Response to Reviewers

Reviewer # 1

Updated Figures, Supplemental Figures, and Supplementary Tables to be included per reviewer #1 comments

Table S6. Transition ions used for PstA protein quantification (peptide ATDEALQIVPR)

Peptide ATDEALQIVPR	y8 - 925.5465+
	y7 - 796.5039+
	b3 - 288.1190+
	b4 - 417.1616+
	b5 - 488.1987+
	b6 - 601.2828+

Table S7. All peptides targeted in PRM mode (note only PstA data reported here)

Description	Tery #	Metagenome ID	Peptide
IdiA periplasmic binding	Tery_3377	TCCM_0877.00000020	IFSEGNNEYPVVAGIPIATVLK
			HYDTDQALYDSFTQK
			FLEHLVSPEAQK
			ILYHDQNIYDPDIDPVEIR
phosphate ATP binding protein PstB	Tery_3540	TCCM_0018.00000090	LGQSGFALSGGQQQR
			NIDQQNSAAALSAEK
			IADVTAFFNAK
			ATDEALQIVPR
			GPLSPTLPSLAYLVYEFSR
phosphate permease PstA	Tery_3539	TCCM_0018.00000080	ATDEALQIVPR
			GPLSPTLPSLAYLVYEFSR
phosphate permease PstC	Tery_3538	TCCM_0018.00000070	VLIPAAFSGIVGGVMLGLGR
			AMGETMAVTMLIGNANSIK
SphX periplasmic binding	Tery_3534	TCCM_0018.00000040	GAIGYIEFGFAK
			INLVGAGASFPAPLYQR
			NSGFEVQVDYQSVGSGAGIER
PstS periplasmic binding less phosphate responsive	Tery_3537	TCCM_0018.00000050	EVYVDILLGNIK
			NDGVTAQITQTEGAIGYVEYGYAK
			VSPELGYIPLPDNVR
			YIEPTFESAEATLGAVALPENLR
flavodoxin		TCCM_0640.00000010	IGLFLGTTTGK
			FVGLALDDDNQAELTDER

