1 Introduction

Biological soil crusts (BSC) are a microbial mat-like surface layer in arid soil. Millmeters in depth, BSC are found in plant interspaces and cover a wide, global geographic range [Garcia-Pichel et al., 2003b]. The ground cover of BSC on the Colorado Plateau has been measured as high as 80% by remote sensing [Karnieli et al., 2003]. The global biomass of BSC cyanobacteria alone is estimated at 54 x 10^{12} g C [Garcia-Pichel et al., 2003b]. BSC play important roles in arid ecosystem productivity and are responsible for significant nitrogen (N) flux (for review of BSC N-fixation see Belnap [2003]). For example, Evans and Belnap [1999] found approximately five times as many BSC samples from sites in North America, Africa and Australia had δ^{15}N values indicative of high N-fixation input relative to the number of samples where δ^{15}N indicated N input was predominantly from atmospheric deposition. Additionally, the presence of BSC is positively correlated with vascular plant survival due in part to BSC ecosystem N contributions (for review of BSC-vacular plant interactions see Belnap et al. [2003]).

Molecular studies of BSC microbial diversity include explorations of vertical BSC microbial diversity with BSC depth [Garcia-Pichel et al., 2003a], BSC *nifH* gene content surveys (e.g. Yeager et al. [2004], Yeager et al. [2012], Yeager et al. [2006] and Steppe et al. [1996]), and next-generation-sequencing (NGS) enabled studies of BSC SSU rRNA gene content across wide geographic ranges [Garcia-Pichel et al., 2013, Steven et al., 2013]. Garcia-Pichel et al. [2003a] found that BSC microbial diversity is organized vertically, likely as the result of vertically oriented environmental gradients (e.g. light and oxygen). *nifH* surveys have been conducted across BSC development stages [Yeager et al., 2004], as well as across seasons, temperatures and precipitation gradients [Yeager et al., 2012]. Mature, more fully developed BSC possess greater numbers of heterocystous cyanobacteria (e.g. *Nostoc, Syctonema*) than developing BSC but both young and old BSC are dominated by non-heterocystous cyanobacteria (*Microcoleus vaginatus* or *M. steenstrupii*) [Yeager et al., 2004, Garcia-Pichel et al., 2013]. Young or recently disturbed BSC are often described as "light" in appearance relative to "dark" mature BSC [Belnap, 2002]. Although an early study of Colorado Plateau BSC *nifH* diversity presented *nifH* genes related to *Gammaproteobacteria* as well as a clade that included *nifH* genes from the anaerobes *Clostridium pssteurianum*, *Desulfovibrio gigas* and *Chromatium buderi*, subsequent studies have found heterocystous cyanobacteria to be the numerically dominant BSC diazotrophs [Yeager et al., 2006, 2004, 2012]. Specifically, Yeager et al. [2006]–in a study of overall BSC *nifH* diversity–categorized 89% of 693 *nifH* sequences derived from Colorado Plateau and New Mexico BSC samples as heterocystous cyanobacterial (non-cyanobacterial *nifH* sequences were largely attributed to alpha- and beta- *proteobacteria*). The heterocystous cyanobacterial BSC diazotrophs fall into three genera, *Scytonema*, *Spirirestis*, and *Nostoc* [Yeager et al., 2006, 2012]. Studies of BSC microbial diversity over broad geographic ranges have elucidated how soil parent material correlates to above and below crust microbial community membership and structure [Steven et al.,

2013] and that the predominant BSC cyanobacterium shifts from *M. vaginatus* to *M. steenstrupii* with increasing mean annual temperature [Garcia-Pichel et al., 2013].

BSC N-fixation rate studies (typically employing the acetylene reduction assay (ARA)) have explored BSC diazotroph activity across various ecological gradients. Reported BSC N-fixation rates vary significantly [Evans and Lange, 2001]. The reasons for this variability are complex and likely include the spatial heterogeneity of BSC [Evans and Lange, 2001] and the impact of recent environmental conditions on N-fixation rates (see Belnap [2001] for discussion). Moreover, the ARA assay is subject to methodological artifacts that preclude cross-study and possibly intra-study but inter-environment type comparisons (see Belnap [2001] for review). Despite the general BSC N-fixation rate measurement variability, mature, dark BSC N-fixation rates have been measured higher than N-fixation rates for younger, light BSC [Belnap, 2002, Yeager et al., 2004]. This difference may be due to the proliferation of heterocystous cyanobacteria in older mats and is consistent with the theory that heterocystous cyanobacteria are the primary BSC diazotrophs. Alternatively, the N-fixation rate differences between young and old BSC might be attributable to methodological artifacts. For instance, Johnson et al. [2005] show that N-fixation rates measured from intact cores of developing BSC may be artifcatually low due to delayed acetylene/ethylene diffusion through the crust in a typical ARA incubation timeframe. When total N-fixation rates were calculated by integrating N-fixation rates over 1-3 mm depth BSC core slices along the full BSC core (thus mitigating ethene/acetylene flux limitations), N-fixation rate differences between young and old BSC were not statistically significant [Johnson et al., 2005].

The influence of microbial community membership and structure on BSC N-fixation is an ongoing research question [Belnap, 2013]. While the presence/abundance of heterocystous cyanobacteria has been proposed as the underlying microbial membership influence on increased N-fixation in mature BSC, it is unclear if the premise that mature BSC fix more N is always correct (see Johnson et al. [2005]). More studies are necessary to elucidate the microbial membership influence on BSC N-fixation and to determine if heterosystous cyanobaceria are the only keystone diazotrophs. To further probe the diversity of diazotrophs in BSC we conducted ^{15}N DNA stable isotope probing (DNA-SIP) experiments with light, developing Colorado Plateau BSC. Although molecular characterizations of BSC *nifH* diversity in other studies have yielded predominantly heterocystous cyanobacterial *nifH* genes, in ths study microbes from young, developing BSC that incorporated N from N_2 into DNA as determined by DNA-SIP were not cyanobacteria but members of the *Gammaproteobacteria*, *Clostridiaceae* and *Deltaproteobacteria*. Further, we explore the distribution of putative diazotrophs uncovered in this study in addition to heterocystous cyanobacteriadia studied by Yeager et al. [2004], Yeager et al. [2006] and Yeager et al. [2012] through collections of NGS SSU rRNA libraries from BSC microbial diversity surveys over a range of spatial scales and soil types [Garcia-Pichel et al., 2013, Steven et al., 2013].

2 Results

2.1 Comparison of sequence collections at "study"-level

2.1.1 Comparisons of OTU content

Of the 4340 OTU centroids established for this study (including sequences from Steven et al. [2013] and [Garcia-Pichel et al., 2013]) 445 and 870 have matches in the Living Tree Project (LTP) (a collection of 16S gene sequences for all sequenced type strains [Yarza et al., 2008]) at greater or equal than 97% and 95% sequence identity, respectively (LTP version 115). Similar numbers of total OTUs were found in each data set explored in this study (i.e. the DNA-SIP data presented here, the data presented by Steven et al. [2013] and by Garcia-Pichel et al. [2013]). Specifically, there were 3079 OTUs (209,354 total sequences after quality control) in the DNA-SIP data, 3203 OTUs (129,033 total sequences after quality control) in the Garcia-Pichel et al. [2013] study, and 2481 OTUs (129,358 total sequences after quality control) in the Steven et al. [2013] study. The DNA-SIP data set shares more OTUs with the Steven et al. [2013] (56% of total OTUs found in either of the two data sets) than it does with the Garcia-Pichel et al. [2013] data (46% of total OTUs between both data sets). The Steven et al. [2013] and Garcia-Pichel et al. [2013] only share 46% of OTUs.

2.1.2 Comparisons of Taxonomic Content

Cyanobacteria and *Proteobacteria* were the top two phylum-level sequence annotations for all three studies but only the DNA-SIP data had more *Proteobacteria* annotations than *Cyanobacteria*. *Proteobacteria* represented the 29.8% of sequence annotations in DNA-SIP data as opposed to 17.8% and 19.2% for the Garcia-Pichel et al. [2013] and Steven et al. [2013] data, respectively. Figure 1 shows the distribution of phylum-level sequence annotations for each study in the nine most abundant phyla across all studies, as determined by raw sequence counts. There is a stark contrast in the total percentage of sequences annotated as *Firmicutes* between the raw environmental samples and the DNA-SIP data. *Firmicutes* represent only 0.21% and 0.23% of total phylum level sequence annotations in the Steven et al. [2013] and Garcia-Pichel et al. [2013] studies, respectively. In the DNA-SIP sequence collection *Firmicutes* make up 19% of phylum level sequence annotations. Also in sharp contrast for the DNA-SIP versus environmental data is the number of putative heterocystous *Cyanobacteria* sequences. Only 0.29% of *Cyanobacteria* sequences in the DNA-SIP data are annotated as belonging to "Subsection IV" which is the heterocystous order of *Cyanobacteria* in the Silva taxonomic nomenclature [Pruesse et al., 2007]. In the Steven et al. [2013] and Garcia-Pichel et al. [2013] studies 15% and 23%, respectively, of *Cyanobacteria* sequences are annotated as belonging to "Subsection IV".

2.2 Ordination of CsCl gradient fraction SSU rRNA libraries

Ordination of Bray-Curtis [Bray and Curtis, 1957] distances between CsCl gradient fraction sequence libraries with principal coordinates analysis shows the labeled gradient fraction libraries diverge from control in the heavy fractions (Figure 2). When the labeled and control gradient fractions are paired such that each pair contains a control fraction and labeled fraction with a density difference below XXX g/mL, the Bray-Curtis distance between the fraction pair is postively correlated to the density of the labeled fraction (p-value: 0.00052, r^2: 0.3315) (inset Figure 2). Additionally, the label/control category for heavy fractions is statistically significant by the Adonis test (p-value: 0.001, r^2: 0.136) [Anderson, 2001]. The first principal axis appears to be correlated with fraction density (Adonis test p-value for density with all CsCl fraction libraries: 0.001, r^2 0.117).

