia*

lineage was excluded and the results remained qualitatively the same (Fig. S5B). The

opposite results of d_R/d_C depending on the amino acid classification scheme suggest

that genetic drift was likely not a primary mechanism driving genome reduction of the

Nonlabens lineage.

Spontaneous mutation rate of the extant Nonlabens is not significantly greater than
that of the extant Leeuwenhoekiella

As discussed above, increased mutation rate is likely another major mechanism

leading to acceleration of sequence evolution and genome reduction, so we employed

the MA/WGS approach to determine the unbiased spontaneous mutation rate of

ZYFB001 in *Leeuwenhoekiella* and of SY33080 in *Nonlabens*. Base-substitution

mutations of both organisms are biased toward intergenic regions compared to coding

regions ($\chi2$ test; $p<0.05$ for *N.* SY33080 and $p<0.001$ for *L.* ZYFB001; Table S6).

This bias is commonly found in other bacteria including the model marine bacterium

Ruegeria pomeroyi DSS-3 in Alphaproteobacteria, which likely results from more active mismatch repair on coding regions or the mutation-prone nature of repetitive DNA in intergenic regions (Lee et al., 2012; Long et al., 2014; Dillon and Cooper, 2016; Sun et al., 2017). Both strains show a significantly greater rate of base-substitution mutations from G/C to A/T than from A/T to G/C (prop.test function of R, $p<0.001$, df=1) (Fig. S6), and this mutational bias is found in most bacterial species that were subject to mutation accumulation experiments (Long et al., 2018). In both organisms, the nonsynonymous to synonymous base-substitution rate ratio is not significantly different from the random expectation ($\chi 2$ test; $p>0.05$; Table S6). Thus, selection does not appear to have a significant impact on the accumulation of mutations in both organisms. This conclusion is further strengthened by the randomly distributed mutations across the genomic regions in all lines (Fig. 3 & Fig. S6) and the comparable number of mutations accumulated among all protein-coding genes in each organism (Fig. 3), the latter supported by a statistical test showing that no genes accumulated an excess of mutations (bootstrapping test, $p>0.05$ for each organism). We determined 366 base-substitution mutations over 2,549 cell divisions in *N.* SY33080 and 604 base-substitution mutations over 3,006 cell divisions in *L.* ZYFB001, leading to a mutation rate of 5.49×10^{-10} and 5.21×10^{-10} per base per cell division for *N.* SY33080 and *L.* ZYFB001, respectively (Table S7 & S8). These mutation rates fall in the middle of the variation determined for other bacteria using MA/WGS (Long et al., 2018). Despite a slight increase, the mutation rate of *N.* SY33080 is not significantly different from that of *L.* ZYFB001 (Kolmogorov-

Smirnov test and Wilcoxon-Mann-Whitney test; $p>0.05$).

Given the same mutation rate per cell division, a more rapidly growing bacterial lineage may give rise to increased evolutionary rate, presumably because their cells underwent more cell divisions and thus collected more mutations (Weller and Wu, 2015; Gibson and Eyre-Walker, 2019). However, the generation time estimation for heterotrophic marine bacteria in the wild has been highly controversial depending on the experimental methods that marine microbial ecologists use, and it is even more problematic to accurately quantify the generation time of a particular heterotrophic taxon and differentiate it among closely related heterotrophic lineages in the wild (Kirchman, 2016). The available lineage-specific growth rate data all compare the generation time of deeply-branching groups, such as the oligotrophic and often free-living groups including the SAR11 clade and the SAR86 clade, as well as the copiotrophic and often algae-associated groups including *Bacteroidetes*, *Rhodobacteraceae*, and *Alteromonadaceae*, and they consistently support that the oligotrophic groups have significantly longer generation time than the copiotrophic groups (Teira et al., 2009; Bakenhus et al., 2017; Sánchez et al., 2017). Because the *Nonlabens* and *Leeuwenhoekiella* lineages are aligned to the oligotrophic and copiotrophic models, respectively, it is reasonable to postulate that members of the *Nonlabens* lineage have longer generation time than members of the *Leeuwenhoekiella* lineage in the ocean. Further, a significantly longer generation of the *N*. SY33080 than that of *L*. ZYFB001 was also observed under optimal laboratory conditions throughout the MA experiment (Wilcoxon-Mann-Whitney test, $p<0.001$)

(Fig. S7). Taken together, *Nonlabens* may have longer generation time than *Leeuwenhoekiella* under both wild and laboratory conditions, and thus the generation time effect is unlikely responsible for the accelerated rate of sequence evolution in the *Nonlabens* lineage.

Losses of DNA repair genes correlated with the increased evolutionary rate of Nonlabens

We noted the above mutation rate measurements were derived from the present-day organisms and are not necessarily applicable to their ancient ancestors, such as the LCA of the *Nonlabens* lineage and the LCA of the *Leeuwenhoekiella* lineage. Because mutator phenotypes (i.e., strains with increased mutation rate) often result from loss of important DNA repair genes that ensure accuracy of DNA replication (Denamur et al., 2000; Batut et al., 2014) or gains of error-prone repair systems (Ysern et al., 1990; Woodgate et al., 1994), gains and losses of these genes can be used as proxies for mutation rate increase in ancestral lineages.

We inferred that three repair genes (*dnaQ*, *umuC* and *umuD*) were lost on the ancestral branch leading to the LCA (N19) of the *Nonlabens* lineage (Fig. 2 & S8). In contrast, only one DNA repair gene *recO* was inferred to be lost on the branch leading to the LCA (N32) of the *Dokdoni*a and *Leeuwenhoekiella* lineages, which is part of the RecF pathway for single-stranded DNA repair (Bork et al., 2001) (Fig. S8). Among the three genes lost on the ancestral node N19, *umuC* and *umuD* encode subunits of the error-prone DNA polymerase V which is highly mutagenic and produces point mutations at a frequency of 2.1×10^{-4} per nucleotide, ~40 times greater

than DNA polymerase III holoenzyme (Maor-Shoshani et al., 2000). They were firstly identified to be involved in SOS response in *E. coli* (Sutton et al., 2000). The proteins encoded by *umuC* and *umuD* constitute the UmuD'2C complex to perform DNA translesion synthesis (TLS), which allows DNA replication to bypass the lethal DNA damage (Sutton and Walker, 2001). It was proposed that two major types of DNA lesions, TT (6-4) photoproducts and TT cis-syn photodimers, may be caused by UV radiation (Tang et al., 2000).

According to theory, high mutation rate is favored by selection in a harsh environment because this increases the chance of acquiring adaptive mutations. Following adaptive events, high mutation rate is disfavored because of the cost of increased accumulation of deleterious mutations (Taddei et al., 1997; Tenaillon et al., 1999; Denamur et al., 2000). This may explain the loss of *umuC* and *umuD* at the LCA (N19) of the *Nonlabens* lineage. Another evidence that SOS response may have been involved during the ecological transition on the branch leading to the LCA of *Nonlabens* is the gain of SOS response associated peptidase (SRAP) gene family on the same branch (Fig. 2), which is involved in the induction of bacterial SOS response (Aravind et al., 2013).

The third repair gene predicted to be lost in the LCA of the *Nonlabens* lineage is *dnaQ* (also called *mutD*; Fig. 2), which encodes the epsilon subunit of DNA polymerase III. It functions as a 3'→5' DNA directed proofreading exonuclease and removes incorrectly incorporated bases during replication (Scheuermann and Echols, 1984), and thus is different from other DNA repair genes acting on replicated DNA

strand(s). Interestingly, it was shown that the proofreading activity of *dnaQ* is inhibited by SOS-induced UV mutagenesis (Woodgate et al., 1987), and that overexpression of *dnaQ* inhibits the SOS mutator activities (Jonczyk et al., 1988). We therefore hypothesized that the loss of *dnaQ* occurred early during the ecological transition, which facilitated the SOS response.

Concluding remarks

Our comprehensive genomic analyses and physiological assays suggest that members of the *Nonlabens* lineage and those of the *Leeuwenhoekiella* lineage take distinct ecological strategies, with the former likely occupying pelagic water niches and the latter favoring to be associated with macroalgae. Ancestral genome reconstruction suggests that this ecological differentiation was initiated early when these two lineages started to diverge from each other. The genome-wide elevated evolutionary rate of *Nonlabens* compared to *Leeuwenhoekiella* provides us an opportunity to investigate the evolutionary forces underlying the niche differentiation between the two lineages. Although genetic drift and/or increased mutation rate are often taken to account for an elevated evolutionary rate at genomic scale in microbial lineages (Itoh et al., 2002; McCutcheon and Moran, 2012; Batut et al., 2014; Wernegreen, 2015), these mechanisms were not supported here based on our rigorous experimental and bioinformatic tests and thus may not explain the accelerated evolutionary rates in *Nonlabens*. However, using evolutionary gains and losses of DNA repair genes as a proxy to represent potential mutation rate changes during the ancestral processes, we proposed that the evolutionary changes of genes involved in

SOS response during the ecological transition from the protected macroalgal habitat to the UV-exposed pelagic niches likely gave rise to the transiently increased mutation rate in the early evolutionary history of *Nonlabens*, which may explain the accelerated evolutionary rate of this lineage.

According to the inferred evolutionary mechanisms, we propose an ecological scenario during the ancient niche transition. The primitive *Nonlabens* cells prior to the LCA of this lineage may have colonized macroalgae which provided a stable source of nutrients to the bacteria, but environmental fluctuations may have made the nutrient sources less predictable. To cope with this stress, some cells began to explore pelagic environments including the nutrient-poor bulk seawater and occasionally phytoplankton niches, where cells with reduced replication burden were favored. During this transition, the cells were subject to greater UV radiation, which caused severe DNA damage such as lesions and subsequently triggered bacterial SOS response. An important consequence was that these ancestral bacteria became mutator; they had an increased chance to evolve adaptive mutations, an increased frequency of accepting genes from external sources, and an increased rate of losing DNA. These genomic changes led to acquisitions of novel functions and reduction of replication costs, which enabled the cells to adapt in pelagic environments. After the lifestyle was switched, the previously important genes like those related to polysaccharide catabolism became dispensable and eventually lost. The proposed evolutionary mechanisms and the related ecological scenario still hold when an updated genome set that contains a few uncultivated members represented by

metagenome-assembled genomes (MAGs) was used, as the genes that likely facilitated the niche separation and SOS response show similar phyletic pattern (Fig. S9 and Fig. 2). A more accurate delineation requires a more thorough sampling of the phylogenetically and ecologically diverse members within these lineages.

Materials and Methods

Leeuwenhoekiella sp. ZYFB001 and *Nonlabens* sp. SY33080 were isolated from the surface marine snow particles in Yellow Sea (121.50 °E, 32.59 °N) and surface seawater in East China Sea (126.51 °E, 28.43 °N), respectively. The potential of the two isolates to utilize various substrates were tested using specialized media, API 20NE strips (BioMérieux), and GN2 MicroPlates Biolog (BIOLOG, Catalog #1011) (See Supplementary Methods *Section 2*).

Whole genome shotgun sequencing of *L.* ZYFB001 and *N.* SY33080 were carried out using both PacBio RS II and Illumina HiSeq 2500 sequencing platforms. The correction, trimming, and assembly of long reads derived from PacBio with Canu 1.4 (Koren et al., 2017) yielded a single contig of 4,053,756 bp for *L.* ZYFB001 and 2,852,234 bp for *N.* SY33080, respectively. These assemblies were further polished by the Illumina reads using BWA (Li and Durbin, 2009) and Pilon (Walker et al., 2014).

A total of 106 publically available *Flavobacteriaceae* genomes were downloaded from Genbank in December 2018, and they were chosen according to their phylogenetic positions in a comprehensive phylogenomic tree of *Flavobacteriaceae*

(Zhang et al., 2019). Among these, *Leeuwenhoekiella* sp. Hel I 48 was recently reclassified into genus *Flavimarina* due to its greater divergence of the 16S rRNA gene sequence (8.4-9.6%) to other *Leeuwenhoekiella* strains (Nedashkovskaya et al., 2014). Genome annotation was performed with in house resources including the Pfam v30.0 (El-Gebali et al., 2018) and CDD v3.15 (Marchler-Bauer et al., 2016) databases using HMMER v3.1 (Eddy, 2009) and RPS-BLAST v2.6.0 (Camacho et al., 2009), respectively, as well as with online resources including the subsystem and KEGG databases via RAST (Overbeek et al., 2013) and BlastKOALA (Kanehisa et al., 2016) online tools, respectively. Genes encoding CAZymes and peptidase were further annotated with dbCAN2 (Zhang et al., 2018) and the specialized MEROPS v12.0 database (Rawlings et al., 2017), respectively. Putative DNA repair genes were identified based on our in-house repair gene database via reciprocal best BLAST hits. The functional polysaccharide utilization locus (PUL) was predicted based on the presence of some essential features: tandem *susCD*-like pairs and adjacent CAZyme genes using in-house scripts (Terrapon et al., 2015) (See Supplementary Methods *Section 3*).

