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 communitites 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_2-fixation see Belnap (2003)). N_2-fixation represents the dominant source of new ecosystem N in more than 80% of BSC from diverse sites across North America, Africa, and Australia (Evans and Belnap, 1999), while atmospheric N deposition was a dominant source of N in only a minority of sites. 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 could alter BSC microbial community structure/membership and therefore it is possible that there will also be changes in diazotroph diversity and N_2-fixation and that these changes can alter the BSC N-budget.

29   BSC N_2-fixation rate studies (typically employing the acetylene reduction assay (ARA)) have explored
30   BSC diazotroph activity across various ecological gradients. Reported BSC N_2-fixation rates vary
31   significantly across samples and studies (Evans and Lange, 2001). The reasons for inter-site and inter-
32   study variability are complex and likely include the spatial heterogeneity of BSC (Evans and Lange, 2001)
33   and the impact of recent environmental change on N_2-fixation rates (see Belnap (2001) for discussion).
34   Moreover, the ARA assay is subject to methodological artifacts that can complicate making robust
35   comparisons across sample types that differ in physical and biological characteristics (see Belnap (2001)
36   for review). Nonetheless, N_2-fixation rates are consistently higher in mature BSC than in young, early
37   successional BSC (Belnap, 2002; Yeager et al., 2004). This difference may be due to the proliferation
38   of heterocystous *Cyanobacteria* in older mats and is consistent with the theory that heterocystous
39   *Cyanobacteria* provide the main source of fixed-N in BSC. Alternatively, the N_2-fixation rate differences
40   between young and old BSC might be attributable to methodological artifacts. For instance, Johnson et al.
41   (2005) show that N_2-fixation in mature mats is maximal at the crust surface (coincident with heterocystous
42   cyanobacteria) while it is maximal below the crust surface in early successional BSC. Diffusional
43   limitation can potentially cause ARA to underestimate N_2-fixation which occurs below the crust surface
44   and as a result ARA may systematically underestimate rates of N_2-fixation in early successional BSC.
45   Diffusion would not be an issue when measuring N_2-fixation rates in mature crust as nitrogenase activity
46   peaks near the surface. Differences of N_2 fixation rates between developing and mature BSC were not
47   statistically significant when aerial rates were estimated by integrating across ARA performed on thin
48   (1-3mm) slices across a BSC depth profile Johnson et al. (2005).

49   Molecular studies of BSC microbial diversity include explorations of the BSC microbial community
50   vertical profile (Garcia-Pichel et al., 2003a), BSC *nifH* gene content surveys (e.g. Yeager et al. (2004),
51   Yeager et al. (2012), Yeager et al. (2006) and Steppe et al. (1996)), and next-generation-sequencing
52   (NGS) enabled studies of BSC SSU rRNA gene content across wide geographic ranges (Garcia-Pichel
53   et al., 2013; Steven et al., 2013). Early successional BSC are often described as "light" in appearance
54   relative to "dark" mature BSC (Belnap, 2002; Yeager et al., 2004). Mature BSC possess greater numbers
55   of heterocystous *Cyanobacteria* (i.e *Scytonema*, *Spirirestis*, and *Nostoc* (Yeager et al., 2006, 2012))
56   than developing BSC but both young and old BSC are dominated by non-heterocystous *Cyanobacteria*
57   (*Microcoleus vaginatus* or *M. steenstrupii*) (Yeager et al., 2004; Garcia-Pichel et al., 2013). Heterocystous
58   *Cyanobacteria* are the numerically dominant BSC diazotrophs in *nifH* clone libraries (Yeager et al., 2006,
59   2004, 2012). Eighty-nine perent of 693 *nifH* sequences derived from Colorado Plateau and New Mexico
60   BSC samples as heterocystous cyanobacterial (non-cyanobacterial *nifH* sequences were largely attributed
61   to alpha- and beta- *proteobacteria*) Yeager et al. (2006). However, an early survey of Colorado Plateau
62   BSC *nifH* diversity recovered *nifH* genes related to *Gammaproteobacteria* as well as a clade that included
63   *nifH* genes from the anaerobes *Clostridium pssteurianum*, *Desulfovibrio gigas* and *Chromatium buderi*,

64   The influence of microbial community membership and structure on BSC N_2-fixation is an ongoing
65   research question (Belnap, 2013). While the presence/abundance of heterocystous *Cyanobacteria* has
66   been proposed as the mechanism behind increased N_2-fixation in mature BSC, it is unclear if mature BSC
67   actually fix more N than early successional BSC (see Johnson et al. (2005)). More studies are necessary
68   to elucidate the microbial membership influence on BSC N_2-fixation and to determine if heterocystous
69   *Cyanobacteria* are the only keystone diazotrophs. The first step in defining structure function relationships
70   with respect to N_2-fixation is a full accounting of BSC diazotrophs. Towards this end we conducted $^{15}N_2$
71   DNA stable isotope probing (DNA-SIP) experiments with light, developing Colorado Plateau BSC. DNA-
72   SIP with $^{15}N_2$ has not been attempted with BSC. DNA-SIP provides an accounting of *active* diazotrophs
73   whereas *nifH* clone libraries account for microbes with the genomic potential for N_2-fixation. Further, we
74   investiage the distribution of these active diazotrophs through collections of SSU rRNA gene sequences
75   from BSC NGS microbial diversity surveys over a range of spatial scales and soil types (Garcia-Pichel
76   et al., 2013; Steven et al., 2013).

3 RESULTS

3.1 ORDINATION OF CSCL GRADIENT FRACTION SSU RRNA SEQUENCE COLLECTIONS SHOWS HEAVY FRACTIONS FROM CONTROL AND LABELED CSCL GRADIENTS ARE DIFFERENT

77 BSC were incubated for 4 days in the presence or absence of $^{15}N_2$ and DNA was extracted for DNA-SIP at
78 2 and 4 days. Fractionation of CsCl gradients permitted separation of DNA on the basis of buoyant density.
79 Ordination of Bray-Curtis (Bray and Curtis, 1957) distances between SSU-rRNA amplicon sequence
80 collections from gradient fractions reveals that labeled gradient fraction (i.e. gradient fractions of DNA
81 from $^{15}N_2$ incubations) sequence collections diverge from control (i.e. DNA from incubations without
82 $^{15}N_2$) at the "heavy" of the CsCl gradients (Figure 1 and Figure S2). Differences among label/control
83 groups with heavy fractions are statistically significant by the Adonis test (p-value: 0.001, r^2: 0.18)
84 (Anderson, 2001).

