



Evolutionary constraints on species diversity in marine bacterioplankton communities

Hsiao-Pei Lu¹ · Yi-Chun Yeh² · Fuh-Kwo Shiah^{1,3,4} · Gwo-Ching Gong^{4,5} · Chih-hao Hsieh^{1,3,6,7}

Received: 27 October 2017 / Revised: 23 July 2018 / Accepted: 11 December 2018 / Published online: 3 January 2019
© International Society for Microbial Ecology 2019

Abstract

Variation in microbial species diversity has typically been explained as the outcome of local ecological factors driving species coexistence, overlooking the roles of evolutionary constraints. Here, we argue that macro-evolutionary niche conservatism and unequal diversification rates among phylum-level lineages are strong determinants of diversity–environment relationships in bacterial systems. That is, apart from stochasticity, environmental effects operate most strongly on phylum composition, which in turn dictates the species diversity of bacterial communities. This concept is demonstrated using bacterioplankton in the surface seawaters of the East China Sea. Furthermore, we show that the species richness of a local bacterioplankton community can generally be estimated based on the relative abundances of phyla and their contributions of species numbers in the global seawater pool—highlighting the important influence of evolutionary constraints on local community diversity.

Introduction

The core objective in community ecology is to assess and explain how species diversity varies along environmental gradients [1–4]. In studies of plant and animal communities, some interesting patterns have been recognized [5],

for example, the latitudinal diversity gradient: species diversity generally increases from the poles towards the equator, showing a positive relationship with temperature, precipitation, as well as productivity [6–8]. Many ecological theories trying to explain such diversity–environment relationships consider mainly in situ resource availability and species interactions [9–11], stressing the importance of local drivers in shaping the diversity of a community [12]. However, it is increasingly recognized that local communities may bear the imprints of macro-scale effects, such as speciation, long-distance dispersal, and unique historical events, which have the potential to explain the differences in local diversity [13–15]. Indeed, as found in plant and animal communities, species diversity in the continental / regional source pool exerts a strong influence on the variation in local diversity [16–19].

Regarding the determinant role of species pools [20], the diversity of local communities depends upon the structure of the regional species pool, which in turn, depends upon the global species pool; importantly, the global species pool eventually is determined by the outcomes of long-term diversification dynamics [15]. Thus, for explaining the variation in species diversity among local communities, in addition to present-day ecological processes (e.g. species sorting), one should account for past opportunities for phylogenetic divergence and adaptation, which determine

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-018-0336-1>) contains supplementary material, which is available to authorized users.

✉ Chih-hao Hsieh
chsieh@ntu.edu.tw

- ¹ Institute of Oceanography, National Taiwan University, Taipei, Taiwan
- ² Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA
- ³ Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan
- ⁴ Institute of Marine Environment and Ecology, National Taiwan Ocean University, Keelung, Taiwan
- ⁵ Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan
- ⁶ Department of Life Science, Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan
- ⁷ National Center for Theoretical Sciences, Taipei, Taiwan

the number of available species associated with each particular environment at the geological time scale [21–24].

While biogeographical studies on plants and animals [25–27] have acknowledged that present-day diversity patterns are strongly influenced by species pools structured by macro-evolutionary mechanisms, studies on microorganisms such as bacteria typically do not consider how long-term, broad-scale processes influence the current observations on local community diversity. Since bacteria have evolved over > 3.5 billion years and occupy a wide variety of habitats on Earth [28], macro-evolutionary drivers should be crucial for the contemporary diversity patterns of bacteria. For example, diversification rates and habitat associations of bacterial lineages likely exert a strong influence on bacterial community structure [29–31]. However, little attention has been paid to the importance of bacterial species pools, which can constrain bacterial diversity across environments. In this study, we attempt to emphasize the effects of long-term, broad-scale processes on local bacterial communities, paying particular attention to the structure of the species source pool and its influence on observed diversity–environment relationships.

To call attention to the importance of long-term, broad-scale processes in local bacterial communities, we emphasize two characteristics of bacterial diversification: (1) niche conservatism within bacterial lineages and (2) unequal diversification rates among bacterial lineages. First, regarding niche conservatism [32], an accumulating volume of evidence indicates that certain ecological characteristics (e.g. habitat requirements and functional traits) are conserved for bacterial taxa within the same lineages, showing phylogenetic signals [33–35]. Moreover, with respect to a high-level taxonomic organization, class- or phylum-level community composition of bacteria can still display spatial and temporal patterns with environmental heterogeneity [36–38], suggesting the existence of niche conservatism. For example, in soil systems, pH and carbon availability have been shown as good predictors of phylum-level abundances [39, 40], implying that most members of a given phylum exhibit similar responses to environmental gradients [41]. Second, regarding diversification rates [42], recent studies have explored the evolution and speciation of prokaryotic life based on time-calibrated phylogenies [43, 44], of which a substantial variation in diversification rates has been detected across higher level bacterial lineages [44]. Considering the aforementioned macro-evolutionary characteristics of bacteria, we attempt to reveal the determinants of species diversity in bacterial communities.

