Combined *in vivo* and *in situ* genome-resolved metagenomics reveals novel nitrogen fixing interactions between non-cyanobacterial diazotrophs and microalgae

Udita Chandola[1], Camille Trottier[1], Marinna Gaudin[2,3], Louis Josselin Lavier Aydat[1], Eric Manirakiza[1], Samuel Menicot[1], Erik Jörg Fischer[4], Isabelle Louvet[5], Thomas Lacour[6], Timothée Chaumier[1], Atsuko Tanaka[7], Georg Pohnert[4], Samuel Chaffron[2,3] and Leila Tirichine[1*]

[1] Nantes Université, CNRS, US2B, UMR 6286, F-44000 Nantes, France.

[2] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, Nantes, France.

[3] Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, F-75016 Paris, France.

[4] Friedrich-Schiller-Universitet Jena, Institute for Inorganic and Analytical Chemistry, Lessingstrasse 8, 7743, Jena, GERMANY

[5] Nantes Université, CNRS, CEISAM, UMR 6230, F-44000 Nantes, France.

[6] Ifremer, PHYTOX, PHYSALG, Rue de l'Ile d'Yeu, BP21105, Nantes Cedex 03, 44311, France

[7] Department of Chemistry, Biology and Marine Science, Faculty of Science, University of the Ryukyus, Okinawa 903-0213, Japan

[*]Correspondence: tirichine-l@univ-nantes.fr; Tel.: +33-276645058

Abstract

Non-cyanobacteria diazotrophs (NCDs) were shown to dominate in surface waters shifting the long-held paradigm of cyanobacteria dominance and raising fundamental questions on how these putative heterotrophic bacteria thrive in sunlit oceans. Here, we report an unprecedented finding in the widely used model diatom *Phaeodactylum tricornutum* (*Pt*) of NCDs sustaining diatom cells in the absence of bioavailable nitrogen. We identified *Pt*NCDs using metagenomics sequencing and detected nitrogenase genes *in silico* and/or by PCR. We demonstrated nitrogen fixation within *Pt*NCDs while establishing their genetic affiliation to NCDs found in the environment. We showed the wide occurrence of this type of interactions with the isolation of NCDs from other microalgae, their identification in the environment, and their predicted associations with photosynthetic microalgae. Our findings will help to understand the different players driving global marine nitrogen fixation. Overall, this study provides evidence for a hitherto unrecognized symbiosis that affects marine element cycling and primary production using a multidisciplinary model-based approach.

Keywords: Biological nitrogen fixation, non-cyanobacterial diazotrophs, *Phaeodactylum tricornutum*, microalgae, symbiosis

Nearly 80% of Earth's atmosphere is in the form of molecular nitrogen (N_2). Similarly, the main form of nitrogen in the oceans which cover 71% of Earth surface is dissolved dinitrogen. The rest is reactive nitrogen in the form of nitrate, ammonia and dissolved organic compounds, which are very scarce, rendering nitrogen one of the main limiting factors of productivity in most areas of the oceans[1]. Despite its dominance, ironically, N_2 is unavailable for use by most organisms because there is a strong triple bond between the two nitrogen atoms that requires high energy input to split the molecule and supply bio-accessible nitrogen[2]. Biological nitrogen fixation which is an energy intensive process is specific to only some prokaryotes and archaea termed diazotrophs. It is catalyzed by the nitrogenase enzyme complex including the nitrogenase reductase encoded by the *nifH* gene and the nitrogenase composed of two pairs of different subunits, α and β, respectively encoded by the *nifD* and *nifK* genes[3]. Therefore, eukaryotes are only able to obtain fixed nitrogen through their interactions with diazotrophs[4].

In aquatic habitats, symbiosis between nitrogen fixing cyanobacteria and various eukaryotes, including several diatom genera such as *Chaetoceros, Climacodium* and *Hemiaulus*, are common and important in oligotrophic habitats of the ocean[5]. Because the nitrogenase complex is sensitive to oxygen, diazotrophs evolved protective nitrogen fixation alternatives that involve conditionally, temporally, or spatially separating oxidative phosphorylation or photosynthesis from nitrogen fixation[6-11]. Cyanobacteria were considered for a very long time as the main diazotrophs in the ocean. However, recent molecular analyses indicated that non-cyanobacterial diazotrophs (NCDs) are also present and active, exhibiting an important presence of diverse *nifH* gene amplicons related to non-cyanobacteria, especially to Proteobacteria and Planctomycetes[12-16].

The overwhelming dominance of NCDs raises important questions on how these presumably heterotrophs proteobacteria thrive in the photic zones of the oceans. One emerging hypothesis is their association with microalgae in symbiotic interactions for nitrogen fixation. In our work, we provide first evidence of occurrence of such interactions between NCDs and several microalgae. More importantly, we identified this interaction in an attractive biological model, the diatom *Phaeodactylum tricornutum (Pt)*, opening important and unique opportunities to investigate the molecular mechanisms of diazotrophy. We sequenced the bacterial meta-community associated with *Pt* and identified previously unknown diverse marine diazotrophs of mainly α-proteobacteria type, supporting the host cells in the absence of any form of usable nitrogen. Further, we isolated, cultured, sequenced and fully assembled most

of the NCDs identified in lab culture. Their assembled genomes clustered phylogenetically to NCDs found in the environment and showed environmental co-occurrence patterns with microalgae including diatoms, green algae and haptophytes. Our work brings evidence for an ecologically important symbiotic interaction between microalgae and NCDs that might explain how both partners cope with periods of nitrogen scarcity, and how heterotrophic NCDs compromise for nitrogen fixation with its high energy cost in fully oxygenated areas of the water column. These new findings will have to be considered in future nitrogen and carbon biogeochemical cycling studies.

Results

Identification of non-cyanobacterial diazotrophs symbiosis with diatoms

A screen of 10 xenic accessions of *Pt* in the absence of any source of usable nitrogen in the growth medium, identified one accession, *P. tricornutum 15,* sampled from East China sea (hereafter *Pt15*) that was able to survive in contrast to the axenic control (Fig. 1a), suggesting that diatom growth was supported by putative bacterial nitrogen fixation providing a usable source of nitrogen to diatom cells. After 3 months of nitrogen starvation, xenic *Pt15* showed similar growth to non-starved cells upon transfer to nitrate replete medium. This indicated that *Pt15* starved cells remained viable and capable of growth under favorable conditions as evidenced by the presence of brown-pigmented cells (Fig. 1b). To demonstrate nitrogen fixation, we used the widely applied Acetylene Reduction Assay (ARA)[17] that showed ethylene production, although low, in xenic *Pt15* cultures starved for nitrogen, but not in plus N condition (+N). No ethylene was detected in axenic *Pt15* confirming that the detected ethylene was not endogenous to the diatom cells but originated from bacterial nitrogen fixation (Supplementary Table 1).

To further investigate nitrogen fixation in minus N condition (-N), we measured Particulate Organic Nitrogen quota (PON) in diatom cells and showed a stable level of PON even after 80 days of growth, supporting the maintenance of diatom cells with infinite amounts of usable nitrogen likely provided by *Pt15* associated bacteria (Supplementary Fig. 1a, Supplementary Table 1). Simultaneously, measured biomass of *Pt15*, indicated a stable cell population suggesting a ground state photosynthetic activity that might support the energy demand of nitrogen fixation (Supplementary Fig. 1b). Taken together, our results suggest bacterial nitrogen fixation that maintain diatom population cells stable and alive.

Light microscopy showed a significant accumulation of bacteria in -N compared to +N condition (Fig. 1c,d,e,f). Transmission electron microscopy confirmed this observation and revealed a peculiar structure with a higher electron density around diatom cells looking like a pellicle of exopolysaccharides exclusively present in starved cells (Fig. 1g,h). Similar structures, although less prominent occurred around some of bacterial cells either as a single bacterium or a group of several bacteria. The probability that these structures are composed of polysaccharides is supported by the TI blue stain, a well-known dye that has an affinity for staining polysaccharides[18].

To identify species composition of *Pt15* bacterial community in -N and +N conditions, bacterial DNA was extracted and shotgun sequenced revealing a total of 153,4 million reads. Using the GTDB reference database, we identified a total of 2 classes, 8 families, 12 orders and 90 genera. At class level, α-proteobacteria represented 98% ± 0,8 in +N and 99% ± 0,5 in -N, based on relative abundance estimations, while *Gammaproteobacteria* were detected with a relative abundance below 1% (Figure 2a). At family level, *Sphingomonadaceae* and *Rhizobiaceae*, both known to contain diazotrophic species, dominated (Figure 2c). At genus level, among the 90 distinct taxa identified, 73 genera were only detected in -N with no reads in +N (Supplementary Table 2). Composition of the community microbiome at genus level showed a dominance of *Sphyngopixys* in +N, while -N showed a significant enrichment in *Mesorhizobium sp.*, *Bradyrhizobium sp.*, *Sphingomonas sp.*, and *Brevundimonas sp.*, all known as nitrogen fixing taxa (Fig. 2d). Compared to +N, *Sphyngopixys* richness decreased in -N although its dominance persisted in this condition. Interestingly, several presumably non-nitrogen fixing species (67 out of 73) were significantly enriched in -N (Supplementary Table 2), suggesting a role of these species in direct or indirect interactions within the bacterial community and/or with the host.

Isolation/identification of *P. tricornutum* associated non-cyanobacterial diazotrophs

Given the unique nature of the identified interaction involving a model diatom species and NCDs, it was important to isolate the diazotrophic bacteria identified *in silico* from metagenomics (MetaG) data in order to further explore this interaction. We used universal media and took advantage of the species assignation provided by MetaG to target the isolation of the desired bacteria according to literature reports[19-22]. Therefore, different growth conditions (media, temperature) were tested (Supplementary File 1) and permitted the isolation of 9 bacterial species (Table 1). Bacterial colonies were identified using either full length 16S rRNA gene or two hypervariable regions of the 16S (V1V2, V3V4). The presence of *nifH* was assayed

using degenerate primers (Supplementary Fig. 2a,b). Five bacteria were identified as nitrogen fixers based on a combination of *nifH* PCR amplification and/or *nifH in silico* detection along with observation of growth on nitrogen free medium (Table 1, supplementary Figure 2). The whole set of *nif* genes including the *nifHDK* operon were identified in *Bradyrhizobium* MAG21, the only bacterium that was sequenced as a single species, while the other MAGs were assembled from MetaG data. Two and four copies of *nifH* were detected in *Bradyrhizobium* MAG21 and *Mesorhizobium* MAG3, respectively. Among the nine isolated bacteria, one species *Methylobacterium* was not detected in metagenomics sequences likely due to its low abundance, emphasizing the importance of combining sequencing and isolation/culturing approaches (Table 1).

To provide stronger support for nitrogen fixation, we monitored $^{15}N_2$ gas-incorporation using mass spectrometry. This allowed a direct measure of nitrogen fixation in two isolated bacteria, *Bradyrhizobium* MAG21 and *Methylobacterium*. The labeling of various amino acids indicated substantial $^{15}N_2$ gas assimilation in both bacteria as well as the positive controls *Pseudomonas stutzeri* and *Azotobacter vinalendii*, while the negative controls, (the axenic Pt15 and *E. coli*) did not show any nitrogen fixation (Supplementary Table 1). This unequivocally establishes the nitrogen-fixing capability of the Pt15-associated bacteria, namely *Bradyrhizobium* and *Methylobacterium*. We could not observe $^{15}N_2$ incorporation in xenic Pt15 cells, which can be attributed to their lower bacterial population density when compared to the cultured high biomass of *Bradyrhizobium* and *Methylobacterium*, as well as the competition with other numerous bacteria for fixed nitrogen. To verify that the ability of Pt15 cells to uptake bioavailable nitrogen was not altered, we cultured xenic Pt15 in artificial seawater containing L-valine labeled with ^{15}N, both in the presence and absence of nitrate. The analysis of endometabolomes using LC-MS distinctly revealed the incorporation of labeled valine in xenic Pt15 when nitrate levels were at 0 µM (Table S1). However, no such incorporation was observed in diatom cells when nitrate was present in the medium (545 µM). This suggests that Pt15 is able to take up nitrogen in form of amino acids from the associated NCDs if available nitrogen is otherwise limited.