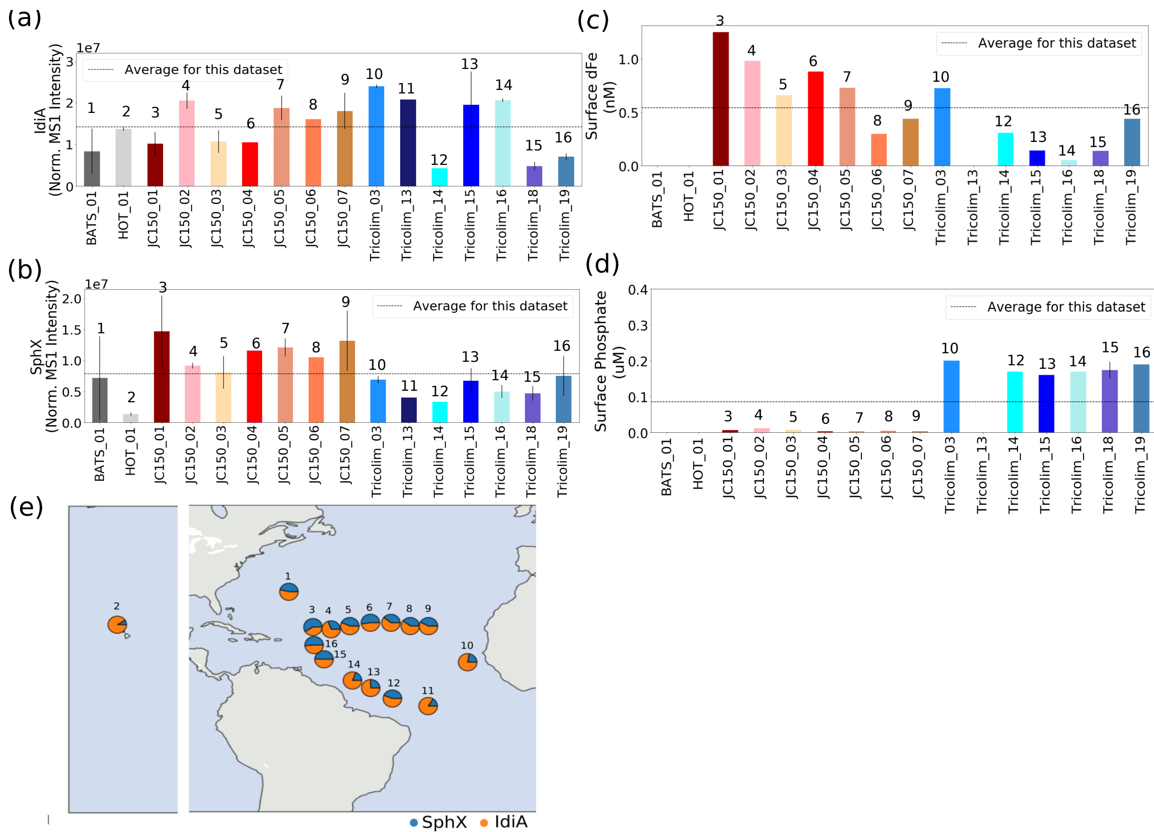

Figure 3. (A) Relative abundance of iron stress protein IdiA (A) and phosphate stress protein SphX (B). IdiA and SphX were among the most abundant proteins in the entire dataset. Error bars are one standard deviation on the mean when multiple samples were available. Dashed lines represent average values across the dataset. Proteins abundances were normalized such that the total MS1 peak area across the entire proteome was the same for each sample. (C) and (D) concentrations of dFe and phosphate nutrients. (E) Relative abundance of IdiA (orange) and SphX (blue) overlaid on the sampling locations.

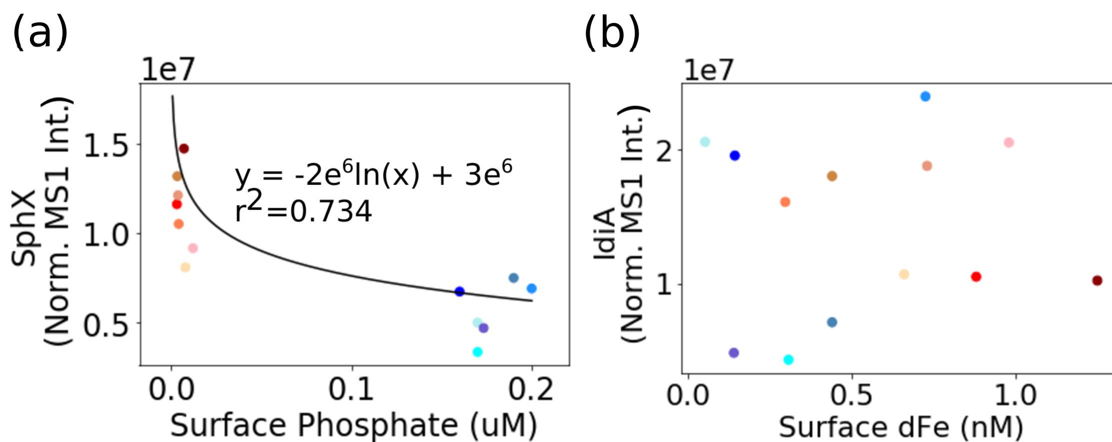

Figure S2. Scatter plot of A) SphX versus surface phosphate and B)IdiA versus surface dFe values. Note that analytical differences between JC150 (red dots) and Tricolim (blue dots) may be forcing the relationship with the surface phosphate values, though all values are above the limit of detection.

Response to reviewer #1 – reproduced here with text formatting for easier reading

General overview

Held et al., present a metaproteomic analysis of Trichodesmium isolates mostly collected in the Tropical and Subtropical Atlantic and use targeted proteomics to quantify putative markers of P and Fe stress to assess correlations with Nitrogenase. Given that Trichodesmium is a dominant fixer of nitrogen in the oceans, mechanisms con- trolling the abundance the major nitrogen fixing enzyme within this cyanobacter is of great interest to ocean biogeochemists (and should be of interest to all). The study involves a survey of Trichodesmium populations and across multiple cruises and dates which supports a more generalizable perspective of Trichodesmium response to low Fe or P. Rather than simply reporting data and immediate findings regarding correlations between proteins between P and Fe levels, the authors make a considerate attempt to extend hypotheses into the biophysical realm by setting up hypothetical scenarios of protein space competition that may lead to reduction in N fixation via membrane overcrowding. Despite lack of experimental evidence and sometime speculative, the authors cite precedent in many cases and have written a thought-provoking manuscript that should lead to testing of alternative hypotheses with regards to how Trichodesmium counters P and Fe limitations at the membrane in addition to regulatory gene expression.

