Taxonomic distribution of metabolic functions underpins nutrient cycling in *Trichodesmium* consortia

Coco Koedooder[1,2,3*], Futing Zhang[1,2], Siyuan Wang[1,2], Subhajit Basu[1,2,4], Sheean T. Haley[5], Nikola Tolic[6], Carrie D. Nicora[6], Tijana Glavina del Rio[7], Sonya T. Dyhrman[5,8], Martha Gledhill[9], Rene M. Boiteau[10,11], Maxim Rubin-Blum[3], Yeala Shaked[1,2]

[1]*The Fredy and Nadine Herrmann Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem, Israel*
[2]*The Interuniversity Institute for Marine Sciences in Eilat, Eilat, Israel*
[3]*Israel Oceanographic and Limnological Research, Haifa, Israel*
[4]*Microsensor Research Group, Max Planck Institute for Marine Microbiology, Bremen, Germany*
[5]*Lamont-Doherty Earth Observatory, Columbia University, New York, NY, United States*
[6]*Biological Sciences Laboratory Pacific Northwest National Laboratory, Richland, WA, United States*
[7]*Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, United States*
[8]*Department of Earth and Environmental Sciences, Columbia University, New York, NY, United States*
[9]*GEOMAR, Helmholtz Center for Ocean Research, Kiel, Germany*
[10]*Environmental Molecular Sciences Laboratory Pacific Northwest National Laboratory, Richland, WA, United States*
[11]*College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, United States*

* *Correspondence: Coco Koedooder (coco.koedooder@mail.huji.ac.il)*

Running Title: Associated Bacteria of Red Sea *Trichodesmium* populations

Keywords: *Trichodesmium*, metagenomes, associated bacteria, siderophore biosynthesis, vitamin B, denitrification, dissimilatory nitrate fixation to ammonia, phosphonate, nitrogen fixation

Abstract

The photosynthetic and diazotrophic cyanobacterium *Trichodesmium* is a key contributor to marine biogeochemical cycles in the subtropical-oligotrophic oceans. *Trichodesmium* forms colonies that harbor a distinct microbial community, which expands their functional potential and is predicted to influence the cycling of carbon, nitrogen, phosphorus and iron (C, N, P, and Fe). To link key traits to taxa and elucidate how community structure influences nutrient cycling, we assessed Red Sea *Trichodesmium* colonies using metagenomics and metaproteomics. This diverse consortium comprises bacteria that typically associate with algae and particles, such as the ubiquitous *Alteromonas macleodii,* but also lineages specific to *Trichodesmium*, such as members from the order Balneolales. These bacteria carry functional traits that would influence resource cycling in the consortium, including siderophore biosynthesis, reduced phosphorus metabolism, vitamins, denitrification, and dissimilatory-nitrate-reduction-to-ammonium (DNRA) pathways. Denitrification and DNRA appeared to be modular as bacteria collectively completed the steps for these pathways. The vast majority of associated bacteria were auxotrophic for vitamins, indicating the interdependency of consortium members. *Trichodesmium* in turn may rely on associated bacteria to meet its high Fe demand as several lineages can synthesize the photolabile siderophores vibrioferrin, rhizoferrin, and petrobactin, enhancing the bioavailability of particulate-Fe to the entire consortium. Our results highlight that *Trichodesmium* is a hotspot for C, N, P, Fe, and vitamin exchange. The functional redundancy of nutrient cycling in the consortium likely underpins its resilience within an ever-changing global environment.

Importance (150 words)

Colonies of the cyanobacteria *Trichodesmium* act as a biological hotspot for the usage and recycling of key resources such as C, N, P and Fe within an otherwise oligotrophic environment. While *Trichodesmium* colonies are known to interact with a unique community of algae and particle-associated microbes, our understanding of the taxa that populate these colonies and the gene functions they encode is still limited. Characterizing the taxa and adaptive strategies that influence consortium physiology and its concomitant biogeochemistry is critical in a future ocean predicted to have increasing particulate fluxes and resource-depleted regions.

Introduction

Trichodesmium spp. are a globally relevant group of cyanobacteria that can form surface blooms visible from space (1). Owing to their high abundance and capability to fix both carbon and nitrogen, they are considered key players in the biogeochemical cycling of C and N within the oligotrophic tropical and subtropical oceans (2–5). A key trait underpinning the functionality of *Trichodesmium* is that its filaments can cluster together to form large colonies of 1-2 mm in size. The formation of colonies allows *Trichodesmium* to capture dust that is deposited on the ocean surface and *Trichodesmium* is reported to actively select for Fe and P-rich particles and shuffle them to their colony core (6–8). Colony formation hereby adds an intriguing spatial component where *Trichodesmium* spp. is predicted to act as a biological hotspot not only for C and N, but also for the uptake and recycling of key limiting nutrients such as Fe and P.

Trichodesmium colonies harbor a unique microbial consortium of epibiotic bacteria. Amplicon sequencing has shown these associated bacteria are distinct from populations in the surrounding oligotrophic surface waters (9–11). At a genetic level, the associated microbial community was found to enrich the total functional potential of the collective colony beyond the traits of N_2 and C fixation (12–14). By enriching the genetic repertoire of metabolic functions, or traits, related to the nutrients C, N, Fe, and P, consortia members were predicted to influence the internal cycling of these nutrients (14–17). To understand the biogeochemical contributions of *Trichodesmium* within the global ocean it is therefore critical to look beyond the metabolic functions present in *Trichodesmium* and towards the processes taking place within the larger microbial community. While metagenomic studies of *Trichodesmium* colonies in the Atlantic (13) and the Pacific Oceans (12, 17, 18) highlighted key functional traits within the consortium, only one study linked pathways to the associated bacteria from 8

56 metagenome-assembled bins (13). Expanding these studies, and characterizing how functional
57 traits are distributed across different taxa is critical to understanding the role of *Trichodesmium*
58 associated bacteria in consortium physiology and its concomitant influence over the cycling of
59 key resources such as C and N.
60
61 Recent evidence from the Red Sea highlights the unique ability of *Trichodesmium* consortia
62 members to acquire particulate Fe from dust through the production of Fe-chelating
63 siderophore molecules (7, 8, 16, 19, 20). Siderophores dissolve particulate Fe from dust and
64 are predicted to increase its bioavailability for both *Trichodesmium* and its associated bacteria
65 (16, 21, 22). In this manner, *Trichodesmium* spp. interacts with their associated bacteria in
66 order to meet the high Fe-requirements for the processes of photosynthesis and N_2 fixation (23).
67 This finding is particularly relevant in light of climate model simulations, which predict an
68 increase in regional dust emissions in areas such as the Red Sea where *Trichodesmium* is known
69 to occur (24, 25). While *Trichodesmium* blooms are a re-occurring phenomenon in this area
70 (26), it represents an under sampled region where the taxonomic composition of its consortium
71 has not been explored genetically. Furthermore, the presence and taxonomic distribution of
72 siderophore biosynthesis pathways within *Trichodesmium* colonies has not yet been
73 interrogated in depth. Understanding which consortium members positively enhance the
74 bioavailability of Fe within the *Trichodesmium* consortium will help elucidate key interactions
75 and community dynamics taking place within the colony that influence the ability of
76 *Trichodesmium* to fix C and N. This information is critical to evaluating *Trichodesmium*
77 contributions to C and N cycling in the future ocean.
78
79 In this study, we conducted an in-depth metagenomic and proteomic study of *Trichodesmium*
80 colonies collected from the Red Sea. To elucidate how the community structure influences the

81 functional dynamics of the *Trichodesmium* consortium we identified the different taxa making
82 up the *Trichodesmium* consortium and examined the presence of functional traits involved in
83 nutrient cycling. We probed our dataset for genes, proteins and pathways involved in the uptake
84 and metabolism of N, P, Fe, and vitamins. Finally, we contextualize these findings by
85 comparing them to previous metagenomic studies of *Trichodesmium* colonies obtained from
86 the Atlantic and the Pacific oceans.
87

88 Results and Discussion

89 A total of 52 metagenomic-assembled genomes (MAGs) were obtained from *Trichodesmium*
90 colonies, of which 42 MAGs were of high quality (90% complete and <5% redundancy
91 (Supplementary Table 1) (27). The relative abundance of the different MAGs was consistent
92 across the three samples with *T. thiebautii* MAG 52 making up 71% ± 6 of the total relative
93 abundance (Supplementary Table 1). In parallel to the metagenomic analysis, a proteomic
94 dataset consisting of 14161 peptide sequences matching to protein sequences of 52 different
95 MAGs further probed the activity within the Red Sea *Trichodesmium* consortia
96 (Supplementary Table 2; Supplementary Table 3).

97 **Figure 1.** Phylogenetic tree of the 52 MAGs assembled from *Trichodesmium* colonies collected in the Red Sea.
98 The tree is rooted by the outgroup *Fuseobacterium nucleatum*. Known bacterial orders are marked in different
99 colours and several MAGs could taxonomically be identified at the genus level. The relative abundance of each
100 MAG within the *Trichodesmium* consortium, excluding the abundance of Trichodesmium MAG 52 (±70% of
101 samples), is presented as a bubble chart. The 11 MAGs that matched to MAGs assembled from a previous
102 metagenomic dataset of Red Sea *Trichodesmium* colonies (28) are highlighted in blue and tentatively represent
103 the core-consortium (see Supplementary Figure 2; Supplementary Table 4). MAGs that did not meet the
104 requirements of a high-quality genome (>90% completeness; <5% redundancy) are indicated in grey. Note that
105 re-occurring MAGs are not necessarily the most abundant MAGs present within the Trichodesmium colony. (a
106 full taxonomic description of each MAG in this dataset can be found in Supplementary Table 1).
107

108 *Trichodesmium* colonies host diverse and flexible associated bacteria

109 *Trichodesmium* colonies consisted of a single *Trichodesmium* species (*T. thiebautii* MAG 52)
110 and a diverse consortium of 51 associated bacteria spanning at least 10 different taxonomic
111 orders (Figure 1). Unlike a previous study (28), our samples did not include any non-

112 diazotrophic *Trichodesmium* populations and the 51 MAGs hereby represent the associated
113 bacteria of *T. thiebautii* MAG 52. The majority of the associated bacterial MAGs belonged to
114 taxonomic orders known to be present in *Trichodesmium* colonies and included the orders
115 Rhodobacterales, Pseudomonadales, Balneolales, Rhodospirillales, Flavobacterales,
116 Enterobacterales, Rhizobiales, and Sphingomonadales (9, 11–13).
117
118 An intercomparison between the MAGs observed here and a previous metagenomic dataset
119 from the Red Sea (28) , identified a total of 11 MAGs present in both datasets (Supplementary
120 Figure 2; Supplementary Table 4). These 11 MAGs included the particle- and algae-associated
121 lineages (29–31) of *Marinobacter* sp. (MAG23) and *Alcanivorax* sp. (MAG 47) from the order
122 Pseudomonadales; *Rhodobacter* sp. (MAG 28) and *Sulfitobacter* sp. (MAG29) from the order
123 Rhodobacterales, and *Alteromonas macleodii* (MAG 10), a ubiquitous bacterium often found
124 in association with phototrophs (32) and previously isolated from *T. erythraeum* IMS101
125 cultures (33). If these observations are consistent across a larger time-period of observations,
126 these MAGs may represent a core component of the *Trichodesmium* consortium. Further
127 comparison of our metagenomic dataset with those obtained from the Atlantic, North or South
128 Pacific oceans, and the Red Sea (12, 13, 17, 18, 28), showed taxonomic similarities at the order
129 level rather than at the species level. While most taxonomic orders re-occurred in all datasets,
130 only one Balneolales genome (MAG 33) was present amongst all *Trichodesmium* metagenomic
131 datasets. As none of the re-occurring MAGs represented highly abundant genomes within our
132 samples (Figure 1), this may partially be the result of a higher sequencing depth (500 million
133 reads) compared to past datasets. Nonetheless, the wide diversity coupled with the recurrence
134 of only a few MAGs indicates the presence of a flexible assemblage of associated bacteria,
135 which may have important implications for the community structure and functioning of the
136 *Trichodesmium* consortium.

137

138 ***Trichodesmium associated bacteria synthesize diverse photolabile siderophores, which may***

139 ***increase iron availability from dust***

140 We explored the different Fe-uptake mechanisms present within the *Trichodesmium*

141 consortium, as N_2 and C fixation in *Trichodesmium* is often constrained by Fe availability (23,

142 34). Despite the high Fe-requirements of *Trichodesmium* (23), *T. thiebautii* (MAG 52) only

143 contained genes for the uptake of free Fe^{+2} and Fe^{+3}, but not organic Fe, suggesting it is limited

144 to a narrow range of forms it can directly take up (Figure 2). In contrast, MAGs of the

145 associated bacteria contained a larger diversity of Fe-uptake genes which included the ability

146 to take up organically bound Fe, such as heme, citrate, and siderophores (Figure 2). The

147 proteome of the *Trichodesmium* consortium indicated that, at the time of sampling,

148 *Trichodesmium* was actively taking up Fe through the presence of a ferric Fe-uptake

149 transporters (FbpC, FbpA) and Fe-storage proteins (Dps, Bfr) (Supplementary Table 2).

