1

Non-cyanobacterial diazotrophs dominate dinitrogen fixation in biological soil crusts at the early stage of crust formation.

Charles Pepe-Ranney [1], **Chantal Koechli** [1], **Ruth Potrafka** [2], **Ferran Garcia-Pichel** [2], **Daniel H Buckley** [1,*]

[1] *Cornell University, Department of Crop and Soil Sciences, Ithaca, NY, USA*
[2] *Arizona State University, School of Life Sciences, Tempe, AZ 85287, USA.*

Correspondence*:
Daniel H Buckley
Cornell University, Department of Crop and Soil Sciences, Ithaca, NY, USA,

1 ABSTRACT

2 Biological soil crusts (BSC) cover a vast global area and are key components of ecosystem productivity
3 in arid soils. In particular, BSC contribute significantly to the nitrogen (N) budget in arid ecosystems
4 via N-fixation. Although BSC N-fixation is largely attributed to heterocystous cyanobacteria (Yeager
5 et al., 2006, 2004, 2012), DNA stable isotope probing with $^{15}N_2$ revealed primarily *Clostridiaceae*
6 and *Proteobacteria* incorporated ^{15}N in mesocosm incubations with light, poorly developed BSC
7 samples. Non-heterocystous BSC diazotrophs are low abundance members of BSC. The maximum
8 relative abundance of putative *Clostridiaceae* and *Proteobacteria* diazotrophs in any SSU rRNA libraries
9 presented by Garcia-Pichel et al. (2013) or Steven et al. (2013) was 0.00225% and 0.00127%, respectively.
10 Heterocystous cyanobacteria relative abundance is correlated with mean annual temperature for *Nostoc*
11 *commune* MCT-1 and MFG-1, and *Scytonema hyalinum* FGP-7A and DC-A (p-values 1.307×10^{-02},
12 1.577×10^{-06} and 3.332×10^{-03}, 3.173×10^{-04}, respectively). However, the direction of the correlation is
13 different for *Nostoc* (decreasing with temperature) and *Scytonema* (increasing with temperature) types.
14 Non-cyanobacterial diazotrophs have not been sampled sufficiently yet in existing BSC SSU rRNA
15 sequence collections to diagnose their temperature relationships or geographic scope. Identifying the full
16 BSC diazotroph diversity is an crucial step towards predicting how climate change and disturbance will
17 and do affect BSC N-fixation.

2 INTRODUCTION

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

27 positively correlated with vascular plant survival due in part to BSC ecosystem N contributions (for review
28 of BSC-vascular plant interactions see Belnap et al. (2003)).

29 Molecular studies of BSC microbial diversity include explorations of the BSC microbial community
30 vertical profile (Garcia-Pichel et al., 2003a), BSC *nifH* gene content surveys (e.g. Yeager et al. (2004),
31 Yeager et al. (2012), Yeager et al. (2006) and Steppe et al. (1996)), and next-generation-sequencing (NGS)
32 enabled studies of BSC SSU rRNA gene content across wide geographic ranges (Garcia-Pichel et al.,
33 2013; Steven et al., 2013). Garcia-Pichel et al. (2003a) found that BSC microbial diversity is organized
34 vertically, likely as the result of vertically oriented environmental gradients (e.g. light and oxygen). *nifH*
35 surveys have been conducted across BSC development stages (Yeager et al., 2004), as well as across
36 seasons, temperatures and precipitation gradients (Yeager et al., 2012). Mature, more fully developed BSC
37 possess greater numbers of heterocystous *Cyanobacteria* (e.g. *Nostoc, Syctonema*) than developing BSC
38 but both young and old BSC are dominated by non-heterocystous *Cyanobacteria* (*Microcoleus vaginatus*
39 or *M. steenstrupii*) (Yeager et al., 2004; Garcia-Pichel et al., 2013). Young or recently disturbed BSC
40 are often described as "light" in appearance relative to "dark" mature BSC (Belnap, 2002; Yeager et al.,
41 2004). Although an early survey of Colorado Plateau BSC *nifH* diversity recovered *nifH* genes related
42 to *Gammaproteobacteria* as well as a clade that included *nifH* genes from the anaerobes *Clostridium*
43 *pssteurianum*, *Desulfovibrio gigas* and *Chromatium buderi*, subsequent studies have found heterocystous
44 *Cyanobacteria* to be the numerically dominant BSC diazotrophs (Yeager et al., 2006, 2004, 2012).
45 Specifically, Yeager et al. (2006)–in a study of overall BSC *nifH* diversity–categorized 89% of 693 *nifH*
46 sequences derived from Colorado Plateau and New Mexico BSC samples as heterocystous cyanobacterial
47 (non-cyanobacterial *nifH* sequences were largely attributed to alpha- and beta- *proteobacteria*). The
48 heterocystous cyanobacterial BSC diazotrophs fall into three genera, *Scytonema*, *Spirirestis*, and *Nostoc*
49 (Yeager et al., 2006, 2012). Studies of BSC microbial diversity over broad geographic ranges have
50 elucidated how soil parent material correlates to above and below crust microbial community membership
51 and structure (Steven et al., 2013) and that the predominant BSC cyanobacterium shifts from *M. vaginatus*
52 to *M. steenstrupii* with increasing mean annual temperature (Garcia-Pichel et al., 2013).

53 BSC N-fixation rate studies (typically employing the acetylene reduction assay (ARA)) have explored
54 BSC diazotroph activity across various ecological gradients. Reported BSC N-fixation rates vary
55 significantly (Evans and Lange, 2001). The reasons for this variability are complex and likely include
56 the spatial heterogeneity of BSC (Evans and Lange, 2001) and the impact of recent environmental
57 conditions on N-fixation rates (see Belnap (2001) for discussion). Moreover, the ARA assay is subject
58 to methodological artifacts that preclude cross-study and possibly intra-study but inter-environment type
59 comparisons (see Belnap (2001) for review). Despite the general BSC N-fixation rate measurement
60 variability, mature, dark BSC N-fixation rates have been measured higher than N-fixation rates for
61 younger, light BSC (Belnap, 2002; Yeager et al., 2004). This difference may be due to the proliferation
62 of heterocystous *Cyanobacteria* in older mats and is consistent with the theory that heterocystous
63 *Cyanobacteria* are the primary BSC diazotrophs. Alternatively, the N-fixation rate differences between
64 young and old BSC might be attributable to methodological artifacts. For instance, Johnson et al. (2005)
65 show that N-fixation rates peak at a lower depth in developing BSC as compared to mature BSC. When
66 N-fixation is measured from intact cores of developing BSC the measurement may be artifactually
67 low due to delayed acetylene/ethylene diffusion through the crust to and from the peak N-fixation rate
68 depth in a typical ARA incubation timeframe. Diffusion would not be an issue when measuring N-
69 fixation rates in mature crust as nitrogenase activity peaks near the surface. When total N-fixation rates
70 were calculated by integrating N-fixation rates over 1-3 mm depth slices along the full BSC core (thus
71 mitigating ethene/acetylene flux limitations), N-fixation rate differences between developing and mature
72 BSC were not statistically significant (Johnson et al., 2005).

73 The influence of microbial community membership and structure on BSC N-fixation is an ongoing
74 research question (Belnap, 2013). While the presence/abundance of heterocystous *Cyanobacteria* has
75 been proposed as the underlying microbial membership influence on increased N-fixation in mature BSC,
76 it is unclear if the premise that mature BSC fix more N is always correct (see Johnson et al. (2005)).
77 More studies are necessary to elucidate the microbial membership influence on BSC N-fixation and to

Table 1. Summary of contextual environmental data.

study	Description of samples	Number of samples
Garcia-Pichel et al. (2013)	Samples from "light", "dark" and "lichen" BSC across a wide geographic range in the southwestern United States and a site mean annual temperature graient.	23 samples total (9 dark, 12 light, 2 lichen)
Steven et al. (2013)	Samples from three different soil types, "sand", "shale" and "gypsum". Both BSC samples and samples taken from sub-BSC soil	42 samples total (11 gypsum, 15 sand, 16 shale; 25 sub-BSC and 17 BSC)

78 determine if heterocystous *Cyanobacteria* are the only keystone diazotrophs. To further probe the diversity
79 of diazotrophs in BSC we conducted $^{15}N_2$ DNA stable isotope probing (DNA-SIP) experiments with
80 light, developing Colorado Plateau BSC. Although molecular characterizations of BSC *nifH* diversity
81 in other studies have yielded predominantly heterocystous cyanobacterial *nifH* genes, in this study
82 microbes from young, developing BSC that incorporated ^{15}N from $^{15}N_2$ into DNA as determined
83 by DNA-SIP were not *Cyanobacteria* but members of the *Gammaproteobacteria*, *Clostridiaceae* and
84 *Deltaproteobacteria*. Further, we track the distribution of putative diazotrophs uncovered in this study in
85 addition to heterocystous *Cyanobacteria* studied by Yeager et al. (2004), Yeager et al. (2006) and Yeager
86 et al. (2012) through collections of NGS SSU rRNA libraries from BSC microbial diversity surveys over a
87 range of spatial scales and soil types (Garcia-Pichel et al., 2013; Steven et al., 2013). Climate change and
88 disturbance alter BSC microbial community structure and membership. Understanding how these changes
89 affect diazotrophs requires BSC diazotroph diversity be identified in full.

3 RESULTS

3.1 SUMMARY OF ENVIRONMENTAL STUDIES

90 We included data from Garcia-Pichel et al. (2013) and Steven et al. (2013) in the results to provide
91 environmental context for the DNA-SIP findings. Table 1 summarizes the relevant background
92 information for the environmental data sets.

