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

Mónica M. Collavino,[1,2] **H James Tripp,**[3]
Ildiko E. Frank,[3] **María L. Vidoz,**[1]
Priscila A. Calderoli,[2] **Mariano Donato,**[4]
Jonathan P. Zehr[3] **and O Mario Aguilar**[2*]

[1]*Instituto de Botánica del Nordeste (IBONE), Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste-CONICET, Corrientes, Argentina.*
[2]*Instituto de Biotecnología y Biología Molecular (IBBM), Universidad Nacional de La Plata-CONICET, La Plata, Argentina.*
[3]*Department of Ocean Sciences, University of California, Santa Cruz, CA, USA.*
[4]*Laboratorio de Sistemática y Biología Evolutiva, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, Argentina.*

Summary

A dataset of 87 020 *nifH* reads and 16 782 unique *nifH* protein sequences obtained over 2 years from four locations across a gradient of agricultural soil types in Argentina were analysed to provide a detailed and comprehensive picture of the diversity, abundance and responses of the N₂-fixing community in relation to differences in soil chemistry and agricultural practices. Phylogenetic analysis revealed an expected high proportion of *Alphaproteobacteria*, *Betaproteobacteria* and *Deltaproteobacteria*, mainly relatives to *Bradyrhizobium* and *Methylosinus/Methylocystis*, but a surprising paucity of *Gammaproteobacteria*. Analysis of variance and stepwise regression modelling suggested location and treatment-specific influences of soil type on diazotrophic community composition and organic carbon concentrations on *nifH* diversity. *nifH* gene abundance, determined by quantitative real-time polymerase chain reaction, was higher in agricultural soils than in non-agricultural soils, and was influenced by soil chemistry under intensive crop rotation but not under monoculture. At some locations, sustainable

increased crop yields might be possible through the management of soil chemistry to improve the abundance and diversity of N₂-fixing bacteria.

Introduction

Biologically available, N is often a limiting nutrient in agricultural soil and other environments (Vitousek and Howarth, 1991). To this end, 4.32×10^5 tons of nitrogen fertilizers are used annually in the productive agricultural region of Argentina, which have recently been dedicated to soybean monoculture, due to financial considerations that also lead to minimal nutrient restoration (FAO, 2012). In an effort to mitigate the negative effects of monoculture on soil quality, principally increased erosion and decreased moisture retention, reduced-till agricultural practices have been widely introduced in Argentina (Viglizzo *et al.*, 2011). Studies have shown that while reduced-till practices can improve soil quality by reducing loss of nutrients and increasing moisture retention, it remains to be seen whether the beneficial effects are fully realized under monoculture and low nutrient replenishment (hereafter named poor no-till practices) as compared with intensive crop rotation and nutrient replacement (named good no-till practices) (Abid and Lal, 2009; Souza *et al.*, 2013). Specifically, it is not known if reduced-till practices under monoculture improve the diversity, abundance and community structure of N₂-fixing microbes in agricultural soil.

Biological N₂ fixation (BNF), which is the biological reduction of molecular N₂ gas to biologically available ammonium, accounts for approximately 128 million tons nitrogen per year and is considered the main route by which fixed nitrogen enters the biosphere by natural processes (Galloway *et al.*, 2004). BNF, catalysed only by Bacteria and Archaea, requires nitrogenase, an evolutionarily conserved protein in N₂-fixing microorganisms. The nitrogenase enzyme is composed of two multisubunit metallo-proteins. Component I contains the active site for N₂ reduction and is composed of two heterodimers encoded by the *nifD* and *nifK* genes. Component II, also known as dinitrogenase reductase, is composed of two identical subunits encoded by the *nifH* gene. The *nifH* gene sequence is highly conserved across the bacterial and archaeal domains; however, because of codon
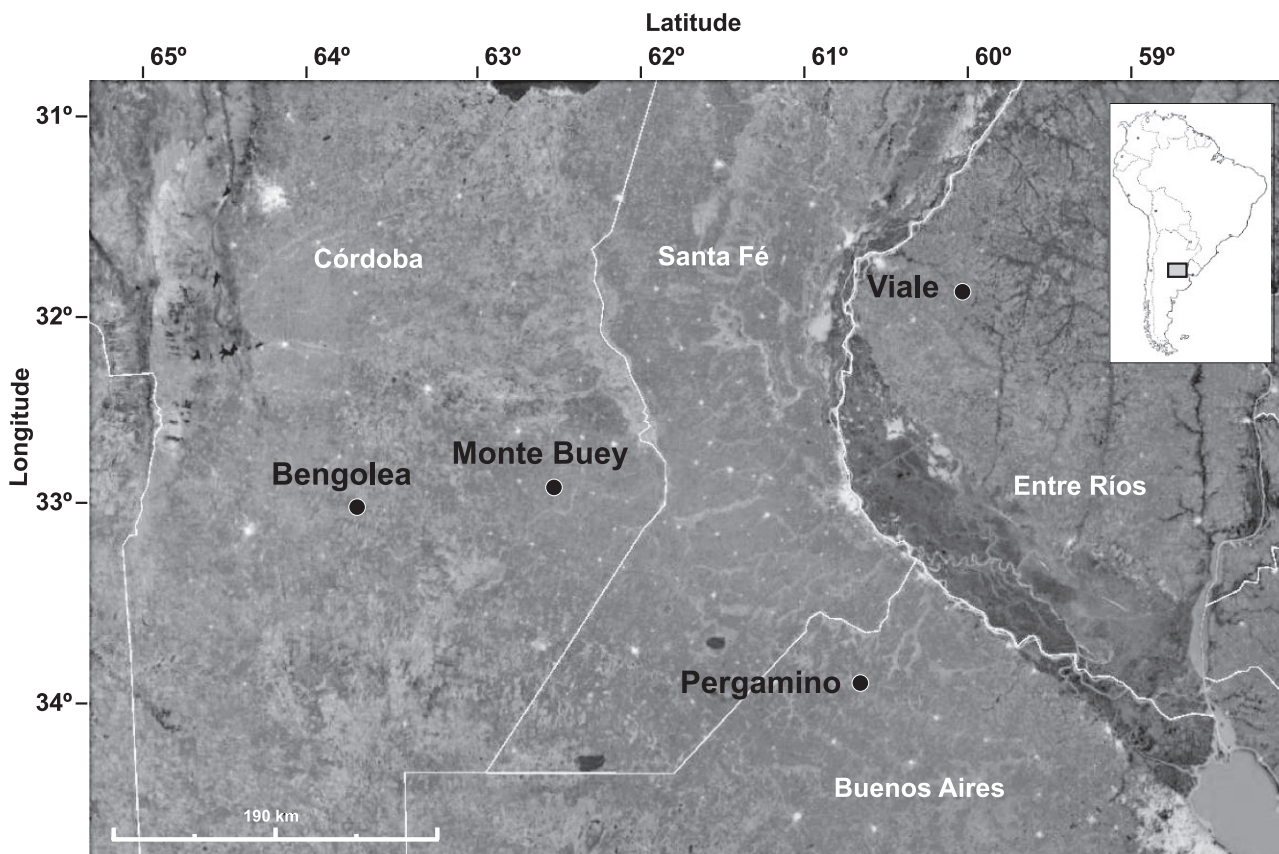
Latitude