2.3 Identities of possible ^{15}N incorporators

The OTUs that have proportion means in heavy fractions that are enriched in labeled versus control gradients are those that have responded to the stable isotope tracer, which would indicate diazotropy in this experiment. We found 38 responders total using a false discovery rate threshold for multiple comparison adjusted p-values of 10%. Of these 38, 26 are annotated as *Firmicutes*, 9 as *Proteobacteria*, 2 as *Acidobacteria* and 1 as *Actinobacteria* (The inset of Figure 3 summarizes the Family level taxanomic profile of stable isotope responders). Figure 3 summarizes the ratio of proportion means for each OTU where means are calculated from proportions in heavy fractions within labeled or controlled gradients and the ratio is labeled over control (see methods). If the OTUs are ranked by descending, moderated proportion mean labeled:control ratios, the top 10 ratios (i.e. the 10 OTUs that were most enriched in the labeled gradients in heavy fractions) are either *Firmicutes* (6 OTUS) or *Proteobacteria* (4 OTUs). Table X summarizes the results from BLAST searching the centroid sequences for these top 10 OTUs against the LTP (version 115). The *Proteobacteria* OTU centroid seqeunces for the top 10 responders all share high identity (>98.48% identity, Table X) with cultivars from genera known to possess diazotrophs including *Klebsiella*, *Shigella*, *Acinetobacter*, and *Ideonella*. None of the *Firmicutes* OTUs in the top 10 responders share greater than 97% sequence identity with sequences in the LTP (relase 115) (see Table X).

2.4 Distribution of BSC Diazotrophs in Environmental Samples

2.4.1 Non-Cyanobacterial Taxa

Clostridiacea Five of the 6 *Firmicutes* in the top 10 responder OTUs belong in the *Clostridiacea*. We only observed one of these strongly responding

Clostrideacea in the data presented by Garcia-Pichel et al. [2013], "OTU.108" (closest BLAST hit in LTP Relase 115 – *Caloramotor proteoclasticus*, BLAST %ID 96.94, Accession X90488). OTU.108 was found in two samples both characterized as "light" crust. One other *Clostrideacea* OTU with a proportion mean ratio (labeled:control) p-value less than 0.10 but outside the top 10 responders was found in the Garcia-Pichel et al. [2013] data and also in a "light" crust sample. None of the strongly responding *Clostridiacea* were found in the sequences provided by Steven et al. [2013].

Figure 11 depicts the phylogenetic breadth of *Clostridiaceae* N responder OTUs from this experiment. The phylogenetic tree was constructed from nearly full-length reference sequences, and edge width demonstrates the placements of short OTU centroid sequences in the backbone tree (see methods for description of placement algorithm and selection criteria for reference sequences). As shown, *Clostridiaceae* N-responder OTU centroid 16S sequences are generally more closely related to environmental than cultivar 16S gene sequences.

Gammaproteobacteria Only "OTU.342" (closest BLAST hit in LTP Release 115, BLAST %ID 100, Accession ZD3440, *Acinetobacter johnsonii*) of the *Proteobacteria* OTUs in the top 10 most strongly responding OTUs was found in the Garcia-Pichel et al. [2013] sequences. None of the strongly responding *Protebacteria* OTUs were found in the Steven et al. [2013] sequences. There were 133 responder OTU-sample occurrences (SIP responding OTU was found in a sample library) in the Steven et al. [2013] data. 83 were in "below crust" samples, 50 in BSC samples.

Other taxa Two potentially diazotroph OTUs were found in an extensive number of environmental samples (61 of 65 samples from the combined data sets of Garcia-Pichel et al. [2013] and Steven et al. [2013]). Both OTUs were annotated as *Acidobacteria* but shared little sequence identity to any cultivar SSU rRNA gene sequences in the LTP (Release 115), with best LTP BLAST hits of 81.91 and 81.32 % identity. Additionally, the evidence for N incorporation for each OTU was weak relative to other putative responders (adjusted p-values of 0.090 and 0.096). Of the remaining 36 stable isotope responder OTUs, only 14 were observed in the environmental data. Figure 8 summarizes the OTU-sample occurrences in both the Steven et al. [2013] and the Garcia-Pichel et al. [2013] data with occurrences distributed into the most relevant sample classes of each respective study.

2.4.2 Heterocystous *Cyanobacteria*

At least one of the six OTUs defined by sequences reported by Yeager et al. [2006] (see Table 1) was found in 21 of the 23 sites surveyed by Garcia-Pichel et al. [2013]. OTUs defined by *Scytonema hyalinum* FGP-7A and *Scytonema hyalinum* DC-A 16S rRNA gene sequences were found in 18 and 17 sites, respectively. *Nostoc commune* MCT-1 and *Spirirestis rafaelensis* LQ-10 defined OTUs we each found in 16 sites. The OTU defined by *Nostoc commune* MFG-1

was found in 12 sites and the OTU defined by *Calothrix* MCC-3A was only found in one site surveyed by Garcia-Pichel et al. [2013]. The opposite BSC relative abundance relationships of *Microcoleus Vaginatus* and *M. Strenstrupii* with site mean annual temperature was a major finding by Garcia-Pichel et al. [2013]. Garcia-Pichel et al. [2013] did not report the relationship of diaztrophic cyanobacteria with temperature although a comment by Belnap [2013] briefly discusses a qualitative positive relationship of *Scytonema* with temperature in the Garcia-Pichel et al. [2013] data. In agreement with the Belnap [2013] interpretation we found a positive relationship of *Scytonema hyalinum* FGP-7A and DC-A OTU relative abundance with mean annual temperature (p-values $3.332\text{x}10^{-03}$ and $3.173\text{x}10^{-04}$, respectively) (Figure 5). We also found *Nostoc commune* MCT-1 and MFG-1 OTU relative abundance was inversely related to mean annual temperature (p-values $1.307\text{x}10^{-02}$ and $1.577\text{x}10^{-06}$, respectively) (Figure 5).

At least one OTU defined by selected 16S rRNA gene sequences presented by Yeager et al. [2006] (Table 1) was found in all but 7 of 42 samples surveyed by Steven et al. [2013] and all of these 7 lacking the Yeager et al. [2006] OTUs were "below crust" samples. Table X summarizes the distribution of Yeager et al. [2006] sequence defined OTUs in Steven et al. [2013] samples. As expected all of the six OTUs defined by Yeager et al. [2006] sequences were more abundant in the crust samples than below crust samples (Figure 6) (maximum p-value for any OTU: $1.96\text{x}10^{-4}$).

2.5 Richness estimates

Figure 7 (inset) summarizes the fraction of observed OTUs of total OTUs as estimated by CatchAll for each sample 16S library. Rarefaction curves for each sample are shown in Figure 7. Qualitatively, rarefaction curves show below crust samples to be more rich than BSC samples in the Steven et al. [2013] data.

3 Discussion

3.1 Ordination of CsCl gradient fraction 16S libraries

The ordination of Bray-Curtis distances between CsCl gradient fraction 16S libraries for each day show that control fractions differ from labeled fractions in the "heavy" range of the CsCl gradients (Figure 2). If each control fraction is paired to the labeled fraction from the same incubation day that it is closest in density to and the Bray-Curtis distances for each pair are plotted against the density of the labeled fraction, there is a positive and statistically significant correlation beweet Bray-Curtis distance and density (see inset Figure 2). Therefore, the "heavy" end of the control and labeled gradients differ and the OTUs enriched in the labeled fractions would have incorporated N into their DNA during the incubation timeframe. If the incubation timeframe is appropriate, the N-incorporators would be likely diazotrophs.

3.2 BSC diazotrophs identified in the study

BSC N-fixation has long been attributed to heterocystous cyanobacteria and molecular microbial ecology surveys of BSC *nifH* gene content have been consistent with this hypothesis finding cyanobacterial *nifH* types to be numerically dominant in *nifH* gene libraries [Yeager et al., 2006, 2004, 2012]. It is possible, however, that PCR-driven molecular surveys of *nifH* gene content have been biased against non-heterocystous cyanobacteria (CITE GABY). Unfortunately, it is impossible to assess or quantify this bias (in either direction) without knowing the *nifH* gene content *de novo*. Perhaps non-PCR based molecular data such as metagenomic DNA sequence libraries will provide additional evidence with respect to the relative abundances of BSC *nifH* gene types. Additionally, heterocysts (the specialized N-fixing cells along the trichome of filamentous heterocystous cyanobacteria such as *Nostoc* and *Scytonema*) may be overepresented with respect to non-heterocyst N-fixing cells in *nifH* libraries because the heterocysts make up a fraction of the total cells along a trichome and even the non-heterocyst cells in a trichome will possess the *nifH* gene. It should also be noted that *nifH* gene content is not directly extrapolable to the taxonomic relative abundances of nitrogenase proteins.

We did not observe evidence for N-fixation by heterocystous cyanobacteria in the "light" crust samples used in this study. One possible explanation for our results is that the the "light", still developing BSC samples used in this study possessed less heterocystous cyanobacteria than dark mature BSC as has been observed in previous comparisons of light and dark BSC [Yeager et al., 2004]. Indeed, only 0.29% of seqeunces from this study's DNA-SIP 16S rRNA gene sequence libraries were from heterocystous cyanobacteria (see results) as opposed to 15% and 23% of total sequences in the Steven et al. [2013] and Garcia-Pichel et al. [2013] data, respectively. It is difficult to compare relative abundance values from CsCl gradient fractions against environmental libraries, but, a three order of magnitude difference between the environmental librares and the CsCl gradient fractions is stark. Nonetheless, we would still expect even low abundance diazotrophs to show evidence for N-incorporation, provided sequence counts were not too sparse in heavy fractions. The OTUs defined by selected heterocystous cyanobacteria sequences presented in Yeager et al. [2006], however, all fall below the sparsity threshold used in our analysis (see methods, Figure 9). Given the sparsity of heterocystous cyanobacteria sequences in the DNA-SIP data set, it is not possible to assess whether heterocystous cyanobacteria incorporated N during the incubation.

