



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

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

3

4 Uditā Chandola¹, Camille Trottier¹, Marinna Gaudin^{2,3}, Erik Manirakiza¹, Samuel Menicot¹,
5 Isabelle Louvet⁴, Thomas Lacour⁵, Timothée Chaumier¹, Atsuko Tanaka⁶, Samuel Chaffron^{2,3}
6 and Leila Tirichine^{1*}

7

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

9

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

11

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

14

15 ⁴ Nantes Université, CNRS, CEISAM, UMR 6230, F-44000 Nantes, France.

16

17 ⁵ Ifremer, PHYTOX, PHYSALG, Rue de l'Île d'Yeu, BP21105, Nantes Cedex 03, 44311,
18 France

19

20 ⁶ Department of Chemistry, Biology and Marine Science, Faculty of Science, University of the
21 Ryukyus, Okinawa 903-0213, Japan

22

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

24

25 **Abstract**

26 Non-cyanobacteria diazotrophs (NCDs) were shown to dominate in surface waters shifting the
27 long-held paradigm of cyanobacteria dominance and raising fundamental questions on how
28 these putative heterotrophic bacteria thrive in sunlit oceans. Here, we report an unprecedented
29 finding in the widely used model diatom *Phaeodactylum tricornutum* (*Pt*) of NCDs sustaining
30 diatom cells in the absence of bioavailable nitrogen. We identified *Pt*NCDs using
31 metagenomics sequencing and detected nitrogenase gene in silico and/or by PCR. We
32 demonstrated nitrogen fixation in *Pt*NCDs and their close genetic affiliation with NCDs from
33 the environment. We showed the wide occurrence of this type of symbiosis with the isolation
34 of NCDs from other microalgae and their identification in the environment and in co-occurrence
35 with photosynthetic microalgae. Overall, this study provides evidence for a previously
36 overlooked symbiosis using a multidisciplinary model-based approach which will consequently
37 help understand the different players driving global marine nitrogen fixation.

38

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

41

42 Nearly 80% of Earth's atmosphere is in the form of molecular nitrogen (N₂). Similarly,
43 the main form of nitrogen in the oceans which cover 71% of Earth surface is dissolved
44 dinitrogen. The rest is reactive nitrogen in the form of nitrate, ammonia and dissolved organic
45 compounds which are very scarce rendering nitrogen one of the main limiting factors of
46 productivity in most areas of the oceans¹. Despite its dominance, ironically, N₂ is unavailable
47 for use by most organisms because there is a strong triple bond between the two nitrogen atoms
48 that requires high energy input to split the molecule and supply bio-accessible nitrogen². The
49 energy intensive reduction of N₂ specific to some prokaryotes termed diazotrophs is catalyzed
50 by the nitrogenase enzyme complex including the nitrogenase reductase gene containing two
51 identical subunits encoded by the *nifH* gene and the nitrogenase composed of two pairs of
52 different subunits (α and β) of dinitrogenase respectively encoded by the *nifD* and *nifK* genes³.
53 Therefore, eukaryotes are only able to obtain fixed nitrogen through their symbiotic interactions
54 with diazotrophs⁴.

55 In aquatic habitats, symbiosis between nitrogen fixing cyanobacteria and various
56 eukaryotes, including several diatom genera such as *Chaetoceros*, *Climacodium* and
57 *Hemiaulus*, are common and important in oligotrophic habitats of the ocean⁵. Because the
58 nitrogenase complex is sensitive to oxygen, diazotrophs that live in the air evolved protective
59 nitrogen fixation alternatives that involve conditionally, temporally, or spatially separating
60 oxidative phosphorylation or photosynthesis from nitrogen fixation⁶⁻¹¹.

61 Cyanobacteria were considered for a very long time as the main diazotrophs in the oceans.
62 However, recent molecular analyses indicated that non-cyanobacterial diazotrophs (NCDs) are
63 also present and active showing an important presence of diverse *nifH* gene amplicons related
64 to non-cyanobacteria, especially to Proteobacteria and Planctomycetes¹²⁻¹⁵. The overwhelming
65 dominance of NCDs raises important questions on how these presumably heterotrophs
66 proteobacteria thrive in the photic zones of the oceans. One emerging hypothesis is their
67 association with microalgae in a symbiotic interaction that benefit to both partners. In our work,
68 we provide first evidence of occurrence of such interactions between NCDs and several
69 microalgae. More importantly, we identified this interaction in an attractive biological model,
70 the diatom *Phaeodactylum tricornerutum* (*Pt*), opening important and unique opportunities to
71 investigate the molecular mechanisms of diazotrophy. We sequenced the bacterial meta-
72 community associated with *Pt* and identified diverse previously unknown marine diazotrophs
73 of mainly α -proteobacteria lineages, supporting the host cells in the absence of any form of
74 usable nitrogen. Further, we isolated, cultured, sequenced and fully assembled most of the

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

83

84 **Results**

85 **Identification of non-cyanobacterial diazotrophs symbiosis with diatoms**

86 A screen of 10 xenic accessions of *Pt* in the absence of any source of usable nitrogen in
87 the growth medium, identified one accession, *P. tricornutum* 15 (hereafter *Pt15*) that was able
88 to survive in comparison to the axenic control (Fig. 1a), suggesting that diatom growth is
89 supported by putative bacterial nitrogen fixation providing an usable source of nitrogen to
90 diatom cells. After 2 months of nitrogen starvation, xenic *Pt15* showed similar growth to non-
91 starved cells upon transfer to nitrate replete medium. This indicates that *Pt15* starved cells
92 stayed alive and capable of growth in favorable conditions, evidenced by their exponential
93 growth when transferred to nitrate replete medium (Fig. 1b). To demonstrate nitrogen fixation,
94 we used the widely applied Acetylene reduction assay (ARA)¹⁶ that showed ethylene
95 production, although low in xenic *Pt15* culture starved for nitrogen, but not in plus N condition
96 (+N). No ethylene was detected in axenic *Pt15* confirming that the detected ethylene is not
97 endogenous to the diatom cells but originated from bacterial nitrogen fixation (Supplementary
98 Table 1, Supplementary Fig. 1a,b,c, d,e,f,g,h).

99 To further investigate nitrogen fixing symbiosis in minus N condition (-N), we measured
100 Particulate Organic Nitrogen quota (PON) in diatom cells and showed a stable level of PON
101 even after 80 days of growth, supporting the maintenance of diatom cells with infinite amounts
102 of usable nitrogen likely provided by *Pt15* associated bacteria (Supplementary Fig. 2a).
103 Simultaneously, measured biomass of *Pt15*, indicated a stable cell population suggesting a
104 ground state photosynthetic activity that might support the energy demand of nitrogen fixation
105 (Supplementary Fig. 2b). Taken together, our results suggest bacterial nitrogen fixation that
106 maintain diatom population cells stable and alive.

107 Light microscopy showed an important accumulation of bacteria in -N compared to +N
108 condition (Fig. 1c,d,e,f). Transmission electron microscopy confirmed this observation and
109 revealed a peculiar structure around diatom cells looking like a pellicle of exopolysaccharides
110 exclusively present in starved cells (Fig. 1g,h). Similar structure, although less prominent
111 occurred around some of bacterial cells either as a single bacterium or a group of several
112 bacteria.

113 To identify species composition of *Pt15* bacterial community in - and +N conditions,
114 bacterial DNA was extracted and shotgun sequenced revealing a total of 153,4 million effective
115 reads. Using the GTDB reference database, we identified a total of 2 classes, 8 families, 12
116 orders and 90 genera. At class level, α -proteobacteria represented $98\% \pm 0,8$ in +N and $99\% \pm$
117 $0,5$ in -N, based on relative abundance estimations, while *Gammaproteobacteria* were detected
118 with a relative abundance below 1% (Figure 2a). At family level, *Sphingomonadaceae* and
119 *Rhizobiaceae*, both known to contain diazotrophic species dominated (Figure 2c). At genus
120 level, among the 90 distinct taxa identified by Kraken2, 73 genera were only detected in -N
121 with no reads in +N (Table S2). Composition of the community microbiome at genus level were
122 examined showing a dominance of *Spyngopixys* in +N, while -N showed a significant
123 enrichment in *Mesorhizobium sp.*, *Bradyrhizobium sp.*, *Sphingomonas sp.*, and *Brevundimonas*
124 *sp.*, all known as nitrogen fixing taxa (Fig. 2d). Compared to +N, *Sphingopixys* richness
125 decreased in -N although its dominance persisted in this condition. Interestingly, several
126 presumably non-nitrogen fixing species (67 out of 73) were significantly enriched in -N
127 (Supplementary Table 2), suggesting a role of these species in direct or indirect interactions
128 within the bacterial community and/or with the host.

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

130 Given the unique nature of the identified symbiosis involving a model diatom species and
131 NCDs, it was important to isolate the diazotrophic bacteria identified in silico from
132 metagenomics (MetaG) data in order to further explore this interaction. We used universal
133 media and took advantage of the species assignment provided by MetaG to target the isolation
134 of the desired bacteria according to literature reports¹⁷⁻²⁰. Therefore, different growth conditions
135 (media, temperature) were tested (Supplementary file 1) and permitted the isolation of 13
136 bacterial species (Table 1). Bacterial colonies were identified using either full length 16S or
137 two hypervariable regions of the 16S (V1V2, V3V4) and the presence of *nifH* was assayed
138 using three combinations of degenerate primers as well as *nifH* specific primers designed from
139 Metagenomics data whenever possible (Supplementary Fig. 3a,b,c). Six bacteria were

140 identified as nitrogen fixers based on a combination of *nifH* PCR amplification and/or *nifH* in
141 silico detection (Table 1). Comparative metagenomics analysis identified two *Sphingopyxis*
142 and *Bradyrhizobium* species which were distinguished using Meta Assembled Genome
143 (MAGs) specific primers designed for each species. The whole set of *nif* genes including the
144 *nifHDK* operon were identified in *Bradyrhizobium* MAG21, the only bacterium that was
145 sequenced as a single species while the other MAGs were assembled from either the MetaG
146 data or sequences from a pool of 3 isolated bacteria. Two and four copies of *nifH* were detected
147 in *Bradyrhizobium* MAG21 and *Mesorhizobium* MAG3 respectively. Among the 11 isolated
148 bacteria, two species *Kocuria* sp., and *Methylobacterium* were not detected in Metagenomics
149 sequences likely due to their low abundance emphasizing the importance of combining
150 sequencing and isolation/culturing methodologies (Table 1). Evidences for nitrogen fixation
151 were reported in the literature for these two genera^{21,22}. To further evidence nitrogen fixing
152 ability of the isolated bacteria, we used nitrogen-free medium with bromothymol blue, a pH
153 sensitive dye to grow the two most abundant bacteria, *Mesorhizobium* and *Bradyrhizobium*
154 MAG8 and found a switch in the color from green (pH=6.7) to blue (pH > 7) indicating nitrogen
155 fixation and the release of NH₃ which alkalizes the medium (Figure 1i).