Diazotrophic MAGs are diverse and closely affiliated to environmentally sampled diazotrophs of Alphaproteobacteria type

To gain insights into *Pt15* associated diazotrophs and their phylogenetic relationship with known taxa, we assembled their genomes and recovered several MAGs with completeness exceeding 97% (Table 1, Supplementary Table 3). To achieve a better identification of the bacterial genetic material, we sought to recover plasmids from non-assembled reads, which are important elements for bacteria survival susceptible to contain nitrogen fixation genes. We used three different plasmid prediction tools and recovered 42 plasmids assigned to different bacterial species with an average length of 97013 bp (Supplementary Table 3). Besides *nifH*, several additional *nif* genes were detected in most of the MAGs, among which two were found in plasmids, *nifU* and *nifS*. Phylogenomic analysis of *Pt15* MAGs, MAGs of known terrestrial and marine diazotrophs from the MicroScope database[24] and MAGs from *Tara* Oceans data using anvi'o[25] revealed two clusters, A (*Mesorhizobium*, *Bradyrhizobium* MAGs, *Georhizobium*, *Phylobacterium*, and *Bevundimonas*) and B (2 *Sphingopyxis* MAGs and *Sphingobium*). Cluster A was affiliated to terrestrial and marine diazotroph MAGs with a closer affiliation to terrestrial MAGs of the same genus (Fig. 3), while cluster B was related to genera from marine and terrestrial environments identified as nitrogen fixers. In both clusters, whenever the related MAGs were not identified as diazotrophs (likely due to incomplete assembly), they were found to belong to Rhizobiales, Parvibaculales, Caulobacterales and Sphingomonodales, orders known to contain nitrogen fixing species[26-28]. For a clearer depiction of Figure 3, an interactive phylogenetic tree is available on the following link: https://ncdstree.univ-nantes.fr. Altogether, our results including (i) survival of the xenic diatom in the absence of usable forms of nitrogen, (ii) nitrogen fixation demonstrated by ARA and ^{15}N-incorporation, (iii) detection of *nifH*, and (iv) the isolation of NCDs from the xenic *Pt15* clustering with known proteobacteria diazotrophs, indicate that an interaction for nitrogen fixation likely occurs in *Pt15* xenic cultures, supporting the survival of diatoms in -N condition.

Functional gene annotation of *Pt15* associated NCDs in nitrate deplete versus replete conditions

To understand the biological functions of the microbial community, genes from -N and +N treatments were assigned to functional categories (Figure 4, supplementary Table S4). There was a significant enrichment of genes related to nitrogen fixation and metabolism in -N compared to +N condition. Typically, in -N, we exclusively identified *nifH* and *nifD*, two key components of the nitrogenase enzyme complex. Likewise, *nifX* and *nifT* were only detected in -N suggesting their critical role in nitrogen fixation. Part of the nif operon, *NifX* was reported

to be important in FeMo-co maturation and its transfer to *NifDK* proteins[29,30]. Other genes related to nitrogen fixation metabolism emerged from the analysis, namely *HemN*, *Fpr*, *FixS*, *FixA*, *FixH*, *GroEL*, *NifS* and *FixJ* (Figure 4), some of which are part of the NIF regulon.

The abundance of several transporters was striking in -N, suggesting an important role in energy/substrate shuttling between the host and the bacterial community. Examples included an *EcfT* gene (energy-coupling factor transporter transmembrane) known for its function in providing the energy necessary to transport a number of different substrates[31], AcpS, an acyl carrier gene, which functions as a multi-carbon units carrier into different biosynthetic pathways[32], TauB, a taurine transporter that might play a role in providing taurine, which is produced by a large number of algae and known as a potential source of carbon, nitrogen and energy[33]. Stress responsive genes were clearly more represented in -N compared to +N including SpoT, which is a general bacterial stress response[34] that allows bacteria to increase its persistence in stressful conditions, here the lack of nitrogen. Other genes found almost exclusively in the starved bacterial community were the PAS domain containing genes (Per-ARNT-Sim)[35], which function as signal sensors by their capacity to bind small molecules, modulators and transducers. They are known to sense oxygen to prevent degradation of oxygen sensitive enzymes suggesting a crucial role in -N condition where there is an enrichment in nitrogenase containing bacteria. Overall, functional analysis of MetaG data corroborated with NCDs enrichment in -N and nitrogen fixation.

Geographical distribution and ecological relevance of NCD-microalgae symbiosis

Next, we sought to assess the prevalence and importance of similar interactions in other microalgae and in the environment. First, we screened a large collection of xenic microalgae as described for *Pt15* and identified 5 species of microalgae (35%) that grew in the absence of nitrogen in the growth medium suggesting the presence of diazotrophs (Supplementary Table 5). As for *Pt15*, the retained microalgae were screened using different nitrogen deprived media under various conditions for diazotrophs isolation. *NifH* degenerate and 16S rRNA gene primers were used to identify the isolated bacterial colonies, which revealed five Alpha-proteobacteria including 2 species of *Thalassiospira sp*. (isolated from two different microalgae), *Stenotrophomonas maltophilia*, *Rhizobium* and *Rhodococcus qingshengii* suggesting that the occurrence of NCDs with microalgae is frequent. PCR detection using *nifH* degenerate primers was successful only in *S. maltophilia*. All isolated bacterial species grew in

nitrogen free medium (Supplementary Table 5). To provide additional support for the presence of nitrogen fixation in bacteria isolated from microalgae other than *P. tricornutum*, we carried out an $^{15}N_2$ incorporation assay followed by LC-MS analysis on *Rhodococcus.* The results clearly indicated that the bacteria assimilated labeled nitrogen, confirming its ability to fix nitrogen (Supplementary Table 1).

Identified NCDs were sequenced in a pool of 3 bacteria using Illumina and PacBio and their genome were assembled, showing overall, a high completeness (>87%, 2 MAGs with 100%) and low contamination (<2%) (Supplementary Table 3). We then used MAGs isolated from *Pt15* and the other microalgae (Table 1, Supplementary Table 3) to reassess their phylogenetic relationship as described for *Pt15* isolated NCDs. The analysis revealed 3 more clusters (C, D and E) in close proximity to diazotrophic MAGs from the environment and/or MAGs belonging to diazotrophic containing orders[36 37 38] (Supplementary Fig. 3). An interactive figure is available at https://ncdstree2.univ-nantes.fr/.

Next, we assessed the occurrence of *Pt15* MAGs in the environment in surface water and Deep Chlorophyll Maximum layer (DCM) using *Tara* Ocean datasets[39,40]. Among the 130 sampling stations, 92 stations showed the occurrence of at least one of the *Pt15* MAGs reflecting a high prevalence (Fig. 5a, Supplementary Fig. 4). *Pt15* closely affiliated *Tara* MAGs were more frequently occurring with 129 recorded stations. The distribution of *Pt15* MAGs was extensive, spanning all latitudes from south Pacific, Indian and Atlantic Oceans to polar regions. The MAGs were found in both surface and DCM samples, mainly within the 0.8-5µm, 0.8-2000, 3-5-20µm and 180-2000 size fractions (Fig. 5a, supplementary Fig. 4). Interestingly, *Pt15* MAGs co-occurred in surface and DCM samples in most cases, although their dominance in surface samples was prominent. Except for few stations, clusters A and B seemed to co-occur suggesting potential ecological interactions between their respective bacteria.

We then asked whether the occurrence of the MAGs isolated from *Pt15* correlated with environmental factors monitored during the *Tara* Oceans campaign and whether they co-occur with other organisms. We considered only two *Pt15* MAGs (*Mesorhizobium* MAG3 and *Bradyrhizobium* MAG21) to assess their correlation with three environmental factors used as indicators of diazotrophy, namely nitrate, iron, and oxygen concentrations. Interestingly, *Pt15* MAGs were detected in low nitrate and almost no detection above 5 µM (Fig. 5c). There are no differences in their occurrence with TARA MAGs firmly identified as NCDs while these differences are highly significant with non diazotrophic MAGs (Fig 5c). These differences were detected only in specific size fractions (Fig. 5c). This measure is in line with nitrate limitation

as a marker of diazotrophs presence [41]. In a similar way, both oxygen and iron showed a consistent pattern when comparing this study's MAGs with TARA NCDs, distinguishing these two categories from non diazotrophic MAGs (Fig 5d,e). By contrast to what is known about oxygen inhibition of nitrogenase, both MAGs seemed to thrive in relatively well oxygenated zones (200 µM) compared to less oxygenated areas or oxygen minimum zones (OMZ, 20- 50 µM O_2). This implies adaptive means by which these bacteria fix nitrogen.

Strikingly, a co-occurrence analysis revealed a clear association of the NCDs isolated from *Pt15* (Fig 5f) with micro-eukaryotes. Most of the Pt15 MAGs were found to co-occur primarily with Haptophytes, Bacillariophyta, and Cryptophytes. The association with Bacillariophyta is consistent with their isolation from a diatom. While we did not demonstrate nitrogen fixation in these predicted associations, this analysis provides evidence for a probable nitrogen fixing interaction between Pt15 MAGs and various species of primarily photosynthetic micro-eukaryotes.

Discussion

Recent reports of NCDs dominance in the euphotic zones of the oceans questioned the long-held dogma that symbiotic N_2 fixation is mainly carried by Cyanobacteria, but also raised fundamental questions on how these presumably heterotroph NCDs can thrive in the sunlit ocean[12-14,42]. The nitrogenase enzyme complex responsible for nitrogen fixation is sensitive to oxygen that irreversibly inactivates the enzyme. Therefore, diazotrophs evolved several mechanisms to protect nitrogenase from oxygen degradation including hyper-respiration, uncoupling of photosynthesis in space and time in photosynthetic diazotrophs, and attachment/interaction with microalgae or debris which were found to be hotspots of nitrogen fixation[43]. Using a multidisciplinary approach, the study described herein brings unprecedented evidence of interactions for nitrogen fixation between microalgae, in particular the model diatom *Phaeodactylum tricornutum* and NCDs. After screening several xenic accessions of *P. tricornutum* in the absence of a usable form of nitrogen, our study revealed that one accession, Pt15, interacts with nitrogen-fixing proteobacteria. Although low, this nitrogen fixation and uptake seemed to maintain the viability of the diatom and its associated bacteria. To compensate for the allocation of energy to the bacterial community, *Pt15* cells seemed to slow down growth. This lab culture interaction may reflect natural environments, where such symbiotic interactions, in extreme conditions (low to no nitrogen) take place until favorable factors (e.g.,

nitrate repletion with upwelling or Sahara dust) permit the recovery and growth of cells we observed in our experimental cultures when transferred to nitrate replete medium.

MetaG sequencing of the associated bacteria revealed not only one diazotrophic species but a group of nitrogen fixers almost exclusively Alpha-proteobacteria mainly represented by Rhizobiaceae, Sphingomonadaceae and Xanthobacteraceae. The nitrogenase gene was identified in all of them either *in silico* or by PCR amplification, supporting further diazotrophy in these bacteria. The *nifH* gene was found in most cases in the chromosome, except for *Georhizobium* where it was predicted to be plasmidic. Interestingly, in the fully sequenced and well assembled *Bradyrhizobium* MAG21, the chromosomal *nif* genes are clustered over 30-40 Kb regions forming genomic islands, a feature found in symbiotic bacteria where these clusters are known as symbiotic islands, which corresponds to an evolutionary strategy adopted by symbiotic bacteria in their quest to interact with eukaryotic hosts[44]. Demonstrated by ARA, the rates of nitrogen fixation in xenic *Pt15* in -N are low compared to the used controls but higher comparable to non-cyanobacteria diazotrophs from the environment, although monitoring techniques and conditions, environment versus lab cultures are different[45,46]. Direct confirmation of the nitrogen fixing capability of the isolated bacteria was obtained with the use of $^{15}N_2$ incorporation monitored by isotope-analysis in LC-MS. This leaves no doubt about the Pt15 associated bacteria's ability to fix nitrogen. This marks thus the first concrete evidence of interactions between diatoms and NCDs. The fact that we established this interaction in the model diatom *P. tricornutum* will allow rapid in depth analysis of the underlying processes.

