environmental microbiology

Environmental Microbiology (2016) **18**(12), 5151–5160 doi:10.1111/1462-2920.13513

Microbial diversity within the *Trichodesmium* holobiont

Mónica Rouco,[1,2] Sheean T. Haley[1] and Sonya T. Dyhrman[1,2]*

[1]*Biology and Paleo Environment Division, Lamont-Doherty Earth Observatory Columbia University, NY, USA.*

[2]*Department of Earth and Environmental Sciences, Columbia University, NY, USA.*

Summary

Nitrogen-fixing cyanobacteria in the genus *Trichodesmium* play a critical role in the productivity of the tropical and subtropical oligotrophic oceans. The ecological success of these populations is likely associated with the diverse microbial interactions occurring within the *Trichodesmium* holobiont, especially between *Trichodesmium* and heterotrophic bacterial epibionts. Yet, the composition of the *Trichodesmium* holobiont and the processes governing microbial assemblage are not well documented. Here, we used high-resolution 16S rDNA amplicon sequencing to examine the diversity of *Trichodesmium* and associated epibionts across different ocean regions and colony morphologies (puffs and rafts). *Trichodesmium* Clade I (i.e., *T. thiebautii*-like) dominated the colonies in all ocean basins regardless of morphology, although the *Trichodesmium* community structure significantly varied between morphologies in some regions. On average, Alphaproteobacteria (i.e., *Thalassobius*), Gammaproteobacteria (i.e., *Pseudoalteromonas*), Sphingobacteria (i.e., *Microscilla* and *Vibrio*) and Flavobacteria dominated the epibiont communities, but community composition and structure significantly differed between regions. Epibionts from the two colony morphologies were taxonomically and functionally distinct in the North Atlantic and North Pacific. These findings suggest that the colony types might define two distinct niches and that epibiont assemblage might be driven in part by selective processes, where epibionts are selected according to their influence on colony metabolism.

Received 31 May, 2016; accepted 25 August, 2016. *For correspondence. Tel. +1 (845) 365-8165; Fax +1 (845) 365-8163; E-mail sdyhrman@ldeo.columbia.edu.

Introduction

Diazotrophs like *Trichodesmium* sp. play a keystone ecological role in the ecosystems where they occur by serving as a source of newly fixed nitrogen in oligotrophic systems (Letelier and Karl, 1996; Capone *et al.*, 1997; Goebel *et al.*, 2007; Luo *et al.*, 2012; Van Mooy *et al.*, 2015). *Trichodesmium* is found as filaments, which often form colonies that can have two main morphologies, termed puffs and rafts (Hynes *et al.*, 2012). The genus *Trichodesmium* comprises six different species and analyses of the 16S and the heterocyst differentiation genes have grouped them in four different clades, with Clade I and Clade III representing the majority of *Trichodesmium* field populations (Hynes *et al.*, 2012; Rouco *et al.*, 2014). Clade I includes *T. thiebautii*, *T. tenue*, *T. hildebrandtii* and *T. spiralis* and Clade III includes *T. erythraeum* and *T. contortum* (Hynes *et al.*, 2012). Many studies have focused on the identification and classification of *Trichodesmium* sp. in the laboratory and the field (Rippka *et al.*, 1979; Anagnostidis and Komárek, 1988; Zehr, 1995; Orcutt *et al.*, 2002; Lundgren *et al.*, 2005; Hewson *et al.*, 2009; Hmelo *et al.*, 2012; Hynes *et al.*, 2012) but have not focused on the distribution of the different *Trichodesmium* clades within the main colony morphotypes in the field.

Within the colony, *Trichodesmium* are only a member of a complex holobiont, a host-microbiome system, where *Trichodesmium* filaments host high concentrations of tightly attached heterotrophic bacterial epibionts (Herbst and Overbeck, 1978; Paerl *et al.*, 1989; Zehr, 1995; Dyhrman *et al.*, 2002; Hewson *et al.*, 2009). In addition, other cyanobacteria have been observed in association with *Trichodesmium* colonies (Hewson *et al.*, 2009; Hynes *et al.*, 2009; Hmelo *et al.*, 2012; Momper *et al.*, 2015). Studies have suggested that the metabolism of *Trichodesmium* and heterotrophic bacterial epibionts is tightly interconnected and that epibionts might play an important role in the cycling of iron and phosphorus within the colony (Achilles *et al.*, 2003; Roe *et al.*, 2012; Van Mooy *et al.*, 2012), with concomitant impacts on ocean biogeochemistry. The nature of these associations, and their impact on *Trichodesmium* physiological ecology and biogeochemical rate processes in the colony, is still not well understood. In addition, the microbial diversity and community structure of the epibionts across environmental gradients, or whether puffs and rafts represent distinct habitats or niches, is not known.

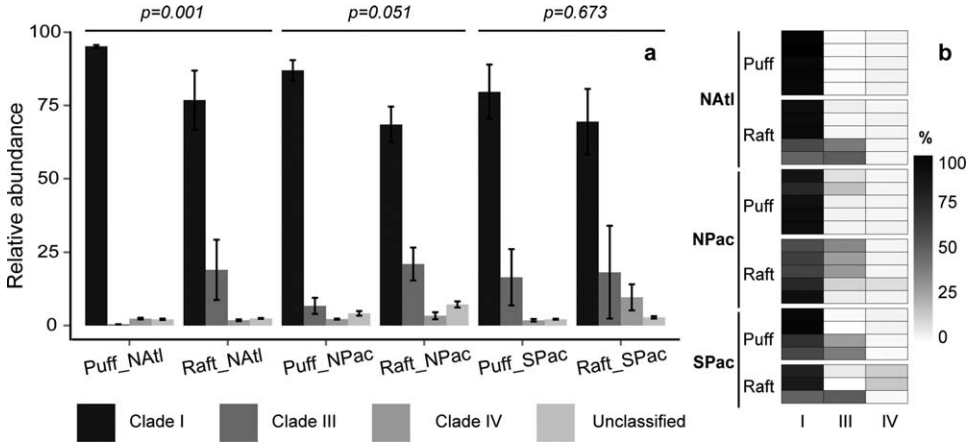

Fig. 1. Relative abundance of *Trichodesmium* clade I, III and IV rDNA 16S sequences. a. Averaged relative abundance as a function of ocean basin and colony morphology. Error bars represent the standard error of the mean. *p*-values resulting from within ocean Permanova analyses on Bray–Curtis dissimilarity matrices are shown on top of the graph. b. Heatmap of the inter-sample variability during each cruise. NAtl (North Atlantic), NPac (North Pacific), SPac (South Pacific).