156 **Diazotrophic MAGs are diverse and closely affiliated to environmentally sampled** 157 **diazotrophs of Alphaproteobacteria type**

158 To gain insights into *Pt15* associated diazotrophs and their phylogenetic relationship with
159 known taxa, we assembled their genomes and recovered several MAGs with completeness
160 higher than 97% (Table 1, Supplementary Table 3). To further improve MAG assembly for
161 some of the isolated bacteria, besides Metagenomics data initially used, MAGs were assembled
162 with additional Illumina and Pacific Pac Bio sequencing of the isolated bacteria including
163 *Bradyrhizobium*, *Georhizobium*, *Sphingopyxis*, and *Kocuria* which were sequenced either
164 alone or in a pool of 3 bacteria. Overall, chromosomal sequences were well assembled with
165 higher completeness as compared to Metagenomic assembled MAGs and nearly no cross
166 contamination of bins. To achieve a better identification of the bacterial genetic material, we
167 sought to recover from non-assembled reads, plasmids which are important elements for
168 bacteria survival susceptible to contain nitrogen fixation genes. We used three different plasmid
169 prediction tools and recovered 42 plasmids assigned to different bacterial species with an
170 average length of 97013 bp (Supplementary Table 3). The nitrogenase gene (*nifH*) was in silico
171 detected in six out of eight *Pt15* associated MAGs among which, one plasmid located *nifH*
172 (Table 1). Besides *nifH*, several additional *nif* genes were detected in most of the MAGs, among

173 which two were found in plasmids, *nifU* and *nifS* (Table 1). Phylogenomic analysis of *Pt15*
174 MAGs, MAGs of known terrestrial and marine diazotrophs from the Microscope database²³
175 and MAGs from *Tara* Oceans data using Anvi'o revealed two clusters, A (*Mesorhizobium*,
176 *Bradyrhizobium* MAGs, *Georhizobium*, *Phylobacterium*, and *Bevundimonas*) and B (2
177 *Sphingopyxis* MAGs and *Sphingobium*). Cluster A was affiliated to terrestrial and marine
178 diazotroph MAGs with a closer affiliation to terrestrial MAGs of the same genus (Fig. 3), while
179 cluster B was related to genera from marine and terrestrial environments identified as nitrogen
180 fixers. In both clusters, whenever the related MAGs are not identified as diazotrophs (likely due
181 to incomplete assembly), they were found to belong to Rhizobiales, Parvibaculales,
182 Caulobacterales and Sphingomonadales, orders known to contain nitrogen fixing species²⁴⁻²⁶.
183 Altogether, our results including (i) survival of the xenic diatom in the absence of usable forms
184 of nitrogen, (ii) nitrogen fixation demonstrated by ARA and BTB (iii) detection of *nifH*, and
185 (iv) the isolation from the xenic *Pt15* of NCDs clustering with known proteobacteria
186 diazotrophs, indicated that a symbiotic nitrogen fixation likely occurred in *Pt15* xenic culture
187 supporting the survival of diatom cells population in -N condition.

188

189 **Functional gene annotation of *Pt15* associated NCDs in nitrate deplete versus replete** 190 **conditions**

191 To understand the biological functions of the microbial community, genes from both
192 conditions were assigned to functional categories (Figure 4, supplementary Table S4). There
193 was a significant enrichment of genes related to nitrogen fixation and metabolism in -N
194 compared to +N condition. Typically, in -N, we exclusively identified *nifH* and *nifD*, two key
195 components of the nitrogenase enzyme complex. Likewise, *nifX* and *nifT* were only detected in
196 -N suggesting their critical role in nitrogen fixation. Part of the *nif* operon, *NifX* was reported
197 to be important in FeMo-co maturation and its transfer to *NifDK* proteins^{27,28}. Other genes
198 related to nitrogen fixation metabolism emerged from the analysis, namely *HemN*, *Fpr*, *FixS*,
199 *FixA*, *FixH*, *GroEL*, *NifS* and *FixJ* (Figure 4), some of which are part of the NIF regulon.

200 The abundance of several transporters was striking in -N, indicating an important role in
201 energy/substrates shuttling between the host and the bacterial community. Examples included
202 *EcfT* gene (energy-coupling factor transporter transmembrane) known for its function in
203 providing the energy necessary to transport a number of different substrates²⁹, *AcpS*, an acyl
204 carrier gene which functions as a multi-carbon units carrier into different biosynthetic

205 pathways³⁰, TauB, a taurine transporter that might play a role in providing taurine, which is
206 produced by a large number of algae and known as a potential source of carbone, nitrogen and
207 energy³¹. Stress responsive genes were clearly more represented in -N compared to +N
208 including SpoT, which is a general bacterial stress response³² that allows bacteria to increase
209 its persistence in stressful conditions, here the lack of nitrogen. Other genes found almost
210 exclusively in the starved bacterial community were the PAS domain containing genes (Per-
211 ARNT-Sim)³³, which function as signal sensors by their capacity to bind small molecules,
212 modulators and transducers. They are known to sense oxygen to prevent degradation of oxygen
213 sensitive enzymes suggesting a crucial role in -N condition where there is an enrichment in
214 nitrogenase containing bacteria. Overall, functional analysis of MetaG data corroborated with
215 NCDs enrichment in -N and nitrogen fixation.

216

217 **Geographical distribution and ecological relevance of NCD-microalgae symbiosis**

218 Next, we sought to assess the prevalence and importance of similar interactions in other
219 microalgae and in the environment. First, we screened a large collection of xenic microalgae as
220 described for *Pt15* and identified 5 species of microalgae (35%) that grew in the absence of
221 nitrogen in the growth medium suggesting the presence of diazotrophs (Supplementary Table
222 5). As for *Pt15*, the retained microalgae were screened using different nitrogen deprived media
223 under various conditions for diazotrophs isolation. *NifH* degenerate and 16S rRNA gene
224 primers were used to identify the isolated bacterial colonies which revealed five Alpha-
225 proteobacteria with *nifH* including 2 species of *Thalassiospira sp.* (isolated from two different
226 microalgae, *Stenotrophomonas maltophilia*, *Stappia sp.*), *Kocuria* and *Rhodococcus*
227 *qingshengii* suggesting that the occurrence of NCDs in a symbiotic interaction with microalgae
228 is frequent (Supplementary Fig. 3a,b, Supplementary Table 5). Identified NCDs were
229 sequenced in a pool of 3 bacteria using Illumina and Pacbio and their genome were assembled,
230 showing overall, a high completeness (>87%, 2 MAGs with 100%) and low contamination
231 (<2%) (Supplementary Table 3). We then used MAGs isolated from *Pt15* and the other
232 microalgae (Table 1, Supplementary Table 3) to reassess their phylogenetic relationship as
233 described for *Pt15* isolated NCDs. The analysis revealed 3 more clusters (C, D and E) in close
234 proximity to diazotrophic MAGs from the environment and/or MAGs belonging to diazotrophic
235 containing orders^{34 35 36}(Supplementary Fig. 4).

236 Next, we assessed the occurrence of *Pt15* MAGs in the environment in surface water and
237 Deep Chlorophyll Maximum layer (DCM) using *Tara* Ocean expedition datasets^{37,38}. Among
238 the 130 sampling stations, 92 stations show the occurrence of at least one of the *Pt15* MAGs
239 reflecting a high occurrence frequency (Fig. 5a, Supplementary fig. 5). *Pt15* closely related
240 TARA MAGs were more occurring with 129 recorded stations. The distribution of *Pt15* MAGs
241 was very large spanning all latitudes from south Pacific, Indian and Atlantic Oceans to polar
242 regions. The MAGs were found in both surface and DCM samples, mainly within the 0.8-5 μ m,
243 0.8-2000, 3-5-20 μ m and 180-2000 size fractions (Fig. 5a, supplementary Fig. 5). Interestingly,
244 *Pt15* MAGs co-occurred in surface and DCM samples in most cases, although their dominance
245 in surface samples was prominent. Except from few stations, clusters A and B seemed to co-
246 occur suggesting probable interactions between their respective bacteria.

247 We then asked whether the occurrence of the MAGs isolated from *Pt15* correlate with
248 environmental factors monitored during the expeditions and whether they co-occur with other
249 organisms. We considered only the top 2 MAGs of *Pt15* (*Mesorhizobium* MAG3 and
250 *Bradyrhizobium* MAG8) to assess their correlation with 4 environmental factors used as
251 indicators of diazotrophy, nitrate, nitrate/phosphate ratio, iron, and oxygen. Interestingly, *Pt15*
252 MAGs were detected in low nitrate and almost no detection above 5 μ M. Similarly, the
253 detection of *Pt15* MAGs coincided with low nitrate to phosphate ratio (Fig. 5b,c,d,e). Both
254 measurements are in line with nitrate limitation and phosphate abundance as markers of
255 diazotrophs presence^{39,40}. Iron availability seems to affect differently the abundance of both
256 MAGs. While, *Mesorhizobium* was observed in low to no detectable levels of iron,
257 *Bradyrhizobium* MAG8 was detected in higher iron concentrations suggesting different
258 requirement for iron in these two species. By contrast to what is known about oxygen inhibition
259 of nitrogenase, both MAGs seem to thrive in relatively well oxygenated zones (200 μ M)
260 compared to less oxygenated areas or oxygen minimum zones (OMZ, 20- 50 μ M O₂). This
261 implies adaptive means by which these bacteria fix nitrogen (Supplementary Fig. 5)

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

269 Discussion

270 Recent reports of NCDs dominance in the euphotic zones of the ocean questioned the
271 long-held dogma that symbiotic N₂ fixation is mainly carried by Cyanobacteria, but also raised
272 fundamental questions on how these presumably heterotroph NCDs can thrive in the sunlit
273 ocean^{12-14,41}. The nitrogenase enzyme complex responsible for nitrogen fixation is sensitive to
274 oxygen that irreversibly inactivates the enzyme. Therefore, diazotrophs evolved several
275 mechanisms to protect nitrogenase from oxygen degradation including hyper-respiration,
276 uncoupling of photosynthesis in space and time in photosynthetic diazotrophs and
277 attachment/interaction with microalgae or debris which were found to be hotspots of nitrogen
278 fixation⁴². Using a multidisciplinary approach, the study described herein brings unprecedented
279 evidence of a symbiotic interaction for nitrogen fixation between microalgae in particular the
280 model diatom *Phaeodactylum tricornutum* and NCDs. Our study revealed among several
281 screened xenic accessions of *P. tricornutum* in the absence of usable form of nitrogen, one
282 accession, *Pt15* interacting with nitrogen fixing proteobacteria. Although low, this nitrogen
283 fixation and uptake seemed to maintain the viability of the diatom and its associated bacteria.
284 To compensate for the allocation of energy to the bacterial community, *Pt15* cells seemed to
285 slow down growth. This lab culture interaction may reflect natural environments, where such
286 symbiotic interactions, in extreme conditions (low to no nitrogen) take place until favorable
287 factors (e.g., nitrate depletion with upwelling or Sahara dust) permit the bloom of cells we
288 observed in our experimental cultures when transferred to nitrate replete medium.

289 MetaG sequencing of the associated bacteria revealed not only one diazotrophic species
290 but a group of nitrogen fixers almost exclusively Alpha-proteobacteria mainly represented by
291 Rhizobiaceae, Sphingomonadaceae and Xanthobacteraceae. The nitrogenase gene was
292 identified in all of them either *in silico* or by PCR amplification, supporting further diazotrophy
293 in these bacteria. The *nifH* gene was found in most cases in the chromosome, except for
294 *Georhizobium* where it was predicted to be plasmidic. Interestingly the chromosomal *nif* genes
295 are clustered over 30-40 Kb regions forming genomic islands, a feature found in symbiotic
296 bacteria where these clusters are known as symbiotic islands, which corresponds to an
297 evolutionary strategy adopted by symbiotic bacteria in their quest to interact with eukaryotic
298 hosts⁴³. Demonstrated by ARA, the rates of nitrogen fixation in xenic *Pt15* in -N is low
299 compared to reports of diazotrophs of cyanobacteria type but similar to nitrogen fixation rates
300 known in proteobacteria⁴⁴.