Bacteria associated with Pt15, identified as NCDs in this study, align with the distribution patterns of *Tara* NCDs in ocean regions characterized by low nitrate, relatively low oxygen, and low iron levels. This observation is in accordance with findings from multiple studies that have reported similar diazotroph distributions under the aforementioned conditions. Indeed, diazotrophs have developed various strategies to counteract the irreversible degradation of the nitrogenase enzyme by oxygen. These strategies include forming associations with other organisms, among other mechanisms, as they seek out low-oxygen environments primarily facilitated by surrounding polymers. Similarly, in regions with low iron levels, NCDs may have an advantage in accessing dissolved iron due to their smaller cell size, resulting in a higher surface area-to-volume ratio compared to other diazotrophs like *Trichodesmium* which are more prevalent in areas such as the North Atlantic, where sporadic Saharan dust inputs seasonally alleviate iron limitations[47 48]. Additionally, these NCDs could potentially employ siderophore

binding proteins capable of binding the hydroxamate siderophore ferrichrome. They may also derive benefits from the release of iron from microalgae they are associated with. Several mechanisms for accessing the organic pool of ligand-bound iron have been reported in various diatoms, including *P. tricornutum*. [49 50 51 52 53].Our study revealed that NCDs interaction with *Pt15* and other microalgae is likely not an isolated phenomenon and is more broadly distributed than previously anticipated. Our MAGs were detected in the environment as closely affiliated with previously identified terrestrial and marine NCDs, supporting diazotrophy and their ecological relevance. These MAGs were identified at global scale in surface and DCM samples in microalgae containing size fractions, which strongly supports the hypothesis of an interaction with them as evidenced by the co-occurrence analysis. This interaction might be taking advantage of hypo-oxic biomes microalgae and bacteria may provide with the secretion of extracellular polymers, mainly polysaccharides, that form biofilms or phycospheres ensuring low O_2 diffusion, which was suggested by our TEM analysis with the formation of pellicles around *Pt* cells and bacteria. Such pellicles were reported to form around two nitrogen fixing bacteria, *P. stutzeri* and *Azospirillum brasilense* in oxygenated surroundings implying the importance of oxygen stress in the formation of these structures[54]. The presence of these sac-like structures might have hindered diatom uptake of released ammonium leading to growth inhibition. Besides biofilm formation, this interaction may combine different strategies for both microalgae and NCDs to avoid/limit oxygen inactivation of nitrogenase such as a conformational switch that protects nitrogenase from O_2 degradation[10,55] or hyper-respiration[56] among other protective mechanisms that need further investigations. In line with our discovery of NCDs interaction with microalgae for nitrogen fixation, a recent study[57] suggested that anaerobic processes, including heterotrophic N_2 fixation, can take place in anoxic microenvironments inside sinking particles even in fully oxygenated marine waters.

The unique interaction we describe here between a model diatom and a bacterial community of diazotrophs available as lab culture, avoided environmental sampling-related issues allowing high quality sequencing and complete to near-complete assembly of several MAGs with taxonomic assignation to the genus/species level. The high-resolution taxonomic assignation of our MAGs represents a reliable reference for genus assignation of the closely affiliated *Tara* Oceans MAGs with unknown taxonomic affiliation. The identification of different diazotrophic taxa with a model diatom opens wide perspectives to investigate the role of individual and combined interactions of bacterial species with the host and within the bacterial community. Associated and often overlooked non-diazotrophic bacteria enriched in -

N, likely contribute to the symbiosis through different means such as buffering the competition between several diazotrophs in the interaction with the host, producing microalgae growth compounds and keeping away antagonists to secure the interaction with the host. Thus, *Pt15* accession with its associated NCDs represent a unique Eukaryotic-bacteria model system and an exciting opportunity to study molecular mechanisms underlying this symbiotic interaction for nitrogen fixation and beyond.

Methods

Microalgae species used and culture media

Screening of Phaeodactylum tricornutum accessions

Ten xenic *P. tricornutum* accessions were screened in presence (545 µM) and absence of nitrate in Enhanced Artificial Sea Water (EASW) in order to select xenic *Pt* ecotype growing in the absence of nitrate. Starting concentration for all the accessions was 100,000 cells per ml. The cells were counted and washed twice with nitrate-free EASW in order to remove all residual nitrate and resuspended in respective media for screening over a period of one month. A growth curve was made by counting the cells in periodic time intervals. The cultures were grown at 19°C with a 12/12 hours photoperiod at 30 to 50 µE/s/m2. A cocktail of antibiotics (Ampicillin 100 µg/µl, Chloramphenicol (14 µg/µl) and Streptomycin 100 µg/µl) was used to make *Pt15* axenic in liquid EASW culture medium with refreshment of antibiotic containing medium every 5 days.

Screening of other microalgae

A total of fourteen microalgae were screened in nitrate-deplete and nitrate-replete enhanced artificial sea water for growth (Supplementary Table 5). About 100,000 cells were counted, resuspended in their respective media and grown as described above. Their growth was assessed for a period of two weeks at regular intervals.

Acetylene reduction assay

Acetylene reduction assay (ARA)[17] was used to quantify the conversion of acetylene to ethylene (C_2H_4) which allows the assessment of nitrogenase activity. ARA was conducted on the following samples: two positive controls, *Azotobacter. vinelandii* (NCIMB 12096), *Pseudomonas stutzeri* BAL361, xenic *Pt15* grown in NFM[58,59] or nitrogen replete medium, all incubated at three different temperatures (19, 25, 30°C). Axenic *Pt15* culture grown in NFM medium + 5 sugars (final concentration of 1% each: D-glucose, D-sucrose, D-fructose, D-galactose and D-mannose) was used in every measurement as a negative control to ensure there

was no endogenous ethylene produced. All the cultures were grown at their respective growth media and conditions. After reaching an appropriate O.D. between 0.2-0.3, 1 ml of the bacterial cells were collected in an Eppendorf's tube and centrifuged at11,000 rpm for 10 minutes at RT to remove the medium. The cells were re-suspended in NFM with five sugars and grown in their respective medium. When the bacterial growth reached an O.D. of 0.2-0.3, the cultures were prepared for acetylene reduction assay. The cultures were secured with a rubber stopper in order to make them air tight. Ten percent of the air was removed from the culture flasks and supplemented with 10% acetylene gas. The cultures were incubated at their appropriate temperatures for a period of 72 hours before monitoring ARA with gas chromatography. A volume of 500 µl was taken from the headspace of the samples using a Hamilton gas tight syringe (SYR 500 ul 750 RN no NDL, NDL large RN 6/pk) and injected into an Agilent gas chromatograph model 7820 model equipped with a Flame Ionization Detector (GC-FID) and using Hydrogen as carrier gas at a constant flow of 7.7 mL/min. Samples were injected into a split/splitless injector, set at 160℃ in splitless mode. A GS-Alumina capillary column, 50 m × 0.53 mm × 0.25 µm (Agilent, 115-3552) was used. The programmed oven temperature was at 160℃ (isotherm program) during 2 min. The FID detector was set to 160℃. A standard curve was made with ethylene gas using 5 volumes (100, 150, 200 µl). The raw data generated from the GC provides value of y, Area [pA*s]. This value is supplemented in the slope-intercept formula, y = mx + c, with 'm' and 'c' representing slope and constant values respectively, denoted in the calibration curve. The amounts of ethylene were calculated as described previously[60]. Acetylene reduction rates were converted to fixed N using the common 3:1 (C2H2 :N2) molar conversion ratio[61,62].

Particulate organic nitrogen measurements

One liter of 200,000 cells/ml culture of xenic *Pt15* was grown in nitrate-free EASW for over a period of 80-days at 19°C and 12h/12h light/dark photo period at 70µE/s/m2. Six time points were used for harvesting samples: 1, 4, 12, 59, 76 and 80 day(s). The cells were counted at each time point using a cell counter. Twenty ml of *Pt15* culture were harvested and injected through a sterile (treated at 450 deg/4hrs) Whatman GF/F glass microfibre filter paper to collect *Pt15* cells and the filters were dried overnight in 60°C incubator and stored at -80°C. The 20 ml filtrate from the cultures was passed through 0.22 µm filters (Sigma ca. GVWP10050), in order to remove bacterial cells and stored at -80°C, before being processed for particulate organic nitrogen measurement analysis. 545 µM nitrate-EASW, 0 µM nitrate-EASW and three heat-

treated Whatman GF/F glass microfibre filters were used as blanks for comparison with test samples.

Test for incorporation of the ^{15}N isotope

Stationary cultures (30 mL) of either the diatoms, *Phaeodactylum tricornutum Pt15* Xenic Plus nitrate (545 µM), *P. tricornutum Pt15* Xenic Minus nitrate (0 µM), *P. tricornutum Pt15* Xenic low nitrate (20 µM), *P. tricornutum Pt15* Axenic Plus nitrate (545 µM) or bacteria, *Bradyrhizobium MAG21*, *Methylobacterium*, *Azotobacter vinalendii*, *Pseudomonas stutzeri BAL361*, *Escherichia coli DH5a* and *Rhodococcus* were put under a $^{15}N_2$ (20 mL, 98 atom %, Sigma-Aldrich, Steinheim, Germany) atmosphere in 50 mL sealed culture flasks. Diatom cultures were incubated at 18 °C under a 14:10 h light: dark regime with an illumination of 60-65 µmol photons m^{-2} s^{-1}. Bacterial cultures were incubated at 100 rpm and 30 °C, except for *E. coli DH5a* which was incubated at 37 °C. After 2.5 days the diatom cultures were filtered (1.2 µm pore size, GF/C, Whatman, GE Healthcare, Little Chalfont, United Kingdom). The cells were washed off the filter using 5 mL ice-cold methanol/ethanol/chloroform 1:3:1 (methanol: HiPerSolv Chromanorm, VWR Chemicals, Dresden, Germany; ethanol: HiPerSolv Chromanorm, VWR Chemicals, Dresden, Germany; chloroform: SupraSolv, VWR Chemicals, Dresden, Germany) and the resulting extracts were treated in an ultrasonic bath for 10 min and centrifuged (16.100 x g, 20 min, 4 °C). The supernatant was stored at -20 °C until analysis. Bacterial cultures were centrifuged (3.893 x g, 20 min, 4 °C) and the cells were resuspended in 5 mL of methanol/dichlormethane/ethylacetate 10:2:3 (dichlormethane: SupraSolv, VWR Chemicals, Dresden, Germany; ethylacetate, HiPerSolv Chromanorm, VWR Chemicals, Dresden, Germany). The suspensions were sonicated (5 min, 9 cycles, 80-85 %) and afterwards centrifuged (3893 x g, 20 min, 4 °C). The supernatant was stored at -20 °C until further use.

Test for uptake of ^{15}N-L-valine-by *P. tricornutum*

To stationary cultures (40 mL) of either *Phaeodactylum tricornutum PT15* Xenic Plus nitrate (545 µM) or *P. tricornutum PT15* Xenic Minus nitrate (0 µM) 100 µL of an aqueous ^{15}N-L-valine- solution (98 atom % ^{15}N, Merck, Saint Louis, USA, 50 µmol L^{-1} in medium) were added to reach a final concentration of 125 nm L^{-1}. The cultures were then incubated, extracted and the samples stored as described in "*Test for incorporation of the ^{15}N isotope".*

Analysis of endometabolomes with LC-MS

Cell extracts were dried in a nitrogen stream at room temperature and resuspended in either 200 µL methanol for diatom extracts or methanol/acetonitrile/water 5:9:1 (acetonitrile: HiPerSolv Chromanorm, VWR Chemicals, Dresden, Germany; water: CHEMSOLUTE, TH.GEYER, Renningen, Germany) for bacterial extracts. Metabolites were separated on a Dionex Ultimate 3000 UHPLC coupled to a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany). Chromatographic separation was performed on a SeQuant ZIC-HILIC column (5 µm, 200 Å, 150 × 2.1 mm, Merck, Germany) equipped with a SeQuant ZIC-HILIC guard column (20 × 2.1 mm, Merck, Germany) at 25 °C using water with 2 % acetonitrile and 0.1 % formic acid as eluent A and 90 % acetonitrile with 10 % water and 1 mmol L^{-1} ammonium acetate as eluent B. Full scan measurements (*m/z* 80-1200, resolution 70k, automatic gain control target $3E^6$, maximum ion injection time 200 ms) were performed in positive and negative ion mode using a heated electrospray ionization source to generate molecular ions.

The method duration was 14.5 min with an MS runtime from 0.5 min to 9.0 min, a flow rate of 0.6 mL min^{-1} and a gradient as follows: 85 % B (0.0-4.0 min), linear gradient to 100 % A (4.0-5.0 min), 100 % A (5.0-9.0 min), linear gradient to 100 % B (9.0-10.0 min), 100 % B (10.0-12.0 min), linear gradient to 85 % B (12.0-12.5 min), 85 % B (12.5-14.5 min). Amino acid standards were run separately for comparison of retention time and mass spectra. Resulting raw data was analyzed with FreeStyle (Thermo Fisher Scientific).

Microscopy

Light microscopy

Living cells were directly used for preparations. Samples were observed with differential interference contrast (DIC) microscopy by BX53 microscope (Olympus, Japan) and Axiocam 705 color (Carl Zeiss, Germany).

Transmission electron microscopy

Samples were prepared as described previously[63]. Embedded samples were cut for making pale yellow ultra-thin sections (70-80nm thickness) with a microtome EM UC6 (Leica Microsystems, Germany). Sections were stained with TI blue (Nisshin EM, Japan) and lead citrate for inspection with an electron microscope JEM-1011KM II (JEOL, Japan).