Homologous gene families of the 108 genomes were identified using OrthoFinder v2.2.1 (Emms and Kelly, 2015). The concatenated amino acid sequence of 163 shared single-copy gene families was used to construct a maximum likelihood phylogenomic tree using IQ-TREE v1.6.2 (Nguyen et al., 2014). The ancestral states and the evolutionary history of each identified orthologous gene family were estimated using the modified Sankoff parsimony algorithm implemented in the

software BadiRate, which takes into account both the branch length of the phylogenomic tree and the uncertainty in the number of family members (Librado et al., 2011) (See Supplementary Methods *Section 4 and Section 5*).

The molecular clock hypothesis between the *Nonlabens* and *Leeuwenhoekiella* lineages was tested using the codeml program implemented in the PAML package (Yang, 2007). To test whether genetic drift was associated with the ancestral genome reduction of the *Nonlabens* lineage, we determined the rates of conservative (d_C) and radical (d_R) changes using methods controlling for the disparity in GC content implemented in RCCalculator (Luo et al., 2017) for each of the 492 single-copy gene families shared by at least 105 (out of 108) genomes. As *Dokdonia* is embedded within the phylogenetic group consisting of *Nonlabens* and *Leeuwenhoekiella*, there is a possibility that it interferes with the d_R/d_C analysis in comparing *Nonlabens* and *Leeuwenhoekiella*. We thus performed the d_R/d_C analysis by either including or excluding *Dokdonia*. To determine the spontaneous mutation rate of *L.* ZYFB001 and *N.* SY33080, we applied the MA/WGS procedure to each of them under optimal condition. In summary, 96 independent MA lines were initiated from a single founder colony of *L.* ZYFB001 and *N.* SY33080, respectively. Every 48 h, cells from a single colony were transferred to a fresh plate. This transfer continued for 232 days, equivalent to 3,006 and 2,549 cell divisions for *L.* ZYFB001 and *N.* SY33080, respectively. Generation time between each transfer was estimated by colony forming units (CFU) every 45 days. Mutation rate μ_{bs} (per base per generation) was calculated by the following equation:

$$\mu_{bs} = \frac{m}{\sum_1^n N \times T}$$

where m is the total number of observed base-substitution mutations across all the MA lines, n is the total number of lines, N is the analyzed sites for each MA line, and T is the number of cell divisions that occurred in the MA study (Sung et al., 2012; Long et al., 2016) (See Supplementary Methods *Section 7, Section 8 and Section 9*).

Data availability

Genomic sequences of *Nonlabens* sp. SY33080 and *Leeuwenhoekiella* sp. ZYFB001 are available at the NCBI GenBank database under the accession number CP050461 and CP050460, respectively. The manually compiled DNA repair gene database as well as the scripts used for ancestral genome reconstruction and mutation rate calculation have been deposited in the online repository (https://github.com/zoopera/flavoevo/).

Conflict of interests

The author declares no competing interests.

Reference

Aravind, L., Anand, S., and Iyer, L.M. (2013) Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol Direct* **8**: 20.

Bakenhus, I., Dlugosch, L., Billerbeck, S., Giebel, H.A., Milke, F., and Simon, M. (2017) Composition of Total and Cell-Proliferating Bacterioplankton Community in Early Summer in the North Sea-Roseobacters Are the Most Active Component. *Front Microbiol* **8**: 1771.

Barbeyron, T., Thomas, F., Barbe, V., Teeling, H., Schenowitz, C., Dossat, C. et al. (2016) Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the

model algae-associated bacterium *Zobellia galactanivorans* Dsij^T. *Environ Microbiol* **18**: 4610-4627.

Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol* **12**: 841.

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E. et al. (2006) Whole genome analysis of the marine *Bacteroidetes* 'Gramella forsetii'reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**: 2201-2213.

Béja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P. et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902-1906.

Bennke, C.M., Krüger, K., Kappelmann, L., Huang, S., Gobet, A., Schüler, M. et al. (2016) Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic. *Environ Microbiol* **18**: 4456-4470.

Bork, J.M., Cox, M.M., and Inman, R.B. (2001) The RecOR proteins modulate RecA protein function

at 5′ ends of single-stranded DNA. *EMBO J* **20**: 7313-7322.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Cole, J.J. (1982) Interactions between bacteria and algae in aquatic ecosystems. *Annu Rev Ecol Syst* **13**: 291-314.

Cottrell, M.T., and Kirchman, D.L. (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Appl Environ Microbiol* **66**: 5116-5122.

Dayhoff, M., Eck, R., and Park, C. (1972) A model of evolutionary change in proteins. In: Dayhoff M. (ed). In *Atlas of protein sequence and structure*. National Biomedical Research Foundation: Silver Spring, MD, USA, pp. 89-99.

Denamur, E., Lecointre, G., Darlu, P., Tenaillon, O., Acquaviva, C., Sayada, C. et al. (2000) Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* **103**: 711-721.

Dillon, M.M., and Cooper, V.S. (2016) The fitness effects of spontaneous mutations nearly unseen by selection in a bacterium with multiple chromosomes. *Genetics* **204**: 1225-1238.

Domozych, D., Ciancia, M., Fangel, J.U., Mikkelsen, M.D., Ulvskov, P., and Willats, W.G. (2012) The cell walls of green algae: a journey through evolution and diversity. *Front Plant Sci* **3**: 82.

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

Eddy, S.R. (2009) A new generation of homology search tools based on probabilistic inference. In *Genome Informatics 2009: Genome Informatics Series Vol 23*: World Scientific, pp. 205-211.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C. et al. (2018) The Pfam protein families database in 2019. *Nucleic Acids Res* **47**: D427-D432.

Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.

Eyre-Walker, A., Keightley, P.D., Smith, N.G., and Gaffney, D. (2002) Quantifying the slightly deleterious mutation model of molecular evolution. *Mol Biol Evol* **19**: 2142-2149.

Fenchel, T. (1988) Marine plankton food chains. *Annu Rev Ecol Syst* **19**: 19-38.

Fernández-Gomez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedros-Alio, C. (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* **7**: 1026-1037.

Fuhrman, J., and Azam, F. (1982) Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results. *Mar Biol* **66**: 109-120.

Gibson, B., and Eyre-Walker, A. (2019) Investigating evolutionary rate variation in bacteria. *J Mol Evol* **87**: 317-326.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D. et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242-1245.

González, J.M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M. et al. (2008) Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**: 8724-8729.

González, J.M., Pinhassi, J., Fernández-Gómez, B., Coll-Lladó, M., González-Velázquez, M., Puigbò, P. et al. (2011) Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp. strain MED134. *Appl Environ Microbiol* **77**: 8676-8686.

Graur, D., Sater, A.K., and Cooper, T.F. (2016) Rates and Patterns of Molecular Evolution. In *Molecular and genome evolution*: Sinauer Associates, Incorporated.

Gügi, B., Le Costaouec, T., Burel, C., Lerouge, P., Helbert, W., and Bardor, M. (2015) Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Mar Drugs* **13**: 5993-6018.

Huber, S.C. (1989) Biochemical mechanism for regulation of sucrose accumulation in leaves during photosynthesis. *Plant Physiol* **91**: 656-662.

Hughes, A.L., and Friedman, R. (2009) More radical amino acid replacements in primates than in rodents: support for the evolutionary role of effective population size. *Gene* **440**: 50-56.

Huizing, H., and Rietema, H. (1975) Xylan and mannan as cell wall constituents of different stages in

the life-histories of some siphoneous green algae. *British Phycological Journal* **10**: 13-16.

Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., and Kandori, H. (2013) A light-driven sodium ion pump in marine bacteria. *Nat Commun* **4**: 1678.

Itoh, T., Martin, W., and Nei, M. (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci USA* **99**: 12944-12948.

Jonczyk, P., Fijalkowska, I., and Ciesla, Z. (1988) Overproduction of the epsilon subunit of DNA polymerase III counteracts the SOS mutagenic response of *Escherichia coli*. *Proc Natl Acad Sci USA* **85**: 9124-9127.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M. et al. (2014) Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* '*Gramella forsetii*'KT0803. *ISME J* **8**: 1492-1502.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726-731.

Kappelmann, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F. et al. (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* **13**: 76-91.

Kirchman, D.L. (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* **39**: 91-100.

Kirchman, D.L. (2016) Growth rates of microbes in the oceans. *Ann Rev Mar Sci* **8**: 285-309.

Kumagai, Y., Yoshizawa, S., Nakajima, Y., Watanabe, M., Fukunaga, T., Ogura, Y. et al. (2018) Solar-panel and parasol strategies shape the proteorhodopsin distribution pattern in marine *Flavobacteriia*. *ISME J*: 1.

Kuo, C.H., and Ochman, H. (2009) Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct* **4**: 35.

Kwon, S.-K., Kim, B.K., Song, J.Y., Kwak, M.-J., Lee, C.H., Yoon, J.H. et al. (2013) Genomic makeup of the marine Flavobacterium *Nonlabens (Donghaeana) dokdonensis* and identification of a novel class of rhodopsins. *Genome Biol Evol* **5**: 187-199.

Lau, S.C., Tsoi, M.M., Li, X., Plakhotnikova, I., Dobretsov, S.V., Wu, M. et al. (2006) *Stenothermobacter spongiae* gen. nov., sp. nov., a novel member of the family *Flavobacteriaceae* isolated from a marine sponge in the Bahamas, and emended description of *Nonlabens tegetincola*. *Int J Syst Evol Microbiol* **56**: 181-185.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci USA* **109**: E2774-E2783.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Librado, P., Vieira, F., and Rozas, J. (2011) BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* **28**: 279-281.

Long, H., Behringer, M.G., Williams, E., Te, R., and Lynch, M. (2016) Similar Mutation Rates but

Highly Diverse Mutation Spectra in Ascomycete and Basidiomycete Yeasts. *Genome Biol Evol* **8**: 3815-3821.

Long, H., Sung, W., Miller, S.F., Ackerman, M.S., Doak, T.G., and Lynch, M. (2014) Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol* **7**: 262-271.

Long, H., Sung, W., Kucukyildirim, S., Williams, E., Miller, S.F., Guo, W. et al. (2018) Evolutionary determinants of genome-wide nucleotide composition. *Nat Ecol Evol*: 1.

Luo, H., Huang, Y., Stepanauskas, R., and Tang, J. (2017) Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. *Nat Microbiol* **2**: 17091.

Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K., and Foster, P.L. (2016) Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **17**: 704-714.

Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T. et al. (2013) The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol* **79**: 6813-6822.

Maor-Shoshani, A., Reuven, N.B., Tomer, G., and Livneh, Z. (2000) Highly mutagenic replication by DNA polymerase V (UmuC) provides a mechanistic basis for SOS untargeted mutagenesis. *Proc Natl Acad Sci USA* **97**: 565-570.

Marais, G.A., Calteau, A., and Tenaillon, O. (2008) Mutation rate and genome reduction in endosymbiotic and free-living bacteria. *Genetica* **134**: 205-210.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S. et al. (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200-D203.

Mathieu-Rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Le Mauff, F. et al. (2013) Exploring the N-glycosylation pathway in Chlamydomonas reinhardtii unravels novel complex structures. *Mol Cell Proteomics* **12**: 3160-3183.

McCutcheon, J.P., and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13.

Muir, J.G., Shepherd, S.J., Rosella, O., Rose, R., Barrett, J.S., and Gibson, P.R. (2007) Fructan and free fructose content of common Australian vegetables and fruit. *J Agric Food Chem* **55**: 6619-6627.

Nedashkovskaya, O.I., Kukhlevskiy, A.D., Zhukova, N.V., and Kim, S.B. (2014) *Flavimarina pacifica* gen. nov., sp. nov., a new marine bacterium of the family *Flavobacteriaceae*, and emended descriptions of the genus *Leeuwenhoekiella*, *Leeuwenhoekiella aequorea* and *Leeuwenhoekiella marinoflava*. *Antonie Van Leeuwenhoek* **106**: 421-429.

Nedashkovskaya, O.I., Vancanneyt, M., Dawyndt, P., Engelbeen, K., Vandemeulebroecke, K., Cleenwerck, I. et al. (2005) Reclassification of [Cytophaga] marinoflava Reichenbach 1989 as *Leeuwenhoekiella marinoflava* gen. nov., comb. nov. and description of *Leeuwenhoekiella aequorea* sp. nov. *Int J Syst Evol Microbiol* **55**: 1033-1038.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274.