301 Our study reveals that NCDs interaction with *Pt15* and other microalgae is likely not an
302 isolated phenomenon and is more broadly distributed than previously anticipated. Our MAGs
303 were detected in the environment clustering with previously identified terrestrial and marine
304 NCDs supporting diazotrophy and their ecological relevance. Our MAGs were identified in
305 surface and DCM samples in microalgae containing size fractions, which strongly supported
306 the hypothesis of an interaction with them as evidenced by the co-occurrence analysis. This
307 interaction might be taking advantage of hypo-oxic biomes microalgae and bacteria may
308 provide with the secretion of extracellular polymers mainly polysaccharides that form biofilm
309 ensuring low O₂ diffusion which was suggested by our TEM analysis with the formation of
310 pellicles around *Pt* cells and bacteria. Such pellicles were reported to form around two nitrogen
311 fixing bacteria, *P. stutzeri* and *Azospirillum brasilense* in oxygenated surroundings implying
312 the importance of oxygen stress in the formation of these structures⁴⁵. Besides biofilm
313 formation, this symbiotic interaction may combine different strategies for both microalgae and
314 NCDs to avoid/limit oxygen inactivation of nitrogenase such as a conformational switch that
315 protects nitrogenase from O₂ degradation^{10,46} or hyper-respiration⁴⁷ among other protective
316 mechanisms that need further investigations. In line with our discovery of NCDs interaction
317 with microalgae for nitrogen fixation, a recent study⁴⁸ suggested that anaerobic processes,
318 including heterotrophic N₂ fixation, can take place in anoxic microenvironments inside sinking
319 particles even in fully oxygenated marine waters.

320 The unique interaction we described here between a model diatom and a bacterial
321 community of diazotrophs available as lab culture, avoided environmental sampling related
322 issues allowing high quality sequencing and complete to nearly complete assembly of several
323 MAGs with an exceptional taxonomic assignation to the genus/species level. The high-
324 resolution taxonomic assignation of our MAGs represents a reliable reference for genus
325 assignation of the closely affiliated *Tara* Oceans MAGs with unknown taxonomic affiliation.
326 The identification of different diazotrophic taxa with a model diatom opens whole perspectives
327 to investigate the role of individual and combined interactions of bacterial species with the host
328 and within the bacterial community. Associated and often overlooked non-diazotrophic bacteria
329 enriched in -N, do certainly contribute to the symbiosis through different means such as
330 buffering the competition between several diazotrophs in the interaction with the host,
331 producing microalgae growth compounds and keeping away antagonists to secure the
332 interaction with the host. Thus, *Pt15* accession with its associated NCDs represent a unique

333 Eukaryotic-bacteria model system and an exciting opportunity to study molecular mechanisms
334 underlying this symbiotic interaction for nitrogen fixation and beyond.

335 **Methods**

336 **Microalgae species used and culture media**

337 *Screening of Phaeodactylum tricornutum accessions*

338 Ten xenic *P. tricornutum* accessions were screened in presence (545 μM) and absence of nitrate
339 in Enhanced Artificial Sea Water (EASW) in order to select xenic *Pt* ecotype growing in the
340 absence of nitrate. Starting concentration for all the accessions was 100,000 cells per ml. The
341 cells were counted and washed twice with nitrate free EASW in order to remove all residual
342 nitrate and resuspended in respective media for screening over a period of one month. A growth
343 curve was made by counting the cells in periodic time intervals. The cultures were grown at
344 19°C with a 12/12 hours photoperiod at 30 to 50 $\mu\text{E/s/m}^2$. A cocktail of antibiotics (Ampicillin
345 100 $\mu\text{g}/\mu\text{l}$, Chloramphenicol (14 $\mu\text{g}/\mu\text{l}$) and Streptomycin 100 $\mu\text{g}/\mu\text{l}$) was used to make *Pt15*
346 axenic in liquid EASW culture medium with refreshment of antibiotic containing medium every
347 5 days.

348 *Screening of other microalgae*

349 A total of eighteen microalgae were screened in nitrate-deplete and nitrate-replete enhanced
350 artificial sea water for growth (Supplementary Table 7). About 100,000 cells were counted,
351 resuspended in their respective media and grown as described above. Their growth was assessed
352 for a period of two weeks at regular intervals.

353 **Acetylene reduction assay**

354 Acetylene reduction assay (ARA)¹⁶ was used to quantify the conversion of acetylene to ethylene
355 (C_2H_4) which allows the assessment of nitrogenase activity. ARA was conducted on the
356 following samples: two positive controls, *Azotobacter. vinelandii* (NCIMB 12096),
357 *Pseudomonas stutzeri* BAL361, xenic *Pt15* grown in nitrogen free medium (NFM)^{49,50} or
358 nitrogen replete medium, all incubated at three different temperatures (19, 25, 30°C). Axenic
359 *Pt15* culture grown in NFM medium + 5 sugars (final concentration of 1% each: D-glucose, D-
360 sucrose, D-fructose, D-galactose and D-mannose) was used in every measurement as a negative
361 control to ensure there was no endogenous ethylene produced. All the cultures were grown at
362 their respective growth media and conditions. After reaching an appropriate O.D. between 0.2-
363 0.3, 1 ml of the bacterial cells were collected in an Eppendorf's tube and centrifuged to 11,000
364 rpm for 10 minutes at RT to remove the medium. The cells were re-suspended in nitrogen free

365 medium with five sugars and grown at their respective medium. When the bacterial growth
366 reached an O.D. of 0.2-0.3, the cultures were prepared for acetylene reduction assay. The
367 cultures were secured with a rubber stopper in order to make them air tight. Ten percent of the
368 air was removed from the culture flasks and supplemented with 10% acetylene gas. The cultures
369 were incubated at their appropriate temperatures for a period of 72 hours before monitoring
370 ARA with gas chromatography. A volume of 500 μ l were taken from the headspace of the
371 samples using a Hamilton gas tight syringe (SYR 500 μ l 750 RN no NDL, NDL large RN 6/pk)
372 and injected into a gas chromatograph 7820 model Agilent equipped with a Flame Ionization
373 Detector (GC-FID) and using Hydrogen as carrier gas at a constant flow of 7.7 mL/min. Sample
374 were injected into a split/splitless injector, set at 160°C in splitless mode. A GS-Alumina
375 capillary column, 50 m \times 0.53 mm \times 0.25 μ m (Agilent, 115-3552) was used. The programmed
376 oven temperature was at 160°C (isotherm program) during 2 min. The FID detector was set at
377 160°C. A standard curve was made with ethylene gas using 5 volumes (100, 150, 200 μ l). The
378 raw data generated from the GC provides value of y, Area [pA*s]. This value is supplemented
379 in the slope-intercept formula, $y = mx + c$, with 'm' and 'c' representing slope and constant
380 values respectively, denoted in the calibration curve. The 'x' value represents amount of
381 ethylene (nmoles/L/hour).

382 **Particulate organic nitrogen measurements**

383 One liter of 200,000 cells/ml culture of xenic *Pt15* was grown in nitrate-free EASW for over a
384 period of 80-days at 19°C and 12h/12h light/dark photo period at 70 μ E/s/m². Seven time points
385 (day(s)) used for harvesting samples: 1, 4, 12, 29, 59, 76 and 80. The cells were counted at each
386 time point using cell counter. Twenty ml of *Pt15* culture were harvested and injected through a
387 sterile (treated at 450 deg/4hrs) Whatman GF/F glass microfibre filter papers to collect *Pt15*
388 cells and the filters were dried overnight at 60°C incubator and stored at -80°C. The 20 ml
389 filtrate from the cultures was passed through 0.22 μ m filters (Sigma ca. GVWP10050), in order
390 to remove bacterial cells and stored at -80°C, before being processed for particulate organic
391 nitrogen measurement analysis. 545 μ M nitrate-EASW, 0 μ M nitrate-EASW and three heat
392 treated Whatman GF/F glass microfibre filter were used as blanks for comparison with test
393 samples.

394 **Microscopy**

395 *Light microscopy*

396 Living cells were directly used for preparations. Samples were observed with differential
397 interference contrast (DIC) microscopy by BX53 microscope (Olympus, Japan) and Axiocam

398 705 color (Carl Zeiss, Germany).

399

400 *Transmission electron microscopy*

401 Samples were prepared as described previously⁵¹. Embedded samples were cut for making
402 pale yellow ultra-thin sections (70-80nm thickness) with a microtome EM UC6 (Leica
403 Microsystems, Germany). Sections were counter-stained with TI blue (Nisshin EM, Japan) and
404 lead citrate for inspection with an electron microscope JEM-1011KM II (JEOL, Japan).

405

406 **Bacterial isolation and culture media used**

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

411 **Determination of nitrogen fixing ability using NfB medium**

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

419 **DNA extraction and PCR amplification of marker genes**

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

425 Wizard® Promega Genomic DNA extraction kit was used for extracting high-quality genomic
426 DNA from bacterial cells. The bacteria were grown in their appropriate media until the culture
427 optical density (OD) reached 0.3. The cells were harvested and washed in 1X PBS twice. The
428 cell pellets were resuspended in 480 µl 50 mM EDTA and treated with 120 µl of 20mg/ml
429 lysozyme (VWR cat no. 0663-5G), incubated at 37°C for 60 min in order to lyse possible gram-

430 positive bacteria. Post incubation, the cells were centrifuged and the supernatant was removed.
431 Nuclear Lysis solution, 600 μ l (Promega cat no. A7941), was used to re-suspend and gently
432 mix the pellet for incubation at 80°C for 5 min. The suspension was allowed to cool at room
433 temperature (RT) for 10 min before adding 1 μ l RNase at 10 mg/ml (VWR cat no. 0675-
434 250MG) and incubation at 37 for 60 min. After incubation, the suspension was cooled at RT
435 for 10 min. Protein lysis solution, 200 μ l (cat no. VWR 0663-5G) was added, vortexed and
436 incubated on ice (4°C for 5 min). The mixture was centrifuged and the clear supernatant was
437 transferred to a clean Eppendorf with 600 μ l RT isopropanol and the tube was mixed gently
438 before being centrifuged. The cell pellet was washed with freshly prepared 70% ethanol and re-
439 centrifuged. The ethanol was carefully removed without disturbing the pellet and air dried
440 before resuspension in appropriate amount of NFW. The quality of the DNA was checked using
441 Nanodrop 260/280 and 230/260 ratios and on 1% agarose gel. The identity of the bacteria was
442 re-confirmed using 16S rRNA gene amplification and sanger sequencing.

443 Three degenerate primers (F1 – nifH3R, nifH2F – nifh3R, PolF – PolR)⁵⁴ for *nifH* gene
444 detection by PCR were standardized. The PCR was done as follows, 95°C for 5 minutes (1
445 cycle), followed by 40 cycles of 95°C at 30 seconds, standardized annealing temperature at 30
446 seconds, elongation at 72°C for 30 seconds and a final extension at 72°C for 7 minutes. The
447 annealing temperature used for F1 – nifH3R, nifH2F – nifh3R, PolF – PolR combinations were
448 48, 52 and 54°C respectively. Fifty ng of DNA was added to a 20 μ l PCR reaction containing,
449 10X Dream Taq buffer, 0.2 mM dNTPs, 2 pmol primers and 1.25 U of *Taq* (Fischer Scientific
450 ca. n° 15689374). The reaction was performed with 35 cycles of 95 °C (1 min), 55 °C (1 min)
451 and 72 °C (1 min). For species identification, 16S PCR fd1-rd1, variable regions PCR v1v2
452 v3v4 and v5v7 were used. All primer sequence details are reported in supplementary Table 6.

