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

Biological soil crusts (BSC) cover a vast global area and are key components of ecosystem productivity in arid soils. In particular, BSC contribute significantly to the nitrogen (N) budget in arid ecosystems via N_2-fixation. N_2-fixation in mature crusts is largely attributed to heterocystous cyanobacteria, however, early successional crusts possess few N-fixing cyanobacteria and this suggests that microorganisms other than cyanobacteria mediate N_2-fixation during the early stages of BSC development. DNA stable isotope probing (DNA-SIP) with $^{15}N_2$ revealed that *Clostridiaceae* and *Proteobacteria* are the most common microorganisms to assimilate ^{15}N in early successional 'light' crusts. The maximum relative abundance of non-cyanobacterial $^{15}N_2$-assimilating taxa in environmental BSC SSU rRNA gene sequence collections was 0.00225% and 0.00127% for taxa that belong to *Clostridiaceae* and *Proteobacteria*, respectively. Their low abundance may explain why these heterotrophic diazotrophs have not previously been characterized in BSC. Diazotrophs play a critical role in BSC formation and characterization of these organisms represents a crucial step towards understanding how antropogenic change will effect the formation and ecological function of BSC in arid ecosystems.

2 INTRODUCTION

Biological soil crusts (BSC) are specialized microbial mat communites that form at the soil surface in arid environmets and fill a variety of important ecological functions in arid ecosystems. BSC occupy 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, N-input via N-fixation versus atmospheric depositon was predominant in five times as many BSC samples from North America, Africa and Australia (Evans and Belnap, 1999). The presence of BSC is positively correlated with vascular plant survival due in part to BSC ecosystem N contributions (for review of BSC-vascular plant interactions see Belnap et al. (2003)). Climate change and disturbance alter BSC microbial community structure and membership and therefore can alter diazotroph diversity and the BSC N-budget.

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). Nonetheless, mature BSC N-fixation rate measurements have been higher than younger, developing BSC N-fixation rate measurements (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 peak at a lower depth in developing BSC as compared to mature BSC. When N-fixation is measured from intact cores of developing BSC the measurement may be artifactually low due to delayed acetylene/ethylene diffusion through the crust to and from the peak N-fixation rate depth in a typical ARA incubation timeframe. Diffusion would not be an issue when measuring N-fixation rates in mature crust as nitrogenase activity peaks near the surface. When total N-fixation rates were calculated by integrating rates over 1-3 mm depth slices along full BSC cores (thus mitigating ethene/acetylene flux limitations), N-fixation rate differences between developing and mature BSC were not statistically significant (Johnson et al., 2005).

Molecular studies of BSC microbial diversity include explorations of the BSC microbial community vertical profile (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). *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; Yeager et al., 2004). Heterocystous *Cyanobacteria* are the numerically dominant BSC diazotrophs in *nifH* clone libraries (Yeager et al., 2006, 2004, 2012) although an early survey of Colorado Plateau BSC *nifH* diversity recovered *nifH* genes related to *Gammaproteobacteria* as well as a clade that included *nifH* genes from the anaerobes *Clostridium pssteurianum*, *Desulfovibrio gigas* and *Chromatium buderi*, 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).

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 mechanism behind increased N-fixation in mature BSC, it is unclear if mature BSC actually fix more N (see Johnson et al. (2005)). More studies are necessary to elucidate the microbial membership influence on BSC N-fixation and to determine if heterocystous *Cyanobacteria* are the only keystone diazotrophs. The first step in defining structure function relationships with respect to N-fixation is a full accounting of BSC diazotrophs. Towards this end we conducted $^{15}N_2$ DNA stable isotope probing (DNA-SIP) experiments with light, developing Colorado Plateau BSC. DNA-SIP with $^{15}N_2$ has not been attempted with BSC. DNA-SIP provides an accounting of *active* diazotrophs whereas *nifH* clone libraries account for microbes with the genomic potential for N-fixation. Further, we track the distribution of putative diazotrophs uncovered in this study 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).

3 RESULTS

3.1 ORDINATION OF CSCL GRADIENT FRACTION SSU RRNA LIBRARIES

78 BSC were incubated for 4 days in the presence or absence of $^{15}N_2$ and DNA was extracted for DNA-SIP at
79 2 and 4 days. Fractionation of CsCl gradients permitted separation of DNA on the basis of buoyant density.
80 Ordination of Bray-Curtis (Bray and Curtis, 1957) distances between SSU-rRNA amplicon sequence
81 collections from gradient fractions reveals that labeled gradient fraction (i.e. gradient fractions of DNA
82 from $^{15}N_2$ incubations) sequence collections diverge from control (i.e. DNA from incubations without
83 $^{15}N_2$) at the "heavy" of the CsCl gradients (Figure 1 and Figure S2). Although the density position
84 of gradient fractions from different gradients do not match perfectly, fraction pairs from corresponding
85 control versus labeled gradients can be constructed by pairing control gradient fractions with their closest
86 density neighbors/fractions from corresponding labeled gradients. If a gradient fraction did not have a
87 mate within a density difference of 0.003 g/mL it remained unpaired. Bray-Curtis distance between
88 the fraction pairs is positively correlated to the density of the labeled fraction (p-value: 4.536e-05, r^2:
89 0.434) (inset Figure S2). Additionally, differences among label/control groups with heavy fractions are
90 statistically significant by the Adonis test (p-value: 0.001, r^2: 0.18) (Anderson, 2001). The first principal
91 axis appears to be correlated with fraction density (Figure S2) and the Adonis test p-value for density
92 versus pairwise Bray-Curtis distances with all CsCl fraction libraries is 0.001 (r^2 0.111).

3.2 IDENTITIES OF OTUS RESPONSIVE TO $^{15}N_2$

93 A statistically significant increase in OTU abundance in heavy fractions of $^{15}N_2$ labeled samples relative
94 to corresponding gradient fractions from controls provides evidence for OTUs that have incorporated
95 ^{15}N into their DNA. Specifically, we compared OTU proportion means between labeled and control
96 samples from heavy gradient fractions using statistics developed to find differentially expressed genes with
97 RNASeq data (McMurdie and Holmes, 2014; Love et al., 2014). p-values were adjusted by the BH method
98 (Benjamini and Hochberg, 1995) and we used a false discovery rate (FDR) cutoff of 0.10 (typical FDR
99 threshold in gene expression data analysis, (Love et al., 2014)) to reject the null hypothesis that labeled
100 versus control proportion mean differences were below a chosen threshold (see methods). With the above
101 methods 38 OTUs had labeled versus control proportion mean difference adjusted p-values below 0.10
102 for one or both incubation days. These OTUs likely incorporated ^{15}N into DNA ($^{15}N_2$ "responders"). Of
103 these 38, 26 are annotated as *Firmicutes*, 9 as *Proteobacteria*, 2 as *Acidobacteria* and 1 as *Actinobacteria*
104 (Figure 3, Figure 2). If the OTUs are ranked by descending, moderated proportion mean labeled:control
105 ratios, the top 10 ratios (i.e. the 10 OTUs that were most enriched in the labeled gradients considering only
106 heavy fractions) are either *Firmicutes* (6 OTUs) or *Proteobacteria* (4 OTUs) (Figure 4). *Proteobacteria*
107 OTU centroid sequences for the top 10 responders all share high identity (>98.48% identity, Table 1)
108 with cultivars from genera known to possess diazotrophs including *Klebsiella*, *Shigella*, *Acinetobacter*,
109 and *Ideonella*. None of the *Firmicutes* OTUs in the top 10 responders share greater than 97% sequence
110 identity with sequences in the LTP database (release 115) (see Table 1).