150

151 The genomes of 10 *Trichodesmium* associated bacteria spread across 6 different taxonomic

152 orders encoded different siderophore biosynthesis gene clusters (BGCs) (Figure 2). The

153 presence of a siderophore biosynthesis pathway was strain-specific and not present in an entire

154 clade or lineage. Similar to previous studies, *T. thiebautii* MAG 52 did not contain any known

155 siderophore biosynthesis pathways (12, 35). Siderophore BGCs could further be separated into

156 4 NIS-type and 6 NRPS-type siderophores (36, 37). Based on the presence of one or more

157 *iucA/iucC* domains, the 4 NIS-types BGCs matched closely to petrobactin (*Alteromonas*

158 *macleodii* MAG 10), vibrioferrin (*Marinobacter salarius* MAG 23; *Limnobacter sp.* MAG 11),

159 and rhizoferrin (Rhodospirillales MAG 08) biosynthesis pathways (Figure 2). From the 6

160 different NRPS-type siderophores, only 2 matched a known BGC, namely that of the

161 membrane-bound amphibactin (100%) (*Alcanivorax venustensis* MAG 47) and one resembling

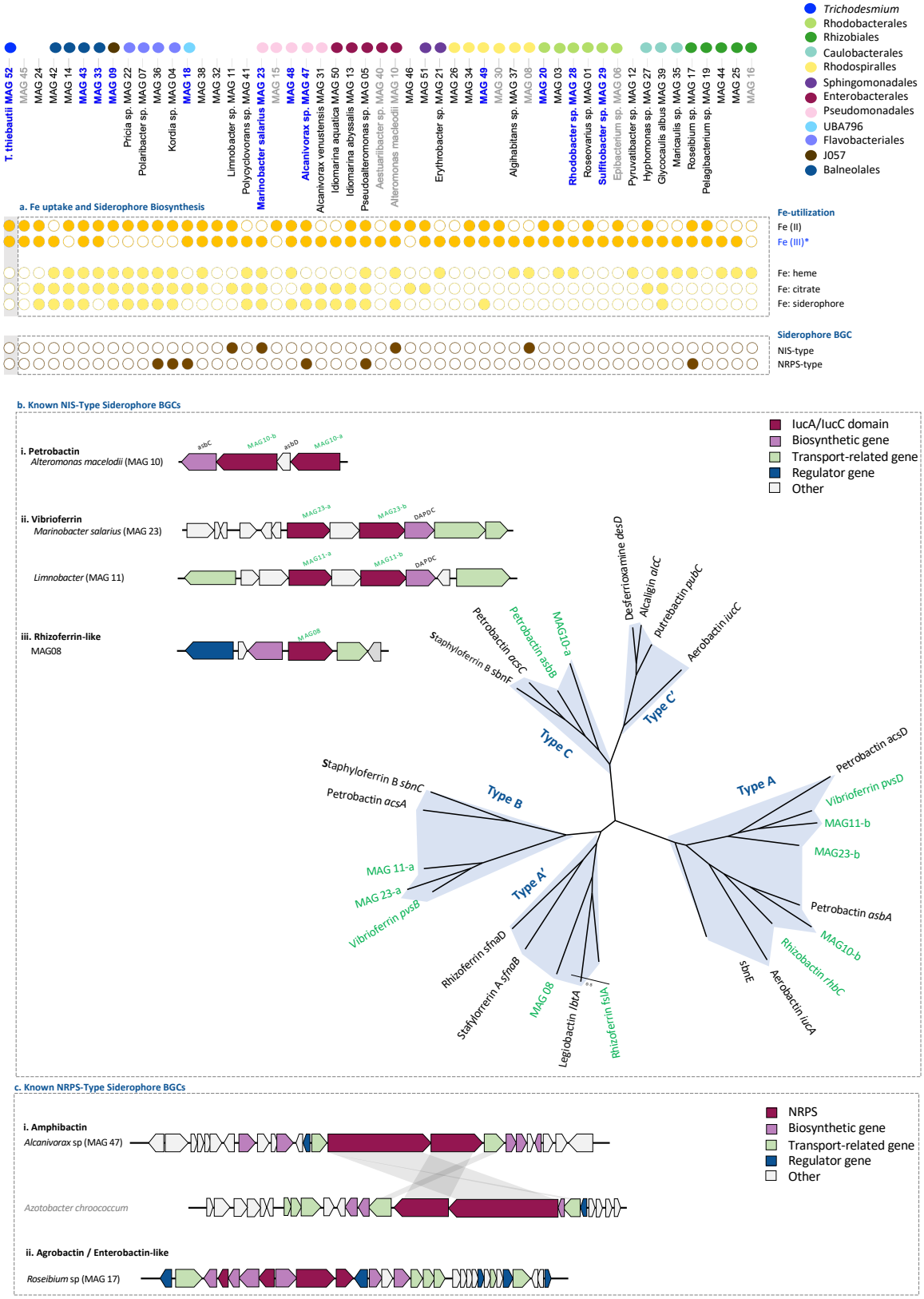
162 the BGC of the siderophores agrobactin (93%) or enterobactin (±83%) (*Roseibium* sp MAG
163 18) (Figure 2). Our results additionally indicated the presence of several novel NRPS-type
164 siderophore biosynthesis pathways (Supplementary Figure 3). NRPS-type-siderophore BGCs
165 are not as well characterized as NIS-type-siderophores BGCs and can be difficult to distinguish
166 from other secondary metabolites such as toxins and antibiotics (38). While the 4 NRPS-type-
167 siderophore BGCs could not be linked to known siderophore BGCs, several unknown
168 metallophores have previously been identified and associated with *Trichodesmium* colonies
169 and we speculate that some may be linked to these uncharacterized NRPS BGCs (19). The
170 presence of multiple NRPS pathways containing a putative siderophore TonB-dependent
171 receptor (SMCOG1082) likely involved in Fe uptake, highlights the potential for novel
172 siderophore pathways within the *Trichodesmium* consortium and indicates the importance of
173 additional culture-based research linking such pathways to a siderophore molecule.
174
175 Our results indicate that the *Trichodesmium* consortium in the Red Sea may interact with
176 photolabile siderophore-producing bacteria to mine Fe from dust. The re-occurring consortium
177 members *Marinobacter salarius* (MAG 23), *Alcanivorax venustensis* (MAG 47), and
178 *Alteromonas macleodii* (MAG 10) are all able to synthesize the NIS-type siderophores
179 petrobactin, vibrioferrin, and rhizoferrin (Figure 2) whose citrate-moiety results in their
180 photolability (39). In the natural illumination of surface waters, where *Trichodesmium* resides,
181 sunlight can reductively disassociate Fe from these siderophore complexes and provide a flux
182 of highly available dissolved inorganic Fe (40, 41). While *Trichodesmium* does not contain a
183 TonB-receptor, the presence of photolabile siderophores implies that Fe is also bioavailable to
184 consortium members that lack a specific TonB-receptor for that particular siderophore. These
185 photolabile siderophores act as a "common good" within the *Trichodesmium* consortia, where
186 only a few members carrying the trait are needed to benefit the entire consortium (42). This

contrasts with the uptake mode of photostable siderophores, such as desferrioxamine-B (DFOB), which require a specialized TonB-siderophore transport system (43). Similarly, amphibactin is a membrane-bound siderophore (44), which prevents its diffusion in the environment and subsequently its ability to act as a public good to the *Trichodesmium* consortium. A recent gene-knockout study in *Alteromonas macleodii* verified the ability of petrobactin to enhance the bioavailability of particulate Fe (22). This further supports the idea that *Trichodesmium* can interact with consortium members that produce photolabile siderophores to acquire Fe from dust.

Most of the siderophore BGCs detected here were identified in ubiquitous and copiotrophic bacteria typically found in association with particles and algae (11, 22) , including *Marinobacter salarius* (MAG 23), *Alcanivorax venustensis* (MAG 47), and *Alteromonas macleodii* (MAG 10) (Figure 2). Siderophore-production within the *Trichodesmium* consortium appears to be linked to a lifestyle associated with a rapidly changing environment rich in N and C, where regulating Fe-bioavailability can provide a competitive advantage. In line with particle and algae associations, siderophore production has also been associated with other ecological functions, including heavy metal detoxification and intra-specific communication (45) (Supplementary Figure 4; Supplementary Table 5). Whether siderophore production can benefit the *Trichodesmium* consortium in functions other than Fe-dissolution from dust requires further investigation. Nonetheless, the presence of a few, yet reoccurring, siderophore producing MAGs augments previous measurements of siderophore production within the consortium (16, 19, 20) and reveals additional layers of *Trichodesmium* colonies' intriguing ability to sequester Fe from dust.

Figure 2 (below). MAGs were probed for Fe-uptake and siderophore biosynthesis genes which are noted in Supplementary Table 5. **(a)** *Trichodesmium* MAG 52 does not contain a wide diversity of Fe-uptake genes, and the majority of MAGs of associated bacteria contain an uptake receptor for organic Fe (heme, citrate, siderophores). Several MAGs of associated bacteria contained an NRPS or NIS-type siderophore biosynthesis pathways (BGCs). **(b)** Siderophore biosynthetic gene clusters (BGCs) of known NIS-type siderophore BGCs can

be clustered according to the presence of one or multiple *iuc*A/*iuc*C domains which enables one to target the closest siderophore BGC match (52). Using this method, we were able to link these pathways to the synthesis of the photolabile **(i)** petrobactin, **(ii)** vibrioferrin and **(iii)** rhizoferrin. **(c)** Siderophore BGCs of known NRPS siderophores. Using antiSMASH 7.0, which aligns BGCs to known siderophores within the MIBiG database, we were able to link 2 of the 6 putative siderophore NRPS-type pathways to **(i)** the membrane-bound amphibactin and **(ii)** agrobactin/enterobactin. (See Supplementary Figure 3 for more details on the siderophore biosynthesis pathways).



223

The *Trichodesmium* consortium is heterogeneous in genes related to the uptake and use of inorganic and organic phosphorous.

We predicted that the majority of bacteria associated with *Trichodesmium* have the ability to metabolize phosphite or phosphonate. This can offer a key competitive advantage amongst consortium members in a P-limited environment (18, 46) including the Gulf of Aqaba in the Red Sea (47). Within a low-P environment, dissolved inorganic phosphorus is scarce, and bacteria must scavenge phosphate from dissolved organic phosphorus to meet their P-requirements (48). *Trichodesmium* is highly efficient at scavenging and utilizing organic-P through the production of alkaline phosphatase (AP), and the uptake of phosphonate and phosphite (49). Multiple studies have shown that *Trichodesmium* colonies in the Atlantic and Pacific Ocean are hotspots for the reduced P-cycling of phosphite or phosphonate (46, 50) suggesting that the majority of consortium members can process reduced-P.

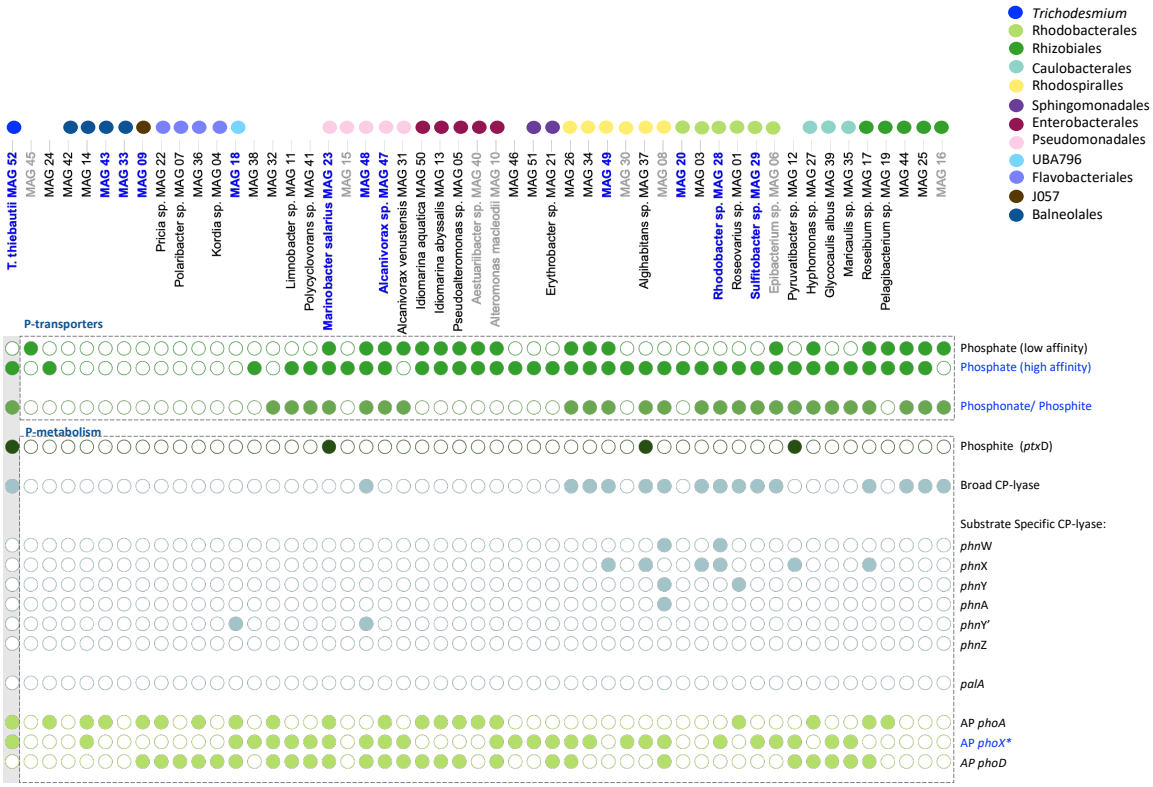


Figure 3. MAGs were probed for several different P-metabolism genes which are noted in Supplementary Table 5. *Trichodesmium* is known to be highly competitive for P, as exemplified by the multitude of **(a)** P-uptake systems present in its genome for phosphate (*pst*S, *pst*B), phosphonate (*phn*CDE), and phosphonate/phosphite (*phn*CDE/*ptx*ABC). **(b)** Within the *Trichodesmium* consortium, the ability to metabolize a broad range of phosphonate using a broad C-P lyase (*phn*GHIJKLM) is more prominent than the substrate specific hydrolysis of

242 2-aminoethylphosphonate (2-AEP; *phnWXYZA*) or phosphonopyruvate (*pal*A) and the phosphite metabolism
243 (*ptx*D).
244
245 Confirming past studies (18, 51), *T. thiebautii* MAG 52 encodes a set of functions to transport
246 and metabolize diverse forms of P (Figure 3). This includes the presence of a high-affinity
247 phosphate transporter (*pst*SCAB), and a phosphonate/phosphite transporter
248 (*phn*CDE/*pxt*ABC). In addition to these transport systems, genes related to utilization and
249 metabolism of organic-P included alkaline phosphatase (AP) (*pho*A, *pho*X), and a carbon-
250 phosphorus (C-P) lyase (*phn*GHIJKLM) which hydrolyses a broad range of phosphonate
251 compounds. *T. thiebautii* MAG 52 also encodes phosphite dehydrogenase (*ptxD)* which
252 catalyzes the oxidation of phosphite to phosphate. The presence of PstB, PhoX, and PhnD
253 within the proteome (Supplementary Table 2), suggests *Trichodesmium* was taking up
254 phosphate, phosphonate, and hydrolyzing phosphoester compounds to meet its P demand.
255 Several of these proteins are regulated by P (51), and their detection in the proteome is
256 consistent with intense competition for P in the low P Red Sea.
257
258 As KEGG annotation cannot accurately distinguish between a phosphonate (*phn*CDE) and a
259 phosphite (*pxt*ABC) transporter, the presence of adjacent P-metabolism genes can further
260 indicate the P-source linked to it. Specifically, the presence of phosphite dehydrogenase (*ptxD)*
261 points to phosphite metabolism while a broad-specificity C-P lyase (*phnGHIJKLM*) or the
262 substrate-specific hydrolysis of phosphonopyruvate (*pal*A) or 2-aminoethylphosphonate (2-
263 AEP) to acetaldehyde (*phn*WX), acetate (*phn*WYA), or glycine (*phn*Y'Z), can point to the
264 metabolism of phosphonate (52, 53). Our results showed that the presence of a
265 *phn*CDE/*ptx*ABC uptake system was specific to certain taxonomic orders including
266 Pseudomonadales, Rhodospirillales, Rhodobacterales, Caulobacterales, and Rhizobiales
267 (Figure 3). Within the *Trichodesmium* consortium, these taxonomic orders may have a
268 competitive advantage over members that lack them, particularly in low P conditions. Of these