In general, similar processes to those influencing β-diversity (variation in community composition across landscape) in macroscopic communities can also shape microbial community assemblages and microbial consortia composition (Hanson *et al.*, 2012; Nemergut *et al.*, 2013; Morris and Hmelo, 2014). If neutral processes govern, biodiversity would arise at random, with organisms at the same trophic level roughly equivalent with respect to fitness within an environment (Hubbell, 2001). In such cases, *Trichodesmium* colonization might result from random habitat discovery by members of the metacommunity, dependent on dispersal rate and habitat affinity (Morris and Hmelo, 2014). In addition, β-diversity could be driven by neutral processes of speciation and drift events due to dispersal limitation (Bell, 2010). Alternatively, if selective or deterministic processes dominate, as proposed by the Baas-Becking's hypothesis 'everything is everywhere and the environment selects' (De Wit and Bouvier, 2006), then *Trichodesmium* epibiont composition will be driven by the differential adaptation of microbes to local environmental conditions and the niche space provided by the colony, as defined by resource competition ability. It has been suggested that both neutral and selective processes likely operate in combination to create and maintain microbial diversity (Nemergut *et al.*, 2013; Morris and Hmelo, 2014), but this has not been examined in the context of the *Trichodesmium* holobiont.

In this study, rDNA 16S amplicons from *Trichodesmium* puff and *Trichodesmium* raft colonies were sequenced from samples collected in the North Atlantic, South Pacific and North Pacific oceans. The study was designed to evaluate the composition of the holobiont as a function of colony morphology and ocean basin and to investigate which evolutionary processes might drive epibiont community assembly within the *Trichodesmium* holobiont.

Results and discussion

A total of 27 *Trichodesmium* samples were collected from the North Atlantic, North Pacific and South Pacific oceans (Supporting Information Fig. S1), rinsed several times to avoid contamination and separated by colony morphology. The diversity of the *Trichodesmium* holobiont was examined to ascertain the extent to which these communities varied between colony morphology and ocean region using high-throughput 16S rDNA amplicon sequencing. This approach ensured robust coverage for all samples in this study, with Good's coverage estimator ranging from 0.94 to 0.98 (Supporting Information Table S1). Overall, the *Trichodesmium* holobiont was comprised primarily of *Trichodesmium* spp. (66.7%) and heterotrophic bacterial epibionts (31.8%), with a low contribution of other organisms such as Archaea or the chloroplasts of eukaryotic phytoplankton (1.5%).

Trichodesmium *diversity as a function of colony morphology and ocean region*

While there are several described species of *Trichodesmium*, falling into four major clades (Hynes *et al.*, 2012), field studies rarely distinguish beyond the genus level. Of the *Trichodesmium* sequences, >97% mapped to one of the three *Trichodesmium* sp. clades assessed (Supporting Information Table S2). Clade I dominated all samples (>80%) (Fig. 1), as has previously been observed in the North Atlantic using 16S rDNA clone libraries (Hmelo *et al.*, 2012), and clade specific qPCR (Rouco *et al.*, 2014). The dominance of Clade I in the other ocean regions suggests that care must be taken in extrapolating culture results from studies of the genome type strain *T. erythraeum* IMS101 to the field, as it is a representative of Clade III. Despite the dominance of Clade I across all ocean basins, the other Clades were routinely detected, particularly in samples of the raft morphology from the North Atlantic and North Pacific (Fig. 1). For example, the relative abundance of Clade III in North Atlantic puffs was always less than 1%, but Clade III represented ~38 and ~50% of the population in two of the North Atlantic raft samples. The clade diversity of the *Trichodesmium*

community, assessed using a two-way Permanova test on Bray–Curtis dissimilarity distance matrices, significantly differed between puffs and rafts in the North Atlantic ($F_{8,1} = 3.27$, $p = 0.001$) and the North Pacific ($F_{8,1} = 5.44$, $p = 0.05$), but not the South Pacific ($F_{5,1} = 0.22$, $p = 0.673$). These findings also suggest that individual *Trichodesmium* colonies might not be clonal and likely are comprised of multiple species from several clades. Although DNA sequencing was performed on samples of ~10–20 pooled colonies, all of the North Atlantic and North Pacific puff samples, for instance, had a similar Clade I and III percentage (Fig. 1b). This pattern would be difficult to replicate if colonies were clonal, since it would require repeated consistent sampling of ~9 Clade I and ~1 Clade III colonies in each sample. *Trichodesmium* species and colony morphologies can have different physiological capabilities (e.g., Hutchins *et al.*, 2007; 2013; Webb *et al.*, 2007; Dyhrman *et al.*, 2009; Chappell and Webb, 2010; Hynes *et al.*, 2012) and the differences in *Trichodesmium* colony clade diversity observed here could underpin physiological differences observed between colonies of these two morphologies in the field, such as those related to phosphorus (Orchard *et al.*, 2010), iron (Achilles *et al.*, 2003; Rubin *et al.* 2011) or CO_2 metabolism (Gradoville *et al.*, 2014).

Trichodesmium species in the laboratory tend to have different colony morphologies (Hynes *et al.*, 2012), and colonies of different morphologies are sometimes separated in the field as a potential proxy for segregating species. Although morphology does not appear to be a good proxy for species or even clade, the colonies could represent a distinct niche for epibiont colonization, either as a result of the physical difference in morphology or because of the distinct mix of *Trichodesmium* clades present, particularly, as shown here, in the North Atlantic and North Pacific.

Epibiont diversity of the Trichodesmium *holobiont*

Although *Trichodesmium* spp. dominated the colony 16S rDNA sequences, >30% of the total 16S rDNA sequences were from putative epibiont populations, predominately heterotrophic bacteria (31.8%). Only a very low number of sequences (12 010, or 1.3%) mapped to 16S sequences of chloroplasts from eukaryotic phytoplankton (Supporting Information Table S3). Of those, 11 720 sequences (1.25%) were from diatoms, which are sometimes observed in tight association with colonies (Hynes *et al.*, 2009). Although previous field studies found moderate numbers of cyanobacteria, such as *Phormidum*, *Plectonema* and *Lyngbya*, associated with *Trichodesmium* colonies (Siddiqui *et al.*, 1992; Hewson *et al.*, 2009; Hmelo *et al.*, 2012), here only 1074 sequences (0.11%) were from cyanobacterial groups other than *Trichodesmium*. The relative abundance of non-*Trichodesmium* cyanobacteria in *Trichodesmium* colonies might vary depending on the total concentration of these organisms in the surrounding water, and how colonies were processed prior to sequencing. These associations warrant further scrutiny, but were a small component of the total observations.

A recent study near Station ALOHA, in the North Pacific, reported high numbers of *Calothrix*-like heterocystous cyanobionts in 75% of the *Trichodesmium* puff colonies (Momper *et al.*, 2015). This association with a cyanobiont that fixes nitrogen with a temporal offset to *Trichodesmium* could influence colony derived nitrogen-fixation budgets depending on its prevalence and distribution (Momper *et al.*, 2015). It has been suggested that this association is a mutualistic relationship where the cyanobiont alleviates *Trichodesmium* nitrogen stress at night while taking advantage of *Trichodesmium* buoyancy to exploit new niches (Momper *et al.*, 2015). Overall, only 0.04% of the sequences in our study represented *Calothrix*-like members, found mostly in the North Pacific and North Atlantic oceans, and a significant differentiation between the raft and puff morphology was not observed (Supporting Information Table S3 – see Cyanobacteria, Family I). Although our data expand the observed distribution of this association to the North Atlantic, it appears that this association is not a consistent feature of puffs in general, and it may be that it is only seasonally present at high abundances, or it is perhaps linked to a particular *Trichodesmium* species that was not routinely present in our samples.

