Detailed responses to reviewers

We thank both reviewers for very useful critiques. We have addressed all of their concerns below, and in a revised manuscript. The most substantive changes in the revised manuscript include the addition of a diazotroph community composition analysis using next generation sequencing of *nifH*-amplified libraries generated from mesocosm and lagoon samples. We feel this analysis significantly strengthens the study, both by verifying that the diazotrophs quantified via qPCR are among the major groups in the experiment, and by supplying the first *nifH*-based community composition analysis in the Noumea lagoon.

Anonymous Referee #1

Referee 1: *P9055, L25-29: The authors suggest that the increase in UCYN-C could be due to their ability to utilize organic phosphorous when DIP concentrations become limiting. However, within the Lagoon, which experienced consistently low DIP, UCYN-C remained in low abundance. What was the availability of organic phosphorous within the lagoon? What other factors could have resulted in the significant increase in UCYN-C during P2 relative to the lagoon? These findings are very interesting but I feel this part of the manuscript requires further discussion.*

Response: DOP concentrations in the lagoon during the mesocosm deployment (day 2-22) were on average 0.105 ± 0.011 μmol L^{-1}, which is slightly less than the DOP concentrations in P1 and the start of P2 inside the mesocosms (Berthelot et al., 2015). Due to the complexity of the DOP pool, we cannot speculate any further about the differences between the DOP available inside and outside the mesocosms. This information has been added into the discussion. Several other potential factors contributing to the UCYN-C bloom may include increasing salinity inside the mesocosm, as UCYN-C abundances were found to be positively correlated to higher salinity both inside and outside the mesocosms, yet salinity increases inside the mesocosms were greater (Bonnet et al, 2015a). More speculative explanations include some component of the mesocosms themselves acting as a factor, such as UCYN-C rich biofilms developing (discussed in the original mansuscript) or a positive response to reduced turbulence (which we have mentioned in the revised discussion). We have added a more thorough discussion of the potential role of salinity into the revised discussion section and below.

In all three mesocosms, which acted as biological replicates, UCYN-C abundances were not only strongly correlated with the decreasing DIP concentrations (p=0.004, r=-0.51), but also increasing temperatures in the mesocosms (p=0.000, r=+0.86) and could be weakly correlated with increasing salinity (p=0.000,r=+0.56), decreasing Het-1 (p=0.002,r=-0.54), Het-2 (p=0.01, r=-0.37) and UCYN-A1 (p=0.000, r=-0.64) abundances. A similar correlation between UCYN-C abundance and both increasing salinity and temperature was observed outside in the Noumea Lagoon (although UCYN-C abundances in the lagoon remained low; see section 3.4). Salinity in each of the three mesocosms and in the lagoon did increase gradually during the experimental period, but were elevated inside the mesocosms in P2, likely due to evaporative loss (Bonnet et al., 2015a). Together, this data suggests that temperatures below 25.6^{o}C are not optimal temperatures for this group, and that it may tolerate slightly elevated salinities better than other diazotrophs present.

Referee 1: *P9059, L18: although the particulars of the mesocosm experimental set-up are referenced elsewhere, it would be good to mention if there are any possible aspects of the experimental design that could result in the differences observed between mesocosms.*

Response: Although we can't address this definitively, we feel the data indicates the mesocosms did replicate each other well with respect to many biogeochemical parameters, as well as picoplankton population dynamics. It is possible that the variations we observed in the timing of net growth and net mortality for each group in the mesocosms reflects natural variations in grazing pressure by copepods and other grazers, as well as natural variation in viral lysis. The text discussing this has been updated in the revised mansuscript and below.

There are no aspects of the experimental design that can be invoked to explain this variability; in fact biogeochemical parameters and picoplankton population dynamics were well replicated in all three mesocosms (Bonnet et al., 2015a). Therefore the dynamic nature of diazotroph growth and mortality rates in each mesocosm most likely results from a combination of grazing pressure and viral lysis, which can be expected to reflect natural variations in the grazers and virus present.

Referee 1: *P9060, L24: are there any studies exploring seasonal variability in Noumea Lagoon diazotroph communities or nitrogen fixation rates? Are Rhizosolenia and Hemiaulus conspicuous members of the phytoplankton community?*

Response: At this time, there are very few studies of the sort, and none that apply quantitative molecular approaches. Garcia et al., (2007) reported on N_2 fixation rates in the Noumea Lagoon and surrounding waters, and found that the temporal variability of the ^{15}N accumulating in the large size fraction (>10 um) was high. We have added this information into the introduction. There have been few prior reports of *Richelia* associated with *Rhizosolenia* and *Hemiaulus* in the local region, which is discussed in the introduction of the original manuscript. There are no studies describing seasonal variability of diazotroph community composition to our knowledge.