269 25 MAGs, 16 could be linked to phosphonate metabolism through the presence of a broad-
270 specificity C-P lyase. In contrast, the presence of phosphite dehydrogenase was species-
271 specific where only 4 MAGs appeared to be able to process phosphite spread across different
272 taxonomic orders (Figure 3). Within the *Trichodesmium* consortium, phosphonate metabolism
273 appears to be the more common reduced P-currency of the two. 7 MAGs contained genes
274 associated with the substrate-specific hydrolysis of 2-AEP but pathways were not complete,
275 making it difficult to draw conclusions (Figure 3). Nonetheless, the presence of a broad-
276 specificity C-P lyase appeared to be the most prevalent strategy for phosphonate utilization
277 within the *Trichodesmium* consortium. This finding is in contrast to a meta-analysis of marine
278 metagenomes where the most common phosphonate degradation strategy within the water
279 column was the substrate-specific hydrolysis of 2-AEP (54). 2-AEP is the most abundant
280 phosphonate source in the marine environment (55), and the differences in enzyme distribution
281 within the Trichodesmium consortium may reflect heterogeneity in the phosphonate
282 compounds present in the colony, relative to the water column. Differences in enzyme
283 distribution may also be a function of the taxonomic composition of *Trichodesmium* associated
284 bacteria relative to those in in the water column. Regardless, these observations suggest that a
285 diversity of P compounds are being cycled by members of the consortia in the low P Red Sea
286 and offer a new reference dataset for future comparisons with *Trichodesmium* consortia from
287 other regions.
288

289 **Denitrification and dissimilatory nitrate reduction to ammonia (DNRA) pathways are**
290 **modular in the *Trichodesmium* consortium.**

291 We explored the presence of multiple N-metabolic pathways as they are predicted to influence
292 the biogeochemical contribution of *Trichodesmium* regarding N_2 fixation. Our results showed
293 that, in the *Trichodesmium* colonies sampled from the Red Sea, N_2-fixation appears to be

294 uniquely attributed to *T. thiebautii* MAG 52 (Figure 4). This contrasts with a previous study
295 that has observed *nifH* genes in both *Trichodesmium* and their associated bacteria (12). In our
296 study, however, only *T. thiebautii* (MAG 52) contained the full N_2 fixation pathway
297 (*nifDKHEB*: M00175) and therefore the ability to fix N_2 within the colony. Similar to other
298 studies (12, 17, 56), genes of the two additional N-transformation pathways of dissimilatory
299 nitrate reduction to ammonium (DNRA) and denitrification were present within the
300 *Trichodesmium* consortium (Figure 4). Respectively, denitrification and DNRA are N-loss and
301 N-recycling processes and are predicted to enable the residing community to conserve and fully
302 utilize the N_2 fixed by *Trichodesmium* (12) but also alter the net contribution of *Trichodesmium*
303 colonies to local N-cycling dynamics (15, 57). Our results indicated a relatively high number
304 of 24 MAGs encoding genes involved in DNRA and 10 MAGs encoding genes involved in
305 denitrification (Figure 4). Intriguingly, none of the MAGs analyzed contained the complete
306 pathways for denitrification and DNRA, with the exception of Rhizobiales for the latter (Figure
307 4). Collectively, MAGs were able to complete all steps within either pathway, indicating that
308 these pathways appear to be modular within the *Trichodesmium* consortium, where a bacteria
309 can perform one step within the pathway but not necessarily all steps. While Rhizobiales (MAG
310 16, 17 and 44) contained the full DNRA pathway, which includes the reduction of nitrate to
311 nitrite (*napAB, narGH*), and nitrite to ammonium (*nrfAH, nirBD*), the remaining MAGs
312 contained only one of two steps (Figure 4). Similar to DNRA, several MAGs from the phyla
313 Alphaproteobacteria contained a part of the denitrification steps, including the reduction of
314 nitrite to nitric-oxide (*nir*KS) and nitric-oxide to nitrous oxide (*nor*BC), while members of the
315 phyla Bacteroidetes could perform the final step within the denitrification pathway (*nosZ)*
316 which converts nitrous oxide to nitrogen (Figure 4).

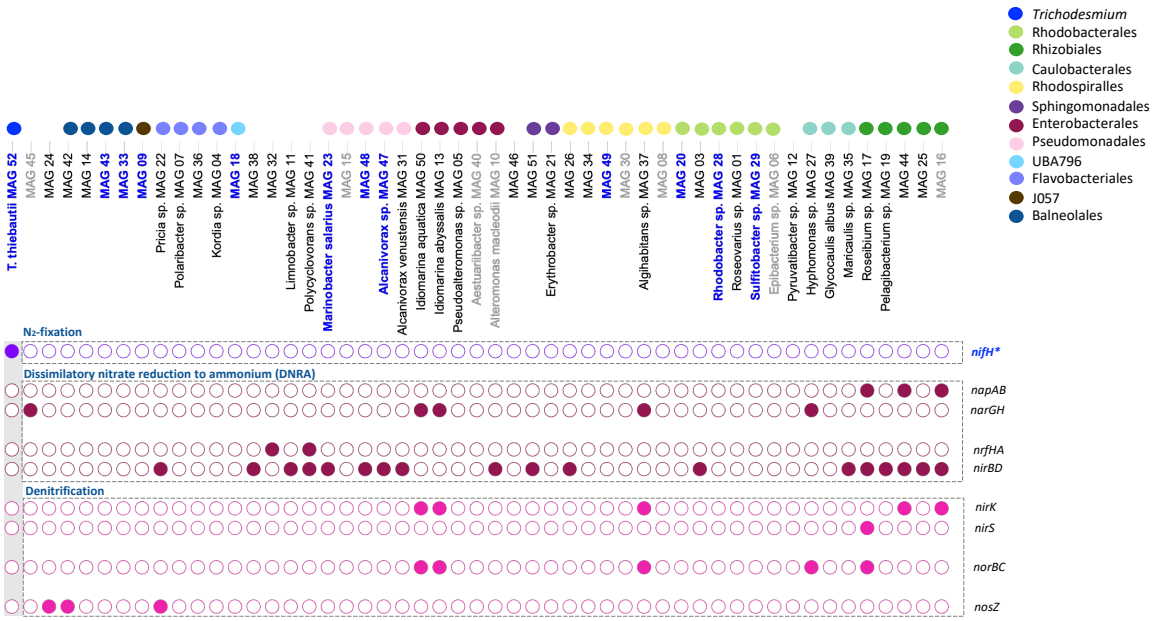


Figure 4. MAGs were probed for N-metabolic pathways and the genes for each step within the pathway are noted in Supplementary Table 5. *Trichodesmium* MAG 52 is the only MAG with a N_2 fixation pathway. Several MAGs contained one or two steps within the pathways of denitrification or dissimilatory nitrate to reduction to ammonia (DNRA) indicating their modularity. Collectively, MAGs completed the denitrification and DNRA pathways, but none contained the genes for the entire pathway.

While the presence of modularity in denitrification and DNRA pathways itself is not uncommon (58–60) finding it present within the *Trichodesmium* consortium highlights the spatial role *Trichodesmium* plays in providing a platform where each step can be performed by a distinct subset of the microbial consortium. While the re-occurring consortium members *Alteromonas macleodii* (MAG 10), *Marinobacter salarius* MAG 23), and *Alcanivorax sp* (MAG 47, MAG 48) encoded genes (*nir*BD) within the DNRA pathway, it did not appear to be a key characteristic within the *Trichodesmium* associated bacteria of the Red Sea. The presence of DNRA and denitrification pathways is rare in the oxic surface ocean, as these processes require anoxic and suboxic conditions (61). In the surface waters, suboxic as well as anoxic micro-environments were detected in marine aggregates, such as marine snow and fecal pellets, where enhanced microbial respiration results in the consumption of O_2 (62, 63). *Trichodesmium* colonies may represent an additional environment where such processes may be taking place within the surface waters of the ocean. While it is not entirely clear when and where these N-cycling pathways are actively taking place within the colony studies have

338 measured DNRA and denitrification processes within natural *Trichodesmium* colonies (15, 57).
339 While the obligatory suboxic conditions for denitrification and DNRA may occur in the
340 presence of dust and the rapid O_2 consumption by bacteria (64), the formation of these
341 microenvironments have thus far not been observed in *Trichodesmium* colonies (57).
342 Altogether, these findings pinpoint the intricate community structure needed for such modular
343 processes to take place within the consortium. The exact tradeoffs between *Trichodesmium* and
344 denitrifying or DNRA bacteria, particularly in light of a changing ocean with higher dust
345 deposition requires further investigation.
346

347 **_Trichodesmium_ consortium members are interdependent for vitamins B1, B7 and B12.**

348 We explored the presence of vitamin auxotrophy within the consortium as an indicator for
349 community interdependence. Previous studies have shown that *Trichodesmium* can synthesize
350 and secrete the vitamin cobalamin (B12) to the environment (17). Our results show that while
351 *Trichodesmium* MAG 52 encodes the biosynthesis pathways for B12, biotin (B7) and thiamin
352 (B1), the majority of associated bacteria (49 of the 52 MAGs) lack a biosynthesis pathway for
353 at least one of these vitamins (Figure 5). Bacteroidetes, Sphingomonadales, and
354 Enterobacterales species, including *Alteromonas macleodii* (MAG 10), were all auxotrophic
355 for vitamin B12. MAGs from the order Rhodospirillales lacked the ability to synthesize vitamin
356 B7, whereas Rhizobiales and Rhodobacterales appeared to be auxotrophic for both vitamin B1
357 and B7. Only *Pseudoalteromonas* encoded all three biosynthesis pathways, including the re-
358 occurring taxa *Alcanivorax* sp. (MAG 47) and *Marinobacter salarius* (MAG 23). These
359 consortium members may also contribute to the synthesis and provision of vitamins B1, B7,
360 and B12 to their auxotrophic counterparts (Figure 5) but, based on overall relative abundance,
361 *Trichodesmium* represents the most likely source of B vitamins. In sum, the majority of
362 associated bacteria (49 of 52 MAGs) were predicted to be auxotrophic for B12 and likely

dependent on *Trichodesmium* to meet their intracellular requirements. Our results indicate that the production and secretion of different vitamins by *Trichodesmium* is a key determinant of community composition, stimulating vitamin-based interactions among auxotrophic consortium members.

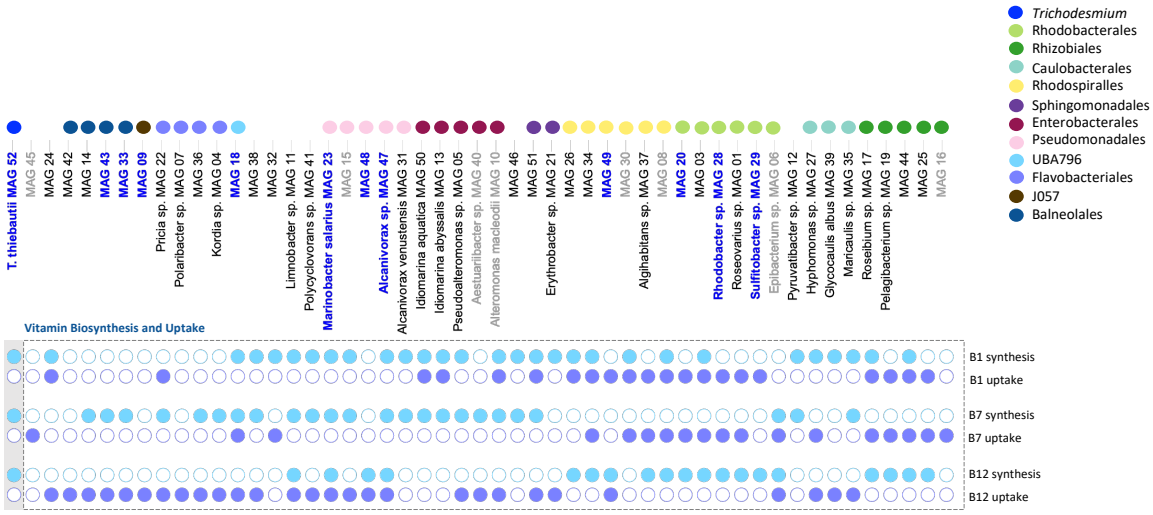


Figure 5. MAGs were probed for vitamin B1, B7, and B12 biosynthesis and uptake pathways. The genes used to indicate the biosynthesis or uptake of these vitamins are noted in Supplementary Table 5. *Trichodesmium* contains the biosynthesis pathways for all three vitamins, but the vast majority of associated bacteria were auxotrophic for one or more vitamins. These results suggest that the associated bacteria rely on *Trichodesmium* to obtain their vitamin requirements.