3.2 COMPARISON OF SEQUENCE COLLECTIONS AT "STUDY"-LEVEL

93 *3.2.1 Comparisons of OTU content:* Of the 4,340 OTU centroids established for this study (including
94 sequences from Steven et al. (2013) and Garcia-Pichel et al. (2013)) 445 and 870 have matches in the
95 Living Tree Project (LTP) (a collection of 16S gene sequences for all sequenced type strains (Yarza et al.,
96 2008)) at greater or equal than 97% and 95% sequence identity, respectively (LTP version 115). Similar
97 numbers of total OTUs were found in each data set explored in this study (i.e. the DNA-SIP data presented
98 here, the data presented by Steven et al. (2013) and by Garcia-Pichel et al. (2013)). Specifically, there were
99 3,079 OTUs (209,354 total sequences after quality control) in the DNA-SIP data, 3,203 OTUs (129,033
100 total sequences after quality control) in the Garcia-Pichel et al. (2013) study, and 2,481 OTUs (129,358
101 total sequences after quality control) in the Steven et al. (2013) study. The DNA-SIP data set shares more
102 OTUs with the Steven et al. (2013) data (56% of total OTUs from combined dataset) than it does with the
103 Garcia-Pichel et al. (2013) data (46% of total OTUs from combined dataset). The Steven et al. (2013) and
104 Garcia-Pichel et al. (2013) data only share 46% of OTUs.

105 *3.2.2 Comparisons of Taxonomic Content:* *Cyanobacteria* and *Proteobacteria* were the top two
106 phylum-level sequence annotations for all three studies but only the DNA-SIP data had more
107 *Proteobacteria* annotations than *Cyanobacteria*. *Proteobacteria* represented the 29.8% of sequence
108 annotations in DNA-SIP data as opposed to 17.8% and 19.2% for the Garcia-Pichel et al. (2013)
109 and Steven et al. (2013) data, respectively. Figure 1 shows the distribution of phylum-level sequence
110 annotations for each study in the nine most abundant phyla across all studies as determined by raw
111 sequence counts. There is a stark contrast in the total percentage of sequences annotated as *Firmicutes*
112 between the raw environmental samples and the DNA-SIP data. *Firmicutes* represent only 0.21% and
113 0.23% of total phylum level sequence annotations in the Steven et al. (2013) and Garcia-Pichel et al.
114 (2013) studies, respectively. In the DNA-SIP sequence collection *Firmicutes* make up 19% of phylum
115 level sequence annotations. Also in sharp contrast for the DNA-SIP versus environmental data is the
116 number of putative heterocystous *Cyanobacteria* sequences. Only 0.29% of *Cyanobacteria* sequences in
117 the DNA-SIP data are annotated as belonging to "Subsection IV" which is the heterocystous order of
118 *Cyanobacteria* in the Silva taxonomic nomenclature (Pruesse et al., 2007). In the Steven et al. (2013) and
119 Garcia-Pichel et al. (2013) studies 15% and 23%, respectively, of *Cyanobacteria* sequences are annotated
120 as belonging to "Subsection IV".

3.3 ORDINATION OF CSCL GRADIENT FRACTION SSU RRNA LIBRARIES

121 Ordination of Bray-Curtis (Bray and Curtis, 1957) distances between CsCl gradient fraction sequence
122 libraries with principal coordinates analysis shows the labeled gradient fraction libraries diverge from
123 control at the "heavy" of the CsCl gradients (Figure 2). When the labeled and control CsCl gradient
124 fraction 16S rRNA gene libraries are paired such that each pair contains a control fraction and labeled
125 fraction from the same incubation day with a density difference below 0.003 g/mL, the Bray-Curtis
126 distance between the fraction pair is positively correlated to the density of the labeled fraction (p-value:
127 0.00052, r^2: 0.3315) (inset Figure 2). Additionally, the label/control category for heavy fractions is
128 statistically significant by the Adonis test (p-value: 0.001, r^2: 0.136) (Anderson, 2001). The first principal
129 axis appears to be correlated with fraction density (Figure 2) and the Adonis test p-value for density versus
130 pairwise Bray-Curtis distances with all CsCl fraction libraries is 0.001 (r^2 0.117).

3.4 IDENTITIES OF POSSIBLE ^{15}N INCORPORATORS

131 The OTUs that have enriched proportion means in labeled gradient heavy fractions versus control gradient
132 heavy fractions are those that have incorporated the stable isotope tracer into their DNA. We found 38
133 OTUs that appeared to incorporate ^{15}N into DNA (or "responders"). Of these 38, 26 are annotated as
134 *Firmicutes*, 9 as *Proteobacteria*, 2 as *Acidobacteria* and 1 as *Actinobacteria* (The inset of Figure 3
135 summarizes the Family level taxonomic profile of stable isotope responders). Figure 3 summarizes the
136 ratio of proportion means for each OTU where means are calculated from proportions in heavy fractions
137 within labeled or controlled gradients and the ratio is labeled over control (see methods). If the OTUs
138 are ranked by descending, moderated proportion mean labeled:control ratios, the top 10 ratios (i.e. the
139 10 OTUs that were most enriched in the labeled gradients considering only heavy fractions) are either
140 *Firmicutes* (6 OTUs) or *Proteobacteria* (4 OTUs). Figure 4 shows the relative abundance values for the
141 top 10 OTUs in heavy fractions of labeled and control gradients. Table 4 summarizes the results from
142 BLAST searching the centroid sequences for these top 10 OTUs against the LTP database *Proteobacteria*
143 OTU centroid sequences for the top 10 responders all share high identity (>98.48% identity, Table 4)
144 with cultivars from genera known to possess diazotrophs including *Klebsiella*, *Shigella*, *Acinetobacter*,
145 and *Ideonella*. None of the *Firmicutes* OTUs in the top 10 responders share greater than 97% sequence
146 identity with sequences in the LTP database (release 115) (see Table 4).

3.5   DISTRIBUTION OF BSC DIAZOTROPHS IN ENVIRONMENTAL SAMPLES

147   *3.5.1   Non-Cyanobacterial Taxa*

148   ***Clostridiacea:*** Five of the 6 *Firmicutes* in the top 10 responder OTUs (above) belong in the *Clostridiacea*.
149   We only observed one of these strongly responding *Clostridiaceae* in the data presented by Garcia-Pichel
150   et al. (2013), "OTU.108" (closest BLAST hit in LTP Release 115 – *Caloramotor proteoclasticus*, BLAST
151   %ID 96.94, Accession X90488). OTU.108 was found in two samples both characterized as "light" crust.
152   One other *Clostridiaceae* OTU with a proportion mean ratio (labeled:control) p-value less than 0.10 but
153   outside the top 10 responders was found in the Garcia-Pichel et al. (2013) data and also in a "light" crust
154   sample. None of the strongly responding *Clostridiacea* were found in the sequences provided by Steven
155   et al. (2013).

156   Figure 5 depicts the phylogenetic breadth of *Clostridiaceae* [15]N responder OTUs from this experiment.
157   The phylogenetic tree was constructed from near full-length reference sequences, and edge width
158   demonstrates the placements of short OTU centroid sequences in the backbone tree (see methods
159   for description of placement algorithm and selection criteria for reference sequences). As shown,
160   *Clostridiaceae* ^{15}N-responder OTU centroid 16S sequences are generally more closely related to
161   environmental than cultivar 16S gene sequences.

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

168   ***Other taxa:*** Two potentially diazotroph OTUs were found in an extensive number of environmental
169   samples (61 of 65 samples from the combined data sets of Garcia-Pichel et al. (2013) and Steven et al.
170   (2013)). Both OTUs were annotated as *Acidobacteria* but shared little sequence identity to any cultivar
171   SSU rRNA gene sequences in the LTP (Release 115), with best LTP BLAST hits of 81.91 and 81.32%
172   identity. Additionally, the evidence for ^{15}N incorporation for each OTU was weak relative to other putative
173   responders (adjusted p-values of 0.090 and 0.096). Of the remaining 36 stable isotope responder OTUs,
174   only 14 were observed in the environmental data. Figure 6 summarizes the OTU-sample occurrences in
175   both the Steven et al. (2013) and the Garcia-Pichel et al. (2013) data with occurrences distributed into the
176   most relevant sample classes of each study.

177   *3.5.2   Heterocystous Cyanobacteria*   At least one OTU defined by Yeager et al. (2006) sequences (see
178   Table 3) was found in 21 of the 23 Garcia-Pichel et al. (2013) sampling sites. Counts of samples with
179   Yeager et al. (2006) sequence defined heterocystous *Cyanobacteria* OTUs are summarized in Table 2.
180   The opposite BSC relative abundance relationships of *Microcoleus Vaginatus* and *M. Strenstrupii* with site
181   mean annual temperature was a major finding by Garcia-Pichel et al. (2013). Garcia-Pichel et al. (2013)
182   did not report the relationship of diazotrophic *Cyanobacteria* with temperature although a comment by
183   Belnap (2013) briefly discusses a qualitative positive relationship of *Scytonema* with temperature in the
184   Garcia-Pichel et al. (2013) data. In agreement with the Belnap (2013) interpretation, we found a positive
185   relationship of *Scytonema hyalinum* FGP-7A and DC-A OTU relative abundance with mean annual
186   temperature (p-values 3.332×10^{-03} and 3.173×10^{-04}, respectively) (Figure 7). We also found *Nostoc*
187   *commune* MCT-1 and MFG-1 OTU relative abundance was inversely related to mean annual temperature
188   (p-values 1.307×10^{-02} and 1.577×10^{-06}, respectively) (Figure 7).

189   At least one Yeager et al. (2006) sequence defined OTU (Table 3) was found in 35 of 42 Steven et al.
190   (2013) samples. The 7 samples that lacked Yeager et al. (2006) OTUs were "below crust" samples. Table 2
191   summarizes the counts of Steven et al. (2013) samples with Yeager et al. (2006) sequence defined OTUs.

Table 2. Counts of heterocystous *Cyanobacteria*l OTU occurrences in Garcia-Pichel et al. (2013) samples (n = 23) and Steven et al. (2013) samples (n = 42)

Isolate	Garcia-Pichel et al. (2013)	Steven et al. (2013)
Calothrix MCC-3A	1	6
Nostoc commune MCT-1	16	23
Nostoc commune MFG-1	12	23
Scytonema hyalinum DC-A	17	30
Scytonema hyalinum FGP-7A	18	27
Spirirestis rafaelensis LQ-10	16	30

192 As expected all of the six OTUs defined by Yeager et al. (2006) sequences were more abundant in the
193 crust samples than below crust samples (Figure 8) (maximum p-value for any OTU: 1.96×10^{-4}).