Referee 1: *P9076, Figure 1: I wonder if the lines are misleading because they imply a known trend between the sampling days.*

Response: We appreciate you pointing this out, and have removed this figure entirely, feeling it is redundant with the text and Figure 2. This data is available in Supplemental Table S3.

Referee 1: *P9078, Figure 3: I didn't see this figure referred to anywhere in the text.*

Response: This was referenced to on line 423, and added to line 738 (line numbers correspond to revised manuscript).

Referee 1: *P9044, L11: include a space between "unicellular" and "cyanobacterial"*

Response: This has been corrected.

Referee 1: *P9047, L15: remove "the"*

Response: This has been corrected.

Referee 1: *P9075, Table 2: missing ")". Perhaps provide the complete terms for DNQ and UD acronyms.*

Response: This has been corrected, and the complete terms for DNQ and UD have been inserted.

Anonymous Referee #2

Referee 2: *...I find it somewhat problematic that there are no indications of whether these nine phylotypes are important in this system. The manuscript deals with community succession and it would have been preferable to have some analyses of the actual diazotrophic communities present in the mesocosms and in the lagoon itself during the experiment (e.g. a nifH clone library or similar). An accompanying manuscript in preparation is mentioned (Berthelot et al. 2015 in prep). It is however, unclear whether the reader can find information on the composition of the diazotrophic community in this paper. Is this the case? And what were the overall findings regarding the community composition id that is the case? On several occasions, e.g. p. 12, l. 12-16; p. 18, l.11-12; Figure 2, the authors talk about abundances of specific phylotypes as fractions of the total diazotrophic community. Such deductions cannot be made as it is unknown/unlikely whether the nine selected primer/probe sets collectively target the entire community. I suggest addressing this matter and supplying a short general description of the present community as well as the relative abundances of the quantified phylotypes if these data are available. If they are not I would strongly suggest making these data if there are DNA left from the study. Alternatively, are there previous data describing the community in this location?*

Response: We have addressed this critique in several ways. All discussion referring to the % of total diazotroph community has been removed, bringing focus onto the absolute abundances measured. Supplemental Table S4 has been removed. Figure 2 has been re-plotted, and not normalized to the total *nifH* genes recovered from targeted assays.

Most significantly, because there was no published data that would have helped us address this concern, we selected a total of 11 samples (three from each mesocosm and two from the lagoon) for PCR amplification of the *nifH* gene, and sequenced these amplicons using a next generation sequencing approach. Details of our methodology have been added to the manuscript in section 2.3. Using this data, we discuss the diazotroph community composition measured using a method different than, but complementary to, qPCR-based measures. Our findings are detailed and discussed primarily in section 3.1, Figure 1, Table 1, and Supplemental tables S4 and S5, but also throughout the revised manuscript. The text from section 2.3 and 3.1 as well as Table 1 and Figure 1 are included at the end of this document. The results of this additional analysis indicate that the diazotrophs targeted in the qPCR assays are major lineages in the New Caledonia lagoon and in the mesocosms. The only significant exception is *Crocosphaera*, as *Crocosphaera*-like sequences recovered in the libraries from mesocosm samples would not amplify reliably with the qPCR assays used. Also of importance, the heterocyst-forming symbionts of diatoms (Het-1, -2, and -3) were practically absent in the *nifH* libraries, yet one of the most abundant diazotrophs quantified inside the mesocosms and in the lagoon, This finding munderscores the limitations of using qualitative PCR approaches.

Referee 2: *p. 12, l. 12-16: I disagree with the use of the term total diazotroph community since the community as such is not investigated. Do the authors have any data describing the relative abundance of this sequence compared to total nifH sequences – maybe in the in prep paper?*
p. 18, l.11-12: As above p. 20, l. 16-19: As above
Figure 2: As above

Response: Please see the above discussion, where we have addressed this concern in detail. All discussion of % of total diazotroph community has been removed.

Referee 2: *p. 2, l. 11: "unicellularcyanobacterial"*

Response: This has been corrected.

Referee 2: *p. 5, l. 15: Delete "the" in "understand the how"*

Response: This has been corrected.