Bacterial isolation and culture media used

A total of eight media recipes were used for bacteria isolation and culture under different conditions, all summarized in Supplementary File 1. All isolated bacteria were checked for purity by repeated streaking and 16S sequencing. Bacterial species were maintained in 20% glycerol at - 80°C for long-term storage.

DNA extraction and PCR amplification of marker genes

For initial identification of bacteria, the isolated colonies were picked under sterile conditions using filter tips, resuspended and mixed thoroughly in 50 µl nuclease-free water (NFW). The suspension was heat/cold treated at 95°C following 4°C, 10 min each and twice for cell lysis. Post-treatment, an additional 50 µl NFW was added to the suspension for a total volume of 100 µl. Five µl of the colony DNA was used as template for PCR.

Wizard® Promega Genomic DNA extraction kit was used for extracting high-quality genomic DNA from bacterial cells. The bacteria were grown in their appropriate media until the culture optical density (OD) reached 0.3. The cells were harvested and washed in 1X PBS twice. The cell pellets were resuspended in 480 µl 50 mM EDTA and treated with 120 µl of 20 mg/ml lyzozyme (VWR cat no. 0663-5G), incubated at 37°C for 60 min in order to lyze possible gram-positive bacteria. Post incubation, the cells were centrifuged and the supernatant was removed. Nuclear Lysis solution, 600 µl (Promega cat no. A7941), was used to re-suspend and gently mix the pellet for incubation at 80°C for 5 min. The suspension was allowed to cool at room temperature (RT) for 10 min before adding 1 µl RNAse at 10 mg/ml (VWR cat no. 0675-250MG) and incubation at 37 for 60 min. After incubation, the suspension was cooled at RT for 10 min. Protein lysis solution, 200 µl (cat no. VWR 0663-5G) was added, vortexed and incubated on ice (4°C for 5 min). The mixture was centrifuged and the clear supernatant was transferred to a clean Eppendorf with 600 µl RT isopropanol and the tube was mixed gently before being centrifuged. The cell pellet was washed with freshly prepared 70% ethanol and re-centrifuged. The ethanol was carefully removed without disturbing the pellet and air dried before resuspension in appropriate amount of NFW. The quality of the DNA was checked using Nanodrop 260/280 and 230/260 ratios and on 1% agarose gel. The identity of the bacteria was re-confirmed using 16S rRNA gene amplification and Sanger sequencing.

For the nifH PCR amplification, the reaction mixture included 2.5 µL of DreamTaq Green Buffer (1X), 0.1 µL of BSA (0.08 µg/µL), 2 µL of each primer (10 µM), 0.2 µL of DreamTaq Green Polymerase (5 U/µL), 0.5 µL of dNTP mix (10 mM), 3 µL of DNA template (normalized

to 10 ng/µL), and H2O adjusted to a total volume of 25 µL. The thermocycling protocol started with an initial denaturation step at 95°C for 5 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 53°C for 30 seconds, and extension at 72°C for 30 seconds, and a final extension step at 72°C for 7 minutes. Following the reaction, amplification results were assessed through 1% agarose gel electrophoresis, and DNA bands of interest were isolated using the NucleoSpin® Gel & PCR Cleanup Kit from Macherey-Nagel. The extracted DNA underwent Sanger sequencing at Eurofins Scientific. Sequences were analyzed using Geneious. For species identification, 16S PCR fd1-rd1, variable regions PCR V1V2 V3V4 and V5V7 were used. The reaction was performed with 35 cycles of 95 °C (1 min), 55 °C (1 min) and 72 °C (1 min). All primer sequence details are reported in supplementary Table 7.

Sample preparation and DNA extraction for Illumina and PacBio sequencing

Duplicate with a starting concentration of 200,000 cells/ml of 3-liter culture volumes of xenic *Pt15* in nitrate-replete and nitrate-deplete enhanced artificial sea water were grown for a period of 2 months (60-days) at 19°C and 12h/12h light/dark photoperiod at 70µE/s/m2 before collection for sequencing. The xenic *Pt15* cells were filtered through a 3 µm filter (MF-millipore hydrophobic nitrocellulose, diameter 47 mm (cat no. SSWP04700) to retain the cells and allow passage of the filtrate containing the bacterial community. The filtrate was filtered twice in order to avoid contamination of the bacterial cells with residual *Pt15* cells. The double filtered bacterial community was collected in 0.22-micron filter (MF-millipore hydrophobic nitrocellulose, diameter 47 mm (cat no. GSWP04700) and resuspended in 1X PBS. DNA was extracted for metagenome sequencing using Quick-DNA fungal/bacterial mini prep kit (cat no. DG005) and according to manufacturer instructions. The bacterial DNA quality was checked using 1% agarose gel to assess intact DNA and quantified using Qubit broad range DNA kit, before being sequenced. Additionally, 16S rRNA sequencing was performed to assess the presence of bacterial DNA and 18s rRNA PCR was run as a negative control to inspect absence of *P. tricornutum* DNA. When bacterial isolates were available, they were sequenced either as a single species or in a pool of 3 bacteria using Illumina and PacBio.

Data processing

Metagenomic analysis

Two different strategies were used to analyze the metagenomic reads: (i) a “Single assemblies” strategy, where assembly and binning was performed for each sample individually. This strategy is known to produce less fragmented assemblies with a limitation of chimeric contigs

formation[65]. (ii) a co-assembly strategy was also performed to complete the results with additional MAGs. The idea is to pool all the available samples reads and build a unique and common assembly for all samples. Six samples from 2 experiments were used: one with sequences obtained at 2g/L salinity (in +N and -N conditions, no replicate) and the other one at 20 g/L salinity with duplicates for each of the two conditions. Both datasets are similar.

Single assembly strategy

DNA raw reads were analyzed using a dedicated metagenomic pipeline ATLAS v2.7[65]. This tool includes four major steps of shotgun analysis: quality control of the raw reads, assembly, binning, taxonomic and functional annotations. Briefly, quality control of raw sequences is performed using utilities in the BBTools suite (https://sourceforge.net/projects/bbmap/): BBDuk for trimming and removing adapters, BBSplit for decontamination. In our case *Phaeodactylum tricornutum* reference genome (accession GCA_000150955.2), including mitochondria and chloroplast sequences, was used for this specific step. Quality-control (QC) reads assemblies were performed using metaSPADES[66] with ATLAS default parameters. The assembled contigs shorter than 500 bases length and without mapped reads were removed to keep high-quality contigs.

Metagenome-assembled genomes (MAGs) were obtained using two distinct binners : maxbin2[67] and metabat2[68]. The bins obtained by the different binning tools were combined using DASTool[69] and dereplicated with dRep[70] using default parameters. Completeness and contamination were computed using CheckM[71]. Only bins with a completeness above 50 % and a contamination rate under 10 % were kept. Prediction of open reading frames (ORFs) was performed, at the assembly level (not only at the MAGs level) using Prodigal[72]. Linclust[73] was used to cluster the translated gene product obtained to generate gene and protein catalogs common to all samples.

Co-assembly strategy

QC reads and co-assembly were analyzed using a shotgun dedicated sequential pipeline MetaWRAP v1.3.2[69]. Briefly, QC reads are pulled together and addressed to the chosen assembler metaSPAdes[66] using MetaWRAP default parameters. Metagenome-assembled genomes (MAGs) were obtained using two distinct binners: maxbin2 and metabat2 and the bins obtained were combined using an internal MetaWRAP dedicated tool. Completeness and contamination were computed using CheckM[71]. Only bins with a completeness above 50 % and a contamination rate under 10 % were kept. MetaWRAP also includes a re-assemble bins

module that collect reads belonging to each bin, and then reassemble them independently with a "permissive" and a "strict" algorithm. Only improved bins are kept in the final set. Using this strategy, we were able to add an additional MAG.

MAGs relative abundances

The QC reads were mapped back to the contigs, and bam files were generated to organize downstream analysis. Median coverage of the genomes was computed to evaluate MAGs relative abundance estimations.

Taxonomic assignation of metagenomic data

We utilized anvio version 7.2[25] for taxonomic profiling of the *P. tricornutum* associated metagenomes in nitrate deplete and replete condition. The workflow uses The Genome Taxonomy Database (GTDB)[74] to determine the taxanomy based on 22 single copy core genes (scgs, anvio/anvio/data/misc/SCG_TAXONOMY). Sequence alignment search was done using DIAMOND[75].

Hybrid assembly of MAGs

Assembly

Metagenomic assembly of Illumina technology reads does not yield complete genomes due to the difficulty of assembling repetitive regions[76]. Whenever Illumina and PacBio reads were available, we used a strategy to improve MAGs assembly by combining both type of reads. Metagenomic assembly of Illumina paired-end and PacBio reads post quality trimming were assembled using metaSPAdes[66] or SPAdes(v3.15.5) with option flag (--meta). To validate the newly assembled genomes, we verified that the long reads covered regions that were not supported by the mapped Illumina sequences only and that we get high quality bins. We obtained more contiguous genomes in which some regions had not been supported by the assembly of Illumina
reads only.

Binning

The contigs were grouped and assigned to the individual genome. To do this, we used MaxBin2 and MetaBAT2 which are clustering methods based on compositional features or on alignment (similarity), or both[77].

Refinement

Bins from MaxBin2 and MetaBAT2 were consolidated into one more solid set of bins. To do this we used Binning_refiner which improves genomic bins by combining different binning programs available at: https://github.com/songweizhi/Binning_refiner [78].

Taxonomic Assignation

GTDB-Tk has already been independently and positively evaluated for the classification of MAGs[79]. After consolidating the bins in the refinement step, we determined the taxonomy of each bin. The gtdbtk classify_wf module was used in this work using a GTDB database https://github.com/ecogenomics/gtdbtk [80].

Gene annotation of metagenomic data and MAGs

Gene annotation of metagenomics data in two conditions was done using anvi'o version 7.2. Three databases were used for a robust gene observations: COG20[81], KEGG[82] and Prokka (https://github.com/tseemann/prokka). DIAMOND was used to search NCBI's Database. In built, anvi'o program annotates a contig database with HMM hits from KOfam, a database of KEGG Orthologs (KOs, https://www.genome.jp/kegg/ko.html). For Prokka, the annotation was done externally and the gene calls were imported to anvi'o platform for the individual metagenome contigs databases.

Phylogenomic analysis of MAGs

The phylogenomic tree was built using anvi'o -7.1 with database of 1888 TARA ocean MAGs (contig-level FASTA files) submitted to Genoscope (https://www.genoscope.cns.fr/tara/). Additional closest terrestrial and marine genomes (Supplementary Table 6) to the individual MAGs were selected using MicroScope database (https://mage.genoscope.cns.fr/microscope/) using genome clustering, utilizing MASH (https://github.com/marbl/Mash) for ANI (Average nucleotide identity) for calculating pairwise genomic distancing, neighbor-joining (https://www.npmjs.com/package/neighbor-joining) for tree-construction and computing the clustering using Louvain Community Detection (https://github.com/taynaud/python-louvain). anvi'o utilizes 'Prodigal'[72] to identify open reading frames for the contig database. The HMM profiling for 71 single core-copy bacterial genes (SCGs) is done using in-built anvi'o database, Bacteria_71 (https://doi.org/10.1093/bioinformatics/btz188) and utilizes MUSCLE[83] for sequence multiple alignment. The bootstrap values are represented from 0 – 1 with replicates of 100.

Plasmid prediction and annotation

Plasmids were predicted from MetaG data using the following plasmid assembly tools: SCAPP[84], metaplasmidSPAdes[85] and PlaScope[86]. Predicted plasmids were annotated using Dfast[87], eggNOG-mapper v2[88] and Prokka 1.14.6[89].

Tara Oceans metagenome-assembled genomes counts tables

MAGs abundance count tables and fasta files were downloaded from open-source released *Tara* Oceans databases at https://www.genoscope.cns.fr/tara/[13,90]. Data included 48 potential diazotrophs among the 1888 prokaryotes MAGs - including 8 cyanobacterial and 40 heterotrophic bacterial diazotrophs - as reported in[91] and downloadable at https://www.genoscope.cns.fr/tara/.

Data included 57 and 31 samples respectively at surface and DCM depths and 6 size fractions (0.22-3µm, 3-20µm, 0.8-5µm, 20-200µm, 0.8-2000µm, 180-2000µm).