111 If we run a second test of differential OTU abundance with the null hypothesis that OTU abundance is
112 greater (above a threshold) in labeled, heavy gradient fractions verus control, heavy gradient fractions, we
113 can count OTUs that likely did not respond the the label. There were 58 and 70 "non-responders" at days
114 2 and 4, respectively. 208 and 233 of 2,127 and 2,160 OTUs passed our sparsity threshold (OTUs had to
115 have counts in at least 62.5% of heavy fractions) for days 2 and 4, respectively..

3.3 COMPARISON OF SEQUENCE COLLECTIONS AT "STUDY"-LEVEL

116 *3.3.1 Comparisons of OTU content:* There were 3,079 OTUs (209,354 total sequences after quality
117 control) in the DNA-SIP data, 3,203 OTUs (129,033 total sequences after quality control) in the Garcia-
118 Pichel et al. (2013) study, and 2,481 OTUs (129,358 total sequences after quality control) in the Steven

119 et al. (2013) study. Of the 4,340 OTU centroids established for this study (including sequences from
120 Steven et al. (2013) and Garcia-Pichel et al. (2013)) 445 have matches in the Living Tree Project (LTP)
121 (a collection of 16S gene sequences for all sequenced type strains (Yarza et al., 2008)) at greater or equal
122 than 97% (LTP version 115). That is, 445 of 4,340 OTUs are closely related to cultivars. The DNA-SIP
123 data set shares 56% OTUs with the Steven et al. (2013) data and 46% of OTUs with the Garcia-Pichel
124 et al. (2013) data (where total OTUs are from the combined data for each pairwise comparison). The
125 Steven et al. (2013) and Garcia-Pichel et al. (2013) share 46% of OTUs.

126 *3.3.2 Comparisons of Taxonomic Content:* *Cyanobacteria* and *Proteobacteria* were the top two
127 phylum-level sequence annotations for all three studies but only the DNA-SIP data had more
128 *Proteobacteria* annotations than *Cyanobacteria*. *Proteobacteria* represented the 29.8% of sequence
129 annotations in DNA-SIP data as opposed to 17.8% and 19.2% for the Garcia-Pichel et al. (2013) and
130 Steven et al. (2013) data, respectively. There is a stark contrast in the total percentage of sequences
131 annotated as *Firmicutes* between the raw environmental samples and the DNA-SIP data. *Firmicutes*
132 represent only 0.21% and 0.23% of total phylum level sequence annotations in the Steven et al. (2013)
133 and Garcia-Pichel et al. (2013) studies, respectively (Figure S1). In the DNA-SIP sequence collection
134 *Firmicutes* make up 19% of phylum level sequence annotations. Also in sharp contrast for the DNA-
135 SIP versus environmental data is the number of putative heterocystous *Cyanobacteria* sequences. Only
136 0.29% of *Cyanobacteria* sequences in the DNA-SIP data are annotated as belonging to "Subsection IV"
137 which is the heterocystous order of *Cyanobacteria* in the Silva taxonomic nomenclature (Pruesse et al.,
138 2007). In the Steven et al. (2013) and Garcia-Pichel et al. (2013) studies 15% and 23%, respectively, of
139 *Cyanobacteria* sequences are annotated as belonging to "Subsection IV".

3.4 DISTRIBUTION OF BSC DIAZOTROPHS IN ENVIRONMENTAL SAMPLES

140 *3.4.1 Clostridiacea:* Five of the 6 *Firmicutes* in the top 10 responder OTUs (above) belong in
141 the *Clostridiacea*. We only observed one of these strongly responding *Clostridiaceae* in the data
142 presented by Garcia-Pichel et al. (2013), "OTU.108" (closest BLAST hit in LTP Release 115 –
143 *Caloramotor proteoclasticus*, BLAST %ID 96.94, Accession X90488). OTU.108 was found in two
144 samples both characterized as "light" crust. One other *Clostridiaceae* OTU with a proportion mean ratio
145 (labeled:control) p-value less than 0.10 but outside the top 10 responders was found in the Garcia-Pichel
146 et al. (2013) data (a "light" crust sample) (Figure 2). None of the strongly responding *Clostridiacea* were
147 found in the sequences provided by Steven et al. (2013). *Clostridiaceae* ^{15}N-responder OTU centroid 16S
148 sequences are generally more closely related to environmental than cultivar 16S gene sequences (Table 1,
149 Figure 5).

150 *3.4.2 Proteobacteria:* One of the *Proteobacteria* OTUs in the 10 most strongly responding OTUs was
151 found in the Garcia-Pichel et al. (2013) sequences (closest BLAST hit in LTP Release 115, BLAST %ID
152 100, Accession ZD3440, *Acinetobacter johnsonii*). None of the strongly responding *Protebacteria* OTUs
153 were found in the Steven et al. (2013) sequences. There were 133 responder OTU-sample occurrences
154 (responder OTU was found in a sample library) in the Steven et al. (2013) data. 83 were in "below crust"
155 samples, 50 in BSC samples (see Figure 2).

156 *3.4.3 Other taxa:* Two ^{15}N-responsive OTUs were found in an extensive number of environmental
157 samples (61 of 65 samples from the combined data sets of Garcia-Pichel et al. (2013) and Steven et al.
158 (2013)). Both OTUs were annotated as *Acidobacteria* but shared little sequence identity to any cultivar
159 SSU rRNA gene sequences in the LTP (Release 115), with best LTP BLAST hits of 81.91 and 81.32%
160 identity. Additionally, the evidence for ^{15}N incorporation for each OTU was weak relative to other putative
161 responders (adjusted p-values of 0.090 and 0.096). Of the remaining 36 stable isotope responder OTUs,
162 only 14 were observed in the environmental data (Figure 2, Figure S5).

4 DISCUSSION

4.1 STUDY-LEVEL DIFFERENCES

163 SIP places focus upon organisms based on isotope incorporation and has the ability to detect activity by
164 low abundance members of the community. DNA from OTUs that incopororate ^{15}N into their biomass
165 moves towards the heavy end of the CsCl gradient and therefore OTUs in "labeled" DNA are enriched
166 in the full data pool relative to bulk DNA. Phylum-level taxonomic annotations of ^{15}N-responsive OTUs
167 (i.e. *Firmicutes* and *Proteobacteria*) are enriched in the DNA-SIP data relative to environmental data
168 (Figure S1).

4.2 ORDINATION OF CSCL GRADIENT FRACTION 16S RRNA GENE SEQUENCE COLLECTIONS

169 The ordination of Bray-Curtis distances between CsCl gradient fraction 16S sequence collections show
170 that control fractions differ from labeled fractions in the "heavy" range of the CsCl gradients (Figure S2).
171 If each control fraction is paired to the labeled fraction from the same incubation day for which it is closest
172 in density, there is a positive and statistically significant correlation between Bray-Curtis distances within
173 fraction pairs and density of the pair (see inset Figure S2). Therefore, the "heavy" end of the control and
174 labeled gradients differ and the OTUs enriched in the labeled fractions (relative to control) would have
175 incorporated ^{15}N into their DNA during the incubation timeframe.