We thank the reviewer for their thoughtful comments on our manuscript, particularly attention paid to the proteomics methodology. Responses to individual comments are provided below in red. Updated text has been provided for major edits specific to this review with changes highlighted in yellow.

Major comments

Abstract line 19, 20 the authors make reference to a '...specific physiological state under nutrient stress'. Given the fact that everything will have a specific physiological state under stress or non-stress conditions, this statement does not carry much impact. Perhaps the authors could be more specific. Did they intend to say that there is a generalized stress response regardless of stressors and that nitrogenase appears to comprise this stress response?

We agree with the reviewer on this important clarification. The intention was to indicate a specific physiological state under co-stress, distinct from that of an organism stressed for just one nutrient. We have added a phrase to clarify this sentence.

Abstract. *Trichodesmium* is a globally important marine microbe that provides fixed nitrogen (N) to otherwise N limited ecosystems. In nature, nitrogen fixation is likely regulated by iron or phosphate availability, but the extent and interaction of these controls are unclear. From metaproteomics analyses using established protein biomarkers for iron and phosphate stress, we found that co-stress is the norm rather than the exception for field *Trichodesmium* colonies. Counter-intuitively, the nitrogenase enzyme was more abundant under co-stress than under single nutrient stress, consistent with the idea that

Trichodesmium has a specific physiological state under nutrient co-stress, as opposed to single nutrient stress. Organic nitrogen uptake was observed to occur simultaneously with nitrogen fixation. Quantification of the phosphate ABC transporter PstA combined with a cellular model of nutrient uptake suggested that *Trichodesmium* is generally confronted by the biophysical limits of membrane space and diffusion rates for iron and phosphate acquisition in the field. Colony formation may benefit nutrient acquisition from particulate and organic nutrient sources, alleviating these pressures. The results highlight that to predict the behavior of *Trichodesmium*, both Fe and P stress must be evaluated and understood simultaneously.

Intro Line 47 – The use of the term 'established' when describing biomarkers is a bit subjective and loses impact when not followed by multiple citations, which in turn, lends credence to the idea that the protein markers are routinely accepted and utilized by the scientific community. Its reasonable for the authors to state that these proteins were utilized as biomarkers in this study based upon scientific precedent from expression studies, but the discriminatory power of these proteins to classify has not been thoroughly validated. The rationale provided in line 47-59 is sufficient for inclusion in the study. The word putative or candidate in lieu of 'established' seems a better fit.

We thank the reviewer for highlighting that protein biomarkers are not yet routinely utilized in the community. Change accepted.

There are several ~~established~~ protein biomarkers for Fe and P stress in *Trichodesmium,* all of which are periplasmic binding proteins involved in nutrient acquisition.

Line 151 -157. Point of clarification. The authors state that standard MS2 peak area was linear between 1 amol and 20 fmol per uL. Further that samples were spiked with standard to 10 fmol per uL and 10uL injected. The assumption is that 10uls was also injected for linearity testing of the standards. Please confirm.

Indeed we confirm that 10uL injections were used for linearity testing of the standard peptides. We now clarify this in the text:

The standard mixture was calibrated to establish the exact concentration of the peptides. A known amount (10 fmol μL^{-1}) of the commercially available Pierce standard peptide mixture (Catalog number 88320) and an apomyoglobin digest was spiked into the standard. The ratio of Pierce (isotopically labelled according to JPT standards) or apomyoglobin (light) to heavy standard peptide MS2 peak area was calculated and used to establish the final concentration of the standard peptide mixture (Fu et al., 2016; Milo, 2013). Multiple peptides were used for this calibration and the standard deviation among them was approximately 10%. Finally, the linearity of the peptide standard was tested by generating a dilution curveand ensuring that the concentration of each peptide versus MS2 peak area was linear between 0.001 and 20 fmol μL^{-1} concentration, using 10uL injections consistent with experimental injection volumes.

Methods: There is no mention of transition ions in the PRM section of the methods. For the sake of reproducibility, the authors should make reference to which transition ions were utilized for quantification (can be supplemental). Further the Trichodesmium genome and version should be referenced that was used for searching.

An additional supplemental file (Table S6) providing the transition ions used for PstA quantification will now be provided. The *Trichodesmium* genome used for the initial (DDA) search is referenced in Section 2.4. In PRM mode specific masses were targeted, and the list of targeted ions will now be provided (Table S7). In the process of compiling these tables we noticed a mistake in our prior manuscript – the peptide quantified belongs to PstA, not PstC. Both PstA and PstC are components of the same phosphate transport permease protein, so results and discussion are not affected however it is important to make this distinction, and this naming has been updated in the text.

