

UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

nifH pyrosequencing reveals the potential for location-specific soil chemistry to influence N₂-fixing community dynamics.

Permalink

<https://escholarship.org/uc/item/56g5f9d0>

Journal

Environmental microbiology, 16(10)

ISSN

1462-2912

Authors

Collavino, Mónica M
Tripp, H James
Frank, Ildiko E
et al.

Publication Date

2014-10-01

DOI

10.1111/1462-2920.12423

Peer reviewed

1 ***nifH* pyrosequencing reveals the potential for location-specific soil chemistry to**
2 **influence N₂-fixing community dynamics**

3
4 Mónica M Collavino^{1,2}, H James Tripp³, Ildiko E Frank³, María L Vidoz¹, Priscila A
5 Calderoli², Mariano Donato⁴, Jonathan P Zehr³, and O Mario Aguilar²

6
7 ¹Instituto de Botánica del Nordeste (IBONE), Facultad de Ciencias Agrarias, Universidad
8 Nacional del Nordeste-CONICET, Corrientes, Argentina

9 ²Instituto de Biotecnología y Biología Molecular (IBBM), Universidad Nacional de La Plata-
10 CONICET, La Plata, Argentina

11 ³Department of Ocean Sciences, University of California, Santa Cruz, CA, USA

12 ⁴Laboratorio de Sistemática y Biología Evolutiva, Facultad de Ciencias Naturales y Museo,
13 Universidad Nacional de La Plata, La Plata, Argentina

14
15 **Corresponding author:** O. Mario Aguilar, 49 esq 115 (1900) La Plata, Argentina.

16 Telephone/Fax number 0054-0221-4229777. E-mail: aguilar@biol.unlp.edu.ar.

17 **Running head:** N₂-fixing communities in agricultural soils

18
19 **Summary**

20 A dataset of 87020 *nifH* reads and 16782 unique *nifH* protein sequences obtained over two
21 years from four locations across a gradient of agricultural soil types in Argentina were
22 analysed to provide a detailed and comprehensive picture of the diversity, abundance, and
23 responses of the N₂-fixing community in relation to differences in soil chemistry and
24 agricultural practices. Phylogenetic analysis revealed an expected high proportion of Alpha-,
25 Beta- and Delta-proteobacteria, mainly relatives to *Bradyrhizobium* and
26 *Methylosinus/Methylocystis*, but a surprising paucity of Gamma-proteobacteria. ANOVA and
27 stepwise regression modelling suggested location and treatment-specific influences of soil
28 type on diazotrophic community composition and organic carbon concentrations on *nifH*
29 diversity. *nifH* gene abundance, determined by qPCR, was higher in agricultural soils than in
30 non-agricultural soils, and was influenced by soil chemistry under intensive crop rotation but
31 not under monoculture. At some locations, sustainable increased crop yields might be
32 possible through the management of soil chemistry to improve the abundance and diversity of
33 N₂-fixing bacteria.

34
35 **Introduction**

36 Biologically available N is often a limiting nutrient in agricultural soil and other environments
37 (Vitousek and Howarth, 1991). To this end, 4.32 x 10⁵ tons of nitrogen fertilizers are used
38 annually in the productive agricultural region of Argentina, which have recently been
39 dedicated to soybean monoculture, due to financial considerations that also lead to minimal

40 nutrient restoration (FAO, 2012). In an effort to mitigate the negative effects of monoculture
41 on soil quality, principally increased erosion and decreased moisture retention, reduced-till
42 agricultural practices have been widely introduced in Argentina (Viglizzo *et al.*, 2011).
43 Studies have shown that while reduced-till practices can improve soil quality by reducing loss
44 of nutrients and increasing moisture retention, it remains to be seen whether the beneficial
45 effects are fully realized under monoculture and low nutrient replenishment (hereafter named
46 poor no-till practices) as compared to intensive crop rotation and nutrient replacement (named
47 good no-till practices) (Abid and Lal, 2009; Souza *et al.*, 2013). Specifically, it is not known
48 if reduced-till practices under monoculture improve the diversity, abundance, and community
49 structure of N₂-fixing microbes in agricultural soil.

50 Biological N₂ fixation (BNF), which is the biological reduction of molecular N₂ gas to
51 biologically-available ammonium, accounts for approximately 128 million tons nitrogen per
52 year and is considered the main route by which fixed nitrogen enters the biosphere by natural
53 processes (Galloway *et al.*, 2004). BNF, catalyzed only by Bacteria and Archaea, requires
54 nitrogenase, an evolutionarily conserved protein in N₂-fixing microorganisms. The
55 nitrogenase enzyme is composed of two multisubunit metallo-proteins. Component I contains
56 the active site for N₂ reduction and is composed of two heterodimers encoded by the *nifD* and
57 *nifK* genes. Component II, also known as dinitrogenase reductase, is composed of two
58 identical subunits encoded by the *nifH* gene. The *nifH* gene sequence is highly conserved
59 across the bacterial and archaeal domains; however, because of codon redundancy for most
60 amino acids, the design of universal *nifH* primers requires a considerable degree of DNA
61 sequence degeneracy. Up to date, several degenerated PCR primers were developed and have
62 been successfully used in culture-independent studies of microbial N₂-fixing communities in
63 terrestrial and aquatic environments for almost 25 years (Zehr and McReynolds, 1989; Zehr *et al.*,
64 1998; Poly *et al.*, 2001b; Rosch *et al.*, 2002; Steward *et al.*, 2004; Bürgmann *et al.*, 2005;
65 Izquierdo and Nüsslein, 2006; Farnelid *et al.*, 2011; Niederberger *et al.*, 2012). As a result,
66 the database for nitrogenase genes (specifically the *nifH* gene) has become one of the largest
67 non-ribosomal gene datasets from uncultivated microorganisms (Zehr *et al.*, 2003) This
68 makes *nifH* gene an outstanding reference tool for studying N₂-fixing microbial communities
69 in soil, most of whose members have not yet been cultured. The advent of pyrosequencing
70 now affords the opportunity to study N₂-fixing soil microbial communities in depth, providing
71 tens or hundreds of thousands of *nifH* amplicon sequences per sample.

72 Like the general survey of soil microbial communities using deep pyrosequencing of 16S
73 rRNA gene (Figuerola *et al.*, 2012), this study is part of a larger effort to provide microbial

74 indicators of the sustainability of no-till practices in four agricultural sites located across a
75 west-east transect in the Argentine Central Pampas (Fig. 1). However, this study is a specific
76 survey of only the N₂-fixing microbial communities in soil, using deep pyrosequencing of the
77 *nifH* gene. The reader is referred to the Introduction and Methods sections of Figuerola *et al.*
78 2012 for a detailed description of the broader BIOSPAS project and the study sites selected.
79 The aim of this work was to examine the abundance, diversity and structure of diazotroph
80 communities in a gradient of Argentinean agricultural soils, under good and poor no-till
81 practices, using deep pyrosequencing-based analysis of the *nifH* gene.

82

83 **Results**

84 **Proteobacteria are predominant in the diazotrophic community of Argentinean soils**

85 A total of 87020 *nifH* reads, comprising 16782 unique protein sequences, were obtained from
86 four Argentinean soils subjected to different management. At 98% sequence similarity, the
87 1558 OTUs for the entire study were imported into an ARB database of *nifH* sequences (Zehr
88 *et al.*, 2003) and assigned to 17 subclusters of the four major *nifH* clusters previously defined
89 by Zehr *et al.*, (2003) (Table 1). The abundances varied widely among the OTUs; about 1000
90 OTUs were observed to be represented by no more than 10 sequences, whereas 1375 had less
91 than 50 sequence counts, and only 19 OTUs were represented by more than 1000 sequences
92 (Fig. S1). The proportion of OTU counts and sequence counts was similar in each subcluster
93 (Table 1). Cluster I was the most abundant in the soil samples, particularly the subclusters 1K
94 (40% of the total number of OTUs), 1A (21%) and 1J (20%) which are mainly composed of
95 sequences related to Rhizobiales and Burkholderiales, *Anaeromyxobacter* and
96 Desulfuromonadales, and Rhizobiales and Rhodospirillales, respectively. The next largest
97 group was subcluster 3B (8%), with sequences mainly related to Desulfovibrionales and
98 Verrucomicrobiales (Table 1). No new *nifH* subclusters were identified in this study; however
99 it is noteworthy that new *nifH* sequences within subclusters were observed, i.e. 285 of the
100 1558 OTUs, most belonging to the subclusters 1A, 3B and 1J, have less than 95% similarity
101 at the amino acid level as compared to the sequences present in the *nifH* database (data not
102 shown).

103 Nineteen of the most abundant OTUs corresponding to about 48% of our *nifH* database were
104 distributed among the four subclusters 1K (12 OTUs), 1J (4 OTUs), 1A (2 OTUs) and 3B (1
105 OTU) (Fig. 2). Most of 1K and 1J OTUs were related to Alphaproteobacteria sequences. OTU
106 3466 of subcluster 1K (Fig. 2) diverged into a separated group, which contains only
107 uncultivated sequences recovered from soil, marine, and plant-associated environments (Zehr

108 *et al.*, 2003; Hsu and Buckley, 2009; Lovell *et al.*, 2000; Moisander *et al.*, 2005). The
109 remaining 1K OTUs clustered with sequences from the Rhizobiales order, which appeared to
110 be highly represented in the database (33% of the total database). These OTUs were more
111 closely related to sequences belonging to *Bradyrhizobium* (7 OTUs) and the
112 *Methylocystis/Methylosinus* methanotrophic group (4 OTUs), which represented 23.6% and
113 9.4% of the total database, respectively. OTUs of subcluster 1J were related to
114 *Rhodospirillales* sequences; one of them was related to *Azospirillum* while the remaining ones
115 clustered with environmental sequences and were related to sequences from phototrophic
116 purple non-sulfur bacteria belonging to Rhizobiales and Rhodospirillales orders.