3.2 OTUS RESPONSIVE TO $^{15}N_2$ ARE PRIMARILY *PROTEOBACTERIA* AND *CLOSTRIDIACEAE*

85 A statistically significant increase in OTU abundance in heavy fractions of $^{15}N_2$ labeled samples relative
86 to corresponding control fractions provides evidence for OTUs that have incorporated ^{15}N into their
87 DNA. Specifically, we compared OTU proportion means between labeled and control samples from
88 heavy gradient fractions using statistics developed to find differentially expressed genes with RNASeq
89 data (McMurdie and Holmes, 2014; Love et al., 2014). OTUs that incorporated ^{15}N into DNA and
90 increased in bouyant density were identified by rejecting the the null hypothesis that the labeled versus
91 control proportion mean ratio for an OTU was below a chosen threshold (see methods). p-values were
92 adjusted by the BH method (Benjamini and Hochberg, 1995) and we used a false discovery rate (FDR)
93 cutoff of 0.10 (typical FDR threshold in gene expression data analysis). A total of 2,127 and 2,160
94 OTUs were detected in days 2 and 4, respectively, and interrogated for evidence of $^{15}N_2$-labelling. Of
95 these OTU, only 208 and 233, respectively, passed a sparsity threshold we applied as an independent
96 filtering step to pre-screen out OTUs not likely to produce significant p-values (see Love et al. (2014)
97 for discussion of independent filtering). Of OTUs passing sparsity criteria 38 were found to be enriched
98 significantly in "heavy" fractions relative to control. These OTUs likely incorporated ^{15}N into DNA ($^{15}N_2$
99 "responders"). Of these 38, 26 are annotated as *Firmicutes*, 9 as *Proteobacteria*, 2 as *Acidobacteria* and
100 1 as *Actinobacteria* (Figure 3, Figure 2). If the responder OTUs are ranked by descending, moderated
101 proportion mean labeled:control ratios, the top 10 ratios (i.e. the 10 OTUs that were most enriched in
102 the labeled gradients considering only heavy fractions) are either *Firmicutes* (6 OTUs) or *Proteobacteria*
103 (4 OTUs) (Figure 4). Centroid sequences of strongly responding *Proteobacteria* OTUs all share high
104 sequence identity (>98.48%, Table 1) with cultivars from genera known to possess diazotrophs including
105 *Klebsiella*, *Shigella*, *Acinetobacter*, and *Ideonella*. None of the *Firmicutes* OTUs in the top 10 responders
106 share greater than 97% sequence identity with sequences in the LTP database (release 115) (see Table 1).
107 OTUs that passed the sparsity threshold but were not classified as ^{15}N-responsive were subsequently
108 tested against the null hypothesis that the OTU proportion mean ratio was above the selected threshold.
109 Rejecting the second null would indicate an OTU did *not* incorporate ^{15}N into biomass. There were 58
110 and 70 "non-responders" at days 2 and 4, respectively. OTUs that did not pass sparsity or could not be
111 classified as either a responder or non-responder are simply ambiguous with respect to ^{15}N labelling.

3.3 ^{15}N-RESPONSIVE OTUS IN ENVIRONMENTAL SAMPLES

112 Five of the 6 *Firmicutes* with the strongest response to ^{15}N-labelling (Table X) belong in the *Clostridiacea*.
113 We only observed one of these strongly responding *Clostridiaceae* in the data presented by Garcia-Pichel
114 et al. (2013), "OTU.108" (closest BLAST hit in LTP Release 115 – *Caloramotor proteoclasticus*, BLAST

115 %ID 96.94, Accession X90488). OTU.108 was found in two samples both characterized as "light" crust.
116 One other *Clostridiaceae* OTU with a proportion mean ratio (labeled:control) p-value less than 0.10 but
117 outside the top 10 responders was found in the Garcia-Pichel et al. (2013) data (a "light" crust sample)
118 (Figure 2). None of the strongly responding *Clostridiacea* were found in the sequences provided by Steven
119 et al. (2013). *Clostridiaceae* ^{15}N-responder OTU are not closely related to cultivars. (Table 1, Figure 5).
120 One of the proteobacterial OTUs with the strongest [15]-N response (Table X) was found in Garcia-Pichel
121 et al. (2013) (closest BLAST hit in LTP Release 115, BLAST %ID 100, Accession ZD3440, *Acinetobacter*
122 *johnsonii*). None of the *strongly responding Protebacteria* OTUs were found in the Steven et al. (2013)
123 sequences. Responder OTUs were found in Steven et al. (2013) samples 133 times. 83 were in "below
124 crust" samples, 50 in crust samples (see Figure 2). Two ^{15}N-responsive OTUs were found in an extensive
125 number of environmental samples (61 of 65 samples from the combined data sets of Garcia-Pichel et al.
126 (2013) and Steven et al. (2013)). Both OTUs were annotated as *Acidobacteria* but shared little sequence
127 identity to any cultivar SSU rRNA gene sequences in the LTP (Release 115), with best LTP BLAST hits
128 of 81.91 and 81.32% identity. Additionally, the ^{15}N-response for each OTU was weak relative to other
129 putative responders (3. Of the remaining 36 stable isotope responder OTUs, only 14 were observed in the
130 environmental data (Figure 2, Figure S5).