3.6 RICHNESS ESTIMATES

194 Figure 9 (inset) summarizes the fraction of observed OTUs over total OTUs as estimated by CatchAll
195 for each sample 16S library. Rarefaction curves for each sample are shown in Figure 9. Qualitatively,
196 rarefaction curves show below crust samples to be more rich than BSC samples in the Steven et al. (2013)
197 data.

4 DISCUSSION

4.1 STUDY-LEVEL DIFFERENCES

198 One striking difference between the environmental datasets (Garcia-Pichel et al., 2013; Steven et al., 2013)
199 and the DNA-SIP data is the increased relative abundance of *Firmicutes* sequence annotations in the DNA-
200 SIP data (Figure 1). The DNA-SIP data also has more *Proteobacteria* sequence annotations than either
201 environmental dataset. (Figure 1). The increased *Firmicutes* and *Proteobacteria* annotations are consistent
202 with the phylum-level taxonomies of the most strongly ^{15}N responding OTUs (see results). At the distal
203 ends of a CsCl DNA-SIP gradient there is little DNA, but, since we are working with compositional data
204 and gradient fraction libraries are not weighted by absolute DNA content, OTUs found at the ends of CsCl
205 gradients are inflated in overall abundance relative to their abundance in the non-fractionated DNA. DNA
206 from OTUs that incopororate ^{15}N into their biomass moves towards the heavy end of the CsCl gradient
207 and therefore OTUs in this "labeled" DNA are enriched in the full data pool relative to environmental
208 DNA.

4.2 ORDINATION OF CSCL GRADIENT FRACTION 16S LIBRARIES

209 The ordination of Bray-Curtis distances between CsCl gradient fraction 16S libraries show that control
210 fractions differ from labeled fractions in the "heavy" range of the CsCl gradients (Figure 2). If each
211 control fraction is paired to the labeled fraction from the same incubation day for which it is closest in
212 density, there is a positive and statistically significant correlation between Bray-Curtis distances within
213 fraction pairs and density of the pair (see inset Figure 2). Therefore, the "heavy" end of the control
214 and labeled gradients differ and the OTUs enriched in the labeled fractions (relative to control) would
215 have incorporated ^{15}N into their DNA during the incubation timeframe. If the incubation timeframe is
216 appropriate, the ^{15}N-incorporators would most likely have incorporated the ^{15}N from atmospheric ^{15}N$_2$.

4.3 BSC DIAZOTROPHS IDENTIFIED IN THE STUDY

217 BSC N-fixation has long been attributed to heterocystous *Cyanobacteria* and molecular microbial ecology
218 surveys of BSC *nifH* gene content have been consistent with this hypothesis finding cyanobacterial
219 *nifH* types to be numerically dominant in *nifH* gene libraries (Yeager et al., 2006, 2004, 2012). It is
220 possible, however, that PCR-driven molecular surveys of *nifH* gene content have been biased against
221 non-heterocystous *Cyanobacteria*. In general the nifH PCR primers used by Yeager et al. (2006, 2004,
222 2012) (19F and nifH3) for the first round of nested PCR have broad specificity and display at least 86% *in*
223 *silico* coverage for *Proteobacteria, Cyanobacteria* and "Cluster III" *nifH* reference sequences (Gaby and
224 Buckley, 2012). In the second round of the nested PCR protocol (Yeager et al., 2006, 2004, 2012), primer
225 nifH11 is' slightly biased against "Cluster III" (50% coverage) but biased in favor of *Proteobacteria* (79%
226 *in silico* coverage against 67% for *Cyanobacteria*) and nifH22 matches *Proteobacteria, Cyanobacteria*
227 and "Cluster III" reference sequences poorly (16%, 23% and 21% *in silico* coverage, respectively) (Gaby
228 and Buckley, 2012). Unfortunately, it is impossible to assess or quantify this bias (in either direction)
229 without knowing the *nifH* gene content *de novo*. Additionally, heterocysts (the specialized N-fixing cells
230 along the trichome of filamentous heterocystous *Cyanobacteria* such as *Nostoc* and *Scytonema*) may be
231 overrepresented with respect to non-cyanobacterial diazotrophs in *nifH* libraries because the heterocysts
232 make up a fraction of the total cells along a trichome and even the non-heterocyst cells in a trichome will
233 possess the *nifH* gene. It should also be noted that *nifH* gene content is not directly extrapolable to the
234 taxonomic relative abundances of nitrogenase proteins.

235 We did not observe evidence for N-fixation by heterocystous *Cyanobacteria* in the "light" crust samples
236 used in this study. One possible explanation for our results is that the "light", still developing BSC
237 samples used in this study possessed less heterocystous *Cyanobacteria* than dark mature BSC, as has
238 been observed in previous comparisons of light and dark BSC (Yeager et al., 2004). Indeed, only 0.29%
239 of sequences from this study's DNA-SIP 16S rRNA gene sequence libraries were from heterocystous
240 *Cyanobacteria* (see results) as opposed to 15% and 23% of total sequences in the Steven et al. (2013)
241 and Garcia-Pichel et al. (2013) data, respectively. It is difficult to compare relative abundance values from
242 CsCl gradient fractions against environmental libraries, but, a three order of magnitude difference between
243 the environmental libraries and the CsCl gradient fractions is stark. Nonetheless, we would still expect
244 even low abundance diazotrophs to show evidence for ^{15}N-incorporation, provided sequence counts were
245 not too sparse in heavy fractions. The OTUs defined by selected heterocystous *Cyanobacteria* sequences
246 presented in Yeager et al. (2006), however, all fall below the sparsity threshold used in our analysis
247 (see methods, Figure 10). Given the sparsity of heterocystous *Cyanobacteria* sequences in the DNA-SIP
248 data set, it is not possible to assess whether heterocystous *Cyanobacteria* incorporated ^{15}N during the
249 incubation.

250 The OTUs that did appear to incorporate ^{15}N during the incubation were predominantly *Proteobacteria*
251 and *Firmicutes*. The *Proteobacteria* OTUs for which ^{15}N-incorporation signal was strongest all shared
252 high sequence identity ($>=98.48\%$ sequence identity) with 16S sequences from cultivars in genera with
253 known diazotrophs (Table 4). The *Firmicutes* that displayed signal for ^{15}N-incorporation (predominantly
254 *Clostridiaceae*) were not closely related to any cultivars (Table 4, Figure 5). These BSC *Clostrodiaceae*
255 diazotrophs represent a gap in culture collections. As culture-based ecophysiological studies have proven
256 useful towards explaining ecological phenomena in BSC 16S rRNA gene sequence libraries (Garcia-
257 Pichel et al., 2013), it would seem that these putative *Clostridiaceae* diazotrophs would be prime
258 candidates for targeted culturing efforts. Assessing the physiological response of these diazotrophic
259 *Clostridiaceae* to temperature would be useful for predicting how climate change will affect the BSC
260 nitrogen budget.

261 Although too undersampled in the environmental data sets to reach statistical conclusions, non-
262 cyanobacterial diazotrophs were found more often in below crust samples (as opposed to BSC samples) in
263 the Steven et al. (2013) data and in "light" BSC samples in the Garcia-Pichel et al. (2013) data (Figure 6).
264 This result generates some hypotheses that are counter to prior discussions regarding BSC diazotroph
265 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 diazotrophs 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 might be 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 8) so the samples used in this study are not necessarily unrepresentative of typical poorly developed BSC simply because they are lacking heterocystous *Cyanobacteria*. Additionally, cyanobacterial *nifH* genes would be found in every heterocystous cyanobacterial cell, not just the heterocysts. Therefore, the relative abundance of heterocystous *Cyanobacteria nifH* in *nifH* gene libraries could easily overwhelm the numbers of *nifH* genes from non-cyanobacterial 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. It is unknown, however, if non-cyanobacterial diazotrophs would be identified by $^{15}N_2$ DNA-SIP 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.

4.4 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 9). 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 9), respectively. Further, the Steven et al. (2013) and Garcia-Pichel et al. (2013) sequence collections 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 OTUs in the combined data, than the Steven et al. (2013) study shares with Garcia-Pichel et al. (2013).

4.5 TEMPERATURE INFLUENCES ON HETEROCYSTOZUS *CYANOBACTERIA* RELATIVE ABUNDANCE

Although few putative diazotrophs identified by DNA-SIP were found in the Garcia-Pichel et al. (2013) and Steven et al. (2013) data, we did observe statistically significant relationships between 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 of *Scytonema hyalinum* FGP-7A and DC-A are positively correlated with mean annual temperature.

Yeager et al. (2012) found *nifH* gene abundance peaks in early summer and falls 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.

310 (2013), which is compositional in nature as opposed to absolute, but it does appear that the structure of
311 heterocystous cyanobacterial diazotroph communities is correlated to mean annual temperature if not the
312 absolute abundance of *nifH* genes.

4.6 ANALYSIS OF NEXT-GENERATION-SEQUENCING DNA-SIP DATA

313 Although DNA-SIP is a powerful technique, analysis of DNA-SIP data is not without ambiguities. One
314 limitation is the artificial boundary in the form of a selected adjusted p-value threshold (or false discovery
315 rate) that marks which OTUs we consider to be enriched in the heavy fractions of labeled CsCl gradients
316 (and thus have likely incorporated ^{15}N into their DNA during the incubation). In reality the metric we
317 use to quantify the magnitude of an OTU's response to a stable isotope is continuous, and there is only
318 an artificial boundary between which OTUs appear to have "responded" and which OTUs have unknown
319 response. For this reason, we have presented all the OTUs that satisfy our "response" criteria but focused
320 on the most strongly responding OTUs. As with any hypothesis-based statistical test, care should be taken
321 when interpreting the significance of results where p-values are near the selected threshold for rejecting
322 the null hypothesis.