The Alphaproteobacteria (29.3%), Gammaproteobacteria (29.1%), Sphingobacteria (18.5%) and Flavobacteria (4.6%) dominated the epibiont communities across all samples (Supporting Information Table S3, Fig. S2), consistent with other lower resolution studies in both the North Atlantic (Hmelo *et al.*, 2012; Van Mooy *et al.*, 2012) and South Pacific oceans (Hewson *et al.*, 2009). Species within these phyla have previously been found in association with other phytoplankton groups such as diatoms (Amin *et al.*, 2012; Sison-Mangus *et al.*, 2014) and cyanobacteria (Stevenson and Waterbury, 2006; Tuomainen *et al.*, 2006). Notably, free-living bacterioplankton taxa dominant in the oligotrophic waters sampled here, such as *Prochlorococcus*, *Synechococcus*, SAR11, SAR86 or Archaea (Venter *et al.*, 2004; Carlson *et al.*, 2009), were either absent from colonies or found at negligible concentrations (Supporting Information Table S3). This suggests that unlike the oligotrophic organisms dominant in seawater, the epibiont communities within the *Trichodesmium* holobiont may be copiotrophic, benefiting from the comparatively nutrient-rich environment on the surface of *Trichodesmium* filaments. In fact, bacteria from the class Flavobacteria appear to utilize the dissolved organic carbon and the amino acids excreted by *Trichodesmium* (Herbst and Overbeck, 1978; Carpenter *et al.*, 1992; Capone *et al.*, 1994). Little is known about the nature of the *Trichodesmium*-epibiont association and its potential ecological

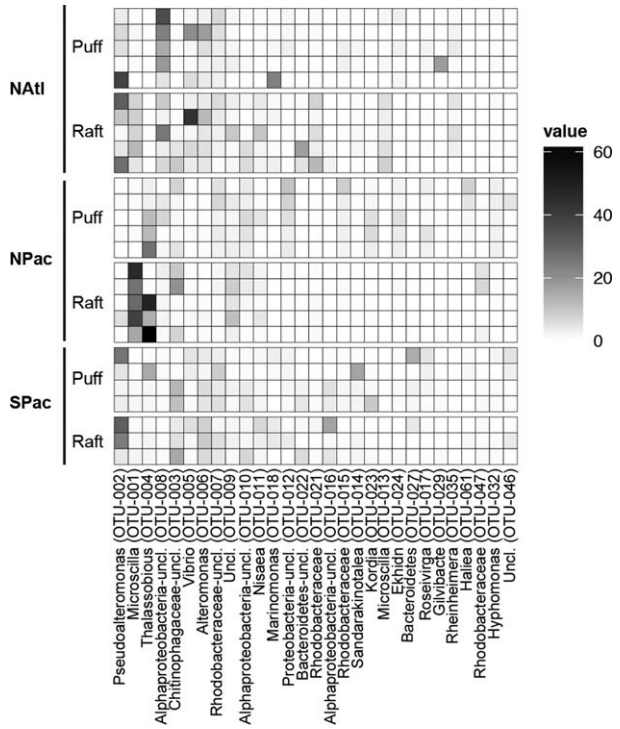

Fig. 2. Top 30 OTUs representing the epibiont community within the *Trichodesmium* holobiont.
On average, the top 10 OTUs represent ~50% of the entire epibiont community. OTU identity is given at the genus level or at the lowest taxonomic level it could be classified.

costs and benefits. Epibionts might lower the local O_2 concentration in the colony, which protects the oxygen-sensitive nitrogenase (Paerl and Pinckney, 1996).

Recent studies suggested that the *Trichodesmium*-epibiont interaction could also play an important role in the nutrient cycling within the colony. For instance, epibionts within the genus *Microscilla* sp. and *Vibrio* sp., which represented a high percentage of the epibiont populations in some of the North Pacific and North Atlantic samples in this study (~8% and ~5% for *Microscilla* sp. and *Vibrio* sp. respectively; Fig. 2 and Supporting Information Table S3), are capable of acquiring iron from a variety of inorganic and organically complex iron sources due to their ability to produce siderophores (Li and Chi, 2004; Roe *et al*., 2012). Species of *Vibrio* sp. present in *Trichodesmium* colonies also produce a class of quorum sensing molecule, acylated homoserine lactones (AHLs), that enhance the activity of the enzyme alkaline phosphate in the colony (Van Mooy *et al*., 2012). This enzyme is thought to play a critical role in supplying bioavailable phosphorus to the colony (Dyhrman *et al*., 2002; Orchard *et al*., 2009; 2010). Associations between epibionts and *Trichodesmium* might thus allow *Trichodesmium* to fulfill certain metabolic requirements that it would not meet otherwise. A recent study suggested that the low coding and large intergenic regions observed in the *Trichodesmium* genome might stem from a small *Trichodesmium* population size in a colony where heterotrophic epibionts are found at high concentrations and where gene functions might be shared (Walworth *et al*., 2015). This 'helper' heterotrophic bacteria model has been proposed to explain the association between the cyanobacteria *Prochlorococcus* sp. and its epibiont community (Morris *et al*., 2011). The observations here are consistent with a 'helper' heterotrophic bacteria model, but metagenome sequencing would further resolve the contribution of the epibionts to metabolic potential within the colony.

Neutral and selective drivers of the microbial assemblage

Despite the potential functional importance of *Trichodesmium*-epibiont associations, little is known about the taxonomic structure or the processes contributing to epibiont assemblage within the colony. The Invsimpson index, which considers richness and relative abundance, suggested an overall higher diversity in puffs versus rafts in the North Pacific and North Atlantic populations (23.9 v 3.8 and 11.3 v 9.4 for puffs and rafts in the North Pacific and North Atlantic respectively). In order to further investigate if epibiont diversity (community composition and relative abundance) differed between colony morphology and ocean basin, epibiont sequences were clustered into operational taxonomic units (OTUs) based on 97% sequencing identity. Cluster analyses identified an average of 641 $\pm$ 194 OTUs per sample (Supporting Information Table S1), approximately 30 times higher than the average richness observed by Hmelo *et al*. (2012), a study which did not use next-generation sequencing. Rank abundance curves followed the power law decay distribution typically observed in microbial communities (Bell, 2001), where ~10 OTUs represented more than 50% of the community whereas low abundant OTUs, below 0.1%, represented most of the diversity in the sample (Fig. 2, Supporting Information Fig. S3).