3.4 COMPARING SEQUENCE COLLECTIONS AT "STUDY"-LEVEL

131 We compared the sequences determined in this study to two previous surveys of SSU rRNA amplicons
132 from BSC communities: the Garcia-Pichel et al. (2013) and Steven et al. (2013) study. There were 3,079
133 OTUs (209,354 total sequences after quality control) in the DNA-SIP data, 3,203 OTUs (129,033 total
134 sequences after quality control) in the Garcia-Pichel et al. (2013) study, and 2,481 OTUs (129,358 total
135 sequences after quality control) in the Steven et al. (2013) study. Of the 4,340 OTU centroids established
136 for this study (including sequences from Steven et al. (2013) and Garcia-Pichel et al. (2013)) 445 have
137 matches in the Living Tree Project (LTP) (a collection of 16S gene sequences for all sequenced type
138 strains (Yarza et al., 2008)) at greater or equal than 97% sequence identity (LTP version 115). That is, 445
139 of 4,340 OTUs (~10%) are closely related to cultivars. The DNA-SIP data set shares 56% OTUs with the
140 Steven et al. (2013) data and 46% of OTUs with the Garcia-Pichel et al. (2013) data (where total OTUs
141 are from the combined data for each pairwise comparison). The Steven et al. (2013) and Garcia-Pichel
142 et al. (2013) studies share 46% of OTUs. *Cyanobacteria* and *Proteobacteria* were the top two phylum-
143 level sequence annotations for all three studies of BSC. Only the DNA-SIP data had more *Proteobacteria*
144 annotations than *Cyanobacteria*. *Proteobacteria* represented the 29.8% of sequence annotations in DNA-
145 SIP data as opposed to 17.8% and 19.2% for the Garcia-Pichel et al. (2013) and Steven et al. (2013)
146 data, respectively. There is a stark contrast in the total percentage of sequences annotated as *Firmicutes*
147 between the raw environmental samples and the DNA-SIP data. *Firmicutes* represent only 0.21% and
148 0.23% of total phylum level sequence annotations in the Steven et al. (2013) and Garcia-Pichel et al.
149 (2013) studies, respectively (Figure S1). In the DNA-SIP sequence collection *Firmicutes* make up 19% of
150 phylum level sequence annotations. Also in sharp contrast for the DNA-SIP versus environmental data is
151 the number of putative heterocystous *Cyanobacteria* sequences. Only 0.29% of *Cyanobacteria* sequences
152 in the DNA-SIP data are annotated as belonging to "Subsection IV" which is the heterocystous order of
153 *Cyanobacteria* in the Silva taxonomic nomenclature (Pruesse et al., 2007). In the Steven et al. (2013) and
154 Garcia-Pichel et al. (2013) studies 15% and 23%, respectively, of *Cyanobacteria* sequences are annotated
155 as belonging to "Subsection IV".

4   DISCUSSION

4.1   ^{15}N-RESPONSIVE OTUS FROM EARLY SUCESSIONAL BSC ARE PRIMARILY *PROTEOBACTERIA* AND *CLOSTRIDIACEAE*

156   BSC N-fixation has long been attributed to heterocystous *Cyanobacteria* and molecular surveys of BSC
157   *nifH* gene content have been consistent with this hypothesis finding cyanobacterial *nifH* types to be
158   numerically dominant in *nifH* gene libraries from BSC (Yeager et al., 2006, 2004, 2012). However, $^{15}N_2$
159   DNA-SIP revealed non-cyanobacterial microbes fixed N_2 in early successional BSC samples. After 2
160   and 4 incubation days in the presence or absenece of $^{15}N_2$ in microcosm headspace, DNA from early
161   successional BSC samples was collected and separated by bouyant density in CsCl density gradients.
162   Heavy CsCl gradient fractions from gradients with ^{15}N-labelled DNA were different in phylogenetic
163   membership/structure than heavy fractions with unlabeled DNA (Figure X). Further, heavy gradient
164   fractions clustered by DNA type (labeled or unlabeled) (Figure X). Therefore, headspace $^{15}N_2$ in early
165   successional BSC microcosms was incorporated into biomass, and, the specific OTUs that incorporated
166   ^{15}N (from $^{15}N_2$) into biomass would be enriched in labeled gradient heavy fractions. *Proteobacteria*
167   and *Clostridiaceae* represented most ^{15}N-responsive OTUs as revealed by a robust statistical framework
168   for quantifying and evaluating differential OTU abundance in microbiome studies (CITE McMurdie),
169   Additionally, *Proteobacteria* and *Clostridiaceae* represented the most strongly ^{15}N-responsive OTUs.
170   SIP places focus upon organisms based on isotope incorporation and has the ability to detect activity by
171   low abundance members of the community. DNA from OTUs that incopororate ^{15}N into their biomass
172   moves towards the heavy end of the CsCl gradient and therefore OTUs in "labeled" DNA are enriched
173   in the full data pool relative to bulk DNA. Phylum-level taxonomic annotations of ^{15}N-responsive OTUs
174   (i.e. *Firmicutes* and *Proteobacteria*) are enriched in the DNA-SIP data relative to environmental data
175   (Figure S1).

176   It is possible that PCR-driven molecular surveys of *nifH* gene content have been biased. We propose
177   three mechanisms that could bias *nifH* clone libraries against heteroptophic diazotrophs. First, polyploidy
178   in *Cyanobacteria* (Griese et al., 2011) would inflate the representation cyanobacteria in community DNA
179   beyond their cell number representation. Second, *nifH* PCR primers could be biased against heterotrophic
180   diazotrophs. In general the nifH PCR primers used by Yeager et al. (2006, 2004, 2012) ("19F" and
181   "nifH3") for the first round of nested PCR have broad specificity and display at least 86% *in silico*
182   coverage for *Proteobacteria, Cyanobacteria* and "Cluster III" (which includes clostridial *nifH*) reference
183   *nifH* sequences (Gaby and Buckley, 2012). In the second round of the nested PCR protocol (Yeager
184   et al., 2006, 2004, 2012), primer "nifH11" is' biased against "Cluster III" (50% *in silico* coverage of
185   reference *nifH* sequences), *Proteobacteria* (79% coverage) and *Cyanobacteria* (67% coverage), and,
186   primer "nifH22" matches *Proteobacteria*, *Cyanobacteria* and "Cluster III" reference sequences poorly
187   (16%, 23% and 21% *in silico* coverage, respectively) (Gaby and Buckley, 2012). Unfortunately, it is
188   difficult to assess or quantify this bias (in either direction) without knowing the *nifH* gene content *de
189   novo*. Third, heterocysts (the specialized N-fixing cells along the trichome of filamentous heterocystous
190   *Cyanobacteria* such as *Nostoc* and *Scytonema*) may be overrepresented with respect to non-cyanobacterial
191   diazotrophs because heterocysts make up a fraction of cells along a trichome and even the non-heterocyst
192   cells in the trichome will possess the *nifH* gene. As a result of polyploidy and the frequency of heterocysts
193   in a cyanobacterial filament, the ratio of cyanobacterial to heterotroph *nifH* gene copies may be on the
194   order of 10^2-10^3 higher than the ratio of heterocysts to heterotrophic diazotroph cells. Regardless, our
195   results suggest that BSC N-fixation may include a significant non-cyanobacterial component that requires
196   further assessment across a more comprehensive sampling of BSC types.