117 Three abundant OTUs were represented by sequences that clustered with Deltaproteobacteria
118 sequences; two of them were classified as 1A and one as 3B. The closest cultivated sequences
119 for the two 1A OTUs were *Anaeromyxobacter* and *Pelobacter/Geobacter*, respectively. The
120 3B OTU clustered with uncultured soil sequences, which are closely related to sequences of
121 the sulphate-reducing bacteria of the *Desulfobacteraceae* family.

122 Some major clusters and subclusters were underrepresented in our samples. Major clusters II
123 and IV were represented by only a few OTUs, which is perhaps understandable given that
124 cluster II contains the alternative (FeV and FeFe) nitrogenases and cluster IV is composed of
125 *nifH* paralogs. By contrast, the scarcity of Gammaproteobacteria in subclusters 1O and 1M
126 was surprising, as well as the paucity of Clostridiales, Pseudomonadales and
127 Enterobacteriales representatives from subgroups 3A and 1G (Table 1).

128 These data revealed that even when soils contain representatives of most of the diversity of
129 *nifH* sequences found in natural environments, sequences belonging to Alpha, Beta and Delta-
130 proteobacteria were predominant.

131 **Geographic location is a major determinant of *nifH* diversity in soil**

132 The diversity of *nifH* in soils was examined using rarefaction curves (Fig. S2) and various
133 estimators of richness and evenness (Table 2). At 2% dissimilarity, none of the rarefaction
134 curves reached saturation. However, the relative coverage was similar among samples,
135 ranging from 59 to 84% and 55 to 77% according to S_{Chao1} and S_{ACE} estimators, respectively.
136 These numbers indicate that although the sequencing effort did not fully cover the potential
137 diversity, a substantial and comparable fraction was assessed in all samples.

138 As shown in Table 2, the richness and diversity of the diazotrophic community were strongly
139 affected by the geographical location of the sampling site, but not by the soil management
140 practices. One-way ANOVA models confirmed significant differences among the four
141 locations (**$p = 3e^{-6}$ for richness and $p = 4e^{-5}$ for diversity**) but not between years or

142 treatments. Two-way ANOVA did not indicate significant interaction effects. Whereas the
143 sandy soils of Bengolea displayed the lowest number of *nifH* OTUs and the lowest Shannon
144 index (H') values, Viale clay soils had the highest richness and diversity (Table 2).

145 Stepwise regression models were applied to assess the relationship between diazotroph
146 diversity measured by the Shannon index and soil chemical composition. Only C content was
147 selected as significant predictor ($p = 0.016$, $r^2 = 0.38$). With treatment factor ($p = 0.004$)
148 added to C content ($p = 3e^{-5}$) as predictor, the model improves considerably ($r^2 = 0.60$)
149 indicating that *nifH* diversity is significantly affected by soil organic carbon.

150 **Soil properties affect the composition of diazotrophic community**

151 There were clear differences in the proportion of the eight predominant subclusters with
152 respect to location. The westernmost (Bengolea) and easternmost (Viale) locations showed
153 the largest differences compared to the two central locations (Monte Buey and Pergamino),
154 likely due to contrasting soil types (Fig. 3.A, top row). As noted before, subcluster 1K
155 dominates all locations. The proportion of subcluster 1J sequences decreases from west
156 (Viale) to east (Bengolea). As mentioned above, the lowest diversity among sites was
157 observed at Bengolea, which consists mainly of subclusters 1K and 1J. Correspondence
158 analysis plots show a more comprehensive and detailed view of the similarities between
159 locations as defined by their subcluster proportions (Fig. 3.B). Bengolea samples, which
160 grouped on the left side of the figure, are unique in their high proportions of 1J and 1K, but
161 also 1B, 1E, and 2. Monte Buey and Pergamino samples share similar cluster proportions and,
162 therefore, overlap in the figure. Viale samples separated along the second component, were
163 characterized by high proportions of 1, 1C, 1D, and 3B clusters. Samples do not group
164 according to treatment (data not shown). Chi-sq independence test ($p = 2e^{-16}$) suggests that
165 locations and *nifH* subclusters are not independent, i.e. subcluster proportions vary
166 significantly across locations.

167 Differences in the proportion of subclusters are subtler across treatments (Fig. 3.A, bottom
168 row). The pie chart shows that the proportion of 1J subcluster is notably high in GAP and
169 PAP samples. The trend in subcluster 1K proportion is NE>GAP>PAP, whereas subcluster
170 3B displays the opposite trend (Fig. 3.A, bottom row).

171 The strong relationship between location and *nifH* cluster proportions was confirmed by
172 linear discriminant analysis. Using the proportion of 13 *nifH* subclusters as predictors, 100%
173 of samples were correctly assigned to location categories (Bengolea, Monte Buey, Pergamino,
174 Viale), while only 88% of samples to treatment categories (NE, GAP, PAP), and 83% of
175 samples to year categories (2010, 2011).

176 The proportion of the four predominant *nifH* subclusters (1A, 1J, 1K, and 3B) is affected
177 differently by soil chemical properties (Fig. 4). The proportion of subclusters 1A and 3B
178 increased with increasing pH and moisture. Subcluster 1J proportion increased, while 3B
179 decreased with decreasing organic C, N, and P levels. Subcluster 1K was unaffected by soil
180 chemistry, explaining its dominance in all samples. The above findings were confirmed by
181 stepwise regression models (Table S1).

182 **Environment influences abundance of soil N₂-fixing bacteria**

183 The relative abundance of N₂-fixing bacteria was determined by qPCR. In general, the
184 number of *nifH* copies varied significantly across sites (Monte Buey<Bengolea<Pergamino; r^2
185 = 0.46, $p = 4e^{-6}$) and were significantly higher in year 2011 than 2010 ($r^2 = 0.30$, $p = 6e^{-5}$)
186 (Fig. 5). In addition to significant main effects of location and year, their interaction also
187 proved to be significant ($p = 0.002$); this three term ANOVA model describes most of the
188 *nifH* abundance variation ($r^2=0.83$), revealing a major influence of the environment on
189 abundance of soil diazotrophs. In contrast, the effect of treatment (PAP, GAP and NE), varied
190 across locations and years; hence, when analyzing all samples together, no relationship was
191 found between *nifH* gene abundance and agricultural treatments. However, when abundance
192 was examined separately at each site, year and treatment effects were found to be significant
193 at Bengolea ($r^2 = 0.84$, $p = 0.001$, $p = 0.015$) and Monte Buey ($r^2 = 0.77$, $p = 0.008$, $p =$
194 0.017). At these two locations, *nifH* abundance was significantly higher under PAP and lower
195 under NE treatment in both years (Fig. 5). This pattern is opposite to the C, P and N levels,
196 i.e. *nifH* abundance is lowest under NE treatment where C, P and N contents are highest.

197 The relationship between soil chemistry and *nifH* abundances was location- and treatment-
198 specific. This is shown by better fit and prediction of regression models calculated for each
199 location or treatment separately compared to modeling all samples together ($r^2=0.57$, r^2 -
200 $pred=0.48$). For two of the four locations, Bengolea ($r^2=0.97$, r^2 - $pred=0.93$) and Pergamino
201 ($r^2=0.86$, r^2 - $pred=0.80$), as well as for NE ($r^2=0.95$, r^2 - $pred=0.88$) and GAP ($r^2=0.94$, r^2 -
202 $pred=0.85$) treatments, soil chemistry parameters showed high r^2 fit and prediction values,
203 suggesting that soil chemistry might influence *nifH* abundance under these conditions. In
204 contrast, low prediction values at Monte Buey ($r^2=0.76$, r^2 - $pred=0.58$), Viale ($r^2=0.84$, r^2 -
205 $pred=0.35$) and PAP treatment ($r^2=0.54$, r^2 - $pred=0.42$) indicate that the models are likely to be
206 overfit to the training data and soil chemistry does not influence *nifH* abundance at these
207 locations and treatment.

208 The location- and treatment dependence of soil chemistry also manifests in the dissimilar
209 correlation patterns between abundance and soil parameters. For example, under NE and GAP

210 treatment strong significant correlation was found between *nifH* abundance and P (-0.66 and -
211 0.64), while pH ($r = -0.74$) and moisture ($r = -0.56$) were the main influencing parameters
212 under PAP treatment. Considering locations, *nifH* abundance was negatively correlated with
213 N ($r = -0.79$), P ($r = -0.88$) and moisture ($r = -0.84$) in Bengolea, whereas in Pergamino the
214 most significant correlation ($r = -0.93$) was with pH. It is notable that the significant
215 correlations between individual soil chemistry parameters and *nifH* abundance were all
216 negative.