Referee 2: *p. 5, l. 19-20: Move parentheses start to surround the 2015*
p. 6, l. 10: As above

Response: These have been corrected.

Referee 2: *p. 20, l. 6: change "NO3" to "NO3-" (may want to check throughout the MS) p. 20, l. 21: change "NH4" to "NO4+" (may want to check throughout the MS)*

Response: These have been corrected here, and in one other place in the manuscript.

Excerpts from the revised manuscript:

2.3 Determination of diazotroph community composition using high throughput sequencing of *nifH* amplicons

In order to evaluate whether the diazotrophs targeted via qPCR assays were representative of the phylotypes present in the lagoon and mesocosms, partial *nifH* fragments (ca. 360 base pairs) were amplified using a nested PCR assay and universal *nifH* primers nifH1-4, as described in Turk-Kubo et al., (2014). Two lagoon samples (day 1 and day 22), and three samples from each mesocosm (day 3, day 23, and the day where UCYN-C abundances were beginning to increase, i.e. days 11-15, see discussion below) were chosen for analysis. For each sample, triplicate PCR reactions were pooled. Internal primers were modified 5' common sequence (CS) linkers (CS1_nifH1F: 5'-ACACTGACGACATGGTTCTACATGYGAYCCNAARGCNGA, CS2_nifH2R: 5'-TACGGTAGCAGAGACTTGGTCTADNGCCATCATYTCNCC) to facilitate library

preparation at the DNA Services (DNAS) Facility at the University of Illinois, Chicago, using the targeted amplicons sequencing (TAS) approach described in Green et al. (2015). These libraries were pooled with other libraries to achieve a target depth of ca. 40,000 sequences per sample. Sequencing of paired end reads was performed at the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign using Illumina MiSeq technology. De-multiplexed raw paired end reads were merged in CLC Genomics workbench, and merged reads between 300-400 base pairs in length were selected. Quality filtering was performed in QIIME (Caporaso et al., 2010) using the usearch quality filter (usearch_qf) pipeline script, which includes steps for denoising, de novo chimera removal using UCHIME (Edgar et al., 2011) and operational taxonomic unit (OTU) determination using usearch6.1 at 97% nucleotide identity (Edgar 2010). Representative nucleotide sequences from OTUs with greater than 100 reads (277 out of 2325 OTUs, representing 92% of all sequences that passed the usearch quality filter) were imported into ARB (Ludwig et al., 2004), translated into protein sequences, where non-*nifH* OTUs or those with frameshifts were discarded. QIIME script exclude_seqs_by_blast.py was used to check for sequences with >92% amino acid identity to known contaminants; none were found. OTUs targeted by each qPCR assay was determined in silico for each group of diazotrophs in ARB by identifying representative sequences that had 0-2 mismatches in either primer or the probe binding region, without exceeding a total of 4 mismatches in all three regions (see Supplement Table S4).

For the characterization of the overall diazotroph community composition in the lagoon and mesocosms, representative sequences from the most highly recovered OTUs (109 OTUs representing 85% of all post quality-filtering sequences) were considered. Translated amino acid sequences were aligned to the existing amino acid alignment in the curated database. Maximum likelihood trees were calculated using translated amino acid sequences from representative sequences and their closest relatives (determined via blastp) in MEGA 6.06 (Tamura et al. 2013), using the JTT matrix based model and bootstrapped with 1000 replicate trees. Distribution of read data across samples for each representative sequence was visualized in the Interactive Tree of Life online tool (Letunic

and Bork, 2006). Raw reads (fastq files) were deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA).

3.1 Diazotroph community structure in VAHINE mesocosms experiment and in the Noumea Lagoon

In order to 1) characterize the diazotroph community composition in the mesocosms and the lagoon using a qualitative approach that is complementary to qPCR and 2) evaluate whether the qPCR assays used in this study, which are widely used in studies of the oligotrophic ocean, target the major lineages of diazotrophs in the Noumea Lagoon, fragments of the *nifH* gene were amplified from a subset of samples using a well-established nested PCR approach (Zehr and Turner, 2001), and sequenced using Illumina MiSeq technology.