197   We did not observe evidence for N-fixation by heterocystous *Cyanobacteria* in the early successional
198   BSC samples used in this study. One possible explanation for our results is that the early successional
199   BSC samples used in this study possessed too few heterocystous *Cyanobacteria* to statistically evaluate
200   their ^{15}N-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%) 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 would create C-rich environment after wetting. Diazotrophs would be uniquely suited to respond quickly in high C:N conditions.**

Although too undersampled in the environmental data sets to reach statistical conclusions, ^{15}N-responsive OTUs were found more often in sub-crust or "light" BSC samples (Figure 2 and Figure S5). This result generates some hypotheses that are counter to prior discussions regarding BSC diazotroph temporal dynamics. 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.

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). Even next-generation sequencing efforts of BSC 16S rRNA genes have only shallowly sampled the full diversity of BSC microbes.

4.2 CONCLUSION

Heterocystous *Cyanobacteria* are key contributors to the BSC N-budget, but, the ^{15}N-responsive OTUs found in this study and the *nifH* gene sequences from Steppe et al. (1996) in addition to the N-fixation rate data presented by Johnson et al. (2005) suggest there may be significant non-cyanobacterial BSC

247   diazotrophs specifically within the *Clostrideaceae* and *Proteobacteria*. It seems clear that heterocystous
248   *Cyanobacteria* increase in abundance with BSC age (Yeager et al., 2004). It is less clear if this transition
249   marks the emergence of diazotrophy versus a re-structuring of the BSC diazotroph community from
250   one dominated by *Firmicutes* and *Proteobacteria* to one predominantly heterocystous *Cyanobacteria*.
251   DNA-SIP is a valuable tool in the molecular microbial ecologist's toolbox for identifying members of
252   microbial community functional guilds (Neufeld et al., 2007). PCR-based surveys of diagnostic marker
253   genes and DNA-SIP are both used to connect microbial phylogenetic types to microbial activities, but
254   they occupy a non-overlapping set of strengths and weaknesses. DNA-SIP does not focus on a specific
255   diagnostic marker but does identify *active* players in the studied process (i.e. N-fixation). Combined these
256   tools can powerfully reveal connections between ecosystem membership/structure and function. Here we
257   supplement previous surveys of BSC *nifH* diversity, a diagnostic marker PCR-driven approach, with $^{15}N_2$
258   DNA-SIP, While we do not confirm previous results, we expand knowledge of BSC diazotroph diversity.
259   Predicting BSC N-fixation with respect to climate change, althered precipitation regimes and physical
260   disturbance requires a careful accounting of diazotrophs including non-cyanobacterial types.

5   MATERIALS AND METHODS

5.1   BSC SAMPLING AND INCUBATION CONDITIONS

261   Light crust samples (37.5 cm^2, average mass 35 g) were incubated in sealed chambers under controlled
262   atmosphere and in the light for 4 days. Crusts were dry prior to time zero and were wetted at initiation of
263   experiment. Treatments included control air (unenriched headspace) and enriched air (>98% atom $^{15}N_2$)
264   headspace. Samples were taken at 2 days and 4 days incubation. Acetylene reduction rates were measured
265   daily. DNA was extracted from 1 g of crust. Samples were taken from Green Butte, Arizona as previously
266   described (site CP3, Beraldi-Campesi et al. (2009)). All samples were from light crusts as described by
267   Johnson et al. (2005).

5.2   DNA EXTRACTION

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

5.3   DNA-SIP

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

5.4   PCR, LIBRARY NORMALIZATION AND DNA SEQUENCING

282   Barcoded PCR of bacterial and archaeal 16S rRNA genes, in preparation for 454 Pyrosequencing, was
283   carried out using primer set 515F/806R (Walters et al., 2011) (primers purchased from Integrated DNA

284 Technologies). The primer 806R contained an 8 bp barcode sequence, a "TC" linker, and a Roche 454
285 B sequencing adaptor, while the primer 515F contained the Roche 454 A sequencing adapter. Each
286 25 μL reaction contained 1x PCR Gold Buffer (Roche), 2.5 mM MgCl$_2$, 200 μM of each of the four
287 dNTPs (Promega), 0.5 mg/mL BSA (New England Biolabs), 0.3 μM of each primers, 1.25 U of Amplitaq
288 Gold (Roche), and 8 μL of template. Each sample was amplified in triplicate. Thermal cycling occurred
289 with an initial denaturation step of 5 minutes at 95C, followed by 40 cycles of amplification (20s at
290 95C, 20s at 53C, 30s at 72C), and a final extension step of 5 min at 72C. Triplicate amplicons were
291 pooled and purified using Agencourt AMPure PCR purification beads, following manufacturers protocol.
292 Once cleaned, amplicons were quantified using PicoGreen nucleic acid quantification dyes (Molecular
293 Probes) and pooled together in equimolar amounts. Samples were sent to the Environmental Genomics
294 Core Facility at the University of South Carolina (now Selah Genomics) to be run on a Roche FLX 454
295 pyrosequencing machine.

5.5 DATA ANALYSIS

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

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

298 *5.5.1 Sequence quality control*   Sequences were initially screened by maximum expected errors at a
299 specific read length threshold (Edgar, 2013) which has been shown to be as effective as denoising 454
300 reads with respect to removing pyrosequencing errors. Specifically, reads were first truncated to 230
301 nucleotides (nt) (all reads shorter than 230 nt were discarded) and any read that exceeded a maximum
302 expected error threshold of 1.0 was removed. After truncation and max expected error trimming, 91% of
303 original reads remained. The first 30 nt representing the forward primer and barcode on high quality,
304 truncated reads were trimmed. Remaining reads were taxonomically annotated using the "UClust"
305 taxonomic annotation framework in the QIIME software package (Caporaso et al., 2010; Edgar, 2010)
306 with cluster seeds from Silva SSU rRNA database (Pruesse et al., 2007) 97% sequence identity OTUs as
307 reference (release 111Ref). Reads annotated as "Chloroplast", "Eukaryota", "Archaea", "Unassigned" or
308 "mitochondria" were culled from the dataset. Finally, reads were aligned to the Silva reference alignment
309 provided by the Mothur software package (Schloss et al., 2009) using the Mothur NAST aligner (DeSantis
310 et al., 2006). All reads that did not appear to align to the expected amplicon region of the SSU rRNA gene
311 were discarded. Quality control parameters removed 34,716 of 258,763 raw reads.