217 **Discussion**

218 This study examined the community structure of N₂-fixing bacteria in soils, and the effects of
219 soil type and management practice on the abundance and structure of the N₂-fixing
220 communities in Argentinean agricultural soils. This database makes a significant contribution
221 to the number of *nifH* sequences coming from soil, which in the recent report by Gaby and
222 Buckley (2011) was estimated to be 3644 unique reads at the Genbank sequence database.
223 The 16782 unique protein sequences and 87020 *nifH* reads we obtained, along with associated
224 metadata, provide the most detailed and comprehensive picture of the diversity, abundance,
225 and environmental responses of the N₂-fixing community in Argentinean soils, which produce
226 much of the world's supply of soybeans. We observed evidence of community members that
227 responded to specific environmental factors, and some evidence of broad community effects
228 related to soil type and soil chemistry.

229 It has been indicated that technical reproducibility and complete removal of sequencing
230 artifacts are important issues in the analysis of sequence data generated by pyrosequencing
231 (Zhou et al., 2011; Knight et al., 2012; Pinto and Raskin, 2012). In this regard, we chose to
232 survey space and time more broadly, without technical replicates. However, to limit the
233 creation of artifacts and to insure their removal, we pooled amplicons from multiple PCR
234 reactions, removed low quality sequences, chimeras and frameshifts, used an equal number of
235 sequences in samples, and removed less frequent OTUs (Schloss et al., 2011). As a result, our
236 dataset is an accurate survey and represent a broad region of agricultural importance for
237 Argentine.

238 We found that Alpha- and Beta-proteobacteria (subclusters 1K and 1J) and Delta-
239 proteobacteria (subcluster 1A) were well represented in our samples as expected, but not
240 Gamma-proteobacteria (Table 1). The low representation of *nifH* sequences related to
241 Gamma-proteobacteria is striking considering that several genera included in this group
242 (*Pseudomonas*, *Enterobacter* and *Azotobacter*) were found to be common components of soil
243 (Wang et al., 2012; Liu et al., 2012). As in any other approach based on PCR amplification,

244 we cannot rule out that underrepresentation of Gamma-proteobacteria sequences should be
245 consequence of primer bias. However, the lack of Gammaproteobacteria has also been
246 reported in the diazotrophic community of some tropical soils examined by both cultivation-
247 dependent and -independent approaches (Izquierdo and Nüsslein, 2006).

248 The Alpha-proteobacterial sequences related to Rhizobiales order accounted for 33% of the
249 total database. Two of the largest groups of sequences in subcluster 1K, groups 657 and 2991
250 (Fig. 2), were found in NE and GAP treatments, but poorly represented or not found in PAP
251 practice soils (data not shown), indicating that they prefer either natural conditions or Good
252 no-till Agricultural Practice, such as corn intercropping. Interestingly, groups 657 and 2991
253 fell into the cluster that groups the photosynthetic *Bradyrhizobium* sp. strains, one of which
254 (clone T1t015) has been reported to be positively affected by the presence of maize residues
255 (Hsu and Buckley, 2009).

256 Sequences related to the methanotrophic group *Methylosinus/Methylocystis* were well
257 represented in our soil samples (9.4% of the total database). The abundance of these
258 sequences in soil has been attributed to their adaptive advantage in poor-carbon soils and to
259 the positive effect of fermentation processes associated with root exudates or stubble retention
260 (Duc *et al.*, 2009; Buckley *et al.*, 2008). However, the distribution of methanotrophic
261 sequences found in our soil was uneven and could not be related to any of these factors (data
262 not shown).

263 The variability in the composition of phylogenetic groups across geographical locations may
264 lend insight into the conditions that affect the different phylogenetic groups. For instance, the
265 distribution of 1B and 1J was mainly associated with the level of soil organic C (Table S1;
266 Fig. 4). Considering that most of the abundant 1J OTUs were related with phototrophic purple
267 non-sulfur bacteria, it could be hypothesized that low carbon content may promote increased
268 phototrophic bacterial populations. Conversely, the proportion of *nifH* sequences from orders
269 Campylobacterales (Cluster 1), Frankia (1D), Desulfovibrionales and Verrucomicrobiales
270 (3B) seems to be associated with high levels of both C and N. Moreover, the proportion of the
271 3B group, as well as the anaerobes Clostridiales (1C), *Anaeromyxobacter* and
272 Desulfuromonadales (1A), seems to be negatively affected by low levels of soil moisture
273 (Table S1). Consequently, these phylotypes were found underrepresented in sandy soils.

274 A significant difference in diazotrophic diversity was observed across the four locations but
275 not among management treatments. In a companion study using the same set of soil samples,
276 diversity of the whole bacterial community did not vary by location or management treatment
277 (Figuerola *et al.*, 2012), suggesting that diversity variation is a distinctive feature of the N₂-

278 fixing community. *nifH* diversity and richness levels were consistently low in the sandy soil
279 from Bengolea, whereas the clay soil from Viale displayed the highest diversity (Table 2).
280 Based on regression analysis, the differences of diversity across sites appear to be mainly
281 associated with levels of soil organic carbon. This result is not surprising, given that
282 heterotrophic N₂-fixers often metabolize organic matter both to fix N₂ and to maintain high
283 respiration rates to avoid O₂ inactivation of nitrogenase (Hill, 1992).
284 Interestingly, *nifH* abundance in agricultural soils was higher than in pristine non-agricultural
285 soils, suggesting that land use or presence of crops somehow promote diazotroph populations.
286 Therefore, our study provided no evidence that no-till production systems either with intense
287 crop rotation or monoculture practices negatively affect the level of the potential N₂ fixation
288 community of soils. It could be that perturbation of soil by cropping promotes increases in
289 abundances unlike the pristine environment which eventually reaches a steady state.
290 Species abundance, diversity and function are important components of the sustainable
291 agriculture systems. Within this ecological framework, we demonstrated at some locations
292 potential for management of soil chemistry to improve the abundance and diversity of N₂-
293 fixing bacteria. Surprisingly, at locations where good agricultural practices are being
294 followed, it could be beneficial from the standpoint of promoting N₂-fixation not to be overly
295 ambitious with nutrient replacement, as well as with practices that affect soil pH and water
296 content. This conclusion is supported by the negative correlations we sometimes found
297 between *nifH* abundance and nutrient levels, pH, and soil moisture. Considering that higher
298 *nifH* abundances are not always associated with higher N₂-fixation rates or diversity,
299 additional research on *nif* expression and N₂-fixation rates would be useful in order to
300 compare diazotroph community structure with N₂-fixation activity.

301

302 **Experimental procedures**

303 **Sample collection and soil characterization**

304 Soil samples were collected from four geographical locations in the Argentinean Pampas.
305 From west to east, the sampling locations were: Bengolea (33°01'31"S; 63°37'53"W) and
306 Monte Buey in Córdoba Province (32°58'14"S; 62°27'06"W); Pergamino in Buenos Aires
307 Province (33°56'36"S; 60°33'57"W); and Viale in Entre Ríos Province (31°52'59"S;
308 59°40'07"W) (Fig. 1). The sites comprise three different soil types, a sandy loam (Entic
309 Haplustoll) in Bengolea, a silty loam (Typic Argiudoll) in Pergamino and Monte Buey and a
310 silty clay (Argic Pelludert) in Viale (Table 3).

311 As previously described by Figuerola *et al.* (2012), three treatments were defined at each of
312 the four sampling geographical locations. Good no-till Agricultural Practices (GAP) treatment
313 is characterized by intensive crop rotation (soybean-maize), nutrient replacement, and low
314 agrochemical use. Poor no-till Agricultural Practices (PAP) treatment is characterized by crop
315 monoculture (soybean), low nutrient replacement, high agrochemical use and lower yields.
316 Grasslands uncultivated for 30 or more years were considered Natural Environment (NE)
317 treatments.

318 Sampling was performed in February (Southern hemisphere summer) of years 2010 and 2011.
319 Each treatment-site combination was sampled in three replicates from 5 m² quadrants
320 separated by at least 50m from each other. Samples consisted of a pool obtained from 16–20
321 randomly selected subsamples from the top 10 cm of bulk soil. These subsamples were
322 immediately combined and homogenized after field collection, and transported to the
323 laboratory at 4°C. Within 3 days after collection, samples were sieved through 4-mm mesh to
324 remove roots and plant detritus, and stored at -80°C until further processing. The replicates of
325 each treatment were independently analyzed for the q-PCR analysis. For pyrosequencing
326 analysis, DNA extraction and amplification were performed independently on each replicate,
327 and amplicons were pooled into a single sample before labelling barcoding, resulting in a total
328 of 24 composite samples (4 sites x 3 treatment x 2 time sampling).

329 The following soil characteristics were used in the subsequent statistical analyses: soil texture
330 (% silt and clay), pH, gravimetric moisture content (%), total organic carbon measured by dry
331 combustion (C, %), total nitrogen obtained by the Kjeldahl method (N, %), and extractable
332 phosphorous determined by the method of Bray and Kurtz (P, ppm) (Table 3). The
333 measurement and analysis of these physico-chemical characteristics have been previously
334 described (Figuerola *et al.*, 2012; Duval *et al.*, 2013). Briefly, there was a clear gradient in
335 soil texture from west to east with increasing clay and decreasing sand content from Bengolea
336 to Viale. Regarding chemical parameters, the lowest C and N values were found at Bengolea
337 under all three treatments, whereas Viale C and N levels surpassed the other sites but only for
338 GAP and PAP treatments. P levels were highest in Monte Buey while the lowest levels were
339 observed at Pergamino. With some exceptions, C, N and P were highest in NE treatment and
340 lowest in PAP. These soil properties show no variation between years. By contrast, moisture
341 content in 2011 was only 2/3 of the 2010 level ($14\% \pm 5.9$ vs. $22.4\% \pm 4.7$) and pH was also
342 lower in 2011 than 2010 (5.91 ± 0.29 vs. 6.62 ± 0.18).