Fig. 1. Map of the four sampling locations in the Argentinean study area. Soil samples were collected from four geographical locations: Bengolea and Monte Buey in Córdoba Province, Pergamino in Buenos Aires Province and Viale in Entre Ríos Province. Template map downloaded from http://www.google.com/earth/ (30 October 2013).

redundancy for most amino acids, the design of universal *nifH* primers requires a considerable degree of DNA sequence degeneracy. Up to date, several degenerated polymerase chain reaction (PCR) primers were developed and have been successfully used in culture-independent studies of microbial N$_2$-fixing communities in terrestrial and aquatic environments for almost 25 years (Zehr and McReynolds, 1989; Zehr *et al.*, 1998; Poly *et al.*, 2001b; Rosch *et al.*, 2002; Steward *et al.*, 2004; Bürgmann *et al.*, 2005; Izquierdo and Nüsslein, 2006; Farnelid *et al.*, 2011; Niederberger *et al.*, 2012). As a result, the database for nitrogenase genes (specifically the *nifH* gene) has become one of the largest non-ribosomal gene datasets from uncultivated microorganisms (Zehr *et al.*, 2003). This makes *nifH* gene an outstanding reference tool for studying N$_2$-fixing microbial communities in soil, most of whose members have not yet been cultured. The advent of pyrosequencing now affords the opportunity to study N$_2$-fixing soil microbial communities in depth, providing tens or hundreds of thousands of *nifH* amplicon sequences per sample.

Like the general survey of soil microbial communities using deep pyrosequencing of 16S rRNA gene (Figuerola

et al., 2012), this study is part of a larger effort to provide microbial indicators of the sustainability of no-till practices in four agricultural sites located across a west-east transect in the Argentine Central Pampas (Fig. 1). However, this study is a specific survey of only the N$_2$-fixing microbial communities in soil, using deep pyrosequencing of the *nifH* gene. The reader is referred to the Introduction and Methods sections of Figuerola *et al.* 2012 for a detailed description of the broader BIOSPAS project and the study sites selected.

The aim of this work was to examine the abundance, diversity and structure of diazotroph communities in a gradient of Argentinean agricultural soils, under good and poor no-till practices, using deep pyrosequencing-based analysis of the *nifH* gene.

Results

Proteobacteria are predominant in the diazotrophic community of Argentinean soils

A total of 87 020 *nifH* reads, comprising 16 782 unique protein sequences, were obtained from four Argentinean soils subjected to different management. At 98% sequence

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

| Cluster | Subcluster | Proportion (%) | | Distribution[a] (%) | Group | Orders[b] |
		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	*Firmicutes*	*Frankia*
	1E	0.5	2.1	96	*Firmicutes*	*Paenibacillus*
	1G	0.1	0.26	46	*Gamma*	*Pseudomonadales and Enterobacteriales*
	1J	20	19.6	100	*Alpha*	*Rhizobiales and Rhodospirillales*
	1 K	51	40.2	100	*Alpha and Beta*	*Rhizobiales and Burkholderiales*
	1 M	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*	*Rhizobiales* and *Opitutales*
	2	0.02	0.1	42	*Firmicutes*	*Paenibacillus*
III	3A	0.02	0.1	29	*Firmicutes*	*Clostridium*
	3B	5	8	100	*Delta and Verrumicrobia*	*Desulfovibrionales and Verrucomicrobiales*
IV	4	0.1	0.4	75	*Archaea*	*Methanococcales*

a. Presence across the 24 samples analysed (e.g. subcluster present in all samples shows a distribution of 100%).
b. Orders closest to the predominant sequences observed in the subcluster.