4.3 BSC DIAZOTROPHS IDENTIFIED IN THE STUDY

176 BSC N-fixation has long been attributed to heterocystous *Cyanobacteria* and molecular microbial ecology
177 surveys of BSC *nifH* gene content have been consistent with this hypothesis finding cyanobacterial *nifH*
178 types to be numerically dominant in *nifH* gene libraries (Yeager et al., 2006, 2004, 2012). Even poorly
179 developed BSC samples have yielded predominantly cyanobacterial *nifH* genes (Yeager et al., 2004).
180 And, "sub-biocrust" samples have yielded *entirely* heterocystous cyanobacterial *nifH* genes (Yeager
181 et al., 2012). It is possible, however, that PCR-driven molecular surveys of *nifH* gene content have
182 been biased against non-heterocystous *Cyanobacteria*. In general the nifH PCR primers used by Yeager
183 et al. (2006, 2004, 2012) (19F and nifH3) for the first round of nested PCR have broad specificity and
184 display at least 86% *in silico* coverage for *Proteobacteria, Cyanobacteria* and "Cluster III" *nifH* reference
185 sequences (Gaby and Buckley, 2012). In the second round of the nested PCR protocol (Yeager et al.,
186 2006, 2004, 2012), primer nifH11 is' slightly biased against "Cluster III" (50% coverage) but biased
187 in favor of *Proteobacteria* (79% *in silico* coverage against 67% for *Cyanobacteria*) and nifH22 matches
188 *Proteobacteria*, *Cyanobacteria* and "Cluster III" reference sequences poorly (16%, 23% and 21% *in silico*
189 coverage, respectively) (Gaby and Buckley, 2012). Unfortunately, it is difficult to assess or quantify this
190 bias (in either direction) without knowing the *nifH* gene content *de novo*. Another potential bias in favor of
191 *Cyanobacteria* in BSC *nifH* gene libraries is heterocysts (the specialized N-fixing cells along the trichome
192 of filamentous heterocystous *Cyanobacteria* such as *Nostoc* and *Scytonema*) may be overrepresented
193 with respect to non-cyanobacterial diazotrophs because heterocysts make up a fraction cells along a
194 trichome and even non-heterocyst cells in a trichome will possess the *nifH* gene. Polyploidy could further
195 exacerbate this bias, as many *Cyanobacteria* are estimated to have multiple genome copies per cell (Griese
196 et al., 2011). Moreover, it should also be noted that *nifH* gene content is not directly extrapolable to
197 the taxonomic relative abundances of nitrogenase proteins. Regardless, our results suggest that BSC N-
198 fixation may include a significant non-cyanobacterial component that requires further assessment across
199 a more comprehensive sampling of BSC types.

200 We did not observe evidence for N-fixation by heterocystous *Cyanobacteria* in the "light" crust samples
201 used in this study. One possible explanation for our results is that the "light", still developing BSC samples
202 used in this study possessed too few heterocystous *Cyanobacteria* to statistically evaluate their ^{15}N-
203 incorporation. Indeed, only 0.29% of sequences 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. Nonetheless, we would still expect even low abundance diazotrophs to show evidence for ^{15}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). Given the sparsity of heterocystous *Cyanobacteria* sequences in the DNA-SIP data set, it is not possible to assess whether heterocystous *Cyanobacteria* incorporated ^{15}N during the incubation. It should be noted that "light" and in particular "sub-biocrust" samples possess much less heterocystous *Cyanobacteria* in general (Figure S3) so the samples used in this study are not necessarily unrepresentative of typical poorly developed BSC simply because they are lacking heterocystous *Cyanobacteria*.

The OTUs that did appear to incorporate ^{15}N during the incubation were predominantly *Proteobacteria* and *Firmicutes*. The *Proteobacteria* OTUs for which ^{15}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 1). The *Firmicutes* that displayed signal for ^{15}N-incorporation (predominantly *Clostridiaceae*) were not closely related to any cultivars (Table 1, Figure 5). These BSC *Clostrodiaceae* diazotrophs represent a gap in culture collections. 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 for predicting how climate change will affect the BSC nitrogen budget. *Gamma-proteobacteria* and spore-forming *Firmicutes* are classic opportunistic lineages that would presumably be suited to the boom/bust BSC environment. The compatible solutes produced and secreted by cyanobacteria in response to dessication and subsequent wetting could produce a local C-rich environment. Diazotrophs would be uniquely suited to respond quickly in C-rich but N-poor conditions.

Although too undersampled in the environmental data sets to reach statistical conclusions, ^{15}N-responsive OTUs were found more often in below crust samples (as opposed to BSC samples) in the Steven et al. (2013) data and in "light" BSC samples in the Garcia-Pichel et al. (2013) data (Figure S5). This result generates some hypotheses that are counter to prior discussions 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 diazotrophs in BSC but rather the *transition* of the diazotroph community from heterotroph dominance to cyanobacteria. Additionally, the soil beneath BSC may contribute significantly to the N budget in arid ecosystems.

4.4 SEQUENCING DEPTH

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 S4). 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 S4), 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). Therefore, is not alarming that few of the ^{15}N-responsive OTUS were 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.

4.5 CONCLUSION

249 Heterocystous *Cyanobacteria* are key contributors to the BSC N-budget, but, the ^{15}N-responsive OTUs
250 found in this study and the *nifH* gene sequences from Steppe et al. (1996) in addition to the N-fixation
251 rate data presented by (Johnson et al., 2005) suggest there may be significant non-cyanobacterial BSC
252 diazotrophs specifically within the *Clostrideaceae* and *Proteobacteria*. It seems clear that heterocystous
253 *Cyanobacteria* increase in abundance with BSC age (Yeager et al., 2004). It is less clear if this transition
254 marks the emergence of diazotrophy versus a re-structuring of the BSC diazotroph community from one
255 dominated by *Firmicutes* and *Proteobacteria* to one predominantly heterocystous *Cyanobacteria*. DNA-
256 SIP is a valuable tool in the molecular microbial ecologist's toolbox for identifying members of microbial
257 community functional guilds (Neufeld et al., 2007). PCR-based surveys of diagnostic marker genes and
258 DNA-SIP are both used to connect microbial phylogenetic types to microbial activities, but they occupy a
259 non-overlapping set of strengths and weaknesses. Combined these tools can powerfully reveal connections
260 between ecosystem membership/structure and function. Here we supplement previous surveys of BSC
261 *nifH* diversity, a diagnostic marker PCR-driven approach, with ^{15}N$_2$ DNA-SIP, and, while we do not
262 confirm previous results, we expand knowledge of BSC diazotroph diversity. Predicting BSC N-fixation
263 with respect to climate change, althered precipitation regimes and physical disturbance requires a careful
264 accounting of diazotrophs including non-cyanobacterial types.