Nwachukwu, A., and Chukwu, M. (2012) The potential of macroalgae as a source of carbohydrates for use in bioethanol fermentation. *Int J Energy Environ* **3**: 761-774.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T. et al. (2013) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**: D206-D214.

Park, S., Yoshizawa, S., Chiura, H.X., Muramatsu, Y., Nakagawa, Y., Kogure, K., and Yokota, A. (2012) Nonlabens marina sp. nov., a novel member of the *Flavobacteriaceae* isolated from the Pacific Ocean. *Antonie Van Leeuwenhoek* **102**: 669-676.

Partensky, F., and Garczarek, L. (2010) *Prochlorococcus*: advantages and limits of minimalism. *Ann Rev Mar Sci* **2**: 305-331.

Pinhassi, J., DeLong, E.F., Béjà, O., González, J.M., and Pedrós-Alió, C. (2016) Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol Mol Biol Rev* **80**: 929-954.

Pinhassi, J., Bowman, J.P., Nedashkovskaya, O.I., Lekunberri, I., Gomez-Consarnau, L., and Pedros-Alio, C. (2006) *Leeuwenhoekiella blandensis* sp. nov., a genome-sequenced marine member of the family *Flavobacteriaceae*. *Int J Syst Evol Microbiol* **56**: 1489-1493.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D. (2017) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**: D624-D632.

Read, S.M., Currie, G., and Bacic, A. (1996) Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydr Res* **281**: 187-201.

Sánchez, O., Koblížek, M., Gasol, J.M., and Ferrera, I. (2017) Effects of grazing, phosphorus and light on the growth rates of major bacterioplankton taxa in the coastal NW Mediterranean. *Environ Microbiol Rep* **9**: 300-309.

Scheuermann, R.H., and Echols, H. (1984) A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proc Natl Acad Sci USA* **81**: 7747-7751.

Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., and Sonnenburg, J.L. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**: 1241-1252.

Sun, Y., Powell, K.E., Sung, W., Lynch, M., Moran, M.A., and Luo, H. (2017) Spontaneous mutations of a model heterotrophic marine bacterium. *ISME J*.

Sung, W., Ackerman, M.S., Miller, S.F., Doak, T.G., and Lynch, M. (2012) Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* **109**: 18488-18492.

Sutton, M.D., and Walker, G.C. (2001) Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc Natl Acad Sci USA* **98**: 8342-8349.

Sutton, M.D., Smith, B.T., Godoy, V.G., and Walker, G.C. (2000) The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu Rev Genet* **34**.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997) Role

of mutator alleles in adaptive evolution. *Nature* **387**: 700.

Tang, M., Pham, P., Shen, X., Taylor, J.S., O'donnell, M., Woodgate, R., and Goodman, M.F. (2000) Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404**: 1014.

Teeling, H., Fuchs, B.M., Bennke, C.M., Krueger, K., Chafee, M., Kappelmann, L. et al. (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* **5**: e11888.

Teira, E., Martínez-García, S., Lønborg, C., and Álvarez-Salgado, X.A. (2009) Growth rates of

different phylogenetic bacterioplankton groups in a coastal upwelling system. *Environ Microbiol Rep* **1**: 545-554.

Tenaillon, O., Toupance, B., Le Nagard, H., Taddei, F., and Godelle, B. (1999) Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**: 485-493.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**: 647-655.

Turvey, J., and Williams, E. (1970) The structures of some xylans from red algae. *Phytochemistry* **9**: 2383-2388.

van Arkel, J., Sévenier, R., Hakkert, J.C., Bouwmeester, H.J., Koops, A.J., and van der Meer, I.M. (2013) Tailor-made fructan synthesis in plants: a review. *Carbohydr Polym* **93**: 48-56.

Versluys, M., Kirtel, O., Toksoy Öner, E., and Van den Ende, W. (2018) The fructan syndrome: evolutionary aspects and common themes among plants and microbes. *Plant Cell Environ* **41**: 16-38.

Viklund, J., Ettema, T.J., and Andersson, S.G. (2011) Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29**: 599-615.

Viola, R., Nyvall, P., and Pedersén, M. (2001) The unique features of starch metabolism in red algae. *Proc R Soc Lond B Biol Sci* **268**: 1417-1422.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**.

Weller, C., and Wu, M. (2015) A generation-time effect on the rate of molecular evolution in bacteria.

Evolution (N Y) **69**: 643-652.

Wernegreen, J.J. (2011) Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. *PLoS ONE* **6**: e28905.

Wernegreen, J.J. (2015) Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann N Y Acad Sci* **1360**: 16.

Woodgate, R., Bridges, B., Herrera, G., and Blanco, M. (1987) Mutagenic DNA repair in *Escherichia coli* XIII Proofreading exonuclease of DNA polymerase III holoenzyme is not operational during UV mutagenesis. *Mutation Research/DNA Repair Reports* **183**: 31-37.

Woodgate, R., Singh, M., Kulaeva, O.I., Frank, E.G., Levine, A.S., and Koch, W.H. (1994) Isolation and characterization of novel plasmid-encoded umuC mutants. *J Bacteriol* **176**: 5011-5021.

Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. et al. (2015) Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9**: 1410-1422.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-1591.

Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W. et al. (2014) Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc Natl Acad Sci USA* **111**: 6732-6737.

Ysern, P., Clerch, B., Castańo, M., Gibert, I., Barbé, J., and Llagostera, M. (1990) Induction of SOS genes in *Escherichia coli* and mutagenesis in Salmonella typhimurium by fluoroquinolones. *Mutagenesis* **5**: 63-66.

Zhang, H., Yoshizawa, S., Sun, Y., Huang, Y., Chu, X., Gonz**á**lez, J.M. et al. (2019) Repeated

evolutionary transitions of flavobacteria from marine to non-marine habitats. *Environ Microbiol* **21**:

648-666.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z. et al. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*.

Zhang, J. (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol* **50**: 56-68.

Zuckerkandl, E., and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H. (eds). In *Evolving genes and proteins*. Academic Press: NY, USA, pp. 97-166.

Figure Legends

Fig. 1   Diversity and abundance of the GH gene families (A) and genome size normalized densities of CAZyme and peptidase genes (B) in the genomes of *Nonlabens* sp. SY33080 and *Leeuwenhoekiella* sp. ZYFB001. In panel B, CAZyme genes are classified into auxiliary activity (AA), carbohydrate-binding module (CBM), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyltransferase (GT), and polysaccharide lyase (PL) categories. Peptidase genes are classified into aspartic, cysteine, metallo, serine, threonine, and unknown families based on the MEROPS annotation.

Fig. 2   Ancestral genome reconstruction using a modified Sankoff algorithm implemented in BadiRate based on a maximum likelihood phylogeny of the marine *Nonlabens*, *Leeuwenhokeiella*, *Dokdonia*, and outgroup lineages (collapsed). The phylogeny was constructed using IQ-TREE based on a concatenation of 163 single-copy orthologous protein sequences. The scale bar at the bottom indicates the number of substitutions per site. Number in parenthesis next to each ancestral node is the number of gene families predicted for that node. Accelerated rate of sequence evolution (defined as increased number of substitutions per site) was observed in the *Nonlabens* lineage compared to the *Leeuwenhokeiella* and the *Dokdonia* lineages

since the Last Common Ancestor (node N33) of these three lineages. Solid and open

circles at the nodes indicate the frequency of the clade defined by that node is greater

and lower than 90%, respectively, in the 1,000 ultrafast bootstrapped replicates. The

genome size, GC content, predicted genome completeness, and presence (solid circle)

or absence (open circle) of signature gene families of each taxon are shown next to

each taxon. Evolutionary gains and losses of SOS response associated genes are

marked with red and blue, respectively, and placed next to the ancestral nodes in

Nonlabens, *Dokdonia* and *Leeuwenhokeiella* lineages. The full version of the figure

with a complete set of outgroup lineages is shown in Fig. S1.

Fig. 3   Base-substitution mutation and insertion/deletion mutation that occurred in

each protein-coding gene are marked with blue and red vertical bars, respectively. The

height of each bar represents the number of mutations across all mutation

accumulation lines. The bars are plotted against the genomic position of that gene in

Nonlabens sp. SY33080 (upper) and *Leeuwenhoekiella* sp. ZYFB001 (lower). The

width of the bar represents the length of the gene. The two types of mutation that

occurred on the remaining genomic regions (intergenic regions and non-protein

coding genes) are marked with blue and red triangles, respectively, which are plotted

against the genomic positions of each occurrence. Note that the genomic position of

insertion/deletion mutation refers to the position of the first mutated site.
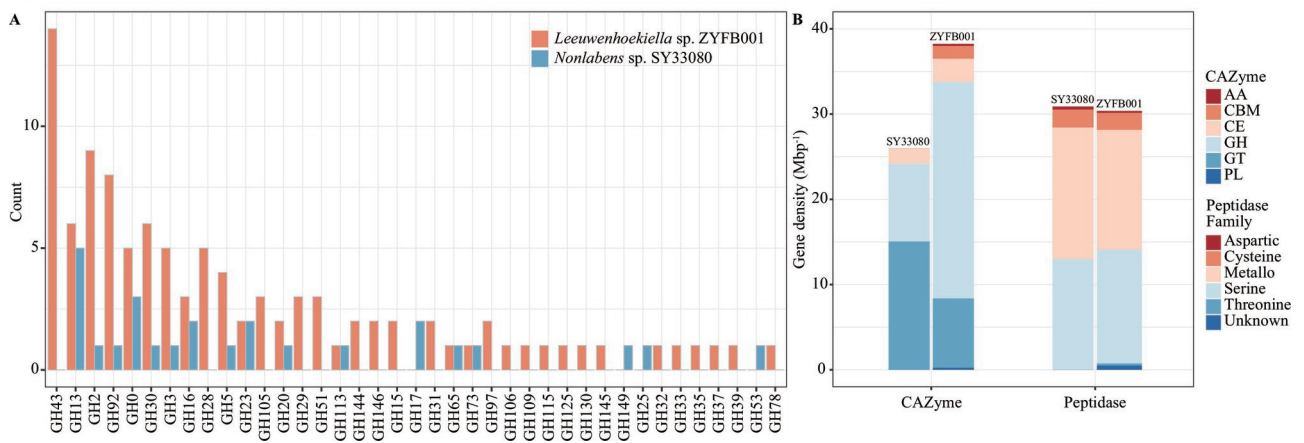
Table Legends

Table 1 Substrate utilization of *Leeuwenhoekiella* sp. ZYFB001 and *Nonlabens* sp.

SY33080 identified using Biolog GN2 Microplates and specialized culture media.

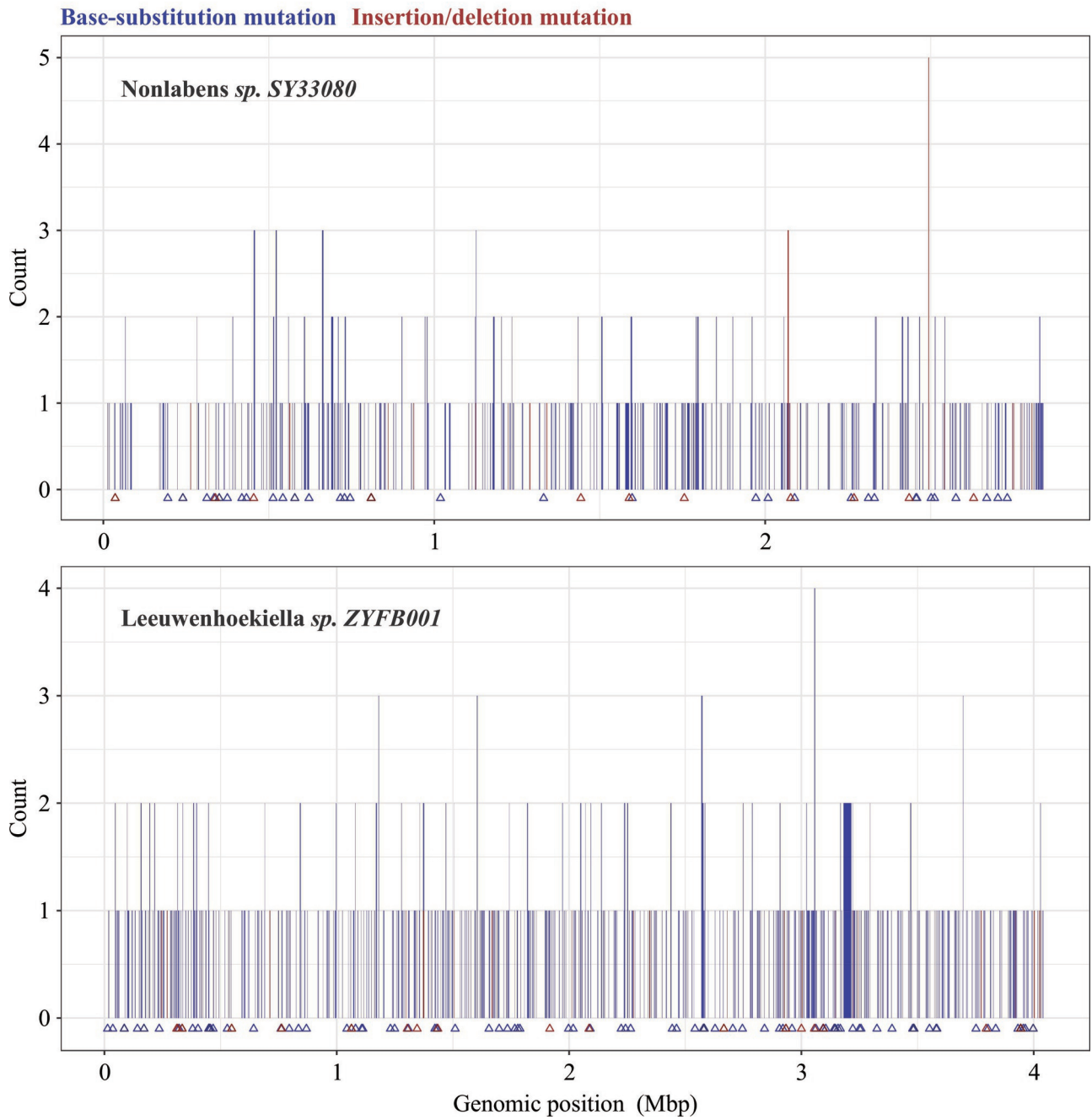Enzyme activities were identified by API 20NE Strips. Substrates marked with bold

are exclusively metabolized by *Leeuwenhoekiella* sp. ZYFB001.