453 **Sample preparation and DNA extraction for Illumina and Pacific Pac Bio sequencing**

454 Duplicate with a starting concentration of 200,000 cells/ml of 3-liter culture volumes of xenic
455 *Pt15* in nitrate-replete and nitrate-deplete enhanced artificial sea water were grown for a period
456 of 2 months (60-days) at 19°C and 12h/12h light/dark photoperiod at 70 μ E/s/m² before
457 collection for sequencing. The xenic *Pt15* cells were filtered through a 3 μ m filter (MF-
458 millipore hydrophobic nitrocellulose, diameter 47 mm (cat log no. SSWP04700) to retain the
459 cells and allow passage of the filtrate containing the bacterial community. The filtrate was
460 filtered twice in order to avoid contamination of the bacterial cells with residual *Pt15* cells. The
461 double filtered bacterial community was collected in 0.22-micron filter (MF-millipore
462 hydrophobic nitrocellulose, diameter 47 mm (cat log no. GSWP04700) and resuspended in 1X

463 PBS. DNA was extracted for metagenome sequencing using Quick-DNA fungal/bacterial mini
464 prep kit (cat no. DG005) and according to manufacturer instructions. The bacterial DNA quality
465 was checked using 1% agarose gel to assess intact DNA and quantified using Qubit broad range
466 DNA kit, before being sequenced. Additionally, 16S rRNA sequencing was performed to assess
467 the presence of bacterial DNA and 18s rRNA PCR was run as a negative control to inspect
468 absence of *P. tricornutum* DNA. When bacterial isolates are available, they are sequenced either
469 as a single species or in a pool of 3 bacteria using Illumina and PacBio.

470 **Data processing**

471 **Metagenomic analysis**

472 Two different strategies were used to analyze the metagenomic reads : (i) a “Single assemblies”
473 strategy, assembly and binning was performed for each sample individually. This strategy is
474 known to produce less fragmented assemblies with a limitation of chimeric contigs formation⁵⁵.
475 (ii) a co-assembly strategy was also performed to complete the results with additional MAGs.
476 The idea is to pool all the available samples reads and build a unique and common assembly
477 for all samples. Samples from 2 experiments were used. An experiment with sequences
478 obtained at 2g/L salinity (one sample in +N condition / one sample in -N condition) and the
479 other experiment at 20 g/L salinity includes duplicates for each of the conditions. Both data sets
480 are similar.

481 *Single assembly strategy*

482 DNA raw reads were analyzed using a dedicated metagenomic pipeline ATLAS v2.7⁵⁵. This
483 tools includes four major steps of shotgun analysis: quality control of the raw reads, assembly,
484 binning, taxonomic and functional annotations. Briefly, quality control of raw sequences is
485 performed using utilities in the BBTools suite (<https://sourceforge.net/projects/bbmap/>):
486 BBDuk for trimming and removing adapters, BBSplit for decontamination. In our case
487 *Phaeodactylum tricornutum* reference genome, including mitochondria and chloroplast
488 sequences, was used for this specific step. Quality-control (QC) reads assemblies were
489 performed using metaSPADES⁵⁶ with ATLAS default parameters. The assembled contigs
490 shorter than 500 bases length and without mapped reads, were removed to keep high-quality
491 contigs.

492 Metagenome-assembled genomes (MAGs) were obtained using two distinct binners :
493 maxbin2⁵⁷ and metabat2⁵⁸. The bins obtained by the different binning tools were combined
494 using DASTool⁵⁹ and dereplicated with dRep⁶⁰ using default parameters. Completeness and

495 contamination were computed using CheckM⁶¹. Only bins with a completeness above 50 % and
496 a contamination rate under 10 % were kept. Prediction of open reading frames (ORFs) was
497 performed, at the assembly level (not only at the MAGs level) using Prodigal⁶². Linclust⁶³ was
498 used to cluster the translated gene product obtained to generate gene and protein catalogs
499 common to all samples.

500 *Co-assembly strategy*

501 QC reads and co-assembly were analyzed using a shotgun dedicated sequential pipeline
502 MetaWRAP v1.3.2⁵⁹. Briefly, QC reads are pulled together and addressed to the chosen
503 assembler MetaSPADES using MetaWRAP default parameters. Metagenome-assembled
504 genomes (MAGs) were obtained using two distinct binners: maxbin2 and metabat2 and the bins
505 obtained were combined using an internal MetaWRAP dedicated tool. Completeness and
506 contamination were computed using CheckM⁶¹. Only bins with a completeness above 50 % and
507 a contamination rate under 10 % were kept. MetaWRAP also includes a re-assemble bins
508 module that collect reads belonging to each bin, and then reassemble them independently with
509 a "permissive" and a "strict" algorithm. Only improved bins are kept in the final set. Using this
510 strategy, we were able to add an additional MAG.

511 **MAGs relative abundances**

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

515 **Taxonomic assignation of metagenomic data**

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

521 **Hybrid assembly of MAGs**

522 *Assembly*

523 Metagenomic assembly of Illumina technology reads does not yield complete
524 genomes due to the difficulty of assembling repetitive regions⁶⁴. Whenever Illumina and Pac
525 bio reads are available, we used a strategy to improve MAGs assembly by combining both type

526 of reads. Metagenomic assembly of Illumina reads paired
527 end and PacBio reads post quality trimming, were assembled using metaSPAdes or
528 SPAdes(v3.15.5) with option flag (--meta). To validate the newly assembled
529 genomes, we verified that the long reads covered regions that were not supported by
530 the mapped Illumina sequences only and that we get high quality bins. We obtained
531 genomes in which some regions had not been supported by the assembly of Illumina
532 reads only.

533 *Binning*

534 The contigs were grouped and assigned to the individual genome. To do this, we used MaxBin2
535 and metaBAT2 which are clustering methods based on compositional features or on alignment
536 (similarity), or both (<https://anaconda.org/bioconda/maxbin2> and
537 <https://anaconda.org/bioconda/metabat2>)⁶⁵.

538 *Refinement*

539 Bins from metabat2 and maxbin2, were consolidated into one more solid set of bins. To do
540 this we used Binning_refiner which improves genomic bins by combining different binning
541 programs available at:https://github.com/songweizhi/Binning_refiner⁶⁶.

542 *Taxonomic Assignment*

543 GTDB-Tk has already been independently and positively evaluated for the classification of
544 MAGs⁶⁷. After consolidating the bins in the refinement step, we determined the taxonomy of
545 each bin. The gtdbtk classify_wf module was used in this work using a GTDB database
546 <https://github.com/ecogenomics/gtdbtk>⁶⁸.

547

548 **Gene annotation of metagenomic data and MAGs**

549 Gene annotation of metagenomics data in two conditions was done using Anvio version 7.2.
550 Three databases were used for a robust gene observations: COG20
551 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383993>,<http://www.ncbi.nlm.nih.gov/COG/>),
552 KEGG and PROKKA (<https://github.com/tseemann/prokka>). DIAMOND to search NCBI's
553 Database. In built, Anvio program annotates a contig database with HMM hits from KOfam, a
554 database of KEGG Orthologs (KOs, <https://www.genome.jp/kegg/ko.html>). For PROKKA, the
555 annotation was done externally and the gene calls were imported to Anvio platform for the
556 individual metagenome contigs databases.

557

558 **Phylogenomic analysis of MAGs**

559 The phylogenomic tree was built using Anvio-7.1 (<https://github.com/merenlab/anvio/releases>)
560 with database of 1888 TARA ocean MAGs (contig-level FASTA files) submitted to Genoscope
561 (<https://www.genoscope.cns.fr/tara/>). Additional closest terrestrial and marine genomes
562 (supplementary table 7) to the individual MAGs were selected using MicroScope database
563 (mage.genoscope.cns.fr/microscope/) using genome clustering, utilizing MASH
564 (<https://github.com/marbl/Mash>) for ANI (Average nucleotide identity) for calculating pairwise
565 genomic distancing, neighbour-joining (<https://www.npmjs.com/package/neighbor-joining>) for
566 tree-construction and computing the clustering using Louvain Community Detection
567 (<https://github.com/taynaud/python-louvain>). Anvi'o utilizes 'prodigal'⁶² to identify open
568 reading frames for the contig database. The HMM profiling for 71 single core-copy bacterial
569 genes (SCGs) is done using in-built Anvi'o database, Bacteria_71
570 (<https://doi.org/10.1093/bioinformatics/btz188>) and utilizes MUSCLE for sequence alignment
571 (<http://www.drive5.com/muscle>). The bootstrap values are represented from 0 – 1 with
572 replicates of 100.

573 **Plasmid prediction and annotation**

574 Plasmids were predicted from MetaG data using the following plasmid assembly tools:
575 SCAPP⁶⁹ (Sequence Contents-Aware Plasmid Peeler), <https://github.com/Shamir-Lab/SCAPP>,
576 metaplasmidSPAdes⁷⁰ (GitHub - ablab/spades: SPAdes Genome Assembler) and PlaScope⁷¹
577 (GitHub - labgem/PlaScope: Plasmid exploration of bacterial genomes). Predicted plasmids
578 were annotated using Dfast⁷² (https://github.com/nigyta/dfast_core), eggNOG-mapper v2⁷³
579 and PROKKA 1.14.6.

580 **Downloading of metagenomes assembled-genomes counts tables**

581 MAGs abundance count tables and fasta files were downloaded from open-source released
582 Foundation Tara Oceans database at <https://www.genoscope.cns.fr/tara/>^{13,74}. Data include
583 2601 MAGs of which 1888 prokaryotes and 713 eukaryotes MAGs in 937 samples.

584 **Biogeography of metagenomes assembled-genomes**

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

589 counts by the total number of reads in a given sample. Different depths (surface and deep
590 chlorophyll maximum) and size fractions (0.22-1.6/3µm, 3/5-20µm, 0.8-5µm, 20-180/200µm,
591 0.8-2000µm, 180-2000µm) were analysed separately to project relative abundances on global
592 maps. Relative abundances of TARA MAGs phylogenetically close to *Pt15* and belonging to
593 the same cluster were summed up for each sample. The sizes of the pie charts reflect the total
594 amount of relative abundances present in each sample.

595 **Environmental parameters preferences**

596 Tara Oceans physico-chemical parameters were downloaded from Pangaea³⁷. MAGs relative
597 abundances (after TSS normalisation) were projected on global maps using R.

598 **Co-occurrence network inference**

599 Co-occurrence network inference was performed using FlashWeave v0.18.0⁷⁵ with Julia v1.5.3,
600 for eukaryotic and prokaryotic Tara MAGs linked to *Pt15* MAGs. To account for the
601 compositional nature of MAGs abundance data, a multiplicative replacement method⁷⁶ was
602 applied to replace zeros, and a Centered Log Ratio transformation⁷⁷ was applied separately on
603 both prokaryotic and eukaryotic abundance matrices. FlashWeave parameters used were
604 max_k=3, n_obs_min=10, and normalize=False as we have applied our own standardisation
605 (CLR).