similarity, the 1558 operational taxonomic units (OTUs) for the entire study were imported into an ARB database of *nifH* sequences (Zehr *et al.*, 2003) and assigned to 17 subclusters of the four major *nifH* clusters previously defined by Zehr and colleagues (2003) (Table 1). The abundances varied widely among the OTUs; about 1000 OTUs were observed to be represented by no more than 10 sequences, whereas 1375 had less than 50 sequence counts, and only 19 OTUs were represented by more than 1000 sequences (Supporting Information Fig. S1). The proportion of OTU counts and sequence counts was similar in each subcluster (Table 1). Cluster I was the most abundant in the soil samples, particularly the subclusters 1K (40% of the total number of OTUs), 1A (21%) and 1J (20%) which are mainly composed of sequences related to *Rhizobiales* and *Burkholderiales*, *Anaeromyxobacter* and *Desulfuromonadales*, and *Rhizobiales* and *Rhodospirillales* respectively. The next largest group was subcluster 3B (8%), with sequences mainly related to *Desulfovibrionales* and *Verrucomicrobiales* (Table 1). No new *nifH* subclusters were identified in this study; however, it is noteworthy that new *nifH* sequences within subclusters were observed, i.e. 285 of the 1558 OTUs, most belonging to the subclusters 1A, 3B and 1J, have less than 95% similarity at the amino acid level as compared with the sequences present in the *nifH* database (data not shown).

Nineteen of the most abundant OTUs corresponding to about 48% of our *nifH* database were distributed among the four subclusters 1K (12 OTUs), 1J (4 OTUs), 1A (2 OTUs) and 3B (1 OTU) (Fig. 2). Most of 1K and 1J OTUs were related to alphaproteobacteria sequences. OTU 3466 of subcluster 1K (Fig. 2) diverged into a separated

group that contains only uncultivated sequences recovered from soil, marine and plant-associated environments (Lovell *et al.*, 2000; Zehr *et al.*, 2003; Moisander *et al.*, 2005; Hsu and Buckley, 2009). The remaining 1K OTUs clustered with sequences from the *Rhizobiales* order, which appeared to be highly represented in the database (33% of the total database). These OTUs were more closely related to sequences belonging to *Bradyrhizobium* (seven OTUs) and the *Methylocystis/Methylosinus* methanotrophic group (four OTUs), which represented 23.6% and 9.4% of the total database respectively. OTUs of subcluster 1J were related to *Rhodospirillales* sequences; one of them was related to *Azospirillum* while the remaining ones clustered with environmental sequences and were related to sequences from phototrophic purple non-sulfur bacteria belonging to *Rhizobiales* and *Rhodospirillales* orders.

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

Some major clusters and subclusters were underrepresented in our samples. Major clusters II and IV were represented by only a few OTUs, which is perhaps understandable given that cluster II contains the alternative (FeV and FeFe) nitrogenases and cluster IV is composed of *nifH* paralogues. By contrast, the scarcity of *Gammaproteobacteria* in subclusters 1O and 1M was
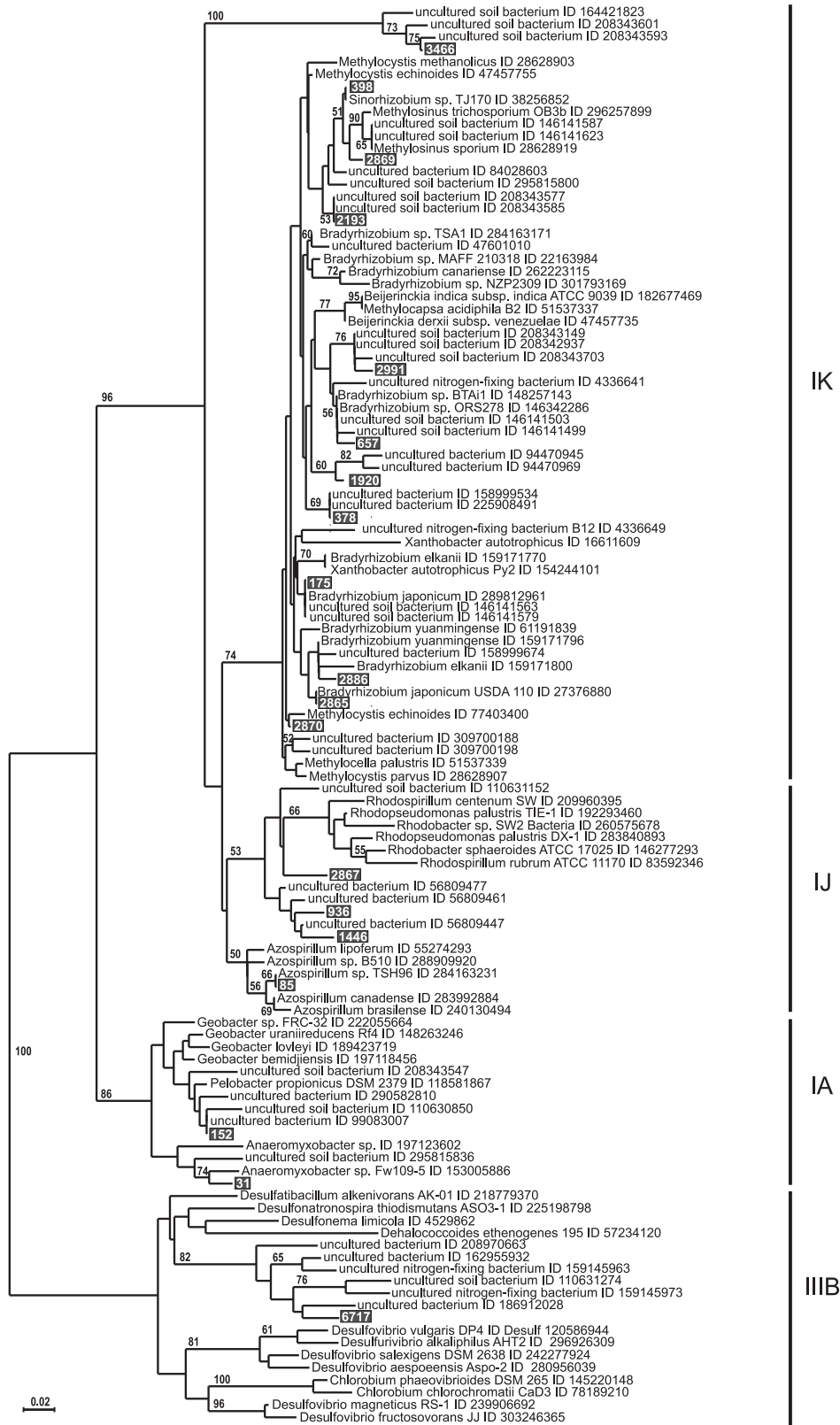
Fig. 2. Phylogenetic relationships of the nineteen largest *nifH* OTUs based on neighbour-joining analysis of partial amino acid sequences. The numbers in shaded boxes identify the largest OTUs observed in this study. Canonical *nifH* clusters are indicated to the right in the figure. The percentage of 500 bootstrap samples that supported each branch is shown. Bootstrap values below 50% are not shown.

