

## NOVEL INSIGHT INTO EVOLUTIONARY PROCESS FROM AVERAGE GENOME SIZE IN MARINE BACTERIOPLANKTONIC BIOTA

NI, J.<sup>1</sup> – XU, M.<sup>2\*</sup> – HE, Z.<sup>3</sup> – SUN, G.<sup>4</sup> – GUO, J.<sup>5</sup>

<sup>1</sup>*Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, Guangdong Institute of Microbiology, Guangzhou 510070, China; State Key Laboratory of Applied Microbiology Southern China, Guangzhou 510070, China.  
(phone: +86-020-8768-4471; fax: +86-020-8768-4471)*

<sup>2</sup>*Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, Guangdong Institute of Microbiology, Guangzhou 510070, China; State Key Laboratory of Applied Microbiology Southern China, Guangzhou 510070, China.  
(phone: +86-020-8768-3656; fax: +86-020-8768-4471)*

<sup>3</sup>*Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA.  
(phone: 405-325-3958)*

<sup>4</sup>*Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, Guangdong Institute of Microbiology, Guangzhou 510070, China; State Key Laboratory of Applied Microbiology Southern China, Guangzhou 510070, China.  
(phone: +86-020-8768-4471; fax: +86-020-8768-4471)*

<sup>5</sup>*Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, Guangdong Institute of Microbiology, Guangzhou 510070, China; State Key Laboratory of Applied Microbiology Southern China, Guangzhou 510070, China.  
(phone: +86-020-8768-4471; fax: +86-020-8768-4471)*

*\* Corresponding author  
e-mail: xumy@gdim.cn;  
(phone: +86-020-8768-3656; fax: +86-020-8768-4471)*

(Received 24<sup>th</sup> Jun 2015; accepted 25<sup>th</sup> Jan 2016)

**Abstract.** Genome evolution is an essential force that shapes biodiversity. Genome reduction has been demonstrated in numerous microorganisms and is interpreted as a means of decreasing material cost during DNA synthesis. However, most of the previous studies have been conducted under pure-culture or endosymbiotic conditions, which greatly restrict horizontal gene transfer between microbial species. In nature, microbes frequently interact with each other and gene exchanges are quite common within the microbiota. The evolutionary trend underlying genome size is still poorly understood. The ocean provides an ideal setting for examining the evolutionary process of microbial genomes under natural conditions. We retrieved bacterial compositional information for marine bacterioplanktonic biota from the Visualization and Analysis of Microbial Population Structures database and compared their average genome size along an ocean depth gradient. The results showed that marine bacterioplanktonic biotas tend to minimize their average genome size in the course of their evolutionary processes and that niche differentiation is probably the major force driving genome size reduction. These findings provide novel insight into the understanding of the evolutionary process underlying microbial genome size and mechanisms for maintaining microbial biodiversity in nature.

**Keywords:** *microbial genome evolution; bacterioplanktonic biota; microbial diversity; evolutionary processes*

## Introduction

Bacteria are the most abundant and diverse organisms on Earth (Kuo et al., 2009). They inhabit almost every environment in the biosphere and play crucial roles in biogeochemical cycling of essential elements (Huse et al., 2008; Zhou et al., 2015; Metcalf et al., 2016). Their diversity and maintenance mechanisms have been extensively studied in various habitats as the potential association between their diversity and function (Vellend and Geber, 2005; Chaffron et al., 2010; Monard et al., 2011; Nemergut et al., 2011; Sul et al., 2013). Currently, many ecological microbial patterns have been identified, such as the cosmopolitan distribution of microbial subgroups (Chaffron et al., 2010; de Wit and Bouvier, 2006), the taxa-area relationship (Bell et al., 2005), and the distance-decay relationship (Green and Bohannan, 2006; Bell, 2010; Ni et al., 2014). These observations have expanded our knowledge regarding microbial diversity and maintenance mechanisms. However, most of these studies have focused on diversity at the species level or were based on the species concept.