343 **DNA extraction**

344 DNA extraction and amplification were performed independently for each replicate sample.
345 Soil DNA was extracted from 0.25 g of soil using FastDNA Spin kit for soil (MP
346 Biomedicals), in accordance with the manufacturer's instructions.

347 **Quantification of *nifH***

348 The relative abundance of the *nifH* gene was quantified via quantitative real-time PCR
349 (qPCR). The qPCR reactions contained 10 ng of soil DNA, 1.2 µl of each primer (5 pM) and
350 10 µl of 2x SYBR Green iCycler iQ mixture (Bio-Rad) and water for 20 µl final reaction
351 volume. The reaction was carried out on an Applied Biosystem 7500 real-time PCR system
352 (Applied Biosystems) using a program of 95°C for 10 min followed by 40 cycles consisting
353 of 15 s at 95°C, 20 s at 55°C, and 20 s at 72°C. Fluorescence was measured at the end of each
354 cycle. 16S rRNA gene abundances were used to normalize values between the different
355 samples. Relative quantities were calculated using Genorm
356 (<http://medgen.ugent.be/~jvdesomp/genorm/>). The following specific primers were used:
357 PolF-PolR for *nifH* (Poly *et al.*, 2001a) and 338F-518R for bacterial 16S rRNA genes (Park
358 and Crowley, 2005). All qPCR reactions were run in duplicate with DNA extracted from
359 replicates soil samples. For each run, the melting curve was analyzed to ensure specific
360 assessment of *nifH* gene.

361

362 **Generation of barcoded *nifH* gene libraries**

363 The diversity and composition of diazotrophic communities were assessed by pyrosequencing
364 analysis of the *nifH* gene. PolF-PolR degenerated universal primers were used to PCR
365 amplify an internal fragment (360 bp) of the *nifH* gene (Poly *et al.*, 2001a). Primer sequences
366 included the Roche 454 tag sequences A (CACGACGTTGTAACGAC) or B
367 (CAGGAAACAGCTATGACC) fused to the 5' end of the forward and reverse *nifH*-
368 primers, respectively. Amplification was performed using a FastStart High Fidelity PCR
369 system (Roche Applied Science, Mannheim, Germany) with the following parameters: initial
370 denaturation at 95°C for 5 min, 20 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 2
371 min, and a final elongation step at 72 °C for 10 min. The PCR reactions were conducted in
372 triplicate to minimize random PCR bias. Amplicons were purified by using QIAquick Gel
373 extraction (Qiagen, CA, USA) and quantified using a Nanodrop Spectrophotometer.
374 Replicates were pooled in equimolar concentrations in a single treatment library. Each of the
375 24 libraries was labelled with a unique oligonucleotide barcode and pyrosequenced using 454
376 GS FLX technology.

377 **Pyrosequencing data analysis**

378 Sequences shorter than the threshold of 300 bp in length or displaying ambiguities were
379 removed from the 24 pyrosequencing-derived datasets. The remaining reads were translated
380 in all six reading frames and compared to a *nifH* reference database
381 (<http://pmc.ucsc.edu/~wwwzehr/research/database/>) using BLASTx from iNquiry software
382 (Bioteam). A cutoff of $1e^{-10}$ for the *E*-value was set to remove nonspecific blast hits. Putative
383 frame shifts were detected and removed with the FrameBot tool (RDP, Ribosomal Database
384 Project) using a profile hidden Markov model (HMM) (Zehr *et al.*, 2003) as reference set. The
385 same HMM was used to align the *nifH* pyrosequences with the hmmlalign program (HMMER
386 2.3.2). The resulting alignment was imported into ARB (Ludwig *et al.*, 2004), evaluated by
387 eye and subjected to minor manual corrections. The regions of the primers were trimmed and
388 incomplete sequences were removed. High-quality sequences were subsequently assessed for
389 chimeras using the UCHIME algorithm (Edgar *et al.*, 2011). Following that, the nucleotide
390 sequences were clustered into OTUs (operational taxonomic units) using CD-HIT Suite
391 program (Huang *et al.*, 2010) with an OTU threshold value of 98% sequence identity at the
392 DNA level. OTU representative sequences were analyzed with *de novo* mode of UCHIME
393 using its relative abundance data. All putative chimeras detected using UCHIME default
394 settings were eliminated. Finally, the sequences in the database were clustered with OTUs
395 defined at 98% amino acid sequence similarity. The conservative OTU cut off used was
396 defined considering that protein-encoding genes of strains of a given species generally have a
397 high sequence similarity (Konstantinidis and Tiedje, 2005). In an attempt to further reduce
398 potential pyrosequencing errors, OTUs represented by less than three sequences were
399 excluded from the database. The relative abundance of sequences of the remaining OTUs was
400 normalized using the subsampling-based method described in mothur
401 (<http://www.mothur.org/wiki/Normalize.shared>), prior to comparative analyses.

402 OTU representative amino acid sequences along with sequences selected from the *nifH*
403 reference database were used to build protein phylogenetic trees. Ambiguously aligned
404 regions were detected by visual inspection and excluded from the analysis. Trees were
405 constructed by performing neighbor-joining analysis (with the Kimura correction) using the
406 ARB program.

407 The data set was deposited in the NCBI-SRA (Sequence Read Archive) with the submission
408 Accession Number SRP029166.

409 **Statistical Data Analyses**

410 Data were analyzed using R (R Core Team, 2013) and by Minitab Statistical Software, 2010.
411 Richness (Chao1, S_{Chao1} and ACE, S_{ACE}), Shannon diversity (H'), and rarefaction were

412 calculated with the vegan package in R (Oksanen *et al.*, 2013) using the diversity, estimate,
413 rarecurve, and specnumber functions.

414 Location, treatment, and year effects were explored graphically by parallel boxplots and their
415 significance was assessed by ANOVA models coupled with Tukey's multiple comparison of
416 means in Minitab. One-way and two-way models were calculated to fit richness and diversity
417 parameters as well as relative abundance values.

418 Relationship between soil chemical parameters and diazotroph community characteristics
419 were investigated calculating correlation and regression models in Minitab. Stepwise
420 regression (both forward and backward) models were calculated to select soil parameters with
421 significant effect on the Shannon diversity and on the relative abundance of diazotrophs.

422 Variation in community composition across locations and treatments was graphically
423 explored by pie charts created in Minitab. The pie charts were created by adding the
424 corresponding subsets (locations or treatments). Samples were subsetted by location and the
425 sum of six samples (three treatments and two years) was calculated at each location for each
426 cluster. Similarly, samples were subsetted by treatment and at each treatment the sum of eight
427 samples (four locations and two years) was calculated.

428 Further details about the *nifH* cluster profiles were obtained by correspondence analysis of the
429 contingency table using the ca package in R (Nenadic *et al.*, 2007). Location and treatment
430 effects were analyzed by chi-squared test and linear discriminant analysis calculated in R.
431 Abundance of selected *nifH* clusters were projected on principal component biplots calculated
432 from scaled soil parameters in order to examine the effect of soil chemistry on the main
433 diazotroph groups.

434

435 **Acknowledgements**

436 This work was supported by the Argentinean National Agency for Science and Technology
437 (FONCyT; project code PICT –PAE 2006-11). MMC, OMA and MLV are funded by the
438 National Research Council for Science and Technology of Argentina (CONICET).

439

440 **References**

441 Abid, M., and Lal R. (2009) Tillage and drainage impact on soil quality: II. Tensile strength
442 of aggregates, moisture retention and water infiltration. *Soil Till Res* **103**: 364–372.

443 Buckley, D.H., Huangyutitham, V., Shi-Fang, H., and Tyrrell, A.N. (2008) ¹⁵N₂-DNA-stable
444 isotope probing of diazotrophic methanotrophs in soil. *Soil Biol Biochem* **40**: 1272–1283.

445 Bürgmann, H., Meier, S., Bunge, M., Widmer, F., and Zeyer, J. (2005) Effects of model root
446 exudates on structure and activity of a soil diazotroph community. *Environ Microbiol.* **7**:
447 1711–1724.

448 Duc, L., Noll, M., Meier, B.E., Burgmann, H., and Zeyer, J. (2009) High Diversity of
449 diazotrophs in the forefield of a receding alpine glacier. *Microb Ecol* **57**: 179–190.

450 Duval, M.E., Galantini J.A., Iglesias, J.O., Canelo, S., Martinez, J.M., and Wall, L. (2013)
451 Analysis of organic fractions as indicators of soil quality under natural and cultivated
452 systems. *Soil Till Res* **131**: 11–19.

453 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight R. (2011) UCHIME improves
454 sensitivity and speed of chimera detection. *Bioinformatics*: **27**: 2194–2200.

455 FAO, Food and Agricultural Organization of the United Nations (2012) Nitrogenous
456 fertilizers. <http://faostat3.fao.org/home/index.html>

457 Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sorensen, S., *et*
458 *al.* (2011) Nitrogenase gene amplicons from global marine surface waters are dominated by
459 genes of non-cyanobacteria. *Plos One* **6**: e19223.

460 Figuerola, E.L.M., Guerrero, L.D., Rosa, S.M., Simonetti, L., Duval, M.E., Galantini, J.A., *et*
461 *al.* (2012) Bacterial indicator of agricultural management for soil under no-till crop
462 production. *Plos One* **7**: e51075.