The OTUs that did appear to incoporate N during the incubation were predominantly *Proteobacteria* and *Firmicutes*. The *Proteobacteria* OTUs for which N-incorporation signal was strongest all shared high sequence identity (>=98.48% sequence identity) with 16S sequences from cultivars in genera with known diazotrophs (Table X). The *Firmicutes* that displayed signal for N-incorporation (predominantly *Clostridiaceae*) were not closely related to any cultivars (Table 10, Figure 11). There appears to be a gap in culture collections for these BSC diazotrophs. As culture-based ecophysiological studies have

proven useful towards explaining ecological phenomena in BSC 16S rRNA gene sequence libraries [Garcia-Pichel et al., 2013], it would seem that these putative *Clostridiaceae* diazotrophs would be prime candidates for targeted culturing efforts. Assessing the physiological response of these diazotrophic *Clostridiaceae* to temperature would be useful towards predicting how climate change will affect the BSC nitrogen budget.

Although too undersampled in the environmental data sets to reach statistical conclusions, non-heterocystous diazotrophs were found more often in below crust samples (as opposed to actual BSC samples) from the Steven et al. [2013] and in "light" BSC samples in the Garcia-Pichel et al. [2013] data (Figure 8). This result generates some hypotheses that are counter to prior conjecture regarding BSC diazotroph temporal dynamics (keeping in mind this phenomenon has not been evaluated statistically). Specifically, the transition of BSC from a light colored, developing crust to a dark, mature crust may not mark the emergence of diazorophs in BSC but rather the transition of the diazotroph community from heterotroph dominance to cyanobacterial. Additionally, the soil beneath BSC may contribute significantly to the N budget in arid ecosystems.

It is unclear why BSC *nifH* gene surveys have overwhelmingly recovered heterocystous, cyanobacterial *nifH* genes which would be in contrast to our results. Even poorly developed BSC samples have yielded predominantly cyanobacterial *nifH* genes [Yeager et al., 2004]. And, "sub-biocrust" samples have yielded *entirely* heterocystous cyanobacterial *nifH* genes [Yeager et al., 2012]. One explanation is that the samples from this study are simply different in diazotrophic community structure than those surveyed in Yeager et al. [2006] Yeager et al. [2004] and Yeager et al. [2012]. Indeed, it appears that the "light" crusts used here had a paucity of heterocystous cyanobacteria from the beginning (see above). It should be noted that "light" and in particular "sub-biocrust" samples possess much less heterocystous cyanobacteria in general (Figure 6) so the samples used in this study are not necessarily unrepresentative of typical poorly developed BSC simply because they're lacking significant numbers of heterocystous cyanobacteria. Additionaly, cyanobacterial *nifH* genes would be found in every heterocystous cyanobacterial cell, not just the heterocysts. Therefore, the relative abundance of heterocystous cyanobacteria in *nifH* gene libraries could easily overwhelm the numbers of *nifH* genes from non-heterocystous diazotrophs. Polyploidy could further exacerbate this bias as many cyanobacteria are estimated to have multiple genome copies per cell [Griese et al., 2011]. In any case, the DNA-SIP discovered diazotrophs for the "light", poorly developed BSC used in the study were not cyanobacterial but it is unknown if non-cyanobacterial diazotrophs would be identified by DNA-SIP with ^{15}N using mature BSC samples. Regardless, our results suggest that BSC N-fixation may include a significant non-cyanobacterial component that requires further assessment across a more comprehensive sampling of BSC types.

3.3 Sequencing depth

While it is somewhat alarming how few of the putative diazotrophs found in this study were also found by Garcia-Pichel et al. [2013] and Steven et al. [2013], it is important to point out that even next-generation sequencing efforts of BSC 16S rRNA genes have only shallowly sampled the full diversity of BSC microbes. Rarefaction curves of all samples from Steven et al. [2013] and Garcia-Pichel et al. [2013] are still sharply increasing especially for "below crust" samples (Figure 7). Parametric richness estimates of BSC diversity indicate the Steven et al. [2013] and Garcia-Pichel et al. [2013] sequencing efforts recovered on average 40.5% (sd. 9.99%) and 45.5% (sd. 11.6%) of existing 16S OTUs from samples (inset Figure 7), respectively. Further, the Steven et al. [2013] and Garcia-Pichel et al. [2013] only share 57.6% of total OTUs found in at least one of the studies. In fact, this study shares more OTUs with Steven et al. [2013], 62.4% of total OTUs between both studies, than the Steven et al. [2013] study shares with Garcia-Pichel et al. [2013].

3.4 Temperature influences on heterocystous cyanobacteria relative abundance

Although few putative diazotrophs identitied by DNA-SIP were found in the Garcia-Pichel et al. [2013] and Steven et al. [2013] data, we did make some new observations regarding the relationship of several heterocystous cyanobacterial OTUs with site mean annual temperature. Specifically, we found *Nostoc commune* MCT-1 and MFG-1 relative abundances were negatively correlated with sample mean annual temperature. Additionally, it appears that the relative abundances *Scytonema hyalinum* FGP-7A are positively correlated with mean annual temperature.

Yeager et al. [2012] found *nifH* gene abundance changes seasonally peaking in early summer and falling in autumn. Although Yeager et al. [2012] also experimentally increased the ambient temperature of several BSC samples over a long period (up to two years), changes in ambient temperature did not influence *nifH* gene abundance as measured by qPCR. We are not able to confirm these results using the data from Garcia-Pichel et al. [2013] which is compositional in nature as opposed to absolute but it does appear that temperature affects the structure of heterocystous cyanobacterial diazotroph communities if not the absolute abundance of *nifH* genes.

3.5 Analysis of next-generation-sequencing DNA-SIP data

Although DNA-SIP is a powerful technique, analysis of DNA-SIP data is not without ambiguities. One limitation is the artificial boundary in the form of a selected adjusted p-value threshold (or false discovery rate) that marks which OTUs we consider to be enriched in the heavy fractions of labeled CsCl gradients (and thus have likely incorporated N into their DNA during the incubation). In reality the metric we use to quantify the magnitude of an OTU's

response to a stable isotope is continuous and there is only an artificial boundary between which OTUs appear to have "responded" and which OTUs have unknown response. For this reason, we have presented all the OTUs that satisfy our "response" criteria but focused on the most strongly responding OTUs. As with any hypthesis-based statistical test, care should be taken when interpreting the significance of results where p-values are near the selected "significance" threshold for rejecting the null hypothesis.

3.6 Conclusion

It would seem unlikely given their ubiquity and abundance that heterocystous cyanobacteria are not key contributors to the BSC N-budget. But, the putative diazotrophs elucidated in this study and in Steppe et al. [1996] in addition to the N-fixation rate data presented by [Johnson et al., 2005] suggest there may be additional and significant non-cyanobacterial BSC diazotrophs specifically within the *Clostrideaceae* and *Proteobacteria*. It seems clear that heterocystous cyanobacteria increase in abundance with BSC age [Yeager et al., 2004]. It is less clear if this transition marks the emergence of diazotrophy versus a re-structuring of the BSC diazotroph community from one dominated by *Firmicutes* and *Proteobacteria* to one predominantly heterocystous cyanobacteria. DNA-SIP is a valuable tool in the molecular microbial ecologists toolbox for identifying members of microbial community functional guilds CITE. PCR-based surveys of diagnostic marker genes and DNA-SIP both target the same generic information but they occupy a non-overlapping set of strenghts and weaknesses. Combined these tools can powerfully untangle connections between ecosystem membership/structure and function. Here we supplement surveys of BSC *nifH* diversity not confirming previous results but expanding our knowledge BSC diazotroph diversity. Evaluating BSC N-fixation due climate change and physical disturbance requires a careful accounting of diazotrophs including non-cyanobacterial types.

4 Materials and Methods

4.1 Field sites

4.2 Soil crust incubation

4.3 DNA extraction

DNA from each sample was extracted using a MoBio PowerSoil DNA Isolation Kit (following manufacturers protocol, but substituting a 2 minute bead beating for the vortexing step), and then gel purified. Extracts were quantified using PicoGreen nucleic acid quantification dyes (Molecular Probes).

4.4 DNA-SIP

Gradient density centrifugation of DNA was undertaken in 6 mL polyallomer centrifuge tubes in a TLA-110 fixed angle rotor (both Beckman Coulter) in CsCl gradients with an average density of 1.725 g mL-1. Average density for all prepared gradients was checked with an AR200 refractometer before runs. Between 2.5- 5 g of DNA extract was added to the CsCl solution, and gradients were run under conditions of 20C for 67 hours at 55,000 rpm (Lueders et al., 2004). Centrifuged gradients were fractionated from bottom to top in 36 equal fractions of 100 L, using a displacement technique similar to Manefield et al. (2002). The density of each fraction was determined using a refractometer. DNA in each fraction was desalted through four washes with 300 L TE per fraction.

4.5 PCR, library normalization and DNA sequencing

Bacterial and archaeal 16S rRNA genes from each fraction were quantified through real-time PCR, using primers Ba519f/Ba907r (Stubner, 2002). Each 25 L reaction contained 1X Quantifast SYBR Green Master Mix (Qiagen), 0.3 M of each primer, and 1 L of a 1:100 dilution of fraction DNA. Thermal cycling occurred with an initial denaturation step of 10 minutes at 95C, followed by 40 cycles of amplification (15s at 95C, then 60s at 60C). After each run, a melt curve was measured and recorded between 60C and 95C. Quantification was achieved through use of a dilution series (108-101 copies/L) from nearly full length 16s rRNA gene amplicons from pure culture DNA of K. pneumoniae.