In total, 636,848 paired end reads were recovered from 11 samples, and 544,209 passed the quality filtering steps described above (Supplemental Table S4). Clustering at 97% nucleotide identity yielded 2,325 OTUs with greater than 4 reads, and 334 OTUs with greater than 100 reads. 277 of these OTUs (representing 479,402 reads or 88% of all reads that passed quality filtering; Supplemental Table S4) remained after removing non-*nifH* reads or those with frameshifts, and were used in downstream analyses. A majority of these 277 OTUs (152 OTUs, representing 311,550 reads, or 65% of the reads selected for analysis) were affiliated with *nifH* cluster 1B Cyanobacteria. Reads affiliated with *nifH* cluster 1G, which is composed primarily of γ-proteobacterial phylotypes, was the second most highly recovered group (88 OTUs, representing 120,586 reads, or 25.2% of the reads selected for analysis). Reads that were closely related to *nifH* cluster 1J/1K, comprised primarily of α- and β-proteobacteria, were also recovered, but only comprised 8.4% of the reads selected for analysis (19 OTUs, 40,032 reads). Cluster 3 and cluster 1O affiliated reads were recovered, but together accounted for less than 2% of the reads selected for analysis (Table 1).

The two OTUs with the highest relative abundance (OTU1890 and OTU2317), accounted for 31% of the reads selected for analysis and were closely related to the prymnesiophyte symbiont UCYN-A2 (Supplemental Table S5). Both OTUs were

present in the lagoon in day 1 and day 22 samples, and were recovered at high relative abundances from all three mesocosms throughout the experiment (Fig. 1). The third most highly recovered OTU (OTU1) was a γ-proteobacteria closely related to γ-24774A11, a heterotrophic diazotroph with widespread occurrence (Moisander et al., 2014; Langlois et al., 2015), that is also preferentially amplified by the *nifH* primers used (Turk et al., 2011). This sequence type was present in the lagoon samples, and had high relative abundances in all three mesocosms in midpoint samplings (days 11, 13, and 15 for M1, M2, and M3, respectively; Fig. 1). OTU2280 (cluster 1J/1K) was the OTU with the fourth highest relative abundance. It does not have high sequence similarity to any uncultivated or cultivated organisms, with the closest relative, an uncultivated rhizosphere isolate (Genbank accession no. KC667160), sharing only 86% nucleotide sequence similarity. This is true for all but one of the 1J/1K OTUs, OTU119, which is closely related (98% nucleotide identity) to an environmental sequence recovered from Heron Reef (Genbank accession no. EF175779).

Also among OTUs that had high recovery were UCYN-A1 (OTU2008, OTU1754, and OTU2), other UCYN-A2 OTUs (OTU2325, OTU1548, and OTU664), *Trichodesmium* (OTU5 and OTU35), UCYN-C (OTU12) and two OTUs affiliated with cluster 1G (OTU2218 and OTU2199) (Fig.1 and Supplemental Table S5). 1G OTUs are recovered from the lagoon sample at day 22, and have high relative abundances in all three mesocosms by the end of the experiment. These sequence types are not closely related to γ-24774A11, thus are not quantified using qPCR assays in the study, and their quantitative importance in the mesocosm environment cannot be determined based on this qualitative measure. It is important to note that high relative abundances in PCR-based libraries is not indicative of high abundances, as often these sequence types dominate PCR libraries yet are present at low abundances in the environment (Hewson et al., 2007, Bonnet et al., 2013, Turk-Kubo et al., 2014, Shizoaki et al., 2014, Bentzon-Tilia et al., 2015), presumably as a result of preferential amplification (Turk et al., 2011).

A majority of the OTUs with high relative abundances (159 out of 277, representing 380,556 out of 479,402 reads) affiliated with the following lineages targeted by qPCR assays used in this study: UCYN-A1, UCYN-A2, UCYN-B, UCYN-C, *Trichodesmium*, and *Richelia* associated with *Rhizosolenia* (Het-1), and γ-24774A11. To