Biogeography of metagenome-assembled genomes

Total sum scaling (TSS) normalisation was performed on reads counts mapped to MAGs, by dividing all counts by the total number of reads in a given sample. Different depths (surface and DCM) and size fractions (0.22-1.6/3µm, 3/5-20µm, 0.8-5µm, 20-180/200µm, 0.8-2000µm, 180-2000µm) were analyzed separately to project relative abundances on global maps. Relative abundances of *Tara* Oceans MAGs phylogenetically close to *Pt15* and belonging to the same cluster were summed up for each sample. Sizes of the pie charts reflect the total amount of relative abundances present in each sample.

Environmental parameters preferences

Tara Oceans physico-chemical parameters were downloaded from Pangaea[39]. MAGs TSS relative abundances were projected and analyzed in relationship with ammonium (µmol/L), chlorophyll A (mg/m3), iron (µmol/L), nitrite (µmol/L), nitrate (µmol/L), phosphate (µmol/L), nitrate/phosphate, oxygen (µmol/kg), salinity (PSS-78), silicate (µmol/L), temperature (°C) and photosynthetically available radiation (mol quanta/m2/day).

Co-occurrence network inference

Co-occurrence network inference was performed using FlashWeave v0.18.0[92] with Julia v1.5.3, for eukaryotic and prokaryotic Tara MAGs linked to *Pt15* MAGs. To account for the compositional nature of MAGs abundance data, a multiplicative replacement method[93] was applied to replace zeros, and a Centered Log Ratio transformation[94] was applied separately on both prokaryotic and eukaryotic abundance matrices. FlashWeave parameters used were max_k=3, n_obs_min=10, and normalize=False.

Acknowledgements

We acknowledge Clara Guillouche and Irene Romero Rodriguez for their technical assistance with PCR screening and Romain le Balch for his help with GC runs. LT acknowledges support from the region of Pays de la Loire (ConnecTalent EPIALG project), Epicycle ANR project (ANR-19-CE20- 0028-02) and µAlgaNIF France-Japan International Research Project. UC was supported by grant 998UMR6286 Connect Talent EpiAlg from Région Pays de la Loire to LT. SC acknowledges support from the CNRS MITI through the interdisciplinary program *Modélisation du Vivant* (GOBITMAP grant), and the H2020 European Commission project AtlantECO (award number 862923). We are grateful to the bioinformatics core facility of Nantes University (BiRD Biogenouest) for its technical support. The LABGeM (CEA/Genoscope &CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la recherche, contrats ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the Microscope annotation platform.

Authors' contributions

LT conceived and designed the study. UC performed most of the experiments. SM contributed to the ARA assay and the screening. LJLA assisted with the screening and performed bacterial isolation. TC, CT, EM, UC performed bioinformatics analysis. MG performed bioinformatics analysis and generated figure 5. TL conducted the PON assay. AT performed the EM study. IL assisted with the ARA assay. EJF performed 15N assay. GP supervised EJF and provided guidance on the 15N assay. SC supervised MG. UC, TC, CT, MG, SC, EJF, GP, AT TL and LT analysed and interpreted the data. LT wrote the manuscript with methodological input from UC, MG, EF, AT, TL, EM and TC. LT supervised and coordinated the study. All authors read, edited and approved the manuscript.

References

1 Moore, C. M., Mills, M.M., Arrigo, K.R. I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, , R. J. Geider, C. G., S. L. Jaccard, T. D. Jickells, J. La Roche, T. M. Lenton, N. M. Mahowald, , E. Marañón, I. M., J. K. Moore, T. Nakatsuka, A. Oschlies, M. A. Saito, T. F. Thingstad, & Ullo, A. T. O. Processes and patterns of oceanic

nutrient limitation. *Nature Geoscience* **6**, 701-710 (2013).

2 Lee, C. C., Ribbe, M. W. & Hu, Y. Cleaving the n,n triple bond: the transformation of dinitrogen to ammonia by nitrogenases. *Met Ions Life Sci* **14**, 147-176 (2014). https://doi.org:10.1007/978-94-017-9269-1_7

3 Einsle, O. & Rees, D. C. Structural Enzymology of Nitrogenase Enzymes. *Chem Rev* **120**, 4969-5004 (2020). https://doi.org:10.1021/acs.chemrev.0c00067

4 Kneip, C., Lockhart, P., Voss, C. & Maier, U. G. Nitrogen fixation in eukaryotes--new models for symbiosis. *BMC Evol Biol* **7**, 55 (2007). https://doi.org:10.1186/1471-2148-7-55

5 Fiore, C. L., Jarett, J. K., Olson, N. D. & Lesser, M. P. Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol* **18**, 455-463 (2010). https://doi.org:10.1016/j.tim.2010.07.001

6 Cornejo-Castillo, F. M. & Zehr, J. P. Hopanoid lipids may facilitate aerobic nitrogen fixation in the ocean. *Proc Natl Acad Sci U S A* **116**, 18269-18271 (2019). https://doi.org:10.1073/pnas.1908165116

7 Inomura, K. *et al.* Quantifying Oxygen Management and Temperature and Light Dependencies of Nitrogen Fixation by Crocosphaera watsonii. *mSphere* **4** (2019). https://doi.org:10.1128/mSphere.00531-19

8 Inomura, K., Wilson, S. T. & Deutsch, C. Mechanistic Model for the Coexistence of Nitrogen Fixation and Photosynthesis in Marine Trichodesmium. *mSystems* **4** (2019). https://doi.org:10.1128/mSystems.00210-19

9 Munoz-Marin, M. D. C. *et al.* The Transcriptional Cycle Is Suited to Daytime N2 Fixation in the Unicellular Cyanobacterium "Candidatus Atelocyanobacterium thalassa" (UCYN-A). *mBio* **10** (2019). https://doi.org:10.1128/mBio.02495-18

10 Schlesier, J., Rohde, M., Gerhardt, S. & Einsle, O. A Conformational Switch Triggers Nitrogenase Protection from Oxygen Damage by Shethna Protein II (FeSII). *J Am Chem Soc* **138**, 239-247 (2016). https://doi.org:10.1021/jacs.5b10341

11 Bothe, H., Tripp, H. J. & Zehr, J. P. Unicellular cyanobacteria with a new mode of life: the lack of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch Microbiol* **192**, 783-790 (2010). https://doi.org:10.1007/s00203-010-0621-5

12 Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* **3**, 804-813 (2018). https://doi.org:10.1038/s41564-018-0176-9

13 Delmont, T. O. P. K., J.S. Veseli, I., Fuessel, J. Murat Eren, A., Foster, R.A., Bowler, C., Wincker, P., Pelletier, E. Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *doi: https://doi.org/10.1101/2021.03.24.436778* (2021).

14 Farnelid, H. *et al.* Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS One* **6**, e19223 (2011). https://doi.org:10.1371/journal.pone.0019223

15 Moisander, P. H. *et al.* Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic Environments. *Front Microbiol* **8**, 1736 (2017). https://doi.org:10.3389/fmicb.2017.01736

16 Harding, K. J. *et al.* Cell-specific measurements show nitrogen fixation by particle-attached putative non-cyanobacterial diazotrophs in the North Pacific Subtropical Gyre. *Nat Commun* **13**, 6979 (2022). https://doi.org:10.1038/s41467-022-34585-y

17 Sullivan, B. W. *et al.* Spatially robust estimates of biological nitrogen (N) fixation imply substantial human alteration of the tropical N cycle. *Proc Natl Acad Sci U S A* **111**, 8101-8106 (2014). https://doi.org:10.1073/pnas.1320646111

18 Nagasato, C., Terauchi, M., Tanaka, A. and Motomura, T. Development and function of plasmodesmata in zygotes of Fucus distichus. *Botanica Marina* **58**, 229–238 (2015).

19 de Lajudie, P. *et al.* Characterization of tropical tree rhizobia and description of Mesorhizobium plurifarium sp. nov. *Int J Syst Bacteriol* **48 Pt 2**, 369-382 (1998). https://doi.org:10.1099/00207713-48-2-369

20 Yang, X. *et al.* Mesorhizobium alexandrii sp. nov., isolated from phycosphere microbiota of PSTs-producing marine dinoflagellate Alexandrium minutum amtk4. *Antonie Van Leeuwenhoek* **113**, 907-917 (2020). https://doi.org:10.1007/s10482-020-01400-x

21 Zhang, X., Tong, J., Dong, M., Akhtar, K. and He, B. Isolation, identification and characterization of nitrogen fixing endophytic bacteria and their effects on cassava production. *PeerJ. 10: e12677* **10** (2022).

22 Bostrom, K. H., Riemann, L., Kuhl, M. & Hagstrom, A. Isolation and gene quantification of heterotrophic N2-fixing bacterioplankton in the Baltic Sea. *Environ Microbiol* **9**, 152-164 (2007). https://doi.org:10.1111/j.1462-2920.2006.01124.x

23 Sy, A. *et al.* Methylotrophic Methylobacterium bacteria nodulate and fix nitrogen in symbiosis with legumes. *J Bacteriol* **183**, 214-220 (2001). https://doi.org:10.1128/JB.183.1.214-220.2001

24 Medigue, C. *et al.* MicroScope-an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. *Brief Bioinform* **20**, 1071-1084 (2019). https://doi.org:10.1093/bib/bbx113

25 Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**, 3-6 (2021). https://doi.org:10.1038/s41564-020-00834-3

26 Lesser, M. P., Morrow, K. M., Pankey, S. M. & Noonan, S. H. C. Diazotroph diversity and nitrogen fixation in the coral Stylophora pistillata from the Great Barrier Reef. *ISME J* **12**, 813-824 (2018). https://doi.org:10.1038/s41396-017-0008-6

27 Videira, S. S., de Araujo, J. L., Rodrigues Lda, S., Baldani, V. L. & Baldani, J. I. Occurrence and diversity of nitrogen-fixing Sphingomonas bacteria associated with rice plants grown in Brazil. *FEMS Microbiol Lett* **293**, 11-19 (2009). https://doi.org:10.1111/j.1574-6968.2008.01475.x

28 Skorupska, A., Janczarek, M., Marczak, M., Mazur, A. & Krol, J. Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb Cell Fact* **5**, 7 (2006). https://doi.org:10.1186/1475-2859-5-7

29 Hernandez, J. A. *et al.* NifX and NifEN exchange NifB cofactor and the VK-cluster, a newly isolated intermediate of the iron-molybdenum cofactor biosynthetic pathway. *Mol Microbiol* **63**, 177-192 (2007). https://doi.org:10.1111/j.1365-2958.2006.05514.x

30 Jasniewski, A. J., Lee, C. C., Ribbe, M. W. & Hu, Y. Reactivity, Mechanism, and Assembly of the Alternative Nitrogenases. *Chem Rev* **120**, 5107-5157 (2020). https://doi.org:10.1021/acs.chemrev.9b00704

31 Wang, T. *et al.* Structure of a bacterial energy-coupling factor transporter. *Nature* **497**, 272-276 (2013). https://doi.org:10.1038/nature12045

32 Geiger, O. & Lopez-Lara, I. M. Rhizobial acyl carrier proteins and their roles in the formation of bacterial cell-surface components that are required for the development of nitrogen-fixing root nodules on legume hosts. *FEMS Microbiol Lett* **208**, 153-162 (2002). https://doi.org:10.1111/j.1574-6968.2002.tb11075.x

33 Cook, A. M. & Denger, K. Metabolism of taurine in microorganisms: a primer in molecular biodiversity? *Adv Exp Med Biol* **583**, 3-13 (2006). https://doi.org:10.1007/978-0-387-33504-9_1

34 Boutte, C. C. & Crosson, S. Bacterial lifestyle shapes stringent response activation. *Trends Microbiol* **21**, 174-180 (2013). https://doi.org:10.1016/j.tim.2013.01.002

35 Stuffle, E. C., Johnson, M. S. & Watts, K. J. PAS domains in bacterial signal transduction. *Curr Opin Microbiol* **61**, 8-15 (2021). https://doi.org:10.1016/j.mib.2021.01.004

36 Madigan, M., Cox, S. S. & Stegeman, R. A. Nitrogen fixation and nitrogenase activities in members of the family Rhodospirillaceae. *J Bacteriol* **157**, 73-78 (1984). https://doi.org:10.1128/jb.157.1.73-78.1984

37 Singh, R. K. *et al.* Unraveling Nitrogen Fixing Potential of Endophytic Diazotrophs of Different Saccharum Species for Sustainable Sugarcane Growth. *International Journal of Molecular Sciences* **23**, 6242 (2022).

38 Sellstedt, A., Richau, K.H. Aspects of nitrogen-fixing Actinobacteria, in particular free-living and symbiotic Frankia. *FEMS Microbiology Letters* **342**, 179–186 (2013).