Barcoded PCR of bacterial and archaeal 16S rRNA genes, in preparation for 454 Pyrosequencing, was carried out using primer set 515F/806R [Walters et al., 2011]. The primer 806R contained an 8 bp barcode sequence, a "TC" linker, and a Roche 454 B sequencing adaptor, while the primer 515F contained the Roche 454 A sequencing adapter. Each 25 μL reaction contained 1x PCR Gold Buffer (Roche), 2.5 mM MgCl2, 200 μM of each of the four dNTPs (Promega), 0.5 mg/mL BSA (New England Biolabs), 0.3 μM of each primers, 1.25 U of Amplitaq Gold (Roche), and 8 μL of template. Template for each sample was added at normalized amounts in an attempt to prevent chimera formation, and each sample was amplified in triplicate. Thermal cycling occurred with an initial denaturation step of 5 minutes at 95C, followed by 40 cycles of amplification (20s at 95C, 20s at 53C, 30s at 72C), and a final extension step of 5 min at 72C. Triplicate amplicons were pooled and purified using Agencourt AMPure PCR purification beads, following manufacturers protocol. Once cleaned, amplicons were quantified using PicoGreen nucleic acid quantification dyes (Molecular Probes) and pooled together in equimolar amounts. Samples were sent to the Environmental Genomics Core Facility at the University of South Carolina (now Selah Genomics) to be run on a Roche FLX 454 pyrosequencing machine.

4.6 Data analysis

4.6.1 Sequence quality control

Sequences were initially screened by maximum expected errors at a specific read length threshold [Edgar, 2013] which has been shown to be as effective as denoising 454 reads with respect to removing pyrosequencing errors. Specifically, reads were first truncated to 230 nt (all reads shorter than 230 nt were discarded) and any read that exceeded a maximum expected error threshold of 1.0 was removed. After truncation and max expected error trimming, 91% of original reads remained. The first 30 nt representing the forward primer and barcode on high quality, truncated reads were trimmed. Remaining reads were taxonomically annotated using the "UClust" taxonomic annotation framework in the QIIME software package [Caporaso et al., 2010, Edgar, 2010] with cluster seeds from Silva SSU rRNA database [Pruesse et al., 2007] 97% sequence identity OTUs as reference (release 111Ref). Reads annotated as "Chloroplast", "Eukaryota", "Archaea", "Unassigned" or "mitochondria" were culled from the dataset. Finally, reads were aligned to the Silva reference alignment provided by the Mothur software package [Schloss et al., 2009] using the Mothur NAST aligner [DeSantis et al., 2006]. All reads that did not appear to align to the expected amplicon region of the SSU rRNA gene were discarded. Quality control parameters removed 34716 of 258763 raw reads.

4.6.2 Sequence clustering

Sequences were distributed into OTUs using the UParse methodology [Edgar, 2013]. Specically, cluster seeds were identified using USearch with a collection of non-redundant reads sorted by count as input. The sequence identity threshold for establishing a new OTU centroid was 97%. After initial cluster centroid selection, select 16S rRNA sequences trimmed to the same 16S position as the other centroids from Yeager et al. [2006] were added to the centroid collection. Specifically, Yeager et al. [2006] Colorado Plateau or Moab, Utah sequences were added which included the 16S sequences for *Calothrix MCC-3A, Nostoc commune MCT-1, Nostoc commune MFG-1, Scytonema hyalinum DC-A, Scytonema hyalinum FGP-7A, Spirirestis rafaelensis LQ-10*. Centroid sequences that matched selected Yeager et al. [2006] sequences with greater than to 97% sequence identity were subsequently removed from the centroid collection. With USearch/UParse, potential chimeras are identified during OTU centroid selection and are not allowed to become cluster centroids effectively removing chimeras from the read pool. All quality controlled reads were then mapped to cluster centroids at an identity threshold of 97% again using USearch. 95.6% of quality controlled reads could be mapped to centroids. Unmapped reads do not count towards sample counts and are essentially removed from downstream analyses. The USearch software version for cluster generation was 7.0.1090.

Accession of representative 16S rRNA seqeunce	Species Name
DQ531701.1	Scytonema hyalinum DC-A
DQ531697.1	Scytonema hyalinum FGP-7A
DQ531696.1	Spirirestis rafaelensis LQ-10
DQ531703.1	Nostoc commune MCT-1
DQ531699.1	Nostoc commune MFG-1
DQ531700.1	Calothrix MCC-3A

Table 1: Chosen 16S sequences for strains in Yeager et al. [2006] included as OTU centroids

4.6.3 Merging data from this study, Garcia-Pichel et al. [2013], and Steven et al. [2013]

As only sequences without corresponding quality scores were publicly available from Garcia-Pichel et al. [2013] and Steven et al. [2013], these data sets were only quality screened by determining if they covered then expected region of the 16S gene (described above). All data (this study, Garcia-Pichel et al. [2013] and Steven et al. [2013]) were included as input to USearch for OTU centroid selection and subsequent mapping to OTU centroids.

4.6.4 Phylogenetic tree

The alignment for the "*Clostridiaceae*" phylogeny was created using SSU-Align which is based on Infernal [Nawrocki and Eddy, 2013, Nawrocki et al., 2009]. Columns in the alignment that were not included in the SSU-Align covariance models or were aligned with poor confidence (less than 95% of characters in a position had posterior probability alignment scores of at least 95%) were masked for phylogenetic reconstruction. Additionally, the aligment was trimmed to coordinates such that all sequences in the alignment began and ended at the same positions. The "*Clostridiaceae*" tree included all top BLAST hits (parameters below) for ^{15}N responders in the Living Tree Project database [Yarza et al., 2008] in addition to BLAST hits within an sequence identity threshold of 97% to ^{15}N responders from the Silva SSURef_NR SSU rRNA database [Pruesse et al., 2007]. Only one SSURef_NR115 hit per study per OTU ("study" was determined by "title" field) was selected for the tree. FastTree [Price et al., 2010] was used to build the tree and split support values are SH-like scores reported by FastTree.

Placement of short sequences into backbone phylogeny Short sequences were mapped to the reference backbone using pplacer [Matsen et al., 2010] (default parameters). pplacer finds the edge placements that maximize phylogenetic likelihood. Prior to being mapped to the reference tree, short sequences were aligned to the reference alignment using Infernal [Nawrocki et al., 2009] against the same SSU-Align covariance model used to align reference sequences.

4.6.5 BLAST searches

BLAST searches were done with the "blastn" program from BLAST+ toolkit [Camacho et al., 2009] version 2.2.29+. Default parameters were always emplyed and the BioPython [Cock et al., 2009] BLAST+ wrapper was used to invoke the blastn program. Pandas [McKinney, 2012] and dplyr [Wickham and Francois, 2014] were used to parse and munge BLAST output tables.

4.6.6 Identifying OTUs that inocorporated ^{15}N into their DNA

SIP is a culture-independent approach towards defining identity-function connections in microbial communities [Buckley, 2011, Neufeld et al.]. Microbes incubated in the presence of ^{13}C or ^{15}N labeled substrates will incorporate the stable heavy isotope into biomass if they participate in it's transformation. Stable isotope labeled nucleic acids can then be separated from unlabeled by buoyant density in a CsCl gradient. As the buoyant density of a macromolecule is dependent on many factors in addition to stable isotope incorporation (e.g. GC-content in nucleic acids [Youngblut and Buckley, 2014]), labeled nucleic acids from one microbial population may have the same buoyant density of unlabeled nucleic acids from another (i.e. each populations nucleic acids would be found at the same point along a density gradient although only one populations nucleic acids are labeled). Therefore it is imperative when employing SIP to compare density gradients with nucleic acids from heavy stable isotope incubations to "control" incubations where everything mimics the experimental conditions except that unlabeld substrates are used. By contrasting "heavy" density gradient fractions in experimental density gradients (hereafter referred to as "labeled" gradients) against heavy fractions in control gradients, the identities of microbes with labeled nucleic acids can be determined

We used an RNA-Seq differential expression statistical framework [Love et al., 2014] to find OTUs enriched in heavy fractions of labelled gradients relative to corresponding density fractions in control gradients (for review of RNA-Seq differential expression statistics applied to microbiome OTU count data see McMurdie and Holmes [2014]). We use the term differential abundance (coined by McMurdie and Holmes [2014]) to denote OTUs that have different proportion means across sample classes (in this case the only sample class is labeled/control). CsCl gradient fractions were categorized as "heavy" or "light". The heavy category denotes fractions with density values above 1.725 g/mL. Since we are only interested in enriched OTUs (labeled versus control), we used a one-sided z-test for differential abundance (i.e. the null hypothesis is the labeled:control proportion mean ratio for an OTU is less than a selected threshold). We selected a null threshold of 0.25 (or a labeled:control proportion mean ratio of 1.19). DESeq2 was used to calculate the moderated $\log_2$ fold change of labeled:control proportion mean ratios and corresponding standard errors. Mean ratio moderation allows for reliable ratio ranking such that high variance and likely statistically insignificant mean ratios are appropriately shrunk and subsequently ranked lower than they would be as raw ratios. To

summarize, OTUs with high labeled:control mean ratios have higher proportion means in heavy fractions of labeled gradients relative to heavy fractions of control gradients, and therefore have likely incorporated atmospheric N into their DNA during the incubation.

4.6.7 Ordination

Principal coordinate ordinations depict the relationship between samples at each time point (day 2 and 4). Weighted Unifrac distances [Lozupone and Knight, 2005] were used as the sample distance metric for ordination. The tree used in the unifrac distance calculations is described above. The Phyloseq [McMurdie and Holmes, 2014] wrapper for Vegan [Oksanen et al., 2013] (both R packages) was used to compute sample values along principal coordinate axes. GGplot2 [Wickham, 2009] was used to display sample points along the first and second principal axes.

4.6.8 Differential abundance in environmental samples

Significance of OTU proportion mean differences with mean annual temperature (for Garcia-Pichel et al. [2013] data) and sample type, BSC or below crust (for Steven et al. [2013] data), was determined using the DESeq2 framework [McMurdie and Holmes, 2014, Love et al., 2014]. A sparsity threshold of 0.40 was set to screen out sparse OTUs.