Although the evolutionary processes of microbial genomes have attracted attention (Andersson and Kurland, 1998; Dufresne et al., 2005; Giovannoni et al., 2005; Nilsson et al., 2005; McCutcheon and Moran, 2010; Wolf and Koonin, 2013), relatively few studies have addressed their role in preserving microbial diversity. Microbial genome size is primarily determined by two opposing processes: acquisition of new genomic fragments by gene duplication or by horizontal gene transfer and deletion of non-essential genomic fragments (Mira et al., 2001; Giovannoni et al., 2005; Nilsson et al., 2005). Several studies have demonstrated that pure-culture and endosymbiotic bacteria tend to reduce their genome size during the evolutionary process (Nilsson et al., 2005; Pérez-Brocal et al., 2006; Marais et al., 2008; Wolf and Koonin, 2013). For instance, Nilsson et al. (2005) report the initial rate of DNA loss is estimated to be 0.05 bp per chromosome per generation in the bacterium *Salmonella enterica* in culture. Phylogenetic analyses also reveal that bacteria with smaller genomes are derived from those with larger ones (Andersson and Kurland, 1998; Wolf and Koonin, 2013). However, this proposed trend has come into question, since interspecies horizontal gene transfer, the primary route by which bacterial species obtain new genes, is restricted under pure culture or endosymbiotic conditions (Ochman et al., 2000). In nature, microbes interact with each other and commonly form a microbial community (i.e. microbiota) within which gene exchanges are quite common (McDaniel et al., 2010; Tuller et al., 2011). For instance, McDaniel et al. (2010) report environmental gene transfer frequencies range from  $6.7 \times 10^{-3}$  to  $4.7 \times 10^{-1}$ , which are 1900 to 459 million times the frequency for transformation and 650000 to 31 million times the frequency of transduction. These exchanges could counteract the influence of genomic fragment deletion. Although Mira et al. (2001) have proposed “a pervasive bias towards higher numbers of deletions than insertions” and that “deletional bias is a major force that shapes bacterial genomes,” this has yet to be proved under natural conditions. The “deletional bias” hypothesis suggests that the average genome size (AGS) of microorganisms would decrease during their evolutionary process in nature.

The evolutionary process on Earth has been maintained in the vertical plane stretching from the deep sea to the ocean surface since ancient times (Carney, 2005; McClain and Hardy, 2010). Therefore, this oceanic vertical plane contains bacterioplanktonic

evolutionary environments, which demonstrate microbial evolutionary processes similar to those exhibited by marine multicellular organisms (Carney, 2005; Brandt et al., 2007; McClain and Hardy, 2010; Plata et al., 2015 and Unpublished data). Thus, the ocean provides an ideal setting for examining the evolutionary process of microbial genomes under natural conditions. Based on the “deletional bias” hypothesis (Mira et al., 2001), the AGS of microorganisms in microbiota would decrease during their evolutionary process in nature. Therefore, we hypothesized that the AGS of the marine bacterioplanktonic biota would decrease from deep sea to surface layers, consistent with the evolutionary processes. To test this, we retrieved bacterial composition information of marine bacterioplanktonic biota from the Visualization and Analysis of Microbial Population Structures (VAMPS) database and compared their AGS along an ocean depth gradient. To the best of our knowledge, this is the first report of AGS analysis at the microbiota level. The findings of this study provide novel insight into the evolutionary processes underlying microbial genome size and microbial biodiversity maintenance mechanisms in nature.

## Materials and Methods

### *Data retrieval*

The bacterial composition information of 63 marine bacterioplanktonic biotas at different water depths range from 0 to 4000 m was retrieved from the VAMPS database ([http://vamps.mbl.edu/portals/icomm/icomm.php/microbis/data\\_exports/download\\_bacteria\\_gz.php](http://vamps.mbl.edu/portals/icomm/icomm.php/microbis/data_exports/download_bacteria_gz.php)) with Project codes KCK\_HOT\_Bv6. The genomic data of 4891 sequenced and annotated bacterial genomes, containing 3257 species were retrieved from the Integrated Microbial Genomes (IMG) system as reference genomic data (Markowitz et al., 2012).

### *Data analysis*

Since the full genomes of most bacterial taxa are not available in public databases, their genome sizes and 16S rRNA gene copy number (GCN) cannot usually be deduced directly from environmental sequence data (Kembel et al., 2012; Tamames et al., 2012; Ni et al., 2013). Hence, the average GCN and genomic size of taxa belonging to the same higher taxonomic category were calculated based on the reference genomic data according to Ni *et al.* (2013) and used as representative of the taxon as the method excludes the bias from the possible weighting effect due to variability in the number of fully sequenced genomes per taxa.

Because of the GCN variation in different bacterial genomes, the organismal proportion of each taxon in the bacterioplanktonic biota was adjusted using the following equation (Kembel et al., 2012; Ni et al., 2013; Angly et al., 2014):

$$p_i = \frac{N_i / (AC_{16S})_i}{\sum_{i=1}^n (N_i / (AC_{16S})_i)} \quad (\text{Eq. 1})$$

where  $p_i$  is the organismal proportion of the  $i^{th}$  taxon,  $N_i$  is the number of 16S rRNA gene reads for the  $i^{th}$  taxon,  $AC_{16S}$  is the average GCN value of the  $i^{th}$  taxon, and  $n$  is the total number of taxa in the bacterioplanktonic biota.