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

		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
	Seqs$_n$	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

SeqN, number of sequences per sample; SeqN$_n$, normalized number of sequences per sample; Sobs, detected number of operational taxonomic units (OTUs) at 2% distance level; RC, relative coverage calculated as OTU number divided by estimated richness (Chao1 or ACE).

surprising, as well as the paucity of *Clostridiales*, *Pseudomonadales* and *Enterobacteriales* representatives from subgroups 3A and 1G (Table 1).

These data revealed that even when soils contain representatives of most of the diversity of *nifH* sequences found in natural environments, sequences belonging to *Alphaproteobacteria*, *Betaproteobacteria* and *Deltaproteobacteria* were predominant.

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

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

As shown in Table 2, the richness and diversity of the diazotrophic community were strongly affected by the geographical location of the sampling site, but not by the soil management practices. One-way analysis of variance (ANOVA) models confirmed significant differences among the four locations ($P = 3e^{-6}$ for richness and $P = 4e^{-5}$ for diversity) but not between years or treatments. Two-way ANOVA did not indicate significant interaction effects. Whereas the sandy soils of Bengolea displayed the lowest number of *nifH* OTUs and the lowest Shannon index (H′) values, Viale clay soils had the highest richness and diversity (Table 2).

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

Soil properties affect the composition of diazotrophic community

There were clear differences in the proportion of the eight predominant subclusters with respect to location. The westernmost (Bengolea) and easternmost (Viale) locations showed the largest differences compared with the two central locations (Monte Buey and Pergamino), likely due to contrasting soil types (Fig. 3A, top row). As noted before, subcluster 1K dominates all locations. The proportion of subcluster 1J sequences decreases from west (Viale) to east (Bengolea). As mentioned above, the lowest diversity among sites was observed at Bengolea, which consists mainly of subclusters 1K and 1J. Correspondence analysis plots show a more comprehensive and detailed view of the similarities between locations as defined by their subcluster proportions (Fig. 3B). Bengolea samples, which grouped on the left side of the figure, are unique in their high proportions of 1J and 1K, but also 1B, 1E and 2. Monte Buey and Pergamino samples share similar cluster proportions and, therefore, overlap in the figure. Viale samples separated along the second component were characterized by high proportions of 1, 1C, 1D and 3B clusters. Samples do not group according to treatment (data not shown). Chi-squared
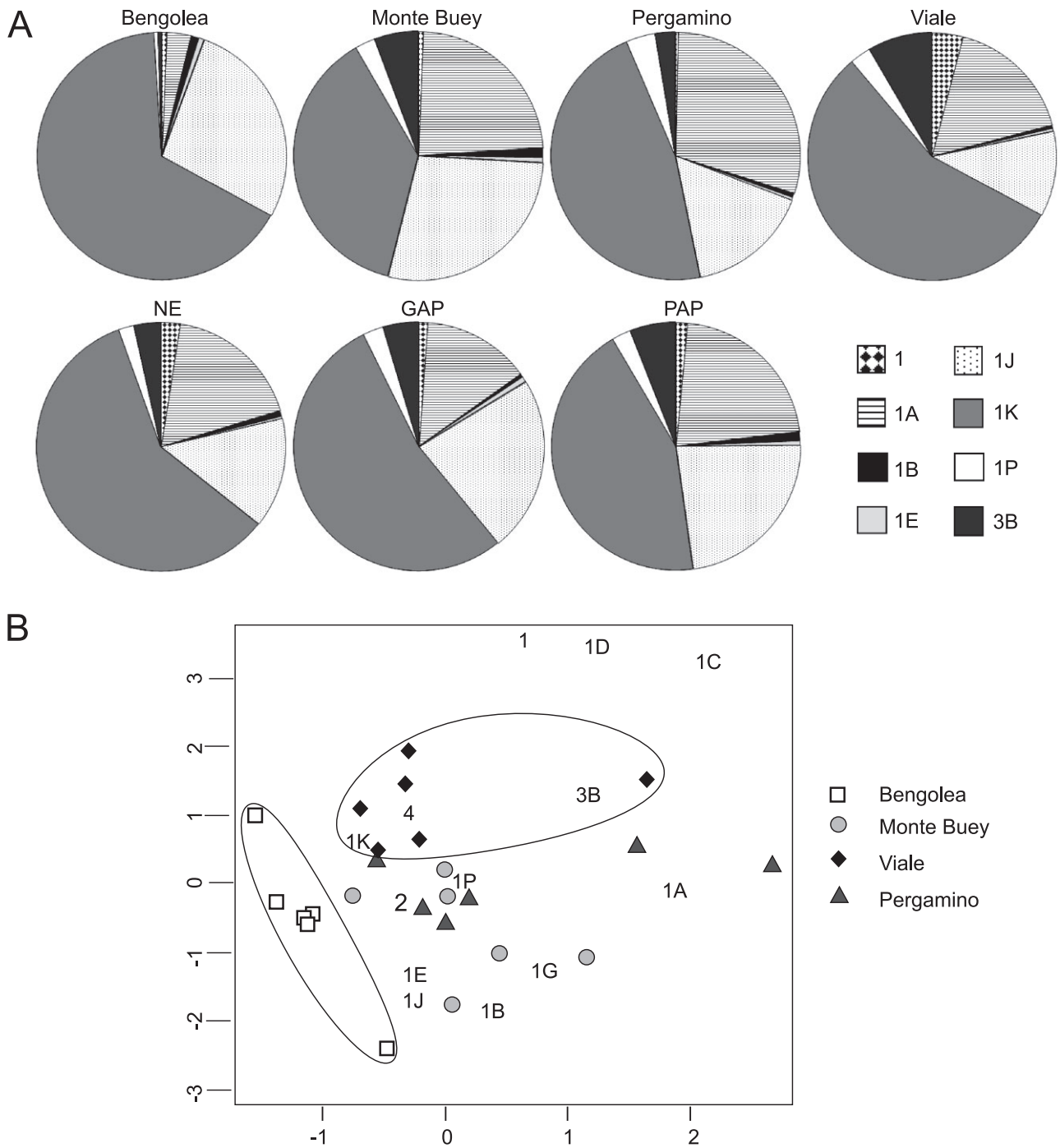
Fig. 3. Proportion of the largest subclusters across locations and treatments.
A. Pie charts show the proportion of the eight predominant subclusters with respect to geographical location (top row) and treatment (bottom row). Subclusters 1K, 1J and 1A dominate at each location and under each treatment (NE = natural environment; GAP = good no-till agricultural practices; PAP = poor no-till agricultural practices).
B. Correspondence analysis plot displays rows (samples) and columns (*nifH* subclusters) of a cross-tabulation of sequence counts. Closeness on the plot indicates similarity between samples in terms of subcluster proportions or similarity between subclusters in terms of their distribution across samples. Bengolea and Viale samples are enclosed by ellipses. The first component (horizontal axis) covers 51% and second component (vertical axis) covers additional 26% of variation.