312 *5.5.2 Sequence clustering*   Sequences were distributed into OTUs using the UParse methodology
313 (Edgar, 2013). Specifically, cluster seeds were identified using USearch with a collection of non-redundant
314 reads sorted by count as input. The sequence identity threshold for establishing a new OTU centroid
315 was 97%. After initial cluster centroid selection, select 16S rRNA gene sequences trimmed to the same
316 alignment positions as the other centroids from Yeager et al. (2006) were added to the centroid collection.
317 Specifically, Yeager et al. (2006) Colorado Plateau or Moab, Utah sequences were added which included
318 the 16S rRNA gene sequences for *Calothrix* MCC-3A (accession DQ531700.1), *Nostoc commune* MCT-1
319 (accession DQ531903), *Nostoc commune* MFG-1 (accession DQ531699.1), *Scytonema hyalinum* DC-A
320 (accession DQ531701.1), *Scytonema hyalinum* FGP-7A (accession DQ531697.1), *Spirirestis rafaelensis*
321 LQ-10 (accession DQ531696.1). Centroid sequences that matched selected Yeager et al. (2006) sequences
322 with greater than to 97% sequence identity were subsequently removed from the centroid collection. With
323 USearch/UParse, potential chimeras are identified during OTU centroid selection and are not allowed to
324 become cluster centroids effectively removing chimeras from the read pool. All quality controlled reads
325 were then mapped to cluster centroids at an identity threshold of 97% again using USearch. 95.6% of
326 quality controlled reads could be mapped to centroids. Unmapped reads do not count towards sample
327 counts and are essentially removed from downstream analyses. The USearch software version for cluster
328 generation was 7.0.1090.

329 *5.5.3 Merging data from this study, Garcia-Pichel et al. (2013), and Steven et al. (2013)* As only
330 sequences without corresponding quality scores were publicly available from Garcia-Pichel et al. (2013)
331 and Steven et al. (2013), these data sets were only quality screened by determining if they covered the
332 expected region of the 16S rRNA gene (described above). All data (this study, Garcia-Pichel et al. (2013)
333 and Steven et al. (2013)) were included as input to USearch for OTU centroid selection and subsequent
334 mapping to OTU centroids.

335 *5.5.4 Phylogenetic tree* The alignment for the "*Clostridiaceae*" phylogeny was created using SSU-
336 Align which is based on Infernal (Nawrocki and Eddy, 2013; Nawrocki et al., 2009). Columns in
337 the alignment that were not included in the SSU-Align covariance models or were aligned with poor
338 confidence (less than 95% of characters in a position had posterior probability alignment scores of
339 at least 95%) were masked for phylogenetic reconstruction. Additionally, the alignment was trimmed
340 to coordinates such that all sequences in the alignment began and ended at the same positions. The
341 "*Clostridiaceae*" tree included all top BLAST hits (parameters below) for ^{15}N *Clostridiaceae* responders
342 in the Living Tree Project database (Yarza et al., 2008) in addition to BLAST hits within a sequence
343 identity threshold of 97% to ^{15}N responders from the Silva SSURef_NR SSU rRNA database (Pruesse
344 et al., 2007). Only one SSURef_NR115 hit per study per OTU ("study" was determined by "title" field)
345 was selected for the tree. FastTree (Price et al., 2010) was used to build the tree and support values are
346 SH-like scores reported by FastTree.

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

352 *5.5.5 BLAST searches* BLAST searches were done with the "blastn" program from BLAST+ toolkit
353 (Camacho et al., 2009) version 2.2.29+. Default parameters were always employed and the BioPython
354 (Cock et al., 2009) BLAST+ wrapper was used to invoke the blastn program. Pandas (McKinney, 2012)
355 and dplyr (Wickham and Francois, 2014) were used to parse and munge BLAST output tables.

356 *5.5.6 Identifying OTUs that incorporated ^{15}N into their DNA* SIP is a culture-independent approach
357 towards defining identity-function connections in microbial communities (Buckley, 2011; Neufeld et al.,
358 2007). Microbes incubated in the presence of ^{13}C or ^{15}N labeled substrates can incorporate the stable
359 heavy isotope into biomass if they participate in the substrate's transformation. Stable isotope labeled
360 nucleic acids can then be separated from unlabeled by buoyant density in a CsCl gradient. As the buoyant
361 density of a macromolecule is dependent on many factors in addition to stable isotope incorporation
362 (e.g. GC-content in nucleic acids (Youngblut and Buckley, 2014)), labeled nucleic acids from one
363 microbial population may have the same buoyant density of unlabeled nucleic acids from another (i.e.
364 each population's nucleic acids would be found at the same point along a density gradient although
365 only one population's nucleic acids are labeled). Therefore it is imperative to compare density gradients
366 with nucleic acids from heavy stable isotope incubations to gradients from "control" incubations where
367 everything mimics the experimental conditions except that unlabeled substrates are used (and all DNA
368 would be unlabeled). By contrasting "heavy" density gradient fractions in experimental density gradients
369 (hereafter referred to as "labeled" gradients) against heavy fractions in control gradients, the identities of
370 microbes with labeled nucleic acids can be determined

371   We used an RNA-Seq differential expression statistical framework (Love et al., 2014) to find OTUs
372 enriched in heavy fractions of labeled gradients relative to corresponding density fractions in control
373 gradients (for review of RNA-Seq differential expression statistics applied to microbiome OTU count data
374 see McMurdie and Holmes (2014)). We use the term differential abundance (coined by McMurdie and

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

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

399 *5.5.7 Ordination* Principal coordinate ordinations depict the relationship between samples at each time
400 point (day 2 and 4). Bray-Curtis distances were used as the sample distance metric for ordination. The
401 Phyloseq (McMurdie and Holmes, 2014) wrapper for Vegan (Oksanen et al., 2013) (both R packages) was
402 used to compute sample values along principal coordinate axes. GGplot2 (Wickham, 2009) was used to
403 display sample points along the first and second principal axes. Adonis tests Anderson (2001) were done
404 with default number of permutations (1000).

5.6 RICHNESS ANALYSES

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

REFERENCES

408 Marti J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*,
409   26(1):32–46, Feb 2001. doi: 10.1111/j.1442-9993.2001.01070.pp.x. URL http://dx.doi.org/
410   10.1111/j.1442-9993.2001.01070.pp.x.
411 J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In
412   *Biological Soil Crusts: Structure Function, and Management*, pages 241–261. Springer Science +
413   Business Media, 2001. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.
414   1007/978-3-642-56475-8_19.
415 J. Belnap. Factors Influencing Nitrogen Fixation and Nitrogen Release in Biological Soil Crusts. In
416   Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*,

volume 150 of *Ecological Studies*, pages 241–261. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-43757-4. doi: 10.1007/978-3-642-56475-8_19. URL http://dx.doi.org/10.1007/978-3-642-56475-8_19.