In line with the previous findings, our analysis focusing on 16S rDNA reference sequences in the SILVA database [45] indicates that different bacterial phyla contain extremely unequal numbers of species-level units, with over a thousandfold difference between species-rich and species-

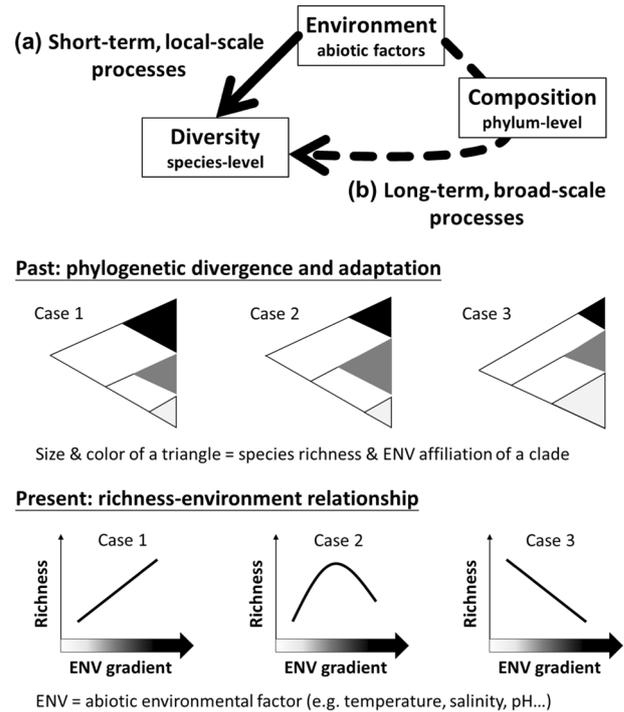


Fig. 1 A conceptual model for explaining diversity–environment relationships. **a** Under short-term, local-scale processes, the variation in species diversity among local communities is regulated by changes in abiotic environmental factors, e.g. temperature. Particular diversity–environment relationships are driven by in situ resource allocations and species interactions. **b** Under long-term broad-scale processes, the environmental effects operate most strongly on phylum composition, which in turn dictates species diversities of local communities. The observed diversity–environment relationships reflect dual macro-evolutionary consequences: distinct phylogenetic lineages (1) have different diversification rates and (2) exhibit unique environmental affiliations. Here, the potential confines of past phylogenetic divergence and adaptation on present-day patterns are illustrated as three hypothetical cases. Notably, the specific diversity–environment relationship may differ depending on the taxonomic groups and environmental variables of consideration in each case

poor lineages (Supplementary Table S1). Moreover, we calculated net diversification rates [42] for phylum-level lineages and detected up to a sixfold difference in diversification rates across bacterial phyla (Supplementary Table S2), with a significant positive relationship between species richness and diversification rates ($r^2 = 0.79–0.88$, $p < 0.001$; based on log-transformed richness). Thus, we anticipate that the species-level diversity of bacterial communities might be constrained by phylum-level composition, as higher taxonomic units preserve the imprint of evolutionary diversification events.

Here, we propose a conceptual model integrating processes at short term, local scale and long term, broad scale for the interpretation of diversity–environment relationships (Fig. 1). For short-term, local-scale processes to operate, we assume that contemporary species are independent units, and local environmental conditions (here only abiotic factors are

considered) strongly affect species sorting and local community assembly, resulting in particular diversity–environment relationships. In contrast, for long-term, broad-scale processes, we assume that species are coherent groups according to phylogenetic relatedness, and local environmental factors operate most strongly on phylum composition, which in turn dictates the species diversity of local communities. Of this path, a specific diversity–environment relationship can be observed when phylogenetic groups have different diversification rates and exhibit distinct (but phylogenetically conserved) environmental preferences. As shown in Fig. 1, we argue that in addition to short-term, local-scale processes, phylogenetic divergence and adaptation in the past likely lead to certain relationships between species richness and environmental variables. For example, in communities dominated by two lineages, if one lineage has adapted to high-pH areas and shows a high diversification rate, and the other has adapted to low-pH areas and shows a low diversification rate, we would expect to detect a positive diversity–pH relationship when both lineages retain their pH-related traits through speciation events (Case 1 of Fig. 1).

To demonstrate this concept, we used bacterioplankton communities in the surface seawaters of the East China Sea (ECS) as a case study. A previous work in this system has documented substantial heterogeneity in both environmental conditions and bacterial community structure [37, 46, 47]. In the ECS ecosystem, in addition to the unequal intra-clade species richness among the bacterial phyla (Supplementary Table S1), we do detect specific environmental responses of phylum-level abundances (Supplementary Table S3) and significant phylogenetic signals in environmental niche preferences (Supplementary Table S4). Accordingly, we anticipate strong relationships among an environment, its phylum-level composition, and its species-level diversity in the ECS ecosystem. Moreover, in line with the species-pool hypothesis, we anticipate that predictors of a local community's species richness will include the relative abundances of phyla and their contributions of species numbers to the source pool. In relation to the sampling scale, we define a hierarchical series of species source pools (Supplementary Figure S1 and Supplementary Table S1) for the prediction of species richness in a local ECS bacterioplankton community. We expect to detect a high consistency between observed and predicted values of species richness, if the structure of the species source pool has a strong influence on local community diversity.

Material and methods

Environmental sampling

A total of 96 surface seawater samples (from 2- to 5-m depth) of ECS were collected from nine cruises during hot

and cold seasons of 2010–2012 (Supporting List S1), using a rosette sampler equipped with 20-L Go-Flo bottles (General Oceanics, Miami, FL, USA). These samples spanned ~6° of latitude and ~7° of longitude (Supplementary Figure S2), covering the spatiotemporal variation of environmental conditions and bacterial communities in the ECS ecosystem.

Bacterioplankton cells in each water sample (~18L) were pre-filtered through a 1.2- μ m pore-size polycarbonate membrane and then collected on a 0.2- μ m pore-size polycarbonate membrane (Millipore, Bedford, MA, USA) onboard [48]. These membranes were frozen in liquid nitrogen onboard and stored at –20°C after each cruise.

Temperature and salinity were recorded by a CTD profiler (Sea-Bird, Bellevue, WA, USA). Nutrients (including phosphate, nitrite, nitrate, and silicate) and chlorophyll *a* were measured according to standard methods [49, 50].