4.7 Richness analyses

Rarefaction curves were created using bioinformatics modules in the PyCognet Python package [Knight et al., 2007]. Parametric richness estimates were made with CatchAll using Only the best model total OTU estimates [Bunge, 2010].

References

Marti J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, Feb 2001. doi: 10.1111/j.1442-9993.2001.01070.pp.x. URL `http://dx.doi.org/10.1111/j.1442-9993.2001.01070.pp.x`.

J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In *Biological Soil Crusts: Structure Function, and Management*, pages 241–261. Springer Science + Business Media, 2001. doi: 10.1007/978-3-642-56475-8_19. URL `http://dx.doi.org/10.1007/978-3-642-56475-8_19`.

J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*, volume 150 of *Ecological*

Studies, pages 241–261. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-43757-4. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.1007/978-3-642-56475-8_19.

J. Belnap, R. Prasse, and K.T. Harper. Influence of Biological Soil Crusts on Soil Environments and Vascular Plants. In Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*, volume 150 of *Ecological Studies*, pages 281–300. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-43757-4. doi: 10.1007/978-3-642-56475-8_21. URL http://dx.doi.org/10.1007/978-3-642-56475-8_21.

Jayne Belnap. Nitrogen fixation in biological soil crusts from southeast Utah USA. *Biology and Fertility of Soils*, 35(2):128–135, Apr 2002. doi: 10.1007/s00374-002-0452-x. URL http://dx.doi.org/10.1007/s00374-002-0452-x.

Jayne Belnap. Some Like It Hot, Some Not. *Science*, 340(6140):1533–1534, 2013. doi: 10.1126/science.1240318. URL http://www.sciencemag.org/content/340/6140/1533.short.

J. Roger Bray and J. T. Curtis. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4):325, Oct 1957. doi: 10.2307/1942268. URL http://dx.doi.org/10.2307/1942268.

Daniel H. Buckley. Stable Isotope Probing Techniques Using 15N. In *Stable Isotope Probing and Related Technologies*, pages 129–147. American Society of Microbiology, jan 2011. doi: 10.1128/9781555816896.ch7. URL http://dx.doi.org/10.1128/9781555816896.ch7.

John Bunge. Estimating the number of species with catchall. In *Biocomputing 2011*, pages 121–130. WORLD SCIENTIFIC, nov 2010. doi: 10.1142/9789814335058_0014. URL http://dx.doi.org/10.1142/9789814335058_0014.

C Camacho, G Coulouris, V Avagyan, N Ma, J Papadopoulos, K Bealer, and TL Madden. BLAST+: architecture and applications. 10:421, Dec 2009.

JG Caporaso, J Kuczynski, J Stombaugh, K Bittinger, FD Bushman, EK Costello, N Fierer, AG Pea, JK Goodrich, JI Gordon, GA Huttley, ST Kelley, D Knights, JE Koenig, RE Ley, CA Lozupone, D McDonald, BD Muegge, M Pirrung, J Reeder, JR Sevinsky, PJ Turnbaugh, WA Walters, J Widmann, T Yatsunenko, J Zaneveld, and R Knight. QIIME allows analysis of high-throughput community sequencing data. 7:335–6, 2010.

PJ Cock, T Antao, JT Chang, BA Chapman, CJ Cox, A Dalke, I Friedberg, T Hamelryck, F Kauff, B Wilczynski, and Hoon MJ de. Biopython: freely available Python tools for computational molecular biology and bioinformatics. 25:1422–3, 2009.

TZ Jr DeSantis, P Hugenholtz, K Keller, EL Brodie, N Larsen, YM Piceno, R Phan, and GL Andersen. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. 34:W394–9, 2006.

RC Edgar. Search and clustering orders of magnitude faster than BLAST. 26: 2460–1, 2010.

RC Edgar. UPARSE: highly accurate OTU sequences from microbial amplicon reads. 10:996–8, 2013.

R. D. Evans and J. Belnap. Long-Term Consequences of Disturbance on Nitrogen Dynamics in an Arid Ecosystem. *Ecology*, 80(1):150–160, Jan 1999. doi: 10.1890/0012-9658(1999)080[0150:ltcodo]2.0.co;2. URL `http://dx.doi.org/10.1890/0012-9658(1999)080[0150:LTCODO]2.0.CO;2`.

R. D. Evans and O. L. Lange. Biological Soil Crusts and Ecosystem Nitrogen and Carbon Dynamics. In *Biological Soil Crusts: Structure Function, and Management*, pages 263–279. Springer Science + Business Media, 2001. doi: 10.1007/978-3-642-56475-8_20. URL `http://dx.doi.org/10.1007/978-3-642-56475-8_20`.

F. Garcia-Pichel, S. L. Johnson, D. Youngkin, and J. Belnap. Small-Scale Vertical Distribution of Bacterial Biomass and Diversity in Biological Soil Crusts from Arid Lands in the Colorado Plateau. *Microbial Ecology*, 46 (3):312–321, Nov 2003a. doi: 10.1007/s00248-003-1004-0. URL `http://dx.doi.org/10.1007/s00248-003-1004-0`.

F. Garcia-Pichel, V. Loza, Y. Marusenko, P. Mateo, and R. M. Potrafka. Temperature Drives the Continental-Scale Distribution of Key Microbes in Topsoil Communities. *Science*, 340(6140):1574–1577, Jun 2013. doi: 10.1126/science.1236404. URL `http://dx.doi.org/10.1126/science.1236404`.

Ferran Garcia-Pichel, Jayne Belnap, Susanne Neuer, and Ferdinand Schanz. Estimates of global cyanobacterial biomass and its distribution. *Algological Studies*, 109(1):213–227, 2003b.

Marco Griese, Christian Lange, and Jrg Soppa. Ploidy in cyanobacteria. *FEMS Microbiology Letters*, 323(2):124–131, sep 2011. doi: 10.1111/j.1574-6968.2011.02368.x. URL `http://dx.doi.org/10.1111/j.1574-6968.2011.02368.x`.

SL Johnson, CR Budinoff, J Belnap, and F Garcia-Pichel. Relevance of ammonium oxidation within biological soil crust communities. 7:1–12, 2005.

A. Karnieli, R.F. Kokaly, N.E. West, and R.N. Clark. Remote Sensing of Biological Soil Crusts. In Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*, volume 150 of *Ecological Studies*, pages 431–455. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-43757-4. doi: 10.1007/978-3-642-56475-8_31. URL `http://dx.doi.org/10.1007/978-3-642-56475-8_31`.

Rob Knight, Peter Maxwell, Amanda Birmingham, Jason Carnes, J Gregory Caporaso, Brett C Easton, Michael Eaton, Micah Hamady, Helen Lindsay, Zongzhi Liu, Catherine Lozupone, Daniel McDonald, Michael Robeson, Raymond Sammut, Sandra Smit, Matthew J Wakefield, Jeremy Widmann, Shandy Wikman, Stephanie Wilson, Hua Ying, and Gavin A Huttley. {PyCogent}: a toolkit for making sense from sequence. *Genome Biol*, 8(8): R171, 2007. doi: 10.1186/gb-2007-8-8-r171. URL `http://dx.doi.org/10.1186/gb-2007-8-8-r171`.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for {RNA}-Seq data with {DESeq}2. Technical report, feb 2014. URL `http://dx.doi.org/10.1101/002832`.

C Lozupone and R Knight. UniFrac: a new phylogenetic method for comparing microbial communities. 71:8228–35, 2005.

Frederick A Matsen, Robin B Kodner, and E Virginia Armbrust. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1):538, 2010. doi: 10.1186/1471-2105-11-538. URL `http://dx.doi.org/10.1186/1471-2105-11-538`.

Wes McKinney. pandas: Python Data Analysis Library. Online, 2012. URL `http://pandas.pydata.org/`.

PJ McMurdie and S Holmes. Waste not, want not: why rarefying microbiome data is inadmissible. 10:e1003531, 2014.

EP Nawrocki and SR Eddy. Infernal 1.1: 100-fold faster RNA homology searches. 29:2933–5, Nov 2013.

EP Nawrocki, DL Kolbe, and SR Eddy. Infernal 1.0: inference of RNA alignments. 25:1335–7, May 2009.

JD Neufeld, J Vohra, MG Dumont, T Lueders, M Manefield, MW Friedrich, and JC Murrell. DNA stable-isotope probing. 2:860–6.

Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community Ecology Package*, 2013. URL `http://CRAN.R-project.org/package=vegan`. R package version 2.0-10.

MN Price, PS Dehal, and AP Arkin. FastTree 2–approximately maximum-likelihood trees for large alignments. 5:e9490, Mar 2010.

E Pruesse, C Quast, K Knittel, BM Fuchs, W Ludwig, J Peplies, and FO Glckner. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. 35:7188–96, 2007.

PD Schloss, SL Westcott, T Ryabin, JR Hall, M Hartmann, EB Hollister, RA Lesniewski, BB Oakley, DH Parks, CJ Robinson, JW Sahl, B Stres, GG Thallinger, Horn DJ Van, and CF Weber. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. 75:7537–41, 2009.

T.F. Steppe, J.B. Olson, H.W. Paerl, R.W. Litaker, and J. Belnap. Consortial N2 fixation: a strategy for meeting nitrogen requirements of marine and terrestrial cyanobacterial mats. *FEMS Microbiology Ecology*, 21 (3):149–156, Nov 1996. doi: 10.1111/j.1574-6941.1996.tb00342.x. URL `http://dx.doi.org/10.1111/j.1574-6941.1996.tb00342.x`.

Blaire Steven, La Verne Gallegos-Graves, Jayne Belnap, and Cheryl R. Kuske. Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. *FEMS Microbiol Ecol*, 86(1):101–113, May 2013. doi: 10.1111/1574-6941.12143. URL `http://dx.doi.org/10.1111/1574-6941.12143`.