4.7 CONCLUSION

323 It would seem unlikely, given their ubiquity and abundance, that heterocystous *Cyanobacteria* are
324 not key contributors to the BSC N-budget. But, the putative diazotrophs elucidated in this study and
325 in Steppe et al. (1996) in addition to the N-fixation rate data presented by (Johnson et al., 2005)
326 suggest there may be additional and significant non-cyanobacterial BSC diazotrophs specifically within
327 the *Clostrideaceae* and *Proteobacteria*. It seems clear that heterocystous *Cyanobacteria* increase in
328 abundance with BSC age (Yeager et al., 2004). It is less clear if this transition marks the emergence
329 of diazotrophy versus a re-structuring of the BSC diazotroph community from one dominated by
330 *Firmicutes* and *Proteobacteria* to one predominantly heterocystous *Cyanobacteria*. DNA-SIP is a valuable
331 tool in the molecular microbial ecologist's toolbox for identifying members of microbial community
332 functional guilds (Neufeld et al., 2007). PCR-based surveys of diagnostic marker genes and DNA-SIP
333 are both used to connect microbial phylogenetic types to microbial activities, but they occupy a non-
334 overlapping set of strengths and weaknesses. Combined these tools can powerfully reveal connections
335 between ecosystem membership/structure and function. Here we supplement previous surveys of BSC
336 *nifH* diversity, a diagnostic marker PCR-driven approach, with ^{15}N$_2$ DNA-SIP, and, while we do not
337 confirm previous results, we expand knowledge of BSC diazotroph diversity. Evaluating BSC N-fixation
338 due to climate change and physical disturbance requires a careful accounting of diazotrophs including
339 non-cyanobacterial types.

5 MATERIALS AND METHODS

5.1 FIELD SITE AND SAMPLE DESCRIPTION

340 Samples were taken from Green Butte, Arizona as previously described (site CP3, Beraldi-Campesi et al.
341 (2009)). All samples were from light crusts as described by Johnson et al. (2005).

5.2 SOIL CRUST INCUBATION

342 Light crust samples (37.5 cm^2, average mass 35 g) were incubated in sealed chambers under controlled
343 atmosphere and in the light for 4 days. Crusts were dry prior to time zero and were wetted at initiation of
344 experiment. Treatments included control air (unenriched headspace) and enriched air (>98% atom ^{15}N$_2$)
345 headspace. Samples were taken at 2 days and 4 days incubation. Acetylene reduction rates were measured
346 daily. DNA was extracted from 1 g of crust.

5.3   DNA EXTRACTION

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

5.4   DNA-SIP

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

5.5   PCR, LIBRARY NORMALIZATION AND DNA SEQUENCING

358 Barcoded PCR of bacterial and archaeal 16S rRNA genes, in preparation for 454 Pyrosequencing, was
359 carried out using primer set 515F/806R (Walters et al., 2011). The primer 806R contained an 8 bp barcode
360 sequence, a ”TC” linker, and a Roche 454 B sequencing adaptor, while the primer 515F contained the
361 Roche 454 A sequencing adapter. Each 25 μL reaction contained 1x PCR Gold Buffer (Roche), 2.5 mM
362 $MgCl_2$, 200 μM of each of the four dNTPs (Promega), 0.5 mg/mL BSA (New England Biolabs), 0.3 μM
363 of each primers, 1.25 U of Amplitaq Gold (Roche), and 8 μL of template. Template for each sample was
364 added at normalized amounts in an attempt to prevent chimera formation, and each sample was amplified
365 in triplicate. Thermal cycling occurred with an initial denaturation step of 5 minutes at 95C, followed
366 by 40 cycles of amplification (20s at 95C, 20s at 53C, 30s at 72C), and a final extension step of 5 min
367 at 72C. Triplicate amplicons were pooled and purified using Agencourt AMPure PCR purification beads,
368 following manufacturers protocol. Once cleaned, amplicons were quantified using PicoGreen nucleic acid
369 quantification dyes (Molecular Probes) and pooled together in equimolar amounts. Samples were sent to
370 the Environmental Genomics Core Facility at the University of South Carolina (now Selah Genomics) to
371 be run on a Roche FLX 454 pyrosequencing machine.

5.6   DATA ANALYSIS

372 *5.6.1   Sequence quality control*   Sequences were initially screened by maximum expected errors at a
373 specific read length threshold (Edgar, 2013) which has been shown to be as effective as denoising 454
374 reads with respect to removing pyrosequencing errors. Specifically, reads were first truncated to 230
375 nucleotides (nt) (all reads shorter than 230 nt were discarded) and any read that exceeded a maximum
376 expected error threshold of 1.0 was removed. After truncation and max expected error trimming, 91% of
377 original reads remained. The first 30 nt representing the forward primer and barcode on high quality,
378 truncated reads were trimmed. Remaining reads were taxonomically annotated using the ”UClust”
379 taxonomic annotation framework in the QIIME software package (Caporaso et al., 2010; Edgar, 2010)
380 with cluster seeds from Silva SSU rRNA database (Pruesse et al., 2007) 97% sequence identity OTUs as
381 reference (release 111Ref). Reads annotated as ”Chloroplast”, ”Eukaryota”, ”Archaea”, ”Unassigned” or
382 ”mitochondria” were culled from the dataset. Finally, reads were aligned to the Silva reference alignment
383 provided by the Mothur software package (Schloss et al., 2009) using the Mothur NAST aligner (DeSantis
384 et al., 2006). All reads that did not appear to align to the expected amplicon region of the SSU rRNA gene
385 were discarded. Quality control parameters removed 34716 of 258763 raw reads.

Table 3. Chosen 16S sequences for strains in Yeager et al. (2006) included as OTU centroids

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

386 *5.6.2 Sequence clustering* Sequences were distributed into OTUs using the UParse methodology
387 (Edgar, 2013). Specifically, cluster seeds were identified using USearch with a collection of non-redundant
388 reads sorted by count as input. The sequence identity threshold for establishing a new OTU centroid was
389 97%. After initial cluster centroid selection, select 16S rRNA sequences trimmed to the same 16S position
390 as the other centroids from Yeager et al. (2006) were added to the centroid collection. Specifically, Yeager
391 et al. (2006) Colorado Plateau or Moab, Utah sequences were added which included the 16S sequences
392 for *Calothrix* MCC-3A, *Nostoc commune* MCT-1, *Nostoc commune* MFG-1, *Scytonema hyalinum* DC-A,
393 *Scytonema hyalinum* FGP-7A, *Spirirestis rafaelensis* LQ-10. Centroid sequences that matched selected
394 Yeager et al. (2006) sequences with greater than to 97% sequence identity were subsequently removed
395 from the centroid collection. With USearch/UParse, potential chimeras are identified during OTU centroid
396 selection and are not allowed to become cluster centroids effectively removing chimeras from the read
397 pool. All quality controlled reads were then mapped to cluster centroids at an identity threshold of 97%
398 again using USearch. 95.6% of quality controlled reads could be mapped to centroids. Unmapped reads
399 do not count towards sample counts and are essentially removed from downstream analyses. The USearch
400 software version for cluster generation was 7.0.1090.

401 *5.6.3 Merging data from this study, Garcia-Pichel et al. (2013), and Steven et al. (2013)* As only
402 sequences without corresponding quality scores were publicly available from Garcia-Pichel et al. (2013)
403 and Steven et al. (2013), these data sets were only quality screened by determining if they covered the
404 expected region of the 16S gene (described above). All data (this study, Garcia-Pichel et al. (2013)
405 and Steven et al. (2013)) were included as input to USearch for OTU centroid selection and subsequent
406 mapping to OTU centroids.

407 *5.6.4 Phylogenetic tree* The alignment for the "*Clostridiaceae*" phylogeny was created using SSU-
408 Align which is based on Infernal (Nawrocki and Eddy, 2013; Nawrocki et al., 2009). Columns in
409 the alignment that were not included in the SSU-Align covariance models or were aligned with poor
410 confidence (less than 95% of characters in a position had posterior probability alignment scores of
411 at least 95%) were masked for phylogenetic reconstruction. Additionally, the alignment was trimmed
412 to coordinates such that all sequences in the alignment began and ended at the same positions. The
413 "*Clostridiaceae*" tree included all top BLAST hits (parameters below) for [15]N *Clostridiaceae* responders
414 in the Living Tree Project database (Yarza et al., 2008) in addition to BLAST hits within a sequence
415 identity threshold of 97% to ^{15}N responders from the Silva SSURef_NR SSU rRNA database (Pruesse
416 et al., 2007). Only one SSURef_NR115 hit per study per OTU ("study" was determined by "title" field)
417 was selected for the tree. FastTree (Price et al., 2010) was used to build the tree and support values are
418 SH-like scores reported by FastTree.

419 *Placement of short sequences into backbone phylogeny* Short sequences were mapped to the reference
420 backbone using pplacer (Matsen et al., 2010) (default parameters). pplacer finds the edge placements that

421 maximize phylogenetic likelihood. Prior to being mapped to the reference tree, short sequences were
422 aligned to the reference alignment using Infernal (Nawrocki et al., 2009) against the same SSU-Align
423 covariance model used to align reference sequences.

424 *5.6.5 BLAST searches*   BLAST searches were done with the "blastn" program from BLAST+ toolkit
425 (Camacho et al., 2009) version 2.2.29+. Default parameters were always employed and the BioPython
426 (Cock et al., 2009) BLAST+ wrapper was used to invoke the blastn program. Pandas (McKinney, 2012)
427 and dplyr (Wickham and Francois, 2014) were used to parse and munge BLAST output tables.