5 MATERIALS AND METHODS

5.1 BSC SAMPLING AND INCUBATION CONDITIONS

265 Light crust samples (37.5 cm^2, average mass 35 g) were incubated in sealed chambers under controlled
266 atmosphere and in the light for 4 days. Crusts were dry prior to time zero and were wetted at initiation of
267 experiment. Treatments included control air (unenriched headspace) and enriched air (>98% atom ^{15}N$_2$)
268 headspace. Samples were taken at 2 days and 4 days incubation. Acetylene reduction rates were measured
269 daily. DNA was extracted from 1 g of crust. Samples were taken from Green Butte, Arizona as previously
270 described (site CP3, Beraldi-Campesi et al. (2009)). All samples were from light crusts as described by
271 Johnson et al. (2005).

5.2 DNA EXTRACTION

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

5.3 DNA-SIP

275 Gradient density centrifugation of DNA was undertaken in 4.7 mL polyallomer centrifuge tubes in a
276 TLA-110 fixed angle rotor (both Beckman Coulter) in CsCl gradients with an average density of 1.725
277 g/mL. Average density for all prepared gradients was checked with an AR200 refractometer before runs.
278 Between 2.5-5 μg of DNA extract was added to the CsCl solution (15mM Tris-HCl, pH 8; 15mM EDTA;
279 15mM KCl), and gradients were run under conditions of 20C for 67 hours at 55,000 rpm (Buckley et al.).
280 Centrifuged gradients were fractionated from bottom to top in 36 equal fractions of 100 μL, using a by
281 syringe pump as described Manefield et al. (2002). The density of each fraction was determined using
282 using an AR200 refractometer modified to accomodate 5ul samples (Buckley et al.). DNA in each fraction
283 was desalted on a filter plate (PALL, AcroPrep Advance 96 Filter Plate, Product Number 8035), using four
284 washes with 300μL TE per fraction. After each wash, the filter plate was spun at 500 g for 10 minutes,
285 with a final spin of 20 minutes. Fractions were resuspended in 50 uL of TE buffer.

5.4 PCR, LIBRARY NORMALIZATION AND DNA SEQUENCING

286 Barcoded PCR of bacterial and archaeal 16S rRNA genes, in preparation for 454 Pyrosequencing, was
287 carried out using primer set 515F/806R (Walters et al., 2011) (primers purchased from Integrated DNA
288 Technologies). The primer 806R contained an 8 bp barcode sequence, a "TC" linker, and a Roche 454
289 B sequencing adaptor, while the primer 515F contained the Roche 454 A sequencing adapter. Each
290 25 μL reaction contained 1x PCR Gold Buffer (Roche), 2.5 mM $MgCl_2$, 200 μM of each of the four
291 dNTPs (Promega), 0.5 mg/mL BSA (New England Biolabs), 0.3 μM of each primers, 1.25 U of Amplitaq
292 Gold (Roche), and 8 μL of template. Each sample was amplified in triplicate. Thermal cycling occurred
293 with an initial denaturation step of 5 minutes at 95C, followed by 40 cycles of amplification (20s at
294 95C, 20s at 53C, 30s at 72C), and a final extension step of 5 min at 72C. Triplicate amplicons were
295 pooled and purified using Agencourt AMPure PCR purification beads, following manufacturers protocol.
296 Once cleaned, amplicons were quantified using PicoGreen nucleic acid quantification dyes (Molecular
297 Probes) and pooled together in equimolar amounts. Samples were sent to the Environmental Genomics
298 Core Facility at the University of South Carolina (now Selah Genomics) to be run on a Roche FLX 454
299 pyrosequencing machine.

5.5 DATA ANALYSIS

300 All code to take raw sequencing data through the presented figures can be found at:

301 `http://nbviewer.ipython.org/github/chuckpr/NSIP_data_analysis`

302 *5.5.1 Sequence quality control* Sequences were initially screened by maximum expected errors at a
303 specific read length threshold (Edgar, 2013) which has been shown to be as effective as denoising 454
304 reads with respect to removing pyrosequencing errors. Specifically, reads were first truncated to 230
305 nucleotides (nt) (all reads shorter than 230 nt were discarded) and any read that exceeded a maximum
306 expected error threshold of 1.0 was removed. After truncation and max expected error trimming, 91% of
307 original reads remained. The first 30 nt representing the forward primer and barcode on high quality,
308 truncated reads were trimmed. Remaining reads were taxonomically annotated using the "UClust"
309 taxonomic annotation framework in the QIIME software package (Caporaso et al., 2010; Edgar, 2010)
310 with cluster seeds from Silva SSU rRNA database (Pruesse et al., 2007) 97% sequence identity OTUs as
311 reference (release 111Ref). Reads annotated as "Chloroplast", "Eukaryota", "Archaea", "Unassigned" or
312 "mitochondria" were culled from the dataset. Finally, reads were aligned to the Silva reference alignment
313 provided by the Mothur software package (Schloss et al., 2009) using the Mothur NAST aligner (DeSantis
314 et al., 2006). All reads that did not appear to align to the expected amplicon region of the SSU rRNA gene
315 were discarded. Quality control parameters removed 34,716 of 258,763 raw reads.

316 *5.5.2 Sequence clustering* Sequences were distributed into OTUs using the UParse methodology
317 (Edgar, 2013). Specifically, cluster seeds were identified using USearch with a collection of non-redundant
318 reads sorted by count as input. The sequence identity threshold for establishing a new OTU centroid
319 was 97%. After initial cluster centroid selection, select 16S rRNA gene sequences trimmed to the same
320 alignment positions as the other centroids from Yeager et al. (2006) were added to the centroid collection.
321 Specifically, Yeager et al. (2006) Colorado Plateau or Moab, Utah sequences were added which included
322 the 16S rRNA gene sequences for *Calothrix* MCC-3A (accession DQ531700.1), *Nostoc commune* MCT-1
323 (accession DQ531903), *Nostoc commune* MFG-1 (accession DQ531699.1), *Scytonema hyalinum* DC-A
324 (accession DQ531701.1), *Scytonema hyalinum* FGP-7A (accession DQ531697.1), *Spirirestis rafaelensis*
325 LQ-10 (accession DQ531696.1). Centroid sequences that matched selected Yeager et al. (2006) sequences
326 with greater than to 97% sequence identity were subsequently removed from the centroid collection. With
327 USearch/UParse, potential chimeras are identified during OTU centroid selection and are not allowed to
328 become cluster centroids effectively removing chimeras from the read pool. All quality controlled reads
329 were then mapped to cluster centroids at an identity threshold of 97% again using USearch. 95.6% of

330 quality controlled reads could be mapped to centroids. Unmapped reads do not count towards sample
331 counts and are essentially removed from downstream analyses. The USearch software version for cluster
332 generation was 7.0.1090.

333 *5.5.3 Merging data from this study, Garcia-Pichel et al. (2013), and Steven et al. (2013)* As only
334 sequences without corresponding quality scores were publicly available from Garcia-Pichel et al. (2013)
335 and Steven et al. (2013), these data sets were only quality screened by determining if they covered the
336 expected region of the 16S rRNA gene (described above). All data (this study, Garcia-Pichel et al. (2013)
337 and Steven et al. (2013)) were included as input to USearch for OTU centroid selection and subsequent
338 mapping to OTU centroids.