Fig. 1

A

emi_15065_fig.1.eps

Fig. 2

emi_15065_fig.2.eps

Fig. 3

emi_15065_fig.3.eps

Ancestral Niche Separation and Evolutionary Rate Differentiation Between Sister Marine Flavobacteria Lineages

Chun-Xu Xue[†,1,2], Hao Zhang[†,2], He-Yu Lin[†,1], Ying Sun[2], Danli Luo[2], Yongjie Huang[2], Xiao-Hua Zhang[*,1], Haiwei Luo[*,2,3]

[1]College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

[2]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

[3]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518000, China

[†]These authors contributed equally to this study.

[*]**Corresponding author**:
Haiwei Luo
E-mail: hluo2006@gmail.com
The Chinese University of Hong Kong

Xiao-Hua Zhang
E-mail: xhzhang@ouc.edu.cn
Ocean University of China

Originality significance statement

Members of the flavobacteria group in *Bacteroidetes* are prevalent in marine environments, representing up to 10-20% of the bacterial cells in some ocean regions. Flavobacteria are generally considered as the specialists for polysaccharide degradation because of the excess of genes encoding carbohydrate active enzymes, and thus play an important role in marine carbon cycle. The present study employs a combination of genomic analyses, physiological assays, ancestral reconstruction and experimental evolution, and addresses the evolutionary processes and mechanisms underlying niche differentiation of two related flavobacteria lineages.

Summary

Marine flavobacteria are specialists for polysaccharide degradation. They dominate in habitats enriched with polysaccharides, but are also prevalent in pelagic environments where polysaccharides are less available. These niches are likely occupied by distinct lineages, but evolutionary processes underlying their niche differentiation remain elusive. Here, genomic analyses and physiological assays indicate that the sister flavobacteria lineages *Leeuwenhoekiella* and *Nonlabens* likely explore polysaccharide-rich macroalgae and polysaccharide-poor pelagic niches, respectively. Phylogenomic analyses inferred that the niche separation likely occurred anciently and coincided with increased sequence evolutionary rate in *Nonlabens* compared to *Leeuwenhoekiella*. Further analyses ruled out the known mechanisms likely driving evolutionary rate acceleration, including reduced selection efficiency, decreased generation time and increased mutation rate. In particular, the mutation rates were determined using an unbiased experimental method, which measures the present-day populations and may not reflect ancestral populations. These data collectively lead to a new hypothesis that an ancestral and transient mutation rate increase resulted in evolutionary rate increase in *Nonlabens*. This hypothesis was supported by inferring that gains and losses of genes involved in SOS response, a mechanism known to drive transiently increased mutation rate, coincided with evolutionary rate acceleration. Our analyses highlight the evolutionary mechanisms underlying niche differentiation of flavobacteria lineages.

Introduction

Approximately one-half of the oceanic primary production is consumed by heterotrophic marine bacteria (Fuhrman and Azam, 1982; Fenchel, 1988). Among these, flavobacteria is one of the most effective groups because of its high abundance in global oceans and its specialized capability for polymeric organic matter utilization (Kirchman, 2002). Marine flavobacteria thrive in diverse ecological niches and are enriched in coastal waters and during phytoplankton blooms, in which they account for up to ~40% and ~70% of bacterial cells, respectively (Cottrell and Kirchman, 2000; Kirchman, 2002; Teeling et al., 2016). The unique role of flavobacteria in high molecular weight organic matter utilization largely results from an excess of genes encoding carbohydrate active enzymes (CAZymes) and peptidases carried in their genomes (Bauer et al., 2006; Fernández-Gomez et al., 2013). CAZymes often cluster with *susCD*-like transporter genes, and form polysaccharide utilization locus (PUL) for binding, degradation and uptake of different classes of polysaccharides (Kappelmann et al., 2019).

Despite their generally recognized role in polysaccharide and peptide degradation, a few studies have revealed niche specialization and adaptation among flavobacteria lineages. For example, although members of the genus *Formosa* have been largely isolated from the nutrient-enriched niches such as brown algae, a few were proposed to inhabit the pelagic ocean (Mann et al., 2013). We recently demonstrated three unidirectional transitions from marine to non-marine habitats during the evolutionary history of flavobacteria, which was likely facilitated by repeated gains and losses of niche-specifying genes (Zhang et al., 2019). The present study investigates the potential niche separation between two well-established sister lineages, *Nonlabens* and *Leeuwenhoekiella*. While the two genera are distributed in

overlapping marine niches such as seawater and invertebrates (Nedashkovskaya et al., 2005; Lau et al., 2006; Pinhassi et al., 2006; Park et al., 2012), they show distinct genomic features, suggestive of niche specialization. Members of *Leeuwenhoekiella* consistently have larger genome size (over 4 Mbp), versatile metabolic capabilities and gliding motility, whereas *Nonlabens* have reduced genomes (around 3 Mbp), decreased gliding motility, and use rhodopsins to acquire energy from sunlight in some members (Fernández-Gomez et al., 2013; Kwon et al., 2013).

A major benefit of using *Nonlabens* and *Leeuwenhoekiella* is that they allow for an in-depth analysis of evolutionary forces associated with their niche separation. As repeatedly shown in published phylogenomic trees in which these two lineages were included (Kumagai et al., 2018; Zhang et al., 2019), the genome-reduced *Nonlabens* evolved at an accelerated rate compared to *Leeuwenhoekiella*, manifested as extraordinarily long branches connected to the former members since the last common ancestor (LCA) shared by the two lineages. In fact, free-living marine bacterial lineages with reduced genome sizes commonly show accelerated rate of sequence evolution at genome-wide scale (Batut et al., 2014), with prominent examples being the genome-reduced *Prochlorococcus* lineages in Cyanobacteria, the SAR11 clade in Alphaproteobacteria, and the SAR86 clade in Gammaproteobacteria. A few evolutionary mechanisms have been invoked to explain the genome reduction in free-living marine bacteria, which highlighted the role of positive selection (Giovannoni et al., 2005), increased mutation rate (Marais et al., 2008), and genetic drift (Luo et al., 2017). In theory, these evolutionary forces may potentially also drive the genome-wide accelerated evolutionary rate.

Among these mechanisms, genetic drift has been subject to rigorous tests. The underlying principle is that, because nonsynonymous substitutions leading to

physicochemically dissimilar amino acid replacements (radical changes) are more likely to be deleterious than those leading to similar amino acid replacements (conservative changes) (Zuckerkandl and Pauling, 1965; Dayhoff et al., 1972), an excess of the former type across genomic regions is evidence for genetic drift as a result of effective population size (N_e) reduction (Zhang, 2000; Eyre-Walker et al., 2002; Hughes and Friedman, 2009; Wernegreen, 2011). Both sides of the causal relationship underlying this drift theory were shown to exist in several important free-living marine bacteria (Luo et al., 2017). For example, the radical nonsynonymous mutations were demonstrated to be under greater functional constraints than the conservative nonsynonymous mutations in all examined marine bacterial populations, indicating the former type of mutations are indeed more deleterious than the latter type (Luo et al., 2017). Further, genome reduction and genomic sequence evolutionary rate acceleration were shown to coincide with an accelerated accumulation of the more deleterious type of mutations (i.e., radical changes) in *Prochlorococcus* and in SAR86, suggesting that genetic drift, rather than natural selection, acted at the time that genome reduction and increased evolutionary rate occurred in these lineages (Luo et al., 2017).

The hypothesis that increased mutation rate as a result of repair gene losses may play a role in accelerated sequence evolution in free-living marine bacteria is straightforward and was used to explain the accelerated sequence evolution in *Prochlorococcus* (Dufresne et al., 2005). The hypothesis that increased mutation rate may drive genome reduction of marine bacterioplankton lineages was supported by a mathematical model (Marais et al., 2008). However, direct experimental measures of unbiased mutation rates in free-living marine bacterial lineages showing accelerated evolutionary rate and possessing reduced genomes have not been available. With

increased mutation rate, not all auxiliary genes can be maintained by purifying

selection (Marais et al., 2008); some of these are pseudogenized through disabling

mutations and subsequently lost by mutations that are commonly biased towards

deletion over insertion (Kuo and Ochman, 2009). Because genome reduction in

marine bacterioplankton lineages are often associated with loss of important repair

genes (Partensky and Garczarek, 2010; Viklund et al., 2011), elevation of mutation

rate is potentially an important mechanism driving genome reduction in marine

bacterioplankton lineages (Marais et al., 2008). In addition to repair gene loss,

increased mutation rates could also be triggered by error-prone DNA replication under

SOS conditions (Maor-Shoshani et al., 2000). However, the relationship between

genome reduction and the bacterial SOS response is yet to be explored.

The mutation accumulation (MA) experiment followed by whole-genome

sequencing (WGS) of the mutant lines is becoming a standard approach for

spontaneous mutation rate determination, and the MA/WGS data of over 20 model

bacteria have become available (Long et al., 2018). In the MA procedure, the

laboratory populations arising from the same ancestor cell independently go through

repeated single-cell bottlenecks over hundreds to several thousand cell divisions. This

procedure minimizes the ability of natural selection to eliminate deleterious mutations

except the lethal ones and thus is considered an approximately unbiased method for

mutation rate determination (Lynch et al., 2016). Currently the common practice of

creating single-cell bottlenecks is by propagating the population on solid medium.

Unfortunately, most marine bacterioplankton lineages with reduced genomes either

remain uncultivated (e.g., SAR86), do not grow on solid media (e.g., SAR11), or are

practically infeasible for long-term propagation on solid medium (e.g.,

Prochlorococcus), and thus are not suitable for the MA/WGS procedure. Members of

Nonlabens and *Leeuwenhoekiella* grow readily on solid medium and thus are amenable to the MA/WGS procedure, providing an unprecedented opportunity to examine the various drivers including mutation rate effect on elevated evolutionary rate and genome reduction in free-living marine bacteria.

Results and Discussion

Niche specialization of present-day populations in Nonlabens and Leeuwenhoekiella

We performed detailed experimental and bioinformatics analyses on a present-day representative member SY33080 in the *Nonlabens* lineage and compared it to strain ZYFB001 in the sister lineage *Leeuwenhoekiella*. Both of them grow on starch, Tween 20, Tween 40, and Tween 80, but neither use gelatin, casein, cellulose, alginate, and chitin. The API 20NE experiment result shows only *L*. ZYFB001 has β-glucosidase and β-galactosidase activities. Furthermore, the GN2 Biolog analysis showed that *L*. ZYFB001 utilizes 34 more carbon compounds than *N*. SY33080 (Table 1), among which 24 are carbohydrates or organic acids and only the remaining three are amino acids, consistent with the more abundant glycoside hydrolase (GH) genes (103 vs. 26) and more diverse glycoside hydrolase gene families (35 vs. 17) in *L*. ZYFB001 than in *N*. SY33080 (Fig. 1A).