Neither ocean basin ($F_{21,2}$= 2.4, $p = 0.12$) nor colony morphology ($F_{21.1} = 0.03$, $p = 0.87$) had a significant effect on community richness, explored by the number of OTUs per sample. However, epibiont communities were not homogeneous across all samples and both region and colony morphology significantly affected community diversity and composition (Supporting Information Table S4). Analyses within each region revealed that puffs and rafts in the North Pacific and North Atlantic contained significantly mutually exclusive epibiont communities (Fig. 3). Morris and Hmelo (2014) suggested that distinct community patterns across colony morphologies could be explained as a result of lottery competition theory (Sale, 1977; Morris and Hmelo, 2014), where each colony is randomly filled by a

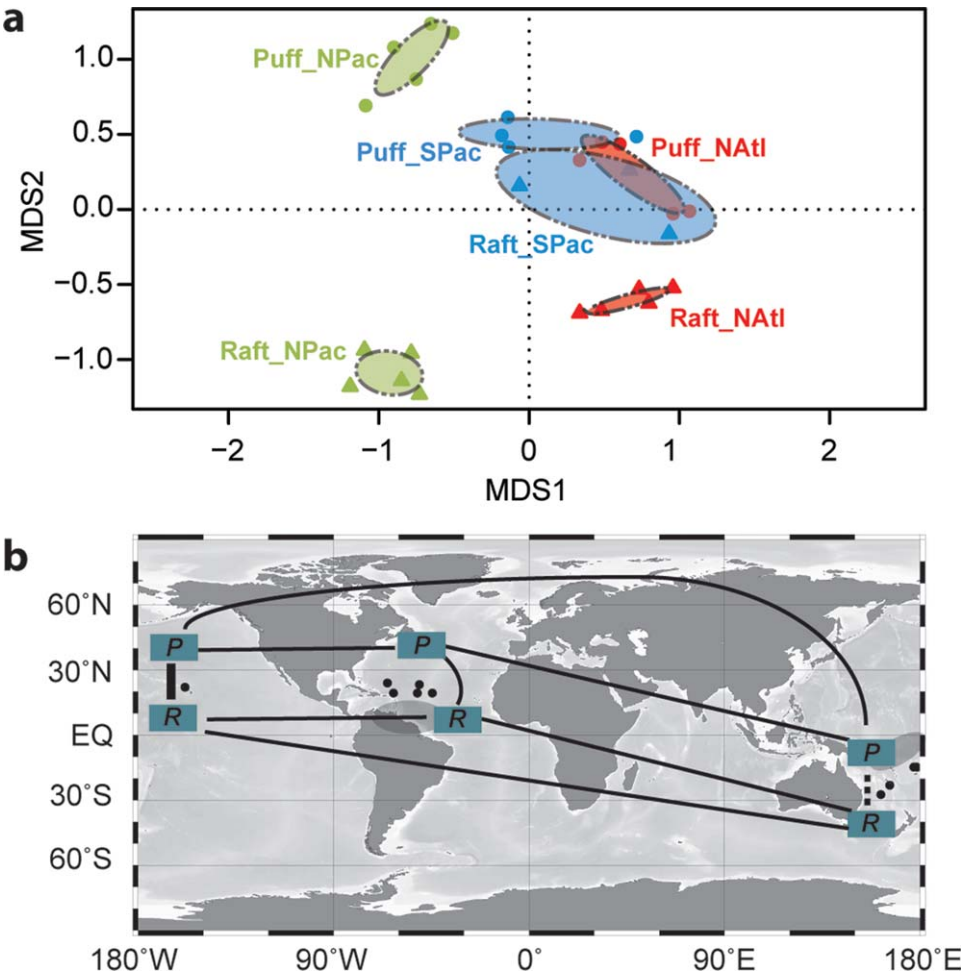

Fig. 3. Epibiont community diversity across ocean basins and colony morphologies.
a. First two dimensions of classic multidimensional scaling of Bray–Curtis dissimilarity distances. Ellipses represent 95% confidence intervals. Each dot represents one sample, color and shape coded by ocean basin and colony morphology. NAtl (North Atlantic), NPac (North Pacific), SPac (South Pacific).
b. Results from pairwise Permanova testing (9999 permutations) of Bray–Curtis dissimilarities of puffs and rafts within each ocean basin and of puffs or rafts across ocean basins, after multiple test correction. Solid line indicates significant differences between pairs ($p < 0.05$) and dashed line indicates absence of significant differences. Black dots represent sampling locations. P (puffs), R (rafts).

group of ecologically and functionally redundant epibionts that compete for settlement on the surface (Simberloff and Dayan, 1991). However, in this study puffs and rafts at each station were sampled from the same water at the same time. If communities were merely defined as a result of the lottery hypothesis, it is unlikely that a distinct raft and puff community would be consistently observed in colonies in the same water mass. Rather, the consistent differences in epibiont diversity between these two morphologies may be driven by selective forces, given that puffs and rafts could represent distinct niches either as a result of the physical difference in morphology or because of the distinct mix of *Trichodesmium* clades present. In fact, statistical differences in epibiont diversity between puffs and rafts were only observed in the North Atlantic and North Pacific, where the *Trichodesmium* diversity between colony morphology was also significantly different (Fig. 1).

An analysis of the PICRUSt predicted metabolic capacities (Langille *et al.*, 2013) encoded by the epibiont communities (a total of 5019 KEGG orthologs: 4927, 4734 and 4650 for the North Atlantic, North Pacific and South Pacific respectively), revealed significant differences between the raft and the puff communities (Fig. 4). These differences are magnified in the North Atlantic (90 KEGG orthologs) and North Pacific oceans (241 KEGG orthologs), relative to the South Pacific (only 4 KEGG orthologs) (Fig. 4, Supporting Information Table S5), consistent with their relative divergence in epibiont and *Trichodesmium* sp. taxonomic diversity. While specific KEGG prediction by PICRUSt have to be interpreted with caution, given the potential biases associated with the availability of reference genomes and the selection of the hypervariable regions of 16S rDNA for taxonomic assignment (Langille *et al.*, 2013; Smith *et al.*, 2014), PICRUSt predictions suggested a number of potential pathways associated with nutrient and trace metal metabolism uniquely enriched in puffs or rafts in the North Pacific and North Atlantic (Fig. 4, Supporting Information Table S5, Fig. S4). For instance, a putative phosphate regulon sensor (K07636), and putative iron-related transporters (TonB-K03832, K02016), were only significantly enriched in the raft communities (Supporting Information Table S5). KEGG pathways involved in lipid transport (Supporting Information Fig. S4) were only enriched in the puff community (Supporting Information Table S5). Although metagenomic

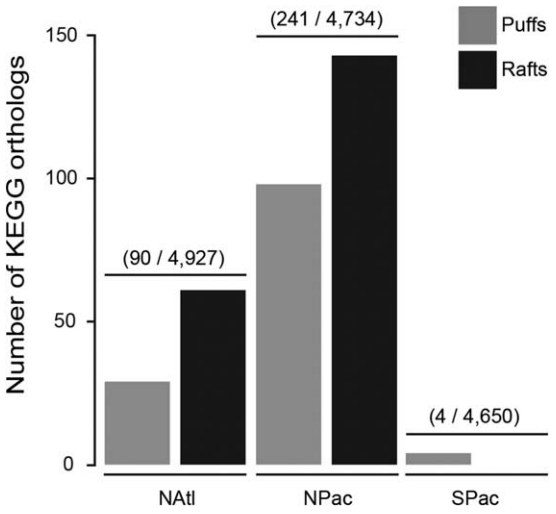

Fig. 4. Number of KEGG pathways predicted to be differentially represented (LDA score > 2) in the puff or raft community within each ocean.
Metagenomes were inferred using PICRUSt, which predicted a total of 5019 KEGG pathways.