The AGS of each bacterioplanktonic biota was calculated as follows:

$$AGM_j = \sum_{i=1}^n (p_i \cdot AG_i) \quad (\text{Eq. 2})$$

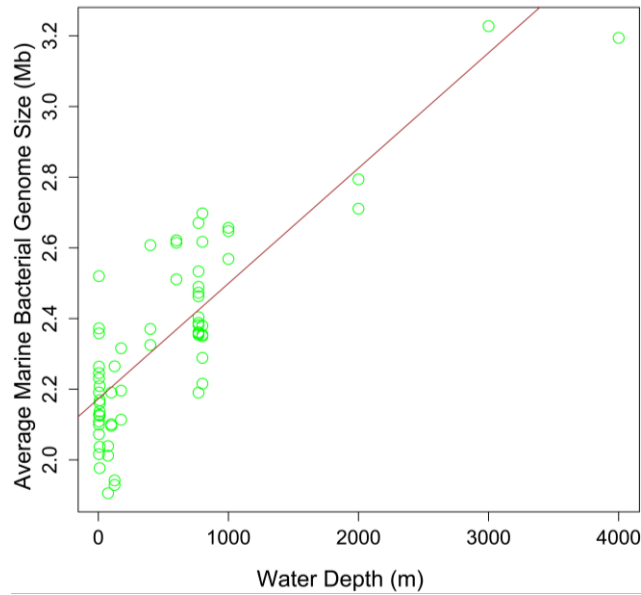
where  $AGS_j$  and  $AG_i$  are the AGS of the  $j^{th}$  bacterioplanktonic biota and the  $i^{th}$  taxon in the  $j^{th}$  bacterioplanktonic biota, respectively.

The Shannon-Wiener, Simpson, and inverse Simpson indices of the bacterioplanktonic biotas, as well as the Bray-Curtis distances between biota at the same depth were calculated using the vegan 2.0-10 package (Dixon, 2003) of the R platform (R Development Core Team, 2006). All calculations and statistical tests were conducted with the R platform.