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

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

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

Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57 (1):289–300, 1995. ISSN 00359246. doi: 10.2307/2346101. URL http://dx.doi.org/10.2307/2346101.

H. Beraldi-Campesi, H. E. Hartnett, A. Anbar, G. W. Gordon, and F. Garcia-Pichel. Effect of biological soil crusts on soil elemental concentrations: implications for biogeochemistry and as traceable biosignatures of ancient life on land. *Geobiology*, 7(3):348–359, jun 2009. doi: 10.1111/j.1472-4669.2009.00204.x. URL http://dx.doi.org/10.1111/j.1472-4669.2009.00204.x.

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

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

DH Buckley, V Huangyutitham, SF Hsu, and TA Nelson. Stable isotope probing with 15N2 reveals novel noncultivated diazotrophs in soil. *Appl Environ Microbiol*, 73:3196–204.

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 +

470   Business Media, 2001. doi: 10.1007/978-3-642-56475-8_20. URL http://dx.doi.org/10.
471   1007/978-3-642-56475-8_20.
472 John Christian Gaby and Daniel H. Buckley. A Comprehensive Evaluation of {PCR} Primers to Amplify
473   the {nifH} Gene of Nitrogenase. {PLoS} {ONE}, 7(7):e42149, jul 2012. doi: 10.1371/journal.pone.
474   0042149. URL http://dx.doi.org/10.1371/journal.pone.0042149.
475 F. Garcia-Pichel, S. L. Johnson, D. Youngkin, and J. Belnap. Small-Scale Vertical Distribution of Bacterial
476   Biomass and Diversity in Biological Soil Crusts from Arid Lands in the Colorado Plateau. *Microbial*
477   *Ecology*, 46(3):312–321, Nov 2003a. doi: 10.1007/s00248-003-1004-0. URL http://dx.doi.
478   org/10.1007/s00248-003-1004-0.
479 F. Garcia-Pichel, V. Loza, Y. Marusenko, P. Mateo, and R. M. Potrafka. Temperature Drives the
480   Continental-Scale Distribution of Key Microbes in Topsoil Communities. *Science*, 340(6140):
481   1574–1577, Jun 2013. doi: 10.1126/science.1236404. URL http://dx.doi.org/10.1126/
482   science.1236404.
483 Ferran Garcia-Pichel, Jayne Belnap, Susanne Neuer, and Ferdinand Schanz. Estimates of global
484   cyanobacterial biomass and its distribution. *Algological Studies*, 109(1):213–227, 2003b.
485 Marco Griese, Christian Lange, and Jrg Soppa. Ploidy in cyanobacteria. *FEMS Microbiology Letters*, 323
486   (2):124–131, sep 2011. doi: 10.1111/j.1574-6968.2011.02368.x. URL http://dx.doi.org/10.
487   1111/j.1574-6968.2011.02368.x.
488 SL Johnson, CR Budinoff, J Belnap, and F Garcia-Pichel. Relevance of ammonium oxidation within
489   biological soil crust communities. 7:1–12, 2005.
490 A. Karnieli, R.F. Kokaly, N.E. West, and R.N. Clark. Remote Sensing of Biological Soil Crusts. In
491   Jayne Belnap and OttoL. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*,
492   volume 150 of *Ecological Studies*, pages 431–455. Springer Berlin Heidelberg, 2003. ISBN 978-
493   3-540-43757-4. doi: 10.1007/978-3-642-56475-8_31. URL http://dx.doi.org/10.1007/
494   978-3-642-56475-8_31.
495 Rob Knight, Peter Maxwell, Amanda Birmingham, Jason Carnes, J Gregory Caporaso, Brett C Easton,
496   Michael Eaton, Micah Hamady, Helen Lindsay, Zongzhi Liu, Catherine Lozupone, Daniel McDonald,
497   Michael Robeson, Raymond Sammut, Sandra Smit, Matthew J Wakefield, Jeremy Widmann, Shandy
498   Wikman, Stephanie Wilson, Hua Ying, and Gavin A Huttley. {PyCogent}: a toolkit for making sense
499   from sequence. *Genome Biol*, 8(8):R171, 2007. doi: 10.1186/gb-2007-8-8-r171. URL http://dx.
500   doi.org/10.1186/gb-2007-8-8-r171.
501 M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for {RNA}-
502   Seq data with {DESeq}2. Technical report, feb 2014. URL http://dx.doi.org/10.1101/
503   002832.
504 Frederick A Matsen, Robin B Kodner, and E Virginia Armbrust. pplacer: linear time maximum-likelihood
505   and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*,
506   11(1):538, 2010. doi: 10.1186/1471-2105-11-538. URL http://dx.doi.org/10.1186/
507   1471-2105-11-538.
508 Wes McKinney. pandas: Python Data Analysis Library. Online, 2012. URL http://pandas.
509   pydata.org/.
510 PJ McMurdie and S Holmes. Waste not, want not: why rarefying microbiome data is inadmissible. 10:
511   e1003531, 2014.
512 EP Nawrocki and SR Eddy. Infernal 1.1: 100-fold faster RNA homology searches. 29:2933–5, Nov 2013.
513 EP Nawrocki, DL Kolbe, and SR Eddy. Infernal 1.0: inference of RNA alignments. 25:1335–7, May
514   2009.
515 JD Neufeld, J Vohra, MG Dumont, T Lueders, M Manefield, MW Friedrich, and JC Murrell. DNA
516   stable-isotope probing. 2:860–6, 2007.
517 Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara,
518   Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community*
519   *Ecology Package*, 2013. URL http://CRAN.R-project.org/package=vegan. R package
520   version 2.0-10.
521 MN Price, PS Dehal, and AP Arkin. FastTree 2–approximately maximum-likelihood trees for large
522   alignments. 5:e9490, Mar 2010.

