HAL
open science

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

Udita Chandola, Camille Trottier, Marinna Gaudin, Erik Manirakiza, Samuel Menicot, Isabelle Louvet, Thomas Lacour, Timothée Chaumier, Atsuko Tanaka, Samuel Chaffron, et al.

To cite this version:

Udita Chandola, Camille Trottier, Marinna Gaudin, Erik Manirakiza, Samuel Menicot, et al.. Combined in vivo and in situ genome-resolved metagenomics reveals novel symbiotic nitrogen fixing interactions between non-cyanobacterial diazotrophs and microalgae. 2022. hal-03781940

HAL Id: hal-03781940

https://hal.science/hal-03781940v1

Preprint submitted on 16 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Udita Chandola[1], Camille Trottier[1], Marinna Gaudin[2,3], Erik Manirakiza[1], Samuel Menicot[1], Isabelle Louvet[4], Thomas Lacour[5], Timothée Chaumier[1], Atsuko Tanaka[6], 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] Nantes Université, CNRS, CEISAM, UMR 6230, F-44000 Nantes, France.

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

[6] 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 triconrnutum* (*Pt*) of NCDs sustaining diatom cells in the absence of bioavailable nitrogen. We identified *Pt*NCDs using metagenomics sequencing and detected nitrogenase gene in silico and/or by PCR. We demonstrated nitrogen fixation in *Pt*NCDs and their close genetic affiliation with NCDs from the environment. We showed the wide occurrence of this type of symbiosis with the isolation of NCDs from other microalgae and their identification in the environment and in co-occurrence with photosynthetic microalgae. Overall, this study provides evidence for a previously overlooked symbiosis using a multidisciplinary model-based approach which will consequently help understand the different players driving global marine nitrogen fixation.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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]. The energy intensive reduction of N_2 specific to some prokaryotes termed diazotrops is catalyzed by the nitrogenase enzyme complex including the nitrogenase reductase gene containing two identical subunits encoded by the *nifH* gene and the nitrogenase composed of two pairs of different subunits (α and β) of dinitrogenase respectively encoded by the *nifD* and *nifK* genes[3]. Therefore, eukaryotes are only able to obtain fixed nitrogen through their symbiotic 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 that live in the air 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 oceans. However, recent molecular analyses indicated that non-cyanobacterial diazotrophs (NCDs) are also present and active showing an important presence of diverse *nifH* gene amplicons related to non-cyanobacteria, especially to Proteobacteria and Planctomycetes[12-15]. 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 a symbiotic interaction that benefit to both partners. 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 diverse previously unknown marine diazotrophs of mainly α-proteobacteria lineages, supporting the host cells in the absence of any form of usable nitrogen. Further, we isolated, cultured, sequenced and fully assembled most of the

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 and 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* (hereafter *Pt15*) that was able to survive in comparison to the axenic control (Fig. 1a), suggesting that diatom growth is supported by putative bacterial nitrogen fixation providing an usable source of nitrogen to diatom cells. After 2 months of nitrogen starvation, xenic *Pt15* showed similar growth to non-starved cells upon transfer to nitrate replete medium. This indicates that *Pt15* starved cells stayed alive and capable of growth in favorable conditions, evidenced by their exponential growth when transferred to nitrate replete medium (Fig. 1b). To demonstrate nitrogen fixation, we used the widely applied Acetylene reduction assay (ARA)[16] that showed ethylene production, although low in xenic *Pt15* culture starved for nitrogen, but not in plus N condition (+N). No ethylene was detected in axenic *Pt15* confirming that the detected ethylene is not endogenous to the diatom cells but originated from bacterial nitrogen fixation (Supplementary Table 1, Supplementary Fig. 1a,b,c, d,e,f,g,h).

To further investigate nitrogen fixing symbiosis 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. 2a). 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. 2b). Taken together, our results suggest bacterial nitrogen fixation that maintain diatom population cells stable and alive.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Light microscopy showed an important 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 around diatom cells looking like a pellicle of exopolysaccharides exclusively present in starved cells (Fig. 1g,h). Similar structure, although less prominent occurred around some of bacterial cells either as a single bacterium or a group of several bacteria.

To identify species composition of *Pt15* bacterial community in - and +N conditions, bacterial DNA was extracted and shotgun sequenced revealing a total of 153,4 million effective 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 by Kraken2, 73 genera were only detected in -N with no reads in +N (Table S2). Composition of the community microbiome at genus level were examined showing a dominance of *Spyngopixys* 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 symbiosis 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[17-20]. Therefore, different growth conditions (media, temperature) were tested (Supplementary file 1) and permitted the isolation of 13 bacterial species (Table 1). Bacterial colonies were identified using either full length 16S or two hypervariable regions of the 16S (V1V2, V3V4) and the presence of *nifH* was assayed using three combinations of degenerate primers as well as *nifH* specific primers designed from Metagenomics data whenever possible (Supplementary Fig. 3a,b,c). Six bacteria were

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

identified as nitrogen fixers based on a combination of *nifH* PCR amplification and/or *nifH* in silico detection (Table 1). Comparative metagenomics analysis identified two *Sphyngopyxis* and *Bradyrhizobium* species which were distinguished using Meta Assembled Genome (MAGs) specific primers designed for each species. 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 either the MetaG data or sequences from a pool of 3 isolated bacteria. Two and four copies of *nifH* were detected in *Bradyrhizobium* MAG21 and *Mesorhizobium* MAG3 respectively. Among the 11 isolated bacteria, two species *Kocuria sp*., and *Methylobacterium* were not detected in Metagenomics sequences likely due to their low abundance emphasizing the importance of combining sequencing and isolation/culturing methodologies (Table 1). Evidences for nitrogen fixation were reported in the literature for these two genera[21,22]. To further evidence nitrogen fixing ability of the isolated bacteria, we used nitrogen-free medium with bromothymol blue, a pH sensitive dye to grow the two most abundant bacteria, *Mesorhizobium* and *Bradyrhizobium* MAG8 and found a switch in the color from green (pH=6.7) to blue (pH > 7) indicating nitrogen fixation and the release of NH3 which alkalinizes the medium (Figure 1i).