WA Walters, JG Caporaso, CL Lauber, D Berg-Lyons, N Fierer, and R Knight. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. 27:1159–61, Apr 2011.

Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL `http://had.co.nz/ggplot2/book`.

Hadley Wickham and Romain Francois. *dplyr: dplyr: a grammar of data manipulation*, 2014. URL `http://CRAN.R-project.org/package=dplyr`. R package version 0.2.

Pablo Yarza, Michael Richter, Jörg Peplies, Jean Euzeby, Rudolf Amann, Karl-Heinz Schleifer, Wolfgang Ludwig, Frank Oliver Glöckner, and Ramon Rosselló-Móra. The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Systematic and Applied Microbiology*, 31(4):241–250, Sep 2008. doi: 10.1016/j.syapm.2008.07.001. URL `http://dx.doi.org/10.1016/j.syapm.2008.07.001`.

Chris M. Yeager, Jennifer L. Kornosky, Rachael E. Morgan, Elizabeth C. Cain, Ferran Garcia-Pichel, David C. Housman, Jayne Belnap, and Cheryl R. Kuske. Three distinct clades of cultured heterocystous cyanobacteria constitute the dominant N2-fixing members of biological soil crusts of the Colorado Plateau USA. *FEMS Microbiology Ecology*, 60(1):85–97, 2006. doi: 10.1111/j.1574-6941.2006.00265.x. URL `http://dx.doi.org/10.1111/j.1574-6941.2006.00265.x`.

Chris M. Yeager, Cheryl R. Kuske, Travis D. Carney, Shannon L. Johnson, Lawrence O. Ticknor, and Jayne Belnap. Response of Biological Soil Crust Diazotrophs to Season Altered Summer Precipitation, and Year-Round Increased Temperature in an Arid Grassland of the Colorado Plateau, USA.

Front. Microbio., 3, 2012. doi: 10.3389/fmicb.2012.00358. URL `http://dx.doi.org/10.3389/fmicb.2012.00358`.

CM Yeager, JL Kornosky, DC Housman, EE Grote, J Belnap, and CR Kuske. Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado Plateau and Chihuahuan Desert. 70: 973–83, 2004.

ND Youngblut and DH Buckley. Intra-genomic variation in G+C content and its implications for DNA stable isotope probing (DNA-SIP). Aug 2014.

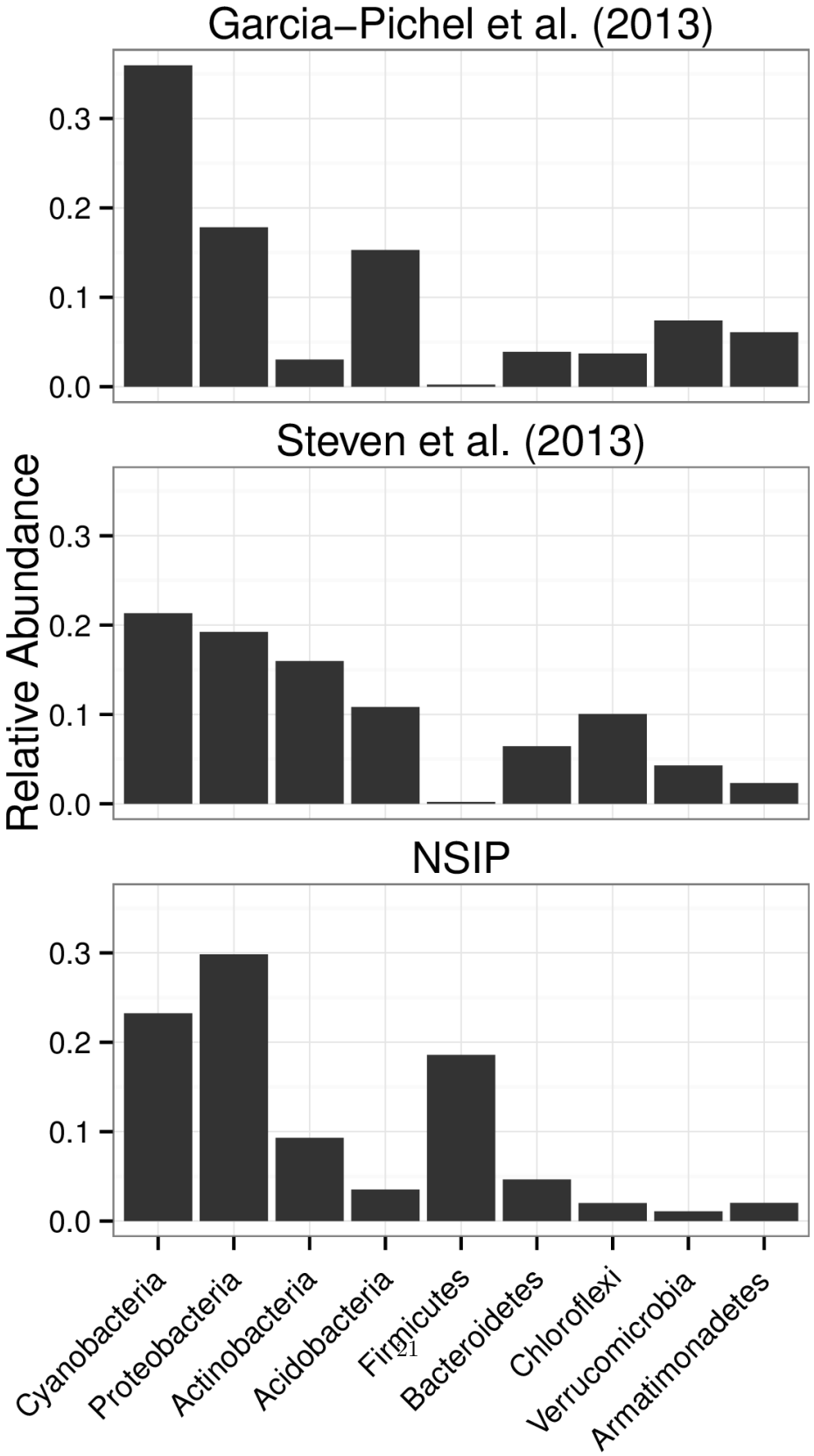

Figure 1: Distribution of sequences into top 9 phyla (phyla ranked by sum of all sequence annotations).

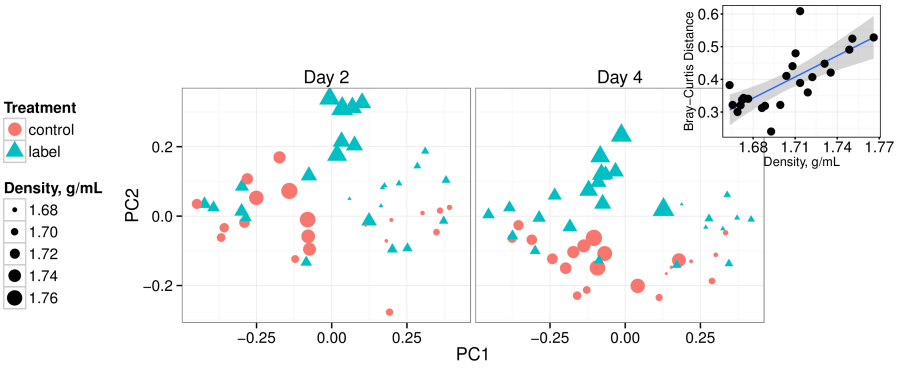

Figure 2: Ordination of Bray-Curtis sample pairwise distances for each incubation time. Point area is proportional to the density of the CsCl gradient fraction for each sequence library, and color reflects control (red) or labeled (blue) treatment. Inset shows Bray-Curtis distances for paired control versus labeled CsCl gradient fractions (i.e. fractions from the same incubation day and same density) against the density of the pair (p-value: 0.000517, r^2: 0.332).