By normalizing the genome size, we found a greater CAZyme gene density (38.27 Mbp^{-1} vs. 25.96 Mbp^{-1}) but a similar peptidase gene density (30.37 Mbp^{-1} vs. 30.88 Mbp^{-1}) in *L*. ZYFB001 compared to *N*. SY33080 (Fig. 1B), giving rise to a starkly elevated peptidase:CAZyme ratio in *N*. SY33080 (1.19) compared to that in *L*. ZYFB001 (0.79). The trend remains consistent when taking into account the remaining genomes of each lineage: the CAZyme gene density in *Leeuwenhoekiella* lineage is significantly higher than that in the *Nonlabens* lineage (Wilcoxon-Mann-

Whitney test, $p<0.001$), whereas the peptidase gene density is not significantly different (Wilcoxon-Mann-Whitney test, $p>0.05$). This leads to a significantly elevated peptidase:CAZyme ratio in the *Nonlabens* lineage compared to the *Leeuwenhoekiella* lineage (Wilcoxon-Mann-Whitney test, $p<0.001$). Since marine *Bacteroidetes* taking a primarily planktonic lifestyle tend to have higher peptidase:CAZyme ratios compared to those adopting a primarily algae-associated lifestyle (Xing et al., 2015; Bennke et al., 2016), our results strongly suggest that the extant members of the *Nonlabens* lineage and the *Leeuwenhoekiella* lineage diversified ecologically towards taking planktonic and algae-associated lifestyles, respectively.

Bacteroidetes utilize complex polysaccharides often through PUL, which consists of *susC*-like TonB-dependent receptors (TBDRs), paired *susD*-like genes, CAZymes, and auxiliary genes such as sulfatase and regulatory genes (Bennke et al., 2016). Among these functional units, SusD binds to the degraded polysaccharides and bring them to TBDR for transportation into the cell (Sonnenburg et al., 2010). *L.* ZYFB001 possesses 59 TBDR genes, 34 of which are co-located with *susD*-like genes. By contrast, *N.* SY33080 only has 22 TBDR genes, two of which form partnership with *susD*-like genes. We identified 15 functional PULs potentially involved in the utilization of a diverse array of polysaccharides in *L.* ZYFB001 but only one PUL in *N.* SY33080 (Table S1).

In summary, *N.* SY33080 possesses a reduced genome, features low catabolic potential for polysaccharides, paralleling the planktonic lifestyle of several model marine flavobacteria strains such as *Polaribacter* sp. MED152, *Polaribacter* sp. Hel1 33 49, and *Dokdonia donghaensis* MED134 (González et al., 2008; González et al., 2011; Xing et al., 2015). In contrast, *L.* ZYFB001 possesses a larger genome and

features highly diverse and abundant genes involved in polysaccharide utilization, likened to the algae-associated paradigm of other model marine flavobacteria strains such as *Zobellia galactanivorans* Dsij[T], *Gramella forsetii* KT0803, *Formosa agariphila* KMM 3901, and *Polaribacter* sp. Hel1 85 (Mann et al., 2013; Kabisch et al., 2014; Xing et al., 2015; Barbeyron et al., 2016).

Niche separation of Leeuwenhoekiella and Nonlabens occurred anciently and accompanied by genome expansion and reduction, respectively

Among the three sister lineages, *Nonlabens* branched off prior to the diversification of *Leeuwenhoekiella* from *Dokdonia* (Fig. 2, Fig. S1). Ancestral genome reconstruction inferred that the LCA (N33) of *Nonlabens*, *Dokdonia* and *Leeuwenhoekiella* carried a genome with 2,216 gene families. It underwent distinct evolutionary processes toward the LCA of *Nonlabens* (N19) and toward the LCA of *Leeuwenhoekiella* (N31). The former reduced the genome size by 3.5%, whereas the latter experienced a stepwise genome expansion via the ancestral node of N32 (LCA of *Dokdonia* and *Leeuwenhoekiella*) by a net increase of 12.5% (Fig. 2, Table S2). Further diversification within *Nonlabens* was not associated with a clear trend of genome reduction, though an overall trend of expansion was reconstructed along the evolution of *Leeuwenhoekiella* (Table S2). Therefore, the dichotomy of lifestyle between extant members of *Nonlabens* and *Leeuwenhoekiella* is a result of two independent processes including early genome reduction in the former lineage and persistent genome expansion in the latter lineage.

Among the 223 gene families lost at the LCA (N19) of *Nonlabens* (Table S3), 173 were consistently annotated by at least two databases. Many of them are involved in the degradation of carbohydrates derived from marine macroalgae (Fig. 2). For example, GH3 family was lost, which consists of β-glucosidase and 1,4-β-xylosidase

for degrading β-glucan in brown algae (Kabisch et al., 2014) and β-xylan in green macroalgae (Domozych et al., 2012), respectively. A second example is the loss of GH32 family with endo- and exo-type of levanase and invertase activities targeting fructan and sucrose, respectively (Versluys et al., 2018). While these carbohydrates are largely found in terrestrial plants including vegetables, fruits and crops (Huber, 1989; Muir et al., 2007; van Arkel et al., 2013), fructan and sucrose were also found in certain species of green macroalgae (e.g. *Cladophorales*) and brown algae (e.g. *Laminaria*), respectively (Nwachukwu and Chukwu, 2012; Versluys et al., 2018). A third example is the loss of trehalase, which enables the utilization of trehalose, the common algal disaccharide accumulated in red algae (Stewart, 1974). A fourth example is the loss of multiple genes responsible for the utilization of xylose, the major component of xylan found in the cell wall of green macroalgae and red algae (Turvey and Williams, 1970; Huizing and Rietema, 1975). This includes the loss of xylose transporter, xylose isomerases catalyzing the interconversion between D-xylose and D-xylulose, and D-xylulose kinase phosphorylating D-xylulose for further degradation. A fifth example is the loss of genes encoding α-amylase and extracellular starch binding protein for the utilization of starch, which is accumulated as granule in red algae (Viola et al., 2001). Furthermore, utilization of polysaccharides is often facilitated by the coordinated activities of genes clustered in PUL, which consists of CAZymes and two types of PUL signature genes, the *susC*-like TonB-dependent receptors and the *susD*-like outer membrane proteins (Bennke et al., 2016). The losses of several copies of both *susC* and *susD* genes further reduce the potential of *Nonlabens* LCA to utilize macroalgae derived polysaccharides.

On the other hand, 145 gene families were predicted to be gained at the LCA of *Nonlabens* (Table S4), of which only 61 were annotated by two or more databases.

Among these, a few are involved in the interaction with microalgae. The genes gained at N19 include GH17 and GH30, which may have β-1,3-glucosidase and β-1,6-glucanase activities, respectively (Fig. S2A & B). They are often adjacently located within PUL for the degradation of laminarin or chrysolaminarin, with GH30 firstly cleaving β-1,6-glucose side chains followed by cleaving the remaining 1,3 bonds via GH17 for further utilization (Kappelmann et al., 2019). Laminarin and chrysolaminarin are used as the storage polysaccharides in brown algae and diatoms, respectively (Kappelmann et al., 2019). Although the backbones of laminarin and chrysolaminarin have the same chemical composition, the former is capped at the reducing end by a D-mannitol (Read et al., 1996), which needs to be removed through the activity of mannitol dehydrogenase as one step of laminarin degradation (Kappelmann et al., 2019). The absence of the mannitol dehydrogenase genes in the LCA of *Nonlabens* suggests that these two glycoside hydrolase genes (GH17 and GH30) were likely involved in the utilization of chrysolaminarin, which makes up to 80% of dry weight in diatom (Gügi et al., 2015). Another example of gain at N19 is the two β-mannosidase gene families in GH2 (Fig. S2C), which produces β-D-mannose residues from hydrolyzing β-D-mannosides. Oligomannosides are the dominant unit linked to membrane-bound proteins through *N*-glycosidic bond in microalgae (Mathieu-Rivet et al., 2013). The gained genes also include GH25, which exclusively encodes a lysozyme for peptidoglycan degradation. There is evidence that GH25 breaks down the cell wall of blue-green algae, allowing the bacteria such as *Cytophaga* (another major lineage of *Bacteroidetes*) to utilize the algal protoplasmic nutrients (Cole, 1982).

Other important gains at N19 include the gene encoding a carotenoid biosynthesis protein (PF04240 in Pfam) (Fig. 2). Carotenoids are likely among the earliest light-

absorbing compounds and are well known for their photoprotective function against UV radiation (Pinhassi et al., 2016). Consequently, the gain of carotenoid synthesis gene may facilitate niche transition to the pelagic water where cells are more exposed to UV light. Another interesting prediction was that *crtI* for β-carotene synthesis experienced an expansion of copy number at N19 via horizontal gene transfer (Fig. S3). β-carotene is one type of carotenoid, which is used for retinal synthesis, an essential component of rhodopsin (Pinhassi et al., 2016). In oligotrophic pelagic environments where labile organic matter is commonly depleted, rhodopsin-containing flavobacteria utilize light as one source of energy to power cellular activities (Yoshizawa et al., 2014). While three types of rhodopsins, including proton-pumping proteorhodopsin (PR), sodium-pumping rhodopsin (NaR) and chloride-pumping rhodopsin (ClR) (Béja et al., 2000; Inoue et al., 2013; Yoshizawa et al., 2014), have been identified in *Nonlabens*, none was predicted to be gained at the LCA (N19) of the *Nonlabens* lineage (Fig. S4). Therefore, acquisition of an additional copy of *crtI* could be a case of pre-adaptation in early *Nonlabens* and facilitate the transition to rhodopsin-based photoheterotrophy in the later stage of their evolution.

Selection efficiency was not reduced despite the genome-wide accelerated sequence evolutionary rate in Nonlabens compared to that in Leeuwenhoekiella

An important observation is that genome reduction in the *Nonlabens* lineage is associated with acceleration of amino acid sequence evolution across genomic regions. This is supported by the severely violated molecular clock between these two lineages (likelihood ratio test, *p*<0.001), and is shown by the extraordinarily longer branches connecting the LCA (N33) of the *Nonlabens*, *Dokdonia* and *Leeuwenhoekiella* lineages to the extant *Nonlabens* compared the branches connecting this LCA to the extant *Leeuwenhoekiella* (Fig. 2). A higher rate of sequence evolution

at genome-wide scale often results from genetic drift, or increased mutation rate, or a combination of these factors (Graur et al., 2016). As genome reduction of *Nonlabens* largely occurred on the long ancestral branch leading to the LCA (N19) of the *Nonlabens* lineage, the power of genetic drift was tested on this branch.

Given this genome reduction event having occurred very anciently, the relative rate of radical versus conservative nonsynonymous substitutions (the d_R/d_C ratio) across the genomic regions can be a proxy for N_e (Zhang, 2000; Eyre-Walker et al., 2002; Hughes and Friedman, 2009; Wernegreen, 2011; Luo et al., 2017). The potential change of N_e associated with this genome reduction event can be approximated by comparing the d_R/d_C on the branch leading to the LCA of the *Nonlabens* lineage to that of its sister branch giving rise to the LCA (N32) of the *Dokdonia* and *Leeuwenhoekiella* lineages. Using the outgroup genomes (Fig. S1) as the reference, pairwise comparisons across 492 single-copy orthologous genes were performed between the target clade (*Nonlabens*) and the reference genomes, as well as between the control clade (*Leeuwenhoekiella* and *Dokdonia*) and the reference genomes, respectively. These d_R/d_C analyses led to opposite interpretations when different classification schemes of the 20 amino acids were used (Fig. S5A, Table S5). While the genome-wide d_R/d_C ratio was consistently greater on the ancestral branch leading to the N19 than that on the branch leading to N32 based on volume and polarity, the d_R/d_C ratio was instead consistently inflated on the ancestral branch leading to N32 when amino acids were classified according to charge. As genomes of *Dokdonia* are smaller and more biased toward low GC content compared to *Leeuwenhoekiella*, which might affect the result, the above analysis was also done after the *Dokdonia* lineage was excluded and the results remained qualitatively the same (Fig. S5B). The opposite results of d_R/d_C depending on the amino acid classification scheme suggest

that genetic drift was likely not a primary mechanism driving genome reduction of the *Nonlabens* lineage.