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 higher than 97% (Table 1, Supplementary Table 3). To further improve MAG assembly for some of the isolated bacteria, besides Metagenomics data initially used, MAGs were assembled with additional Illumina and Pacific Pac Bio sequencing of the isolated bacteria including *Bradyrhizobium*, *Georhizobium*, *Sphyngopyxis*, and *Kocuria* which were sequenced either alone or in a pool of 3 bacteria. Overall, chromosomal sequences were well assembled with higher completeness as compared to Metagenomic assembled MAGs and nearly no cross contamination of bins. To achieve a better identification of the bacterial genetic material, we sought to recover from non-assembled reads, plasmids 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). The nitrogenase gene (*nifH*) was in silico detected in six out of eight *Pt15* associated MAGs among which, one plasmid located *nifH* (Table 1). Besides *nifH*, several additional *nif* genes were detected in most of the MAGs, among

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which two were found in plasmids, *nifU* and *nifS* (Table 1). Phylogenomic analysis of *Pt15* MAGs, MAGs of known terrestrial and marine diazotrophs from the Microscope database[23] and MAGs from *Tara* Oceans data using Anvi'o revealed two clusters, A (*Mesorhizobium*, *Bradyrhizobium* MAGs, *Georhizobium*, *Phylobacterium*, and *Bevundimonas*) and B (2 *Sphyngopyxis* 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 are 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[24-26]. 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 BTB (iii) detection of *nifH*, and (iv) the isolation from the xenic *Pt15* of NCDs clustering with known proteobacteria diazotrophs, indicated that a symbiotic nitrogen fixation likely occurred in *Pt15* xenic culture supporting the survival of diatom cells population 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 both conditions 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[27,28]. 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, indicating an important role in energy/substrates shuttling between the host and the bacterial community. Examples included *EcfT* gene (energy-coupling factor transporter transmembrane) known for its function in providing the energy necessary to transport a number of different substrates[29], AcpS, an acyl carrier gene which functions as a multi-carbon units carrier into different biosynthetic

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pathways[30], 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 carbone, nitrogen and energy[31]. Stress responsive genes were clearly more represented in -N compared to +N including SpoT, which is a general bacterial stress response[32] 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)[33], 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 with *nifH* including 2 species of *Thalassiospira sp*. (isolated from two different microalgae, *Stenotrophomonas maltophilia*, *Stappia sp.)*, *Kocuria* and *Rhodococcus qingshengii* suggesting that the occurrence of NCDs in a symbiotic interaction with microalgae is frequent (Supplementary Fig. 3a,b, Supplementary Table 5). 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[34 35 36](Supplementary Fig. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Next, we assessed the occurrence of *Pt15* MAGs in the environment in surface water and Deep Chlorophyll Maximum layer (DCM) using *Tara* Ocean expedition datasets[37,38]. Among the 130 sampling stations, 92 stations show the occurrence of at least one of the *Pt15* MAGs reflecting a high occurrence frequency (Fig. 5a, Supplementary fig. 5). *Pt15* closely related TARA MAGs were more occurring with 129 recorded stations. The distribution of *Pt15* MAGs was very large 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. 5). Interestingly, *Pt15* MAGs co-occurred in surface and DCM samples in most cases, although their dominance in surface samples was prominent. Except from few stations, clusters A and B seemed to co-occur suggesting probable interactions between their respective bacteria.

We then asked whether the occurrence of the MAGs isolated from *Pt15* correlate with environmental factors monitored during the expeditions and whether they co-occur with other organisms. We considered only the top 2 MAGs of *Pt15 (Mesorhizobium* MAG3 and *Bradyrhizobium* MAG8*)* to assess their correlation with 4 environmental factors used as indicators of diazotrophy, nitrate, nitrate/phosphate ratio, iron, and oxygen. Interestingly, *Pt15* MAGs were detected in low nitrate and almost no detection above 5 µM. Similarly, the detection of *Pt15* MAGs coincided with low nitrate to phosphate ratio (Fig. 5b,c,d,e). Both measurements are in line with nitrate limitation and phosphate abundance as markers of diazotrophs presence[39,40]. Iron availability seems to affect differently the abundance of both MAGs. While, *Mesorhizobium* was observed in low to no detectable levels of iron, *Bradyrhizobium* MAG8 was detected in higher iron concentrations suggesting different requirement for iron in these two species. By contrast to what is known about oxygen inhibition of nitrogenase, both MAGs seem 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 (Supplementary Fig. 5)

Strikingly, a co-occurrence analysis revealed a clear association of the NCDs isolated from *Pt15* (Fig 5f) with essentially chromista including diatoms, haptophytes, and chlorophytes. *Bradyrhizobium* MAG8 showed the highest co-occurrence followed by *Brevundimonas* MAG9. The most shared co-occurrence is with haptophytes, Bikosea and Pycnococcus. Although, nitrogen fixation is not demonstrated in these identified associations, this analysis comforted a likely symbiotic interaction for nitrogen fixation between investigated NCDs and different species of mainly photosynthetic micro-eukaryotes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Recent reports of NCDs dominance in the euphotic zones of the ocean 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[12-14,41]. 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 diazotrops and attachment/interaction with microalgae or debris which were found to be hotspots of nitrogen fixation[42]. Using a multidisciplinary approach, the study described herein brings unprecedented evidence of a symbiotic interaction for nitrogen fixation between microalgae in particular the model diatom *Phaeodactylum tricornutum* and NCDs. Our study revealed among several screened xenic accessions of *P. tricornutum* in the absence of usable form of nitrogen, one accession, *Pt15* interacting 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 bloom 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 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[43]. Demonstrated by ARA, the rates of nitrogen fixation in xenic *Pt15* in -N is low compared to reports of diazotrophs of cyanobacteria type but similar to nitrogen fixation rates known in proteobacteria[44].