339 *5.5.4 Phylogenetic tree* The alignment for the "*Clostridiaceae*" phylogeny was created using SSU-
340 Align which is based on Infernal (Nawrocki and Eddy, 2013; Nawrocki et al., 2009). Columns in
341 the alignment that were not included in the SSU-Align covariance models or were aligned with poor
342 confidence (less than 95% of characters in a position had posterior probability alignment scores of
343 at least 95%) were masked for phylogenetic reconstruction. Additionally, the alignment was trimmed
344 to coordinates such that all sequences in the alignment began and ended at the same positions. The
345 "*Clostridiaceae*" tree included all top BLAST hits (parameters below) for ^{15}N *Clostridiaceae* responders
346 in the Living Tree Project database (Yarza et al., 2008) in addition to BLAST hits within a sequence
347 identity threshold of 97% to ^{15}N responders from the Silva SSURef_NR SSU rRNA database (Pruesse
348 et al., 2007). Only one SSURef_NR115 hit per study per OTU ("study" was determined by "title" field)
349 was selected for the tree. FastTree (Price et al., 2010) was used to build the tree and support values are
350 SH-like scores reported by FastTree.

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

356 *5.5.5 BLAST searches* BLAST searches were done with the "blastn" program from BLAST+ toolkit
357 (Camacho et al., 2009) version 2.2.29+. Default parameters were always employed and the BioPython
358 (Cock et al., 2009) BLAST+ wrapper was used to invoke the blastn program. Pandas (McKinney, 2012)
359 and dplyr (Wickham and Francois, 2014) were used to parse and munge BLAST output tables.

360 *5.5.6 Identifying OTUs that incorporated ^{15}N into their DNA* SIP is a culture-independent approach
361 towards defining identity-function connections in microbial communities (Buckley, 2011; Neufeld et al.,
362 2007). Microbes incubated in the presence of ^{13}C or ^{15}N labeled substrates can incorporate the stable
363 heavy isotope into biomass if they participate in the substrate's transformation. Stable isotope labeled
364 nucleic acids can then be separated from unlabeled by buoyant density in a CsCl gradient. As the buoyant
365 density of a macromolecule is dependent on many factors in addition to stable isotope incorporation
366 (e.g. GC-content in nucleic acids (Youngblut and Buckley, 2014)), labeled nucleic acids from one
367 microbial population may have the same buoyant density of unlabeled nucleic acids from another (i.e.
368 each population's nucleic acids would be found at the same point along a density gradient although
369 only one population's nucleic acids are labeled). Therefore it is imperative to compare density gradients
370 with nucleic acids from heavy stable isotope incubations to gradients from "control" incubations where
371 everything mimics the experimental conditions except that unlabeled substrates are used (and all DNA
372 would be unlabeled). By contrasting "heavy" density gradient fractions in experimental density gradients
373 (hereafter referred to as "labeled" gradients) against heavy fractions in control gradients, the identities of
374 microbes with labeled nucleic acids can be determined

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

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

403 *5.5.7 Ordination*   Principal coordinate ordinations depict the relationship between samples at each time
404 point (day 2 and 4). Bray-Curtis distances were used as the sample distance metric for ordination. The
405 Phyloseq (McMurdie and Holmes, 2014) wrapper for Vegan (Oksanen et al., 2013) (both R packages) was
406 used to compute sample values along principal coordinate axes. GGplot2 (Wickham, 2009) was used to
407 display sample points along the first and second principal axes. Adonis tests Anderson (2001) were done
408 with default number of permutations (1000).

5.6 RICHNESS ANALYSES

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

REFERENCES

412 Marti J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*,
413   26(1):32–46, Feb 2001. doi: 10.1111/j.1442-9993.2001.01070.pp.x. URL http://dx.doi.org/
414   10.1111/j.1442-9993.2001.01070.pp.x.
415 J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In
416   *Biological Soil Crusts: Structure Function, and Management*, pages 241–261. Springer Science +

417   Business Media, 2001. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.
418   1007/978-3-642-56475-8_19.
419 J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In
420   Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*,
421   volume 150 of *Ecological Studies*, pages 241–261. Springer Berlin Heidelberg, 2003. ISBN 978-
422   3-540-43757-4. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.1007/
423   978-3-642-56475-8_19.
424 J. Belnap, R. Prasse, and K.T. Harper. Influence of Biological Soil Crusts on Soil Environments and
425   Vascular Plants. In Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function,
426   and Management*, volume 150 of *Ecological Studies*, pages 281–300. Springer Berlin Heidelberg, 2003.
427   ISBN 978-3-540-43757-4. doi: 10.1007/978-3-642-56475-8_21. URL http://dx.doi.org/10.
428   1007/978-3-642-56475-8_21.
429 Jayne Belnap. Nitrogen fixation in biological soil crusts from southeast Utah USA. *Biology and Fertility*
430   *of Soils*, 35(2):128–135, Apr 2002. doi: 10.1007/s00374-002-0452-x. URL http://dx.doi.org/
431   10.1007/s00374-002-0452-x.
432 Jayne Belnap. Some Like It Hot, Some Not. *Science*, 340(6140):1533–1534, 2013. doi: 10.1126/science.
433   1240318. URL http://www.sciencemag.org/content/340/6140/1533.short.
434 Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful
435   Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57
436   (1):289–300, 1995. ISSN 00359246. doi: 10.2307/2346101. URL http://dx.doi.org/10.
437   2307/2346101.
438 H. Beraldi-Campesi, H. E. Hartnett, A. Anbar, G. W. Gordon, and F. Garcia-Pichel. Effect of biological
439   soil crusts on soil elemental concentrations: implications for biogeochemistry and as traceable
440   biosignatures of ancient life on land. *Geobiology*, 7(3):348–359, jun 2009. doi: 10.1111/j.1472-4669.
441   2009.00204.x. URL http://dx.doi.org/10.1111/j.1472-4669.2009.00204.x.
442 J. Roger Bray and J. T. Curtis. An Ordination of the Upland Forest Communities of Southern Wisconsin.
443   *Ecological Monographs*, 27(4):325, Oct 1957. doi: 10.2307/1942268. URL http://dx.doi.org/
444   10.2307/1942268.
445 Daniel H. Buckley. Stable Isotope Probing Techniques Using 15N. In *Stable Isotope Probing and*
446   *Related Technologies*, pages 129–147. American Society of Microbiology, jan 2011. doi: 10.1128/
447   9781555816896.ch7. URL http://dx.doi.org/10.1128/9781555816896.ch7.
448 DH Buckley, V Huangyutitham, SF Hsu, and TA Nelson. Stable isotope probing with 15N2 reveals novel
449   noncultivated diazotrophs in soil. *Appl Environ Microbiol*, 73:3196–204.
450 John Bunge. Estimating the Number of Species with Catchall. In *Biocomputing 2011*, pages 121–130.
451   WORLD SCIENTIFIC, nov 2010. doi: 10.1142/9789814335058_0014. URL http://dx.doi.
452   org/10.1142/9789814335058_0014.
453 C Camacho, G Coulouris, V Avagyan, N Ma, J Papadopoulos, K Bealer, and TL Madden. BLAST+:
454   architecture and applications. 10:421, Dec 2009.
455 JG Caporaso, J Kuczynski, J Stombaugh, K Bittinger, FD Bushman, EK Costello, N Fierer, AG Pea,
456   JK Goodrich, JI Gordon, GA Huttley, ST Kelley, D Knights, JE Koenig, RE Ley, CA Lozupone,
457   D McDonald, BD Muegge, M Pirrung, J Reeder, JR Sevinsky, PJ Turnbaugh, WA Walters, J Widmann,
458   T Yatsunenko, J Zaneveld, and R Knight. QIIME allows analysis of high-throughput community
459   sequencing data. 7:335–6, 2010.
460 PJ Cock, T Antao, JT Chang, BA Chapman, CJ Cox, A Dalke, I Friedberg, T Hamelryck, F Kauff,
461   B Wilczynski, and Hoon MJ de. Biopython: freely available Python tools for computational molecular
462   biology and bioinformatics. 25:1422–3, 2009.
463 TZ Jr DeSantis, P Hugenholtz, K Keller, EL Brodie, N Larsen, YM Piceno, R Phan, and GL Andersen.
464   NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. 34:W394–9,
465   2006.
466 RC Edgar. Search and clustering orders of magnitude faster than BLAST. 26:2460–1, 2010.
467 RC Edgar. UPARSE: highly accurate OTU sequences from microbial amplicon reads. 10:996–8, 2013.
468 R. D. Evans and J. Belnap. Long-Term Consequences of Disturbance on Nitrogen Dynamics in an Arid
469   Ecosystem. *Ecology*, 80(1):150–160, Jan 1999. doi: 10.1890/0012-9658(1999)080[0150:ltcodo]2.