Spontaneous mutation rate of the extant Nonlabens is not significantly greater than that of the extant Leeuwenhoekiella

As discussed above, increased mutation rate is likely another major mechanism leading to acceleration of sequence evolution and genome reduction, so we employed the MA/WGS approach to determine the unbiased spontaneous mutation rate of ZYFB001 in *Leeuwenhoekiella* and of SY33080 in *Nonlabens*. Base-substitution mutations of both organisms are biased toward intergenic regions compared to coding regions ($\chi 2$ test; $p<0.05$ for *N.* SY33080 and $p<0.001$ for *L.* ZYFB001; Table S6). This bias is commonly found in other bacteria including the model marine bacterium *Ruegeria pomeroyi* DSS-3 in Alphaproteobacteria, which likely results from more active mismatch repair on coding regions or the mutation-prone nature of repetitive DNA in intergenic regions (Lee et al., 2012; Long et al., 2014; Dillon and Cooper, 2016; Sun et al., 2017). Both strains show a significantly greater rate of base-substitution mutations from G/C to A/T than from A/T to G/C (prop.test function of R, $p<0.001$, df=1) (Fig. S6), and this mutational bias is found in most bacterial species that were subject to mutation accumulation experiments (Long et al., 2018). In both organisms, the nonsynonymous to synonymous base-substitution rate ratio is not significantly different from the random expectation ($\chi 2$ test; $p>0.05$; Table S6). Thus, selection does not appear to have a significant impact on the accumulation of mutations in both organisms. This conclusion is further strengthened by the randomly distributed mutations across the genomic regions in all lines (Fig. 3 & Fig. S6) and the comparable number of mutations accumulated among all protein-coding genes in each organism (Fig. 3), the latter supported by a statistical test showing that no genes

accumulated an excess of mutations (bootstrapping test, $p>0.05$ for each organism).

We determined 366 base-substitution mutations over 2,549 cell divisions in *N.* SY33080 and 604 base-substitution mutations over 3,006 cell divisions in *L.* ZYFB001, leading to a mutation rate of 5.49×10^{-10} and 5.21×10^{-10} per base per cell division for *N.* SY33080 and *L.* ZYFB001, respectively (Table S7 & S8). These mutation rates fall in the middle of the variation determined for other bacteria using MA/WGS (Long et al., 2018). Despite a slight increase, the mutation rate of *N.* SY33080 is not significantly different from that of *L.* ZYFB001 (Kolmogorov-Smirnov test and Wilcoxon-Mann-Whitney test; $p>0.05$).

Given the same mutation rate per cell division, a more rapidly growing bacterial lineage may give rise to increased evolutionary rate, presumably because their cells underwent more cell divisions and thus collected more mutations (Weller and Wu, 2015; Gibson and Eyre-Walker, 2019). However, the generation time estimation for heterotrophic marine bacteria in the wild has been highly controversial depending on the experimental methods that marine microbial ecologists use, and it is even more problematic to accurately quantify the generation time of a particular heterotrophic taxon and differentiate it among closely related heterotrophic lineages in the wild (Kirchman, 2016). The available lineage-specific growth rate data all compare the generation time of deeply-branching groups, such as the oligotrophic and often free-living groups including the SAR11 clade and the SAR86 clade, as well as the copiotrophic and often algae-associated groups including *Bacteroidetes*, *Rhodobacteraceae*, and *Alteromonadaceae*, and they consistently support that the oligotrophic groups have significantly longer generation time than the copiotrophic groups (Teira et al., 2009; Bakenhus et al., 2017; Sánchez et al., 2017). Because the *Nonlabens* and *Leeuwenhoekiella* lineages are aligned to the oligotrophic and

copiotrophic models, respectively, it is reasonable to postulate that members of the *Nonlabens* lineage have longer generation time than members of the *Leeuwenhoekiella* lineage in the ocean. Further, a significantly longer generation of the *N*. SY33080 than that of *L*. ZYFB001 was also observed under optimal laboratory conditions throughout the MA experiment (Wilcoxon-Mann-Whitney test, *p*<0.001) (Fig. S7). Taken together, *Nonlabens* may have longer generation time than *Leeuwenhoekiella* under both wild and laboratory conditions, and thus the generation time effect is unlikely responsible for the accelerated rate of sequence evolution in the *Nonlabens* lineage.

Losses of DNA repair genes correlated with the increased evolutionary rate of Nonlabens

We noted the above mutation rate measurements were derived from the present-day organisms and are not necessarily applicable to their ancient ancestors, such as the LCA of the *Nonlabens* lineage and the LCA of the *Leeuwenhoekiella* lineage. Because mutator phenotypes (i.e., strains with increased mutation rate) often result from loss of important DNA repair genes that ensure accuracy of DNA replication (Denamur et al., 2000; Batut et al., 2014) or gains of error-prone repair systems (Ysern et al., 1990; Woodgate et al., 1994), gains and losses of these genes can be used as proxies for mutation rate increase in ancestral lineages.

We inferred that three repair genes (*dnaQ*, *umuC* and *umuD*) were lost on the ancestral branch leading to the LCA (N19) of the *Nonlabens* lineage (Fig. 2 & S8). In contrast, only one DNA repair gene *recO* was inferred to be lost on the branch leading to the LCA (N32) of the *Dokdoni*a and *Leeuwenhoekiella* lineages, which is part of the RecF pathway for single-stranded DNA repair (Bork et al., 2001) (Fig. S8). Among the three genes lost on the ancestral node N19, *umuC* and *umuD* encode

subunits of the error-prone DNA polymerase V which is highly mutagenic and produces point mutations at a frequency of 2.1×10^{-4} per nucleotide, ~40 times greater than DNA polymerase III holoenzyme (Maor-Shoshani et al., 2000). They were firstly identified to be involved in SOS response in *E. coli* (Sutton et al., 2000). The proteins encoded by *umuC* and *umuD* constitute the UmuD'2C complex to perform DNA translesion synthesis (TLS), which allows DNA replication to bypass the lethal DNA damage (Sutton and Walker, 2001). It was proposed that two major types of DNA lesions, TT (6-4) photoproducts and TT cis-syn photodimers, may be caused by UV radiation (Tang et al., 2000).

According to theory, high mutation rate is favored by selection in a harsh environment because this increases the chance of acquiring adaptive mutations. Following adaptive events, high mutation rate is disfavored because of the cost of increased accumulation of deleterious mutations (Taddei et al., 1997; Tenaillon et al., 1999; Denamur et al., 2000). This may explain the loss of *umuC* and *umuD* at the LCA (N19) of the *Nonlabens* lineage. Another evidence that SOS response may have been involved during the ecological transition on the branch leading to the LCA of *Nonlabens* is the gain of SOS response associated peptidase (SRAP) gene family on the same branch (Fig. 2), which is involved in the induction of bacterial SOS response (Aravind et al., 2013).

The third repair gene predicted to be lost in the LCA of the *Nonlabens* lineage is *dnaQ* (also called *mutD*; Fig. 2), which encodes the epsilon subunit of DNA polymerase III. It functions as a 3'→5' DNA directed proofreading exonuclease and removes incorrectly incorporated bases during replication (Scheuermann and Echols, 1984), and thus is different from other DNA repair genes acting on replicated DNA strand(s). Interestingly, it was shown that the proofreading activity of *dnaQ* is

inhibited by SOS-induced UV mutagenesis (Woodgate et al., 1987), and that overexpression of *dnaQ* inhibits the SOS mutator activities (Jonczyk et al., 1988). We therefore hypothesized that the loss of *dnaQ* occurred early during the ecological transition, which facilitated the SOS response.

Concluding remarks

Our comprehensive genomic analyses and physiological assays suggest that members of the *Nonlabens* lineage and those of the *Leeuwenhoekiella* lineage take distinct ecological strategies, with the former likely occupying pelagic water niches and the latter favoring to be associated with macroalgae. Ancestral genome reconstruction suggests that this ecological differentiation was initiated early when these two lineages started to diverge from each other. The genome-wide elevated evolutionary rate of *Nonlabens* compared to *Leeuwenhoekiella* provides us an opportunity to investigate the evolutionary forces underlying the niche differentiation between the two lineages. Although genetic drift and/or increased mutation rate are often taken to account for an elevated evolutionary rate at genomic scale in microbial lineages (Itoh et al., 2002; McCutcheon and Moran, 2012; Batut et al., 2014; Wernegreen, 2015), these mechanisms were not supported here based on our rigorous experimental and bioinformatic tests and thus may not explain the accelerated evolutionary rates in *Nonlabens*. However, using evolutionary gains and losses of DNA repair genes as a proxy to represent potential mutation rate changes during the ancestral processes, we proposed that the evolutionary changes of genes involved in SOS response during the ecological transition from the protected macroalgal habitat to the UV-exposed pelagic niches likely gave rise to the transiently increased mutation rate in the early evolutionary history of *Nonlabens*, which may explain the accelerated evolutionary rate of this lineage.

According to the inferred evolutionary mechanisms, we propose an ecological scenario during the ancient niche transition. The primitive *Nonlabens* cells prior to the LCA of this lineage may have colonized macroalgae which provided a stable source of nutrients to the bacteria, but environmental fluctuations may have made the nutrient sources less predictable. To cope with this stress, some cells began to explore pelagic environments including the nutrient-poor bulk seawater and occasionally phytoplankton niches, where cells with reduced replication burden were favored. During this transition, the cells were subject to greater UV radiation, which caused severe DNA damage such as lesions and subsequently triggered bacterial SOS response. An important consequence was that these ancestral bacteria became mutator; they had an increased chance to evolve adaptive mutations, an increased frequency of accepting genes from external sources, and an increased rate of losing DNA. These genomic changes led to acquisitions of novel functions and reduction of replication costs, which enabled the cells to adapt in pelagic environments. After the lifestyle was switched, the previously important genes like those related to polysaccharide catabolism became dispensable and eventually lost. The proposed evolutionary mechanisms and the related ecological scenario still hold when an updated genome set that contains a few uncultivated members represented by metagenome-assembled genomes (MAGs) was used, as the genes that likely facilitated the niche separation and SOS response show similar phyletic pattern (Fig. S9 and Fig. 2). A more accurate delineation requires a more thorough sampling of the phylogenetically and ecologically diverse members within these lineages.

Materials and Methods

Leeuwenhoekiella sp. ZYFB001 and *Nonlabens* sp. SY33080 were isolated from

the surface marine snow particles in Yellow Sea (121.50 °E, 32.59 °N) and surface

seawater in East China Sea (126.51 °E, 28.43 °N), respectively. The potential of the

two isolates to utilize various substrates were tested using specialized media, API

20NE strips (BioMérieux), and GN2 MicroPlates Biolog (BIOLOG, Catalog #1011)

(See Supplementary Methods *Section 2*).

Whole genome shotgun sequencing of *L.* ZYFB001 and *N.* SY33080 were

carried out using both PacBio RS II and Illumina HiSeq 2500 sequencing platforms.

The correction, trimming, and assembly of long reads derived from PacBio with Canu

1.4 (Koren et al., 2017) yielded a single contig of 4,053,756 bp for *L.* ZYFB001 and

2,852,234 bp for *N.* SY33080, respectively. These assemblies were further polished

by the Illumina reads using BWA (Li and Durbin, 2009) and Pilon (Walker et al.,

2014).

A total of 106 publically available *Flavobacteriaceae* genomes were downloaded

from Genbank in December 2018, and they were chosen according to their

phylogenetic positions in a comprehensive phylogenomic tree of *Flavobacteriaceae*

(Zhang et al., 2019). Among these, *Leeuwenhoekiella* sp. Hel I 48 was recently

reclassified into genus *Flavimarina* due to its greater divergence of the 16S rRNA

gene sequence (8.4-9.6%) to other *Leeuwenhoekiella* strains (Nedashkovskaya et al.,

2014). Genome annotation was performed with in house resources including the Pfam

v30.0 (El-Gebali et al., 2018) and CDD v3.15 (Marchler-Bauer et al., 2016) databases

using HMMER v3.1 (Eddy, 2009) and RPS-BLAST v2.6.0 (Camacho et al., 2009),

respectively, as well as with online resources including the subsystem and KEGG

databases via RAST (Overbeek et al., 2013) and BlastKOALA (Kanehisa et al., 2016)

online tools, respectively. Genes encoding CAZymes and peptidase were further

annotated with dbCAN2 (Zhang et al., 2018) and the specialized MEROPS v12.0

database (Rawlings et al., 2017), respectively. Putative DNA repair genes were identified based on our in-house repair gene database via reciprocal best BLAST hits. The functional polysaccharide utilization locus (PUL) was predicted based on the presence of some essential features: tandem *susCD*-like pairs and adjacent CAZyme genes using in-house scripts (Terrapon et al., 2015) (See Supplementary Methods *Section 3*).

Homologous gene families of the 108 genomes were identified using OrthoFinder v2.2.1 (Emms and Kelly, 2015). The concatenated amino acid sequence of 163 shared single-copy gene families was used to construct a maximum likelihood phylogenomic tree using IQ-TREE v1.6.2 (Nguyen et al., 2014). The ancestral states and the evolutionary history of each identified orthologous gene family were estimated using the modified Sankoff parsimony algorithm implemented in the software BadiRate, which takes into account both the branch length of the phylogenomic tree and the uncertainty in the number of family members (Librado et al., 2011) (See Supplementary Methods *Section 4 and Section 5*).