463 Gaby, J.C., and Buckley, D.H. (2011) A global census of nitrogenase diversity. *Environ*
464 *Microbiol* **13**: 1790–1799.

465 Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S. *et*
466 *al.* (2004) Nitrogen cycles: Past, present, and future, *Biogeochemistry* **70**: 153–226.

467 Hsu, S-F, and Buckley, D.H. (2009) Evidence for the functional significance of diazotroph
468 community structure in soil. *ISME J* **3**: 124–136.

469 Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010) CD-HIT Suite: a web server for
470 clustering and comparing biological sequences. *Bioinformatics* **26**: 680-682.

471 Izquierdo, J.A., and Nüsslein, K. (2006) Distribution of extensive *nifH* gene diversity across
472 physical soil microenvironments. *Microbial Ecol* **51**: 441-452.

473 Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J.A., *et al.* (2012) Unlocking
474 the potential of metagenomics through experimental design. *Nat Biotechnol* **30**: 513–520.

475 Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species
476 definition for prokaryotes. *Proc. Natl. Acad. Sci.* **102**: 2567–2572.

477 Liu, J., Peng, M., and Li, Y. (2012) Phylogenetic diversity of nitrogen-fixing bacteria and the
478 *nifH* gene from mangrove rhizosphere soil. *Can J Microbiol* **58**: 531–539.

479 Lovell, C.R., Piceno, Y.M., Quattro, J.M., and Bagwell, C.E. (2000) Molecular analysis of
480 diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. Appl
481 Environ Microbiol **66**: 3814–3822.

482 Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., and Yadhukumar, *et al.* (2004)
483 ARB: a software environment for sequence data. Nucl Acids Res **32**: 1363–1371.

484 Minitab 16 Statistical Software (2010) Computer software. State College, PA: Minitab, Inc.
485 (<http://www.minitab.com>).

486 Moisander, P.H., Piehler, M.F., and Paerl, H.W. (2005) Diversity and activity of epiphytic
487 nitrogen-fixers on standing dead stems of the salt marsh grass *Spartina alterniflora*. Aquat
488 Microb Ecol **39**: 271–279.

489 Nenadic, O., and Greenacre, M. (2007) Correspondence Analysis in R, with two- and three-
490 dimensional graphics: The ca package. Journal of Statistical Software **20**: 1-13.

491 Niederberger, T.D., Sohm, J.A., Tirindelli, J., Gunderson, T., Capone, D.G., Carpenter, E.J.,
492 and Cary, S.C. (2012) Diverse and highly active diazotrophic assemblages inhabit
493 ephemerally wetted soils of the Antarctic Dry Valleys. FEMS Microbiol Ecol **82**: 376-90.

494 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al.*
495 (2013) Vegan: Community Ecology Package. R package version 2.0-7. [http://CRAN.R-](http://CRAN.R-project.org/package=vegan)
496 [project.org/package=vegan](http://CRAN.R-project.org/package=vegan).

497 Park, J.W., and Crowley, D.E. (2005) Normalization of soil DNA extraction for accurate
498 quantification of real-time PCR and of target genes by DGGE. Biotechniques **38**: 579–586.

499 Pinto, A.J., and Raskin, L. (2012) PCR Biases Distort Bacterial and Archaeal Community
500 Structure in Pyrosequencing Datasets. PLoS ONE **7**: e43093.

501 Poly, F., Monrozier, L.J., and Bally, R. (2001a) Improvement in the RFLP procedure for
502 studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. Res Microbiol
503 **152**: 95–103.

504 Poly, F., Ranjard, L., Nazaret, S., Gourbiere, F., and Jocteur-Monrozier, L. (2001b)
505 Comparison of *nifH* gene pools in soils and soil microenvironments with contrasting
506 properties. Appl Environ Microbiol **67**: 2255–2262.

507 R Core Team (2013) R: A language and environment for statistical computing. R Foundation
508 for Statistical Computing, Vienna, Austria (<http://www.R-project.org>).

509 Rosch, C., Mergel, A., and Bothe, H. (2002) Biodiversity of denitrifying and dinitrogen-
510 fixing bacteria in an acid forest soil. Appl Environ Microbiol **68**: 3818–3829.

511 Souza, R.C., Cantão, M.E., Ribeiro Vasconcelos, A.T., Nogueira, M.A., and Hungria, M.
512 (2013) Soil metagenomics reveals differences under conventional and no-tillage with crop

513 rotation or succession. *Appl Soil Ecol* **72**: 49-61.

514 Schloss, P.D., Gevers, D., and Westcott, S.L. (2011) Reducing the effects of PCR
515 amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* **6**: e27310.

516 Steward, G.F., Zehr, J.P., Jellison, R., Montoya, J.P., and Hollibaugh, J.T. (2004) Vertical
517 distribution of nitrogen-fixing phylotypes in a meromictic, hypersaline lake. *Microb. Ecol.* **47**:
518 30-40.

519 Viglizzo, E.F., Frank, F.C., Carreño, L.V., Jobbágy, E.G., Pereyra, H., and Clatt, J., *et al.*
520 (2011) Ecological and environmental footprint of 50 years of agricultural expansion in
521 Argentina. *Global Change Biol* **17**: 959-973.

522 Vitousek, P.M., and Howarth, R.W. (1991) Nitrogen limitation on land and in the sea: how
523 can it occur? *Biogeochemistry* **13**: 87–115.

524 Zehr, J.P., and McReynolds, L.A. (1989) Use of degenerate oligonucleotides for amplification
525 of the *nifH* gene from the marine cyanobacterium *Trichodesmium* spp. *Appl. Environ.*
526 *Microbiol.* **55**: 2522-2526.

527 Zehr, J.P., Mellon, M., and Zani, S. (1998) New nitrogen-fixing microorganisms detected in
528 oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl Environ Microbiol* **64**:
529 3444–3450.

530 Zehr, J.P., Jenkins, B.D., Short, S.M., and Steward, G.F. (2003) Nitrogenase gene diversity
531 and microbial community structure: a cross-system comparison. *Environ Microbiol* **57**: 539-
532 554.

533 Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.H., Tu, Q., *et al.*, (2011) Reproducibility and
534 quantitation of amplicon sequencing-based detection. *ISME J* **5**: 1303–1313.

535

536 **Table and figure legends**

537 **Table 1. *nifH* phylotypes obtained from Argentinean soils.**

538 **Table 2. Richness, diversity and relative sample coverage for *nifH* libraries.** The richness
539 estimators (Chao1, S_{Chao1} and ACE, S_{ACE}) and Shannon's diversity index (H') were calculated
540 for *nifH* libraries with OTUs defined at 98% amino acid sequence similarity.

541 **Table 3. Soil properties according to site and agricultural management at the sampling**
542 **dates.**

543 **Table S1. Soil parameters that explain significant variation in the proportion of *nifH***
544 **subclusters.**

545 **Fig. 1. Map of the four sampling locations in the Argentinean study area.** Soil samples
546 were collected from four geographical locations: Bengolea and Monte Buey in Córdoba

547 Province, Pergamino in Buenos Aires Province and Viale in Entre Ríos Province. Template
548 map downloaded from www.google.com/earth/ (30 October 2013).

549 **Fig. 2. Phylogenetic relationships of nineteen largest *nifH* OTUs based on neighbour-**
550 **joining analysis of partial amino acid sequences.** The numbers in shaded boxes identify the
551 largest OTUs observed in this study. The percentage of 500 bootstrap samples that supported
552 each branch is shown. Bootstrap values below 50% are not shown.

553 **Fig. 3. Proportion of the largest subclusters across locations and treatments.**

554 **A.** Pie charts show the proportion of the eight predominant subclusters with respect to
555 geographical location (top row) and treatment (bottom row). Subclusters 1K, 1J, and 1A
556 dominate at each location and under each treatment (NE = Natural Environment; GAP =
557 Good no-till Agricultural Practices; PAP = Poor no-till Agricultural Practices).

558 **B.** Correspondence analysis plot displays rows (samples) and columns (*nifH* subclusters) of a
559 cross-tabulation of sequence counts. Closeness on the plot indicates similarity between
560 samples in terms of subcluster proportions or similarity between subclusters in terms of their
561 distribution across samples. Bengolea and Viale samples are enclosed by ellipses. The first
562 component (horizontal axis) covers 51% and second component (vertical axis) covers
563 additional 26% of variation.

564 **Fig. 4. Proportion of the four largest subclusters related to location, treatment, year, and**
565 **soil chemistry.** Proportion is indicated by circle size at each sample point plotted on the first
566 two principal components of the five soil parameters (represented by arrows). Component 1
567 (horizontal axis) covers 49% and component 2 covers 30% of the variance in soil chemistry.
568 Each sample is labeled by its location (B = Bengolea, M = Monte Buey, P = Pergamino, V =
569 Viale), treatment (NE = Natural Environment, GAP = Good no-till Agricultural Practices,
570 PAP = Poor no-till Agricultural Practices), and year (10 = 2010, 11 = 2011). Circle size is
571 scaled independently for each of the four subclusters.

572 **Fig. 5. Location, year, and treatment effect on *nifH* gene relative abundance**

573 Rows of panels correspond to years (2010, 2011) and columns to locations (Bengolea, Monte
574 Buey, Pergamino, and Viale). Colors indicate soil management treatment: Natural
575 Environment (NE) in white, Good no-till Agricultural Practices (GAP) in gray and Poor no-
576 till Agricultural Practices (PAP) in dark gray. Boxes display the range of three biological and
577 two technical replicates. Levels normalized to 16S rRNA expression are indicated as relative
578 units. A value of 1 was assigned to the lowest detected value (Monte Buey-NE-2010 sample).