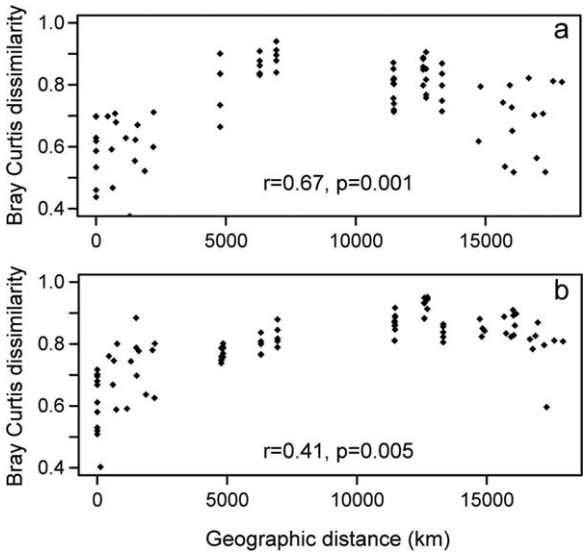

Fig. 5. Distance-decay curves for the epibiont communities obtained from mantel test analyses.
a. Raft morphology.
b. Puff morphology.

analyses are necessary to fully resolve the metabolic potential conferred by these epibiont communities, these results suggest that the two morphologies might be colonized by epibionts which confer distinct metabolic function. These results also emphasize the potential role of selective processes in defining the epibiont community composition, suggesting that puffs and rafts in the North Pacific and North Atlantic might not be colonized by ecologically and functionally redundant epibionts, but by those epibionts best suited to exploit that particular colony morphology.

Differences in epibiont diversity between regions were also examined and analyzed for each colony morphology independently. Permanova analyses revealed that epibiont diversity varied across ocean basins (Fig. 3). In addition, results from the mantel tests showed a significant distance-decay pattern (Fig. 5), indicating that the dissimilarity in epibiont communities in both the puff and raft morphology increased with increasing geographic distance. Distance-decay relationships have been observed for communities of microorganisms in a range of habitats and at various taxonomic resolutions (Green *et al*., 2004; Horner-Devine *et al*., 2004; Hewson *et al*., 2006; Casteleyn *et al*., 2010; Hanson *et al*., 2012), and can originate as a result of neutral and/or selective processes (Hanson *et al*., 2012; Nemergut *et al*., 2013). If ecological equivalence could be assumed for each sample across regions (i.e., the samples were collected from sites of identical environmental conditions and comprised an identical *Trichodesmium* sp. host composition), a distance-decay relationship could be explained solely by the neutral forces of differential speciation, community drift or dispersal limitation. However, different ocean regions represent distinct physicochemical environments (Sohm *et al*., 2011), and disparate physiological strategies have already been observed between *Trichodesmium* sp. in the North Pacific and North Atlantic oceans (Hynes *et al*., 2009; Chappell *et al*., 2012). This suggests that the distance-decay pattern can also be driven by selective processes, following the Baas-Becking hypothesis 'everything is everywhere and the environment selects' (De Wit and Bouvier, 2006), where the epibiont community will represent the best-fit community. The influence of neutral and selective processes cannot be separated to explain differences in epibiont diversity across oceans, and it is likely that these two processes occur in combination. Taken together, these data suggest that associations between *Trichodesmium* and the epibionts are not random, but rather the epibionts might be selected according to their metabolic potential to best exploit regional biogeochemistry or local micro-environment within the colony, including *Trichodesmium* spp. composition.

Conclusion

This study characterized the *Trichodesmium* holobiont with unprecedented resolution, and significant differences were observed in the epibiont communities between both ocean basins and colony morphologies. These data suggest that the heterotrophic epibionts are essential members of the *Trichodesmium* holobiont, and that while neutral processes might contribute to the structure of these communities, selective processes might play an important role in structuring the epibionts that best suits the *Trichodesmium* community and best exploits the local environmental

conditions. The distinct *Trichodesmium* holobiont populations associated with colony morphology and ocean region likely contribute to the known heterogeneity in the activities and physiological ecology of *Trichodesmium*. This may explain the difficulties associated with the prediction of *Trichodesmium* bloom formation and nitrogen fixation patterns in the ocean, which do not typically consider the metabolic potential conferred by the epibiont community. An increased effort to identify these epibiont communities and their metabolic potential will help to further unravel the ecological relevance of these associations as well as their present and future contributions to global nutrient cycles.

Experimental procedures

Field sample collection

A total of 27 samples were collected during a cruise transect in the South Pacific (SPAC, March 2007), North Atlantic (X804, May 2008) and as part of a time series sampling in the North Pacific at station ALOHA (PHORII, September 2013) (Supporting Information Fig. S1). *Trichodesmium* colonies were collected from the near surface (approximately within the top 25 m) using a handheld 130 μm net. Single colonies were picked and transferred into 0.2 μm filtered local surface water collected at 5 m with a Rosette sampling device. To avoid potential contamination, and to assure that only organisms tightly associated with *Trichodesmium* were sequenced, colonies were rinsed in fresh 0.2 μm filtered local surface water three more times. Then, colonies were separated into two morphologies, 'rafts' (colonies with a parallel organization of trichomes), and 'puffs' (colonies with a radial organization of the trichomes). Between 10 and 20 washed *Trichodesmium* puffs or rafts were filtered onto 47 mm, 10 μm polycarbonate filters, which were then placed in 2 ml cryovials, snap-frozen and stored in liquid nitrogen until DNA extraction was performed.