606 **Acknowledgements**

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

618

619 References

- 620 1 Moore, C. M., Mills, M.M., Arrigo, K.R. I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, ,
621 R. J. Geider, C. G., S. L. Jaccard, T. D. Jickells, J. La Roche, T. M. Lenton, N. M. Mahowald, , E.
622 Marañón, I. M., J. K. Moore, T. Nakatsuka, A. Oschlies, M. A. Saito, T. F. Thingstad, & Ulloa, A.
623 T. O. Processes and patterns of oceanic
624 nutrient limitation. *Nature Geoscience* **6**, 701-710 (2013).
- 625 2 Lee, C. C., Ribbe, M. W. & Hu, Y. Cleaving the n,n triple bond: the transformation of
626 dinitrogen to ammonia by nitrogenases. *Met Ions Life Sci* **14**, 147-176 (2014).
627 https://doi.org/10.1007/978-94-017-9269-1_7
- 628 3 Einsle, O. & Rees, D. C. Structural Enzymology of Nitrogenase Enzymes. *Chem Rev* **120**, 4969-
629 5004 (2020). <https://doi.org/10.1021/acs.chemrev.0c00067>
- 630 4 Kneip, C., Lockhart, P., Voss, C. & Maier, U. G. Nitrogen fixation in eukaryotes--new models
631 for symbiosis. *BMC Evol Biol* **7**, 55 (2007). <https://doi.org/10.1186/1471-2148-7-55>
- 632 5 Fiore, C. L., Jarett, J. K., Olson, N. D. & Lesser, M. P. Nitrogen fixation and nitrogen
633 transformations in marine symbioses. *Trends Microbiol* **18**, 455-463 (2010).
634 <https://doi.org/10.1016/j.tim.2010.07.001>
- 635 6 Cornejo-Castillo, F. M. & Zehr, J. P. Hopanoid lipids may facilitate aerobic nitrogen fixation in
636 the ocean. *Proc Natl Acad Sci U S A* **116**, 18269-18271 (2019).
637 <https://doi.org/10.1073/pnas.1908165116>
- 638 7 Inomura, K. *et al.* Quantifying Oxygen Management and Temperature and Light
639 Dependencies of Nitrogen Fixation by *Crocospaera watsonii*. *mSphere* **4** (2019).
640 <https://doi.org/10.1128/mSphere.00531-19>
- 641 8 Inomura, K., Wilson, S. T. & Deutsch, C. Mechanistic Model for the Coexistence of Nitrogen
642 Fixation and Photosynthesis in Marine Trichodesmium. *mSystems* **4** (2019).
643 <https://doi.org/10.1128/mSystems.00210-19>
- 644 9 Munoz-Marin, M. D. C. *et al.* The Transcriptional Cycle Is Suited to Daytime N₂ Fixation in the
645 Unicellular Cyanobacterium "Candidatus Atelocyanobacterium thalassa" (UCYN-A). *mBio* **10**
646 (2019). <https://doi.org/10.1128/mBio.02495-18>
- 647 10 Schlesier, J., Rohde, M., Gerhardt, S. & Einsle, O. A Conformational Switch Triggers
648 Nitrogenase Protection from Oxygen Damage by Shethna Protein II (FeSII). *J Am Chem Soc*
649 **138**, 239-247 (2016). <https://doi.org/10.1021/jacs.5b10341>
- 650 11 Bothe, H., Tripp, H. J. & Zehr, J. P. Unicellular cyanobacteria with a new mode of life: the lack
651 of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch Microbiol* **192**,
652 783-790 (2010). <https://doi.org/10.1007/s00203-010-0621-5>
- 653 12 Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are
654 abundant in surface ocean metagenomes. *Nat Microbiol* **3**, 804-813 (2018).
655 <https://doi.org/10.1038/s41564-018-0176-9>
- 656 13 Delmont, T. O. P. K., J.S. Veseli, I., Fuessel, J. Murat Eren, A., Foster, R.A., Bowler, C., Wincker,
657 P., Pelletier, E. Heterotrophic bacterial diazotrophs are more abundant than their
658 cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *doi:*
659 <https://doi.org/10.1101/2021.03.24.436778> (2021).
- 660 14 Farnelid, H. *et al.* Nitrogenase gene amplicons from global marine surface waters are
661 dominated by genes of non-cyanobacteria. *PLoS One* **6**, e19223 (2011).
662 <https://doi.org/10.1371/journal.pone.0019223>
- 663 15 Moisaner, P. H. *et al.* Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic
664 Environments. *Front Microbiol* **8**, 1736 (2017). <https://doi.org/10.3389/fmicb.2017.01736>
- 665 16 Sullivan, B. W. *et al.* Spatially robust estimates of biological nitrogen (N) fixation imply
666 substantial human alteration of the tropical N cycle. *Proc Natl Acad Sci U S A* **111**, 8101-8106
667 (2014). <https://doi.org/10.1073/pnas.1320646111>

- 668 17 de Lajudie, P. *et al.* Characterization of tropical tree rhizobia and description of
669 Mesorhizobium plurifarum sp. nov. *Int J Syst Bacteriol* **48 Pt 2**, 369-382 (1998).
670 <https://doi.org/10.1099/00207713-48-2-369>
- 671 18 Yang, X. *et al.* Mesorhizobium alexandrii sp. nov., isolated from phycosphere microbiota of
672 PSTs-producing marine dinoflagellate Alexandrium minutum amtk4. *Antonie Van*
673 *Leeuwenhoek* **113**, 907-917 (2020). <https://doi.org/10.1007/s10482-020-01400-x>
- 674 19 Zhang, X., Tong, J., Dong, M., Akhtar, K. and He, B. Isolation, identification and
675 characterization of nitrogen fixing endophytic bacteria and their effects on cassava
676 production. *PeerJ*. *10*: e12677 **10** (2022).
- 677 20 Bostrom, K. H., Riemann, L., Kuhl, M. & Hagstrom, A. Isolation and gene quantification of
678 heterotrophic N₂-fixing bacterioplankton in the Baltic Sea. *Environ Microbiol* **9**, 152-164
679 (2007). <https://doi.org/10.1111/j.1462-2920.2006.01124.x>
- 680 21 Ouyabe, M., Kikuno, H., Tanaka, N., Babil, P. and Shiwachi, H. Isolation and identification of
681 nitrogen-fixing bacteria
682 associated with Dioscorea alata L. and
683 Dioscorea esculenta L. *Microb. Resour. Syst.* **35**, 3-11 (2019).
- 684 22 Sy, A. *et al.* Methylophilic Methylobacterium bacteria nodulate and fix nitrogen in
685 symbiosis with legumes. *J Bacteriol* **183**, 214-220 (2001).
686 <https://doi.org/10.1128/JB.183.1.214-220.2001>
- 687 23 Medigue, C. *et al.* MicroScope-an integrated resource for community expertise of gene
688 functions and comparative analysis of microbial genomic and metabolic data. *Brief Bioinform*
689 **20**, 1071-1084 (2019). <https://doi.org/10.1093/bib/bbx113>
- 690 24 Lesser, M. P., Morrow, K. M., Pankey, S. M. & Noonan, S. H. C. Diazotroph diversity and
691 nitrogen fixation in the coral Stylophora pistillata from the Great Barrier Reef. *ISME J* **12**, 813-
692 824 (2018). <https://doi.org/10.1038/s41396-017-0008-6>
- 693 25 Videira, S. S., de Araujo, J. L., Rodrigues Lda, S., Baldani, V. L. & Baldani, J. I. Occurrence and
694 diversity of nitrogen-fixing Sphingomonas bacteria associated with rice plants grown in Brazil.
695 *FEMS Microbiol Lett* **293**, 11-19 (2009). <https://doi.org/10.1111/j.1574-6968.2008.01475.x>
- 696 26 Skorupska, A., Janczarek, M., Marczak, M., Mazur, A. & Krol, J. Rhizobial exopolysaccharides:
697 genetic control and symbiotic functions. *Microb Cell Fact* **5**, 7 (2006).
698 <https://doi.org/10.1186/1475-2859-5-7>
- 699 27 Hernandez, J. A. *et al.* NifX and NifEN exchange NifB cofactor and the VK-cluster, a newly
700 isolated intermediate of the iron-molybdenum cofactor biosynthetic pathway. *Mol Microbiol*
701 **63**, 177-192 (2007). <https://doi.org/10.1111/j.1365-2958.2006.05514.x>
- 702 28 Jasniewski, A. J., Lee, C. C., Ribbe, M. W. & Hu, Y. Reactivity, Mechanism, and Assembly of the
703 Alternative Nitrogenases. *Chem Rev* **120**, 5107-5157 (2020).
704 <https://doi.org/10.1021/acs.chemrev.9b00704>
- 705 29 Wang, T. *et al.* Structure of a bacterial energy-coupling factor transporter. *Nature* **497**, 272-
706 276 (2013). <https://doi.org/10.1038/nature12045>
- 707 30 Geiger, O. & Lopez-Lara, I. M. Rhizobial acyl carrier proteins and their roles in the formation
708 of bacterial cell-surface components that are required for the development of nitrogen-fixing
709 root nodules on legume hosts. *FEMS Microbiol Lett* **208**, 153-162 (2002).
710 <https://doi.org/10.1111/j.1574-6968.2002.tb11075.x>
- 711 31 Cook, A. M. & Denger, K. Metabolism of taurine in microorganisms: a primer in molecular
712 biodiversity? *Adv Exp Med Biol* **583**, 3-13 (2006). [https://doi.org/10.1007/978-0-387-33504-
713 9_1](https://doi.org/10.1007/978-0-387-33504-9_1)
- 714 32 Boutte, C. C. & Crosson, S. Bacterial lifestyle shapes stringent response activation. *Trends*
715 *Microbiol* **21**, 174-180 (2013). <https://doi.org/10.1016/j.tim.2013.01.002>
- 716 33 Stuffle, E. C., Johnson, M. S. & Watts, K. J. PAS domains in bacterial signal transduction. *Curr*
717 *Opin Microbiol* **61**, 8-15 (2021). <https://doi.org/10.1016/j.mib.2021.01.004>