Sequencing of bacterial communities

Total genomic DNA was extracted with the Meta-G-Nome™ DNA Isolation Kit (Epicentre, Madison, WI, USA), according to the manufacturer's instructions. To determine the structure of bacterial communities, the hypervariable V6 region of the 16S rRNA gene was amplified using bacterial universal primers (967F and 1064R) [51] and sequenced on a Roche 454 GS FLX Sequencing System (Branford, CT, USA). The specific details regarding PCR amplification and sequencing preparation have been described previously [47]. Raw sequence data have been deposited in the NCBI Sequence Read Archive under the accession number SRX183038.

Sequence processing

Sequences were processed using the Quantitative Insights Into Microbial Ecology (QIIME v. 1.9.1) platform [52]. To minimize the effects of random-sequencing errors, we eliminated 1) sequences that did not perfectly match with the primers and barcodes; 2) sequences that contained > 1 undetermined nucleotides; and 3) sequences with an average quality score < 25. After removing low-quality sequences, OTU (operational taxonomic unit) picking and taxonomic assignment against the SILVA.v123 reference set [45] was carried out using Usearch61 [53] with chimera checking. Sequences affiliated to archaea and chloroplasts were removed. We obtained > 4,000,000 qualified and annotated sequences across 96 sampling sites, with maximum and minimum sequencing depths of ~86,000 and ~10,000. To fairly compare the community structure across the sites, all community analyses were performed based on 100 rarefied OTU tables with equal number of sequences (i.e. 10,000) per site through random sampling (without

replacement) of the original OTU table in the QIIME platform [52].

Bacterial community structure

To represent the community variation across the sampling sites, both diversity and composition of bacterial taxa were calculated. For diversity analysis, we grouped sequences into 99, 97, and 94% OTUs (roughly referring to subspecies-, species-, and genus-level taxa) [54] against the SILVA.v123 reference set [45]. The observed OTU richness of each local community was calculated as the average number of OTUs detected in the 100 randomly rarefied OTU tables. Considering numerous rare taxa in microbial communities, in addition to the observed OTU richness, Chao1 [55] and ACE [56] indices were calculated as the richness estimators. For composition analysis, the phylum-level composition was summarized based on the taxonomic assignment of each OTU, which was retrieved according to consensus taxonomy given by the SILVA database [57].

Path modeling

With the notions from phylogeny-based biogeography [21, 24], we hypothesize that different extents of diversification radiations and habitat associations of bacterial lineages exert strong influences on the contemporary bacterial diversity across environments. Specifically, we formulated a causal model integrating the effects from both short-term, local-scale and long-term, broad-scale processes on the variation in local community diversity (Fig. 1). Considering the interconnection among environmental conditions, phylum-level composition, and species-level diversity, path modeling was used as a means of analyzing systems involving multiple causal relationships to provide directed dependencies among these three components [58].

Focusing on the ECS ecosystem, we evaluated the significance of paths among seawater environment, bacterioplankton composition, and bacterioplankton diversity, with the package “plspm” [59] and the package “sem” [60] in the R statistical computing platform [61]. Notably, although path modeling allows to assess the tenability of the model based on reasonable causal hypotheses, it has restrictive assumptions such as the linearity between predictor and criterion variables, and non-collinearity among predictor variables [62]. Thus, before path modeling, relationships between community diversity and each environmental variable (including temperature, salinity, phosphate, nitrite, nitrate, silicate, and chlorophyll a) were examined using univariate linear regression models. Regressions and corresponding residuals were checked graphically to screen for linearity (based on either original or log-transformed values). In general, bacterial species richness was

negatively correlated with temperature while positively correlated with log-transformed nutrient concentrations, such as phosphate (Supplementary Figure S3). Because measured environmental variables strongly covaried with each other (Supplementary Figure S4), we reduced the variables into a set of uncorrelated values through principal component analysis (PCA). The first principal component (PC1), accounting for 61% of the total variance, was used as a proxy for representing the overall environmental heterogeneity in path modeling. PC1 is positively correlated with nutrient concentrations while negatively correlated with temperature (Supplementary Figure S4). For phylum-level composition, as Proteobacteria and Cyanobacteria accounted for over 90% of the total abundance in the ECS communities (Supplementary Figure S5); their relative dominance would determine the whole compositional variation. Thus, in path modeling, we used the log ratio of Proteobacteria% to Cyanobacteria% as a proxy for representing the overall phylum-level compositional variation. For diversity variation in the path modeling, in addition to species-level richness (97% OTUs), the path coefficients were recalculated based on subspecies-level (99% OTUs) and genus-level (94% OTUs) richness as well as Chao1 and ACE diversity estimators.

Variation partitioning

In addition to path modeling, we performed variation partitioning [63, 64] to assess the relative explanatory power of environmental factors (E) and phylum composition (P) on the variation in species richness, with the package “vegan” [65] in the R statistical computing platform [61]. This method can provide complementary results to the findings from path modeling. Specifically, the variation in species richness is partitioned into four independent components: pure E, pure P, E + P, and undetermined. Notably, the shared component (E + P; not an interaction term) simply reflects the variation that could be explained by both the explanatory matrices. The E matrix contains temperature, salinity, silicate, phosphate, nitrite, nitrate, and chlorophyll a. The P matrix contains Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes, Marinimicrobia, Planctomycetes, Proteobacteria, and Verrucomicrobia. Here, collinear variables in the explanatory tables do not need to be removed prior to variation partitioning, since collinearity has no impact on the associated statistics such as R^2 and p values [66].