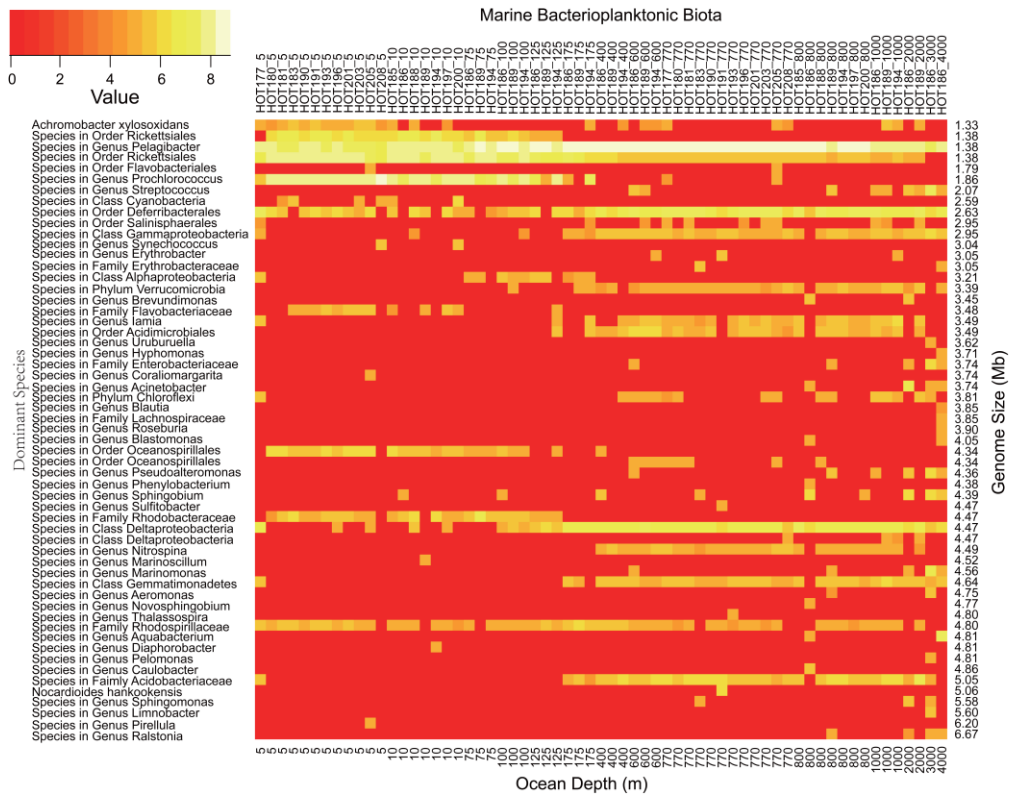
## Results

Since the complete genome information of many bacteria is not available in public databases, approximated genome sizes and GCN were calculated according to previous reports (Ni et al., 2013); these calculated values were generally consistent with our results. Next, we calculated the AGS of each marine bacterioplanktonic biota. The AGS was significantly, positively correlated with ocean depth ( $R^2 = 0.694$ ,  $F = 138.1$ ,  $p < 2.2 \times 10^{-16}$ ; Fig. 1), with a regression equation of  $y = 0.000316x + 2.171$ , where  $y$  is the AGS (Mb) and  $x$  is ocean depth (m). The correlation between the average gene count per genome of each marine bacterioplanktonic biota and ocean depth exhibited the same trend ( $R^2 = 0.662$ ,  $F = 119.6$ ,  $p = 5.17 \times 10^{-16}$ ). These results indicate that the AGS of marine bacterioplanktonic biota decreases from deep sea to surface, consistent with our hypothesis. Similarly, *Prochlorococcus* lineages exhibited a similar evolutionary progression in their genomes. Based on genome data from the IMG system (Markowitz et al., 2012), ecotype eMIT9313, the earliest lineage of *Prochlorococcus* branching from the last common ancestor of *Prochlorococcus* and *Synechococcus*, contains the largest genome (2.41 Mb), followed by ecotype eNATL2A (1.84 Mb), strains restricted to the deep euphotic zone. However, ecotypes eMIT9312 and eMED4, the most recently derived lineages, only contain genomes of 1.71 Mb and 1.66 Mb, respectively, and have spread to the upper euphotic zone (Dufresne et al., 2005; Morris et al., 2011).

In order to understand the mechanism underlying changes in AGS, we compared the composition of dominant species (organismal abundance >1%) in the bacterioplanktonic biota. We observed that the relative abundance of bacteria with smaller genome sizes was higher at the surface than in the deep sea (Fig. 2), indicating that the bacterial composition with different genome size of marine bacterioplanktonic biota changes at various ocean depths, forming a marine bacterioplanktonic biota AGS gradient along the ocean depth gradient.