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Our study reveals 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 clustering with previously identified terrestrial and marine NCDs supporting diazotrophy and their ecological relevance. Our MAGs were identified in surface and DCM samples in microalgae containing size fractions, which strongly supported 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 biofilm 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[45]. Besides biofilm formation, this symbiotic 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[10,46] or hyper-respiration[47] among other protective mechanisms that need further investigations. In line with our discovery of NCDs interaction with microalgae for nitrogen fixation, a recent study[48] 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 described 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 nearly complete assembly of several MAGs with an exceptional 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 whole 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, do certainly 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 eighteen microalgae were screened in nitrate-deplete and nitrate-replete enhanced artificial sea water for growth (Supplementary Table 7). 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)[16] 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 nitrogen free medium (NFM)[49,50] 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 to 11,000 rpm for 10 minutes at RT to remove the medium. The cells were re-suspended in nitrogen free

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

medium with five sugars and grown at 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 were 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 a gas chromatograph 7820 model Agilent equipped with a Flame Ionization Detector (GC-FID) and using Hydrogen as carrier gas at a constant flow of 7.7 mL/min. Sample 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 at 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 'x' value represents amount of ethylene (nmoles/L/hour).

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. Seven time points (day(s)) used for harvesting samples: 1, 4, 12, 29, 59, 76 and 80. The cells were counted at each time point using 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 papers to collect *Pt15* cells and the filters were dried overnight at 60°C incubator and stored at -80°C. The 20 ml filtrate from the cultures was passed through 0.22 um 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 uM nitrate-EASW, 0 uM nitrate-EASW and three heat treated Whatman GF/F glass microfibre filter were used as blanks for comparison with test samples.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

705 color (Carl Zeiss, Germany).

Transmission electron microscopy

Samples were prepared as described previously [51]. Embedded samples were cut for making pale yellow ultra-thin sections (70-80nm thickness) with a microtome EM UC6 (Leica Microsystems, Germany). Sections were counter- 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 recipe 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.

Determination of nitrogen fixing ability using NfB medium

Bacterial isolates were incubated in liquid nitrogen free medium with bromothymol blue (NfB)[52] at 30°C for 16 days and monitored for color change every day. Two positive controls, a slow and fast growing strains, respectively, *Pseudomonas stutzeri* and *Azotobacter vinelandii* were used. Isolates free NfB medium was used as a negative control. The NfB medium contains a pH sensitive dye, bromothymol blue which switches from green (pH= 6-6.7) to blue (pH>6.7)[53]. The switch in the color is due to nitrogen fixation which consumes H+ protons alkalinizing the medium.

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 ul 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 20mg/ml lyzozyme (VWR cat no. 0663-5G), incubated at 37°C for 60 min in order to lyze possible gram-

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 ul (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.

Three degenerate primers (F1 – nifH3R, nifH2F – nifh3R, PolF – PolR)[54] for *nifH* gene detection by PCR were standardized. The PCR was done as follows, 95°C for 5 minutes (1 cycle), followed by 40 cycles of 95°C at 30 seconds, standardized annealing temperature at 30 seconds, elongation at 72°C for 30 seconds and a final extension at 72°C for 7 minutes. The annealing temperature used for F1 – nifH3R, nifH2F – nifh3R, PolF – PolR combinations were 48, 52 and 54°C respectively. Fifty ng of DNA was added to a 20 µl PCR reaction containing, 10X Dream Taq buffer, 0.2 mM dNTPs, 2 pmol primers and 1.25 U of *Taq* (Fischer Scientific ca. n° 15689374). The reaction was performed with 35 cycles of 95 °C (1 min), 55 °C (1 min) and 72 °C (1 min). For species identification, 16S PCR fd1-rd1, variable regions PCR v1v2 v3v4 and v5v7 were used. All primer sequence details are reported in supplementary Table 6.

Sample preparation and DNA extraction for Illumina and Pacific Pac Bio 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 log 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 log no. GSWP04700) and resuspended in 1X

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 are available, they are 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, 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[55]. (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. Samples from 2 experiments were used. An experiment with sequences obtained at 2g/L salinity (one sample in +N condition / one sample in -N condition) and the other experiment at 20 g/L salinity includes duplicates for each of the conditions. Both data sets are similar.

Single assembly strategy

DNA raw reads were analyzed using a dedicated metagenomic pipeline ATLAS v2.7[55]. This tools 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, including mitochondria and chloroplast sequences, was used for this specific step. Quality-control (QC) reads assemblies were performed using metaSPADES[56] 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[57] and metabat2[58]. The bins obtained by the different binning tools were combined using DASTool[59] and dereplicated with dRep[60] using default parameters. Completeness and

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

contamination were computed using CheckM[61]. 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[62]. Linclust[63] 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[59]. Briefly, QC reads are pulled together and addressed to the chosen assembler MetaSPADES 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[61]. 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.