523 E Pruesse, C Quast, K Knittel, BM Fuchs, W Ludwig, J Peplies, and FO Glckner. SILVA: a comprehensive
524   online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.
525   35:7188–96, 2007.
526 PD Schloss, SL Westcott, T Ryabin, JR Hall, M Hartmann, EB Hollister, RA Lesniewski, BB Oakley,
527   DH Parks, CJ Robinson, JW Sahl, B Stres, GG Thallinger, Horn DJ Van, and CF Weber.
528   Introducing mothur: open-source, platform-independent, community-supported software for describing
529   and comparing microbial communities. 75:7537–41, 2009.
530 T.F. Steppe, J.B. Olson, H.W. Paerl, R.W. Litaker, and J. Belnap. Consortial N2 fixation: a strategy
531   for meeting nitrogen requirements of marine and terrestrial cyanobacterial mats. *FEMS Microbiology*
532   *Ecology*, 21(3):149–156, Nov 1996. doi: 10.1111/j.1574-6941.1996.tb00342.x. URL http://dx.
533   doi.org/10.1111/j.1574-6941.1996.tb00342.x.
534 Blaire Steven, La Verne Gallegos-Graves, Jayne Belnap, and Cheryl R. Kuske. Dryland soil microbial
535   communities display spatial biogeographic patterns associated with soil depth and soil parent material.
536   *FEMS Microbiol Ecol*, 86(1):101–113, May 2013. doi: 10.1111/1574-6941.12143. URL http:
537   //dx.doi.org/10.1111/1574-6941.12143.
538 WA Walters, JG Caporaso, CL Lauber, D Berg-Lyons, N Fierer, and R Knight. PrimerProspector: de novo
539   design and taxonomic analysis of barcoded polymerase chain reaction primers. 27:1159–61, Apr 2011.
540 Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-
541   387-98140-6. URL http://had.co.nz/ggplot2/book.
542 Hadley Wickham and Romain Francois. *dplyr: dplyr: a grammar of data manipulation*, 2014. URL
543   http://CRAN.R-project.org/package=dplyr. R package version 0.2.
544 Pablo Yarza, Michael Richter, Jörg Peplies, Jean Euzeby, Rudolf Amann, Karl-Heinz Schleifer, Wolfgang
545   Ludwig, Frank Oliver Glöckner, and Ramon Rosselló-Móra. The All-Species Living Tree project: A
546   16S rRNA-based phylogenetic tree of all sequenced type strains. *Systematic and Applied Microbiology*,
547   31(4):241–250, Sep 2008. doi: 10.1016/j.syapm.2008.07.001. URL http://dx.doi.org/10.
548   1016/j.syapm.2008.07.001.
549 Chris M. Yeager, Jennifer L. Kornosky, Rachael E. Morgan, Elizabeth C. Cain, Ferran Garcia-Pichel,
550   David C. Housman, Jayne Belnap, and Cheryl R. Kuske. Three distinct clades of cultured heterocystous
551   cyanobacteria constitute the dominant N2-fixing members of biological soil crusts of the Colorado
552   Plateau USA. *FEMS Microbiology Ecology*, 60(1):85–97, 2006. doi: 10.1111/j.1574-6941.2006.00265.
553   x. URL http://dx.doi.org/10.1111/j.1574-6941.2006.00265.x.
554 Chris M. Yeager, Cheryl R. Kuske, Travis D. Carney, Shannon L. Johnson, Lawrence O. Ticknor, and
555   Jayne Belnap. Response of Biological Soil Crust Diazotrophs to Season Altered Summer Precipitation,
556   and Year-Round Increased Temperature in an Arid Grassland of the Colorado Plateau, USA. *Front.*
557   *Microbio.*, 3, 2012. doi: 10.3389/fmicb.2012.00358. URL http://dx.doi.org/10.3389/
558   fmicb.2012.00358.
559 CM Yeager, JL Kornosky, DC Housman, EE Grote, J Belnap, and CR Kuske. Diazotrophic community
560   structure and function in two successional stages of biological soil crusts from the Colorado Plateau
561   and Chihuahuan Desert. 70:973–83, 2004.
562 ND Youngblut and DH Buckley. Intra-genomic variation in G+C content and its implications for DNA
563   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 Names*	BLAST %ID
OTU.342	*Acinetobacter johnsonii*	100.0
OTU.263	*Azospirillum picis*	98.48
OTU.137	*Azospirillum rugosum, A. lipoferum*	98.98
OTU.3	*Bacillus azotoformans*	100.0
OTU.140	*Bacillus korlensis, B. beringensis*	100.0
OTU.108	*Caloramator proteoclasticus*	96.94
OTU.61	*Clostridium drakei, C. carboxidivorans*	95.92
OTU.11	*Clostridium drakei, C. carboxidivorans*	95.94
OTU.1673	*Clostridium drakei, C. carboxidivorans*	95.9
OTU.1747	*Clostridium hydrogeniformans, C. algidicarnis*	94.36
OTU.327	*Clostridium hydrogeniformans, C. amylolyticum*	94.92
OTU.330	*Clostridium lundense*	96.94
OTU.75	*Clostridium lundense*	96.97
OTU.2175	*Clostridium paraputrificum, C. lundense*	95.96
OTU.643	*Clostridium tagluense, C. estertheticum subsp. laramiense, C. estertheticum subsp. estertheticum, C. bowmanii, C. algoriphilum*	97.45
OTU.17	*Clostridium thiosulfatireducens, C. sulfidigenes, C. subterminale*	95.45
OTU.176	*Delftia tsuruhatensis, D. lacustris*	100.0
OTU.78	*Desulfocella halophila, Bryobacter aggregatus*	80.31
OTU.55	*Desulfocella halophila, Bryobacter aggregatus*	81.03
OTU.2404	*Domibacillus robiginosus*	99.49
OTU.3712	*Eubacterium tarantellae, Clostridium perfringens*	96.43
OTU.4167	*Fonticella tunisiensis*	93.43
OTU.4037	*Fonticella tunisiensis*	93.85
OTU.57	*Fonticella tunisiensis, Caloramator proteoclasticus*	93.88
OTU.575	*Gracilibacter thermotolerans*	94.42
OTU.37	*Ilyobacter delafieldii, Clostridium nitrophenolicum, C. aciditolerans*	96.43
OTU.14	*Pantoea rwandensis, P. rodasii, Kluyvera intermedia, K. cryocrescens, Klebsiella variicola, K. pneumoniae subsp. rhinoscleromatis, K. pneumoniae subsp. pneumoniae, Erwinia aphidicola, Enterobacter soli, E. ludwigii, E. kobei, E. hormaechei, E. cloacae subsp. dissolvens, E. cancerogenus, E. asburiae, E. amnigenus, E. aerogenes, Buttiauxella warmboldiae, B. noackiae, B. izardii, B. agrestis*	99.49
OTU.259	*Parasporobacterium paucivorans*	98.47
OTU.321	*Pseudomonas beteli*	100.0
OTU.54	*Shigella sonnei, S. flexneri, Escherichia fergusonii, E. coli*	100.0
OTU.116	*Streptomyces ziwulingensis, S. viridodiastaticus, S. viridochromogenes, S. violascens, S. violarus, S. violaceus, S. violaceorubidus, S. violaceoruber, S. violaceolatus, S. violaceochromogenes, S. vinaceusdrappus, S. variabilis, S. tuirus, S. tricolor, S. thinghirensis, S. tendae, S. spectabilis, S. silaceus, S. rutgersensis, S. rubrogriseus, S. roseoviolaceus, S. rochei, S. resistomycificus, S. plicatus, S. phaeoluteichromatogenes, S. parvulus, S. paradoxus, S. pactum, S. olivaceus, S. mutabilis, S. misionensis, S. minutiscleroticus, S. matensis, S. massasporeus, S. marokkonensis, S. malachitospinus, S. luteogriseus, S. longispororuber, S. lienomycini, S. levis, S. labedae, S. koyangensis, S. jietaisiensis, S. janthinus, S. intermedius, S. iakyrus, S. hydrogenans, S. humiferus, S. hawaiiensis, S. griseorubens, S. griseoincarnatus, S. griseoflavus, S. griseochromogenes, S. griseoaurantiacus, S. gougerotii, S. glomeratus, S. ghanaensis, S. geysiriensis, S. fulvissimus, S. flavofungini, S. flaveolus, S. eurythermus, S. erythrogriseus, S. diastaticus subsp. diastaticus, S. daghestanicus, S. collinus, S. coerulescens, S. coeruleorubidus, S. coeliflavus, S. coelescens,*	100.0