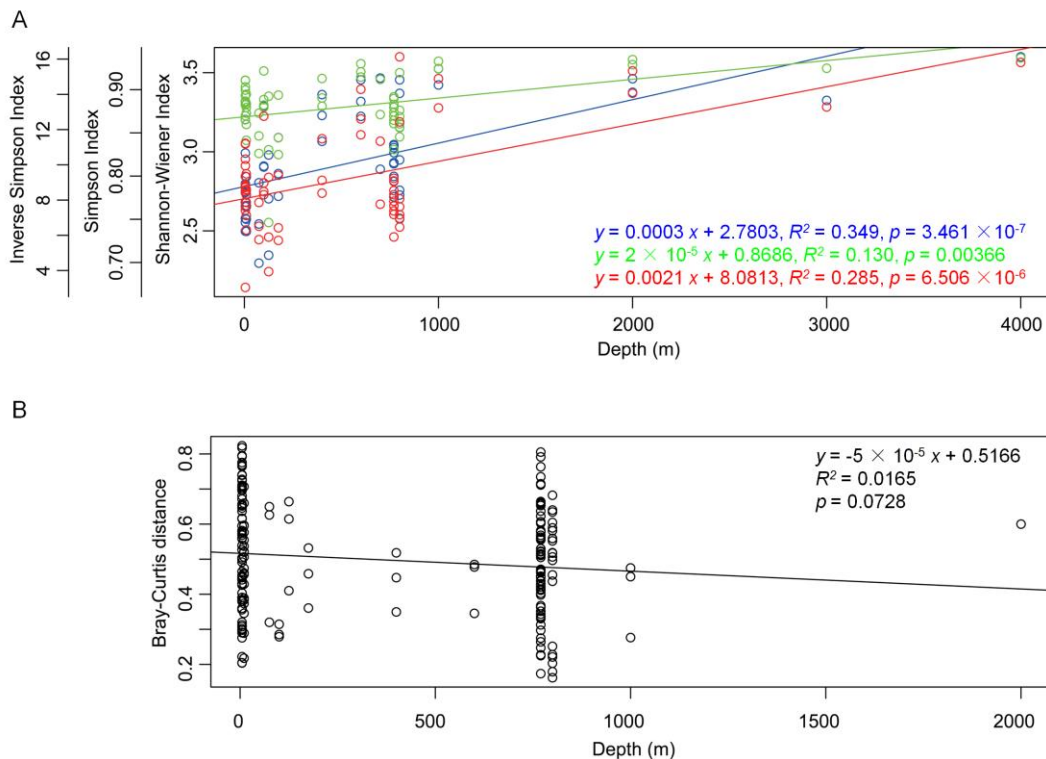


**Figure 1.** The correlation between the average genome size of marine bacterioplanktonic biota and ocean water depth.



**Figure 2.** Heatmap showing abundance of dominant species in marine bacterioplanktonic biota. The heatmap was generated based on the logarithmic transformation of the relative abundance of the dominant species.

The physical and chemical properties of marine water columns show considerable variability at the upper layer of the ocean, but remain relatively constant at greater depth (Robison, 2004; Giovannoni et al., 2014). These observations indicate that oceanic niches differentiated from deep sea layers to the surface. Oceanic niche differentiation has been demonstrated to cause genome divergence in two *Prochlorococcus* ecotypes (Rocap et al., 2003). In addition, genome reduction has been widely observed in symbiotic bacteria that have adapted to a specific symbiotic niche (Gil et al., 2002; Pál et al., 2006; Pérez-Brocal et al., 2006; McCutcheon et al., 2010). Hence, it is plausible that bacterioplanktonic biotas continuously differentiate to better suit specific marine niches formed during evolutionary progression. For example, a bacterial group might reduce its genomic material to better suit a specific marine niche, especially an oligotrophic marine zone (Morris et al., 2002) since genomic reduction could conserve materials and energy for biosynthesis during the process of bacterial reproduction. If this hypothesis is correct, the marine bacterioplanktonic biota alpha-diversity indices, such as the Shannon-Wiener index, the Simpson index, and the inverse Simpson index, would increase with water depth, while the beta-diversity indices (e.g., Bray-Curtis distances) between biotas at the same depth (isobaths) should decrease along the water depth gradient. Our results support this hypothesis with the exception of the Bray-Curtis distance showing a weak non-significant decreasing trend (Fig. 3).



**Figure 3.** Alpha-diversity indices (A) and Bray-Curtis distances (B) of marine bacterioplanktonic biota along the ocean water depth gradient. Blue, green, and red in panel (A) indicate the Shannon-Wiener index, Simpson index, and inverse Simpson index, respectively. Since only one sample was retrieved at 3,000 m and 4,000 m, the Bray-Curtis distance between the biota at the same depth was not calculated.

## Discussion

Genomic reduction could conserve materials and energy for biosynthesis during the process of bacterial reproduction. Since bacterial generation time is relatively short, this reduction is probably quite significant in terms of bacterial competition, especially in oligotrophic ocean environments. The dominance of SAR11, which has the smallest genome known for a free-living heterotrophic cell, in the ocean surface bacterioplanktonic biota, exemplifies the advantage of genome reduction (Morris et al., 2002). Based on the linear relationship between bacterial genome size and gene counts (Mira, 2001; Giovannoni et al., 2005), our study demonstrates that marine bacterioplanktonic biota show an evolutionary trend towards efficiently utilizing their genetic material to adapt to different marine niches. These results suggest that bacteria in current ecosystems may be subjected to powerful selection to minimize the material and energetic cost of cellular replication, concurring with the streamlining hypothesis (Giovannoni et al., 2005). Morris et al. (2012) proposed a Black Queen hypothesis to explain how selection leads to the dependence of free-living organisms with gene loss on co-occurring microbes. However, it is unclear why bacteria possessed large genomes at the start of their evolutionary processes.