470 0.co;2. URL `http://dx.doi.org/10.1890/0012-9658(1999)080[0150:LTCODO]2.`
471 `0.CO;2.`
472 R. D. Evans and O. L. Lange. Biological Soil Crusts and Ecosystem Nitrogen and Carbon Dynamics.
473 In *Biological Soil Crusts: Structure Function, and Management*, pages 263–279. Springer Science +
474 Business Media, 2001. doi: 10.1007/978-3-642-56475-8_20. URL `http://dx.doi.org/10.`
475 `1007/978-3-642-56475-8_20.`
476 John Christian Gaby and Daniel H. Buckley. A Comprehensive Evaluation of {PCR} Primers to Amplify
477 the {nifH} Gene of Nitrogenase. {PLoS} {ONE}, 7(7):e42149, jul 2012. doi: 10.1371/journal.pone.
478 0042149. URL `http://dx.doi.org/10.1371/journal.pone.0042149.`
479 F. Garcia-Pichel, S. L. Johnson, D. Youngkin, and J. Belnap. Small-Scale Vertical Distribution of Bacterial
480 Biomass and Diversity in Biological Soil Crusts from Arid Lands in the Colorado Plateau. *Microbial*
481 *Ecology*, 46(3):312–321, Nov 2003a. doi: 10.1007/s00248-003-1004-0. URL `http://dx.doi.`
482 `org/10.1007/s00248-003-1004-0.`
483 F. Garcia-Pichel, V. Loza, Y. Marusenko, P. Mateo, and R. M. Potrafka. Temperature Drives the
484 Continental-Scale Distribution of Key Microbes in Topsoil Communities. *Science*, 340(6140):
485 1574–1577, Jun 2013. doi: 10.1126/science.1236404. URL `http://dx.doi.org/10.1126/`
486 `science.1236404.`
487 Ferran Garcia-Pichel, Jayne Belnap, Susanne Neuer, and Ferdinand Schanz. Estimates of global
488 cyanobacterial biomass and its distribution. *Algological Studies*, 109(1):213–227, 2003b.
489 Marco Griese, Christian Lange, and Jrg Soppa. Ploidy in cyanobacteria. *FEMS Microbiology Letters*, 323
490 (2):124–131, sep 2011. doi: 10.1111/j.1574-6968.2011.02368.x. URL `http://dx.doi.org/10.`
491 `1111/j.1574-6968.2011.02368.x.`
492 SL Johnson, CR Budinoff, J Belnap, and F Garcia-Pichel. Relevance of ammonium oxidation within
493 biological soil crust communities. 7:1–12, 2005.
494 A. Karnieli, R.F. Kokaly, N.E. West, and R.N. Clark. Remote Sensing of Biological Soil Crusts. In
495 Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*,
496 volume 150 of *Ecological Studies*, pages 431–455. Springer Berlin Heidelberg, 2003. ISBN 978-
497 3-540-43757-4. doi: 10.1007/978-3-642-56475-8_31. URL `http://dx.doi.org/10.1007/`
498 `978-3-642-56475-8_31.`
499 Rob Knight, Peter Maxwell, Amanda Birmingham, Jason Carnes, J Gregory Caporaso, Brett C Easton,
500 Michael Eaton, Micah Hamady, Helen Lindsay, Zongzhi Liu, Catherine Lozupone, Daniel McDonald,
501 Michael Robeson, Raymond Sammut, Sandra Smit, Matthew J Wakefield, Jeremy Widmann, Shandy
502 Wikman, Stephanie Wilson, Hua Ying, and Gavin A Huttley. {PyCogent}: a toolkit for making sense
503 from sequence. *Genome Biol*, 8(8):R171, 2007. doi: 10.1186/gb-2007-8-8-r171. URL `http://dx.`
504 `doi.org/10.1186/gb-2007-8-8-r171.`
505 M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for {RNA}-
506 Seq data with {DESeq}2. Technical report, feb 2014. URL `http://dx.doi.org/10.1101/`
507 `002832.`
508 Frederick A Matsen, Robin B Kodner, and E Virginia Armbrust. pplacer: linear time maximum-likelihood
509 and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*,
510 11(1):538, 2010. doi: 10.1186/1471-2105-11-538. URL `http://dx.doi.org/10.1186/`
511 `1471-2105-11-538.`
512 Wes McKinney. pandas: Python Data Analysis Library. Online, 2012. URL `http://pandas.`
513 `pydata.org/.`
514 PJ McMurdie and S Holmes. Waste not, want not: why rarefying microbiome data is inadmissible. 10:
515 e1003531, 2014.
516 EP Nawrocki and SR Eddy. Infernal 1.1: 100-fold faster RNA homology searches. 29:2933–5, Nov 2013.
517 EP Nawrocki, DL Kolbe, and SR Eddy. Infernal 1.0: inference of RNA alignments. 25:1335–7, May
518 2009.
519 JD Neufeld, J Vohra, MG Dumont, T Lueders, M Manefield, MW Friedrich, and JC Murrell. DNA
520 stable-isotope probing. 2:860–6, 2007.
521 Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara,
522 Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community*