Interactions between copiotrophic microbes underpin nutrient cycling in *Trichodesmium* consortia

A literature study of the bacterial lineages present within the *Trichodesmium* consortium describes these taxonomic orders as typical primary colonizers of marine surfaces such as algae or sinking particles (29–31). Strains of *Alteromonas*, *Marinobacter,* and *Pseudomonas* often co-occur with phototrophs such as cyanobacteria as well as oil spills, likely as a result of their ability to degrade hydrocarbons and aromatic compounds (32, 65). While the prevalence of these copiotrophic lifestyles within the *Trichodesmium* consortium has been observed in previous studies (11, 13) it is still unclear whether their association with *Trichodesmium* is specific or reflective of a more general interaction between algae or particle associations. Copiotrophy is a widely encompassing term, but it generally concerns opportunistic bacteria

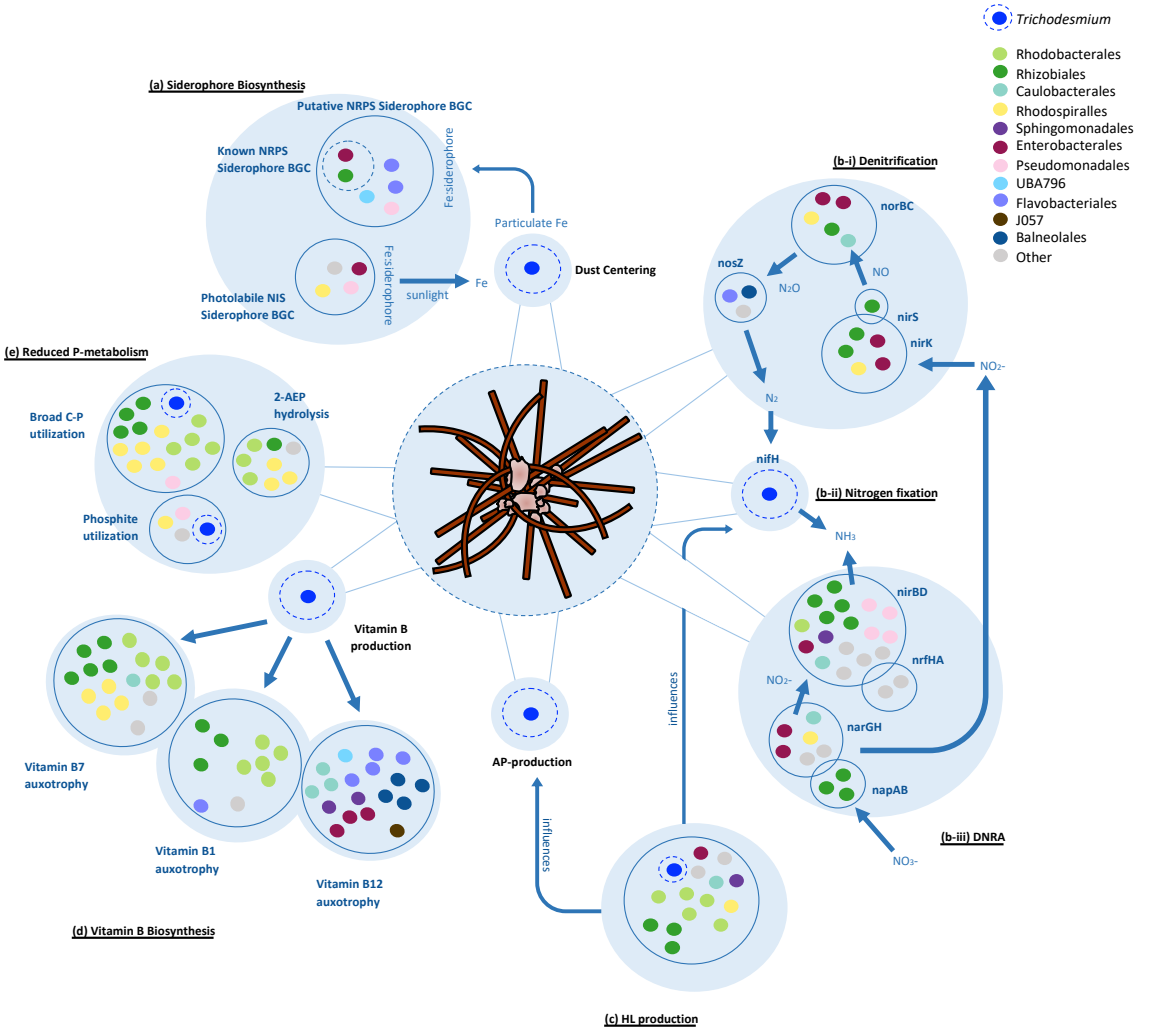
385 with a versatile metabolic capacity to use a wide range of substrates, acting as important
386 remineralizers by rapidly responding towards areas enriched in organic matter (66, 67). It is
387 therefore possible to consider the majority of the consortium as active remineralizers of the C,
388 N, P, Fe, and vitamins *Trichodesmium* is providing. Viewing *Trichodesmium* as a hotspot for
389 nutrients is supported by our proteomic samples where the majority of protein hits attributed
390 to *T. thiebautii* MAG 52 were related to photosynthetic (K05376; K05377; K02284) and
391 diazotrophic machinery (K03839, K02588, K02591) indicating that the processes of N and C
392 fixation were actively taking place at the time of collection (Supplementary Table 2). In turn,
393 the majority of protein hits associated to consortia members were involved in the uptake of
394 amino-acids (K01999; K09969), sugars (K02055; K02027; K10552), and peptides (K02035;
395 K02032) which is reflective of typical copiotrophic activity (Supplementary Table 3).
396
397 Copiotrophic bacteria are also characterised by their strong ability to interact within their
398 associations (68). This is supported by the prominence of several bacterial-interactive and
399 particle-associated traits in the MAGs studied here (68, 69) (Supplementary Figure 4;
400 Supplementary Table 5). The presence of bacterial-associated interactions taking place
401 between *Trichodesmium,* and their consortium members could further be implied through the
402 presence of several proteins involved in pilus formation (K02658, K02662, K02669) and
403 secretion systems (K11003; K03072; K03070, K03110, K03073; K03217) within *T. thiebautii*
404 MAG 52 (Supplementary Table 2). Collectively, these findings are indicative of an intricate
405 network of interaction between *Trichodesmium* and its associated bacteria. For example, *T.*
406 *thiebautii* MAG 52 and several MAGs, including re-occurring members from the order
407 Rhodobacterales, contained a homoserine lactone biosynthesis pathway (Supplementary
408 Figure 4a). Previous studies have predicted that consortium members can directly interact with
409 *Trichodesmium* through the production of homoserine lactones (HLs), where the presence of

410 HLs was shown to modulate N_2 fixation rates (14) and increase alkaline phosphatase (AP)
411 activity in *Trichodesmium* (70). The functional association of poorly studied bacterial clades,
412 in particular Balneolales (Supplementary Table 1), however, remained more elusive and
413 indicated that we still lack a clear understanding of these bacteria and their role within the
414 *Trichodesmium* consortium. Further investigations will be required to untangle the different
415 dynamics of both general and specific interactions taking place between members of the
416 *Trichodesmium* consortium.
417

418 ## *Conclusions and Future Perspectives*

419 The complex *Trichodesmium* consortium represents an ideal system to study the intricate
420 interactions taking place within a particle-rich system (Figure 6), asking how these processes
421 may impact the C, N, P and Fe-cycles in a rapidly changing ocean where both microbial
422 community composition and changes in microbial abundance are expected to occur (25, 71).
423 The *Trichodesmium* consortium is redundant for multiple traits related to nutrient cycling
424 including siderophore production, denitrification, DNRA, and reduced phosphate uptake. The
425 processes of denitrification and DNRA were modular, indicating an interactive and spatial
426 coupling between different consortium members. The associated bacteria were largely
427 auxotrophic for one or more vitamins that *Trichodesmium* can synthesize, and this is likely a
428 major trait influencing the community structure (Figure 6). The vast majority of the associated
429 bacteria could further be characterized according to a copiotrophic lifestyle where
430 *Trichodesmium* represents a hotspot for C, N, P, Fe, and vitamins to the residing community.
431 On the other hand, Fe-uptake through the production of siderophores appeared to be a key
432 beneficial function *Trichodesmium* can obtain from its residing community. *Trichodesmium*
433 harbored several bacteria that can secrete photolabile siderophores, supporting the idea that
434 *Trichodesmium* can meet their high Fe-demand by interacting with siderophore-producing

435 bacteria to mine Fe from dust particles collected within the colony. Other, as of yet
436 uncharacterized, traits will likely further elucidate the factors that drive the community
437 structure of *Trichodesmium* colonies. Our results highlight the presence of a dynamic and
438 flexible consortium where functional traits are well conserved within the *Trichodesmium*
439 consortium, underpinning the resilience of the colony within an ever-changing ocean.



440
441 **Figure 6. Schematic summary of the potential interactions taking place within the *Trichodesmium* consortia**
442 **as discussed in this study.** Each blue bubble indicates a different function, and each blue circle pools together
443 the different MAGs containing a particular trait for that function. Each MAG is represented by a coloured dot
444 reflective of the taxonomic order it is associated with (grey dots represent MAGs that were not attributed to a
445 specific taxonomic order). *Trichodesmium* is highlighted in cobalt-blue, and the important functions
446 *Trichodesmium* is known to display are highlighted with a dotted circle. (See Supplementary Table 5 for the full
447 list of functional traits and their presence/absence). **(a)** Several siderophore biosynthesis genes are present in a
448 diverse set of MAGs. Siderophores are predicted to mine Fe from dust that has been collected and centered within
449 *Trichodesmium* colonies. 6 of the 4 were identified as known NIS-type siderophores of which petrobactin,
450 vibrioferrin and rhizoferrin are photolabile. In the presence of sunlight, the Fe complexed to these siderophores
451 disassociates and acts as a public good to the larger consortium. The genes and pathways for **(b-i)** denitrification,
452 **(b-ii)** N_2 fixation, and **(b-iii)** dissimilatory- nitrate- reduction- to- nitrate (DNRA) are present in several MAGs of
453 the consortium. These processes appear to be modular within the *Trichodesmium* consortium as different
454 taxonomic groups complete the steps for these pathways. **(c)** Several MAGs contain a homoserine lactone

synthesis pathway (HL), including *Trichodesmium*. HL acts as an intercommunicating quorum sensing molecule and has been shown to influence the production of alkaline phosphatase (AP) and N_2 fixation in *Trichodesmium*. **(d)** *Trichodesmium* is capable of producing the vitamins cobalamin (B12), thiamin (B1) and biotin (B7). Different subsets of MAGs lack the biosynthesis pathways for one or more of these vitamins with very few capable of synthesizing all three (associated bacteria can also produce vitamins but is not depicted here - see Figure 2). Consortia members and *Trichodesmium* likely exhibit an interdependent relationship for these vitamins. **(e)** *Trichodesmium* colonies have been shown to take up and utilize phosphonate and phosphite in a low-P environment and MAGs containing these pathways may have a competitive advantage over members that lack them.

Material and Methods

Trichodesmium metagenomic sampling and extraction

Trichodesmium puff colonies were hand-picked during the spring bloom (6th May, 2019) and separated into three samples for metagenomics (~100-200 colonies each) using a 100 µm phytoplankton net at 20 m depth in the Gulf of Aqaba (Eilat, Israel) (29.56ºN, 34.95ºE)**.** Colonies were washed 3 times, by gently picking colonies with a pipette and placing them in new petri dishes with fresh 0.2 µm filtered seawater, before being filtered on a 0.2 µm, 45 mm polycarbonate (PC) filter using a vacuum pump. Filters were flash frozen in liquid N_2 and kept at -80ºC. DNA was extracted from the three *Trichodesmium* samples using the DNeasy Plant Pro Kit (Qiagen), with a minor modification to the lysis procedure. The kit-provided tissue disruption tubes were not used. Rather, ~250 µL zirconia/silica beads (0.5 mm) were added to each sample tube before the addition of Solution CD1, and then samples were vortexed for 5 min. The resulting lysate was processed as per the remainder of the manufacturer instructions. DNA was quantified fluorometrically using the dsDNA HS kit (ThermoFisher Scientific) on a Qubit. From the three samples, metagenomic libraries were prepared and sequenced with 0.7 billion reads 2 × 151 bp reads on Illumina NovaSeq S4 at the DOE Joint Genome Institute (California, USA). This study can be found under JGI Gold Study ID number Gs0149370.

Metagenomic raw read assembly and binning

Sequences were analysed from this study and compared to previously published metagenomic studies from the Red Sea (PRJNA804487) (28) and metagenomes of *Trichodesmium* colonies

486 isolated from the Pacific (PRJNA435427) (17) (PRJNA358796) (12) and the Atlantic
487 (PRJNA330990) (13) Oceans. All raw sequences were analyzed using the same protocol
488 described below.
489
490 Metagenomes were assembled, binned, and quantified using the ATLAS (v2) pipeline (72).
491 Briefly, raw sequences underwent quality control through the BBTools suite (73, 74) and were
492 assembled using metaSPAdes (75) (k-mer lengths: 21, 33, 55, 99, 121 bp). MAGs were binned
493 from each sample using MetaBAT 2 (76), MaxBin 2.0 (77) and VAMB (78). The completeness
494 and redundancy of each bin were subsequently assessed using CheckM2 v0.1.2 (79). MAGs
495 were considered of high-quality if they reached 90% completeness and 10% redundancy and
496 were highlighted throughout our results and figures (27). A non-redundant set of bins was
497 produced using DAS Tool (80) and dRep (81) based on an average nucleotide identity (ANI)
498 cutoff of 97.5%. Genes were predicted using Prodigal (82). The coverage of each MAG was
499 quantified across each sample by mapping reads back to each MAG. We refer to the MAGs
500 from this study as MAG XX, bacterial MAGs from a previous Red Sea study as MAG R-XX
501 (28) and bacterial MAGs from other metagenomic studies from the Pacific and Atlantic Oceans
502 (12, 13, 17) as MAG T-XX.
503
504 *Phylogenetic Diversity*
505 Metagenomically assembled genomes (MAGs) were taxonomically characterized using the
506 genome taxonomy database tool kit GTDB-tk v2.1.1 (83). MAGs were classified according to
507 GTDB taxonomy and NCBI taxonomy (Supplementary Table 1; Supplementary Table 4).
508 GTDB taxonomy is used throughout the analysis, but to simplify comparisons with the
509 literature, we used the larger NCBI taxonomic order Rhodospirillales which encompasses the
510 smaller GTDB taxonomic groups of Kiloniellales and Thalassobaculaceae.

511

512 Phylogeny was inferred using GTo-Tree (v.1.16.12; default settings) (84) from a concatenation
513 of 74 conserved single-copy HMM markers for bacteria using the best-fit model Q.pfam+R7
514 model in IQ-Tree (v2.1.3), using the Bayesian information criterion (BIC) (85, 86).
515 Shimodaira–Hasegawa approximate likelihood-ratio test (SH-aLRT) and ultrafast bootstrap
516 approximation (UFBoot) branch support values were estimated from 1000 bootstraps. The tree
517 was visualized using FigTree (v1.4.4) and rooted using the genome of *Fuseobacterium*
518 *nucleatum* (PRJNA1419) as an outgroup according to (87). MAG were considered to re-occur
519 in another metagenomic dataset if its genome sequence identity was >97.5% similar
520 (Supplementary Figure 1).