Delineation of species source pools

Based on the species-pool concept [20], the species richness of a local community is expected to be constrained by the structure of the species source pool. The

main difficulty for testing the species-pool hypothesis is to define a set of species potentially able to occur in targeted local communities. For the ECS bacterioplankton communities, we defined four hierarchical species source pools (Supplementary Figure S1 and Supplementary Table S1) in relation to the sampling scale: (a) the ECS-surface-seawater pool containing the set of species detected in the surface seawaters of ECS in this study; (b) the global-surface-seawater pool containing the set of species detected in the surface seawaters of global oceans; (c) the global-marine-environment pool containing the set of species detected across the various marine environments including seawaters, sediments, biofilms, and host-associated habitats of global oceans; and d) the whole-contemporary-earth pool containing the set of species detected in a variety of environments of the contemporary earth, such as waters, soils, and guts from both lands and oceans. Since primer pairs would greatly affect the coverage of species for each taxonomic group [67], we only collected 16S rDNA sequences generated by the same pair of primers as the ECS data. For the b and c pools, data were obtained from the International Census of Marine Microbes [68]; see Supporting List S2 for the detailed list. For the d pool, data were obtained from the representative sequences of the Visualization and Analysis of Microbial Population Structures [69], which incorporated > 2000 datasets from all online projects. Those sequence data were assigned OTUs and taxonomy against the SILVA.v123 reference set as the ECS data.

Prediction of species richness

According to the species-pool hypothesis, species richness at a smaller scale is primarily determined by the availability of the species at a corresponding larger scale called ‘proportional sampling’ [13, 20, 70]. Borrowing this concept, we assume that ‘proportional sampling’ is phylum-dependent for bacterial communities, since a bacterial species pool typically contains species derived from multiple phylum-level lineages.

For each species source pool mentioned above (pools a–d), we estimated the species richness contributed by each bacterial phylum by rarefaction with the equal number of sequences (i.e. 10,000) to control the sampling effect. For simplicity, we assume that all species of a certain phylum in the species source pool are functionally equivalent and have an equal chance to occur in a local community. With these assumptions, we predicted the species richness of each ECS bacterioplankton community using the following formula:

$$\text{Predicted species richness} = \sum_{i=1}^n \text{phylum } P_i \times \text{phylum } S_i$$

where P_i = the proportion of a certain phylum i in a local community, S_i = the species richness of a certain phylum i in the species source pool, and n = the number of distinct phyla. That is, the species richness is determined by multiplying the relative abundance of various phyla present in a local community with the number of species contributed by those phyla to the species source pool. The highest or lowest predicted value would be equal to the rarefied species richness of the most species-rich or species-poor phylum when the community contains 100% of the given phylum. Notably, the resolution of this prediction formula depends on the level of variation in intra-clade species richness among different phyla as well as on the variation in phylum composition across local communities. In the present case, to evaluate the predictions based on the four hierarchical species source pools, Pearson’s correlation coefficient was calculated between the observed and predicted species richness, assuming there is a linear relationship between the observed and predicted values.

Results

Based on simple correlation analysis, we detected significant relationships among seawater environment (PC1), bacterioplankton phylum composition, and bacterioplankton species diversity (Fig. 2a). In terms of phylum composition, communities were predominated by Proteobacteria and Cyanobacteria (Supplementary Figure S5). Importantly, their relative dominance ratio (i.e. the log (Proteo/Cyano)) increased significantly with environmental PC1 (Fig. 2a). In terms of species diversity, the number of 97% OTUs (i.e. observed species richness) ranged between 89 and 783 (355 ± 144 ; mean \pm SD) among communities, showing a positive correlation with environmental PC1 (Fig. 2a). Moreover, a strong relationship between the phylum dominance ratio and the observed species richness was detected (Fig. 2a), suggesting a synchronous change in phylum composition and species diversity in the ECS bacterioplankton communities. Specifically, the dominance of Proteobacteria is associated with communities featured by higher species richness, whereas the dominance of Cyanobacteria is associated with that by lower species richness.

However, path modeling results showed significant effects from seawater environment to bacterioplankton phylum composition and from phylum composition to bacterioplankton species diversity, whereas the environment–diversity path was not significant (Fig. 2b). When removing the environment–diversity path, the model still showed a good fit to our data, as indicated by the non-significant χ^2 test ($N = 96$, $\chi^2 = 2.23$, d.f. = 1, $p > 0.1$). These results suggest that the environmental effects operate

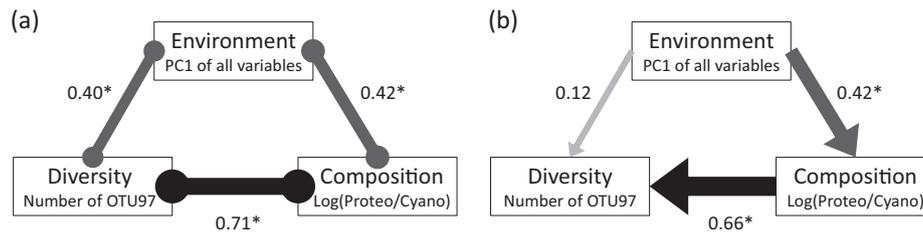


Fig. 2 Relationships among environmental variables, phylum-level composition, and species-level richness (97% OTUs) of the ECS bacterioplankton communities: **a** Results of simple correlation;

b results of path modeling. Values along lines or arrows indicate standardized coefficients ($*p < 0.01$)

most strongly on phylum composition, which in turn dictates the species diversity of bacterioplankton communities. This conclusion remains valid for 99 and 94% of OTUs (subspecies- and genus-level richness; Supplementary Figures S6, S7) as well as for Chao1 and ACE diversity estimators (Supplementary Figure S8, S9), suggesting that the fine-level taxonomic diversity is generally dictated by the phylum composition, and this determinant might remain when considering the unsampled rare taxa in the community.

Moreover, variation partitioning results showed that the pure environmental effect only accounted for a tiny fraction (3%) of the variation in species richness, whereas the pure effect of phylum composition contributed over 40% of the variation (Supplementary Figure S10), with a large amount of the variation (35%) shared by both the effects. Here, the shared explanatory power by environmental factors and phylum composition may be treated as a phylogenetically constrained component of environmental influence, as the conceptual model we proposed (Fig. 1). In line with the findings from path modeling, the results of variation partitioning indicate that phylum composition might be the primary determinant of species diversity observed in local communities.