Figure 1. Ordination of heavy gradient fraction sequence collections by Bray-Curtis distances.
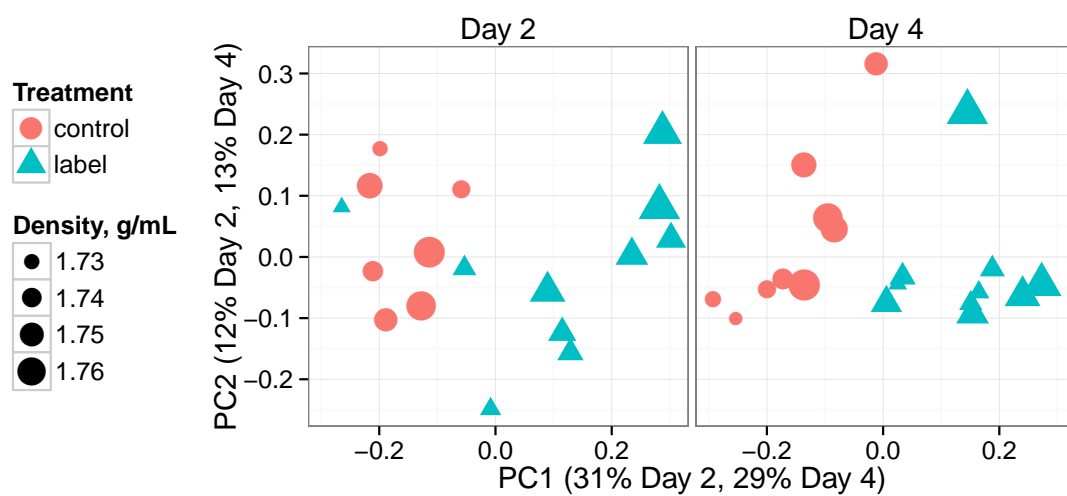
Figure 2. Phylogenetic trees of OTUs passing sparsity threshold for selected phyla. *A)* Point denotes OTU is classified as a ^{15}N "responder". *B)* Heatmap of moderated log$_2$ proportion mean ratios (labeled:control gradients) for each OTU at each incubation day. High values indicate ^{15}N incorporation. *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).
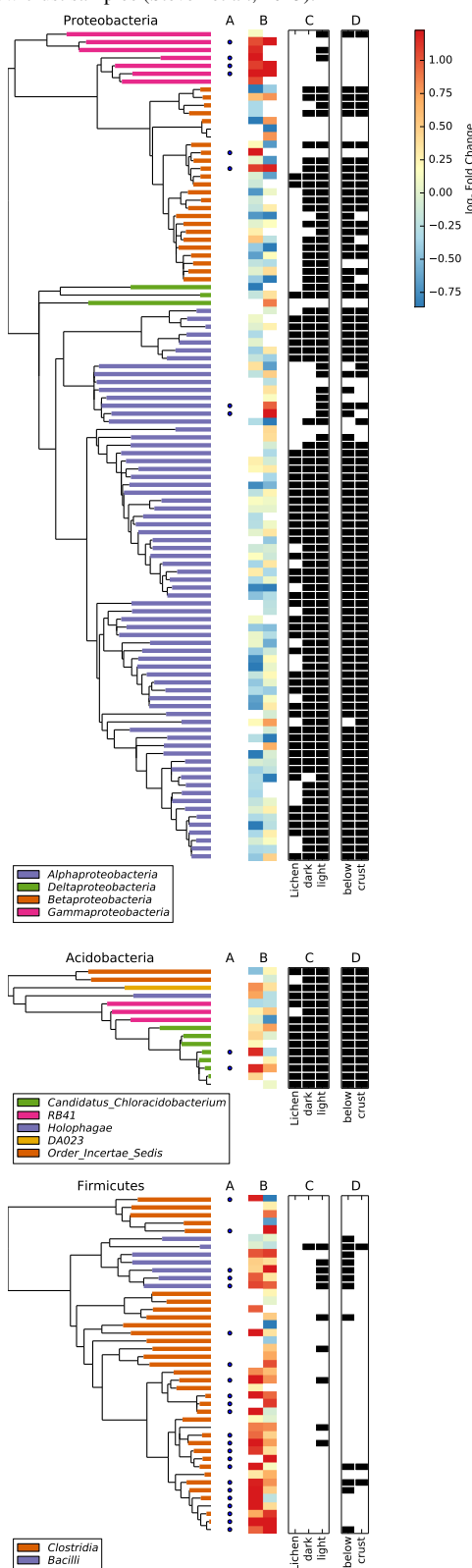
Figure 3. Moderated log$_2$ of proportion mean ratios for labeled versus control gradients (heavy fractions only, densities ¿1.725 g/mL). All OTUs found in at least 62.5% of heavy fractions at a specific incubation day are shown. Red color denotes a proportion mean ratio that has a corresponding adjusted 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.