521

522 *Functional Diversity*

523 The amino-acid sequence for each MAG was annotated using hmmscan and GhostKoala under
524 default parameters (50% identity cutoff), utilizing the PFAM and KEGG databases,
525 respectively (88, 89). Secondary metabolites (e.g. biosynthesis of siderophores and homoserine
526 lactones) were screened for each MAG using AntiSMASH 7.0. In certain cases, the Rapid
527 Annotation using Subsystem Technology (RAST) server was used to annotate amino-acid
528 sequences using the collection of protein families, FIGfams (90, 91). In addition, each MAG
529 was assessed using METABOLIC (v4.0) which lists the presence of interactive KEGG
530 modules for each MAG using a default completeness cutoff of 0.75 (92). Based on these
531 different annotation formats, we inspected each MAG for the presence of functional traits. For
532 each functional pathway or gene, the annotation method and coinciding KEGG and PFAM ID
533 number can be found in Supplementary Table 5. The presence and absence of different traits
534 for each MAG was displayed using the Interactive Tree Of Life (iTOL) visualization tool (93).

535

536 *Iron metabolism*

537 To explore the relationship between *Trichodesmium* and its associated bacteria for the
538 bioavailability of Fe from dust, we searched MAGs for the KEGG IDs of multiple Fe-uptake
539 pathways. Fe-uptake receptors were separated according to abiotic Fe^{+2} (*feoB*:K04759), Fe^{+3}
540 (*afuA*:K02012), and organic Fe uptake which is further separated into Fe-heme
541 (*hemR*:K16087), Fe-citrate (*fecA*:K16091) and Fe-siderophores (*fevS*:K02016). The presence
542 of an Fe:siderophore receptor K02016 should be taken cautiously as its KEGG annotation does
543 not clearly distinguish between siderophores or cobalamin as a substrate.

544

545 Siderophore biosynthesis pathways within MAGs were identified using a conservative
546 approach that required a combination of both KEGG and HMM annotation for each MAG and
547 the use of AntiSMASH 7.0 and FeGenie tools (94) (Supplementary Figure 4). Pathways were
548 separately analyzed for both Non-Ribosomal Peptide Synthetase (NRPS) Type and NRPS
549 Independent Synthetase (NIS) Type biosynthesis pathways. NIS-type siderophore biosynthesis
550 pathways are more easily identifiable of the two, and we searched for the presence of the Fe-
551 uptake chelate (*iuc)*A/*iuc*C domain (PF04183). NIS genes were clustered into different types
552 using known NIS biosynthesis genes as defined and described by Carroll and Moore (36).
553 Briefly, genes were aligned using the online Multiple Alignment Fast Fourier Transformation
554 (MAFFT v7.490; L-INS-i) software (95). A phylogenetic tree was constructed from this
555 alignment using IQtree2 (v2.1.3). The resulting best-fit substitution model Q.pfam+F+R4 was
556 selected using Bayesian information criteria (BIC). Branches were assigned using Shimodira–
557 Hasegawa approximate likelihood-ratio test (SH-aLRT) and ultrafast bootstrap approximation
558 (UFBoot) branch support values were estimated from 1,000 bootstraps. The resulting
559 consensus tree was visualized using FigTree (v1.4.4).

560

561 The identification of NRPS-type siderophore biosynthesis pathways required FeGenie, which
562 utilizes biosynthetic pHMMs, and was used to highlight putative NRPS-type siderophores that
563 were further inspected using AntiSMASH 7.0. MAGs containing an NRPS pathway coupled
564 with a siderophore transport-receptor (SMCOG1082) within the gene-cluster were marked as
565 putative NRPS-siderophore producers. These pathways were further compared to known
566 NRPS siderophore pathways using the MiBIG database (96) and the KEGG pathway map for
567 NRPS-type siderophores biosynthesis gene cluster (BGC) map01053.
568

569 *Vitamin B1, B7, and B12 biosynthesis and uptake*

570 We assume that a bacterial strain is auxotrophic for thiamin (B1), biotin (B7) or cobalamin
571 (B12) if their coinciding genome lacks a cobalamin biosynthesis pathway and contains a
572 transporter for its uptake. We searched MAGs for the KEGG IDs and determined the presence
573 of a vitamin biosynthesis pathway using a KEGG map for each vitamin. The presence of a
574 thiamin biosynthesis pathway (M00127) was determined by the presence of *thi*DCEGL
575 (K00946; K00788; K03149; K00941; K03147). A thiamin uptake system was present if a
576 coinciding transporter system (M00192) was observed. The presence of a biotin biosynthesis
577 pathway (M00581) was determined by the presence of *bio*ABDFH (K01012; K01935;
578 K00833; K00652; K02170). A biotin uptake system was present if a coinciding transporter
579 system (M00581) was observed. The presence of a cobalamin biosynthesis pathway was
580 determined by the presence of *cob*ACDUQS (K19221; K02226; K02227; K02231:K02233).
581 A putative vitamin B12 uptake system was considered present in a MAG if the specific
582 transporter *btu*B (K16092) was observed.

583

584 *Nitrogen metabolism*

585 N-metabolism pathways om MAGs were compiled from a selected set of KEGG ID numbers
586 using METABOLIC-C (v4.0) and the KEGG pathway map ko00910. MAGs were queried for
587 the presence of genes related to N_2 fixation, dissimilatory nitrate reduction to ammonium
588 (DNRA), and denitrification. The ability to fix N_2 was determined through the presence of
589 *nif*HKD (K02588, K02591, K02586). DNRA can be separated into two steps and includes the
590 reduction of nitrate to nitrite (*nap*AB: K02567, K02568*; nar*GH: K00370-K00374) and nitrite
591 to ammonia (*nrf*AH: K15876, K03385*, nir*BD: K00362, K00363). Denitrification can further
592 be separated into the reduction of nitrite to nitric oxide (*nir*S:K15864*; nir*K: K00368), nitric
593 oxide to nitrous oxide (*nor*BC: K04561, K02305) and nitrous oxide to N_2 (*nos*Z: K00376).

594

595 *Phosphorus metabolism*

596 The presence of multiple P-metabolic pathways within the *Trichodesmium* consortium were
597 explored. We explored several transporters for high-affinity phosphate (*pst*SCAB: K02036-
598 K02038, K02040), low-affinity phosphate (*pitA*:K03306), and a
599 phosphite/phosphonate transporter (*phn*CDE: K02041; K02042; K02044). We searched
600 MAGs for KEGG ID numbers for the presence of phosphite dehydrogenase (*ptx*D:K18916),
601 which oxidizes phosphite to phosphate, and the broad specificity C-P lyase (*phn*GHIJKLM:
602 K06162-6, K05780-1), which hydrolyses phosphonate bonds in organic P. Additionally,
603 MAGs were queried for the presence of the substrate-specific phosphonate metabolism
604 enzymes including *phn*A (K19670), *phn*X (K05306), and *phn*W (K03430) and collectively
605 these genes were used to indicate the functional potential for reduced P-metabolism. Lastly,
606 MAGs were examined for the presence of alkaline phosphatases (AP) *pho*A (K01077), *pho*X

607 (K07093), and *pho*D (K01113). AP hydrolyses phosphoesters enabling the metabolism of
608 organic-P.
609

610 ### *Bacterial and Particle Interaction traits*

611 Known traits that are indicative of putative bacterial interactions were previously summarized
612 and identified by Zoccarato *et al* for marine bacterial genomes (68). In addition to siderophore
613 biosynthesis and vitamin B1, B7 and B12 biosynthesis (see previously), MAGs were probed
614 for the presence of homoserine lactone (HLs) biosynthesis genes using AntiSMASH 7.0. The
615 presence of auxin efflux genes was determined according to RAST annotation. MAGs were
616 screened for KEGG ID numbers for the presence of the quorum sensing regulation gene *lux*R
617 (K07782) and *bja*R1 (K18098), secretion systems (KEGG map03070), and motility and
618 adhesion genes (KEGG map02020) (Supplementary Table 5). Traits involved in particle-
619 interaction were previously described and summarized by Debeljak *et al* (69). In addition to
620 Fe-uptake and siderophore biosynthesis pathways, MAGs were screened for KEGG ID
621 numbers involved in metal-response, metal-efflux and metal-storage pathways (Supplementary
622 Table 5).
623

624 ### *Trichodesmium proteomic sampling and extraction*

625 In parallel to the metagenomic sampling, 20 puff colonies were picked, in triplicate, for
626 proteomic analysis at the Environmental Molecular Sciences Laboratory (EMSL) at Pacific
627 Northwest National Laboratory according to standardized protocols. Colonies were washed 3
628 times with filtered seawater, as described above. Samples were centrifuged, and the resulting
629 pellet was diluted in 200 µl 8 M urea and transferred to 2 mL pre-filled Micro-Organism Lysing
630 Mix glass bead tubes and bead beat in a Bead Ruptor Elite bead mill homogenizer (OMNI
631 International, Kennesaw Georgia) at speed 5.5 for 45 sec. After bead beating, the lysate was

632 transferred to a new 4 mL tube and immediately placed in an ice block and spun at 1,000 x g
633 for 10 mins at 4°C. 200 µl of the lysed samples were transferred into 2 mL centrifuge tubes. A
634 bicinchoninic acid (BCA) assay (Thermo Scientific, MA USA) was performed to determine
635 protein concentration. Following the assay, 10 mM dithiothreitol (DTT) was added to the
636 samples and incubated at 60°C for 30 mins with constant shaking at 800 rpm. Samples were
637 diluted 8-fold for preparation for digestion with 100 mM NH_4HCO_3, 1 mM $CaCl_2$ and
638 sequencing grade trypsin (Promega, WI) was added to all protein samples at a 1:50 (w/w)
639 trypsin-to-protein ratio for 3 hrs at 37˚C, 450 rpm. Digested samples were desalted using a 4-
640 probe positive pressure Gilson GX-274 ASPEC™ system (Gilson Inc., WI) with Discovery
641 C18 50 mg/1 mL solid phase extraction tubes (Supelco, MO), using the following protocol: 3
642 mL of methanol was added for conditioning followed by 3 mL of 0.1% trifluoroacetic acid
643 (TFA) in H_2O. The samples were then loaded onto each column followed by 4 mL of 95:5:
644 H_2O:ACN, 0.1% TFA. Samples were eluted with 1mL 80:20 ACN:H_2O, 0.1% TFA. The
645 samples were concentrated using a speed vac and a final BCA was performed to determine the
646 peptide concentration and samples were fractionated into 12 fractions for LC-MS/MS analysis.
647
648 MS analysis was performed using a Q-Exactive Plus mass spectrometer (Thermo Scientific)
649 outfitted with a homemade nano-electrospray ionization interface. Electrospray emitters used
650 150 µm o.d. × 20 µm i.d. chemically etched fused silica (126). The ion transfer tube
651 temperature and spray voltage were 300°C and 2.2 kV, respectively. Data was collected for
652 120 min following a 10 min delay after completion of sample trapping and start of gradient.
653 FT-MS spectra were acquired from 300 to 1800 m/z at a resolution of 70 k (AGC target 3e6)
654 and the top 12 FT-HCD-MS/MS spectra were acquired in data-dependent mode with an
655 isolation window of 1.5 m/z at a resolution of 17.5 k (AGC target 1e5) using a normalized
656 collision energy of 30, dynamic exclusion time of 30 s, and detected charge state of an ion 2 or

657 higher. The resulting proteomic samples from this study was deposited in MassIVE
658 (MSV000091416).
659

660 ### *Proteomic Analysis*

661 Peptide matching was performed using MSGF+ software (97) against the amino-acid
662 sequences obtained from the matching metagenome dataset through PROKKA (98) using the
663 ATLAS pipeline (v2) (72) resulting in 201374 protein sequences. MSGF+ was used in
664 target/decoy mode with 20 ppm parent ion tolerance, partial tryptic rule, and M-oxidation
665 (+15.9949) as a dynamic modification. Best matches from the MSGF+ searches were filtered
666 at 1% FDR based on the target/decoy model; this set of peptides was used in consequent
667 quantitation analysis. Spectral counts were performed without consideration of peptide
668 specificity for proteins and MAGs. 92% of peptides mapped to unique proteins while 93% of
669 peptides were found in only one MAG. Spectral count normalization was performed using the
670 normalized spectral abundance factor (NSAF) approach (99). Briefly, counts were divided by
671 the matching peptide length and their relative abundance was calculated from the total number
672 of spectral hits per sample.
673

674 ## Acknowledgements

675 This research was performed under the Facilities Integrating Collaborations for User Science
676 (FICUS) program (proposal: 10.46936/fics.proj.2018.50403/60000053) and used resources at
677 the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which
678 are DOE Office of Science User Facilities. Both facilities are sponsored by the Biological and
679 Environmental Research program and operated under Contract Nos. DE-AC02-05CH11231
680 (JGI) and DE-AC05-76RL01830 (EMSL). This research was supported by Grant no. 2020041
681 awarded to YS and RB from the United States-Israel Binational Science Foundation (BSF).

This research was supported by Grant no. 2020041 awarded to YS and RB from the United States-Israel Binational Science Foundation (BSF) and the Israel Ministry of Science and Technology (MOST) Grant no. 001126 to MR-B. Additional support was provided by a grant from the Simons Foundation award ID 721225 to STD.

Competing Interests

The authors declare no conflicting interests.

Data Availability Statement

Raw sequences are available under the Gold Study ID number Gs0149370. The 52 metagenomic assembled genomes analysed in this study are available at NCBI under the BioProject number PRJNA944101. Proteomic samples are available at MassIVE under the project accession number MSV000091416.

Author Contributions

YS, RB, MG, SB, conceived of this study. SB, FZ, SW collected the DNA and protein samples. SB performed the experiments together with FZ, SW and YS. SD, SH, TG extracted and sequenced the DNA. CDN extracted and processed protein samples that were analyzed by NT and RB. MRB and CK performed the metagenomic analysis. CK wrote the manuscript under guidance of YS, MRB, RB, MG, SD, with further input from SW, SB, FZ, NT.

1. Bergman B, Sandh G, Lin S, Larsson J, Carpenter EJ. 2013. *Trichodesmium*– a widespread marine Cyanobacterium with unusual nitrogen fixation properties. FEMS Microbiol Rev 37:286–302.
2. Zehr JP. 2011. Nitrogen fixation by marine cyanobacteria. Trends Microbiol 19:162–173.
3. Capone DG, Burns JA, Montoya JP, Subramaniam A, Mahaffey C, Gunderson T, Michaels AF, Carpenter EJ, Capone C, Burns JA, Montoya JP, Subramaniam A, Mahaffey C, Gunderson T, Michaels AF. 2005. Nitrogen fixation by *Trichodesmium* spp: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. Global Biogeochem Cycles 19:1–17.
4. Hutchins DA, Fu FX, Zhang Y, Warner ME, Feng Y, Portune K, Bernhardt PW, Mulholland MR. 2007. CO_2 control of *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios: Implications for past, present, and future ocean biogeochemistry. Limnol Oceanogr 52:1293–1304.
5. Thompson AW, Zehr JP. 2013. Cellular interactions: lessons from the nitrogen-fixing cyanobacteria. J Phycol 49:1024–1035.