579 **Fig. S1. Abundance of 1558 OTUs**

580 Abundance is quantified as sequence count in each OTU. The hockey-stick curve indicates
581 that only 19 OTUs are represented by more than 1000 sequences.

582 **Fig. S2. Rarefaction curves grouped by location**

583 Number of sequences (horizontal axis) plotted against number of OTUs (vertical axis). Each
584 curve corresponds to a sample labeled by its location (B = Bengolea, M = Monte Buey, P =
585 Pergamino, V = Viale), treatment (NE = Natural Environment, GAP = Good no-till
586 Agricultural Practices, PAP = Poor no-till Agricultural Practices), and year (10 = 2010, 11 =
587 2011).

588

589 **Tables**590 **Table 1. *nifH* phylotypes obtained from Argentinean soils.**

591

Cluster	Subcluster	Proportion (%)		Distribution ¹ (%)	Group	Orders ²
		Sequence count	OTU count			
I	1	1.6	1.9	96	Epsilon	Campylobacterales
	1A	18	21.4	100	Delta	Anaeromyxobacter and Desulfuromonadales
	1B	0.8	1.4	87	Cyanobacteria	Nostocales and Chroococcales
	1C	0.2	0.8	62	Firmicutes	Clostridiales
	1D	0.1	0.8	50	Actinobacteria	<i>Frankia</i>
	1E	0.5	2.1	96	Firmicutes	<i>Paenibacillus</i>
	1G	0.1	0.26	46	Gamma	Pseudomonadales and Enterobacteriales
	1J	20	19.6	100	Alpha	Rhizobiales and Rhodospirillales
	1K	51	40.2	100	Alpha and Beta	Rhizobiales and Burkholderiales
	1M	0.005	0.06	12	Gamma	Methylococcales
	1O	0.01	0.06	4	Gamma	Chromatiales
	1P	2.4	2.7	100	Beta	Rhodocyclales
	II	2C	0.01	0.1	21	Alpha
2		0.02	0.1	42	Firmicutes	<i>Paenibacillus</i>
III	3A	0.02	0.1	29	Firmicutes	<i>Clostridium</i>
	3B	5	8	100	Delta and Verrucomicrobia	Desulfovibrionales and Verrucomicrobiales
IV	4	0.1	0.4	75	Archaea	Methanococcales

592

593 ¹ presence across the 24 samples analyzed (e.g. subcluster present in all samples shows a
594 distribution of 100%)595 ² orders closest to the predominant sequences observed in the subcluster

596

597 **Table 2. Richness, diversity and relative sample coverage for *nifH* libraries.** The richness
 598 estimators (Chao1, S_{Chao1} and ACE, S_{ACE}) and Shannon's diversity index (H') were calculated
 599 for *nifH* libraries with OTUs defined at 98% amino acid sequence similarity.
 600

	Bengolea			Monte Buey			Pergamino			Viale			
	NE	GAP	PAP	NE	GAP	PAP	NE	GAP	PAP	NE	GAP	PAP	
2010	SeqN	3142	3193	3763	2460	5870	9067	2632	2950	2756	4752	5278	6330
	SeqN _n	1927	1927	1927	1927	1927	1927	1927	1927	1927	1927	1927	1927
	Sobs	252	214	223	391	520	493	237	268	161	501	485	530
	H'	3.46	3.37	3.57	4.77	4.68	4.42	3.81	3.61	3.28	4.59	4.58	4.98
	S_{Chao1}	314	330	319	650	746	786	338	392	191	694	721	742
	S_{ACE}	353	355	319	703	778	774	353	430	209	728	744	767
	RC _{Chao1}	80	65	70	60	70	63	70	68	84	72	67	71
	RC _{ACE}	71	60	70	56	67	64	67	62	77	69	65	69
2011	Seq	2905	2098	4081	2652	2788	2980	1927	2162	2584	3762	4274	2614
	Seq _n	1927	1927	1927	1927	1927	1927	1927	1927	1927	1927	1927	1927
	Sobs	156	183	259	328	285	342	282	297	344	453	427	414
	H'	2.27	3.66	3.55	4.10	4.13	4.26	3.81	4.18	4.48	4.65	4.23	4.68
	S_{Chao1}	233	305	384	503	422	505	428	501	553	661	577	687
	S_{ACE}	237	331	424	540	445	530	471	502	579	710	616	703
	RC _{Chao1}	67	60	68	65	68	68	66	59	62	68	74	60
	RC _{ACE}	66	55	61	61	64	65	60	59	59	64	69	59

601
 602 **Abbreviations:** SeqN, number of sequences per sample; SeqN_n, normalized number of sequences per
 603 sample; Sobs, detected number of operational taxonomic units (OTUs) at 2% distance level; RC, relative
 604 coverage calculated as OTU number divided by estimated richness (Chao1 or ACE).
 605

606 **Table 3. Soil properties according to site and agricultural management at the sampling**
 607 **dates.**

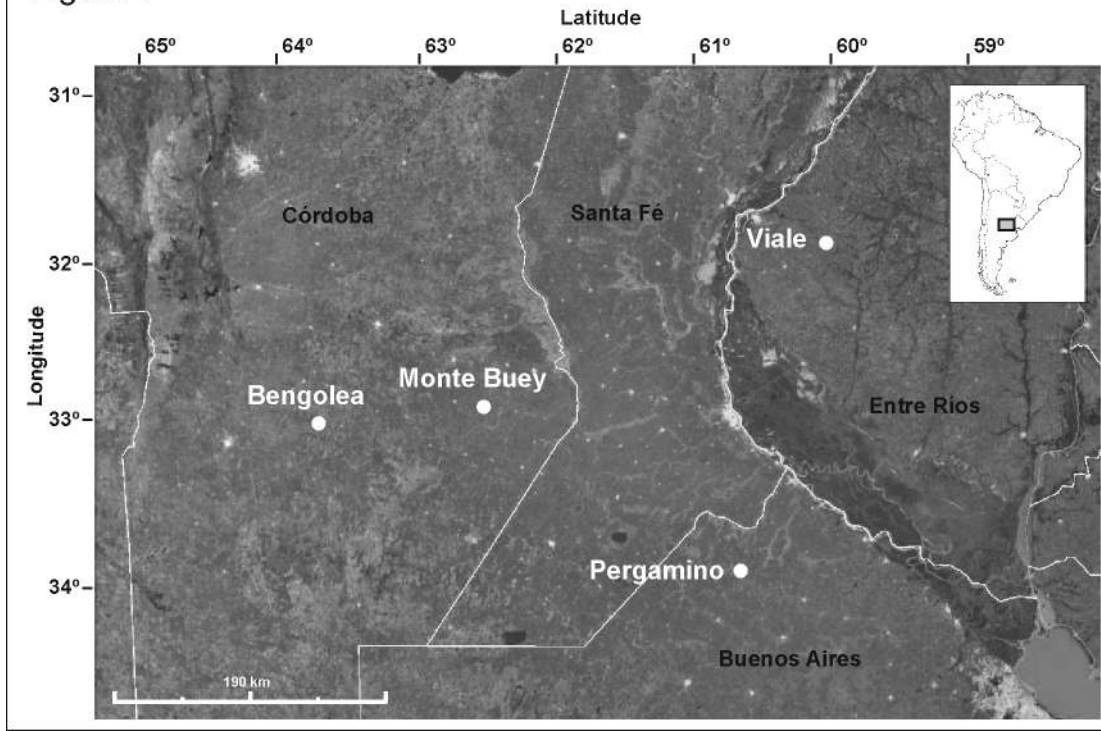
	Soil classification	Bengolea			Monte Buey			Pergamino			Viale		
		NE	GAP	PAP	NE	GAP	PAP	NE	GAP	PAP	NE	GAP	PAP
	Texture	Sandy loam			Silt loam			Silt loam			Silty clay loam		
2010	pH	6.7	6.5	6.6	6.4	6.5	6.6	6.6	6.4	6.6	7	6.9	6.7
	Carbon %	1.69	1.44	1.19	3.74	2.27	1.73	3.21	1.75	1.73	3.52	3.41	2.48
	N %	0.14	0.13	0.11	0.28	0.18	0.13	0.25	0.14	0.14	0.20	0.23	0.16
	P (ppm)	43.10	37.20	26.20	395.10	122.30	17.30	16.60	25.30	22.00	24.30	50.40	43.10
	Moisture %	18.56	15.48	18.00	26.44	25.13	25.32	26.60	22.53	17.41	23.37	28.69	22.08
2011	pH	6.1	5.6	6.1	5.6	5.5	6.2	5.8	5.6	6	6.1	6.5	5.9
	Carbon %	1.71	1.45	1.28	3.83	2.37	1.91	2.99	1.77	1.91	3.57	3.58	2.65
	N %	0.13	0.12	0.10	0.33	0.19	0.14	0.25	0.16	0.14	0.27	0.30	0.20
	P (ppm)	20.93	24.93	8.9	356.13	164.93	8.16	4.76	13.33	2.00	10.23	20.43	36.33
	Moisture %	11.35	16.26	5.63	10.22	15.60	7.95	26.65	19.50	14.98	10.58	19.83	9.92