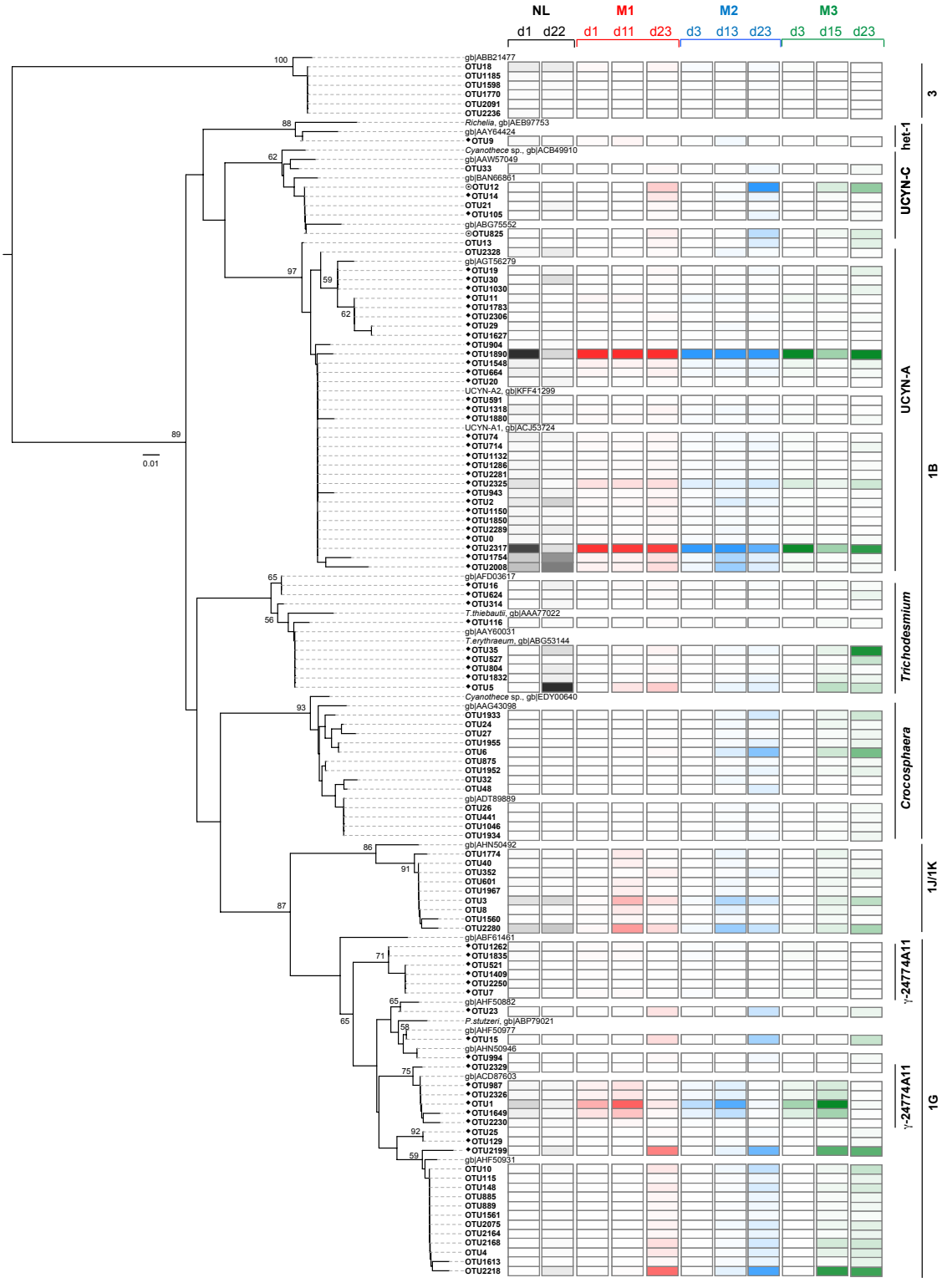
determine whether the qPCR assays used would target the diazotrophs present, representative sequences were identified that contained between 0-2 mismatches in the primer and probe binding regions, without exceeding a total of 4 mismatches for all three regions. For UCYN-A, *Trichodesmium*, Het-1, and γ-24774A11 lineages, nearly all of the sequence types present met these criteria (between 97-100%), thus would be quantified in the qPCR assays (Table 1). Only 26% of the recovered *Cyanothece*-like sequences would successfully be targeted by the UCYN-C assay, however, this coverage increases to 85% when including two of the most highly recovered OTUs that have a third mismatch at the 5' end of the probe binding region, thus are still likely to be quantified.