6. Wang S, Koedooder C, Zhang F, Kessler N, Eichner M, Shi D, Shaked Y. 2022. Colonies of the marine cyanobacterium *Trichodesmium* optimize dust utilization by selective collection and retention of nutrient-rich particles. iScience 25:103587.
7. Kessler N, Armoza-Zvuloni R, Wang S, Basu S, Weber PK, Stuart RK, Shaked Y. 2019. Selective collection of iron-rich dust particles by natural *Trichodesmium* colonies. The ISME Journal 2019 14:1 14:91–103.
8. Rubin M, Berman-Frank I, Shaked Y. 2011. Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. Nature Geoscience 2011 4:8 4:529–534.
9. Hmelo L, Van Mooy B, Mincer T. 2012. Characterization of bacterial epibionts on the cyanobacterium Trichodesmium. Aquatic Microbial Ecology 67:1–14.
10. Rouco M, Haley ST, Dyhrman ST. 2016. Microbial diversity within the Trichodesmium holobiont. Environ Microbiol 18:5151–5160.
11. Lee MD, Walworth NG, McParland EL, Fu FX, Mincer TJ, Levine NM, Hutchins DA, Webb EA. 2017. The *Trichodesmium* consortium: Conserved heterotrophic co-occurrence and genomic signatures of potential interactions. ISME Journal 11:1813–1824.
12. Gradoville MR, Crump BC, Letelier RM, Church MJ, White AE. 2017. Microbiome of *Trichodesmium* colonies from the North Pacific Subtropical Gyre. Front Microbiol 8.
13. Frischkorn KR, Rouco M, Van Mooy BAS, Dyhrman ST. 2017. Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean. The ISME Journal 2017 11:9 11:2090–2101.
14. Frischkorn KR, Rouco M, Van Mooy BAS, Dyhrman ST. 2018. The *Trichodesmium* microbiome can modulate host N_2 fixation. Limnol Oceanogr Lett 3:401–408.
15. Klawonn I, Eichner MJ, Wilson ST, Moradi N, Thamdrup B, Kümmel S, Gehre M, Khalili A, Grossart HP, Karl DM, Ploug H. 2019. Distinct nitrogen cycling and steep chemical gradients in *Trichodesmium* colonies. The ISME Journal 2019 14:2 14:399–412.
16. Basu S, Gledhill M, de Beer D, Prabhu Matondkar SG, Shaked Y. 2019. Colonies of marine cyanobacteria *Trichodesmium* interact with associated bacteria to acquire iron from dust. Communications Biology 2019 2:1 2:1–8.
17. Frischkorn KR, Haley ST, Dyhrman ST. 2018. Coordinated gene expression between *Trichodesmium* and its microbiome over day-night cycles in the North Pacific Subtropical Gyre. ISME Journal 12:997–1007.
18. Frischkorn KR, Krupke A, Guieu C, Louis J, Rouco M, Salazar Estrada AE, Van Mooy BAS, Dyhrman ST. 2018. *Trichodesmium* physiological ecology and phosphate reduction in the western tropical South Pacific. Biogeosciences 15:5761–5778.
19. Gledhill M, Basu S, Shaked Y. 2019. Metallophores associated with *Trichodesmium erythraeum* colonies from the Gulf of Aqaba. Metallomics 11:1547–1557.
20. Basu S, Shaked Y. 2018. Mineral iron utilization by natural and cultured *Trichodesmium* and associated bacteria. Limnol Oceanogr 63:2307–2320.
21. Polyviou D, Baylay AJ, Hitchcock A, Robidart J, Moore CM, Bibby TS. 2018. Desert dust as a source of iron to the globally important diazotroph *Trichodesmium*. Front Microbiol 8:2683.
22. Manck LE, Park J, Tully BJ, Poire AM, Bundy RM, Dupont CL, Barbeau KA. 2021. Petrobactin, a siderophore produced by *Alteromonas*, mediates community iron acquisition in the global ocean. The ISME Journal 2021 16:2 16:358–369.
23. Berman-Frank I, Quigg A, Finkel Z V., Irwin AJ, Haramaty L. 2007. Nitrogen-fixation strategies and Fe requirements in cyanobacteria. Limnol Oceanogr 52:2260–2269.
24. Jickells TD, An ZS, Andersen KK, Baker AR, Bergametti C, Brooks N, Cao JJ, Boyd PW, Duce RA, Hunter KA, Kawahata H, Kubilay N, LaRoche J, Liss PS, Mahowald

N, Prospero JM, Ridgwell AJ, Tegen I, Torres R. 2005. Global iron connections between desert dust, ocean biogeochemistry, and climate. Science (1979) 308:67–71.

25. Kok JF, Ward DS, Mahowald NM, Evan AT. 2018. Global and regional importance of the direct dust-climate feedback. Nature Communications 9:1 9:1–11.

26. Post AF, Dedej Z, Gottlieb R, Li H, Thomas DN, El-Absawi M, El-Naggar A, El-Gharabawi M, Sommer U. 2002. Spatial and temporal distribution of *Trichodesmium* spp. in the stratified Gulf of Aqaba, Red Sea. Mar Ecol Prog Ser 239:241–250.

27. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology 35:8 35:725–731.

28. Koedooder C, Landou E, Zhang F, Wang S, Basu S, Berman-Frank I, Shaked Y, Rubin-Blum M. 2022. Metagenomes of Red Sea subpopulations challenge the use of marker genes and morphology to assess *Trichodesmium* diversity. Front Microbiol 13:1712.

29. Dang H, Lovell CR. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA Genes. Appl Environ Microbiol 66:467.

30. Dang H, Li T, Chen M, Huang G. 2008. Cross-ocean distribution of Rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. Appl Environ Microbiol 74:52–60.

31. Duret MT, Lampitt RS, Lam P. 2019. Prokaryotic niche partitioning between suspended and sinking marine particles. Environ Microbiol Rep 11:386–400.

32. Koch H, Germscheid N, Freese HM, Noriega-Ortega B, Lücking D, Berger M, Qiu G, Marzinelli EM, Campbell AH, Steinberg PD, Overmann J, Dittmar T, Simon M, Wietz M. 2020. Genomic, metabolic, and phenotypic variability shapes ecological differentiation and intraspecies interactions of *Alteromonas macleodii*. Scientific Reports 2020 10:1 10:1–14.

33. Hou S, López-Pérez M, Pfreundt U, Belkin N, Stüber K, Huettel B, Reinhardt R, Berman-Frank I, Rodriguez-Valera F, Hess WR. 2018. Benefit from decline: The primary transcriptome of *Alteromonas macleodii* str. Te101 during *Trichodesmium* demise. ISME Journal 12:981–996.

34. Held NA, Webb EA, McIlvin MM, Hutchins DA, Cohen NR, Moran DM, Kunde K, Lohan MC, Mahaffey C, Malcolm E, Saito MA. 2020. Co-occurrence of Fe and P stress in natural populations of the marine diazotroph *Trichodesmium*. Biogeosciences 17:2537–2551.

35. Chappell PD, Webb EA. 2010. A molecular assessment of the iron stress response in the two phylogenetic clades of *Trichodesmium*. Environ Microbiol 12:13–27.

36. Carroll CS, Moore MM. 2018. Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases. Crit Rev Biochem Mol Biol. Taylor and Francis Ltd https://doi.org/10.1080/10409238.2018.1476449.

37. Gulick AM. 2017. Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens. Nat Prod Rep 34:981–1009.
38. Reitz ZL, Medema MH. 2022. Genome mining strategies for metallophore discovery. Curr Opin Biotechnol 77:102757.
39. Butler A, Harder T, Ostrowski AD, Carrano CJ. 2021. Photoactive siderophores: structure, function and biology. J Inorg Biochem 221:111457.
40. Amin SA, Green DH, Küpper FC, Carrano CJ. 2009. Vibrioferrin, an unusual marine siderophore: iron binding, photochemistry, and biological implications. Inorg Chem 48:11451–11458.
41. Amin SA, Green DH, Hart MC, Küpper FC, Sunda WG, Carrano CJ. 2009. Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. Proc Natl Acad Sci U S A 106:17071–17076.
42. Cordero OX, Ventouras LA, DeLong EF, Polz MF. 2012. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. Proc Natl Acad Sci U S A 109:20059–20064.
43. Barbeau K, Rue EL, Trick CG, Bruland KW, Butler A. 2003. Photochemical reactivity of siderophores produced by marine heterotrophic bacteria and cyanobacteria based on characteristic Fe(III) binding groups. Limnol Oceanogr 48:1069–1078.
44. Martinez JS, Carter-Franklin JN, Mann EL, Martin JD, Haygood MG, Butler A. 2003. Structure and membrane affinity of a suite of amphiphilic siderophores produced by a marine bacterium. Proc Natl Acad Sci U S A 100:3754–3759.
45. Kramer J, Özkaya Ö, Kümmerli R. 2020. Bacterial siderophores in community and host interactions. Nat Rev Microbiol. Nature Research https://doi.org/10.1038/s41579-019-0284-4.
46. Van Mooy BAS, Krupke A, Dyhrman ST, Fredricks HF, Frischkorn KR, Ossolinski JE, Repeta DJ, Rouco M, Seewald JD, Sylva SP. 2015. Phosphorus cycling: a major role of planktonic phosphate reduction in the marine phosphorus redox cycle. Science 348:783–785.
47. Mackey KRM, Labiosa RG, Calhoun M, Street JH, Post AF, Paytan A. 2007. Phosphorus availability, phytoplankton community dynamics, and taxon-specific phosphorus status in the Gulf of Aqaba, Red Sea. Limnol Oceanogr 52:873–885.
48. Karl DM, Björkman KM. 2015. Dynamics of dissolved organic phosphorus. Biogeochemistry of marine dissolved organic matter: Second Edition 233–334.
49. Orchard ED, Webb EA, Dyhrman ST. 2009. Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. Environ Microbiol 11:2400–2411.
50. Frischkorn KR, Krupke A, Guieu C, Louis J, Rouco M, Salazar Estrada AE, Van Mooy BAS, Dyhrman ST. 2018. *Trichodesmium* physiological ecology and phosphate reduction in the western tropical South Pacific. Biogeosciences 15:5761–5778.
51. Frischkorn KR, Haley ST, Dyhrman ST. 2019. Transcriptional and proteomic choreography under phosphorus deficiency and re-supply in the N2 fixing cyanobacterium *Trichodesmium erythraeum*. Front Microbiol 10:330.
52. Stosiek N, Talma M, Klimek-Ochab M. 2020. Carbon-Phosphorus Lyase—the State of the Art. Appl Biochem Biotechnol 190:1525–1552.
53. Kim AD, Baker AS, Dunaway-Mariano D, Metcalf WW, Wanner BL, Martin BM. 2002. The 2-aminoethylphosphonate-specific transaminase of the 2-aminoethylphosphonate degradation pathway. J Bacteriol 184:4134–4140.
54. Lockwood S, Greening C, Baltar F, Morales SE. 2022. Global and seasonal variation of marine phosphonate metabolism. The ISME Journal 2022 16:9 16:2198–2212.
55. Hilderbrand RL. 2018. The effects of synthetic phosphonates on living systems: the role of phosphonates in living systems 139–169.

56. Coates CJ, Wyman M. 2017. A denitrifying community associated with a major, marine nitrogen fixer. Environ Microbiol 19:4978–4992.
57. Eichner M, Basu S, Wang S, de Beer D, Shaked Y. 2020. Mineral iron dissolution in Trichodesmium colonies: The role of O2 and pH microenvironments. Limnol Oceanogr 65:1149–1160.
58. Pajares S, Ramos R. 2019. Processes and microorganisms involved in the marine nitrogen cycle: knowledge and gaps. Front Mar Sci 6:739.
59. Graf DRH, Jones CM, Hallin S. 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O Emissions. PLoS One 9:e114118.
60. Kreimer A, Borenstein E, Gophna U, Ruppin E. 2008. The evolution of modularity in bacterial metabolic networks. Proc Natl Acad Sci U S A 105:6976–6981.
61. Hutchins DA, Capone DG. 2022. The marine nitrogen cycle: new developments and global change. Nature Reviews Microbiology 2022 20:7 20:401–414.
62. Smriga S, Ciccarese D, Babbin AR. 2021. Denitrifying bacteria respond to and shape microscale gradients within particulate matrices. Communications Biology 2021 4:1 4:1–9.
63. Stief P, Kamp A, Thamdrup B, Glud RN. 2016. Anaerobic nitrogen turnover by sinking diatom aggregates at varying ambient oxygen levels. Front Microbiol 7:98.
64. Paerly HW, Bebout BM, Prufert LE. 1989. Bacteria associations with marine *Oscillatoria* sp. (*Trichodesmium* sp.) populations: ecophysiological implications. J Phycol 25:773–784.
65. McGenity TJ. 2019. Taxonomy, genomics and ecophysiology of hydrocarbon-degrading microbes. Springer International Publishing. Retrieved 18 January 2023.
66. Ho A, Di Lonardo DP, Bodelier PLE. 2017. Revisiting life strategy concepts in environmental microbial ecology. FEMS Microbiol Ecol 93:6.
67. Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J, Ferriera S, Lapidus A, Anderson I, Kyrpides N, Munk AC, Detter C, Han CS, Brown M V., Robb FT, Kjelleberg S, Cavicchioli R. 2009. The genomic basis of trophic strategy in marine bacteria. Proc Natl Acad Sci U S A 106:15527–33.
68. Zoccarato L, Sher D, Miki T, Segrè D, Grossart HP. 2022. A comparative whole-genome approach identifies bacterial traits for marine microbial interactions. Communications Biology 2022 5:1 5:1–13.
69. Debeljak P, Blain S, Bowie A, van der Merwe P, Bayer B, Obernosterer I. 2021. Homeostasis drives intense microbial trace metal processing on marine particles. Limnol Oceanogr 66:3842–3855.
70. Van Mooy BAS, Hmelo LR, Sofen LE, Campagna SR, May AL, Dyhrman ST, Heithoff A, Webb EA, Momper L, Mincer TJ. 2011. Quorum sensing control of phosphorus acquisition in *Trichodesmium* consortia. The ISME Journal 2012 6:2 6:422–429.
71. Azam. F., Fenchel T, Field JG, Azam F, Field JG, Graf JS, Meyer-Rei LA, Thingstad F. 1983. The Ecological Role of Water-Column Microbes in the Sea. Ecol Prog Ser, Mar 10:257–263.
72. Kieser S, Brown J, Zdobnov EM, Trajkovski M, McCue LA. 2020. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. BMC Bioinformatics 21.
73. B. Bushnell. BBTools software package. sourceforge.net/projects/bbmap/. Retrieved 16 December 2021.

74. Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read merging via overlap. PLoS One 12:e0185056.
75. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834.
76. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019:e7359.
77. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.
78. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, Jensen LJ, Nielsen HB, Petersen TN, Winther O, Rasmussen S. 2021. Improved metagenome binning and assembly using deep variational autoencoders. Nat Biotechnol 39:555–560.
79. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2022. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. bioRxiv 2022.07.11.499243.
80. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 2018 3:7 3:836–843.
81. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME Journal 2017 11:12 11:2864–2868.
82. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:1–11.
83. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927.
84. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162–4164.
85. Kalyaanamoorthy S, Minh B, Wong T, von Haeseler A, Jermiin L. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.
86. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274.
87. Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllsi GJ, Williams TA. 2021. A rooted phylogeny resolves early bacterial evolution. Science (1979) 372.
88. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–731.
89. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR. 2004. The Pfam protein families database. Nucleic Acids Res 32:D138–D141.
90. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9:1–15.

91. Meyer F, Overbeek R, Rodriguez A. 2009. FIGfams: yet another set of protein families. Nucleic Acids Res 37:6643–6654.
92. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, Karaoz U, Anantharaman K. 2022. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome 10:33.
93. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49: W293–W296.
94. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, Merino N. 2020. FeGenie: A comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Front Microbiol 11:37.
95. Katoh K, Kuma KI, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–518.
96. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH. 2019. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48: D454–D458.
97. Kim S, Gupta N, Pevzner PA. 2008. Spectral probabilities and generating functions of tandem mass spectra: A strike against decoy databases. J Proteome Res 7:3354–3363.
98. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.
99. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, Washburn MP. 2006. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. Methods 40:303–311.

FIGURES

Figure 1. Phylogenetic tree of the 52 MAGs assembled from *Trichodesmium* colonies collected in the Red Sea. The tree is rooted by the outgroup *Fuseobacterium nucleatum*. Known bacterial orders are marked in different colors and several MAGs could taxonomically be identified at the genus level. The relative abundance of each MAG within the *Trichodesmium* consortium, excluding the abundance of *Trichodesmium* MAG 52 (±70% of samples), is presented as a bubble chart. The 11 MAGs that matched to MAGs assembled from a previous metagenomic dataset of Red Sea *Trichodesmium* colonies (28) are highlighted in blue and tentatively represent the core-consortium (see Supplementary Figure 2; Supplementary Table 4). MAGs that did not meet the requirements of a high-quality genome (>90% completeness; <5% redundancy) are indicated in grey. Note that re-occurring MAGs are not necessarily the most abundant MAGs present within the *Trichodesmium* colony. (a full taxonomic description of each MAG in this dataset can be found in Supplementary Table 1).

Figure 2. MAGs were probed for Fe-uptake and siderophore biosynthesis genes which are noted in Supplementary Table 5. **(a)** *T. thiebautii* MAG 52 does not contain a wide diversity of Fe-uptake genes, and the majority of MAGs of associated bacteria contain an uptake receptor for organic Fe (heme, citrate, siderophores). Several MAGs of associated bacteria contained an NRPS or NIS-type siderophore biosynthesis pathways (BGCs). **(b)** Siderophore biosynthetic gene clusters (BGCs) of known NIS-type siderophore BGCs can be clustered according to the presence of one or multiple iucA/iucC domains which enables one to target the closest siderophore BGC match (36). Using this method, we were able to link these pathways to the synthesis of the photolabile **(i)** petrobactin, **(ii)** vibrioferrin and **(iii)** rhizoferrin.

1015 (c) Siderophore BGCs of known NRPS siderophores. Using antiSMASH 7.0, which aligns
1016 BGCs to known siderophores within the MIBiG database, we were able to link 2 of the 6
1017 putative siderophore NRPS-type pathways to **(i)** the membrane-bound amphibactin and **(ii)**
1018 agrobactin/enterobactin. (See Supplementary Figure 3 for more details on the siderophore
1019 biosynthesis pathways).
1020
1021 **Figure 3.** MAGs were probed for several different P-metabolism genes which are noted in
1022 Supplementary Table 5. *Trichodesmium* is known to be highly competitive for P, as
1023 exemplified by the multitude of **(a)** P-uptake systems for phosphate (*pst*S, *pst*B), phosphonate
1024 (*phn*CDE), and phosphonate/phosphite (*phn*CDE/*ptxABC*). **(b)** Within the *Trichodesmium*
1025 consortium, the ability to metabolize a range of phosphonate compounds using a broad
1026 specificity C-P lyase (*phn*GHIJKLM), is more prominent than the substrate-specific hydrolysis
1027 of 2-aminoethylphosphonate (2-AEP; *phn*WXYZA) or phosphonopyruvate (*pal*A) and
1028 phosphite metabolism (*ptx*D).
1029
1030 **Figure 4.** MAGs were probed for N-metabolic pathways and the genes for each step within the
1031 pathway are noted in Supplementary Table 5. *T. thiebautii* MAG 52 is the only MAG with a
1032 N_2 fixation pathway. Several MAGs contained one or two steps within the pathways of
1033 denitrification or dissimilatory nitrate to reduction to ammonia (DNRA) indicating their
1034 modularity. Collectively, MAGs completed the denitrification and DNRA pathways, but none
1035 contained the genes for the entire pathway.
1036
1037 **Figure 5.** MAGs were probed for vitamin B1, B7, and B12 biosynthesis and uptake pathways.
1038 The genes used to indicate the biosynthesis or uptake of these vitamins are noted in
1039 Supplementary Table 5. *Trichodesmium* contains the biosynthesis pathways for all three
1040 vitamins but the vast majority of associated bacteria were auxotrophic for one or more
1041 vitamins. These results suggest that the associated bacteria rely on *Trichodesmium* to obtain
1042 their vitamin requirements.
1043
1044 **Figure 6. Schematic summary of the potential interactions taking place within the**
1045 **Trichodesmium consortia as discussed in this study.** Genes for each trait are noted in
1046 Supplementary Table 5**.** Each blue bubble indicates a different function, and each blue circle
1047 pools together the different MAGs containing a particular trait for that function. Each MAG is
1048 represented by a colored dot reflective of the taxonomic order it is associated with (grey dots
1049 represent MAGs that were not attributed to a specific taxonomic order). *Trichodesmium* is
1050 highlighted in cobalt-blue, and the important functions *Trichodesmium* is known to display are
1051 highlighted with a dotted circle. (See Supplementary Table 5 for the full list of functional traits
1052 and their presence/absence). **(a)** Several siderophore biosynthesis genes are present in a diverse
1053 set of MAGs. Siderophores are predicted to mine Fe from dust that has been collected and
1054 centered within *Trichodesmium* colonies. 6 of the 4 were identified as known NIS-type
1055 siderophores of which petrobactin, vibrioferrin and rhizoferrin are photolabile. In the presence
1056 of sunlight, the Fe complexed to these siderophores disassociates and acts as a public good to
1057 the larger consortium. The genes and pathways for **(b-i)** denitrification, **(b-ii)** N_2 fixation, and
1058 **(b-iii)** dissimilatory- nitrate- reduction- to- nitrate (DNRA) are present in several MAGs of the
1059 consortium. These processes appear to be modular within the *Trichodesmium* consortium as
1060 different taxonomic groups complete the steps for these pathways. **(c)** Several MAGs contain
1061 a homoserine lactone synthesis pathway (HL), including *Trichodesmium*. HL acts as an
1062 intercommunicating quorum sensing molecule and has been shown to influence the production
1063 of alkaline phosphatase (AP) and N_2 fixation in *Trichodesmium*. (associated bacteria can also
1064 produce AP but is not depicted here - see Figure 3). **(d)** *Trichodesmium* is capable of producing

1065 the vitamins cobalamin (B12), thiamin (B1) and biotin (B7). Different subsets of MAGs lack
1066 the biosynthesis pathways for one or more of these vitamins with very few capable of
1067 synthesizing all three (associated bacteria can also produce vitamins but is not depicted here -
1068 see Figure 5). Consortia members and *Trichodesmium* likely exhibit an interdependent
1069 relationship for these vitamins. **(e)** *Trichodesmium* colonies have been shown to take up and
1070 utilize phosphonate and phosphite in a low-P environment and MAGs containing these
1071 pathways may have a competitive advantage over members that lack them.
1072

1073 ## SUPPLEMENTARY FIGURES

1074 **Supplementary Figure 1. Relative abundance of bacterial MAGs from sampled**
1075 ***Trichodesmium* colonies indicate the associated bacterial distribution is consistent.** The 51
1076 MAGs are stacked, and colors reflect the different bacterial orders each MAG belongs to and
1077 'other' in grey. The exact relative abundance value (%) for each MAG and its standard
1078 deviation can be viewed in Supplementary Table 1.
1079

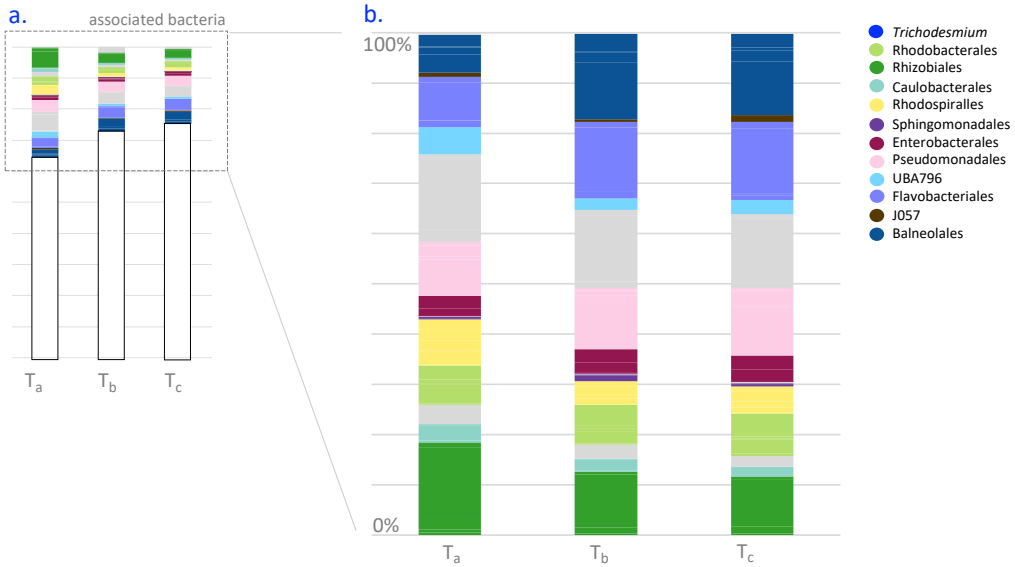
1080 **Supplementary Figure 2. Phylogenetic tree of the 52 MAGs together with MAGs**
1081 **assembled from 3 other Trichodesmium metagenomic datasets from colonies collected in**
1082 **the Red Sea (28), the Pacific (12, 17) and the Atlantic (13) Oceans.** MAGs from the 3 other
1083 datasets are highlighted in grey. MAGs deriving from samples of the Atlantic or Pacific Ocean
1084 are listed as MAG T-XX and those from the Red Sea are listed as MAG R-XX. The 11 MAGs
1085 that matched to those assembled from previous metagenomic datasets, using a 97.5 % identity
1086 cutoff, are highlighted in blue. Known bacterial orders are marked using different colored
1087 squares. The tree is rooted by the outgroup *Fuseobacterium nucleatum*. The full taxonomic
1088 description of each MAG in this dataset and others can be found in Supplementary Table 4.
1089

1090 **Supplementary Figure 3. (a)** NIS and NRPS-type siderophore biosynthesis pathways
1091 identified in MAGs of associated bacteria. (b) The siderophore BGC analysis tools used to
1092 probe and identify the different siderophore BGCs. Most tools are limited in their ability to
1093 discover siderophore BGCs while antiSMASH 7.0 was the most conclusive but also time-
1094 consuming (53). **(c)** Putative NRPS-type siderophore BGCs that did not match to a known
1095 siderophore within the MIBiG database but did containing a TonB-siderophore receptor
1096 (SMCOG 1208). This made us speculate that these are putative siderophores whose identity
1097 remains to be resolved.
1098

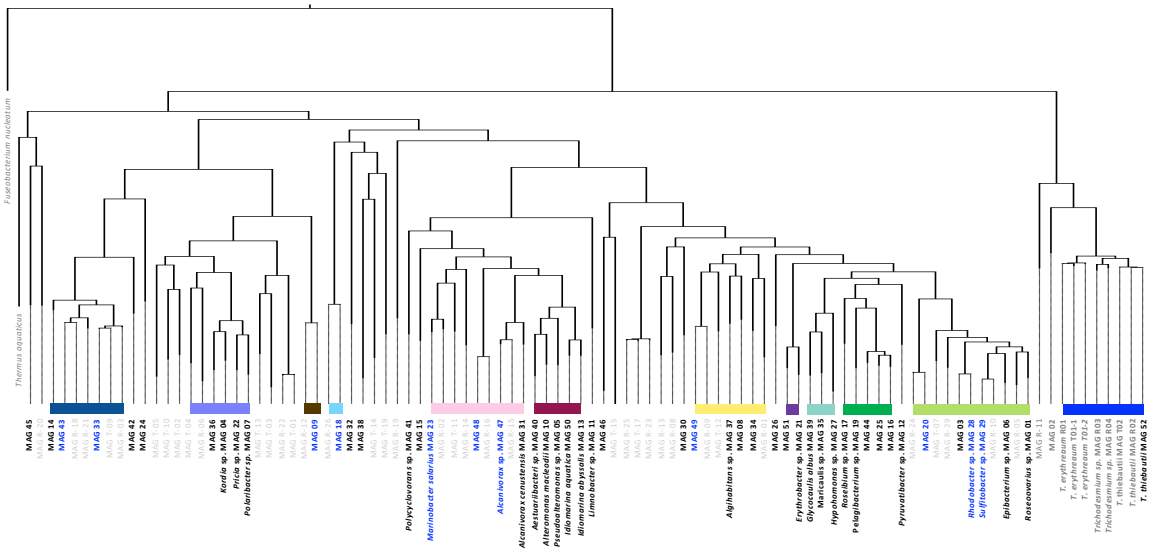
1099 **Supplementary Figure 4. (a) MAGs were probed for bacterial-interactive traits** (68)**.**
1100 MAGs from the phyla Proteobacter, contain the presence several interactive traits, including
1101 quorum sensing, gliding motility, adhesion, cell-to-cell signaling, and antibiotic resistance
1102 marker genes. Intriguingly, both the Trichodesmium MAG and several Proteobacter MAGs
1103 (e.g. Rhodobacter) contain genes for homoserine lactones (HLs), which are putative quorum
1104 sensing molecules. **(b) MAGs were probed for particle-interactive traits** (69)**.** In
1105 comparison to *T. thiebautii* MAG 52, many MAGs from associated bacteria contained a wide
1106 variety of heavy metal efflux systems and metal-transcription factors. While *Trichodesmium*,
1107 like most Cyanobacteria, is known to be sensitive to heavy metals, our results indicate that the
1108 consortium is attuned towards an environment rich in metals, including particulate dust. Genes
1109 for each trait are noted in Supplementary Table 5
1110