DNA extraction

DNA was extracted from each filter using the Power Plant Pro DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA), which has been shown to be the most robust DNA extraction kit for *Trichodesmium* samples (Hynes, 2009). Sample filters were transferred to bead tubes provided by the kit and extracted following the manufacturer's instructions. DNA was eluted in 50–100 μl of elution solution and the DNA samples were kept at −20°C until DNA sequencing was performed.

rDNA 16S library construction

DNA from each sample was sent to Argonne National Laboratory (Lemont, IL) for paired-end sequencing (2 × 150 bp) of partial 16S rDNA genes using the Illumina Miseq platform. Genomic DNA was amplified using the Earth Microbiome Project barcoded primer set, adapted for the MiSeq platform by adding nine extra bases in the adapter region of the forward amplification primer that support paired-end sequencing. As suggested by Kozich *et al*. (2013) and Mizrahi-Man *et al*. (2013) for short-read sequencing strategies, the V4 region of the 16S rDNA gene (515F-806R) was amplified, using region-specific primers that included the Illumina flowcell adapter sequences. The reverse amplification primer also contained a twelve base barcode sequence for the later distinction of individual sample sequences (Caporaso *et al*., 2011; 2012). Each 25 μl PCR reaction contained 12 μl of MoBio PCR Water (Certified DNA-Free), 10 μl of 5 Prime HotMasterMix (1x), 1 μl of Forward Primer (5 μM concentration, 200 pM final), 1 μl Golay Barcode Tagged Reverse Primer (5 μM concentration, 200 pM final) and 1μl of template DNA. The conditions for PCR were as follows: 94°C for 3 min to denature the DNA, with 35 cycles at 94°C for 45 s, 50°C for 60 s and 72°C for 90 s; with a final extension of 10 min at 72°C to ensure complete amplification. The PCR amplicons were quantified using PicoGreen (Invitrogen, Carlsbad, CA). Once quantified, different volumes of each of the products were pooled into a single tube so that each amplicon was represented equally. This pool was then cleaned up using UltraClean® PCR Clean-Up Kit (MoBIO), and then quantified using a Qubit (Invitrogen). After quantification, the molarity of the pool was determined and diluted down to 2 nM, denatured and then diluted to a final concentration of 6.75 pM with a 10% PhiX spike for sequencing on the Illumina MiSeq. Sequencing was performed as described previously (Caporaso *et al*., 2012).

rDNA 16S sequence analyses

rDNA 16S sequences were processed using modules implemented in the Mothur v.1.34.0 software following Kozich *et al*. (2013). Briefly, any sequences with ambiguous bases or homopolymers longer than eight bases were removed from the data set and sequences were trimmed to a uniform length of 253 bp. Sequences were aligned using the SILVA-compatible alignment database available within Mothur and chimeric sequences were removed using Uchime (Edgar *et al*., 2011). Unique sequences were then classified by a Bayesian approach using the Mothur-formatted version of the RDP training set (v.9) with an 80% cut-off. Sequences unique to *Trichodesmium* sp. (624 599 sequences, 67.7%) were extracted and assigned to three different *Trichodesmium* clades (Clade I, III and IV) by mapping to rDNA 16S sequences obtained from Hynes *et al*. (2012), after Janson *et al*. (1999) and Lundgren *et al*. (2005) (Supporting Information Table S2), using the classify.seqs command implemented in Mothur. Herein, Clade I is represented by sequences mapping to *T. thiebautii* and *T. tenue Z-1*, Clade III is represented by sequences mapping to *T. havanum* F34-5, *T. erythraeum* K-02#2 and *T. erythraeum* 21-75 and Clade IV is represented by sequences mapping to *T. contortum* and *T. tenue* (accession numbers AF013028 and AF013029 respectively). A bacterial-epibiont-only dataset (298 174 sequences, 31.8%) was created by removing any sequences classified as 'chloroplast' (which included *Trichodesmium* sp. sequences and other photosynthesizers), 'mitochondrial', 'archaeal' or 'unknown'. Bacterial-epibiont-only sequences were clustered into OTUs based on 97% sequencing identity. Depth coverage was assessed using Good's coverage estimator calculated in Mothur. Venn diagrams were created within Mothur to visualize the number of OTUs shared between ocean basins or colony morphology. To analyze the effects of morphology and ocean basin on both epibiont community composition (species

richness) and community structure (species relative abundances) a two-way Permanova was performed using a Bray–Curtis dissimilary distance matrix (Anderson, 2001). Permanova tests were also performed separately with both morphology and ocean basin as fixed effects in the *Trichodesmium* and the epibiont dataset and a Benjamini–Hochberg correction for multiple testing was applied, limiting the overall false discovery rate to 5% (Benjamini and Hochberg, 1995). Variance in epibiont community structure described by Bray–Curtis dissimilarities was visualized using principal coordinate analysis in R. Mantel tests testing for correlations between community dissimilarity and geographic distance were conducted for each colony morphology. All statistical tests were performed in R.

Metabolic predictions of the epibiont community were carried out on OTUs at the 97% similarity level, using PICRUSt (Langille *et al*., 2013) through the Galaxy server (Goecks *et al*., 2010). OTUs were assigned with Mothur against the Greengenes database (v5). The mean Nearest Sequenced Taxon Index, calculated by PICRUSt to provide a reliability estimation of the metagenome predictions, was in a normal range for all samples (0.099 ± 0.04), according to Langille *et al*. (2013). The Linear Discriminant Analysis (LDA) Effect Size (LEfSe) algorithm (Segata *et al*., 2011) was used through the Galaxy server to identify significant differentially abundant microbial relevant pathways enriched in the puff or raft morphology within each ocean basin. KEGG pathways were considered differentially represented if their LDA score was higher than 2 (Segata *et al*., 2011). Raw sequences for each sample have been submitted to NCBI Sequence Read Archive, BIOPROJECT PRJNA314461 and SRA accession ID SRP072053.

Acknowledgements

This research was supported by the National Science Foundation, as part of the Center for Microbial Oceanography Research and Education (C-MORE) (NSF/EF04-24599) EDventures program, the Chemical Oceanography Program and the Biological Oceanography Program. This research was also in part funded by a grant from the Simons Foundation (SCOPE Award ID 329108, S.T.D.). This is a contribution of the Simons Collaboration on Ocean Processes and Ecology (SCOPE). We thank the captain, crew, science party and leadership personnel of all the cruises, which supported the collection of the samples, in particular Elizabeth Orchard. A special thanks is also given to Bethan Jones, Lydia Baker, Stephen Harris, Katherine Brooks and Jeff Bowman, for helpful discussions. Authors declare no conflict of interest.

References

Achilles, K.M., Church, T.M., Wilhelm, S.W., Luther, G.W.I., and Hutchins, D.A. (2003) Bioavailability of iron to *Trichodesmium* colonies in the western subtropical Atlantic Ocean. *Limnol Oceanogr* **48:** 2250–2255.

Amin, S.A., Parker, M.S., and Armbrust, E. (2012) Interactions between diatoms and bacteria. *Microbiol Mol Biol Rev* **76:** 667–684.

Anagnostidis, K., and Komárek, J. (1988) Modern approach to the classification system of cyanophytes. 3 - Oscillatoriales. *Arch Fur Hydrobiol Suppl* **80:** 327–472.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26:** 32–46.