The molecular clock hypothesis between the *Nonlabens* and *Leeuwenhoekiella* lineages was tested using the codeml program implemented in the PAML package (Yang, 2007). To test whether genetic drift was associated with the ancestral genome reduction of the *Nonlabens* lineage, we determined the rates of conservative (d_C) and radical (d_R) changes using methods controlling for the disparity in GC content implemented in RCCalculator (Luo et al., 2017) for each of the 492 single-copy gene families shared by at least 105 (out of 108) genomes. As *Dokdonia* is embedded within the phylogenetic group consisting of *Nonlabens* and *Leeuwenhoekiella*, there is a possibility that it interferes with the d_R/d_C analysis in comparing *Nonlabens* and *Leeuwenhoekiella*. We thus performed the d_R/d_C analysis by either including or

excluding *Dokdonia*. To determine the spontaneous mutation rate of *L.* ZYFB001 and *N.* SY33080, we applied the MA/WGS procedure to each of them under optimal condition. In summary, 96 independent MA lines were initiated from a single founder colony of *L.* ZYFB001 and *N.* SY33080, respectively. Every 48 h, cells from a single colony were transferred to a fresh plate. This transfer continued for 232 days, equivalent to 3,006 and 2,549 cell divisions for *L.* ZYFB001 and *N.* SY33080, respectively. Generation time between each transfer was estimated by colony forming units (CFU) every 45 days. Mutation rate μ_{bs} (per base per generation) was calculated by the following equation:

$$\mu_{bs} = \frac{m}{\sum_{1}^{n} N \times T}$$

where *m* is the total number of observed base-substitution mutations across all the MA lines, *n* is the total number of lines, *N* is the analyzed sites for each MA line, and *T* is the number of cell divisions that occurred in the MA study (Sung et al., 2012; Long et al., 2016) (See Supplementary Methods *Section 7, Section 8 and Section 9*).

Data availability

Genomic sequences of *Nonlabens* sp. SY33080 and *Leeuwenhoekiella* sp. ZYFB001 are available at the NCBI GenBank database under the accession number CP050461 and CP050460, respectively. The manually compiled DNA repair gene database as well as the scripts used for ancestral genome reconstruction and mutation rate calculation have been deposited in the online repository (https://github.com/zoopera/flavoevo/).

Acknowledgements

Conflict of interests

The author declares no competing interests.

Reference

Aravind, L., Anand, S., and Iyer, L.M. (2013) Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol Direct* **8**: 20.

Bakenhus, I., Dlugosch, L., Billerbeck, S., Giebel, H.A., Milke, F., and Simon, M. (2017) Composition of Total and Cell-Proliferating Bacterioplankton Community in Early Summer in the North Sea-Roseobacters Are the Most Active Component. *Front Microbiol* **8**: 1771.

Barbeyron, T., Thomas, F., Barbe, V., Teeling, H., Schenowitz, C., Dossat, C. et al. (2016) Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the

model algae-associated bacterium *Zobellia galactanivorans* Dsij[T]. *Environ Microbiol* **18**: 4610-4627.

Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol* **12**: 841.

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E. et al. (2006) Whole genome analysis of the marine *Bacteroidetes* 'Gramella forsetii' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**: 2201-2213.

Béja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P. et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902-1906.

Bennke, C.M., Krüger, K., Kappelmann, L., Huang, S., Gobet, A., Schüler, M. et al. (2016) Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic. *Environ Microbiol* **18**: 4456-4470.

Bork, J.M., Cox, M.M., and Inman, R.B. (2001) The RecOR proteins modulate RecA protein function

at 5′ ends of single-stranded DNA. *EMBO J* **20**: 7313-7322.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Cole, J.J. (1982) Interactions between bacteria and algae in aquatic ecosystems. *Annu Rev Ecol Syst* **13**: 291-314.

Cottrell, M.T., and Kirchman, D.L. (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Appl Environ Microbiol* **66**: 5116-5122.

Dayhoff, M., Eck, R., and Park, C. (1972) A model of evolutionary change in proteins. In: Dayhoff M. (ed). In *Atlas of protein sequence and structure*. National Biomedical Research Foundation: Silver Spring, MD, USA, pp. 89-99.

Denamur, E., Lecointre, G., Darlu, P., Tenaillon, O., Acquaviva, C., Sayada, C. et al. (2000) Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* **103**: 711-721.

Dillon, M.M., and Cooper, V.S. (2016) The fitness effects of spontaneous mutations nearly unseen by selection in a bacterium with multiple chromosomes. *Genetics* **204**: 1225-1238.

Domozych, D., Ciancia, M., Fangel, J.U., Mikkelsen, M.D., Ulvskov, P., and Willats, W.G. (2012) The cell walls of green algae: a journey through evolution and diversity. *Front Plant Sci* **3**: 82.

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

Eddy, S.R. (2009) A new generation of homology search tools based on probabilistic inference. In

Genome Informatics 2009: Genome Informatics Series Vol 23: World Scientific, pp. 205-211.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C. et al. (2018) The Pfam protein families database in 2019. *Nucleic Acids Res* **47**: D427-D432.

Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.

Eyre-Walker, A., Keightley, P.D., Smith, N.G., and Gaffney, D. (2002) Quantifying the slightly deleterious mutation model of molecular evolution. *Mol Biol Evol* **19**: 2142-2149.

Fenchel, T. (1988) Marine plankton food chains. *Annu Rev Ecol Syst* **19**: 19-38.

Fernández-Gomez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedros-Alio, C. (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* **7**: 1026-1037.

Fuhrman, J., and Azam, F. (1982) Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results. *Mar Biol* **66**: 109-120.

Gibson, B., and Eyre-Walker, A. (2019) Investigating evolutionary rate variation in bacteria. *J Mol Evol* **87**: 317-326.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D. et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242-1245.

González, J.M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M. et al. (2008) Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**: 8724-8729.

González, J.M., Pinhassi, J., Fernández-Gómez, B., Coll-Lladó, M., González-Velázquez, M., Puigbò, P. et al. (2011) Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp. strain MED134. *Appl Environ Microbiol* **77**: 8676-8686.

Graur, D., Sater, A.K., and Cooper, T.F. (2016) Rates and Patterns of Molecular Evolution. In *Molecular and genome evolution*: Sinauer Associates, Incorporated.

Gügi, B., Le Costaouec, T., Burel, C., Lerouge, P., Helbert, W., and Bardor, M. (2015) Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Mar Drugs* **13**: 5993-6018.

Huber, S.C. (1989) Biochemical mechanism for regulation of sucrose accumulation in leaves during photosynthesis. *Plant Physiol* **91**: 656-662.

Hughes, A.L., and Friedman, R. (2009) More radical amino acid replacements in primates than in rodents: support for the evolutionary role of effective population size. *Gene* **440**: 50-56.

Huizing, H., and Rietema, H. (1975) Xylan and mannan as cell wall constituents of different stages in the life-histories of some siphoneous green algae. *British Phycological Journal* **10**: 13-16.

Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., and Kandori, H. (2013) A light-driven sodium ion pump in marine bacteria. *Nat Commun* **4**: 1678.

Itoh, T., Martin, W., and Nei, M. (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci USA* **99**: 12944-12948.

Jonczyk, P., Fijalkowska, I., and Ciesla, Z. (1988) Overproduction of the epsilon subunit of DNA polymerase III counteracts the SOS mutagenic response of *Escherichia coli*. *Proc Natl Acad Sci USA* **85**: 9124-9127.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M. et al. (2014) Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* 'Gramella

forsetii'KT0803. *ISME J* **8**: 1492-1502.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726-731.

Kappelmann, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F. et al. (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* **13**: 76-91.

Kirchman, D.L. (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* **39**: 91-100.

Kirchman, D.L. (2016) Growth rates of microbes in the oceans. *Ann Rev Mar Sci* **8**: 285-309.

Kumagai, Y., Yoshizawa, S., Nakajima, Y., Watanabe, M., Fukunaga, T., Ogura, Y. et al. (2018) Solar-panel and parasol strategies shape the proteorhodopsin distribution pattern in marine *Flavobacteriia*. *ISME J*: 1.

Kuo, C.H., and Ochman, H. (2009) Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct* **4**: 35.

Kwon, S.-K., Kim, B.K., Song, J.Y., Kwak, M.-J., Lee, C.H., Yoon, J.H. et al. (2013) Genomic makeup of the marine Flavobacterium *Nonlabens (Donghaeana) dokdonensis* and identification of a novel class of rhodopsins. *Genome Biol Evol* **5**: 187-199.

Lau, S.C., Tsoi, M.M., Li, X., Plakhotnikova, I., Dobretsov, S.V., Wu, M. et al. (2006) *Stenothermobacter spongiae* gen. nov., sp. nov., a novel member of the family *Flavobacteriaceae* isolated from a marine sponge in the Bahamas, and emended description of *Nonlabens tegetincola*. *Int J Syst Evol Microbiol* **56**: 181-185.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci USA* **109**: E2774-E2783.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Librado, P., Vieira, F., and Rozas, J. (2011) BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* **28**: 279-281.

Long, H., Behringer, M.G., Williams, E., Te, R., and Lynch, M. (2016) Similar Mutation Rates but Highly Diverse Mutation Spectra in Ascomycete and Basidiomycete Yeasts. *Genome Biol Evol* **8**: 3815-3821.

Long, H., Sung, W., Miller, S.F., Ackerman, M.S., Doak, T.G., and Lynch, M. (2014) Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol* **7**: 262-271.

Long, H., Sung, W., Kucukyildirim, S., Williams, E., Miller, S.F., Guo, W. et al. (2018) Evolutionary determinants of genome-wide nucleotide composition. *Nat Ecol Evol*: 1.

Luo, H., Huang, Y., Stepanauskas, R., and Tang, J. (2017) Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. *Nat Microbiol* **2**: 17091.

Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K., and Foster, P.L. (2016) Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **17**: 704-714.

Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T. et al. (2013) The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol* **79**: 6813-6822.

Maor-Shoshani, A., Reuven, N.B., Tomer, G., and Livneh, Z. (2000) Highly mutagenic replication by

DNA polymerase V (UmuC) provides a mechanistic basis for SOS untargeted mutagenesis. *Proc Natl Acad Sci USA* **97**: 565-570.

Marais, G.A., Calteau, A., and Tenaillon, O. (2008) Mutation rate and genome reduction in endosymbiotic and free-living bacteria. *Genetica* **134**: 205-210.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S. et al. (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200-D203.

Mathieu-Rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Le Mauff, F. et al. (2013) Exploring the N-glycosylation pathway in Chlamydomonas reinhardtii unravels novel complex structures. *Mol Cell Proteomics* **12**: 3160-3183.

McCutcheon, J.P., and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13.

Muir, J.G., Shepherd, S.J., Rosella, O., Rose, R., Barrett, J.S., and Gibson, P.R. (2007) Fructan and free fructose content of common Australian vegetables and fruit. *J Agric Food Chem* **55**: 6619-6627.

Nedashkovskaya, O.I., Kukhlevskiy, A.D., Zhukova, N.V., and Kim, S.B. (2014) *Flavimarina pacifica* gen. nov., sp. nov., a new marine bacterium of the family *Flavobacteriaceae*, and emended descriptions of the genus *Leeuwenhoekiella*, *Leeuwenhoekiella aequorea* and *Leeuwenhoekiella marinoflava*. *Antonie Van Leeuwenhoek* **106**: 421-429.

Nedashkovskaya, O.I., Vancanneyt, M., Dawyndt, P., Engelbeen, K., Vandemeulebroecke, K., Cleenwerck, I. et al. (2005) Reclassification of [Cytophaga] marinoflava Reichenbach 1989 as *Leeuwenhoekiella marinoflava* gen. nov., comb. nov. and description of *Leeuwenhoekiella aequorea* sp. nov. *Int J Syst Evol Microbiol* **55**: 1033-1038.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274.

Nwachukwu, A., and Chukwu, M. (2012) The potential of macroalgae as a source of carbohydrates for use in bioethanol fermentation. *Int J Energy Environ* **3**: 761-774.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T. et al. (2013) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**: D206-D214.

Park, S., Yoshizawa, S., Chiura, H.X., Muramatsu, Y., Nakagawa, Y., Kogure, K., and Yokota, A. (2012) Nonlabens marina sp. nov., a novel member of the *Flavobacteriaceae* isolated from the Pacific Ocean. *Antonie Van Leeuwenhoek* **102**: 669-676.

Partensky, F., and Garczarek, L. (2010) *Prochlorococcus*: advantages and limits of minimalism. *Ann Rev Mar Sci* **2**: 305-331.