- 718 34 Madigan, M., Cox, S. S. & Stegeman, R. A. Nitrogen fixation and nitrogenase activities in
719 members of the family Rhodospirillaceae. *J Bacteriol* **157**, 73-78 (1984).
720 <https://doi.org/10.1128/jb.157.1.73-78.1984>
- 721 35 Singh, R. K. *et al.* Unraveling Nitrogen Fixing Potential of Endophytic Diazotrophs of Different
722 Saccharum Species for Sustainable Sugarcane Growth. *International Journal of Molecular*
723 *Sciences* **23**, 6242 (2022).
- 724 36 Sellstedt, A., Richau, K.H. Aspects of nitrogen-fixing Actinobacteria, in particular free-living
725 and symbiotic Frankia. *FEMS Microbiology Letters* **342**, 179–186 (2013).
- 726 37 Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data.
727 *Sci Data* **2**, 150023 (2015). <https://doi.org/10.1038/sdata.2015.23>
- 728 38 Karsenti, E. *et al.* A holistic approach to marine eco-systems biology. *PLoS biology* **9**,
729 e1001177 (2011). <https://doi.org/10.1371/journal.pbio.1001177>
- 730 39 Sanudo-Wilhelmy, S. A. *et al.* Phosphorus limitation of nitrogen fixation by Trichodesmium in
731 the central Atlantic Ocean. *Nature* **411**, 66-69 (2001). <https://doi.org/10.1038/35075041>
- 732 40 Ward BA, D. S., Moore CM, Follows MJ. Iron, phosphorus, and nitrogen supply ratios define
733 the biogeography of
734 nitrogen fixation. . *Limnol Oceanogr.* **58**, 2059–2075 (2013).
- 735 41 Bombar, D., Paerl, R. W. & Riemann, L. Marine Non-Cyanobacterial Diazotrophs: Moving
736 beyond Molecular Detection. *Trends Microbiol* **24**, 916-927 (2016).
737 <https://doi.org/10.1016/j.tim.2016.07.002>
- 738 42 Geisler, E., Bogler, A., Rahav, E. & Bar-Zeev, E. Direct Detection of Heterotrophic Diazotrophs
739 Associated with Planktonic Aggregates. *Sci Rep* **9**, 9288 (2019).
740 <https://doi.org/10.1038/s41598-019-45505-4>
- 741 43 Ochman, H. & Moran, N. A. Genes lost and genes found: evolution of bacterial pathogenesis
742 and symbiosis. *Science* **292**, 1096-1099 (2001). <https://doi.org/10.1126/science.1058543>
- 743 44 Farnelid, H. *et al.* Active nitrogen-fixing heterotrophic bacteria at and below the chemocline
744 of the central Baltic Sea. *ISME J* **7**, 1413-1423 (2013). <https://doi.org/10.1038/ismej.2013.26>
- 745 45 Wang, D., Xu, A., Elmerich, C. & Ma, L. Z. Biofilm formation enables free-living nitrogen-fixing
746 rhizobacteria to fix nitrogen under aerobic conditions. *ISME J* **11**, 1602-1613 (2017).
747 <https://doi.org/10.1038/ismej.2017.30>
- 748 46 Maier, R. J. & Moshiri, F. Role of the *Azotobacter vinelandii* nitrogenase-protective shethna
749 protein in preventing oxygen-mediated cell death. *J Bacteriol* **182**, 3854-3857 (2000).
750 <https://doi.org/10.1128/JB.182.13.3854-3857.2000>
- 751 47 Brian, A. C. J., C. W. . The respiratory system of *Azotobacter vinelandii*. *Eur. J. Biochem.* **20**,
752 29–35 (1971).
- 753 48 Chakraborty, S. *et al.* Quantifying nitrogen fixation by heterotrophic bacteria in sinking
754 marine particles. *Nat Commun* **12**, 4085 (2021). <https://doi.org/10.1038/s41467-021-23875-6>
- 755 49 Jorgensen, J. H., Pfaller, M.A., Carroll, K.C., Funke, G., Landry, M.L., Richter, S.S and Warnock.,
756 D.W. . *Manual of Clinical Microbiology*. 11th edn, Vol. 1 (2015).
- 757 50 Isenberg, H. D. *Clinical Microbiology Procedures Handbook*. 2nd Edition edn, Vol. 1 (Amer
758 Society for Microbiology, 1992).
- 759 51 Tanaka A, D. M. A., Amato A, Montsant A, Mathieu B, Rostaing P, Tirichine L & Bowler C. .
760 Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom
761 *Phaeodactylum tricornutum*. *Protist* **166**, 506-521 (2015).
- 762 52 Baldani, J. I., Reis, V.M., Videira, S.S. *et al.* The art of isolating nitrogen-fixing bacteria from
763 non-leguminous plants using N-free semi-solid media: a practical guide for microbiologists.
764 *Plant Soil* **384**, 413–431
- 765 53 Jha, C. K., Patel, D., Rajendran, N. & Saraf, M. Combinatorial assessment on dominance and
766 informative diversity of PGPR from rhizosphere of *Jatropha curcas* L. *J Basic Microbiol* **50**,
767 211-217 (2010). <https://doi.org/10.1002/jobm.200900272>

- 768 54 Gaby, J. C. & Buckley, D. H. A comprehensive aligned nifH gene database: a multipurpose tool
769 for studies of nitrogen-fixing bacteria. *Database (Oxford)* **2014**, bau001 (2014).
770 <https://doi.org/10.1093/database/bau001>
- 771 55 Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M. & McCue, L. A. ATLAS: a Snakemake
772 workflow for assembly, annotation, and genomic binning of metagenome sequence data.
773 *BMC Bioinformatics* **21**, 257 (2020). <https://doi.org/10.1186/s12859-020-03585-4>
- 774 56 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile
775 metagenomic assembler. *Genome Res* **27**, 824-834 (2017).
776 <https://doi.org/10.1101/gr.213959.116>
- 777 57 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to
778 recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).
779 <https://doi.org/10.1093/bioinformatics/btv638>
- 780 58 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
781 reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
782 <https://doi.org/10.7717/peerj.7359>
- 783 59 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication,
784 aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).
785 <https://doi.org/10.1038/s41564-018-0171-1>
- 786 60 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate
787 genomic comparisons that enables improved genome recovery from metagenomes through
788 de-replication. *ISME J* **11**, 2864-2868 (2017). <https://doi.org/10.1038/ismej.2017.126>
- 789 61 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing
790 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
791 *Genome Res* **25**, 1043-1055 (2015). <https://doi.org/10.1101/gr.186072.114>
- 792 62 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
793 identification. *BMC Bioinformatics* **11**, 119 (2010). <https://doi.org/10.1186/1471-2105-11-119>
- 794 63 Steinegger, M. & Soding, J. Clustering huge protein sequence sets in linear time. *Nat*
795 *Commun* **9**, 2542 (2018). <https://doi.org/10.1038/s41467-018-04964-5>
- 796 64 Giguere, D. J., Bahcheli, A.T., Joris, B.R., Paulssen, J.M., Gieg, L.M., Flatley, M.W., Gloor, G.B.
797 Complete and validated genomes from a metagenome. *doi:*
798 <https://doi.org/10.1101/2020.04.08.032540> (2020).
- 799 65 Meyer, F. *et al.* Tutorial: assessing metagenomics software with the CAMI benchmarking
800 toolkit. *Nat Protoc* **16**, 1785-1801 (2021). <https://doi.org/10.1038/s41596-020-00480-3>
- 801 66 Song, W. Z. & Thomas, T. Binning_refiner: improving genome bins through the combination
802 of different binning programs. *Bioinformatics* **33**, 1873-1875 (2017).
803 <https://doi.org/10.1093/bioinformatics/btx086>
- 804 67 Coil, D. A. *et al.* Genomes from bacteria associated with the canine oral cavity: A test case for
805 automated genome-based taxonomic assignment. *PLoS One* **14**, e0214354 (2019).
806 <https://doi.org/10.1371/journal.pone.0214354>
- 807 68 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify
808 genomes with the Genome Taxonomy Database. *Bioinformatics* (2019).
809 <https://doi.org/10.1093/bioinformatics/btz848>
- 810 69 Pellow, D. *et al.* SCAPP: an algorithm for improved plasmid assembly in metagenomes.
811 *Microbiome* **9**, 144 (2021). <https://doi.org/10.1186/s40168-021-01068-z>
- 812 70 Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. Plasmid detection and assembly in
813 genomic and metagenomic data sets. *Genome Res* **29**, 961-968 (2019).
814 <https://doi.org/10.1101/gr.241299.118>
- 815 71 Royer, G. *et al.* PlaScope: a targeted approach to assess the plasmidome from genome
816 assemblies at the species level. *Microb Genom* **4** (2018).
817 <https://doi.org/10.1099/mgen.0.000211>

- 818 72 Tanizawa, Y., Fujisawa, T., Kaminuma, E., Nakamura, Y. & Arita, M. DFAST and DAGA: web-
819 based integrated genome annotation tools and resources. *Biosci Microbiota Food Health* **35**,
820 173-184 (2016). <https://doi.org/10.12938/bmfh.16-003>
- 821 73 Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-
822 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
823 Metagenomic Scale. *Mol Biol Evol* (2021). <https://doi.org/10.1093/molbev/msab293>
- 824 74 Delmont, T., Gaia, M., Hinsinger DD., et. al. . Functional repertoire convergence of distantly
825 related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *Cell*
826 *Genomics* **2** (2022).
- 827 75 Tackmann, J., Matias Rodrigues, J. F. & von Mering, C. Rapid Inference of Direct Interactions
828 in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. *Cell Syst*
829 **9**, 286-296 e288 (2019). <https://doi.org/10.1016/j.cels.2019.08.002>
- 830 76 Fernández, J. A. M., Vidal, C.B. & Glahn, V.P. . Dealing with Zeros and Missing Values in
831 Compositional Data Sets Using Nonparametric Imputation. *Mathematical Geology* **35**,
832 pages253–278 (2003).
- 833 77 Quinn, T. P. *et al.* A field guide for the compositional analysis of any-omics data. *Gigascience*
834 **8** (2019). <https://doi.org/10.1093/gigascience/giz107>

835

836

837 Legend

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

857 the membrane. c, chloroplast, v, vacuole, n, nucleus. Scale bars, 10 μm (c-f) and 2 μm (g, h). i,
858 color change from green (negative control, pH = 6.7) to blue (control and assayed strains, pH
859 > 6.7) due to pH change from acidic to alkaline induced by nitrogen fixation and release of
860 NH_3^+ monitored 16 days post-inoculation. From left to right, *Azotobacter vinelandii* (NCIMB
861 12096), *Pseudomonas stutzeri* (NCIMB885), NfB medium without inoculum, *Mesorhizobium*
862 MAG3, *Bradyrhizobium* MAG8. PC, Positive Control, PC, Negative Control.