Bell, G. (2001) Neutral Macroecology. *Science* **293:** 2413–2419.

Bell, T. (2010) Experimental tests of the bacterial distance-decay relationship. *ISME J* **4:** 1357–1365.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57:** 289–300.

Capone, D., Zehr, J., Paerl, H., Bergman, B., and Carpenter, E.J. (1997) *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276:** 1221–1229.

Capone, D.G., Ferrier, M.D., and Carpenter, E.J. (1994) Amino-acid cycling in colonies of the planktonic marine cyanobacterium *Trichodesmium-thiebautii*. *Appl Environ Microbiol* **60:** 3989–3995.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., *et al*. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* **108:** 4516–4522.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., *et al*. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6:** 1621–1624.

Carlson, C.A., Morris, R., Parsons, R., Treusch, A.H., Giovannoni, S.J., and Vergin, K. (2009) Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3:** 283–295.

Carpenter, E.J., Bergman, B., Dawson, R., Siddiqui, P.J.A., Soderback, E., and Capone, D.G. (1992) Glutamine synthetase and nitrogen cycling in colonies of the marine diazotrophic cyanobacteria *Trichodesmium* spp. *Appl Environ Microbiol* **58:** 3122–3129.

Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A., Kotaki, Y., Rhodes, L., *et al*. (2010) Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc Natl Acad Sci USA* **107:** 12952–12957.

Chappell, P.D., and Webb, E.A. (2010) A molecular assessment of the iron stress response in the two phylogenetic clades of *Trichodesmium*. *Environ Microbiol* **12:** 13–27.

Chappell, P.D., Moffett, J.W., Hynes, A.M., and Webb, E.A. (2012) Molecular evidence of iron limitation and availability in the global diazotroph *Trichodesmium*. *ISME J* **6:** 1728–1739.

De Wit, R., and Bouvier, T. (2006) "Everything is everywhere, but, the environment selects"; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8:** 755–758.

Dyhrman, S.T., Webb, E.A., Anderson, D.M., Moffett, J.W., and Waterbury, J.B. (2002) Cell-specific detection of phosphorus stress in *Trichodesmium* from the Western North Atlantic. *Limnol Oceanogr* **47:** 1832–1836.

Dyhrman, S.T., Benitez-Nelson, C.R., Orchard, E.D., Haley, S.T., and Pellechia, P.J. (2009) A microbial source of phosphonates in oligotrophic marine systems. *Nat Geosci* **2:** 696–699.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194–2200.

Goebel, N.L., Edwards, C.A., Church, M.J., and Zehr, J.P. (2007) Modeled contributions of three types of diazotrophs to nitrogen fixation at Station ALOHA. *ISME J* **1:** 606–619.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11:** R86.

Gradoville, M.R., White, A.E., Bottjer, D., Church, M.J., and Letelier, R.M. (2014) Diversity trumps acidification: lack of evidence for carbon dioxide enhancement of *Trichodesmium* community nitrogen or carbon fixation at station ALOHA. *Limnol Oceanogr* **59:** 645–659.

Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., David, B., Dangerfield, M., *et al.* (2004) Spatial scaling of microbial eukaryote diversity. *Nature* **432:** 747–750.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497–506.

Herbst, V., and Overbeck, J. (1978) Metabolic coupling between the alga *Oscillatoria redekei* and accompanying bacteria. *Naturwissenschaften* **65:** 598–599.

Hewson, I., Steele, J.A., Capone, D.G., and Fuhrman, J.A. (2006) Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Mar Ecol Prog Ser* **311:** 67–77.

Hewson, I., Poretsky, R.S., Dyhrman, S.T., Zielinski, B., White, A.E., Tripp, H.J., *et al.* (2009) Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *ISME J* **3:** 1286–1300.

Hmelo, L.R., Van Mooy, B., and Mincer, T. (2012) Characterization of bacterial epibionts on the cyanobacterium *Trichodesmium*. *Aquat Microb Ecol* **67:** 1–14.

Horner-Devine, M.C., Lage, M., Hughes, J.B., and Bohannan, B.J.M. (2004) A taxa-area relationship for bacteria. *Nature* **432:** 750–753.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press, p. 375.

Hutchins, D.A., Fu, F.X., Zhang, Y., Warner, M.E., Feng, Y., Portune, K., *et al.* (2007) CO_2 control of *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios: implications for past, present, and future ocean biogeochemistry. *Limnol Oceanogr* **52:** 1293–1304.

Hutchins, D.A., Fei-Xue, F., Webb, E.A., Walworth, N., and Tagliabue, A. (2013) Taxon-specific response of marine nitrogen fixers to elevated carbon dioxide concentrations. *Nat Geosci* **6:** 790–795.

Hynes, A. (2009) Diversity of the marine cyanobacterium *Trichodesmium*: characterization of the Woods Hole culture collection and quantification of field populations. PhD Thesis. Massachusetts Institute of Technology and Woods Hole Oceanographic Institution.

Hynes, A.M., Chappell, P.D., Dyhrman, S.T., Doney, S.C., and Webb, E.A. (2009) Cross-basin comparison of phosphorus stress and nitrogen fixation in *Trichodesmium*. *Limnol Oceanogr* **54:** 1438–1448.

Hynes, A.M., Webb, E.A., Doney, S.C., and Waterbury, J.B. (2012) Comparison of cultured *Trichodesmium* (Cyanophyceae) with species characterized from the field. *J Phycol* **48:** 196–210.

Janson, S., Bergman, B., Carpenter, E.J., Giovannoni, S.J., and Vergin, K. (1999) Genetic analysis of natural populations of the marine diazotrophic cyanobacterium *Trichodesmium*. *FEMS Microbiol Ecol* **30:** 57–65.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol* **79:** 5112–5120.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31:** 814–821.

Letelier, R.M., and Karl, D.M. (1996) Role of *Trichodesmium* spp. in the productivity of the subtropical North Pacific Ocean. *Mar Ecol Prog Ser* **133:** 263–273.

Li, J., and Chi, Z. (2004) Siderophores from marine microorganisms and their applications. *J Ocean Univ China* **3:** 40–47.

Lundgren, P., Janson, S., Jonasson, S., Singer, A., and Bergman, B. (2005) Unveiling of novel radiations within *Trichodesmium* cluster by hetR gene sequence analysis. *Appl Environ Microbiol* **71:** 190–196.

Luo, Y.W., Doney, S.C., Anderson, L.A., Benavides, M., Bode, A., Bonnet, S., *et al.* (2012) Database of diazotrophs in global ocean: abundances, biomass and nitrogen fixation rates. *Earth Syst Sci Data Discuss* **5:** 47–106.

Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. (2013) Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* **8:** e53608.

Momper, L.M., Reese, B.K., Carvalho, G., Lee, P., and Webb, E.A. (2015) A novel cohabitation between two diazotrophic cyanobacteria in the oligotrophic ocean. *ISME J* **9:** 882–893.