independence test ($P = 2e-16$) suggests that locations and *nifH* subclusters are not independent, i.e. subcluster proportions vary significantly across locations.

Differences in the proportion of subclusters are subtler across treatments (Fig. 3A, bottom row). The pie chart shows that the proportion of 1J subcluster is notably high in good no-till agricultural practices (GAP) and poor no-till agricultural practices (PAP) samples. The trend in subcluster 1K proportion is natural environment (NE) > GAP > PAP, whereas subcluster 3B displays the opposite trend (Fig. 3A, bottom row).

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

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

Environment influences abundance of soil N$_2$-fixing bacteria

The relative abundance of N$_2$-fixing bacteria was determined by quantitative real-time PCR (qPCR). In general, the number of *nifH* copies varied significantly across sites (Monte Buey < Bengolea < Pergamino; $r^2 = 0.46$, $P = 4e^{-6}$) and were significantly higher in year 2011 than 2010 ($r^2 = 0.30$, $P = 6e^{-5}$) (Fig. 5). In addition to significant main effects of location and year, their interaction also proved to be significant ($P = 0.002$); this three-term ANOVA model describes most of the *nifH* abundance variation ($r^2 = 0.83$), revealing a major influence of the environment on abundance of soil diazotrophs. In contrast, the effect of treatment (PAP, GAP and NE) varied across locations and years; hence, when analysing all samples together, no relationship was found between *nifH* gene abundance and agricultural treatments. However, when abundance was examined separately at each site, year and treatment effects were found to be significant at Bengolea ($r^2 = 0.84$, $P = 0.001$, $P = 0.015$) and Monte Buey ($r^2 = 0.77$, $P = 0.008$, $P = 0.017$). At these two locations, *nifH* abundance was significantly higher under PAP and lower under NE treatment in both years (Fig. 5). This pattern is opposite to the C, P and N levels, i.e. *nifH*

abundance is lowest under NE treatment where C, P and N contents are highest.

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

The location and treatment dependence of soil chemistry also manifests in the dissimilar correlation patterns between abundance and soil parameters. For example, under NE and GAP treatment strong significant correlation was found between *nifH* abundance and P (-0.66 and -0.64), while pH ($r = -0.74$) and moisture ($r = -0.56$) were the main influencing parameters under PAP treatment. Considering locations, *nifH* abundance was negatively correlated with N ($r = -0.79$), P ($r = -0.88$) and moisture ($r = -0.84$) in Bengolea, whereas in Pergamino, the most significant correlation ($r = -0.93$) was with pH. It is notable that the significant correlations between individual soil chemistry parameters and *nifH* abundance were all negative.