The results from path modeling and variation partitioning suggest that the species richness of a local ECS bacterioplankton community might be predicted based on the phylum composition with the known structure of the species source pool (i.e. our prediction formula). Here, the intra-clade species richness of each bacterial phylum was estimated based on rarefaction from the four hierarchical species source pools (Supplementary Figure S11). We detected a high correlation between observed and predicted species richness (Fig. 3), regardless of the species source pools (Supplementary Figure S12). Notably, while the estimated numbers of species from distinct phyla vary across the four pools, their rankings are very similar (Supplementary Figure S11); thus, we can detect consistent high observation–prediction correlations based on all the four pools (Supplementary Figure S12). However, in terms of absolute values, we noted that both the pools a and b can provide

roughly accurate predictions for the local species richness in the ECS, whereas the predictions are two- and threefold overestimated with pools c and d (Supplementary Figure S12). Here, the interesting thing is that the global seawater pool can give predictions as good as the ECS seawater pool, indicating that with respect to species that potentially occur in our targeted bacterioplankton communities, the surface seawaters at a regional (i.e. ECS) to global scale may be considered as a biogeographically homogeneous space for bacteria to maintain a species pool.

Notably, although the species richness of the whole communities showed a positive correlation with environmental PC1, the number of species from each individual phylum (Proteobacteria or Cyanobacteria) did not show a significant relationship with environmental PC1 (Supplementary Table S5). Specifically, when considering the whole community diversity (of which unequal numbers of species are derived from various phylum-level lineages), the diversity–environment relationships were significant based on PC1 and most environmental variables (Supplementary Table S5), whereas these diversity–environment relationships were relatively weak and non-significant when considering the species richness within a given phylum (Supplementary Table S5).

Discussion

In this study, we applied the species-pool hypothesis to the diversity–environment relationship of bacterial communities. We found that environmental effects on bacterioplankton species diversity might operate most strongly on phylum composition, which in turn dictates the species diversity of local communities (Fig. 2). Our results support the importance of considering intra-clade diversity of different phylogenetic lineages for interpreting the diversity–environment relationship [22, 23]. Specifically, two dominant bacterial phyla, Proteobacteria and Cyanobacteria, are involved in the present case, in which the diversity–environment relationship emerges because of dual circumstances, which are as follows: 1) these two phyla exhibit

opposite preferences along environmental gradients and 2) they contribute unequal numbers of species to the species source pool. Therefore, as a consequence of evolutionary constraints, the species richness of a local bacterioplankton community can generally be estimated with a known phylum composition plus an appropriate species source pool (Fig. 3). While the influence of long-term, broad-scale processes on the diversity–environment relationship has been demonstrated for plants and animals [25–27], here we, for the first time, show it in bacteria.

Based on the 16S rDNA-based phylogeny, we found that the contributions of species numbers to the contemporary species pool vary greatly among bacterial lineages, with Proteobacteria and Cyanobacteria, respectively, contributing > 30 and ~ 2% of the species to the species source pools, regardless of the defined range of the pool (Supplementary Table S1). Actually, topologically imbalanced phylogenetic trees (with a few species-rich lineages and many species-poor lineages) have long been recognized in macro-organisms [71, 72]. More importantly, since phylogenetic niche conservatism would largely constrain species source pools across environments [24], researchers have suggested that global diversity patterns should be associated with the diversification and adaptation of lineages; for example, the well-recognized latitudinal diversity gradient might be related to old and vigorous lineages in the tropics, and relatively young and limited lineages might be adapted to temperate areas [21, 26, 73]. In agreement with the notions in macro-organisms, our results showing a strong connection between phylum-level composition and species-level diversity in marine bacterioplankton stress the importance of accounting for evolutionary constraints when explaining modern bacterial diversity patterns.

Regarding the latitudinal diversity gradient of marine bacterioplankton, we speculate that the relative dominance of Proteobacteria vs Cyanobacteria may at least partly determine the variation in species richness in surface seawaters since these two phyla are dominant in not only the ECS ecosystem but also the global oceans [74–76]. Previous studies have found that, unlike the diversity of macro-organisms that generally increases from the poles towards the equator, the diversity of marine bacterioplankton seems to peak at mid to high latitudes [74, 77]. Here, we suppose that one possible explanation for the lower diversity in tropical waters vs higher diversity in temperate waters is associated with the relative dominance of Proteobacteria vs Cyanobacteria, as observed in the ECS ecosystem. Indeed, the abundance of Cyanobacteria has been found to increase toward the equator [75], with a relatively high proportion of Proteobacteria in mid-latitudes. In terms of the species-diversity pattern, since Proteobacteria contributes the majority of species richness to the global oceans [74, 76], the latitudinal diversity gradient of marine bacterioplankton

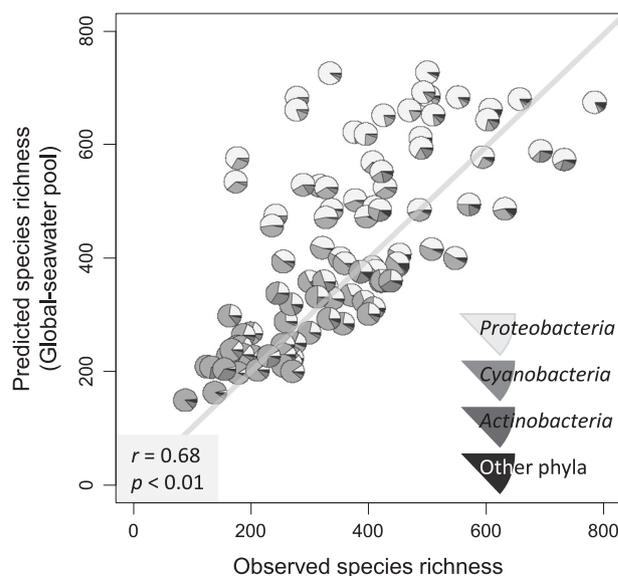


Fig. 3 Correlation between observed and predicted species richness of the ECS bacterioplankton communities. The predicted species richness is calculated based on the global-surface-seawater pool. For each community, the pie chart shows the phylum composition. Other predictions based on a hierarchical series of species pools (cf. Supplementary Figure S1) are shown in Supplementary Figure S8. The gray line indicates the 1:1 line