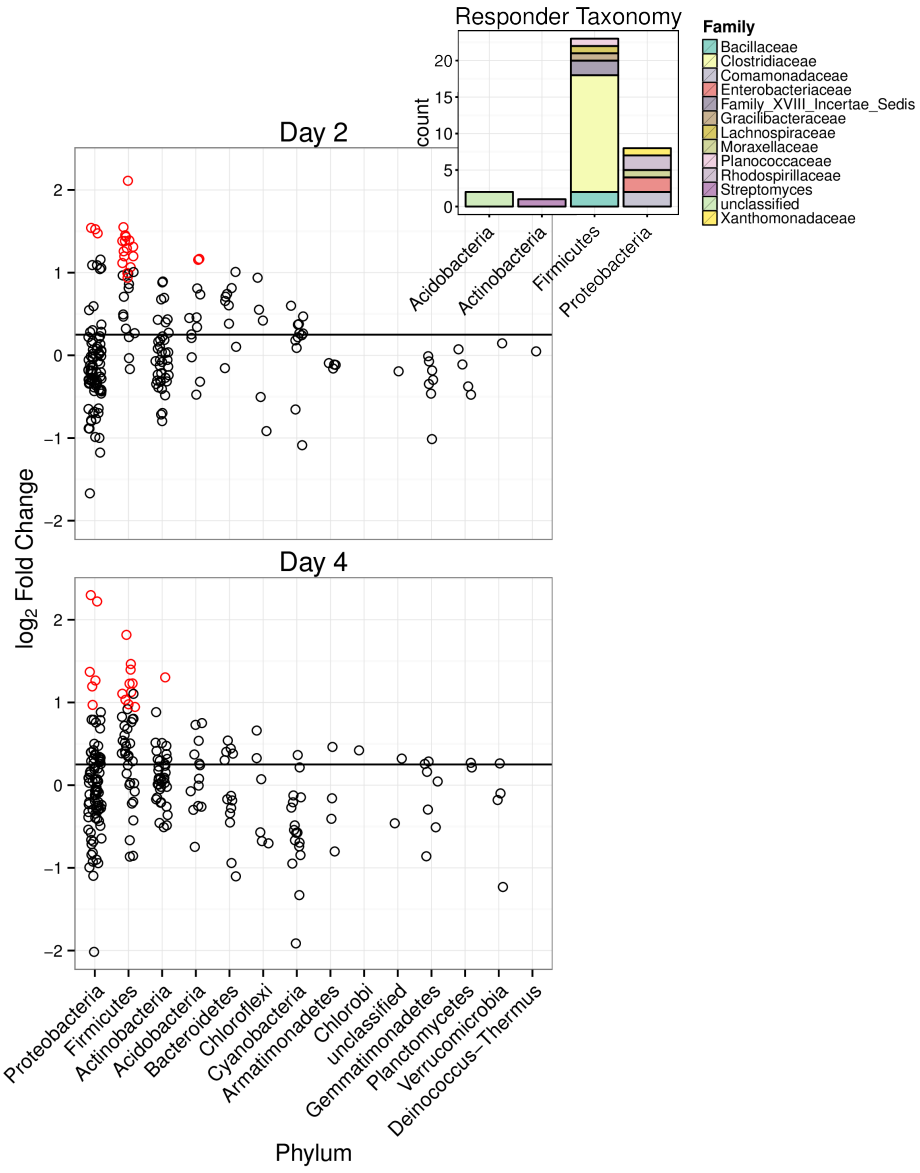

Figure 3: Moderated $\log_2$ of proportion mean ratios for labeled versus control gradients (heavy fractions only, densities ¿1.725 g/mL). All OTUs found in at least 62.5% of heavy fractions at a specific incubation day are shown. Red color denotes a proportion mean ratio that has a corresponding adjusted p-value below a false discovery rate of 10% (the null model is that the proportion mean is ratio is below 0.25). The horizontal line is the proportion mean threshold for the null model, 0.25. The inset figure summarizes the taxonomy of OTUs that with proportion mean ratio p-vaules under 0.10 for at least one time point.

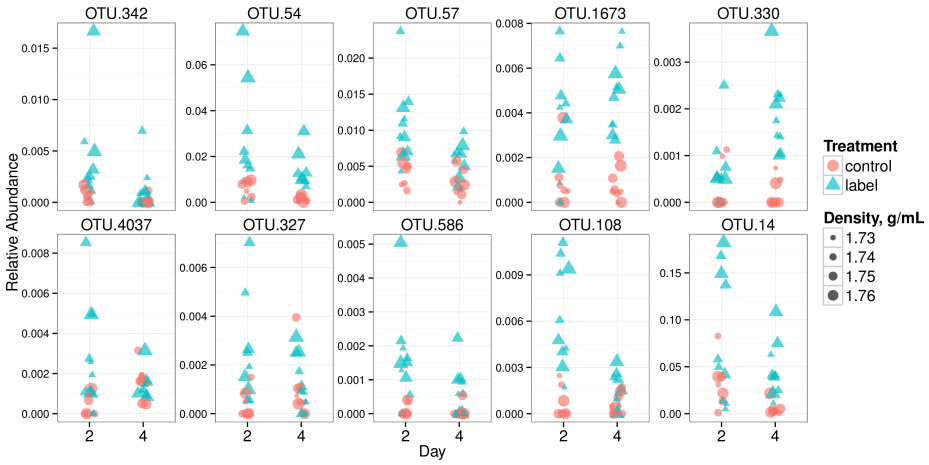

Figure 4: Relative abundance values in heavy fractions (density greater or equal to 1.725 g/mL) for the top 10 ^{15}N "responders" (putative diazotrophs, see results for selection criteria of top 10) at each incubation day. See Table X for BLAST results of top 10 responders against the LTP database (release 115). Point area is proportional to CsCl gradient fraction density, and color signifies control (red) or labeled (blue) treatment.

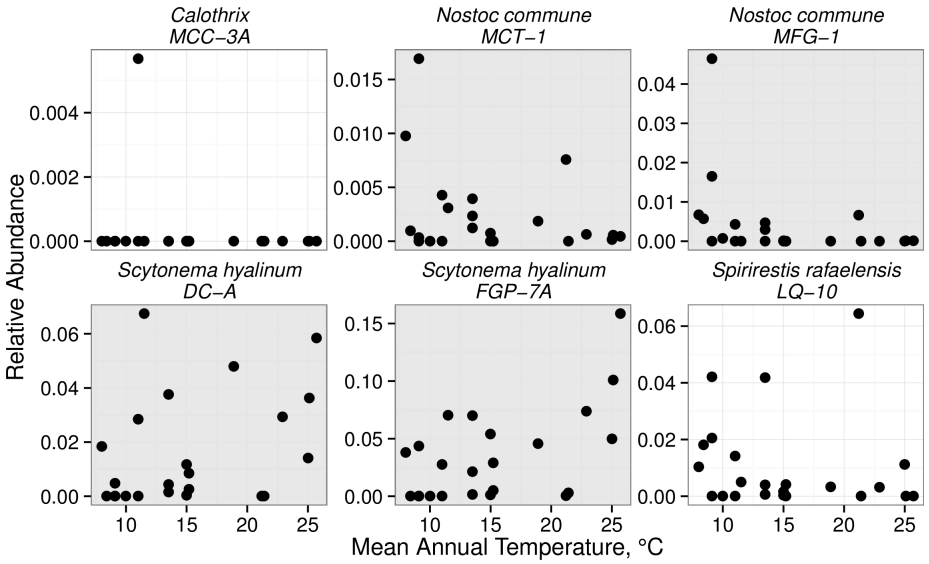

Figure 5: Relative abundance of selected heterocystous cyanobacterial OTUs with centroids from sequences described in Yeager et al. [2006] (see methods for selection criteria) in Steven et al. [2013] data set.

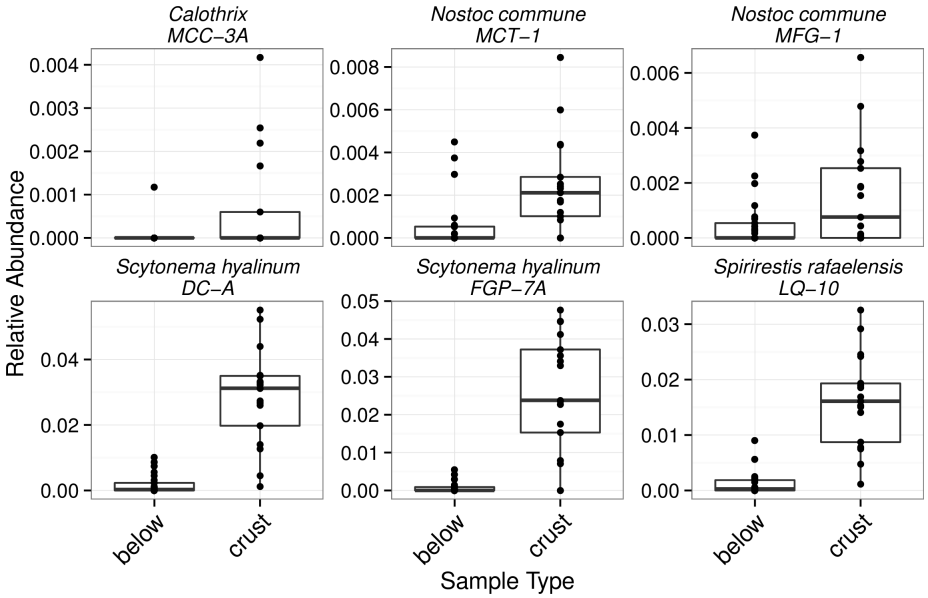

Figure 6: Relative abundance of selected heterocystous cyanobacterial OTUs with centroids from sequences described in Yeager et al. [2006] (see methods for selection criteria) in Steven et al. [2013] data set.

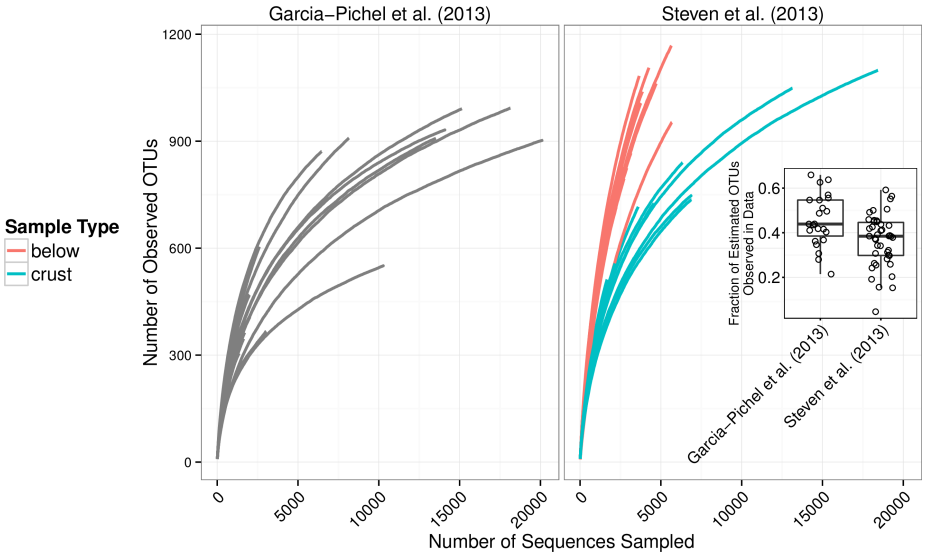

Figure 7: Rarefaction curves for all samples presented by Garcia-Pichel et al. [2013] and Steven et al. [2013]. Inset is boxplot of estimated sampling effort for all samples in Garcia-Pichel et al. [2013] and Steven et al. [2013] (number of observed OTUs divided by number of CatchAll Bunge [2010] estimated total OTUs)