Discussion

This study examined the community structure of N$_2$-fixing bacteria in soils, and the effects of soil type and management practice on the abundance and structure of the N$_2$-fixing communities in Argentinean agricultural soils. This database makes a significant contribution to the number of *nifH* sequences coming from soil, which in the recent report by Gaby and Buckley (2011) was estimated to be 3644 unique reads at the GenBank sequence database. The 16 782 unique protein sequences and 87 020 *nifH* reads we obtained, along with associated metadata, provide the most detailed and comprehensive picture of the diversity, abundance and environmental responses of the N$_2$-fixing community in Argentinean soils, which produce much of the world's supply of soybeans. We observed evidence of community members that responded to specific environmental factors, and some evidence of broad community effects related to soil type and soil chemistry.
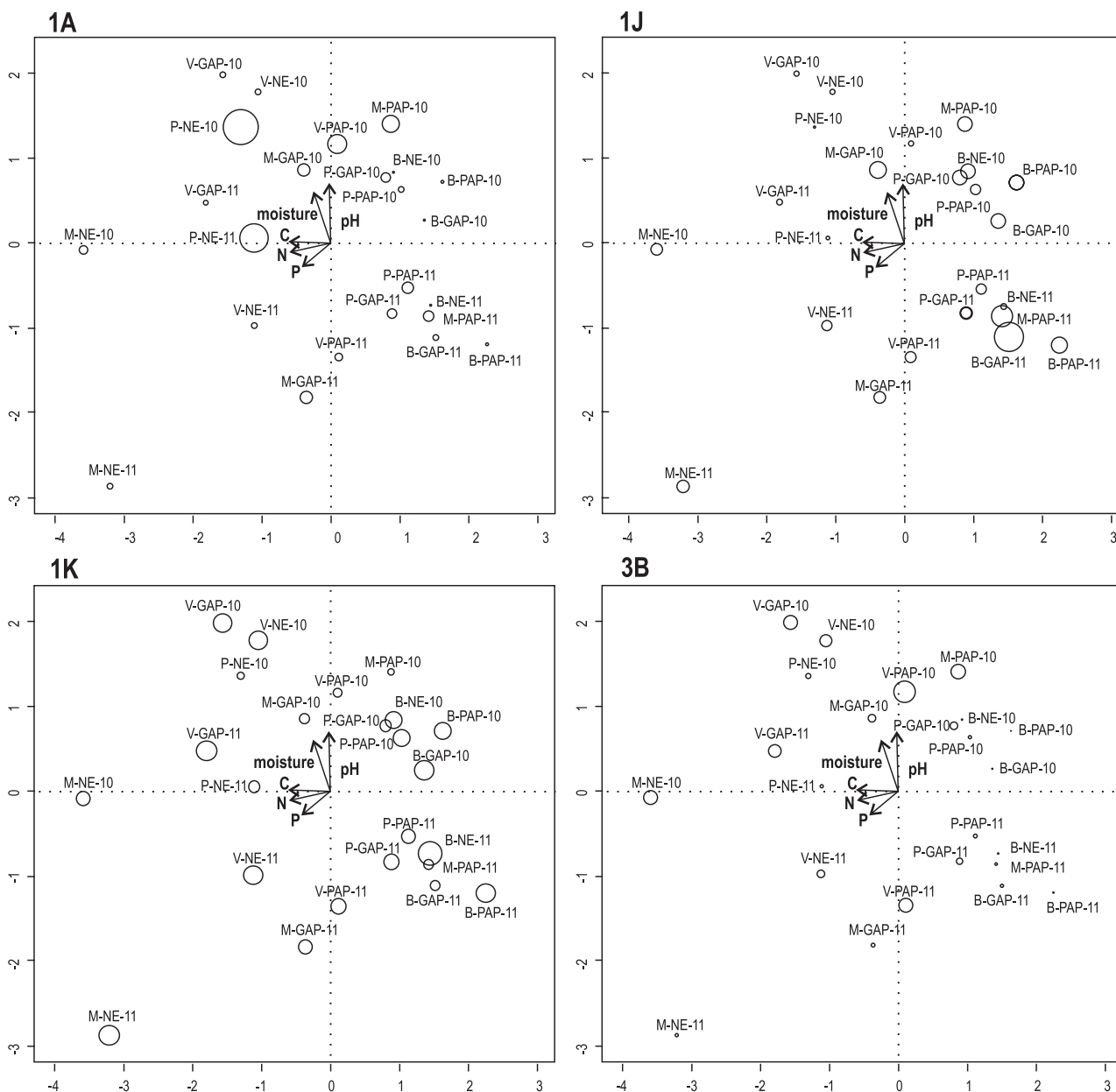
Fig. 4. Proportion of the four largest subclusters related to location, treatment, year and soil chemistry. Proportion is indicated by circle size at each sample point plotted on the first two principal components of the five soil parameters (represented by arrows). Component 1 (horizontal axis) covers 49% and component 2 covers 30% of the variance in soil chemistry. Each sample is labelled by its location (B = Bengolea, M = Monte Buey, P = Pergamino, V = Viale), treatment (NE = natural environment, GAP = good no-till agricultural practices, PAP = poor no-till agricultural practices), and year (10 = 2010, 11 = 2011). Circle size is scaled independently for each of the four subclusters.

It has been indicated that technical reproducibility and complete removal of sequencing artefacts are important issues in the analysis of sequence data generated by pyrosequencing (Zhou *et al.*, 2011; Knight *et al.*, 2012; Pinto and Raskin, 2012). In this regard, we chose to survey space and time more broadly, without technical replicates. However, to limit the creation of artefacts and to insure their removal, we pooled amplicons from multiple PCR reactions, removed low-quality sequences, chime-

ras and frameshifts, used an equal number of sequences in samples and removed less frequent OTUs (Schloss *et al.*, 2011).