39 Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* **2**, 150023 (2015). https://doi.org:10.1038/sdata.2015.23

40 Karsenti, E. *et al.* A holistic approach to marine eco-systems biology. *PLoS biology* **9**, e1001177 (2011). https://doi.org:10.1371/journal.pbio.1001177

41 Ward BA, D. S., Moore CM, Follows MJ. Iron, phosphorus, and nitrogen supply ratios define the biogeography of

nitrogen fixation. . *Limnol Oceanogr.* **58**, 2059–2075 (2013).

42 Bombar, D., Paerl, R. W. & Riemann, L. Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. *Trends Microbiol* **24**, 916-927 (2016). https://doi.org:10.1016/j.tim.2016.07.002

43 Geisler, E., Bogler, A., Rahav, E. & Bar-Zeev, E. Direct Detection of Heterotrophic Diazotrophs Associated with Planktonic Aggregates. *Sci Rep* **9**, 9288 (2019). https://doi.org:10.1038/s41598-019-45505-4

44 Ochman, H. & Moran, N. A. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096-1099 (2001). https://doi.org:10.1126/science.1058543

45 Farnelid, H. *et al.* Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. *ISME J* **7**, 1413-1423 (2013). https://doi.org:10.1038/ismej.2013.26

46 Hamersley R, T. K., Leinweber A, Gruber N, Zehr JP et al . . Nitrogen fixation within the water column associated with two hypoxic basins within the Southern California Bight. *Aquat Microb Ecol.* **63**, 193–205 (2011).

47 Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., Raven, J. A., et al. . Phytoplankton in a changing world: cell size and elemental stoichiometry. *J. Plankton Res* **32**, 119-137 (2010).

48 Sohm, J. A., Webb, E. A. & Capone, D. G. Emerging patterns of marine nitrogen fixation. *Nat Rev Microbiol* **9**, 499-508 (2011). https://doi.org:10.1038/nrmicro2594

49 Soria-Dengg S , H. U. Ferrioxamines B and E as iron sources for the marine diatom Phaeodactylum tricornutum. *Marine Ecology Progress Series* **127**, 269–177 (1995).

50 Kazamia, E. *et al.* Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci Adv* **4**, eaar4536 (2018). https://doi.org:10.1126/sciadv.aar4536

51 Maldonado MT , P. N. Reduction and transport of organically bound iron by Thalassiosira oceanica (Bacillariophyceae). *Journal of Phycology* **37**, 298–310 (2001).

52 Coale, T. H. *et al.* Reduction-dependent siderophore assimilation in a model pennate diatom. *Proc Natl Acad Sci U S A* **116**, 23609-23617 (2019). https://doi.org:10.1073/pnas.1907234116

53 Groussman, R. D., Parker, M. S. & Armbrust, E. V. Diversity and Evolutionary History of Iron Metabolism Genes in Diatoms. *PLoS One* **10**, e0129081 (2015). https://doi.org:10.1371/journal.pone.0129081

54 Wang, D., Xu, A., Elmerich, C. & Ma, L. Z. Biofilm formation enables free-living nitrogen-fixing rhizobacteria to fix nitrogen under aerobic conditions. *ISME J* **11**, 1602-1613 (2017). https://doi.org:10.1038/ismej.2017.30

55 Maier, R. J. & Moshiri, F. Role of the Azotobacter vinelandii nitrogenase-protective shethna protein in preventing oxygen-mediated cell death. *J Bacteriol* **182**, 3854-3857 (2000). https://doi.org:10.1128/JB.182.13.3854-3857.2000

56 Brian, A. C. J., C. W. . The respiratory system of Azotobacter vinelandii. *Eur. J. Biochem.* **20**, 29–35 (1971).

57 Chakraborty, S. *et al.* Quantifying nitrogen fixation by heterotrophic bacteria in sinking marine particles. *Nat Commun* **12**, 4085 (2021). https://doi.org:10.1038/s41467-021-23875-6

58 Jorgensen, J. H., Pfaller, M.A., Carroll, K.C., Funke, G., Landry, M.L., Richter, S.S and Warnock., D.W. . *Manual of Clinical Microbiology*. 11th edn, Vol. 1 (2015).

59 Isenberg, H. D. *Clinical Microbiology Procedures Handbook*. 2nd Edition edn, Vol. 1 (Amer Society for Microbiology, 1992).

60 Paudel, D. *et al.* Isolation, Characterization, and Complete Genome Sequence of a Bradyrhizobium Strain Lb8 From Nodules of Peanut Utilizing Crack Entry Infection. *Front Microbiol* **11**, 93 (2020). https://doi.org:10.3389/fmicb.2020.00093

61 Seitzinger, S. P., and Garber, J. H. . Nitrogen-fixation and N-15(2) calibration of the acetylene-reduction assay in coastal marine-sediments. . *Mar. Ecol. Progr. Ser* **65–73**, 65–73 (1987).

62 Hardy RWF, H. R., Jackson EK, Burns RC. . The Acetylene-Ethylene Assay for N2 fixation: laboratory and field evaluation. *Plant Phys* **43**, 1185–1207 (1968).

63 Tanaka A, D. M. A., Amato A, Montsant A, Mathieu B, Rostaing P, Tirichine L & Bowler C. . Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom Phaeodactylum tricornutum. *Protist* **166**, 506-521 (2015).

64 Gaby, J. C. & Buckley, D. H. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database (Oxford)* **2014**, bau001 (2014). https://doi.org:10.1093/database/bau001

65 Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M. & McCue, L. A. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics* **21**, 257 (2020). https://doi.org:10.1186/s12859-020-03585-4

66 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824-834 (2017). https://doi.org:10.1101/gr.213959.116

67 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016). https://doi.org:10.1093/bioinformatics/btv638

68 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019). https://doi.org:10.7717/peerj.7359

69 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018). https://doi.org:10.1038/s41564-018-0171-1

70 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868 (2017). https://doi.org:10.1038/ismej.2017.126

71 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015). https://doi.org:10.1101/gr.186072.114

72 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010). https://doi.org:10.1186/1471-2105-11-119

73 Steinegger, M. & Soding, J. Clustering huge protein sequence sets in linear time. *Nat Commun* **9**, 2542 (2018). https://doi.org:10.1038/s41467-018-04964-5

74 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004 (2018). https://doi.org:10.1038/nbt.4229

75 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60 (2015). https://doi.org:10.1038/nmeth.3176

76 Giguere, D. J., Bahcheli, A.T., Joris, B.R., Paulssen, J.M., Gieg, L.M., Flatley, M.W., Gloor, G.B. Complete and validated genomes from a metagenome. *doi: https://doi.org/10.1101/2020.04.08.032540* (2020).

77 Meyer, F. *et al.* Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit. *Nat Protoc* **16**, 1785-1801 (2021). https://doi.org:10.1038/s41596-020-00480-3

78 Song, W. Z. & Thomas, T. Binning_refiner: improving genome bins through the combination of different binning programs. *Bioinformatics* **33**, 1873-1875 (2017). https://doi.org:10.1093/bioinformatics/btx086

79 Coil, D. A. *et al.* Genomes from bacteria associated with the canine oral cavity: A test case for automated genome-based taxonomic assignment. *PLoS One* **14**, e0214354 (2019). https://doi.org:10.1371/journal.pone.0214354

80 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* (2019). https://doi.org:10.1093/bioinformatics/btz848

81 Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**, D261-269 (2015). https://doi.org:10.1093/nar/gku1223

82 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

83 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004). https://doi.org:10.1093/nar/gkh340

84 Pellow, D. *et al.* SCAPP: an algorithm for improved plasmid assembly in metagenomes. *Microbiome* **9**, 144 (2021). https://doi.org:10.1186/s40168-021-01068-z

85 Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. Plasmid detection and assembly in genomic and metagenomic data sets. *Genome Res* **29**, 961-968 (2019). https://doi.org:10.1101/gr.241299.118

86 Royer, G. *et al.* PlaScope: a targeted approach to assess the plasmidome from genome assemblies at the species level. *Microb Genom* **4** (2018). https://doi.org:10.1099/mgen.0.000211

87 Tanizawa, Y., Fujisawa, T., Kaminuma, E., Nakamura, Y. & Arita, M. DFAST and DAGA: web-based integrated genome annotation tools and resources. *Biosci Microbiota Food Health* **35**, 173-184 (2016). https://doi.org:10.12938/bmfh.16-003

88 Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* (2021). https://doi.org:10.1093/molbev/msab293

89 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014). https://doi.org:10.1093/bioinformatics/btu153

90 Delmont, T., Gaia, M., Hinsinger DD., et. al. . Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *Cell Genomics* **2** (2022).

91 Delmont, T. O. *et al.* Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *ISME J* **16**, 927-936 (2022). https://doi.org:10.1038/s41396-021-01135-1

92 Tackmann, J., Matias Rodrigues, J. F. & von Mering, C. Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. *Cell Syst* **9**, 286-296 e288 (2019). https://doi.org:10.1016/j.cels.2019.08.002

93 Fernández, J. A. M., Vidal, C.B.& Glahn, V.P. . Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Mathematical Geology* **35**, pages253–278 (2003).

94 Quinn, T. P. *et al.* A field guide for the compositional analysis of any-omics data. *Gigascience* **8** (2019). https://doi.org:10.1093/gigascience/giz107

Legend

Fig. 1. Macro/microscopic phenotype of *P. tricornutum* cells under nitrate starvation. a, Illustration demonstrating the screen comparison of two accessions of *P. tricornutum*: xenic *Pt4* in +N and not surviving in -N (left) in comparison to the candidate accession, the xenic *Pt15* (right) in duplicates. b, from right to left: axenic *Pt15* not surviving in -N (negative control), xenic *Pt15* in -N condition and xenic *Pt15* transferred to +N medium after 3 months in -N. c-h, micrographs of *P. tricornutum* xenic cultures on plates. Cells were grown in -N (e, f, h) or +N culture plates (c, d, g). Living cells were observed by DIC (c-f), and fixed cells by transmission electron microscopy (g, h). d, Diatom cells grown in +N possessed yellow-brown color of photosynthetic pigments. Mainly in aggregated cells of fusiform and oval shapes, some bacteria were observed (arrows). d, Single or multiple whitish droplets, oil bodies, were developed next to chloroplasts (arrowheads). e, f, Diatom oval cells aggregated with huge number of bacteria (arrows) in -N culture. f, Diatom cells grown without nitrate also developed oil bodies in the oval cells. g, h, Electron micrographs indicate cross-sectioned cells with or without nitrate. Organelles with high electron density are chloroplast, and major spaces were occupied by oil bodies or vacuoles, which showed low electron densities. Number of bacteria cells (arrows) were obviously larger in -N (h) compared to +N culture (g). h, Interestingly, both *Pt* cells and some bacteria are encased in sacs surrounded with a likely EPS made membrane (white arrows) which showed darker electron density around diatom cells compared to bacteria. A low electron density space (stars) separates *Pt* cells and bacteria from the membrane. c, chloroplast, v, vacuole, n, nucleus. Scale bars, 10 μm (c-f) and 2 μm (g, h).

Fig. 2. Comparative relative abundance estimation of *Pt15* associated bacteria. Relative abundance is shown at a, class, b, family c, order and d genus levels between -N and +N conditions.

Fig. 3. Phylogenetic tree of *Pt15* assembled bacterial genomes amongst 1888 TARA bacterial MAGs and MAGs from MicroScope.

Fig. 4. Functional gene annotation of *Pt15* associated NCDs in nitrate deplete versus replete conditions. a, Relative abundance of selected genes involved in nitrogen fixation and assimilation in nitrate deplete and nitrate replete conditions. b**,** illustration representing a combined view of 50 top abundant gene in nitrate deplete and replete condition with ordering representing top relatively abundant genes in nitrate deplete conditions. Dashed line indicates the relative percentage at 50%.

Fig. 5. Geographical distribution and ecological relevance of isolated and assembled bacterial diazotrophs genomes correlating to TARA data. Geographical distribution of phylogenetically closest Pt15 associated TARA ocean MAGs in a, surface and b, DCM samples. c, d, relative abundance of the top 2 most occurring MAGs of Pt15 bacterial community (c, d, *Mesorhizobium* MAG3, e, f *Bradyrhizobium* MAG8 respectively) in the environment in respect to nitrate, oxygen and iron concentrations. f, Co-occurrence of Pt15 MAGs in the environment with different species of micro-eukaryotes inferred through FlashWeave, FDR <0.01)[92].

Table 1. List of isolated bacteria from xenic *Pt15* in -N condition.

NCDs are indicated in bold. NA, not assembled, ND, not determined.

Competing interests

The author(s) declare no competing interests

Supplementary figures legend

Supplementary Fig 1. Particulate Organic Nitrogen (PON) measurements. a, PON kinetic and b, corresponding xenic *Pt15* biomass for 6 time points over a period of 80 days in nitrate deplete and replete conditions.