Line 196 "This clustering indicated direct regulatory links between C and N fixation.." perhaps use suggests in lieu of indicates due to probabilistic nature of the association of protein covariance. Similar strong language should be avoided without follow-up direct experimentation which is beyond the scope of the study.

Agreed and updated, particularly in this section (3.1) which relies on multivariate statistics:

A self-organizing map analysis identified groups of proteins with similar profiles, i.e. proteins whose abundances changed cohesively, suggestive of proteins that may be regulated similarly (Reddy et al., 2016). This revealed the central importance of nitrogen fixation to *Trichodesmium*. The nitrogenase proteins were among the most abundant in the proteome and were located in clusters 1 and 2 (Figure 2 and Table S3). Also in these clusters were nitrogen metabolism proteins including glutamine synthetase, glutamine hydrolyzing guanosine monophosphate (GMP) synthase and glutamate racemase. This is consistent with previous reports finding that N assimilation is synchronized with nitrogen fixation (Carpenter et al., 1992).

Nitrogen fixation was closely linked to carbon fixation. Many photosystem proteins clustered with the nitrogenase proteins, including phycobilisome proteins, photosystem proteins, and the citric acid cycle protein 2-oxoglutarate dehydrogenase. This clustering indicated direct regulatory links between C and N fixation. The nitrogen regulators P-II and NtcA were also present in this cluster and may mediate this association. In non-nitrogen fixing cyanobacteria, high abundance of the nitrogen regulators NtcA and P-II is suggestive of nitrogen stress (Flores and Herrero, 2005; Saito et al., 2014). In diazotrophs, the role of these regulators is unclear because they do not respond to nitrogen compounds such as ammonia as they do in other cyanobacteria (Forchhammer and De Marsac, 1994). Here, clustering of NtcA and P-II with C and N fixation proteins suggests that they play a role in balancing these processes in field populations, though the details of this role has yet to be elucidated.

Line 228 – 229 "This observation contrasts with the current paradigm that Trichodesmium down regulates nitrogen fixation when it is Fe or P stressed. . ." This statement needs to cite the current published paradigm.

Change accepted:

This observation contrasts with the current paradigm that *Trichodesmium* down regulates nitrogen fixation when it is Fe or P stressed (Frischkorn et al., 2018, Ruoco, et al., 2018, Bergmann et al., 2012, Shi et al., 2007).

Line 267 – 278 when considering membrane protein space, the authors make a considerate attempt. Space and physics are often ignored; however, because the authors measured whole cell protein abundance and did not attempt to isolate the plasma membrane fraction, the % occupancy estimates are bound to be overestimates. A statement providing limitations of the estimate are needed here. Limitation of space and crowding on a membrane makes sense. Assigning the protein number to the membrane alone is not accurate.

This is an important point and a sentence has been added to the text to indicate this limitation:

"The calculation assumes that 100% of the PstC protein quantified is present in the plasma membrane, however it should be recognized that some fraction is likely present in the cytoplasm, leading to the possibility of over-estimation of membrane space occupied."

Line 321 "Thus, we conclude that in certain scenarios, lack of membrane space could indeed limit Fe and perhaps P acquisition by Trichodesmium." There is no disagreement with the rationale and calculations that led to this statement, but the statement is hypothetical in nature and the term 'hypothetical' should be included. The authors do a nice job in addressing model limitations in the next paragraph and go on to describe an artificial scenario where space limitation could produce further nutrient uptake limitation as additional proteins are made and transported to the membrane. One has to assume uptake activity does not change or is influenced by intracellular events or membrane compositional changes that lead to conformational changes; however, the idea that a generalized stress response to Fe or P could lead to a negative effect on uptake and more limitation due to space limits is fun to ponder.

We thank the reviewer for their positive comment and interest in this section. We have updated Line 321 to highlight uncertainty in the statement:

Thus, we conclude that in certain scenarios, lack of membrane space could hypothetically limit Fe and perhaps P acquisition by *Trichodesmium*.

Line 369 – 385, Conclusions. The conclusion section is written as a perspective which is fine given that the conclusions linked specifically to the data are stated within the results

and discussion. In line 372, the use of the word 'norm' is understandable given the common phrase 'the norm rather than the exception', but this might be contentious because normal is being assigned to all Trichodesmium based on 16 sampling sites mainly focused in the Atlantic. Fe and P stress may be more common that previously accepted or realized. If the authors feel strongly about this phrasing and believe the audience will be receptive and not over-interpret, then it is fine. Otherwise, perhaps it can be a bit more tempered. The same phrase was used in the Abstract, but due to space limitations the fact that authors were inferring norm based on their samples seems reasonable.