We found that *Alphaproteobacteria* and *Betaproteobacteria* (subclusters 1K and 1J) and *Deltaproteobacteria* (subcluster 1A) were well represented in our samples as expected, but not *Gammaproteobacteria* (Table 1). The low representation of *nifH* sequences related to *Gammaproteobacteria* is striking considering that
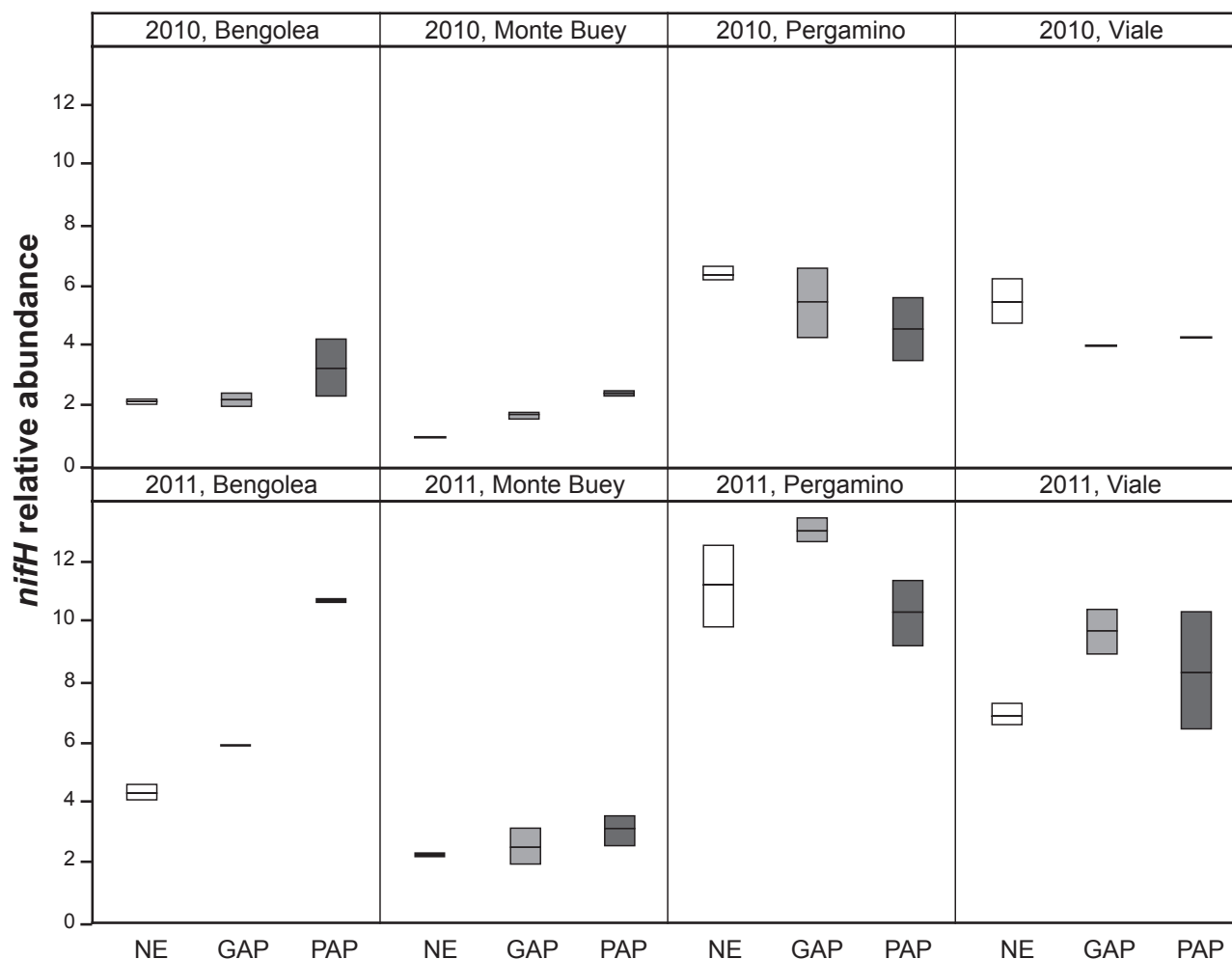
Fig. 5. Location, year and treatment effect on *nifH* gene relative abundance. Rows of panels correspond to years (2010, 2011) and columns to locations (Bengolea, Monte Buey, Pergamino, and Viale). Colours indicate soil management treatment: natural environment (NE) in white, good no-till agricultural practices (GAP) in grey and poor no-till agricultural practices (PAP) in dark gray. Boxes display the range of three biological and two technical replicates. Levels of *nifH* copies indicated as relative units are normalized to 16S rRNA gene abundance. A value of 1 was assigned to the lowest detected value (Monte Buey-NE-2010 sample).

several genera included in this group (*Pseudomonas*, *Enterobacter* and *Azotobacter*) were found to be common components of soil (Shu *et al.*, 2012; Liu *et al.*, 2012). As in any other approach based on PCR amplification, we cannot rule out that primer bias could explain the under-representation of *Gammaproteobacteria* sequences. However, the lack of *Gammaproteobacteria* has also been reported in the diazotrophic community of some tropical soils examined by both cultivation-dependent and -independent approaches (Izquierdo and Nüsslein, 2006).

The alphaproteobacterial sequences related to *Rhizobiales* order accounted for 33% of the total database. Two of the largest groups of sequences in subcluster 1K, groups 657 and 2991 (Fig. 2), were found in NE and GAP treatments, but poorly represented or not found in PAP practice soils (data not shown), indicating that they prefer

either natural conditions or GAP, such as corn inter-cropping. Interestingly, groups 657 and 2991 fell into the cluster that groups the photosynthetic *Bradyrhizobium* sp. strains, one of which (clone T1t015) has been reported to be positively affected by the presence of maize residues (Hsu and Buckley, 2009).