Supplementary Fig 2. a, Agarose gel pictures of PCR amplified products. a, amplification of *nifH* gene in isolated bacteria using degenerate primers. 1, *Stenotrophomonas maltophilia* MAG20, 2, *Bradyrhizobium* sp. (MAG 21), 3, Positive control (PC) *Azotobacter vinelandii,* Negative control (NC). b, 1, *Paracoccus sp*., 2, *Bradyrhizobium sp.,* (MAG 21), *Azotobacter vinelandii* (PC), Negatice control (NC). A black arrow indicates the *nifH* PCR band.

Supplementary Fig 3. Phylogenomic tree representing clusters C, D and E MAGs from *Pt15* and other microalgae with 1888 TARA ocean bacterial MAGs.

Supplementary File 1. Media recipes used for isolation of bacteria from *Phaeodactylum tricornutum* and other microalgae.

Supplementary Table 1. Combined table representing amount of ethylene (Acetylene reduction assay), N15 incorporation assay and PON detection in xenic *Pt15* in nitrate deplete condition.

Supplementary Table 2. Details of taxonomy for bacterial metagenomics data in nitrate deplete and nitrate replete conditions (replicates).

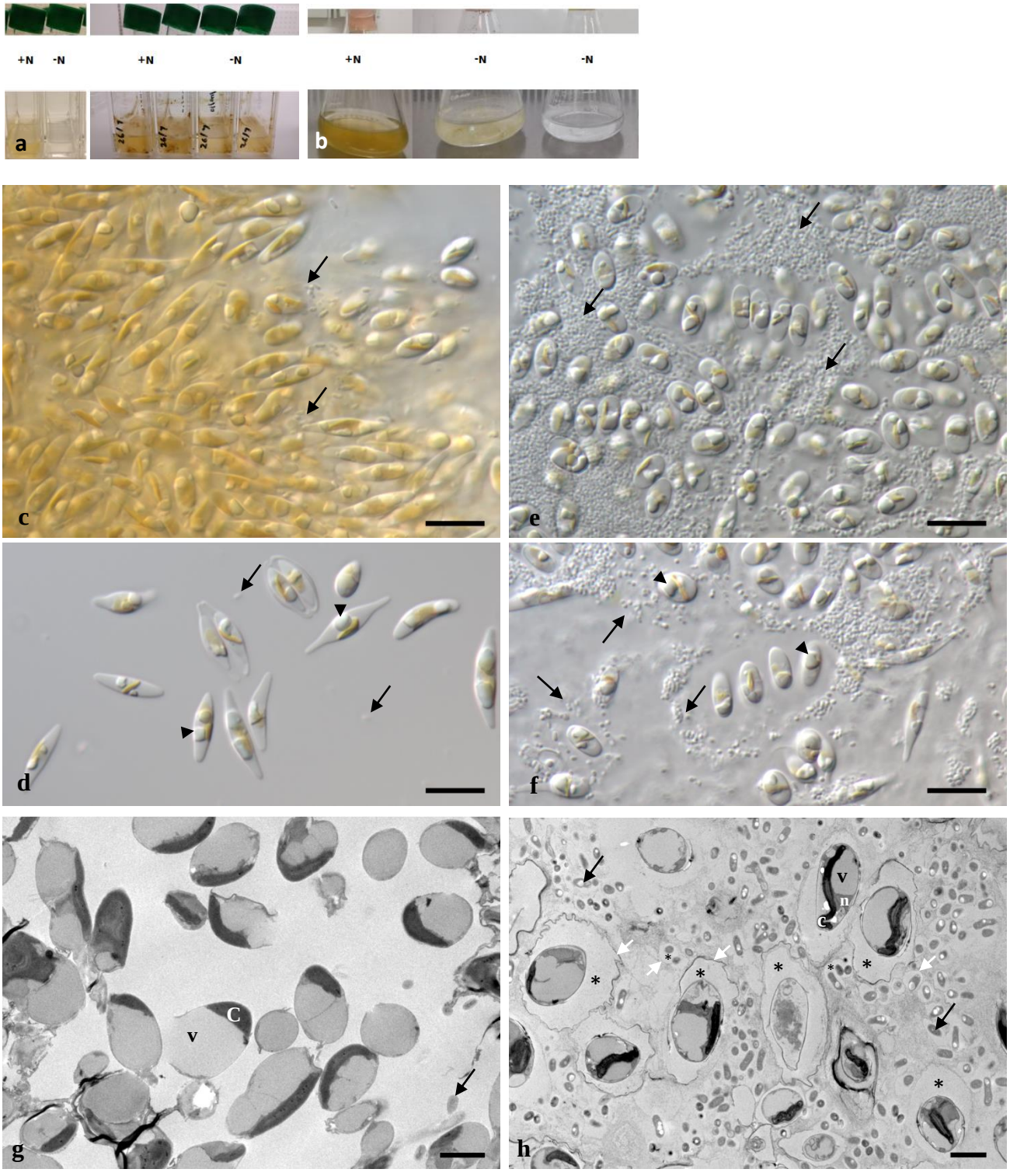
Supplementary Table 3. MAG features and species-specific plasmid detection from *Pt15* meta-genomic data.

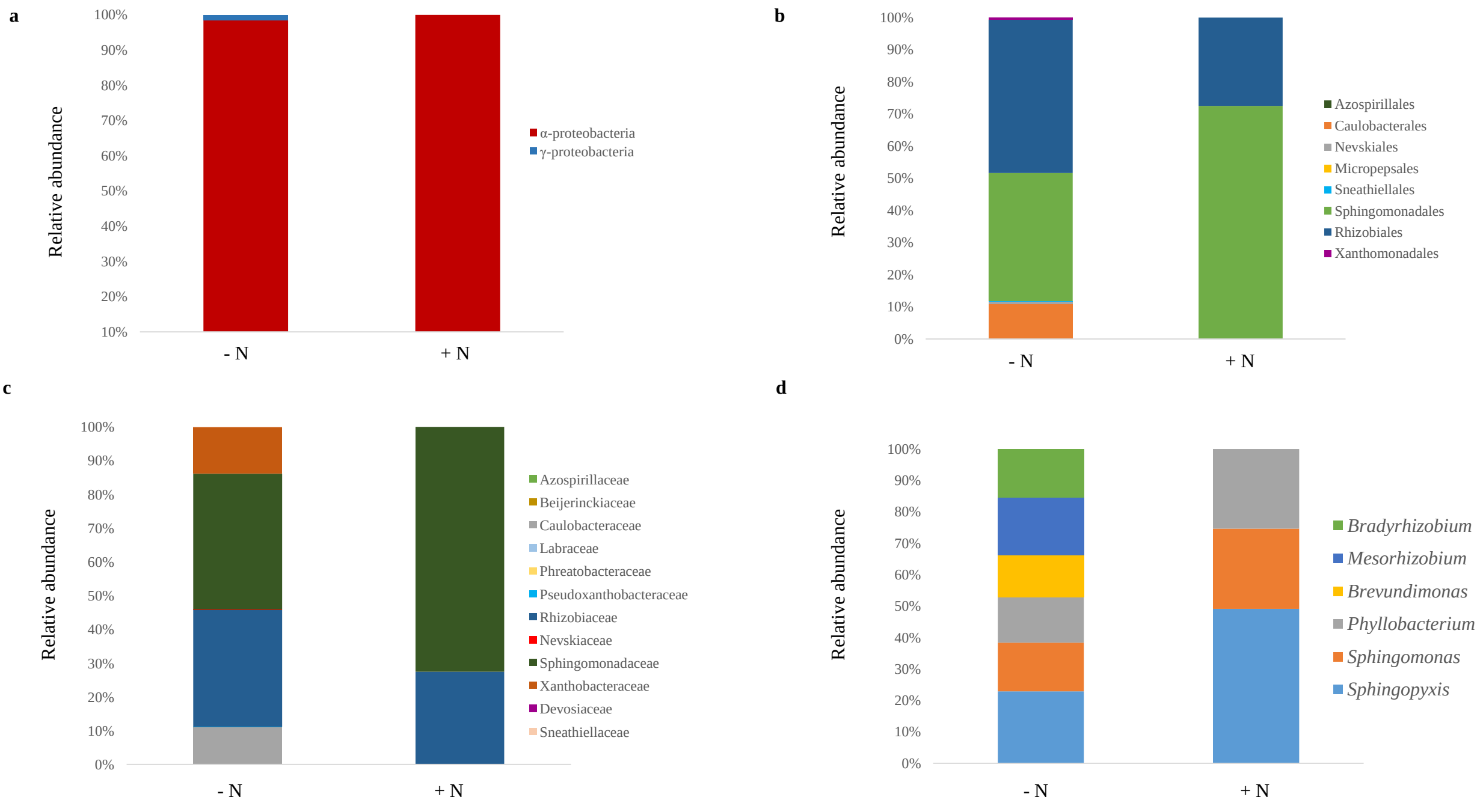
Supplementary Table 4. Gene annotation using anvio_COG20, anvio_KEGG and anvio_Prokka of the metagenomics data in nitrate deplete and replete conditions (replicates).

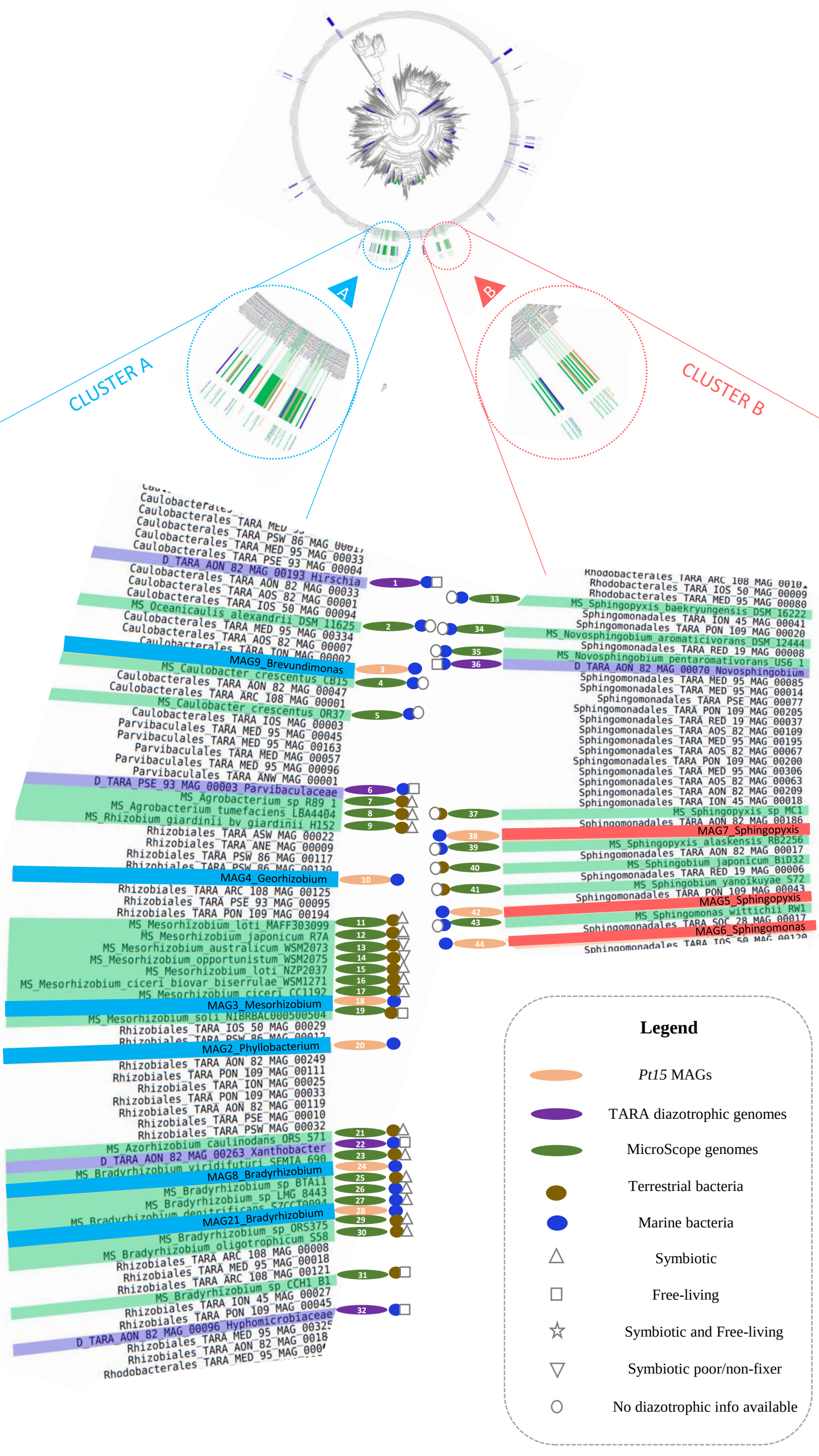
Supplementary Table 5. Details of Microalgae collection screened in nitrate deplete condition and bacteria isolated from specific microalgae.