863

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

867

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

870

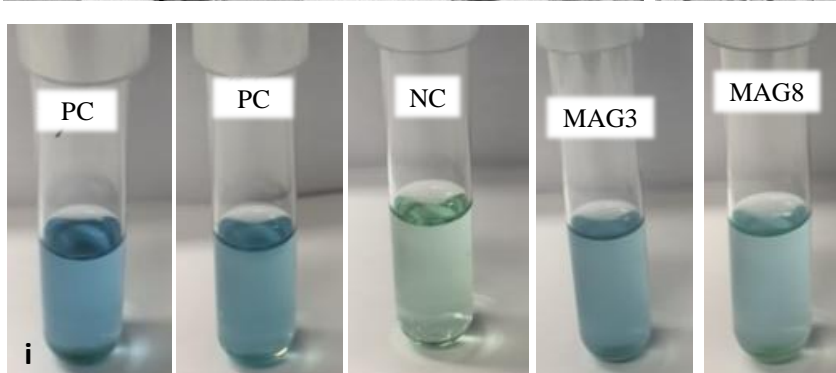
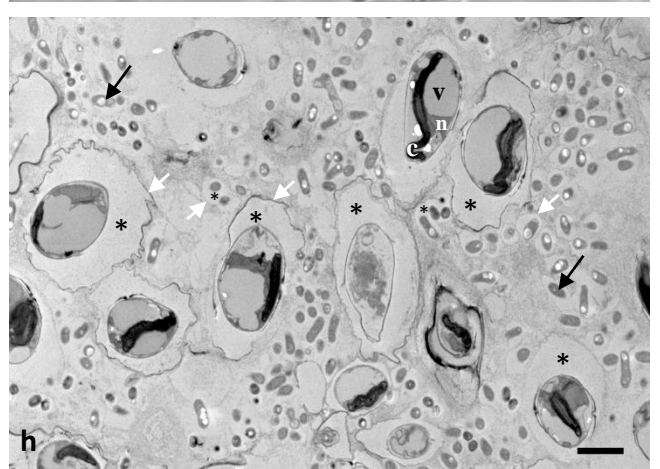
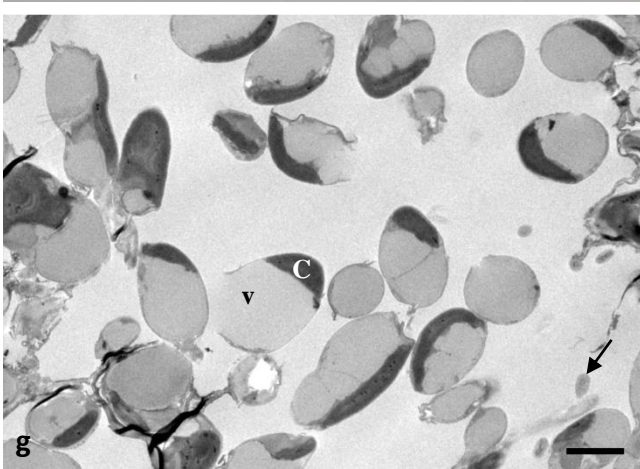
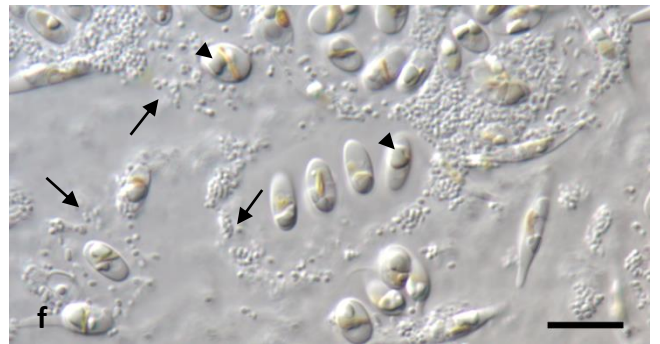
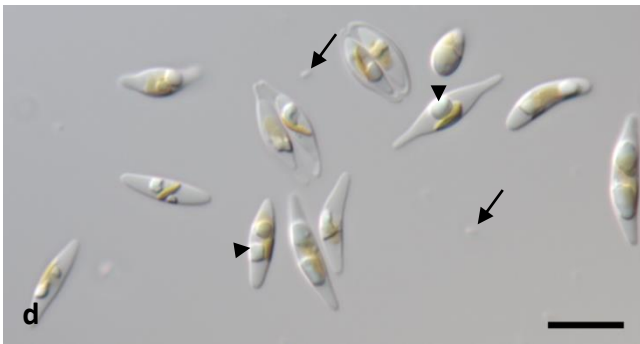
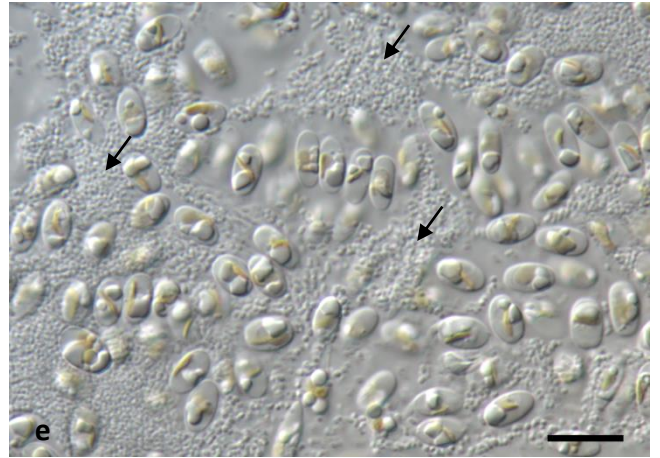
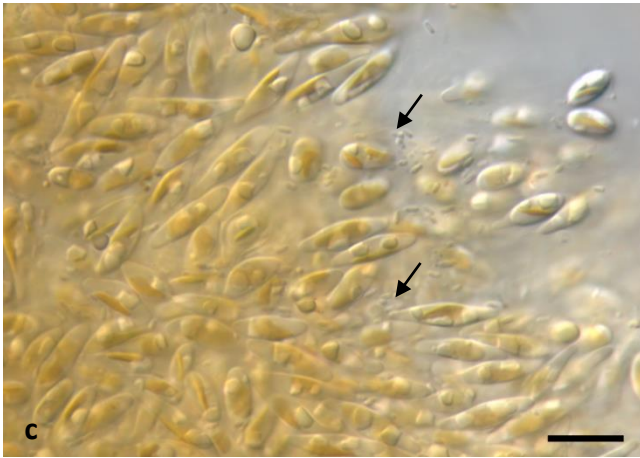
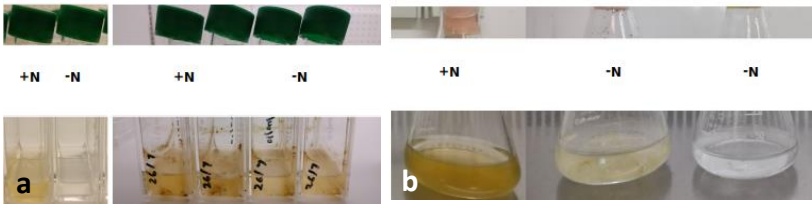
871 **Fig. 4. Functional gene annotation of *Pt15* associated NCDs in nitrate deplete versus**
872 **replete conditions.** a, Relative abundance of selected genes involved in nitrogen fixation and
873 assimilation in nitrate deplete and nitrate replete condition. b, illustration representing a
874 combined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

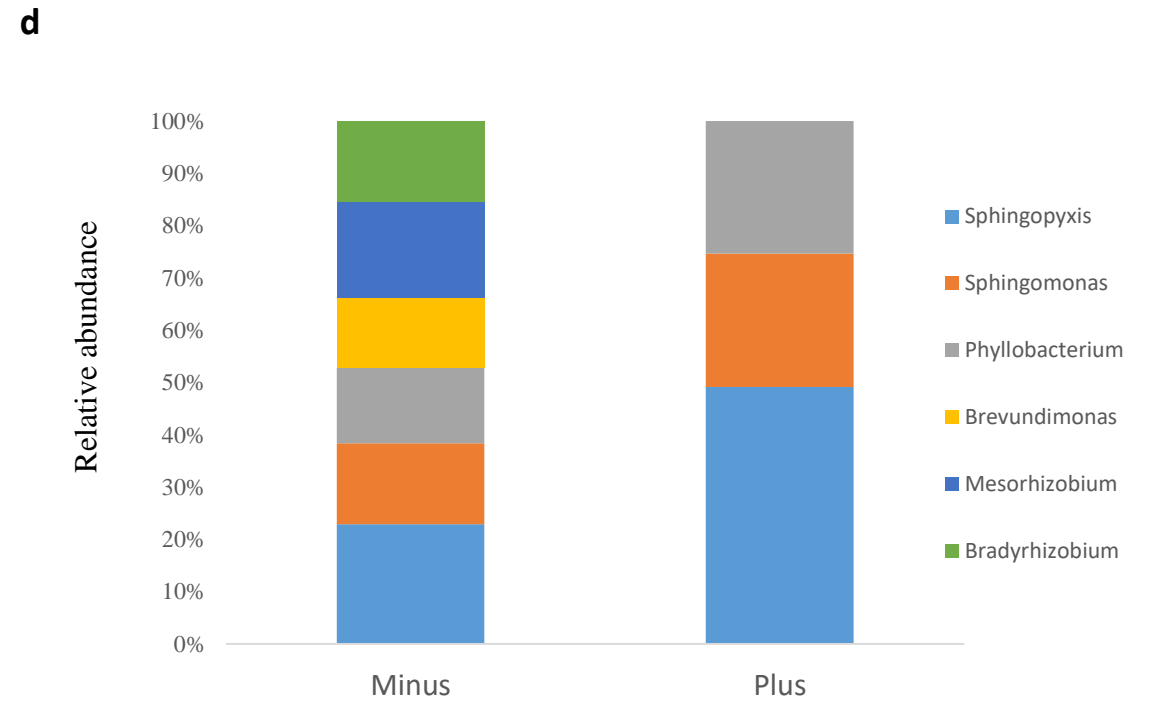
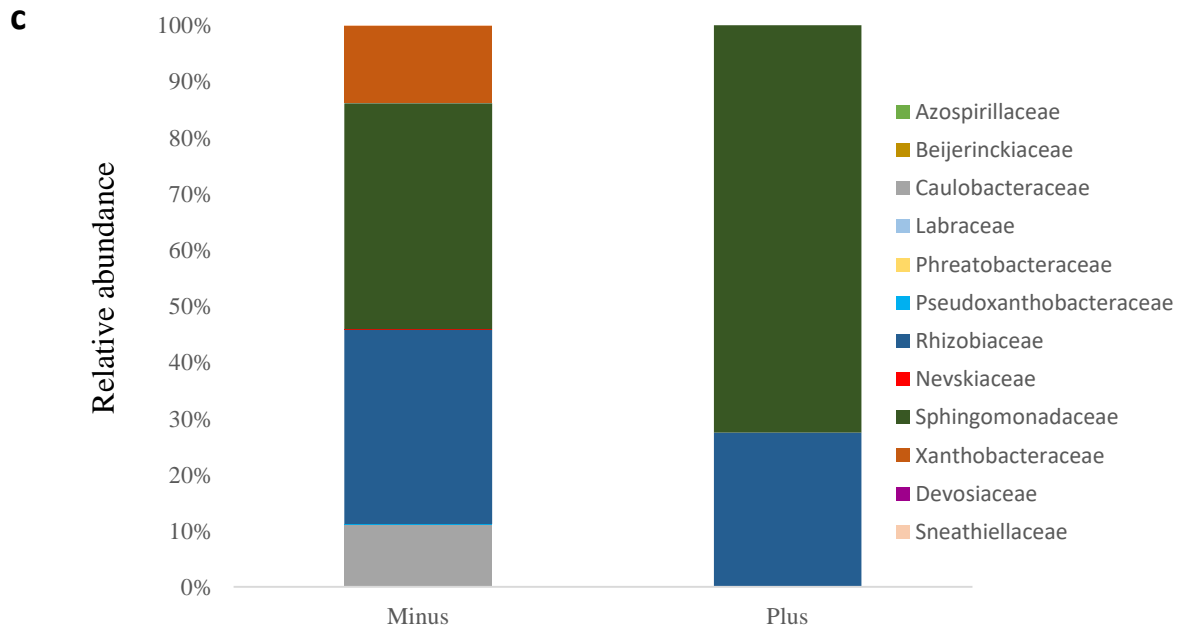
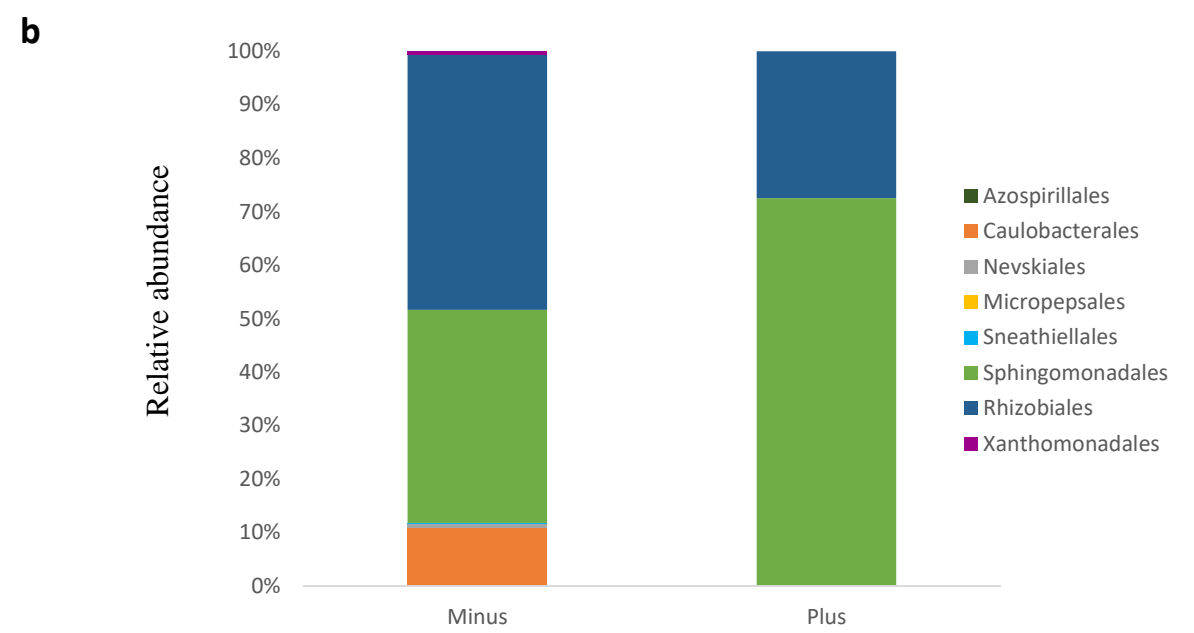
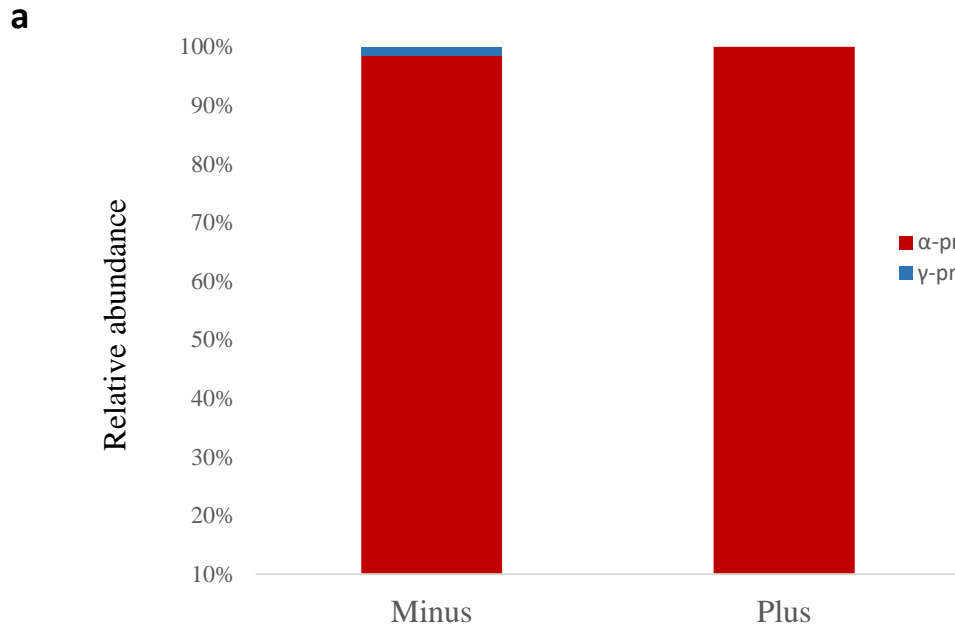
877