We thank the reviewer for this comment. We have altered our wording in this section to avoid over-reach (similar changes will also be made elsewhere in the paper):

Trichodesmium's colonial lifestyle likely produces challenges for dissolved Fe and P acquisition, which must be compensated for by production of multiple nutrient transport systems, such as for particulate iron and organic phosphorous, at a considerable cost. While laboratory studies have largely focused on single nutrient stresses in free filaments, these metaproteomic observations and accompanying nutrient uptake model demonstrate that Fe and P co-stress may be the norm rather than the exception, particularly in the North Atlantic ocean. This means that the emphasis on single limiting nutrients in culture studies and biological models may not capture the complexities of *Trichodesmium's* physiology in situ. Thus, biogeochemical models should consider incorporating Fe and P co-stress conditions. Specifically, in this study and in others there is evidence that nitrogen fixation is optimal under co-limited or co-stressed conditions, implying that an input of either Fe or P could counter-intuitively decrease N_2 driven new production (Garcia et al., 2015; Walworth et al., 2016).

Conclusion Line 378-385, membrane space limitation is likely to be confronted by all cells, not only Trichodesmium, but the idea is understood. The authors make a case for including co-stress based on protein observations and if nutrient stress is truly occurring (which can be difficult to define to everyone's liking and measure in situ; hence use of protein markers) then including these parameters in biological models makes sense.

We agree that membrane space is likely confronted by all cells, though few oceanographic studies have demonstrated or discussed this. In the interest of not over-reaching our argument we leave the reviewer's point out of our main text but hope that this work will stimulate future investigations.

Figure 3. This is a very informative and nice figure. If possible it would be great to see Figure S2 incorporated for ease of reading and comparison.

Both reviewer #1 and reviewer #2 suggested this change, so we now include the Figure S2 plots in the main text as part of Figure 3. Figure S2 now provides scatter plots of IdiA relating dFe and SphX versus phosphate concentrations. Per these changes and the comments from the reviewers the discussion in Section 3.2 has been updated as below.

3.2 *Trichodesmium* is simultaneously iron and phosphate stressed throughout its habitat

A surprising emergent observation from the *Trichodesmium* metaproteomes was the co-occurrence of the iron (IdiA) and phosphate (SphX) stress biomarkers across the samples. The ubiquitous and highly abundant presence of these proteins relative to total protein implied that co-stress may be the norm rather than the exception for *Trichodesmium* colonies in the field, particularly in the North Atlantic. Even though low-level basal expression of IdiA and SphX has been observed, it was clear that the colonies were devoting a large fraction of their cellular resources to Fe and P uptake, respectively (see Tables S8 and S9) (Webb et al., 2001, Webb et al., 2007, Chappell et al., 2010, Orchard et al., 2010, Snow et al., 2015, Walworth et al., 2016, Frischkorn et al., 2019). This, combined with the responsiveness of IdiA and SphX to nutrient availability in laboratory experiments on *Trichodesmium* filaments laboratory, indicated that co-stress was occurring.

Interestingly, biomarker abundance was not necessarily associated with nutrient concentrations in the surface ocean, suggesting that the colonies were experiencing stress despite variation in nutrient availability (Figure 3 C-D). SphX abundance varied up to 7.5 fold and may have been negatively associated with dissolved phosphate concentrations, though analytical differences across the field expeditions may have forced this relationship (Figure S2). Oceanographically, SphX was most abundant in the P-deplete, summer-stratified North Atlantic gyre (JC150 expedition) compared with winter waters near the Amazon river plume (Tricolim expedition) or at station ALOHA, where phosphate concentrations were greater (Hynes et al., 2009; Sañudo-Wilhelmy et al., 2001; Wu et al., 2000). IdiA varied up to 8 fold but there was no observable relationship with dFe concentrations at the surface. Instead, IdiA may be responsive to other factors such as the varying iron requirements of the populations/species examined here. For instance, it should be highlighted that in this study only *Trichodesmium* colonies were examined, so factors such as colony size may affect iron availability and biomarker expression. Additionally, because the surface ocean iron inventory is low, transient inputs such as from the Sahara desert can dramatically impact iron availability on short time scales, and the time scale of these inputs relative to changes in biomarker abundance is not well understood (Kunde et al., 2019). Carefully calibrated datasets relating IdiA and SphX abundance to nutrient-limited growth rates of *Trichodesmium* in both the filamentous and colonial forms would facilitate further interpretation of this data.