Supplementary Table 6. List of primers used in the study.

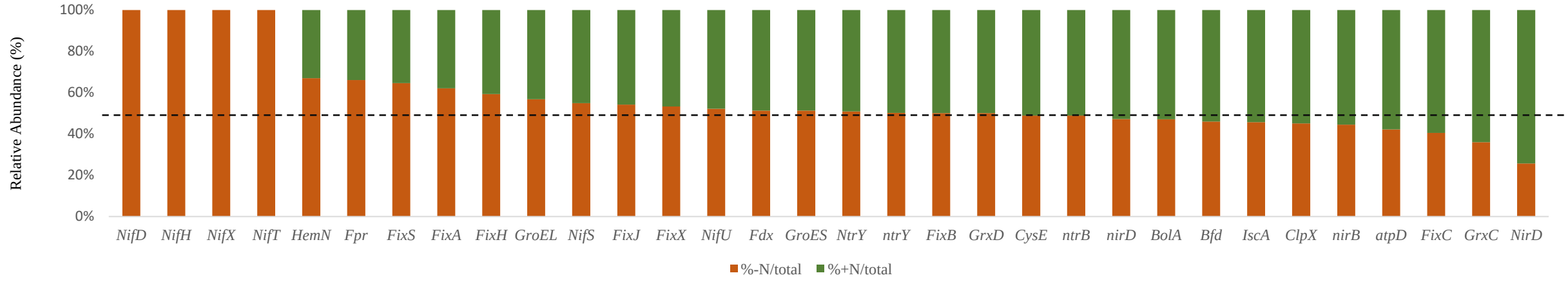
Supplementary Table 7. Features of the genomes selected from Mircoscope for phylogenetic studies.


+N
-N
+N
-N
+N
-N
-N
a
b
c
e
d
f
g
C
v
h
v
n
c

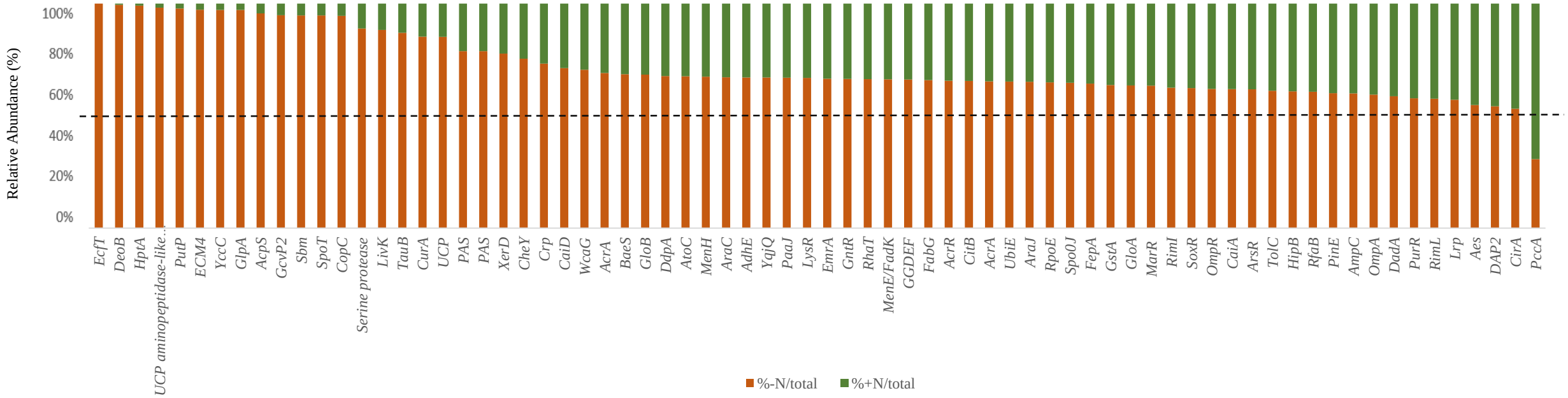

a
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
- N
+ N
α-proteobacteria
γ-proteobacteria
b
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
- N
+ N
Azospirillales
Caulobacterales
Nevskiales
Micropepsales
Sneathiellales
Sphingomonadales
Rhizobiales
Xanthomonadales
c
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
- N
+ N
Azospirillaceae
Beijerinckiaceae
Caulobacteraceae
Labraceae
Phreatobacteraceae
Pseudoxanthobacteraceae
Rhizobiaceae
Nevskiaceae
Sphingomonadaceae
Xanthobacteraceae
Devosiaceae
Sneathiellaceae
d
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
- N
+ N
Bradyrhizobium
Mesorhizobium
Brevundimonas
Phyllobacterium
Sphingomonas
Sphingopyxis

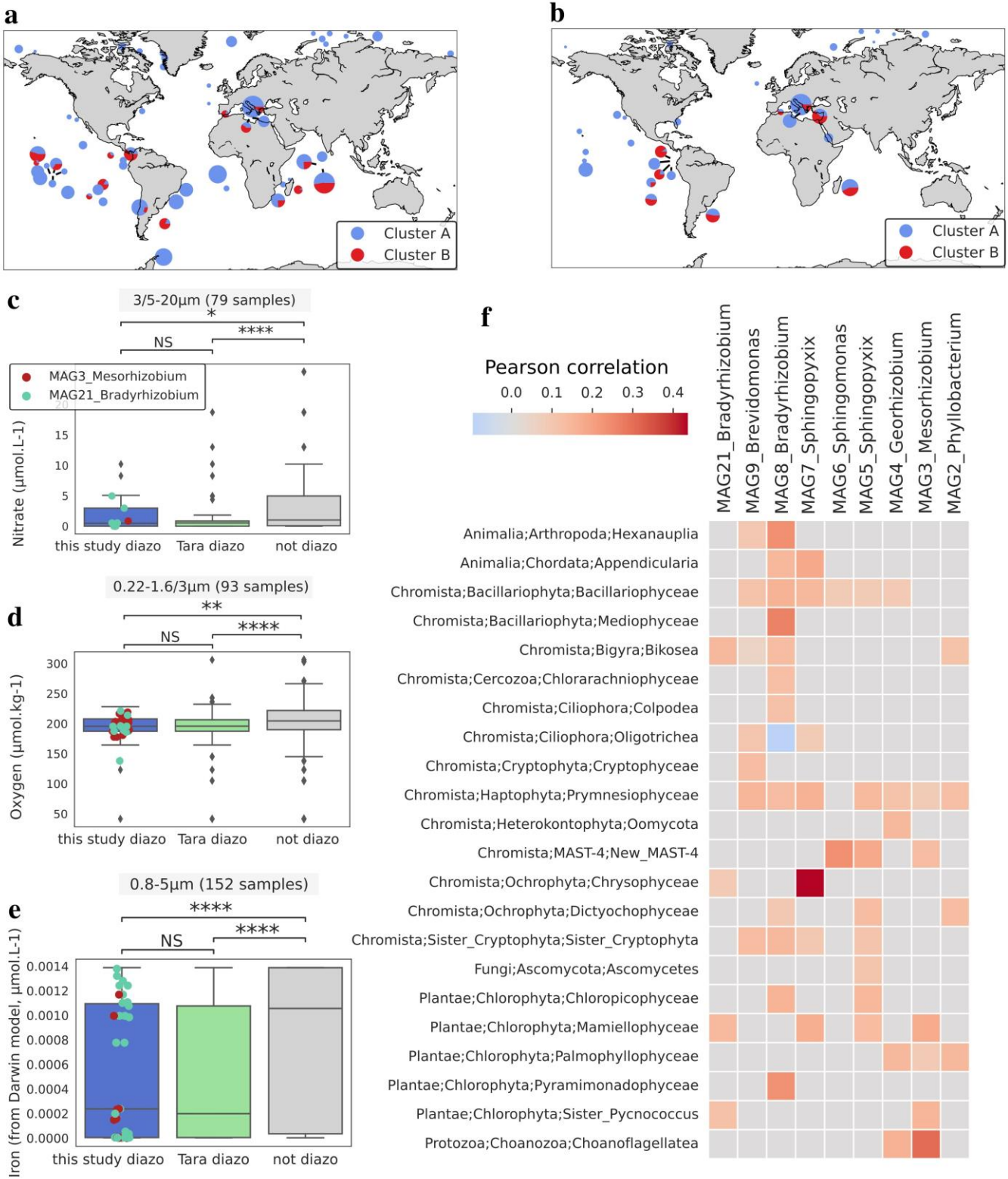

CLUSTER A
CLUSTER B
A
B
MAG9_Brevundimonas
MAG4_Georhizobium
MAG3_Mesorhizobium
MAG2_Phyllobacterium
MAG8_Bradyrhizobium
MAG21_Bradyrhizobium
MAG7_Sphingopyxis
MAG5_Sphingopyxis
MAG6_Sphingomonas
Legend
Pt15 MAGs
TARA diazotrophic genomes
MicroScope genomes
Terrestrial bacteria
Marine bacteria
Symbiotic
Free-living
Symbiotic and Free-living
Symbiotic poor/non-fixer
No diazotrophic info available

a


Relative Abundance (%)
100%
80%
60%
40%
20%
0%
NifD NifH NifX NifT HemN Fpr FixS FixA FixH GroEL NifS FixJ FixX NifU Fdx GroES NtrY ntrY FixB GrxD CysE ntrB nirD BolA Bfd IscA ClpX nirB atpD FixC GrxC NirD
%-N/total
%+N/total

b


Relative Abundance (%)
100%
80%
60%
40%
20%
0%
EcfT DeoB HptA UCP aminopeptidase-like... PutP ECM4 YccC GlpA AcpS GcvP2 Sbm SpoT CopC Serine protease LivK TauB CurA UCP PAS PAS XerD CheY Crp CaiD WcaG AcrA BaeS GloB DdpA AtoC MenH AraC AdhE YqjQ PaaJ LysR EmrA GntR RhaT MenE/FadK GGDEF FabG AcrR CitB AcrA UbiE AraJ RpoE Spo0J FepA GstA GloA MarR RimI SoxR OmpR CaiA ArsR TolC HipB RfaB PinE AmpC OmpA DadA PurR RimL Lrp Aes DAP2 CirA PccA
%-N/total
%+N/total


a
Cluster A
Cluster B
b
Cluster A
Cluster B
c
3/5-20µm (79 samples)
*
****
NS
MAG3_Mesorhizobium
MAG21_Bradyrhizobium
Nitrate (µmol.L-1)
this study diazo
Tara diazo
not diazo
d
0.22-1.6/3µm (93 samples)
**
****
NS
Oxygen (µmol.kg-1)
this study diazo
Tara diazo
not diazo
e
0.8-5µm (152 samples)
****
****
NS
Iron (from Darwin model, µmol.L-1)
this study diazo
Tara diazo
not diazo
f
Pearson correlation
0.0 0.1 0.2 0.3 0.4
MAG21_Bradyrhizobium
MAG9_Brevidomonas
MAG8_Bradyrhizobium
MAG7_Sphingopyxix
MAG6_Sphingomonas
MAG5_Sphingopyxix
MAG4_Georhizobium
MAG3_Mesorhizobium
MAG2_Phyllobacterium
Animalia;Arthropoda;Hexanauplia
Animalia;Chordata;Appendicularia
Chromista;Bacillariophyta;Bacillariophyceae
Chromista;Bacillariophyta;Mediophyceae
Chromista;Bigyra;Bikosea
Chromista;Cercozoa;Chlorarachniophyceae
Chromista;Ciliophora;Colpodea
Chromista;Ciliophora;Oligotrichea
Chromista;Cryptophyta;Cryptophyceae
Chromista;Haptophyta;Prymnesiophyceae
Chromista;Heterokontophyta;Oomycota
Chromista;MAST-4;New_MAST-4
Chromista;Ochrophyta;Chrysophyceae
Chromista;Ochrophyta;Dictyochophyceae
Chromista;Sister_Cryptophyta;Sister_Cryptophyta
Fungi;Ascomycota;Ascomycetes
Plantae;Chlorophyta;Chloropicophyceae
Plantae;Chlorophyta;Mamiellophyceae
Plantae;Chlorophyta;Palmophyllophyceae
Plantae;Chlorophyta;Pyramimonadophyceae
Plantae;Chlorophyta;Sister_Pycnococcus
Protozoa;Choanozoa;Choanoflagellatea