Taxanomic assignation of metagenomic data

We utilized Anvio version 7.2 (https://anvio.org) for taxanomic profiling of the *P. tricornutum* associated metagenomes in nitrate deplete and replete condition. The workflow uses The Genome Taxonomy Database (GTDB, doi:10.1038/nbt.4229) to determine the taxanomy based on 22 single copy core genes (scgs, anvio/anvio/data/misc/SCG_TAXONOMY). Sequence alignment search is done using DIAMOND (doi:10.1038/nmeth.3176).

Hybrid assembly of MAGs

Assembly

Metagenomic assembly of Illumina technology reads does not yield complete genomes due to the difficulty of assembling repetitive regions[64]. Whenever Illumina and Pac bio reads are available, we used a strategy to improve MAGs assembly by combining both type

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of reads. Metagenomic assembly of Illumina reads paired end and PacBio reads post quality trimming, were assembled using metaSPAdes 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 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 (https://anaconda.org/bioconda/maxbin2 and https://anaconda.org/bioconda/metabat2)[65].

Refinement

Bins from metabat2 and maxbin2, 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 [66].

Taxonomic Assignation

GTDB-Tk has already been independently and positively evaluated for the classification of MAGs[67]. 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 [68].

Gene annotation of metagenomic data and MAGs

Gene annotation of metagenomics data in two conditions was done using Anvio version 7.2. Three databases were used for a robust gene observations: COG20 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383993,http://www.ncbi.nlm.nih.gov/COG/), KEGG and PROKKA (https://github.com/tseemann/prokka). DIAMOND to search NCBI's Database. In built, Anvio 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 Anvio platform for the individual metagenome contigs databases.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phylogenomic analysis of MAGs

The phylogenomic tree was built using Anvio-7.1 (https://github.com/merenlab/anvio/releases) 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 7) to the individual MAGs were selected using MicroScope database (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, neighbour-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'[62] 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 for sequence alignment (http://www.drive5.com/muscle). 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[69] (Sequence Contents-Aware Plasmid Peeler), https://github.com/Shamir-Lab/SCAPP, metaplasmidSPAdes[70] (GitHub - ablab/spades: SPAdes Genome Assembler) and PlaScope[71] (GitHub - labgem/PlaScope: Plasmid exploration of bacterial genomes). Predicted plasmids were annotated using Dfast[72] (https://github.com/nigyta/dfast_core), eggNOG-mapper v2[73] and PROKKA 1.14.6.

Downloading of metagenomes assembled-genomes counts tables

MAGs abundance count tables and fasta files were downloaded from open-source released Foundation Tara Oceans database at <https://www.genoscope.cns.fr/tara/> [13,74]. Data include 2601 MAGs of which 1888 prokaryotes and 713 eukaryotes MAGs in 937 samples.

Biogeography of metagenomes assembled-genomes

MAGs abundance count tables and *.fasta* files were downloaded from Tara Oceans databases available at <https://www.genoscope.cns.fr/tara/>[13,74]. This resource included 2601 MAGs, including 1888 prokaryotic and 713 eukaryotic MAGs profiled across 937 samples. Total sum scaling (TSS) normalisation was performed on reads counts mapped to MAGs, by dividing all

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

counts by the total number of reads in a given sample. Different depths (surface and deep chlorophyll maximum) and size fractions (0.22-1.6/3µm, 3/5-20µm, 0.8-5µm, 20-180/200µm, 0.8-2000µm, 180-2000µm) were analysed separately to project relative abundances on global maps. Relative abundances of TARA MAGs phylogenetically close to *Pt15* and belonging to the same cluster were summed up for each sample. The 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[37]. MAGs relative abundances (after TSS normalisation) were projected on global maps using R.

Co-occurrence network inference

Co-occurrence network inference was performed using FlashWeave v0.18.0[75] 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[76] was applied to replace zeros, and a Centered Log Ratio transformation[77] was applied separately on both prokaryotic and eukaryotic abundance matrices. FlashWeave parameters used were max_k=3, n_obs_min=10, and normalize=False as we have applied our own standardisation (CLR).

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. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

17 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

18 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

19 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).

20 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

21 Ouyabe, M., Kikuno, H.,Tanaka, N., Babil, P. and Shiwachi, H. Isolation and identification of nitrogen-fixing bacteria

associated with Dioscorea alata L. and

Dioscorea esculenta L. *Microb. Resour. Syst.* **35**, 3-11 (2019).

22 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

23 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

24 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

25 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

26 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

27 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

28 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

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

30 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

31 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

32 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

33 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

34 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

35 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).

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

37 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

38 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

39 Sanudo-Wilhelmy, S. A. *et al.* Phosphorus limitation of nitrogen fixation by Trichodesmium in the central Atlantic Ocean. *Nature* **411**, 66-69 (2001). https://doi.org:10.1038/35075041

40 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).

41 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

42 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

43 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

44 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

45 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

46 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

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

48 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

49 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).

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

51 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).

52 Baldani, J. I., Reis, V.M., Videira, S.S. et al. The art of isolating nitrogen-fixing bacteria from non-leguminous plants using N-free semi-solid media: a practical guide for microbiologists. *Plant Soil* **384**, 413–431

53 Jha, C. K., Patel, D., Rajendran, N. & Saraf, M. Combinatorial assessment on dominance and informative diversity of PGPR from rhizosphere of Jatropha curcas L. *J Basic Microbiol* **50**, 211-217 (2010). https://doi.org:10.1002/jobm.200900272

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

54 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

55 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

56 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

57 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

58 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

59 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

60 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

61 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

62 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

63 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

64 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).

65 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

66 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

67 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

68 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

69 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

70 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

71 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

72 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

73 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

74 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).

75 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

76 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).