523 *Ecology Package*, 2013. URL http://CRAN.R-project.org/package=vegan. R package
524   version 2.0-10.
525 MN Price, PS Dehal, and AP Arkin. FastTree 2–approximately maximum-likelihood trees for large
526   alignments. 5:e9490, Mar 2010.
527 E Pruesse, C Quast, K Knittel, BM Fuchs, W Ludwig, J Peplies, and FO Glckner. SILVA: a comprehensive
528   online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.
529   35:7188–96, 2007.
530 PD Schloss, SL Westcott, T Ryabin, JR Hall, M Hartmann, EB Hollister, RA Lesniewski, BB Oakley,
531   DH Parks, CJ Robinson, JW Sahl, B Stres, GG Thallinger, Horn DJ Van, and CF Weber.
532   Introducing mothur: open-source, platform-independent, community-supported software for describing
533   and comparing microbial communities. 75:7537–41, 2009.
534 T.F. Steppe, J.B. Olson, H.W. Paerl, R.W. Litaker, and J. Belnap. Consortial N2 fixation: a strategy
535   for meeting nitrogen requirements of marine and terrestrial cyanobacterial mats. *FEMS Microbiology*
536   *Ecology*, 21(3):149–156, Nov 1996. doi: 10.1111/j.1574-6941.1996.tb00342.x. URL http://dx.
537   doi.org/10.1111/j.1574-6941.1996.tb00342.x.
538 Blaire Steven, La Verne Gallegos-Graves, Jayne Belnap, and Cheryl R. Kuske. Dryland soil microbial
539   communities display spatial biogeographic patterns associated with soil depth and soil parent material.
540   *FEMS Microbiol Ecol*, 86(1):101–113, May 2013. doi: 10.1111/1574-6941.12143. URL http:
541   //dx.doi.org/10.1111/1574-6941.12143.
542 WA Walters, JG Caporaso, CL Lauber, D Berg-Lyons, N Fierer, and R Knight. PrimerProspector: de novo
543   design and taxonomic analysis of barcoded polymerase chain reaction primers. 27:1159–61, Apr 2011.
544 Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-
545   387-98140-6. URL http://had.co.nz/ggplot2/book.
546 Hadley Wickham and Romain Francois. *dplyr: dplyr: a grammar of data manipulation*, 2014. URL
547   http://CRAN.R-project.org/package=dplyr. R package version 0.2.
548 Pablo Yarza, Michael Richter, Jörg Peplies, Jean Euzeby, Rudolf Amann, Karl-Heinz Schleifer, Wolfgang
549   Ludwig, Frank Oliver Glöckner, and Ramon Rosselló-Móra. The All-Species Living Tree project: A
550   16S rRNA-based phylogenetic tree of all sequenced type strains. *Systematic and Applied Microbiology*,
551   31(4):241–250, Sep 2008. doi: 10.1016/j.syapm.2008.07.001. URL http://dx.doi.org/10.
552   1016/j.syapm.2008.07.001.
553 Chris M. Yeager, Jennifer L. Kornosky, Rachael E. Morgan, Elizabeth C. Cain, Ferran Garcia-Pichel,
554   David C. Housman, Jayne Belnap, and Cheryl R. Kuske. Three distinct clades of cultured heterocystous
555   cyanobacteria constitute the dominant N2-fixing members of biological soil crusts of the Colorado
556   Plateau USA. *FEMS Microbiology Ecology*, 60(1):85–97, 2006. doi: 10.1111/j.1574-6941.2006.00265.
557   x. URL http://dx.doi.org/10.1111/j.1574-6941.2006.00265.x.
558 Chris M. Yeager, Cheryl R. Kuske, Travis D. Carney, Shannon L. Johnson, Lawrence O. Ticknor, and
559   Jayne Belnap. Response of Biological Soil Crust Diazotrophs to Season Altered Summer Precipitation,
560   and Year-Round Increased Temperature in an Arid Grassland of the Colorado Plateau, USA. *Front.*
561   *Microbio.*, 3, 2012. doi: 10.3389/fmicb.2012.00358. URL http://dx.doi.org/10.3389/
562   fmicb.2012.00358.
563 CM Yeager, JL Kornosky, DC Housman, EE Grote, J Belnap, and CR Kuske. Diazotrophic community
564   structure and function in two successional stages of biological soil crusts from the Colorado Plateau
565   and Chihuahuan Desert. 70:973–83, 2004.
566 ND Youngblut and DH Buckley. Intra-genomic variation in G+C content and its implications for DNA
567   stable isotope probing (DNA-SIP). Aug 2014.

6 FIGURES AND LONG TABLES

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

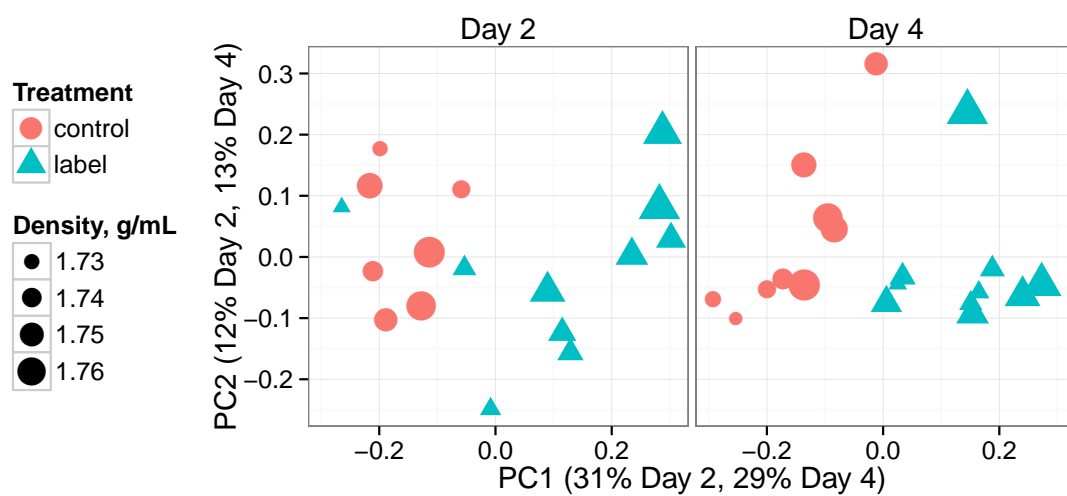**Figure 1.** Ordination of heavy gradient fraction sequence collections by Bray-Curtis distances.