Pinhassi, J., DeLong, E.F., Béjà, O., González, J.M., and Pedrós-Alió, C. (2016) Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol Mol Biol Rev* **80**: 929-954.

Pinhassi, J., Bowman, J.P., Nedashkovskaya, O.I., Lekunberri, I., Gomez-Consarnau, L., and Pedros-Alio, C. (2006) *Leeuwenhoekiella blandensis* sp. nov., a genome-sequenced marine member of the family *Flavobacteriaceae*. *Int J Syst Evol Microbiol* **56**: 1489-1493.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D. (2017) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**: D624-D632.

Read, S.M., Currie, G., and Bacic, A. (1996) Analysis of the structural heterogeneity of laminarin by

electrospray-ionisation-mass spectrometry. *Carbohydr Res* **281**: 187-201.

Sánchez, O., Koblížek, M., Gasol, J.M., and Ferrera, I. (2017) Effects of grazing, phosphorus and light on the growth rates of major bacterioplankton taxa in the coastal NW Mediterranean. *Environ Microbiol Rep* **9**: 300-309.

Scheuermann, R.H., and Echols, H. (1984) A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proc Natl Acad Sci USA* **81**: 7747-7751.

Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., and Sonnenburg, J.L. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**: 1241-1252.

Sun, Y., Powell, K.E., Sung, W., Lynch, M., Moran, M.A., and Luo, H. (2017) Spontaneous mutations of a model heterotrophic marine bacterium. *ISME J*.

Sung, W., Ackerman, M.S., Miller, S.F., Doak, T.G., and Lynch, M. (2012) Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* **109**: 18488-18492.

Sutton, M.D., and Walker, G.C. (2001) Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc Natl Acad Sci USA* **98**: 8342-8349.

Sutton, M.D., Smith, B.T., Godoy, V.G., and Walker, G.C. (2000) The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu Rev Genet* **34**.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997) Role of mutator alleles in adaptive evolution. *Nature* **387**: 700.

Tang, M., Pham, P., Shen, X., Taylor, J.S., O'donnell, M., Woodgate, R., and Goodman, M.F. (2000) Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404**: 1014.

Teeling, H., Fuchs, B.M., Bennke, C.M., Krueger, K., Chafee, M., Kappelmann, L. et al. (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* **5**: e11888.

Teira, E., Martínez-García, S., Lønborg, C., and Álvarez-Salgado, X.A. (2009) Growth rates of different phylogenetic bacterioplankton groups in a coastal upwelling system. *Environ Microbiol Rep* **1**: 545-554.

Tenaillon, O., Toupance, B., Le Nagard, H., Taddei, F., and Godelle, B. (1999) Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**: 485-493.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**: 647-655.

Turvey, J., and Williams, E. (1970) The structures of some xylans from red algae. *Phytochemistry* **9**: 2383-2388.

van Arkel, J., Sévenier, R., Hakkert, J.C., Bouwmeester, H.J., Koops, A.J., and van der Meer, I.M. (2013) Tailor-made fructan synthesis in plants: a review. *Carbohydr Polym* **93**: 48-56.

Versluys, M., Kirtel, O., Toksoy Öner, E., and Van den Ende, W. (2018) The fructan syndrome: evolutionary aspects and common themes among plants and microbes. *Plant Cell Environ* **41**: 16-38.

Viklund, J., Ettema, T.J., and Andersson, S.G. (2011) Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29**: 599-615.

Viola, R., Nyvall, P., and Pedersén, M. (2001) The unique features of starch metabolism in red algae. *Proc R Soc Lond B Biol Sci* **268**: 1417-1422.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. et al. (2014) Pilon: an

integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**.

Weller, C., and Wu, M. (2015) A generation-time effect on the rate of molecular evolution in bacteria.

Evolution (N Y) **69**: 643-652.

Wernegreen, J.J. (2011) Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. *PLoS ONE* **6**: e28905.

Wernegreen, J.J. (2015) Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann N Y Acad Sci* **1360**: 16.

Woodgate, R., Bridges, B., Herrera, G., and Blanco, M. (1987) Mutagenic DNA repair in *Escherichia coli* XIII Proofreading exonuclease of DNA polymerase III holoenzyme is not operational during UV mutagenesis. *Mutation Research/DNA Repair Reports* **183**: 31-37.

Woodgate, R., Singh, M., Kulaeva, O.I., Frank, E.G., Levine, A.S., and Koch, W.H. (1994) Isolation and characterization of novel plasmid-encoded umuC mutants. *J Bacteriol* **176**: 5011-5021.

Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. et al. (2015) Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9**: 1410-1422.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-1591.

Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W. et al. (2014) Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc Natl Acad Sci USA* **111**: 6732-6737.

Ysern, P., Clerch, B., Castaño, M., Gibert, I., Barbé, J., and Llagostera, M. (1990) Induction of SOS genes in *Escherichia coli* and mutagenesis in Salmonella typhimurium by fluoroquinolones. *Mutagenesis* **5**: 63-66.

Zhang, H., Yoshizawa, S., Sun, Y., Huang, Y., Chu, X., González, J.M. et al. (2019) Repeated

evolutionary transitions of flavobacteria from marine to non-marine habitats. *Environ Microbiol* **21**:

648-666.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z. et al. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*.

Zhang, J. (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol* **50**: 56-68.

Zuckerkandl, E., and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H. (eds). In *Evolving genes and proteins*. Academic Press: NY, USA, pp. 97-166.

Figure Legends

Fig. 1   Diversity and abundance of the GH gene families (A) and genome size normalized densities of CAZyme and peptidase genes (B) in the genomes of *Nonlabens* sp. SY33080 and *Leeuwenhoekiella* sp. ZYFB001. In panel B, CAZyme genes are classified into auxiliary activity (AA), carbohydrate-binding module (CBM), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyltransferase (GT), and polysaccharide lyase (PL) categories. Peptidase genes are classified into aspartic, cysteine, metallo, serine, threonine, and unknown families based on the MEROPS annotation.

Fig. 2   Ancestral genome reconstruction using a modified Sankoff algorithm implemented in BadiRate based on a maximum likelihood phylogeny of the marine *Nonlabens*, *Leeuwenhokeiella*, *Dokdonia*, and outgroup lineages (collapsed). The phylogeny was constructed using IQ-TREE based on a concatenation of 163 single-copy orthologous protein sequences. The scale bar at the bottom indicates the number of substitutions per site. Number in parenthesis next to each ancestral node is the number of gene families predicted for that node. Accelerated rate of sequence evolution (defined as increased number of substitutions per site) was observed in the *Nonlabens* lineage compared to the *Leeuwenhokeiella* and the *Dokdonia* lineages since the Last Common Ancestor (node N33) of these three lineages. Solid and open circles at the nodes indicate the frequency of the clade defined by that node is greater and lower than 90%, respectively, in the 1,000 ultrafast bootstrapped replicates. The

genome size, GC content, predicted genome completeness, and presence (solid circle) or absence (open circle) of signature gene families of each taxon are shown next to each taxon. Evolutionary gains and losses of SOS response associated genes are marked with red and blue, respectively, and placed next to the ancestral nodes in *Nonlabens*, *Dokdonia* and *Leeuwenhokeiella* lineages. The full version of the figure with a complete set of outgroup lineages is shown in Fig. S1.

Fig. 3   Base-substitution mutation and insertion/deletion mutation that occurred in each protein-coding gene are marked with blue and red vertical bars, respectively. The height of each bar represents the number of mutations across all mutation accumulation lines. The bars are plotted against the genomic position of that gene in *Nonlabens* sp. SY33080 (upper) and *Leeuwenhoekiella* sp. ZYFB001 (lower). The width of the bar represents the length of the gene. The two types of mutation that occurred on the remaining genomic regions (intergenic regions and non-protein coding genes) are marked with blue and red triangles, respectively, which are plotted against the genomic positions of each occurrence. Note that the genomic position of insertion/deletion mutation refers to the position of the first mutated site.

Table Legends

Table 1 Substrate utilization of *Leeuwenhoekiella* sp. ZYFB001 and *Nonlabens* sp. SY33080 identified using Biolog GN2 Microplates and specialized culture media. Enzyme activities were identified by API 20NE Strips. Substrates marked with bold

are exclusively metabolized by *Leeuwenhoekiella* sp. ZYFB001.

Table 1   Substrate utilization of *Leeuwenhoekiella* sp. ZYFB001 and *Nonlabens* sp. SY33080 identified using Biolog GN2 Microplates and specialized culture media. Enzyme activities were identified by API 20NE Strips. Substrates marked with bold are exclusively metabolized by *Leeuwenhoekiella* sp. ZYFB001.

Biolog GN2 Microplate								
Carbohydrates	*L.*	*N.*	Organic acid	*L.*	*N.*	Amino acid	*L.*	*N.*
Glycogen	+	+	D-Galactonic Acid Lactone	-	-	Succinamic Acid (Organic acid)	-	-
α-Cyclodextrin	+	+	**D-Galacturonic Acid**	+	-	**L-Alaninamide**	+	-
Dextrin	+	+	**D-Gluconic Acid**	+	-	D-Alanine	-	-
D-Raffinose	+	-	D- Glucosaminic Acid	-	-	**L-Alanine**	+	-
D-Cellobiose	+	-	**D-Glucuronic Acid**	+	-	L-Alanyl Glycine	+	+
Gentiobiose	+	-	D-Saccharic Acid	-	-	L-Asparagine	+	+
α-D-Lactose	+	-	L-Aspartic Acid	+	+	L-Histidine	-	-
Lactulose	+	-	L-Glutamic Acid	+	+	**L-Proline**	+	-
Maltose	+	-	Glycyl-L-Aspartic Acid	+	+	L-Leucine	+	+
Sucrose	+	-	Glycyl-L-Glutamic Acid	+	+	L-Ornithine	+	+
D-Trehalose	+	-	L-Pyroglutamic Acid	-	-	L-Phenylalanine	-	-
Turanose	+	-	γ-Aminobutyric Acid	-	-	L-Proline	+	+
D-Melibiose	+	-	Urocanic Acid	-	-	D-Serine	+	+
N-Acetyl-D- Glucosamine	-	-	**Succinic Acid**	+	-	L-Serine	-	-
Adonitol	-	-	Acetic Acid	+	+	L-Threonine	+	+
L-Arabinose	-	-	Cis-Aconitic Acid	-	-	Other substrates	+	-
D-Arabitol	-	-	Citric Acid	-	-	**Pyruvic Acid Methyl Ester**	+	-
Erythritol	-	-	Formic Acid	-	-	**Succinic Acid Mono-Methyl Ester**	+	-
D-Fructose	+	-	α- Hydroxybutyric Acid	-	-	**D-, L-, α-Glycerol Phosphate**	+	-
L-Fucose	-	-	β- Hydroxybutyric Acid	-	-	**Glycerol**	-	-
D-Galactose	+	-	γ- Hydroxybutyric Acid	-	-	2,3-Butanediol	-	-
α-D-Glucose	+	-	p-Hydroxyphenlyacetic Acid	-	-	D-, L-Carnitine	-	-
m-Inositol	-	-	Itaconic Acid	-	-	Phenylethylamine	-	-
D-Mannitol	-	-	α-Ketobutyric Acid	+	+	Putrescine	-	-
D-Mannose	+	-	α-Ketoglutaric Acid	-	-	2-Aminoethanol	+	-
β-Methyl-D-Glucoside	+	-	α-Ketovaleric Acid	+	+	**Uridine**	-	-
α-D-Glucose-1-Phosphate	+	-	**D-, L-Lactic Acid**	+	-	Inosine	-	-
D-Glucose- 6-Phosphate	+	-	Malonic Acid	-	-	Thymidine	-	-
D-Psicose	+	-	Propionic Acid	-	-	Glucuronamide		
L-Rhamnose	+	-	Quinic Acid	-	-			
D-Sorbitol	+	-	Sebacic Acid	-	-			
Xylitol	-	-	**Bromosuccinic Acid**	+	-			
API 20NE Strips								
β-glucosidase	+	-	**β-galactosidase**	+	-			
Specialized Culture Media								
Starch	+	+	Chitin	-	-	Tween 20	+	+
Cellulose	-	-	Gelatin	-	-	Tween 40	+	+
Alginate	-	-	Casein	-	-	Tween 80	+	+

Author/s:
Xue, C-X; Zhang, H; Lin, H-Y; Sun, Y; Luo, D; Huang, Y; Zhang, X-H; Luo, H

Title:
Ancestral niche separation and evolutionary rate differentiation between sister marine flavobacteria lineages

Date:
2020-06-08

Persistent Link:
http://hdl.handle.net/11343/275876

File Description:
Accepted version