The benefits afforded by genome reduction are not without a cost. Since the majority of bacterial genomic sequences are functional protein-coding regions (Mira et al., 2001), AGS reduction in marine bacterioplanktonic biota could result in substantial gene loss. At the species level, gene loss might cause physiological deficiencies. For instance, the lack of catalase and additional protective mechanisms in *Prochlorococcus* genomes leads to high sensitivity to oxidative damage due to hydrogen peroxide (Morris et al., 2011). However, in a bacterioplanktonic community, *Prochlorococcus* could eliminate oxidative damage by coexisting with heterotrophs at the mixed surface layer of the oligotrophic ocean (Morris et al., 2011). Tripp and colleagues (Tripp et al., 2008) provided another example of this cost-benefit; SAR11 possesses the smallest genome among free-living heterotrophic bacteria and is the dominant and ubiquitous lineage in the euphotic zone. However, SAR11 is deficient in assimilatory sulphate reduction genes. Hence, SAR11 requires exogenously reduced sulphurs that originate from other bacterioplanktons for growth (Tripp et al., 2008). Another considerable issue is trade-off of benefits between the acquisition of a function after loss and keeping it. It is reasonable to keep the function if the cost of acquiring the function far outweighs the cost of keeping it and the function is essential for bacterial surviving. However, it is hard to distinguish which function is re-acquired after loss and which one is always retained. Therefore, it cannot compare the cost between re-acquiring a function after loss and keeping it. The trade-off of benefits between the acquisition of a function after loss and keeping it need to be further studied.

A possible explanation for the lower correlation coefficient ( $R^2 = 0.694$ ) between AGS and water depth is the disruption of the vertical bacterioplanktonic distribution pattern by turbulence, a driver of vertical plankton distribution at the subsurface upper ocean layer (Macías et al., 2013). However, although turbulence temporally changes nutrient and bacterioplanktonic community compositions (Lewis et al., 1986; Macías et al., 2013), it does not seem to alter the distribution pattern of some bacterioplanktonic taxa, such as *Prochlorococcus*, at the ocean surface (Morris et al., 2011). The upper euphotic zone is dominated by eMED4 and eMIT9312, which are high-light adapted, while the lower euphotic

zone is dominated by low-light adapted ecotypes including eNATL2A, eMIT9313, and eSS120 (Morris et al., 2011). In deep ocean layers, water features such as temperature, pressure, and salinity are relatively constant (Robison, 2004). Thus, although vertical turbulence would temporally disturb the bacterioplanktonic community composition, it could not change the distribution pattern of bacterioplanktonic biota in a vertical direction.

Many factors (e.g. water temperature and bacterial density) affect the AGS of marine bacterioplanktonic biota. Water temperature could constitute a crucial factor influencing the AGS. For instance, the maximal values of SAR11 relative abundance in marine bacterioplanktonic biota occurred during the summer (Morris et al., 2002). This increase in the relative abundance of this small genome free-living heterotrophic cell could reduce the AGS of the bacterioplanktonic biota. Therefore, water temperature could change the abundance of other bacterial species thus indirectly affecting the AGS. Density is another possible factor influencing AGS. Bacterial density decreases with water depth, especially at 0-1000 m (Morris et al., 2002). The decrease in bacterial density probably reduces mutuality among different bacterial species, which is crucial for the survival of bacteria with small genomes (Morris et al., 2012). Under these conditions, bacteria must maintain a genomic size sufficient to ensure they retain essential genes. The effect of these factors on AGS remains to be confirmed.

The Bray-Curtis distance values demonstrate a weak non-significant decreasing trend (*Fig. 3*). Such non-significance may be due to small sampling areas; the bacterioplanktonic biota habitat at the sampling depth probably did not completely differentiate into different niches. The decreasing trend would be more obvious if the sampling sites were at a greater horizontal distance from each other. Moreover, the evolutionary mechanisms of marine bacterioplanktonic biota could be further elucidated at the level of gene composition via metagenomic sequencing.

Based on the environmental, genome, and evolutionary data, we postulate that niche differentiation is probably the major driving force for AGS reduction in marine bacterioplanktonic biota. These finding, together with future studies, could provide considerable information for protecting the bio- and genetic diversity of microbiota.

**Acknowledgments.** We would like to thank Editage (<http://www.editage.cn/>) for English language editing. This research was supported by the Guangdong Natural Science Foundation (2014A030310281; 2014A030308019) and the Youth Science Foundation of Guangdong Academy of Sciences (qnjj201504) for Jijia Ni, Guangdong Province - Chinese Academy of Sciences strategic cooperative project (2013B091500081), the Special Fund for Agro-scientific Research in the Public Interest (201503108), the Guangdong Province Special Fund for Application Research (2015B020235011) and Guangdong Provincial Science and Technology Foundation for Marine and Fishery Development (A201501D01) for Meiyang Xu. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## REFERENCES

- [1] Andersson, S. G. E., Kurland, C. G. (1998): Reductive evolution of resident genomes. - *Trends in Microbiology* 6: 263-268.
- [2] Angly, F. E., Dennis, P. G., Skarszewski, A., Vanwonterghem, I., Hugenholtz, P., Tyson, G. W. (2014): CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. - *Microbiome* 2: 11.