77 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 and BTB assay. 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the membrane. c, chloroplast, v, vacuole, n, nucleus. Scale bars, 10 µm (c-f) and 2 µm (g, h). i, color change from green (negative control, pH = 6.7) to blue (control and assayed strains, pH > 6.7) due to pH change from acidic to alkaline induced by nitrogen fixation and release of NH3+ monitored 16 days post-inoculation. From left to right, *Azotobacter vinalendii* (NCIMB 12096), *Pseudomonas stutzeri* (NCIMB885), NfB medium without inoculum, *Mesorhizobium* MAG3, *Bradyrhizobium* MAG8. PC, Positive Control, PC, Negative Control.

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.

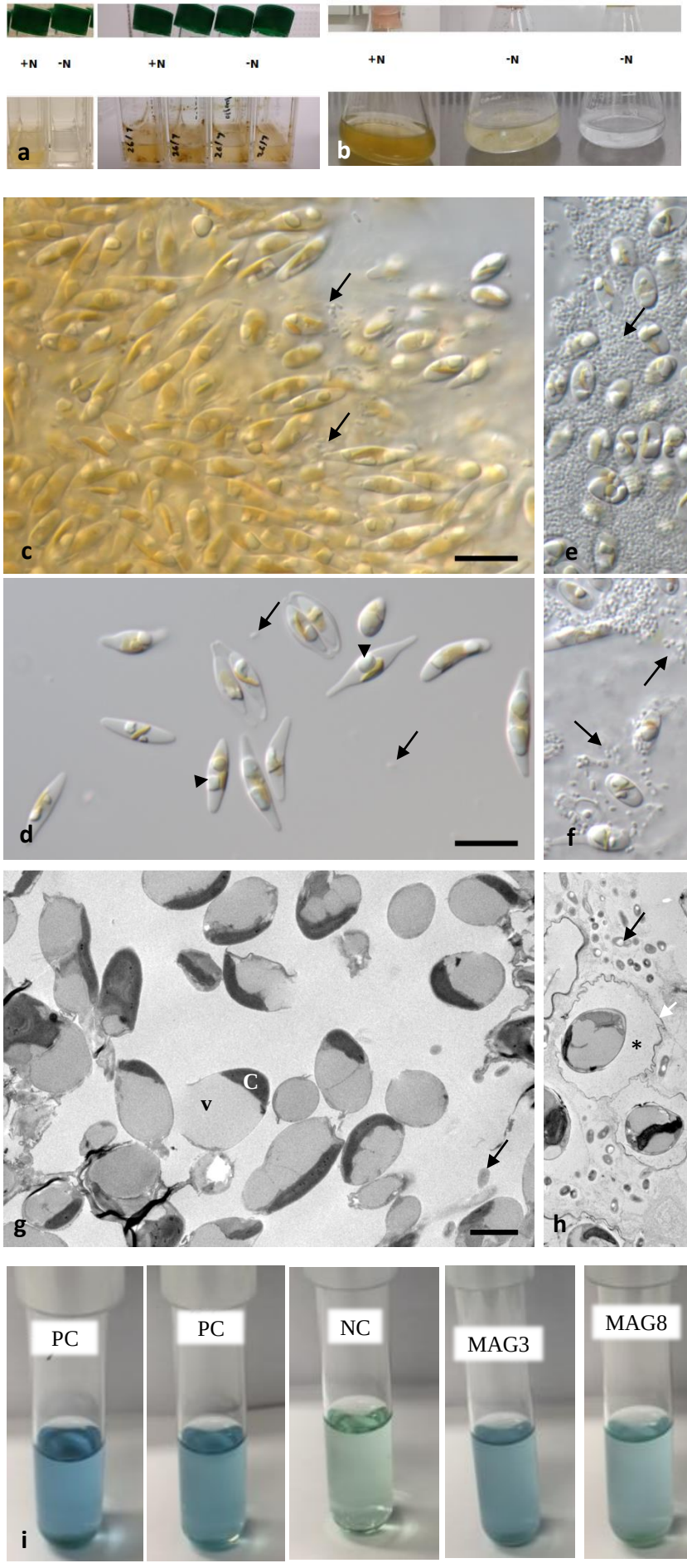
Fig. 3. Phylogenetic tree of *Pt15* assembled bacterial genomes amongst 1888 TARA bacterial MAGs and MAGs from microscope.