608
 609

610 **Table S1. Soil parameters that explain significant variation in the proportion of *nifH***
 611 **subclusters.**

	<u>Subcluster</u>	<u>Significant parameters</u>	<u>r² fit</u>	<u>r² predicted</u>
613	<u>1</u>	<u>C, N</u>	<u>0.73</u>	<u>0.60</u>
614	<u>1A</u>	<u>pH, moisture</u>	<u>0.33</u>	<u>0.12</u>
615	<u>1B</u>	<u>C</u>	<u>0.10</u>	<u>0</u>
	<u>1C</u>	<u>moisture</u>	<u>0.12</u>	<u>0</u>
616	<u>1D</u>	<u>C, N</u>	<u>0.35</u>	<u>0</u>
617	<u>1E</u>	<u>=</u>		
	<u>1G</u>	<u>pH, P</u>	<u>0.52</u>	<u>0.38</u>
618	<u>1J</u>	<u>C, P</u>	<u>0.44</u>	<u>0.32</u>
	<u>1K</u>	<u>moisture</u>	<u>0.17</u>	<u>0</u>
619	<u>1P</u>	<u>=</u>		
620	<u>2</u>	<u>moisture, N</u>	<u>0.20</u>	<u>0</u>
	<u>3B</u>	<u>moisture, C, N</u>	<u>0.49</u>	<u>0.15</u>
621	<u>4</u>	<u>=</u>		
622				

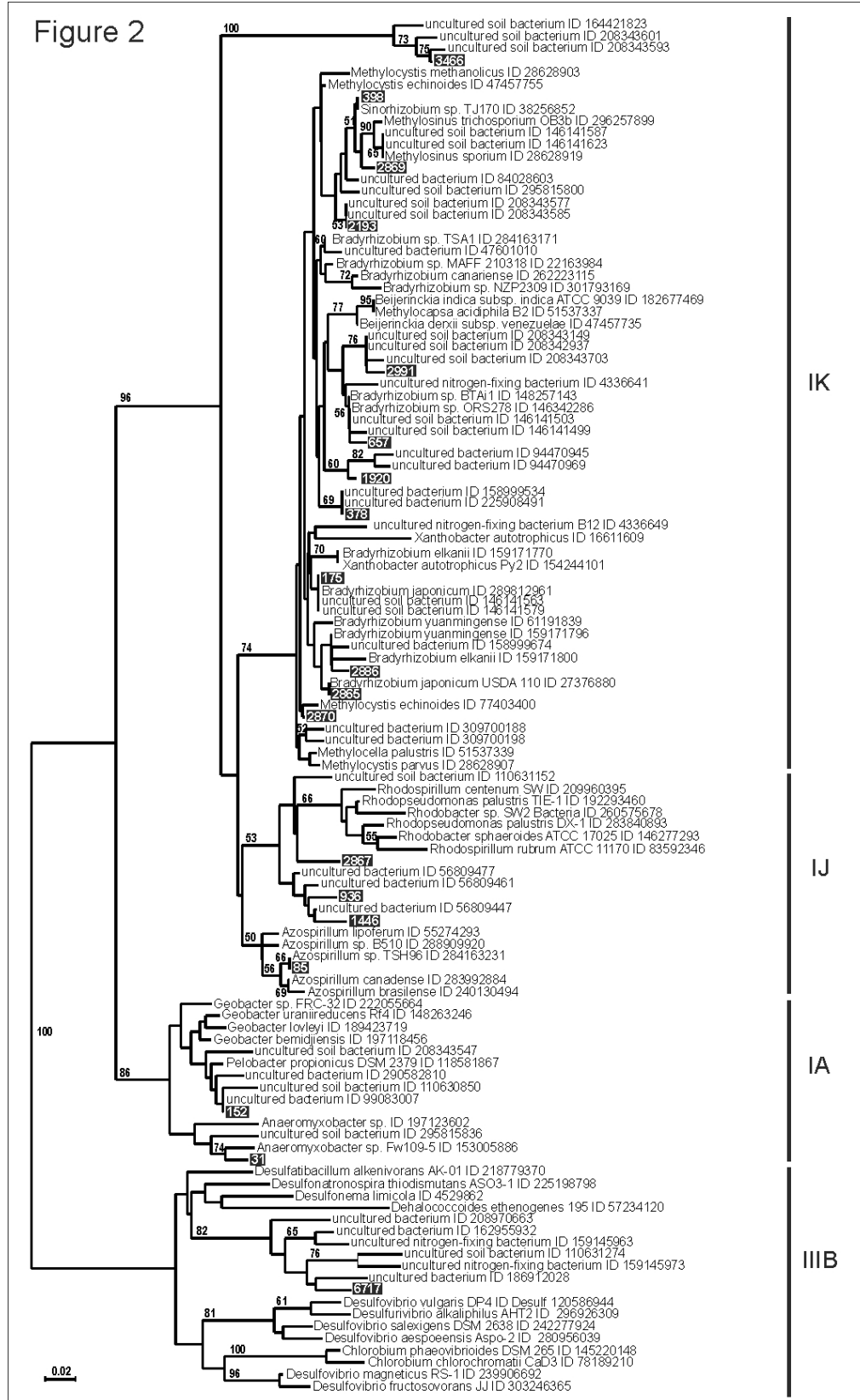
Figure 1



623

624

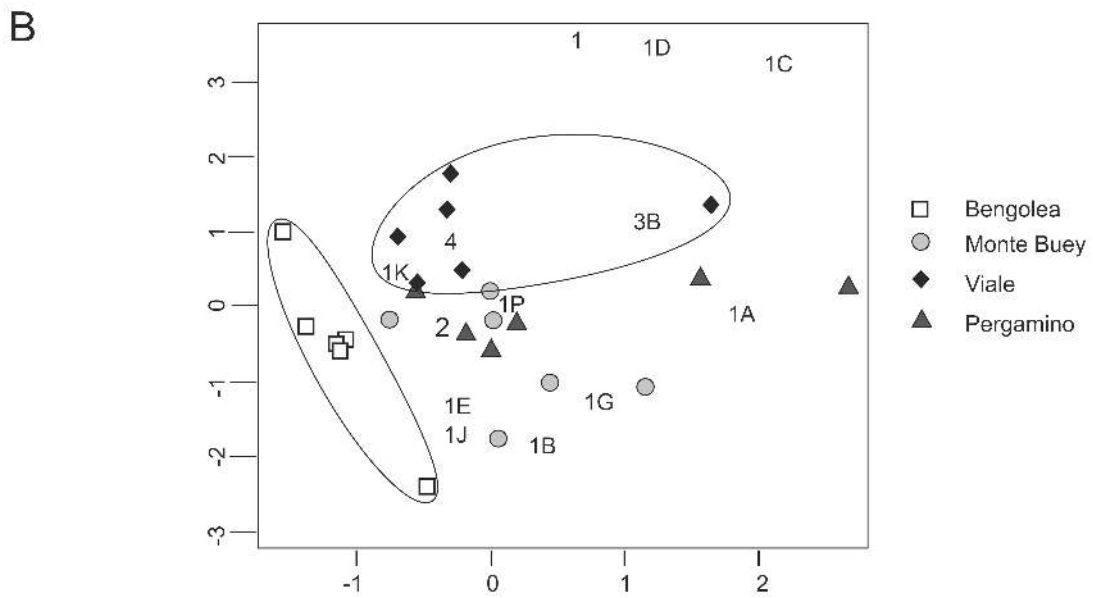
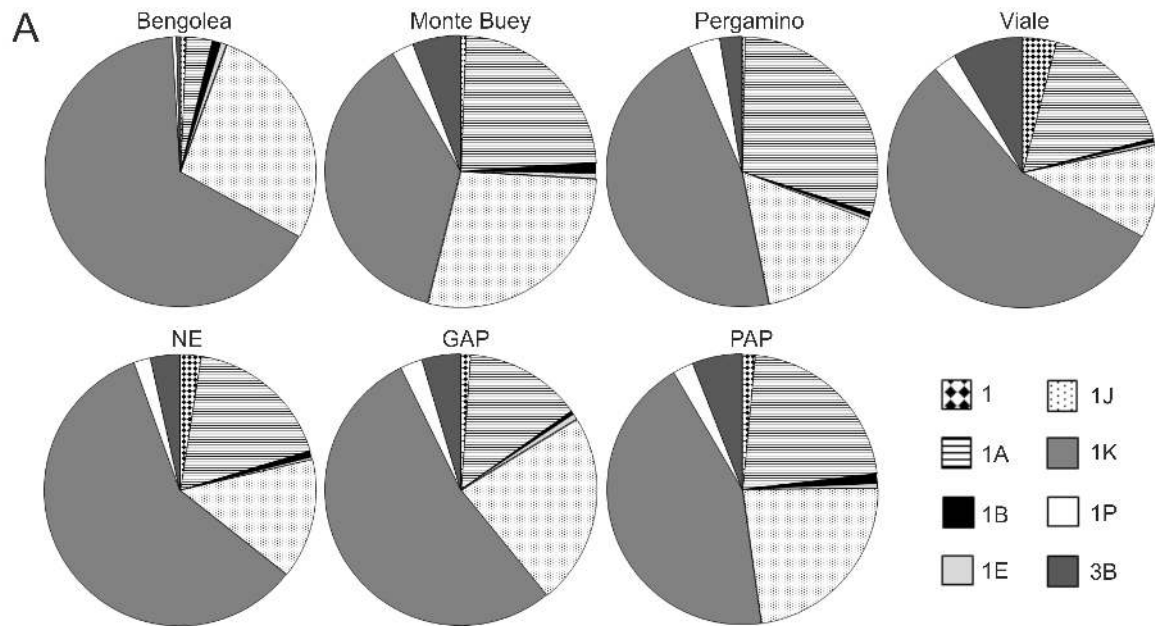
Figure 2