- [3] Bell, T., Ager, D., Song, J. I., Newman, J. A., Thompson, I. P., Lilley, A. K., van der Gast C. J. (2005): Larger islands house more bacterial taxa. - *Science* 308: 1884.
- [4] Bell, T. (2010): Experimental tests of the bacterial distance-decay relationship. - *ISME Journal* 4: 1357-1365.
- [5] Brandt, A., De Broyer, C., De Mesel, I., Ellingsen, K.E., Gooday, A.J., Hilbig, B., Linse, K., Thomson, M. R., Tyler, P. A. (2007): The biodiversity of the deep Southern Ocean benthos. - *Philosophical Transactions B* 362: 39-66.
- [6] Carney, R. S. (2005): Zonation of deep biota on continental margins. - *Oceanography and Marine Biology* 43: 211-278.
- [7] Chaffron, S., Rehrauer, H., Pernthaler, J., von Mering, C. (2010): A global network of coexisting microbes from environmental and whole-genome sequence data. - *Genome Research* 20: 947-959.
- [8] de Wit, R., Bouvier, T. (2006): "Everything is everywhere, but, the environment selects"; what did Baas Becking and Beijerinck really say? - *Environmental Microbiology* 8: 755-758.
- [9] Dixon, P. (2003): VEGAN, a package of R functions for community ecology. - *Journal of Vegetation Science* 14: 927-930.
- [10] Dufresne, A., Garczarek, L., Partensky, F. (2005): Accelerated evolution associated with genome reduction in a free-living prokaryote. - *Genome Biology* 6: R14.
- [11] Gil, R., Sabater-Muñoz, B., Latorre, A., Silva, F. J., Moya, A. (2002): Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. - *Proceedings of the National Academy of Sciences of the United States of America* 99: 4454-4458.
- [12] Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappé, M. S., Short, J. M., Carrington, J. C., Mathur, E. J. (2005): Genome Streamlining in a cosmopolitan oceanic bacterium. - *Science* 309: 1242-1245.
- [13] Giovannoni, S. J., Thrash, J. C., Temperton, B. (2014): Implications of streamlining theory for microbial ecology. - *ISME Journal* 8: 1553-1565.
- [14] Green, J., Bohannan, B. J. M. (2006): Spatial scaling of microbial biodiversity. - *Trends in Ecology & Evolution* 21: 501-507.
- [15] Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A., Sogin, M. L. (2008): Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. - *PLoS Genetics* 4: e1000255.
- [16] Kembel, S. W., Wu, M., Eisen, J. A., Green, J. L. (2012): Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. - *PLoS Computational Biology* 8: e1002743.
- [17] Kuo, C., Moran, N. A., Ochman, H. (2009): The consequences of genetic drift for bacterial genome complexity. - *Genome Research* 19: 1450-1454.
- [18] Lewis, M. R., Harrison, W. G., Oakey, N. S., Hebert, D., Platt, T. (1986): Vertical nitrate fluxes in the oligotrophic ocean. - *Science* 234: 870-873.
- [19] Macías, D., Rodríguez-Santana, Á., Ramírez-Romero, E., Bruno, M., Pelegrí, J. L., Sangrà, P., Aguiar-González, B., García, C. M. (2013): Turbulence as a driver for vertical plankton distribution in the subsurface upper ocean. - *Scientia Marina* 77: 541-549.
- [20] Marais, G. A. B., Calteau, A., Tenaillon, O. (2008): Mutation rate and genome reduction in endosymbiotic and free-living bacteria. - *Genetica* 134: 205-210.
- [21] Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova N. N., Kyrpides, N. C. (2012): IMG: the integrated microbial genomes database and comparative analysis system. - *Nucleic Acids Research* 40: D115-D122.