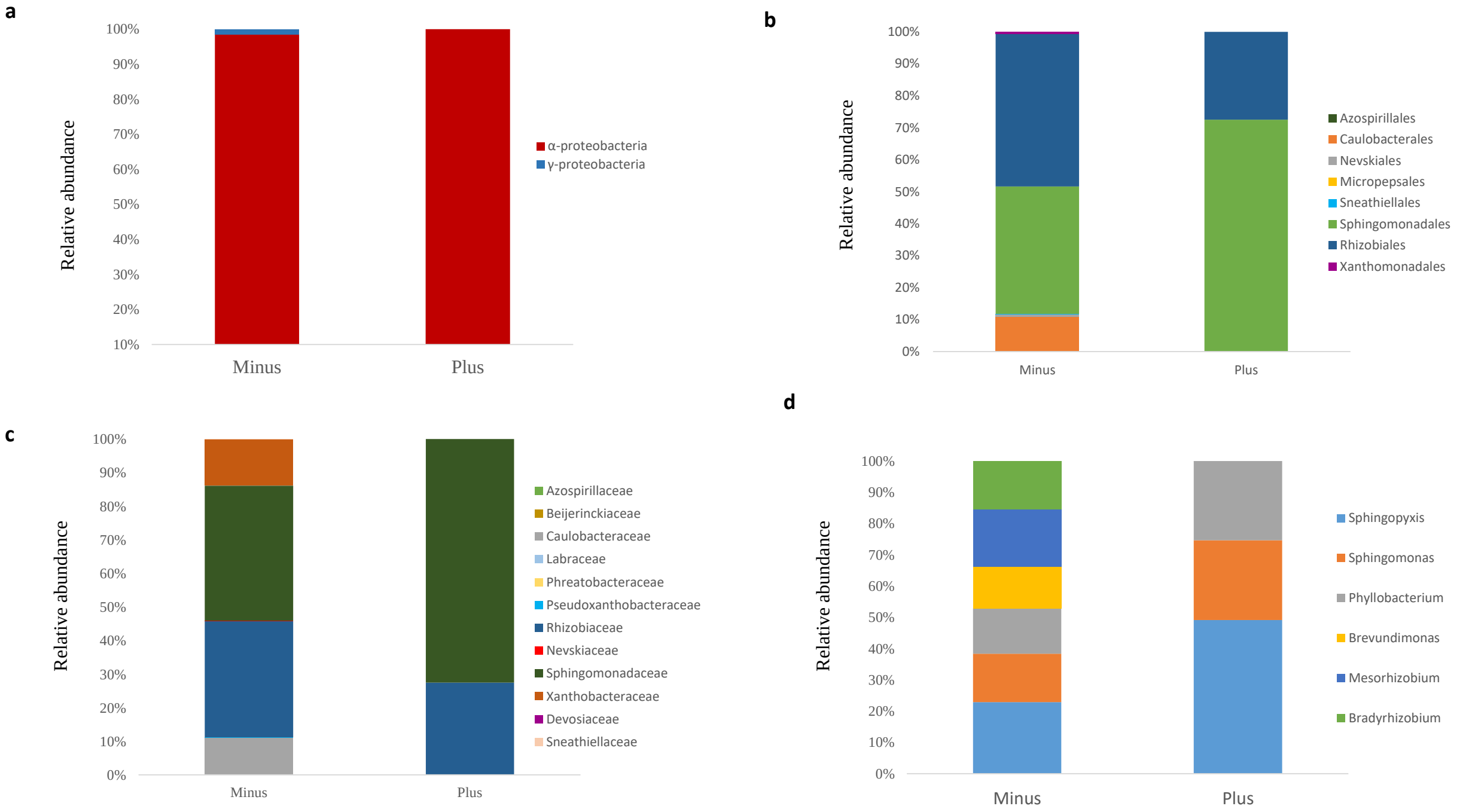
Fig. 4. Functional gene annotation of *Pt15* associated NCDs in nitrate deplete versus replete conditions. a, Relative abundance of selected genes involved in nitrogen fixation and assimilation in nitrate deplete and nitrate replete condition. b**,** illustration representing a combined view of 50 top abundant gene in nitrate deplete and replete condition with ordering representing top relatively abundant genes in nitrate deplete condition. 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. 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 concentrations and nitrate phosphate ratios. g, Co-occurrence of *Pt15* MAGs in the environment with different species of micro-eukaryotes.

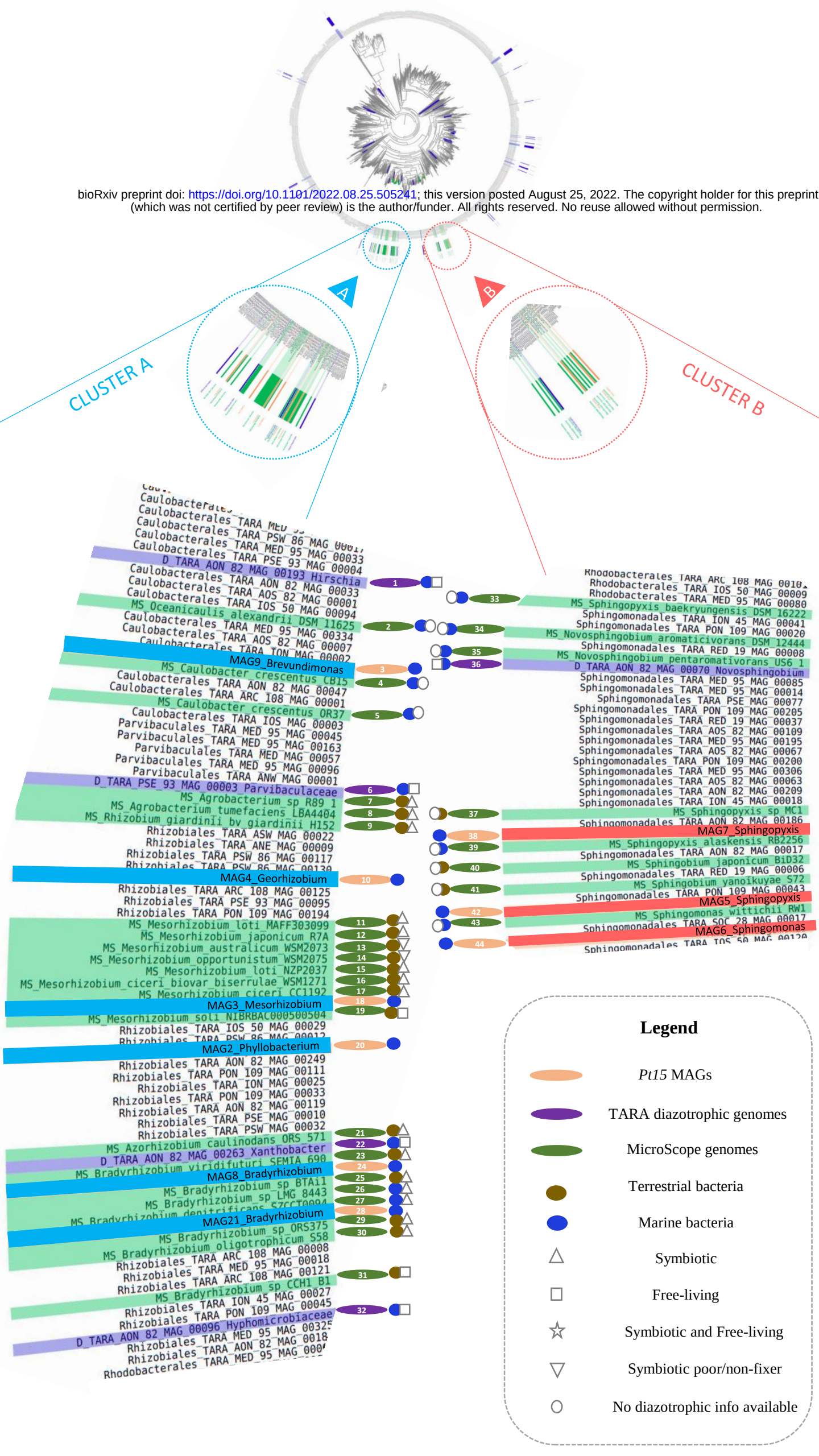
Table 1. List of isolated bacteria from xenic *Pt15* in -N condition.