Figure 9 is unnecessary, but certainly would be welcomed by visual learners if space is not an issue. Otherwise the concepts are described in the discussion.

As the first author is a visual learner and space is not an issue, she would prefer to leave this in, but welcomes editorial advice on the matter!

Minor comments

Line 63-64"..suggesting nutrient stress was driven not only by biogeochemical gradients but also by Trichodesmium's inherent physiology". The term 'inherent physiology' is

very broad and does not add substance to the sentence. Trichodesmium is respond- ing to stress in the study and saying something like "... .but also by Trichodesmium's response to stress" puts the sentence in a category that doesn't contain every bio- chemical reaction in the organism.

Rephrased: "suggesting nutrient stress was driven not only by biogeochemical gradients but also by *Trichodesmium's* response to nutrient depletion"

Line 75. Table S1 does not correspond to the supplemental table indicated. Please correct the designation.

Change accepted – should read Table S2

Line 100 "... ...vacuum centrifugation to 1 100 μg μL-1 concentration." Assume this was estimated based on starting concentration of protein and not actually measured?

Yes, this is estimated from starting concentration and will now be noted in the text:

The resulting peptide mixture was concentrated by vacuum centrifugation to 1 μg μL^{-1} concentration estimated from the starting protein concentration.

Line 107 "C18 columns packed in house." Please add column size, diameter, c18 particle size and supplier.

This section has been updated with more details per the reviewer:

2.3 Sample acquisition

The global proteomes were analysed by online comprehensive active-modulation two-dimensional liquid chromatography (LC x LC-MS) using high and low pH reverse phase chromatography with inline PLRP-S (200μm x 150mm, 3μm bead size, 300A pore size, NanoLCMS Solutions) and C18 columns packed in house (100 mm x 150 mm, 3 μm particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH packed in a New Objective PicoFrit column). The first dimension utilized an 8 hour pH = 10 gradient (10mM ammonium formate and 10mM ammonium formate in in 90% acetonitrile), and was trapped every 30min on alternating dual traps, then eluted at 500nL/min onto the C18 column with a 30 min gradient (0.1% formic acid and 0.1% formic acid in 99.9% acetonitrile). 10 μg of protein was injected per run directly onto the first column using a Thermo Dionex Ultimate3000 RSLCnano system (Waltham, MA), and an additional RSLCnano pump was used for the second dimension gradient. The samples were then analyzed on a Thermo Orbitrap Fusion mass spectrometer with a Thermo Flex ion source (Waltham, MA). MS1 scans were monitored between 380-1580 m/z, with a 1.6 m/z MS2 isolation window (CID mode), 50 millisecond maximum injection time and 5 second dynamic exclusion time.

Line 110 A little more detail regarding the parameters would be useful. Based on the search parameters the instrument was likely operated in orbitrap/ion trap mode with

HCD? This would be of interest to include and assume more details are in located in Pride.

Details are located in Pride but the reviewer is right that more should be included in the main text.

Added sentence: "MS1 scans were monitored between 380-1580 m/z, with a 1.6 m/z MS2 isolation window (CID mode), 50 millisecond maximum injection time and 5 second dynamic exclusion time."

Line 120 should include the term "local FDR" if local FDR was used.

These are global FDRs (will now be noted).

Line 179-180 is repeated from methods section. Can be deleted.

Change accepted

Line 185 ribosomal and phycobilisomal

Change accepted

Line 237 – 257 This is quite fascinating although not the focus of the study. urtA substrate specificity is poorly defined outside of urea. Curious if the authors also found urease protein elevation in stressed samples.

Unfortunately urease protein abundances were very patchy (and were generally in low abundance) so we were unable to draw any specific conclusions. We did make an effort to note that increased abundance of the urea ABC transporter indicates general use of organic N sources including urea but also possibly TMA or other compounds.

Figure S2 "Note that the phosphate concentrations from the Tricolim cruise were not measured at the. . ." Do the authors mean to say Tricolim_13?

Per the reviewers we now include the data Figure S2 in the main text as part of Figure 3, and have updated the caption accordingly.

Supplemental figures. Please read through legends for spelling errors.

Change accepted.