428 *5.6.6 Identifying OTUs that incorporated ^{15}N into their DNA*   SIP is a culture-independent approach
429 towards defining identity-function connections in microbial communities (Buckley, 2011; Neufeld et al.,
430 2007). Microbes incubated in the presence of ^{13}C or ^{15}N labeled substrates can incorporate the stable
431 heavy isotope into biomass if they participate in the substrate's transformation. Stable isotope labeled
432 nucleic acids can then be separated from unlabeled by buoyant density in a CsCl gradient. As the buoyant
433 density of a macromolecule is dependent on many factors in addition to stable isotope incorporation
434 (e.g. GC-content in nucleic acids (Youngblut and Buckley, 2014)), labeled nucleic acids from one
435 microbial population may have the same buoyant density of unlabeled nucleic acids from another (i.e.
436 each population's nucleic acids would be found at the same point along a density gradient although
437 only one population's nucleic acids are labeled). Therefore it is imperative to compare density gradients
438 with nucleic acids from heavy stable isotope incubations to gradients from "control" incubations where
439 everything mimics the experimental conditions except that unlabeled substrates are used (and all DNA
440 would be unlabeled). By contrasting "heavy" density gradient fractions in experimental density gradients
441 (hereafter referred to as "labeled" gradients) against heavy fractions in control gradients, the identities of
442 microbes with labeled nucleic acids can be determined

443   We used an RNA-Seq differential expression statistical framework (Love et al., 2014) to find OTUs
444 enriched in heavy fractions of labeled gradients relative to corresponding density fractions in control
445 gradients (for review of RNA-Seq differential expression statistics applied to microbiome OTU count data
446 see McMurdie and Holmes (2014)). We use the term differential abundance (coined by McMurdie and
447 Holmes (2014)) to denote OTUs that have different proportion means across sample classes (in this case
448 the only sample class is labeled/control). CsCl gradient fractions were categorized as "heavy" or "light".
449 The heavy category denotes fractions with density values above 1.725 g/mL. Since we are only interested
450 in enriched OTUs (labeled versus control), we used a one-sided z-test for differential abundance (the null
451 hypothesis is the labeled:control proportion mean ratio for an OTU is less than a selected threshold). P-
452 values were corrected with the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). We
453 selected a $\log_2$ fold change null threshold of 0.25 (or a labeled:control proportion mean ratio of 1.19).
454 DESeq2 was used to calculate the moderated $\log_2$ fold change of labeled:control proportion mean ratios
455 and corresponding standard errors. Mean ratio moderation allows for reliable ratio ranking such that
456 high variance and likely statistically insignificant mean ratios are appropriately shrunk and subsequently
457 ranked lower than they would be as raw ratios. To summarize, OTUs with high moderated labeled:control
458 proportion mean ratios have higher proportion means in heavy fractions of labeled gradients relative to
459 heavy fractions of control gradients, and therefore have likely incorporated ^{15}N into their DNA during the
460 incubation.

461 *5.6.7 Ordination*   Principal coordinate ordinations depict the relationship between samples at each time
462 point (day 2 and 4). Bray-Curtis distances were used as the sample distance metric for ordination. The
463 Phyloseq (McMurdie and Holmes, 2014) wrapper for Vegan (Oksanen et al., 2013) (both R packages) was
464 used to compute sample values along principal coordinate axes. GGplot2 (Wickham, 2009) was used to
465 display sample points along the first and second principal axes.

466 *5.6.8 Differential abundance in environmental samples* Significance of OTU proportion mean
467 differences with mean annual temperature (for Garcia-Pichel et al. (2013) data) and sample type ("BSC"
468 or "below crust" Steven et al. (2013) data) was determined using the DESeq2 framework (McMurdie and
469 Holmes, 2014; Love et al., 2014). A sparsity threshold of 0.40 was set to screen out sparse OTUs. No
470 p-value correction was done for differential abundance in environmental samples as only six OTUs were
471 considered for any test.

5.7 RICHNESS ANALYSES

472 Rarefaction curves were created using bioinformatics modules in the PyCogent Python package (Knight
473 et al., 2007). Parametric richness estimates were made with CatchAll using only the best model for total
474 OTU estimates (Bunge, 2010).

REFERENCES

475 Marti J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*,
476   26(1):32–46, Feb 2001. doi: 10.1111/j.1442-9993.2001.01070.pp.x. URL http://dx.doi.org/
477   10.1111/j.1442-9993.2001.01070.pp.x.
478 J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In
479   *Biological Soil Crusts: Structure Function, and Management*, pages 241–261. Springer Science +
480   Business Media, 2001. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.
481   1007/978-3-642-56475-8_19.
482 J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In
483   Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*,
484   volume 150 of *Ecological Studies*, pages 241–261. Springer Berlin Heidelberg, 2003. ISBN 978-
485   3-540-43757-4. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.1007/
486   978-3-642-56475-8_19.
487 J. Belnap, R. Prasse, and K.T. Harper. Influence of Biological Soil Crusts on Soil Environments and
488   Vascular Plants. In Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function,*
489   *and Management*, volume 150 of *Ecological Studies*, pages 281–300. Springer Berlin Heidelberg, 2003.
490   ISBN 978-3-540-43757-4. doi: 10.1007/978-3-642-56475-8_21. URL http://dx.doi.org/10.
491   1007/978-3-642-56475-8_21.
492 Jayne Belnap. Nitrogen fixation in biological soil crusts from southeast Utah USA. *Biology and Fertility*
493   *of Soils*, 35(2):128–135, Apr 2002. doi: 10.1007/s00374-002-0452-x. URL http://dx.doi.org/
494   10.1007/s00374-002-0452-x.
495 Jayne Belnap. Some Like It Hot, Some Not. *Science*, 340(6140):1533–1534, 2013. doi: 10.1126/science.
496   1240318. URL http://www.sciencemag.org/content/340/6140/1533.short.
497 Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful
498   Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57
499   (1):289–300, 1995. ISSN 00359246. doi: 10.2307/2346101. URL http://dx.doi.org/10.
500   2307/2346101.
501 H. Beraldi-Campesi, H. E. Hartnett, A. Anbar, G. W. Gordon, and F. Garcia-Pichel. Effect of biological
502   soil crusts on soil elemental concentrations: implications for biogeochemistry and as traceable
503   biosignatures of ancient life on land. *Geobiology*, 7(3):348–359, jun 2009. doi: 10.1111/j.1472-4669.
504   2009.00204.x. URL http://dx.doi.org/10.1111/j.1472-4669.2009.00204.x.
505 J. Roger Bray and J. T. Curtis. An Ordination of the Upland Forest Communities of Southern Wisconsin.
506   *Ecological Monographs*, 27(4):325, Oct 1957. doi: 10.2307/1942268. URL http://dx.doi.org/
507   10.2307/1942268.
508 Daniel H. Buckley. Stable Isotope Probing Techniques Using 15N. In *Stable Isotope Probing and*
509   *Related Technologies*, pages 129–147. American Society of Microbiology, jan 2011. doi: 10.1128/
510   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.

John Christian Gaby and Daniel H. Buckley. A Comprehensive Evaluation of {PCR} Primers to Amplify the {nifH} Gene of Nitrogenase. {PLoS} {ONE}, 7(7):e42149, jul 2012. doi: 10.1371/journal.pone.0042149. URL http://dx.doi.org/10.1371/journal.pone.0042149.

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

564 from sequence. *Genome Biol*, 8(8):R171, 2007. doi: 10.1186/gb-2007-8-8-r171. URL http://dx.
565 doi.org/10.1186/gb-2007-8-8-r171.
566 M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for {RNA}-
567 Seq data with {DESeq}2. Technical report, feb 2014. URL http://dx.doi.org/10.1101/
568 002832.
569 Frederick A Matsen, Robin B Kodner, and E Virginia Armbrust. pplacer: linear time maximum-likelihood
570 and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*,
571 11(1):538, 2010. doi: 10.1186/1471-2105-11-538. URL http://dx.doi.org/10.1186/
572 1471-2105-11-538.
573 Wes McKinney. pandas: Python Data Analysis Library. Online, 2012. URL http://pandas.
574 pydata.org/.
575 PJ McMurdie and S Holmes. Waste not, want not: why rarefying microbiome data is inadmissible. 10:
576 e1003531, 2014.
577 EP Nawrocki and SR Eddy. Infernal 1.1: 100-fold faster RNA homology searches. 29:2933–5, Nov 2013.
578 EP Nawrocki, DL Kolbe, and SR Eddy. Infernal 1.0: inference of RNA alignments. 25:1335–7, May
579 2009.
580 JD Neufeld, J Vohra, MG Dumont, T Lueders, M Manefield, MW Friedrich, and JC Murrell. DNA
581 stable-isotope probing. 2:860–6, 2007.
582 Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara,
583 Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community
584 Ecology Package*, 2013. URL http://CRAN.R-project.org/package=vegan. R package
585 version 2.0-10.
586 MN Price, PS Dehal, and AP Arkin. FastTree 2–approximately maximum-likelihood trees for large
587 alignments. 5:e9490, Mar 2010.
588 E Pruesse, C Quast, K Knittel, BM Fuchs, W Ludwig, J Peplies, and FO Glckner. SILVA: a comprehensive
589 online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.
590 35:7188–96, 2007.
591 PD Schloss, SL Westcott, T Ryabin, JR Hall, M Hartmann, EB Hollister, RA Lesniewski, BB Oakley,
592 DH Parks, CJ Robinson, JW Sahl, B Stres, GG Thallinger, Horn DJ Van, and CF Weber.
593 Introducing mothur: open-source, platform-independent, community-supported software for describing
594 and comparing microbial communities. 75:7537–41, 2009.
595 T.F. Steppe, J.B. Olson, H.W. Paerl, R.W. Litaker, and J. Belnap. Consortial N2 fixation: a strategy
596 for meeting nitrogen requirements of marine and terrestrial cyanobacterial mats. *FEMS Microbiology
597 Ecology*, 21(3):149–156, Nov 1996. doi: 10.1111/j.1574-6941.1996.tb00342.x. URL http://dx.
598 doi.org/10.1111/j.1574-6941.1996.tb00342.x.
599 Blaire Steven, La Verne Gallegos-Graves, Jayne Belnap, and Cheryl R. Kuske. Dryland soil microbial
600 communities display spatial biogeographic patterns associated with soil depth and soil parent material.
601 *FEMS Microbiol Ecol*, 86(1):101–113, May 2013. doi: 10.1111/1574-6941.12143. URL http:
602 //dx.doi.org/10.1111/1574-6941.12143.
603 WA Walters, JG Caporaso, CL Lauber, D Berg-Lyons, N Fierer, and R Knight. PrimerProspector: de novo
604 design and taxonomic analysis of barcoded polymerase chain reaction primers. 27:1159–61, Apr 2011.
605 Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-
606 387-98140-6. URL http://had.co.nz/ggplot2/book.
607 Hadley Wickham and Romain Francois. *dplyr: dplyr: a grammar of data manipulation*, 2014. URL
608 http://CRAN.R-project.org/package=dplyr. R package version 0.2.
609 Pablo Yarza, Michael Richter, Jörg Peplies, Jean Euzeby, Rudolf Amann, Karl-Heinz Schleifer, Wolfgang
610 Ludwig, Frank Oliver Glöckner, and Ramon Rosselló-Móra. The All-Species Living Tree project: A
611 16S rRNA-based phylogenetic tree of all sequenced type strains. *Systematic and Applied Microbiology*,
612 31(4):241–250, Sep 2008. doi: 10.1016/j.syapm.2008.07.001. URL http://dx.doi.org/10.
613 1016/j.syapm.2008.07.001.
614 Chris M. Yeager, Jennifer L. Kornosky, Rachael E. Morgan, Elizabeth C. Cain, Ferran Garcia-Pichel,
615 David C. Housman, Jayne Belnap, and Cheryl R. Kuske. Three distinct clades of cultured heterocystous
616 cyanobacteria constitute the dominant N2-fixing members of biological soil crusts of the Colorado