1111 ## SUPPLEMENTARY TABLES

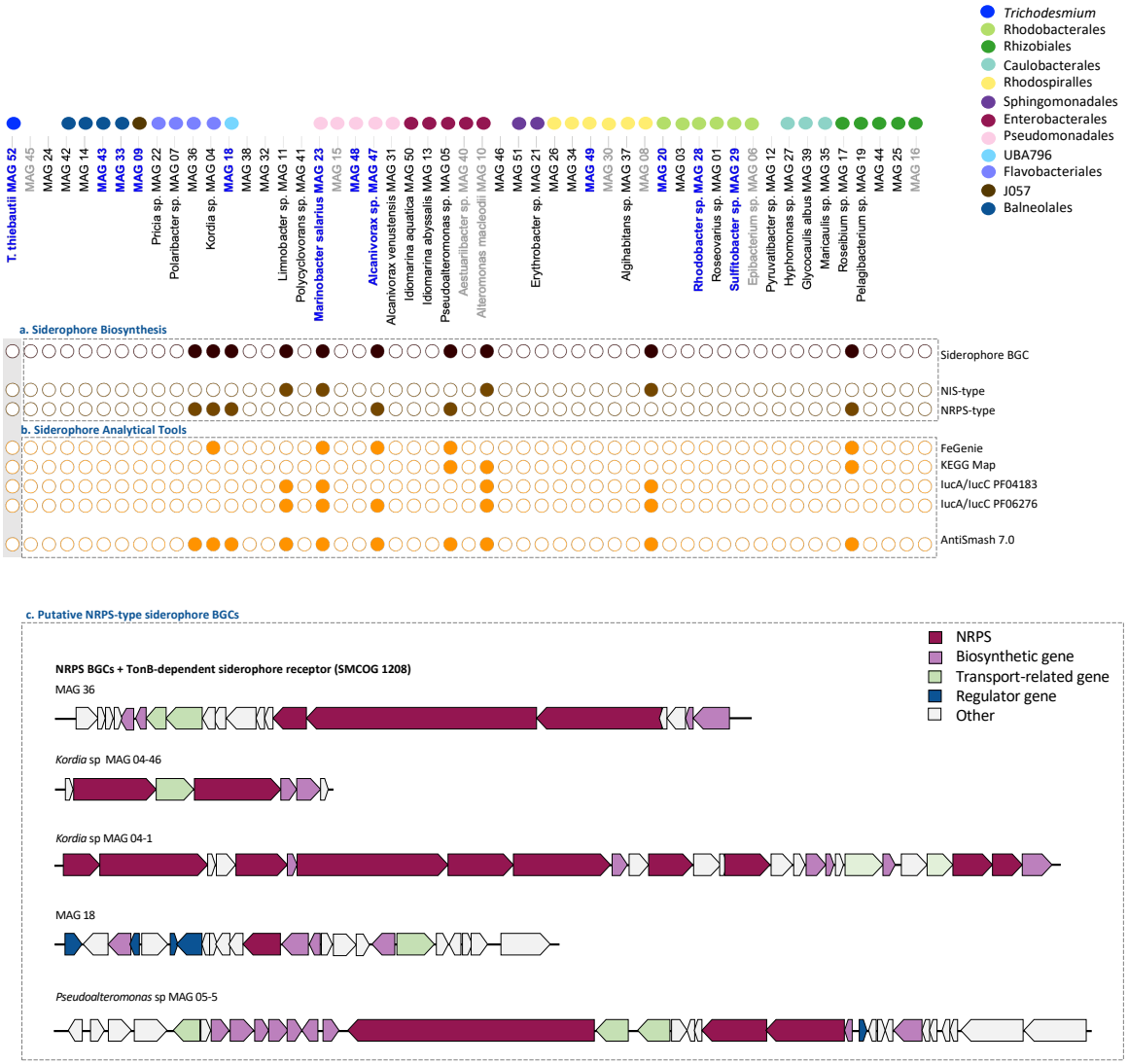
1112 **Supplementary Table 1.** General information about the MAGs making up the Red Sea
1113 *Trichodesmium* consortium. The relative abundance of associated bacteria (%) and standard
1114 deviation of the triplicates (SD) excludes the abundance of *T. thiebautii* (MAG 52). The relative
1115 abundance of MAGs was consistent across our samples. MAGs that did not meet the high-
1116 quality standard of 90% completeness and <5% redundancy are marked in green (10 of 52) and
1117 the functional analysis from these MAGs needs to be taken with more caution. The % of the
1118 genomes annotated by RAST is also indicated, showcasing how MAGs from the phyla
1119 Bacteroidetes are marked with a low annotation percentage. The closest taxonomic identity is
1120 given using both NCBI and GTDB-taxonomy. Throughout this paper we use GTDB-taxonomy
1121 except for the order Rhodospirillales.
1122
1123 **Supplementary Table 2:** Top proteomic hits matching to *T. thiebautii* MAG52. Spectrum
1124 counts that mapped to KEGG ID numbers mentioned in Supplementary Table 5 are highlighted
1125 in blue. Most proteins are geared towards the processes of photosynthesis, nitrogen fixation,
1126 carbon metabolism, amino acid uptake, protein synthesis, and phosphate uptake. Total spectral
1127 counts (T-total) and their associated KEGG or PFAM ID number and their description are
1128 presented together with the average relative contribution (T0%) of each protein hit to all
1129 Trichodesmium-related proteins and for all three samples (%T1-3). Collectively, the top 50
1130 proteins represented >50% of the total spectral counts.
1131
1132 **Supplementary Table 3:** Top proteomic hits matching to MAGs of *Trichodesmium* associated
1133 bacteria. Many proteins are geared towards a typical heterotrophic and copiotrophic lifestyle
1134 including proteins related to amino-acid uptake and carbon metabolism are highlighted in bold.
1135 Total spectral counts (%) and their associated KEGG or PFAM ID numbers and their
1136 description are presented together with the average relative contribution (Avg%) of each
1137 protein hit to all Trichodesmium-related proteins and for all three samples (%T1-3).
1138
1139 **Supplementary Table 4. Taxonomic identity of MAGs used in this study are compared to**
1140 **MAGs assembled from other metagenomic studies of *Trichodesmium* colonies** (Red Sea
1141 samples: MAG R-XX (28), the Pacific (12, 17) and Atlantic samples (13): MAG T-XX. Their
1142 closest taxonomic identity is given using both NCBI and GTDB-taxonomy. Throughout this
1143 paper we use GTDB-taxonomy except for the order Rhodospirillales. Highlighted in blue are
1144 the 11 MAGs from this dataset that matched to those assembled from these other datasets (in
1145 bold), using a 97.5 % identity cutoff. The phylogenetic tree of all MAGs from all datasets
1146 combined can be seen in Supplementary Figure 2.
1147
1148 **Supplementary Table 5.** The presence and absence of functional traits separated into larger
1149 groups including Fe-uptake, P-metabolism, N-metabolism, Vitamin B synthesis and uptake,
1150 metal regulators, metal efflux, metal storage, signaling, secretion systems. motility and
1151 adhesion genes. The associated analysis tool, KEGG id number or module is given for each
1152 trait. The total number of MAGs containing a specific trait and (MAGs %) is calculated and
1153 show below. Traits that were detected within our proteomic samples are also highlighted in
1154 blue.
1155



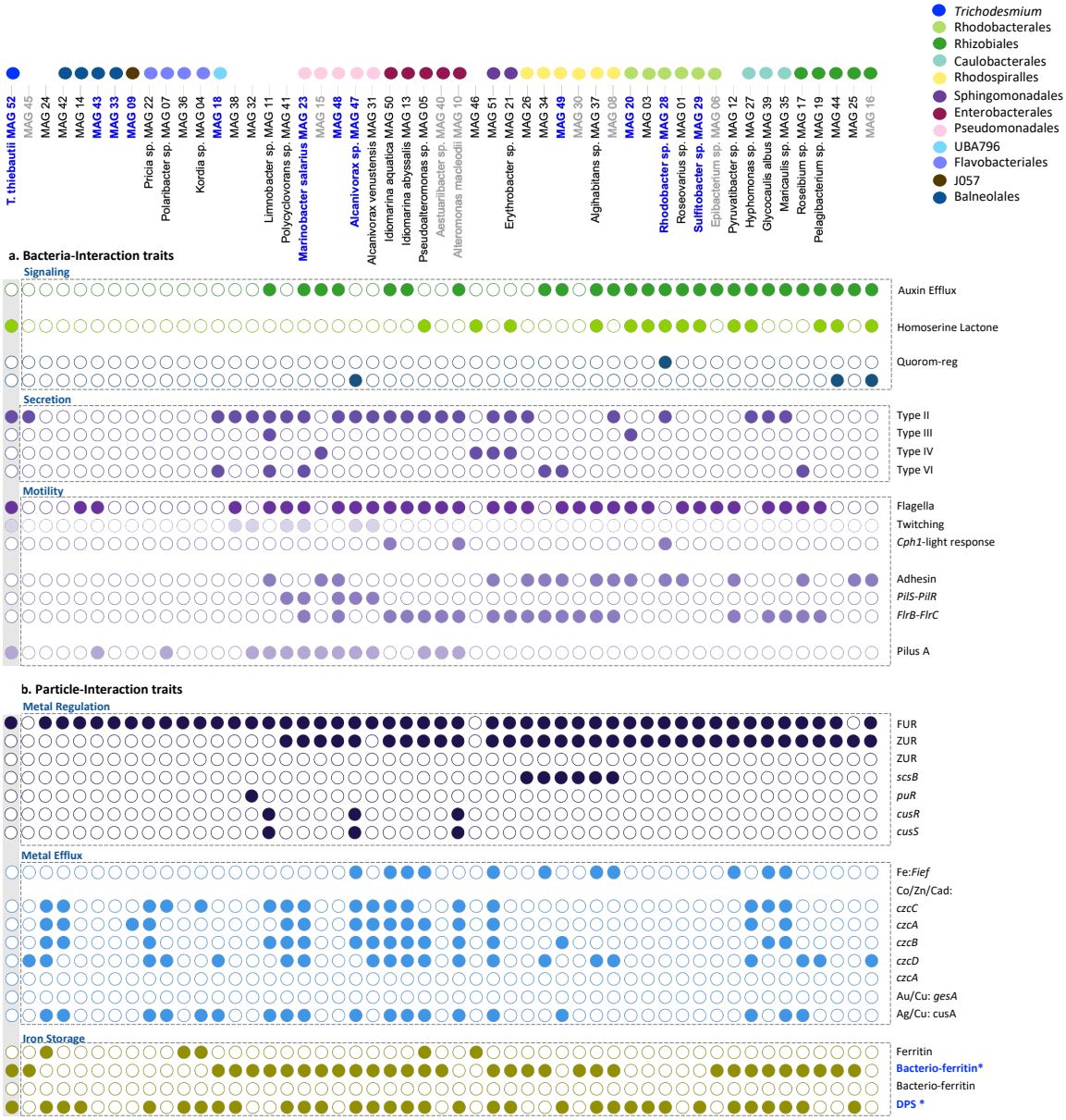
Supplementary Figure 1. Relative abundance of bacterial MAGs from sampled *Trichodesmium* colonies indicate the associated bacterial distribution is consistent. The 51 MAGs are stacked, and colors reflect the different bacterial orders each MAG belongs to and 'other' in grey. The exact relative abundance value (%) for each MAG and its standard deviation can be viewed in Supplementary Table 1.



Phylogenetic tree of the 52 MAGs together with MAGs assembled from 3 other *Trichodesmium* metagenomic datasets from colonies collected in the Red Sea (28)**, the Pacific and the Atlantic Ocean.** MAGs from the 3 other datasets are highlighted in grey. MAGs deriving from samples of the Atlantic or Pacific Ocean are listed as MAG T-XX and those from the Red Sea are listed as MAG R-XX. The 11 MAGs that matched to those assembled from previous metagenomic datasets, using a 97.5 % identity cutoff, are highlighted in blue. Known bacterial orders are marked using different colored squares. The tree is rooted by the outgroup *Fuseobacterium nucleatum*. The full taxonomic description of each MAG in this dataset and others can be found in Supplementary Table 4.



Supplementary Figure 3. (a) NIS and NRPS-type siderophore biosynthesis pathways identified in MAGs of associated bacteria. **(b)** The siderophore BGC analysis tools used to probe and identify the different siderophore BGCs. Most tools are limited in their ability to discover siderophore BGCs while antiSMASH 7.0 was the most conclusive but also time-consuming. **(c)** Putative NRPS-type siderophore BGCs that did not match to a known siderophore within the MIBiG database but did containing a TonB-siderophore receptor (SMCOG 1208). This made us speculate that these are putative siderophores whose identity remains to be resolved.



Supplementary Figure 4. (a) MAGs were probed for bacterial-interactive traits (68). MAGs from the phyla Proteobacter, contain the presence several interactive traits, including quorum sensing, gliding motility, adhesion, cell-to-cell signaling, and antibiotic resistance marker genes. Intriguingly, both the Trichodesmium MAG and several Proteobacter MAGs (e.g. Rhodobacter) contain genes for homoserine lactones (HLs), which are putative quorum sensing molecules. **(b)** MAGs were probed for particle-interactive traits (69). In comparison to T. thiebautii MAG 52, many MAGs from associated bacteria contained a wide variety of heavy metal efflux systems and metal-transcription factors. While *Trichodesmium*, like most Cyanobacteria, is known to be sensitive to heavy metals, our results indicate that the consortium is attuned towards an environment rich in metals, including particulate dust. Genes for each trait are noted in Supplementary Table 5.