625

626

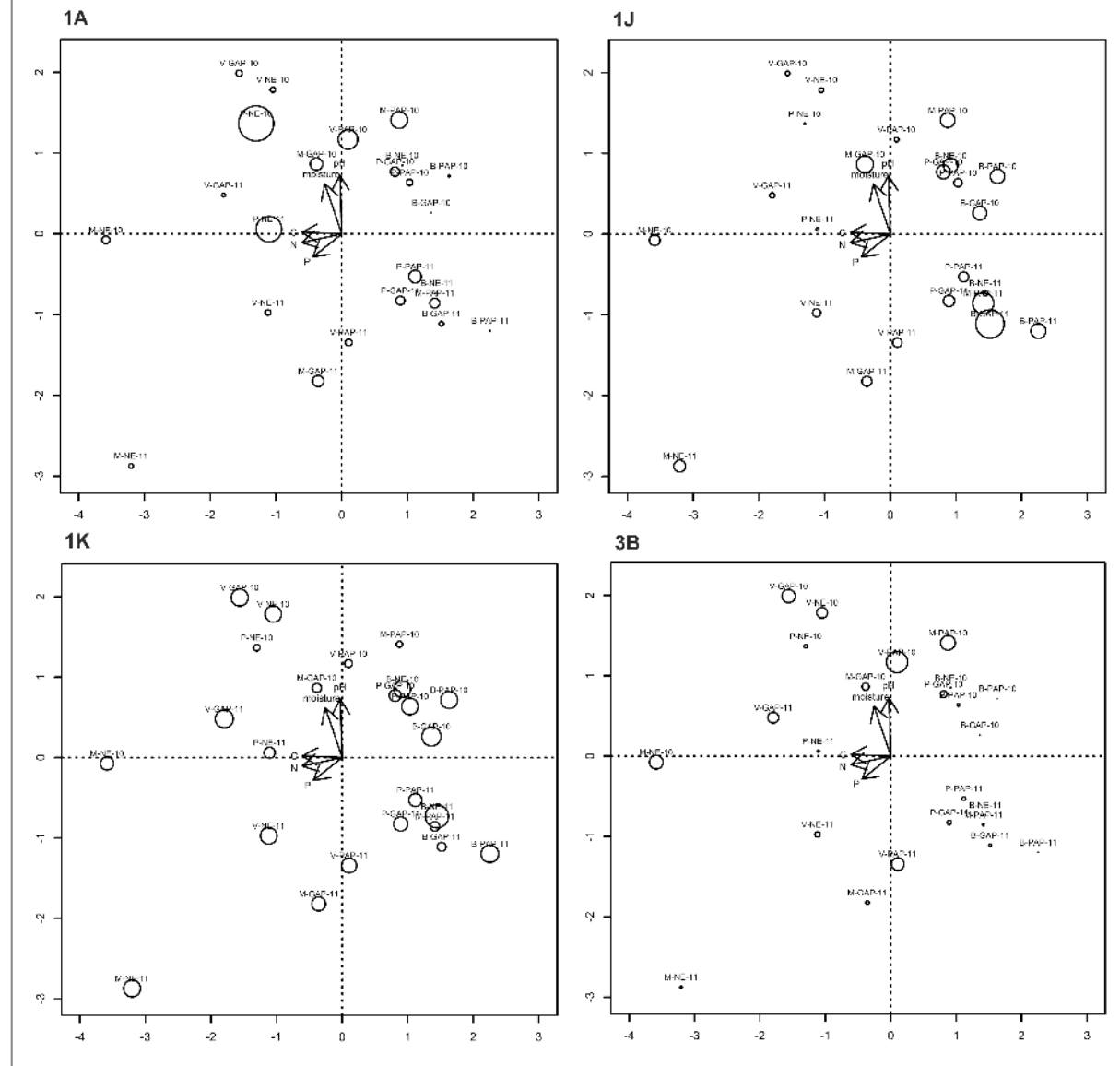
Figure 3



627

628

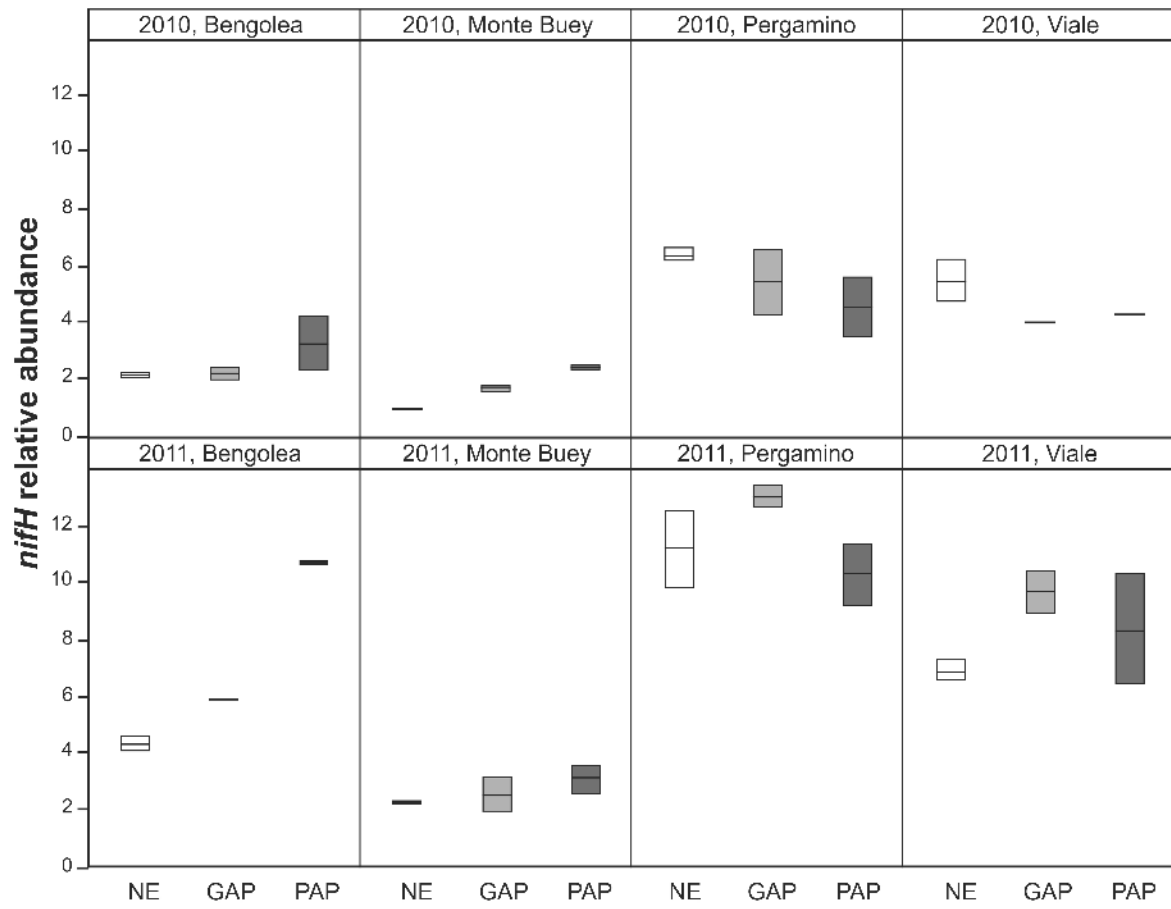
Figure 4



629

630

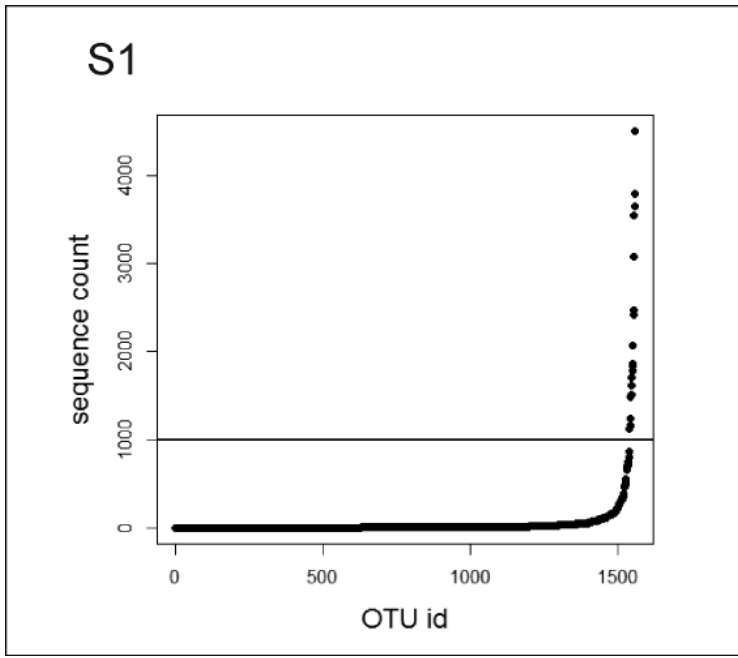
Figure 5



631

632

633



634