617   Plateau USA. *FEMS Microbiology Ecology*, 60(1):85–97, 2006. doi: 10.1111/j.1574-6941.2006.00265.
618   x. URL http://dx.doi.org/10.1111/j.1574-6941.2006.00265.x.
619 Chris M. Yeager, Cheryl R. Kuske, Travis D. Carney, Shannon L. Johnson, Lawrence O. Ticknor, and
620   Jayne Belnap. Response of Biological Soil Crust Diazotrophs to Season Altered Summer Precipitation,
621   and Year-Round Increased Temperature in an Arid Grassland of the Colorado Plateau, USA. *Front.*
622   *Microbio.*, 3, 2012. doi: 10.3389/fmicb.2012.00358. URL http://dx.doi.org/10.3389/
623   fmicb.2012.00358.
624 CM Yeager, JL Kornosky, DC Housman, EE Grote, J Belnap, and CR Kuske. Diazotrophic community
625   structure and function in two successional stages of biological soil crusts from the Colorado Plateau
626   and Chihuahuan Desert. 70:973–83, 2004.
627 ND Youngblut and DH Buckley. Intra-genomic variation in G+C content and its implications for DNA
628   stable isotope probing (DNA-SIP). Aug 2014.

6 FIGURES AND LONG TABLES
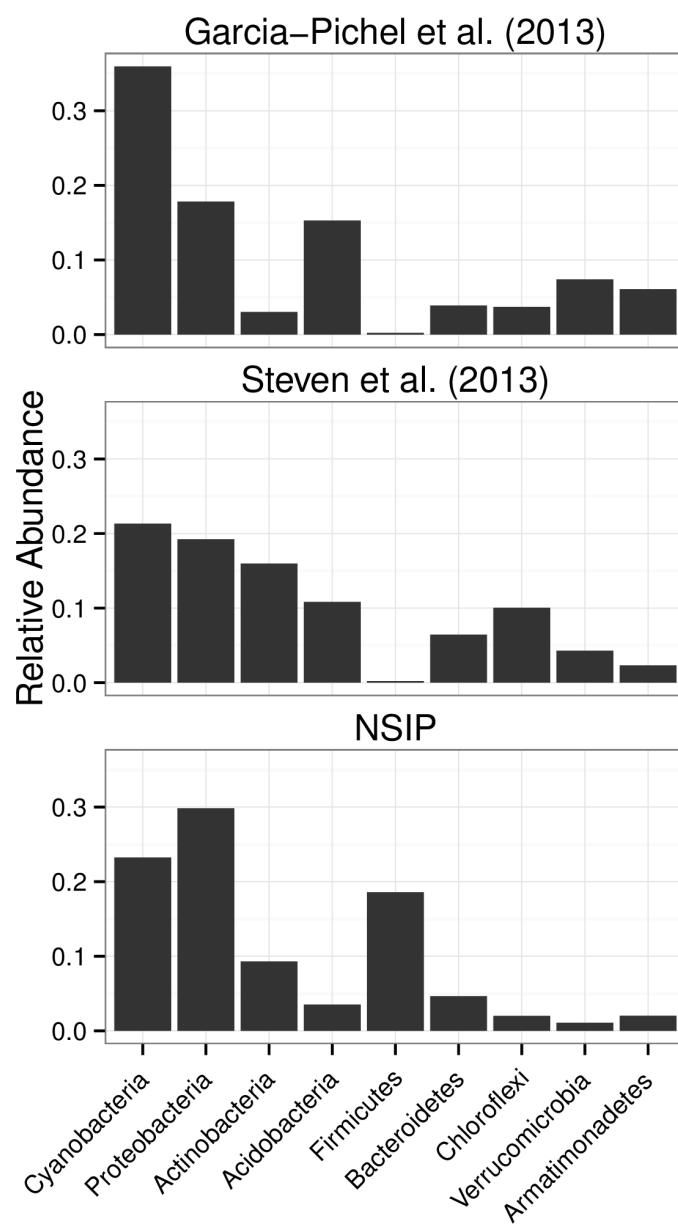
**Figure 1.** Distribution of sequences into top 9 phyla (phyla ranked by sum of all sequence annotations).

Table 4.[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
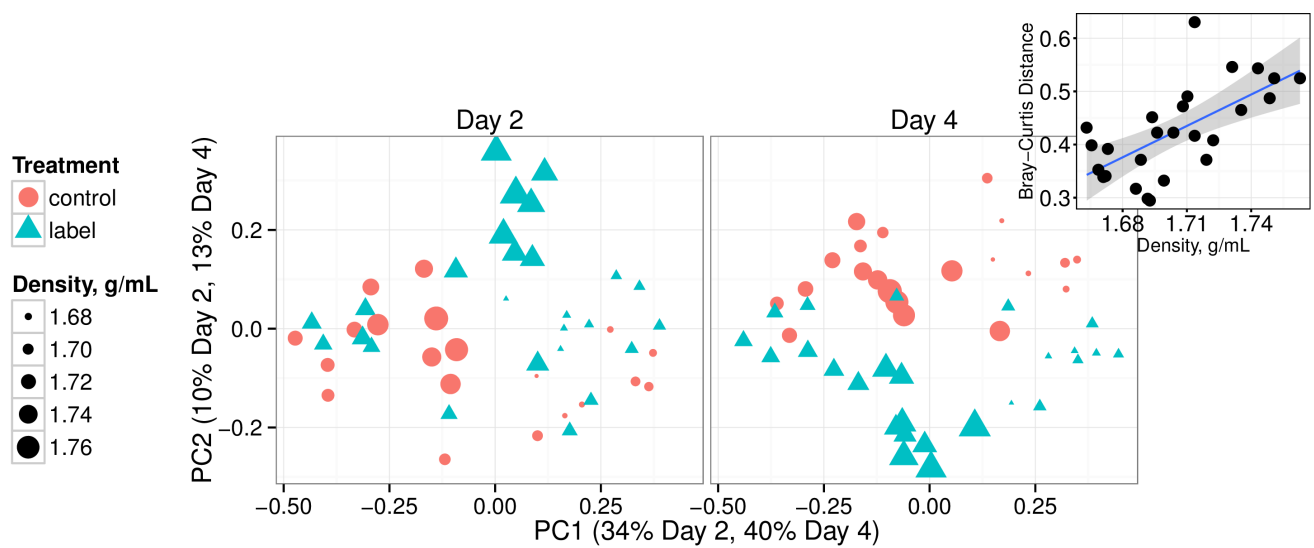
Figure 2. Ordination of Bray-Curtis sample pairwise distances for each incubation time. Point area is proportional to the density of the CsCl gradient fraction for each sequence library, and color/shape reflects control (red triangles) or labeled (blue circles) treatment. Inset shows Bray-Curtis distances 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: $4.526e^{-5}$, r^2: 0.434).
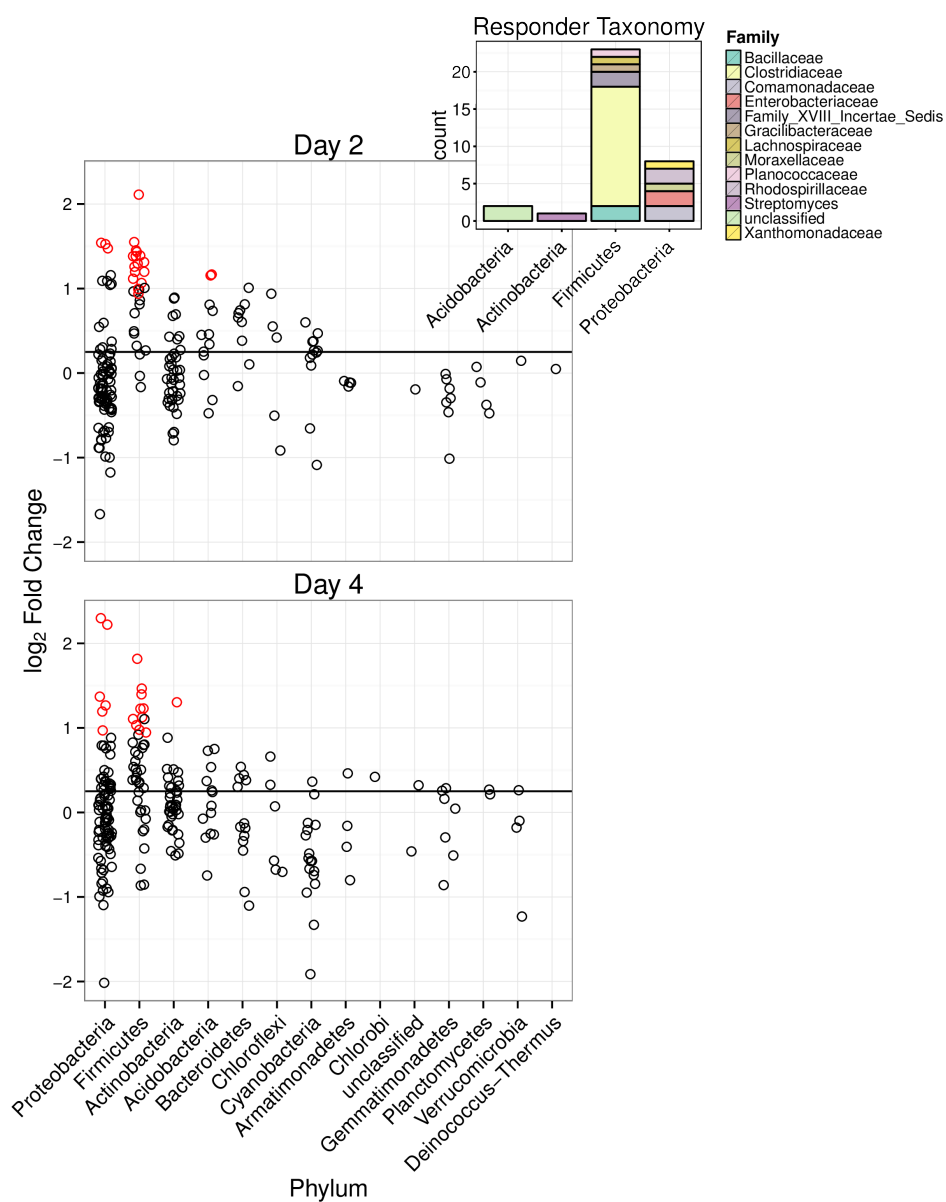
Figure 3. Moderated log$_2$ of proportion mean ratios for labeled versus control gradients (heavy fractions only, densities ¿1.725 g/mL). All OTUs found in at least 62.5% of heavy fractions at a specific incubation day are shown. Red color denotes a proportion mean ratio that has a corresponding adjusted 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.
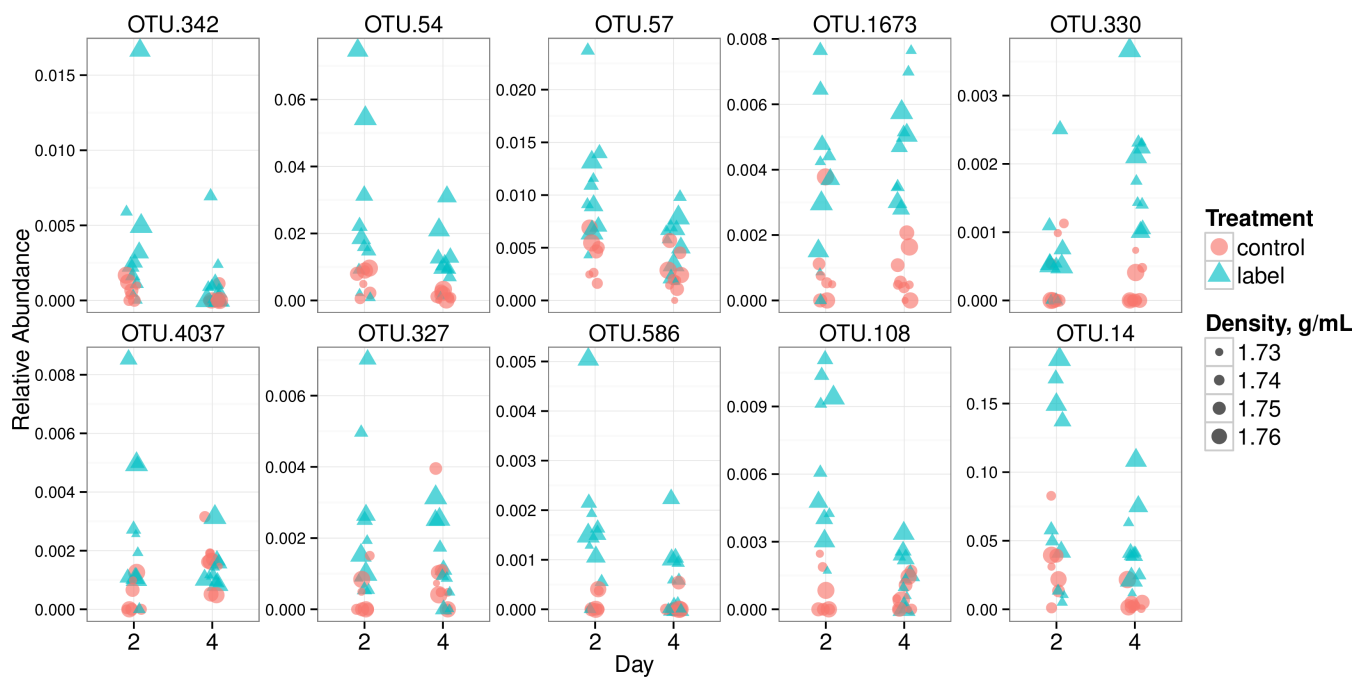
Figure 4. Relative abundance values in heavy fractions (density greater or equal to 1.725 g/mL) for the top 10 [15]N "responders" (putative diazotrophs, see results for selection criteria of top 10) at each incubation day. See Table 4 for BLAST results of top 10 responders against the LTP database (release 115). Point area is proportional to CsCl gradient fraction density, and color signifies control (red) or labeled (blue) treatment.