- [22] McClain, C. R., Hardy, S. M. (2010): The dynamics of biogeographic ranges in the deep sea. - *Proceedings of the Royal Society B* 227: 3533-3546.
- [23] McCutcheon, J. P., Moran, N. A. (2010): Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. - *Genome Biology and Evolution* 2: 708-718.
- [24] McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B., Paul, J. H. (2010): High frequency of horizontal gene transfer in the oceans. - *Science* 330: 50.
- [25] Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., Hyde, E. R., Song, S. J., Amir, A., Larsen, P., Sangwan, N., Haarmann, D., Humphrey, G. C., Ackermann, G., Thompson, L. R., Lauber, C., Bibat, A., Nicholas, C., Gebert, M. J., Petrosino, J. F., Reed, S. C., Gilbert, J. A., Lynne, A. M., Bucheli, S. R., Carter, D. O., Knight, R. (2016): Microbial community assembly and metabolic function during mammalian corpse decomposition. - *Science* 351: 158-162.
- [26] Mira, A., Ochman, H., Moran, N. A. (2001): Deletional bias and the evolution of bacterial genomes. - *Trends in Genetics* 17: 589-596.
- [27] Monard, C., Vandenkoornhuise, P., Le Bot, B., Binet, F. (2011): Relationship between bacterial diversity and function under biotic control: the soil pesticide degraders as a case study. - *ISME Journal* 5: 1048-1056.
- [28] Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M., Zinser, E. R. (2011): Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. - *PLoS ONE* 6: e16805.
- [29] Morris, J. J., Lenski, R. E., Zinser, E. R. (2012): The Black Queen hypothesis: evolution of dependencies through adaptive gene loss. - *mBio* 3: e00036-12.
- [30] Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., Giovannoni, S. J. (2002): SAR11 clade dominates ocean surface bacterioplankton communities. - *Nature* 420: 806-810.
- [31] Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., Fierer, N., Townsend, A. R., Cleveland, C. C., Stanish, L., Knight, R. (2011): Global patterns in the biogeography of bacterial taxa. - *Environmental Microbiology* 13: 135-144.
- [32] Ni, J., Yan, Q., Yu, Y. (2013): How much metagenomic sequencing is enough to achieve a given goal? - *Scientific Reports* 3: 1968.
- [33] Ni, J., Yan, Q., Yu, Y., Zhang, T. (2014): Fish gut microecosystem: a model for detecting spatial pattern of microorganisms. - *Chinese Journal of Oceanology and Limnology* 32: 54-57.
- [34] Nilsson, A. I., Koskineemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C., Andersson, D. I. (2005): Bacterial genome size reduction by experimental evolution. - *Proceedings of the National Academy of Sciences of the United States of America* 102: 12112-12116.
- [35] Ochman, H., Lawrence, J. G., Groisman, E. A. (2000): Lateral gene transfer and the nature of bacterial innovation. - *Nature* 405: 299-304.
- [36] Pál, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., Hurst, L. D. (2006): Chance and necessity in the evolution of minimal metabolic networks. - *Nature* 440: 667-670.
- [37] Pérez-Brocá, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., Silva, F. J., Moya, A., Latorre, A. (2006): A small microbial genome: the end of a long symbiotic relationship. - *Science* 314: 312-313.
- [38] Plata, G., Henry, C. S., Vitkup, D. (2015): Long-term phenotypic evolution of bacteria. *Nature* 517: 369-72.
- [39] R Development Core Team. (2006): R: A language and environment for statistical computing. - R Foundation for Statistical Computing: Vienna, Austria.

- [40] Robison, B. H. (2004): Deep pelagic biology. - *Journal of Experimental Marine Biology and Ecology* 300: 253-272.
- [41] Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R., Chisholm, S. W. (2003): Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. - *Nature* 424: 1042-1047.
- [42] Sul, W. J., Oliver, T. A., Ducklow, H. W., Amaral-Zettler, L. A., Sogin, M. L. (2013): Marine bacteria exhibit a bipolar distribution. - *Proceedings of the National Academy of Sciences of the United States of America* 110: 2342-2347.
- [43] Tamames, J., de la Peña, S., de Lorenzo, V. (2012): COVER: a priori estimation of coverage for metagenomic sequencing. - *Environmental Microbiology Reports* 4: 335-341.
- [44] Tripp, H. J., Kitner, J. B., Schwabach, M. S., Dacey, W. H., Wilhelm, L. J., Giovannoni, S. J. (2008): SAR11 marine bacteria require exogenous reduced sulphur for growth. - *Nature* 452: 741-744.
- [45] Tuller, T., Girshovich, Y., Sella, Y., Kreimer, A., Freilich, S., Kupiec, M., Gophna, U., Ruppin, E. (2011): Association between translation efficiency and horizontal gene transfer within microbial communities. - *Nucleic Acids Research* 39: 4743-4755.
- [46] Vellend, M., Geber, M. A. (2005): Connections between species diversity and genetic diversity. - *Ecology Letters* 8: 767-781.
- [47] Wolf, Y. I., Koonin, E. V. (2013): Genome reduction as the dominant mode of evolution. - *Bioessays* 35: 829-837.
- [48] Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S. G., Alvarez-Cohen, L. (2015): High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. - *mbio* 6: e02288-14.