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

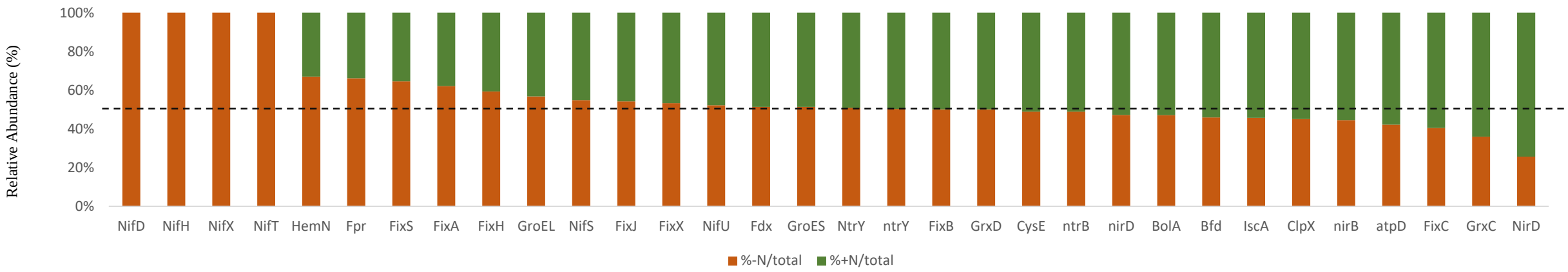

+N
-N
+N
-N
+N
-N
-N
a
b
c
e
d
f
C
v
g
v
n
c
h
PC
PC
NC
MAG3
MAG8
i


a
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
Minus
Plus
α-proteobacteria
γ-proteobacteria
b
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Minus
Plus
Azospirillales
Caulobacterales
Nevskiales
Micropepsales
Sneathiellales
Sphingomonadales
Rhizobiales
Xanthomonadales
c
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Minus
Plus
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%
Minus
Plus
Sphingopyxis
Sphingomonas
Phyllobacterium
Brevundimonas
Mesorhizobium
Bradyrhizobium

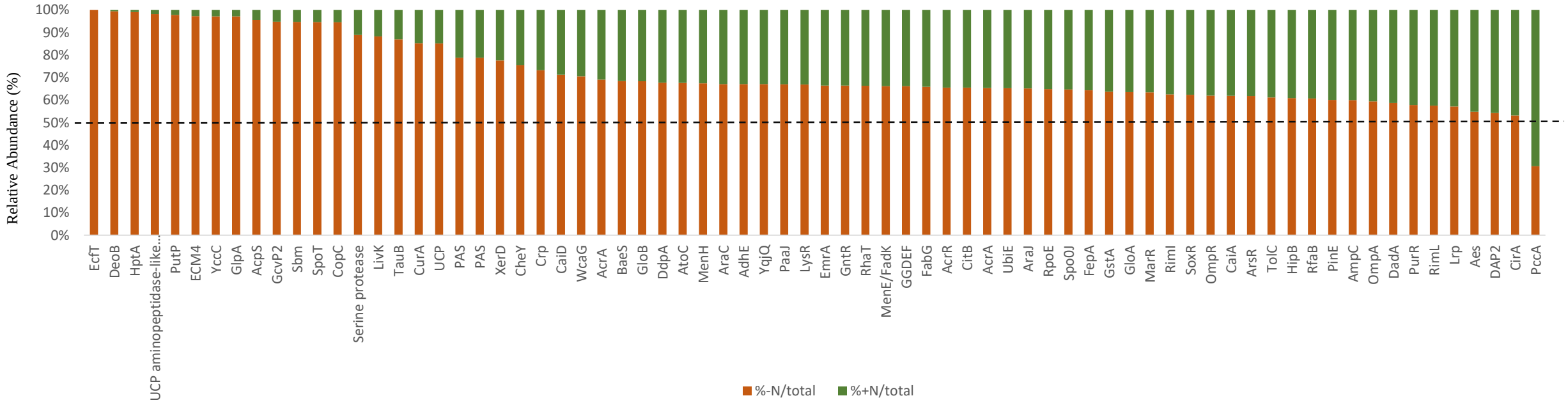
bioRxiv preprint doi: https://doi.org/10.1101/2022.08.25.505241; this version posted August 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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%
90%
80%
70%
60%
50%
40%
30%
20%
10%
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 SpoOJ 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