in surface seawaters, to some extent, is accepted as our conceptual model (Fig. 1). However, general conclusions require more samples from broad marine regions to span environmental gradients [78, 79]. Moreover, there is a need to consider biotic factors (such as interactions with viruses and eukaryotic microbes), which might have important roles in determining bacterial community diversity [80, 81].

It is notable that the species-pool hypothesis is a kind of “null model” [13, 20, 70], which is complementary, rather than alternative, to ecological theories based on local-scale processes. Our findings emphasize that large-scale processes acting at the evolutionary time scale should be considered when interpreting contemporary species-diversity patterns; nevertheless, local-scale processes may still have strong effects on the observed diversity variation. In the present case, despite a high correlation between observed and predicted species richness, some predicted values clearly departed from the observed values, with large mismatches detected when the communities were dominated by Proteobacteria (Fig. 3). Focusing on those communities, the observed species richness varies greatly from ~200 to ~800, whereas this variation is not clearly associated with the proportion of Proteobacteria (Supplementary Figure S13), thereby resulting in bad predictions. Interestingly, some environmental factors are significantly associated with the prediction–observation mismatches of species richness (Supplementary Figure S14), implying the potential local-scale influence on the observed variation in

community diversity. Moreover, we acknowledge that the assumption for predicting species richness (i.e. species of the same lineage are functionally equivalent and have an equal chance to occur in a local community) is unrealistic, especially for Proteobacteria, which contains species with diverse traits and lifestyles [82, 83]. For Proteobacteria that have a high diversification rate (Supplementary Table S2), it may demand a remarkable variation in traits or niches among species [84, 85]. In fact, compared with other phyla, most proteobacterial species in the ECS pool have a relatively narrow niche breadth (Supplementary Figure S15), indicating the high specialization and turnover of proteobacterial species across the sampling sites. Thus, when investigating Proteobacteria-dominated communities, class-level (or finer units) relative abundances and species pools may be required for better predictions of species richness of a local community. Further research on the precise demarcation of evolutionarily and ecologically meaningful units (i.e. taxa with distinct diversification rates and habitat associations) would allow us to generalize our hypothesis regarding macro-evolutionary effects on local community diversity.

Finally, we explore potential reasons for the unequal intra-clade species richness in Proteobacteria vs Cyanobacteria (Supplementary Table S1). Clade age and diversification rate have been suggested as the two main determinants of variation in species richness across phylogenetic groups [23, 44, 86]. In bacterial systems, the variation in diversification rates seems to explain the most variation in species richness among phylum-level lineages, with Proteobacteria vs Cyanobacteria exhibiting high vs low diversification rates (Supplementary Table S2). Corresponding to the “flexibility hypothesis” that supposes high net speciation rates in the lineages with flexible character traits [87], we speculate that the different net speciation rates between Proteobacteria and Cyanobacteria are probably due to their distinct metabolic strategies to acquire nutrients and energy [88]. Despite noteworthy exceptions, Proteobacteria and Cyanobacteria, respectively, represent two primary functional groups, i.e. organic-matter decomposers and primary producers, in ecosystems [89]. Compared to proteobacterial species showing a high flexibility in gene combinations for assimilating environmental organic matter into biomass, cyanobacterial species are characterized by a complex and conserved set of genes to fix inorganic carbon dioxide through oxygenic photosynthesis [34]. Accordingly, the high vs low intra-clade species richness in Proteobacteria vs Cyanobacteria might reflect different levels of evolutionary flexibility for genome-wide divergences in these two lineages [88].

In conclusion, our findings suggest that the species diversity of a local bacterial community is strongly influenced by habitat associations and diversification rates of

deep phylogenetic lineages, highlighting the view of contemporary diversity patterns as an epiphenomenon resulting from long-term evolutionary diversification events [90]. With an understanding of the structure of the seawater species pool, the species richness of a local ECS bacterioplankton community is generally predictable. These results do not imply that local-scale ecological processes are unimportant, but rather that extending the study framework to account for evolutionary constraints is helpful for interpreting the observed diversity–environment patterns. Our conceptual model may lead to a more comprehensive understanding of the origin and variation of bacterial community diversity.

Acknowledgements We thank Hon-Tsen Yu for providing facilities and advice on laboratory work, and the Genome Research Center in National Yang-Ming University for sequencing service. Comments from David Armitage have greatly improved this work. This work was supported by the National Center for Theoretical Sciences, Foundation for the Advancement of Outstanding Scholarship, and the Ministry of Science and Technology, Taiwan.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Currie DJ, Mittelbach GG, Cornell HV, Field R, Guegan JF, Hawkins BA, et al. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol Lett.* 2004;7:1121–34.
- Gaston KJ. Global patterns in biodiversity. *Nature.* 2000;405:220–7.
- Huston M. A general hypothesis of species diversity. *Am Nat.* 1979;113:81–101.
- Ricklefs RE. Disintegration of the ecological community. *Am Nat.* 2008;172:741–50.
- MacArthur RH. Patterns of species diversity. *Biol Rev.* 1965;40:510–33.
- Currie DJ. Energy and large-scale patterns of animal-species and plant-species richness. *Am Nat.* 1991;137:27–49.
- Pianka ER. Latitudinal gradients in species diversity - a review of concepts. *Am Nat.* 1966;100:33–46.
- Qian H. Environment-richness relationships for mammals, birds, reptiles, and amphibians at global and regional scales. *Ecol Res.* 2010;25:629–37.
- Hawkins BA, Field R, Cornell HV, Currie DJ, Guegan JF, Kaufman DM, et al. Energy, water, and broad-scale geographic patterns of species richness. *Ecology.* 2003;84:3105–17.
- Menge BA, Sutherland JP. Species diversity gradients: synthesis of roles of predation, competition, and temporal heterogeneity. *Am Nat.* 1976;110:351–69.
- Palmer MW. Variation in species richness: towards a unification of hypotheses. *Folia Geobot Phytotaxon.* 1994;29:511–30.
- Shmida A, Wilson MV. Biological determinants of species diversity. *J Biogeogr.* 1985;12:1–20.