Figure 5. 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.
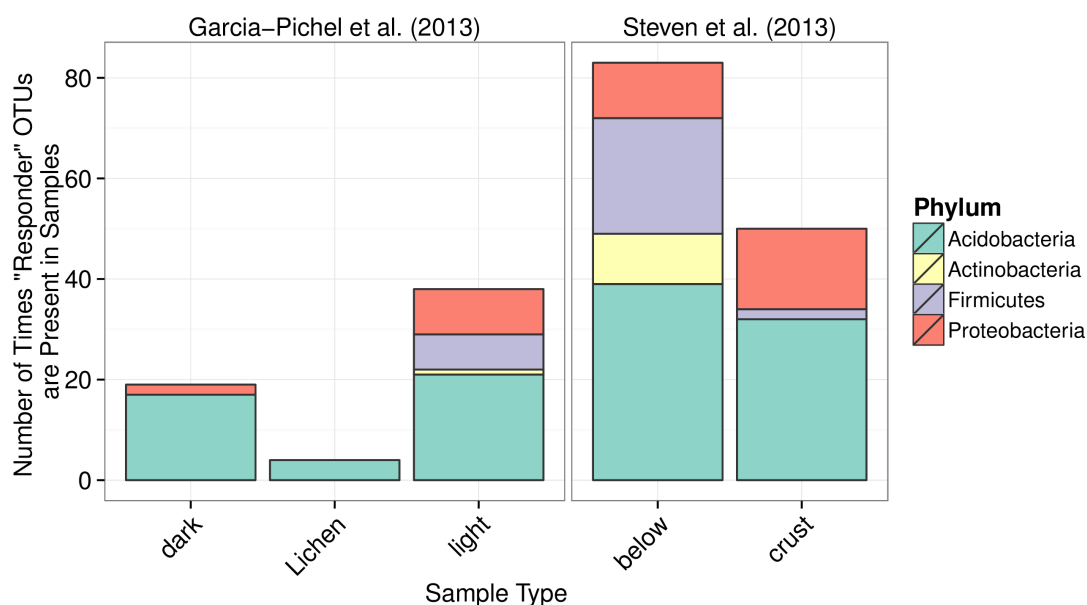
Figure 6. Counts of "responder" OTU occurrences in samples from Steven et al. (2013) and Garcia-Pichel et al. (2013). Steven et al. (2013) collected BSC samples (25 samples total) and samples from soil beneath BSC (17 samples total, "below" column in figure). Garcia-Pichel et al. (2013) collected samples from "dark" (9 samples total) and "light" (12 samples total) crusts in addition to "lichen" (2 samples total) dominated crusts.
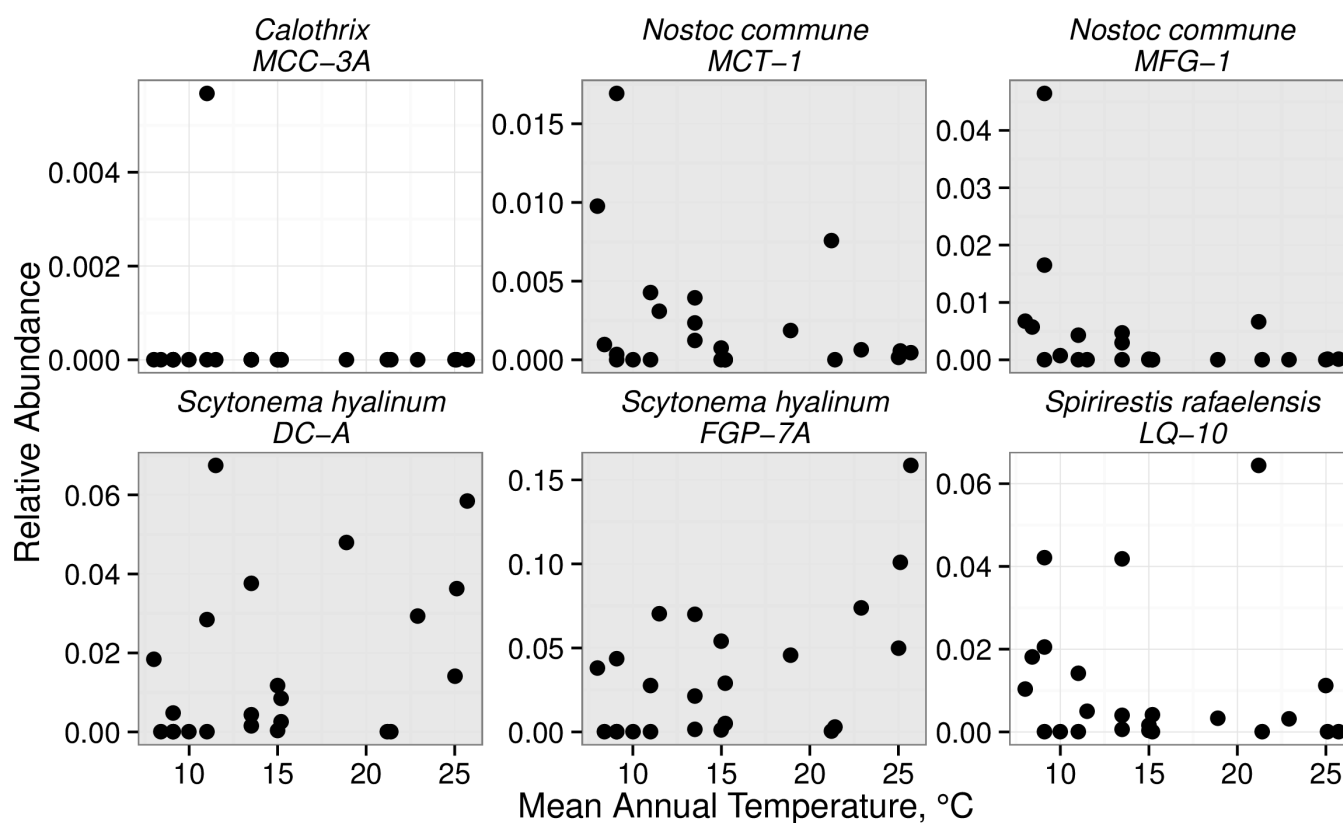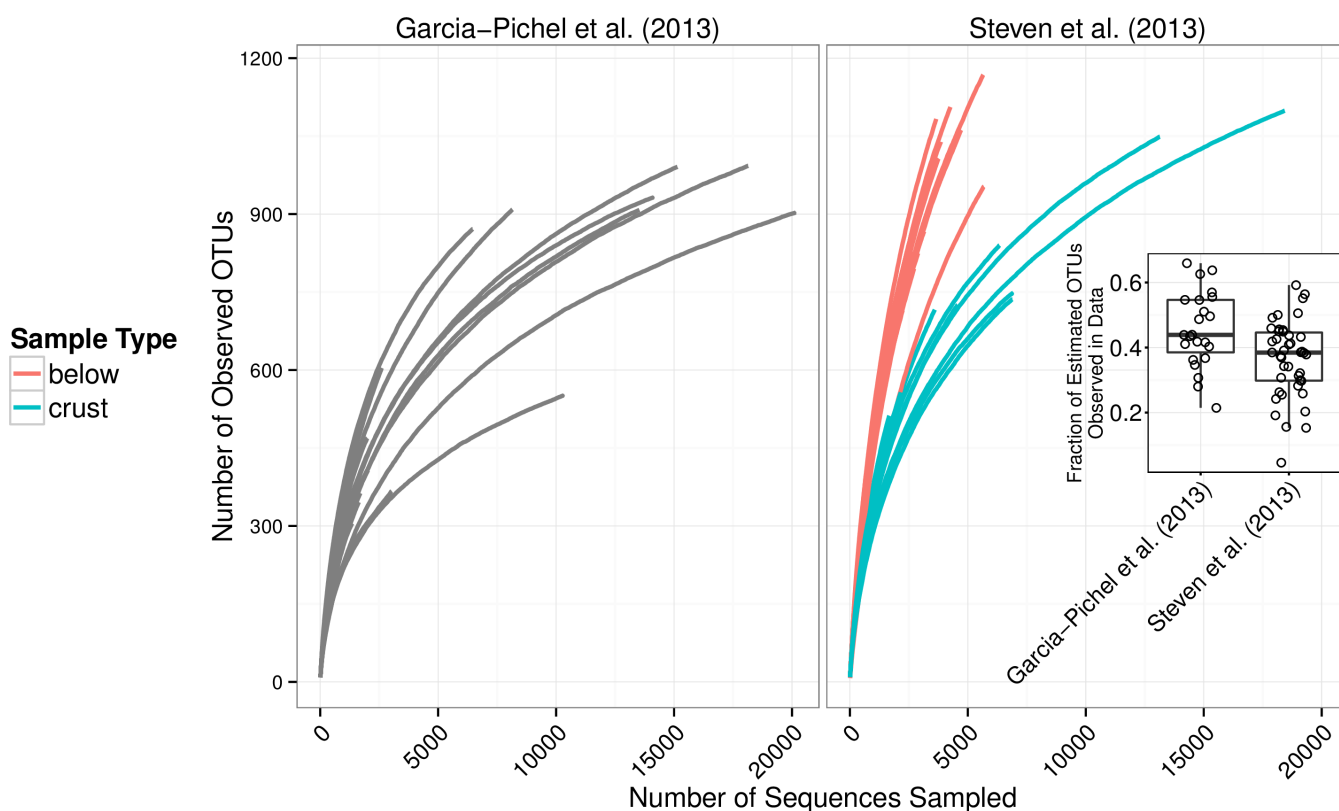