Morris, J.J., and Hmelo, L.R. (2014) Organizing principles for marine microbial consortia. PF Kemp [ed.]. Eco-DAS IX Symposium Proceedings. Waco, TX: Association for the Science of Limnology and Oceanography.

Morris, J.J., Johnson, Z.I., Szul, M.J., Keller, M., and 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.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'neill, S.P., Bilinski, T.M., Stanish, L.F., *et al.* (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* **77:** 342–356.

Orchard, E.D., Webb, E.A., and Dyhrman, S.T. (2009) Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. *Environ Microbiol* **11:** 2400–2411.

Orchard, E.D., Ammerman, J.W., Lomas, M.W., and Dyhrman, S.T. (2010) Dissolved inorganic and organic phosphorus uptake in *Trichodesmium* and the microbial community: the importance of phosphorus ester in the Sargasso Sea. *Limnol Oceanogr* **55:** 1390–1399.

Orcutt, K.M., Rasmussen, U., Webb, E.A., Waterbury, J.B., Gundersen, K., and Bergman, B. (2002) Characterization of *Trichodesmium* spp. by genetic techniques. *Appl Environ Microbiol* **68:** 2236–2245.

Paerl, H.W., and Pinckney, J.L. (1996) A mini-review of microbial consortia: their roles in aquatic production and biogeochemical cycling. *Microb Ecol* **31:** 225–247.

Paerl, H.W., Bebout, B.M., and Prufert, L.E. (1989) Bacterial associations with marine *Oscilatoria sp.* (*Trichodesmium* sp.) populations: ecophysiological implications. *J Phycol* **25:** 773–784.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111:** 1–61.

Roe, K.L., Barbeau, K., Mann, E.L., and Haygood, M.G. (2012) Acquisition of iron by *Trichodesmium* and associated bacteria in culture. *Environ Microbiol* **14:** 1681–1695.

Rouco, M., Joy-Warren, H., McGillicuddy, D.J., Waterbury, J.B., and Dyhrman, S.T. (2014) *Trichodesmium* sp. clade distributions in the western North Atlantic Ocean. *Limnol Oceanogr* **59:** 1899–1909.

Rubin, M., Berman-Frank, I., and Shaked, Y. (2011) Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. *Nat Geosci* **4:** 529–534.

Sale, P.F. (1977) Maintenance of high diversity in coral reef fish communities. *Am Nat* **111:** 337–359.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12:** R60.

Siddiqui, P.J.A., Bergman, B., and Carpenter, E.J. (1992) Filamentous cyanobacterial associates of the marine planktonic cyanobacterium *Trichodesmium*. *Phycologia* **31:** 326–337.

Simberloff, D., and Dayan, T. (1991) The guild concept and the structure of ecological communities. *Annu Rev Ecol Syst* **22:** 115–143.

Sison-Mangus, M.P., Jiang, S., Tran, K.N., and Kudela, R.M. (2014) Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J* **8:** 63–76.

Smith, M.I., Turpin, W., Tyler, A.D., Silverberg, M.S., and Croitoru, K. (2014) Microbiome analysis - from technical advances to biological relevance. *F1000Prime Rep* **6:** 51.

Sohm, J.A., Webb, E.A., and Capone, D.G. (2011) Emerging patterns of marine nitrogen fixation. *Nat Rev Microbiol* **9:** 499–508.

Stevenson, B.S., and Waterbury, J.B. (2006) Isolation and identification of an epibiotic bacterium associated with heterocystous *Anabaena* cells. *Biol Bull* **210:** 73–77.

Tuomainen, J., Hietanen, S., Kuparinen, J., Martikainen, P., and Servomaa, K. (2006) Community Structure of the Bacteria Associated with *Nodularia* sp. (Cyanobacteria) Aggregates in the Baltic Sea. *Microb Ecol* **52:** 513–522.

Van Mooy, B., Hmelo, L.R., Sofen, L.E., Campagna, S.R., May, A.L., Dyhrman, S.T., *et al.* (2012) Quorum sensing control of phosphorus acquisition in *Trichodesmium* consortia. *ISME J* **6:** 422–429.

Van Mooy, B., Krupke, A., Dyhrman, S.T., Fredricks, H.F., Frischkorn, K.R., Ossolinski, J.E., *et al.* (2015) Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science* **348:** 783–785.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Walworth, N., Pfreundt, U., Nelson, W.C., Mincer, T., Heidelberg, J.F., Fu, F., *et al.* (2015) *Trichodesmium* genome maintains abundant, widespread noncoding DNA in situ, despite oligotrophic lifestyle. *Proc Natl Acad Sci* **112:** 4251–4256.

Webb, E., Jakuba, R., Moffett, J., and Dyhrman, S. (2007) Molecular assessment of phosphorus and iron physiology in *Trichodesmium* populations from the western Central and western South Atlantic. *Limnol Oceanogr* **52:** 2221–2232.

Zehr, J.P. (1995) Nitrogen fixation in the sea: why only *Trichodesmium*? In *Molecular Ecology of Aquatic Microbes*. Joint,I.R. (ed). Berlin: Springer, pp. 333–364.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Sampling locations and number of samples of each morphology collected in each ocean basin.

Fig. S2. Average phylogenetic compositon at the class level for puffs and rafts in each ocean basin.

Fig. S3. Rank abundance curves for individual samples. All samples follow the power law decay distribution,where low abundant OTUs, below 0.1%, represented most of the diversity in the sample, whereas $\sim$10 OTUs represent more than 50% of the community.

Fig. S4. Proportion of enriched KEGG pathways (LDA score $>$2) falling into different KEGG modules in the puff or raft community within each ocean. Metagenomes were inferred using PICRUSt, which predicted a total of 5019 KEGG pathways. The 'Other' category corresponds to proteins not falling into any of the specific KEGG modules reported here. The 'Uncharacterized' category refers to those KEGG pathways without an annotation.

Table S1. Number of sequences, OTUs and Good's coverage estimator per sample. The sample ID is preceded by the ocean basin (NAtl, NPac and SPac, for the North Atlantic, North Pacific and South Pacific samples respectively), followed by the station number and colony morphology (P and R for the puff and raft morphology respectively).

Table S2. GenBank accession numbers for *Trichodesmium* rDNA 16S sequences used to map field sequences. Sequences were selected following Hynes *et al.* (2012).

Table S3. 1) Taxonomic overview of the epibiont community per sample. 2) OTU identification and total number of reads associated to each OTU per sample.

Table S4. Results of two-way permutation ANOVA (Permanova) of Bray–Curtis dissimilarity matrix between epibiont communities collected across ocean region and colony morphologies (9999 permutations).

Table S5. Total number of differentially enriched KEGG pathways in puffs and rafts in each ocean. Metabolic inference was performed with PICRUSt and statistical analyses with LEfSe. KEGG pathways are considered significantly different for LEfSe LDA score $>$ 2 (Segata *et al.* 2010).