Table 1. ***In silico qPCR coverage analysis.*** Taxonomic assignment for all *nifH* amplicons reads that passed the quality filtering steps (first column), and the number of OTUs affiliated with each group that are successfully targeted by qPCR assays used in this study. Partial *nifH* sequences were classified according to the convention defined in Zehr et al. (2003). OTU – operational taxonomic unit.

		targeted by qPCR	
	no. OTUs (no. sequences)	**no. OTUs (no. sequences)**	**% OTUs (% sequences)**
1B	**152 (311550)**		
UCYN-A	68 (260221)	60 (258005)	88% (99%)
UCYN-B	32 (15917)	0 (0)	0% (0%)
UCYN-C	18 (11324)	9 (2915)	50% (26%) ⊙
Tricho.	18 (20034)	14 (19386)	78% (97%)
Het-1	2 (1738)	2 (1738)	100% (100%)
Het-2	0 (0)	0 (0)	0% (0%)
Het-3	0 (0)	0 (0)	0% (0%)
other	14 (2316)	na	
1G	**88 (120586)**		
γ-24774A11	22 (51594)	18 (50833)	82% (99%)
other	66 (68992)	na	
1J/1K	**19 (40032)**		
3	**16 (6825)**		
1O	**2 (409)**		

⊙ 61% (85%) when a third mismatch at the 5' end (probe) is allowed

Figure 1. ***Maximum likelihood tree calculated using partial*** **nifH** ***amino acid sequences recovered from the Noumea Lagoon (NL) and mesocosms (M1, M2, and M3).*** Relative abundances of *nifH* reads associated with each operational taxonomic unit (OTU) are indicated for each sample by shaded boxes, with intense shading indicating high relative abundances, and light shading indicating low relative abundances. Trees were bootstrapped using 1000 replicate trees, and nodes with values >50 are displayed. Branch lengths were inferred using the JTT model, and the scale bar indicates the number of substitutions per site. OTUs that are targeted by qPCR assays used in this study are marked with a black diamond (◆), and two UCYN-C sequences that are likely to amplify are marked with a circle (⊙). *nifH* cluster designations according to the convention in Zehr et al. (2003) are notated at the right. d1 – day 1; d11 – day 11, d13 – day 13, d15 – day 15, d22 – day 22, d23 – day 23.


NL
M1
M2
M3
d1 d22 d1 d11 d23 d3 d13 d23 d3 d15 d23
gb|ABB21477
OTU18
OTU1185
OTU1598
OTU1770
OTU2091
OTU2236
Richelia, gb|AEB97753
gb|AAY64424
OTU9
Cyanothece sp., gb|ACB49910
gb|AAW57049
OTU33
gb|BAN66861
OTU12
OTU14
OTU21
OTU105
gb|ABG75552
OTU825
OTU13
OTU2328
gb|AGT56279
OTU19
OTU30
OTU1030
OTU11
OTU1783
OTU2306
OTU29
OTU1627
OTU904
OTU1890
OTU1548
OTU664
OTU20
UCYN-A2, gb|KFF41299
OTU591
OTU1318
OTU1880
UCYN-A1, gb|ACJ53724
OTU74
OTU714
OTU1132
OTU1286
OTU2281
OTU2325
OTU943
OTU2
OTU1150
OTU1850
OTU2289
OTU0
OTU2317
OTU1754
OTU2008
gb|AFD03617
OTU16
OTU624
OTU314
T.thiebautii, gb|AAA77022
OTU116
gb|AAY60031
T.erythraeum, gb|ABG53144
OTU135
OTU527
OTU804
OTU1832
OTU5
Cyanothece sp., gb|EDY00640
gb|AAG43098
OTU1933
OTU24
OTU27
OTU1955
OTU6
OTU875
OTU1952
OTU32
OTU48
gb|ADT89889
OTU26
OTU441
OTU1046
OTU1934
gb|AHN50492
OTU1774
OTU40
OTU352
OTU601
OTU1967
OTU3
OTU8
OTU1560
OTU2280
gb|ABF61461
OTU1262
OTU1835
OTU521
OTU1409
OTU2250
OTU7
gb|AHF50882
OTU23
P.stutzeri, gb|ABP79021
gb|AHF50977
OTU15
gb|AHN50946
OTU994
OTU2329
gb|ACD87603
OTU987
OTU2326
OTU1
OTU1649
OTU2230
OTU25
OTU129
OTU2199
gb|AHF50931
OTU10
OTU115
OTU148
OTU885
OTU889
OTU1561
OTU2075
OTU2164
OTU2168
OTU4
OTU1613
OTU2218
100
88
62
97
59
62
89
0.01
65
56
93
86
91
87
71
65
65
58
75
92
59
3
het-1
UCYN-C
UCYN-A
Trichodesmium
Crocosphaera
1B
1J/1K
γ-24774A11
γ-24774A11
1G