13. Eriksson O. The species-pool hypothesis and plant community diversity. *Oikos*. 1993;68:371–4.
14. Ricklefs RE. Community diversity - relative roles of local and regional processes. *Science*. 1987;235:167–171.
15. Tokeshi M. Species coexistence: ecological and evolutionary perspectives. Oxford: Blackwell Science; 1999.
16. Huston MA. Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals. *Oikos*. 1999;86:393–401.
17. Partel M, Zobel M, Zobel K, vanderMaarel E. The species pool and its relation to species richness: Evidence from Estonian plant communities. *Oikos*. 1996;75:111–7.
18. Partel M, Zobel M. Small-scale plant species richness in calcareous grasslands determined by the species pool, community age and shoot density. *Ecography*. 1999;22:153–9.
19. Ricklefs RE, He FL. Region effects influence local tree species diversity. *Proc Natl Acad Sci USA*. 2016;113:674–9.
20. Zobel M. The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends Ecol Evol*. 1997;12:266–9.
21. Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, et al. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol Lett*. 2007;10:315–31.
22. Ricklefs RE. Evolutionary diversification and the origin of the diversity–environment relationship. *Ecology*. 2006;87:S3–13.
23. Ricklefs RE. History and diversity: explorations at the intersection of ecology and evolution. *Am Nat*. 2007;170:S56–70.
24. Wiens JJ, Donoghue MJ. Historical biogeography, ecology and species richness. *Trends Ecol Evol*. 2004;19:639–44.
25. Donoghue MJ. A phylogenetic perspective on the distribution of plant diversity. *Proc Natl Acad Sci USA*. 2008;105:11549–55.
26. Buckley LB, Davies TJ, Ackerly DD, Kraft NJB, Harrison SP, Anacker BL, et al. Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. *Proc R Soc Lond B Biol Sci*. 2010;277:2131–8.
27. Partel M. Local plant diversity patterns and evolutionary history at the regional scale. *Ecology*. 2002;83:2361–6.
28. Cavalier-Smith T. Cell evolution and Earth history: stasis and revolution. *Philos T R Soc B*. 2006;361:969–1006.
29. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol*. 2012;10:497–506.
30. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol*. 2006;4:102–112.
31. Vass M, Langenheder S. The legacy of the past: effects of historical processes on microbial metacommunities. *Aquat Microb Ecol*. 2017;79:13–19.
32. Ricklefs RE. Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proc Natl Acad Sci USA*. 2010;107:1265–72.
33. Goberna M, Verdu M. Predicting microbial traits with phylogenies. *ISME J*. 2016;10:959–67.
34. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J*. 2013;7:830–8.
35. Morrissey EM, Mau RL, Schwartz E, Caporaso JG, Dijkstra P, van Gestel N, et al. Phylogenetic organization of bacterial activity. *ISME J*. 2016;10:2336–40.
36. Herlemann DPR, Lundin D, Andersson AF, Labrenz M, Jurgens K. Phylogenetic signals of salinity and season in bacterial community composition across the salinity gradient of the Baltic Sea. *Front Microbiol*. 2016;7:1883.
37. Lu HP, Yeh YC, Sastri AR, Shiah FK, Gong GC, Hsieh CH. Evaluating community-environment relationships along fine to broad taxonomic resolutions reveals evolutionary forces underlying community assembly. *ISME J*. 2016;10:2867–78.
38. Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB, et al. The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol*. 2010;8:523–9.
39. Fierer N, Bradford MA, Jackson RB. Toward an ecological classification of soil bacteria. *Ecology*. 2007;88:1354–64.
40. Lauber CL, Hamady M, Knight R, Fierer N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol*. 2009;75:5111–20.
41. Philippot L, Bru D, Saby NPA, Cuhel J, Arrouays D, Simek M, et al. Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environ Microbiol*. 2009;11:3096–104.
42. Magallon S, Sanderson MJ. Absolute diversification rates in angiosperm clades. *Evolution*. 2001;55:1762–80.
43. Marin J, Battistuzzi FU, Brown AC, Hedges SB. The timetree of prokaryotes: new insights into their evolution and speciation. *Mol Biol Evol*. 2017;34:437–46.
44. Scholl JP, Wiens JJ. Diversification rates and species richness across the Tree of Life. *Proc Biol Sci*. 2016;283:1334.
45. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590–6.
46. Gong GC, Wen YH, Wang BW, Liu GJ. Seasonal variation of chlorophyll a concentration, primary production and environmental conditions in the subtropical East China Sea. *Deep Sea Res Part 2 Top Stud Oceanogr*. 2003;50:1219–36.
47. Yeh YC, Peres-Neto PR, Huang SW, Lai YC, Tu CY, Shiah FK, et al. Determinism of bacterial metacommunity dynamics in the southern East China Sea varies depending on hydrography. *Ecography*. 2015;38:198–212.
48. Fuhrman JA, Comeau DE, Hagstrom A, Chan AM. Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Appl Environ Microbiol*. 1988;54:1426–9.
49. Gong GC, Chen YLL, Liu KK. Chemical hydrography and chlorophyll a distribution in the East China Sea in summer: Implications in nutrient dynamics. *Cont Shelf Res*. 1996;16:1561–90.
50. Gong GC, Shiah FK, Liu KK, Wen YH, Liang MH. Spatial and temporal variation of chlorophyll a, primary productivity and chemical hydrography in the southern East China Sea. *Cont Shelf Res*. 2000;20:411–36.
51. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA*. 2006;103:12115–20.
52. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
53. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
54. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*. 2014;64:346–51.
55. Chao A. Non-parametric estimation of the number of classes in a population. *Scand J Stat*. 1984;11:265–70.
56. Chao A, Lee S-M. Estimating the number of classes via sample coverage. *J Am Stat Assoc*. 1992;87:210–7.
57. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked

- and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 2007;35:7188–96.
58. Wright S. The method of path coefficients. *Ann Math Stat.* 1934;5:161–215.
 59. Sanchez G. *PLS Path Modeling with R.* Berkeley: Trowchez Editions; 2013.
 60. Fox J. Structural equation modeling with the sem package in R. *Struct Equ Modeling.* 2006;13:465–86.
 61. R Development Core Team. *R: A language and environment for statistical computing.* 3.3.2 edn. Vienna, Austria: The R Foundation for Statistical Computing Platform; 2016.
 62. Petraitis PS, Dunham AE, Niewiarowski PH. Inferring multiple causality: The limitations of path analysis. *Funct Ecol.* 1996;10:421–31.
 63. Borcard D, Legendre P, Drapeau P. Partialling out the spatial component of ecological variation. *Ecology.* 1992;73:1045–55.
 64. Peres-Neto PR, Legendre P, Dray S, Borcard D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology.* 2006;87:2614–25.
 65. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D et al. *Vegan: Community Ecology Package.* R package version 2.4-4. 2017.
 66. Dray S, Pelissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, et al. Community ecology in the age of multivariate multiscale spatial analysis. *Ecol Monogr.* 2012;82: 257–75.
 67. Soergel DA, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* 2012;6:1440–4.
 68. Amaral-Zettler L, Artigas LF, Baross J, Bharathi L, Boetius A, Chandramohan D et al. (2010). A global census of marine microbes. *Life in the world's oceans: diversity, distribution and abundance.* Oxford: Blackwell Publishing Ltd; McIntyre AD 2010. p. 223–45.
 69. Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren AM, et al. VAMPS: a website for visualization and analysis of microbial population structures. *Bmc Bioinformatics.* 2014;15:41.
 70. Cornell HV, Lawton JH. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *J Anim Ecol.* 1992;61:1–12.
 71. Dial KP, Marzluff JM. Nonrandom diversification within taxonomic assemblages. *Syst Zool.* 1989;38:26–37.
 72. Scotland RW, Sanderson MJ. The significance of few versus many in the tree of life. *Science.* 2004;303:643–3.
 73. Willig MR, Kaufman DM, Stevens RD. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu Rev Ecol Evol Syst.* 2003;34:273–309.
 74. Ladau J, Sharpton TJ, Finucane MM, Jospin G, Kembel SW, O'Dwyer J, et al. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* 2013;7:1669–77.
 75. Milici M, Deng ZL, Tomasch J, Decelle J, Wos-Oxley ML, Wang H, et al. Co-occurrence analysis of microbial taxa in the Atlantic Ocean reveals high connectivity in the free-living bacterioplankton. *Front Microbiol.* 2016a;7:649.
 76. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science.* 2015;348:1261359.
 77. Milici M, Tomasch J, Wos-Oxley ML, Wang H, Jauregui R, Camarinha-Silva A, et al. Low diversity of planktonic bacteria in the tropical ocean. *Sci Rep.* 2016b;6:19054.
 78. Hendershot JN, Read QD, Henning JA, Sanders NJ, Classen AT. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. *Ecology.* 2017;98:1757–63.
 79. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature.* 2017;551:457–63.
 80. Saleem M, Fetzer I, Dormann CF, Harms H, Chatzinotas A. Predator richness increases the effect of prey diversity on prey yield. *Nat Commun.* 2012;3:1305.
 81. Yang JW, Wu W, Chung CC, Chiang KP, Gong GC, Hsieh CH. Predator and prey biodiversity relationship and its consequences on marine ecosystem functioning-interplay between nano-flagellates and bacterioplankton. *ISME J.* 2018;12:1532–42.
 82. Ettema TJ, Andersson SG. The alpha-proteobacteria: the Darwin finches of the bacterial world. *Biol Lett.* 2009;5:429–32.
 83. Kloesges T, Popa O, Martin W, Dagan T. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol Biol Evol.* 2011;28:1057–74.
 84. Jezkova T, Wiens JJ. What explains patterns of diversification and richness among animal phyla? *Am Nat.* 2017;189:201–12.
 85. Castro-Insua A, Gomez-Rodriguez C, Wiens JJ, Baselga A. Climatic niche divergence drives patterns of diversification and richness among mammal families. *Sci Rep.* 2018;8:8781.
 86. McPeck MA, Brown JM. Clade age and not diversification rate explains species richness among animal taxa. *Am Nat.* 2007;169: E97–106.
 87. Ricklefs RE, Renner SS. Species richness within families of flowering plants. *Evolution.* 1994;48:1619–36.
 88. Cohan FM. Bacterial speciation: genetic sweeps in bacterial species. *Curr Biol.* 2016;26:R112–5.
 89. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science.* 2016;353:1272–7.
 90. Armitage DW. Experimental evidence for a time-integrated effect of productivity on diversity. *Ecol Lett.* 2015;18:1216–25.