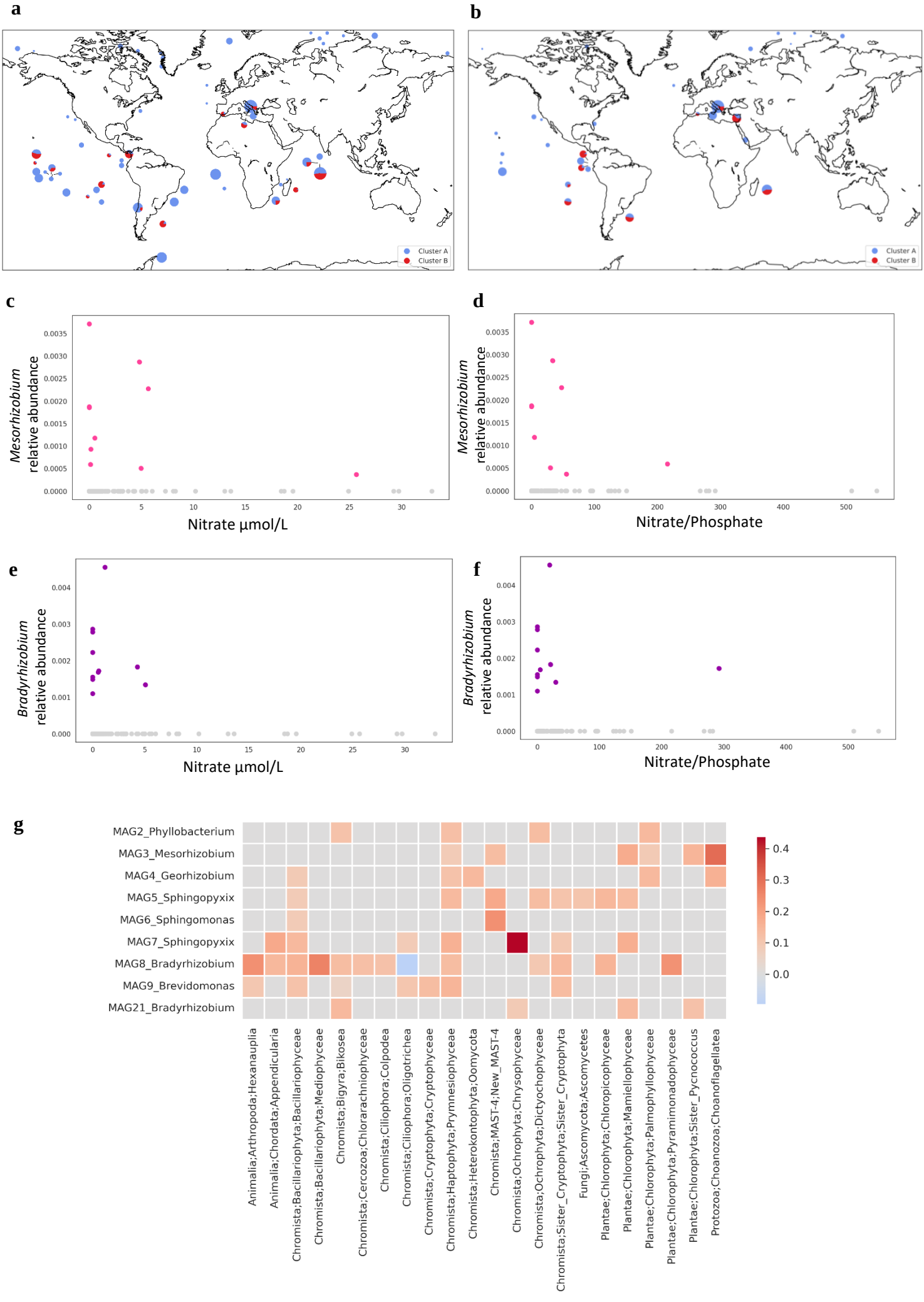

a
Cluster A
Cluster B
b
Cluster A
Cluster B
c
Mesorhizobium relative abundance
Nitrate µmol/L
d
Mesorhizobium relative abundance
Nitrate/Phosphate
e
Bradyrhizobium relative abundance
Nitrate µmol/L
f
Bradyrhizobium relative abundance
Nitrate/Phosphate
g
MAG2_Phyllobacterium
MAG3_Mesorhizobium
MAG4_Georhizobium
MAG5_Sphingopyxix
MAG6_Sphingomonas
MAG7_Sphingopyxix
MAG8_Bradyrhizobium
MAG9_Brevidomonas
MAG21_Bradyrhizobium
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

Table_1

Genus	MAG No.	Size (Mbp)	No. of contigs	*nifH* PCR identification	*nifH in silico* identification	Predicted plasmids
Mesorhizobium	**MAG3**	6.85	75	No	Yes	Yes
Bradyrhizobium sp1	**MAG 8**	7.40	109	Yes	Yes	Yes
Bradyrhizobium sp2	**MAG 21**	8.42	5	Yes	Yes	Yes
Georhizobium	**MAG13**	4.4	5	No	Yes	Yes*
Sphingopyxis sp1	**MAG5**	4.00	52	Yes	Yes	Yes**
Sphingopyxis sp2	MAG 7	5.00	23	ND	No	Yes
Sphingomonas	**MAG6**	3.31	17	Yes	Yes	Yes
Kocuria	MAG19	3.1	6	Yes	No	ND
Rhizobium sp. strain	NA	ND	ND	No	No	ND
Methylobacterium	NA	ND	ND	No	No	ND
Arthrobacter	MAG14	3.46	12	No	No	ND
Pseudoarthrobacter	NA	ND	ND	No	No	ND
Microccocacea	NA	ND	ND	No	No	ND

**nifH* identified
** other nif genes