Figure 2. Phylogenetic trees of OTUs passing sparsity threshold for selected phyla. *A)* Point denotes OTU is classified as a ^{15}N "responder". *B)* Heatmap of moderated $\log_2$ proportion mean ratios (labeled:control gradients) for each OTU at each incubation day. High values indicate ^{15}N incorporation. *C)* Presence/absence of OTUs (black indicates presence) in lichen, light, or dark environmental samples (Garcia-Pichel et al., 2013). *D)* Presence/absence of OTUs (black indicates presence) in crust and below crust samples (Steven et al., 2013).
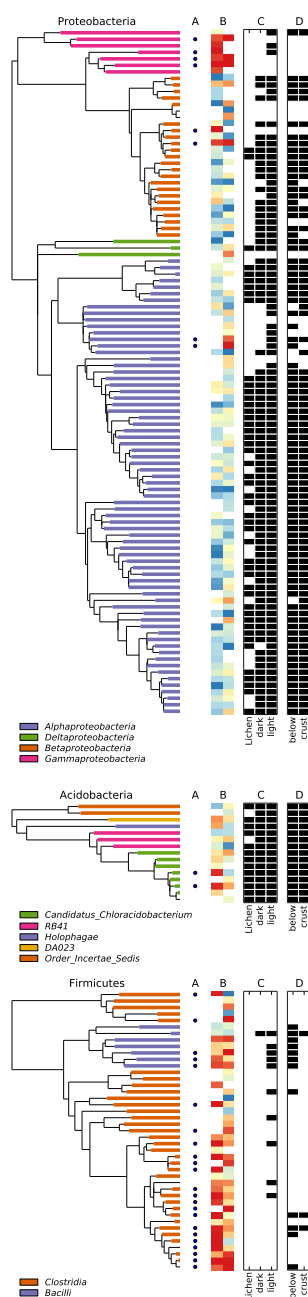
Figure 3. Moderated log$_2$ of proportion mean ratios for labeled versus control gradients (heavy fractions only, densities ¿1.725 g/mL). All OTUs found in at least 62.5% of heavy fractions at a specific incubation day are shown. Red color denotes a proportion mean ratio that has a corresponding adjusted 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.
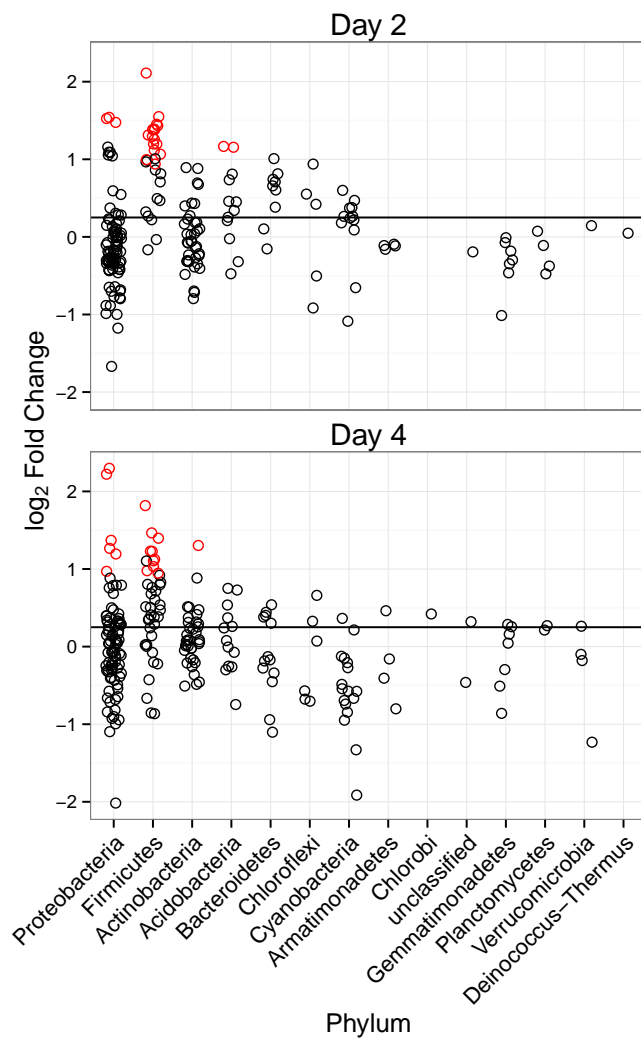
Figure 4. Relative abundance values in heavy fractions (density greater or equal to 1.725 g/mL) for the top 10 ^{15}N "responders" (putative diazotrophs, see results for selection criteria of top 10) at each incubation day. See Table 1 for BLAST results of top 10 responders against the LTP database (release 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.

7 SUPPLEMENTAL FIGURES

Figure S1. Distribution of sequences into top 9 phyla (phyla ranked by sum of all sequence annotations).

Figure S2. 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 S3. 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 S4. 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)
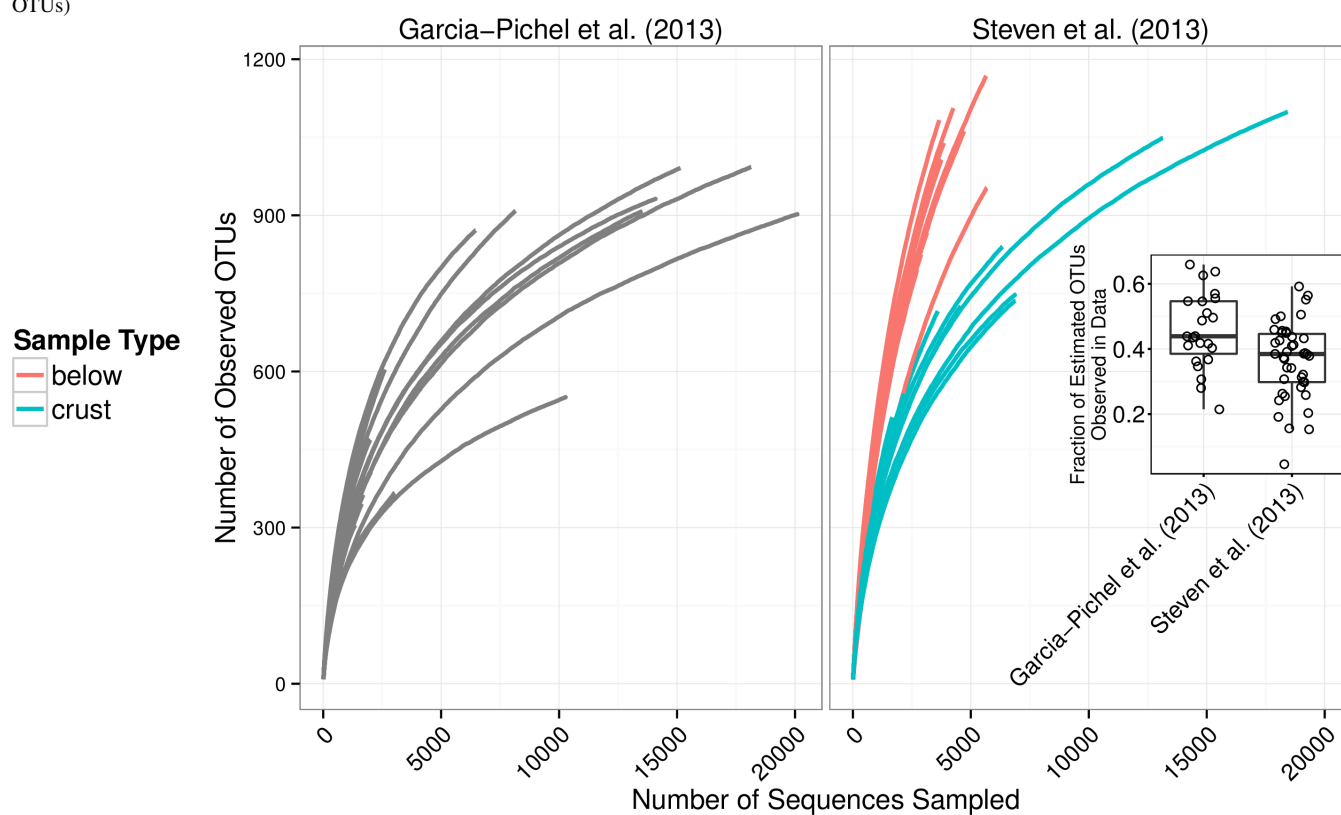
Figure S5. Counts of "responder" OTU occurrences in samples from Steven et al. (2013) and Garcia-Pichel et al. (2013). Steven et al. (2013) collected BSC samples (25 samples total) and samples from soil beneath BSC (17 samples total, "below" column in figure). Garcia-Pichel et al. (2013) collected samples from "dark" (9 samples total) and "light" (12 samples total) crusts in addition to "lichen" (2 samples total) dominated crusts.