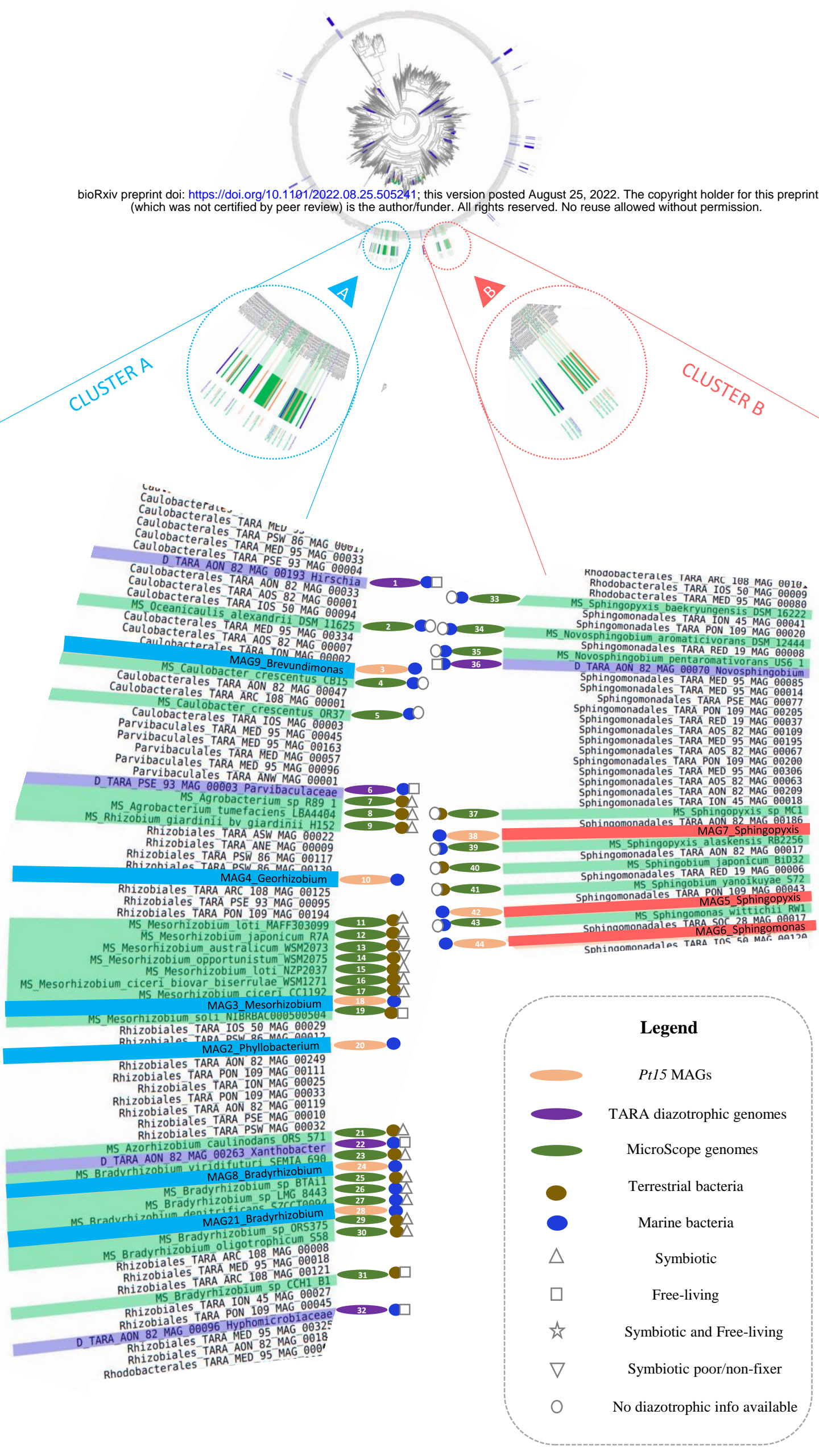
878 **Fig. 5. Geographical distribution and ecological relevance of isolated and assembled**
879 **bacterial diazotrophs genomes correlating to TARA data.** Geographical distribution of
880 phylogenetically closest *Pt15* associated TARA ocean MAGs in a, surface and b, DCM
881 samples. Relative abundance of the top 2 most occurring MAGs of *Pt15* bacterial community
882 (c,d, *Mesorhizobium* MAG3, e, f *Bradyrhizobium* MAG8 respectively) in the environment in
883 respect to nitrate concentrations and nitrate phosphate ratios. g, Co-occurrence of *Pt15* MAGs
884 in the environment with different species of micro-eukaryotes.

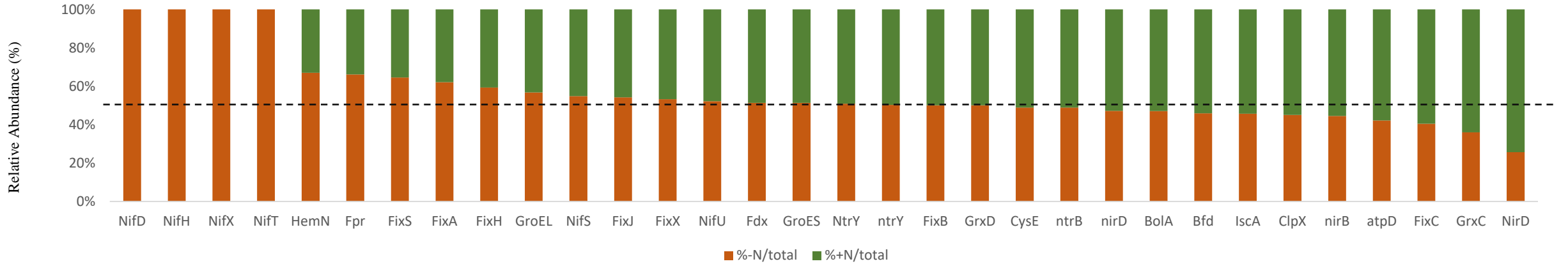
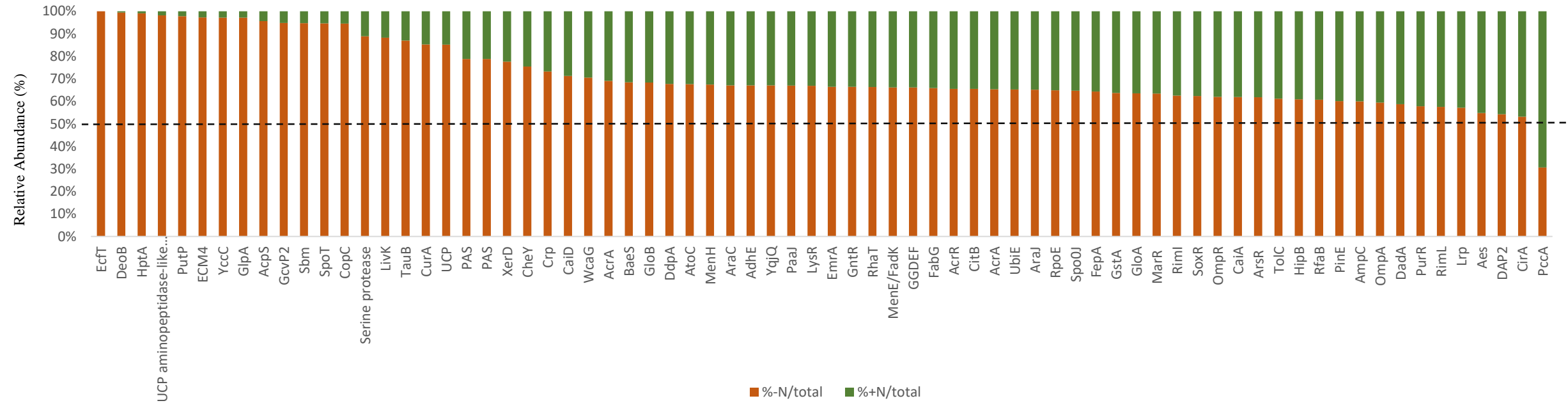
885 **Table 1. List of isolated bacteria from xenic *Pt15* in -N condition.**

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







a**b**

Table_1

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

**nifH* identified

** other nif genes