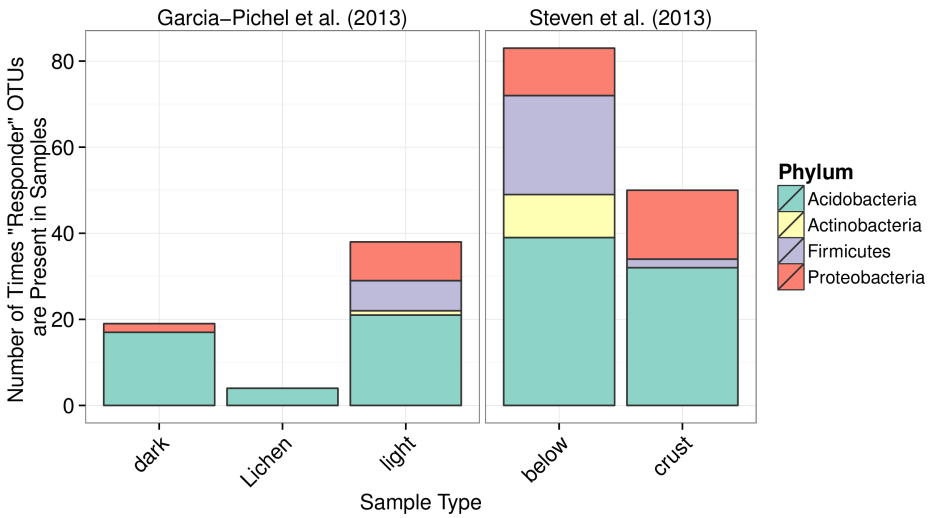

Figure 8: Counts of "responder" OTU occurrences in samples from Steven et al. [2013] and Garcia-Pichel et al. [2013]. Steven et al. [2013] collected BSC samples (25 samples total) and samples from soil beneath BSC (17 samples total, "below" column in figure). Garcia-Pichel et al. [2013] collected samples from "dark" (9 samples total) and "light" (12 samples total) crusts in addition to "lichen" (2 samples total) dominated crusts.

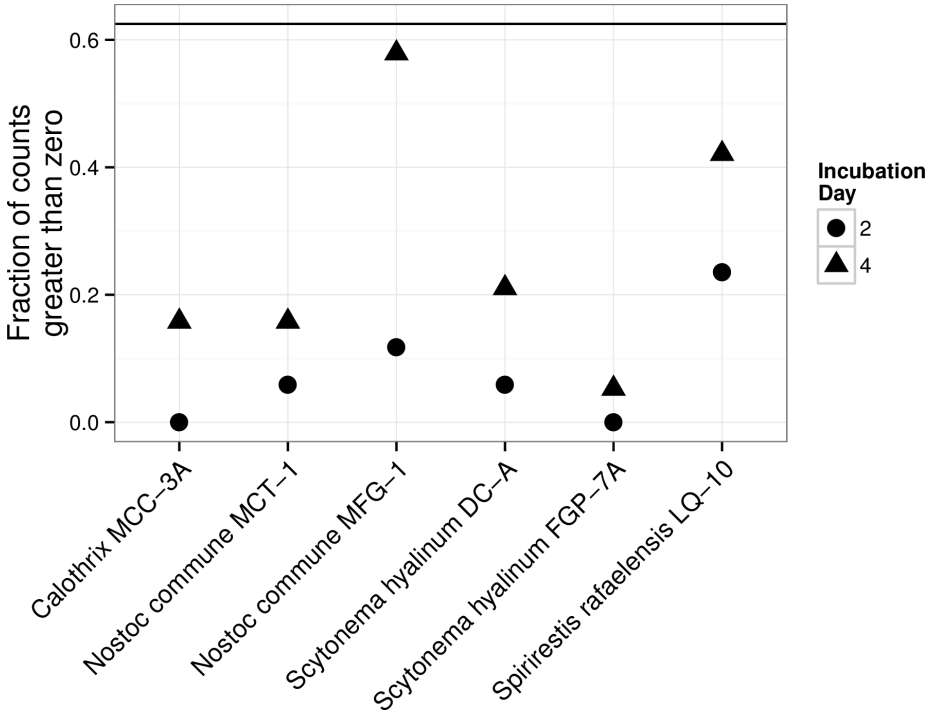

Figure 9: Relative abundance of selected heterocystous cyanobacterial OTUs with centroids from sequences described in Yeager et al. [2006] (see methods for selection criteria) in Steven et al. [2013] data set.

Table 1: ^{15}N responders BLAST against Living Tree Project

OTU ID	***Species Name***	**BLAST percent identity**	**accession**
OTU.108	*Caloramator proteoclasticus*	96.94	X90488
OTU.14	*Pantoea rwandensis*	99.49	JF295055
	Pantoea rodasii	99.49	JF295053
	Kluyvera intermedia	99.49	AF310217
	Kluyvera cryocrescens	99.49	AF310218
	Klebsiella variicola	99.49	AJ783916
	Klebsiella pneumoniae subsp. rhinoscleromatis	99.49	Y17657
	Klebsiella pneumoniae subsp. pneumoniae	99.49	X87276
	Erwinia aphidicola	99.49	FN547376
	Enterobacter soli	99.49	GU814270
	Enterobacter ludwigii	99.49	AJ853891
	Enterobacter kobei	99.49	AJ508301
	Enterobacter hormaechei	99.49	AJ508302
	Enterobacter cloacae subsp. dissolvens	99.49	Z96079
	Enterobacter cancerogenus	99.49	Z96078
	Enterobacter asburiae	99.49	AB004744
	Enterobacter amnigenus	99.49	AB004749
	Enterobacter aerogenes	99.49	AB004750
	Buttiauxella warmboldiae	99.49	AJ233406
	Buttiauxella noackiae	99.49	AJ233405
	Buttiauxella izardii	99.49	AJ233404
	Buttiauxella agrestis	99.49	AJ233400
OTU.1673	*Clostridium drakei*	95.9	Y18813
	Clostridium carboxidivorans	95.9	FR733710
OTU.327	*Clostridium hydrogeniformans*	94.92	DQ196623
	Clostridium amylolyticum	94.92	EU037903
OTU.330	*Clostridium lundense*	96.94	AY858804
OTU.342	*Acinetobacter johnsonii*	100.0	Z93440
OTU.4037	*Fonticella tunisiensis*	93.85	HE604099
OTU.54	*Shigella sonnei*	100.0	FR870445
	Shigella flexneri	100.0	X96963
	Escherichia fergusonii	100.0	AF530475
	Escherichia coli	100.0	X80725
OTU.57	*Fonticella tunisiensis*	93.88	HE604099
	Caloramator proteoclasticus	93.88	X90488
OTU.586	*Vitreoscilla filiformis*	98.48	HM037993
	Ottowia pentelensis	98.48	EU518930
	Ideonella dechloratans	98.48	X72724
	Diaphorobacter nitroreducens	98.48	AB064317
	Comamonas terrigena	98.48	AF078772

Figure 10: Replace this text with your caption

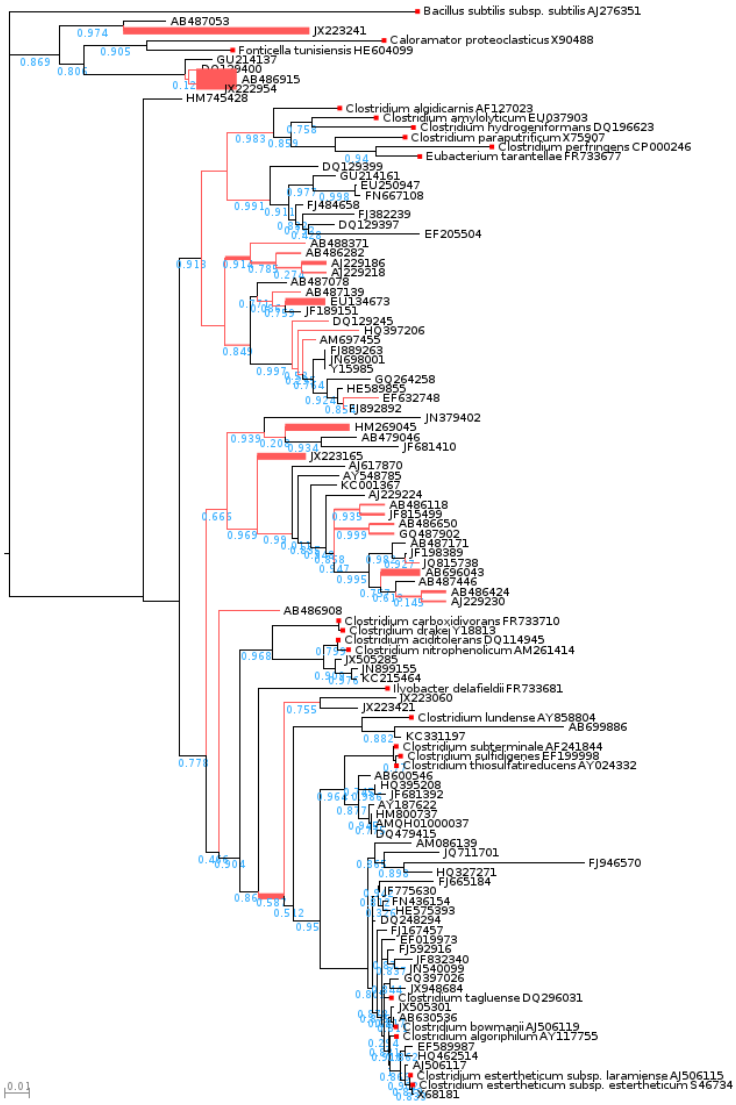

Figure 11: See methods for selection criteria for sequences in backbone tree. Edge width is proportional to number of short putative *Clostridiaceae* diazotroph sequences placed at that position. Placement of short sequences can be spread across multiple edges Matsen et al. [2010]. Reference sequences from cultivars have boxes at tips and full species names. Tips with only accession annotations are from environmental reference sequences.