Figure 7. Relative abundance of selected heterocystous cyanobacterial OTUs with centroids from sequences described in Yeager et al. (2006) (see methods for selection criteria) in Steven et al. (2013) data set.
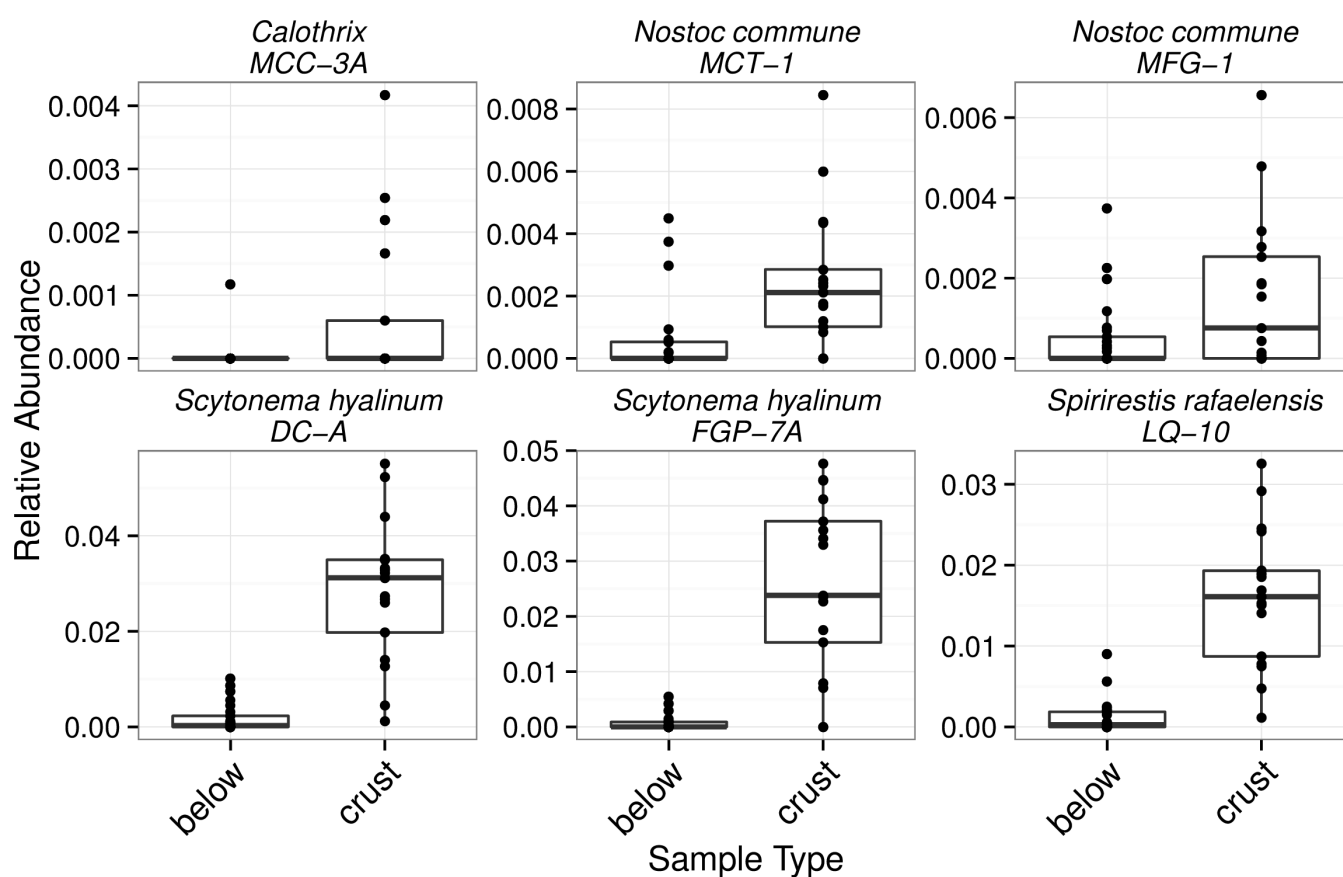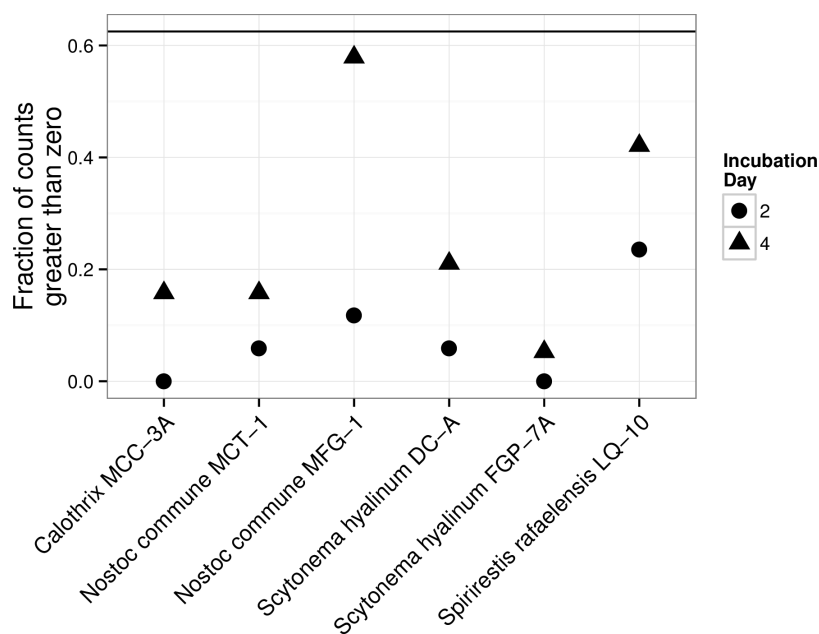
Figure 8. Relative abundance of selected heterocystous cyanobacterial OTUs with centroids from sequences described in Yeager et al. (2006) (see methods for selection criteria) in Steven et al. (2013) data set.

Figure 9. 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 10. 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. Horizontal line is the sparsity threshold for independent OTU filtering prior to adjusting p-values when identifying OTUS enriched in labeled gradients (heavy fractions).