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

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

		Bengolea			Monte Buey			Pergamino			Viale		
		NE	GAP	PAP	NE	GAP	PAP	NE	GAP	PAP	NE	GAP	PAP
	Soil classification	Entic Haplustoll			Typic Argiudoll			Typic Argiudoll			Argic Pelludert		
	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

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

A significant difference in diazotrophic diversity was observed across the four locations but not among management treatments. In a companion study using the same set of soil samples, diversity of the whole bacterial community did not vary by location or management treatment (Figuerola *et al.*, 2012), suggesting that diversity variation is a distinctive feature of the N$_2$-fixing community. *nifH* diversity and richness levels were consistently low in the sandy soil from Bengolea, whereas the clay soil from Viale displayed the highest diversity (Table 2). Based on regression analysis, the differences of diversity across sites appear to be mainly associated with levels of soil organic carbon. This result is not surprising, given that heterotrophic N$_2$ fixers often metabolize organic matter both to fix N$_2$ and to maintain high respiration rates to avoid O$_2$ inactivation of nitrogenase (Hill, 1992).

Interestingly, *nifH* abundance in agricultural soils was higher than in pristine non-agricultural soils, suggesting

that land use or presence of crops somehow promote diazotroph populations. Therefore, our study provided no evidence that no-till production systems either with intense crop rotation or monoculture practices negatively affect the level of the potential N$_2$ fixation community of soils. It could be that perturbation of soil by cropping promotes increases in abundances unlike the pristine environment which eventually reaches a steady state.

Species abundance, diversity and function are important components of the sustainable agriculture systems. Within this ecological framework, we demonstrated at some locations potential for management of soil chemistry to improve the abundance and diversity of N$_2$-fixing bacteria. Surprisingly, at locations where good agricultural practices are being followed, it could be beneficial from the standpoint of promoting N$_2$ fixation not to be overly ambitious with nutrient replacement, as well as with practices that affect soil pH and water content. This conclusion is supported by the negative correlations we sometimes found between *nifH* abundance and nutrient levels, pH and soil moisture. Considering that higher *nifH* abundances are not always associated with higher N$_2$ fixation rates or diversity, additional research on *nif* expression and N$_2$ fixation rates would be useful in order to compare diazotroph community structure with N$_2$ fixation activity.

Experimental procedures

Sample collection and soil characterization

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

loam (Typic Argiudoll) in Pergamino and Monte Buey and a silty clay (Argic Pelludert) in Viale (Table 3).

As previously described by Figuerola and colleagues (2012), three treatments were defined at each of the four sampling geographical locations. GAP treatment is characterized by intensive crop rotation (soybean-maize), nutrient replacement, and low agrochemical use. PAP treatment is characterized by crop monoculture (soybean), low nutrient replacement, high agrochemical use and lower yields. Grasslands uncultivated for 30 or more years were considered NE treatments.

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

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

DNA extraction

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

Quantification of nifH

The relative abundance of the *nifH* gene was quantified via qPCR. The qPCR reactions contained 10 ng of soil DNA,

1.2 μl of each primer (5 pM) and 10 μl of 2× SYBR Green iCycler iQ mixture (Bio-Rad) and water for 20 μl final reaction volume. The reaction was carried out on an Applied Biosystem 7500 real-time PCR system (Applied Biosystems) using a programme of 95°C for 10 min followed by 40 cycles consisting 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 cycle. 16S rRNA gene abundances were used to normalize values between the different samples. Relative quantities were calculated using GENORM (http://medgen.ugent.be/~jvdesomp/genorm/). The following specific primers were used: PolF-PolR for *nifH* (Poly *et al.*, 2001a) and 338F-518R for bacterial 16S rRNA genes (Park and Crowley, 2005). All qPCR reactions were run in duplicate with DNA extracted from replicates soil samples. For each run, the melting curve was analysed to ensure specific assessment of *nifH* gene.

Generation of bar-coded nifH gene libraries

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

Pyrosequencing data analysis

Sequences shorter than the threshold of 300 bp in length or displaying ambiguities were removed from the 24 pyrosequencing-derived datasets. The remaining reads were translated in all six reading frames and compared with a *nifH* reference database (http://pmc.ucsc.edu/~wwwzehr/research/database/) using BLASTX from INQUIRY software (Bioteam). A cut-off of 1e⁻¹⁰ for the *E*-value was set to remove non-specific blast hits. Putative frame shifts were detected and removed with the FrameBot tool (RDP, Ribosomal Database Project) using a profile hidden Markov model (HMM) (Zehr *et al.*, 2003) as reference set. The same HMM was used to align the *nifH* pyrosequences with the HMMALIGN programme (HMMER 2.3.2). The resulting alignment was imported into ARB (Ludwig *et al.*, 2004), evaluated by eye and subjected to minor manual corrections. The regions of the primers were trimmed and incomplete sequences were removed. High-quality sequences were subsequently assessed for chimeras using the UCHIME algorithm (Edgar *et al.*, 2011). Following that, the nucleotide sequences were

clustered into OTUs using CD-HIT SUITE programme (Huang *et al.*, 2010) with an OTU threshold value of 98% sequence identity at the DNA level. OTU representative sequences were analysed with *de novo* mode of UCHIME using its relative abundance data. All putative chimeras detected using UCHIME default settings were eliminated. Finally, the sequences in the database were clustered with OTUs defined at 98% amino acid sequence similarity. The conservative OTU cut-off used was defined considering that protein-encoding genes of strains of a given species generally have a high sequence similarity (Konstantinidis and Tiedje, 2005). In an attempt to further reduce potential pyrosequencing errors, OTUs represented by less than three sequences were excluded from the database. The relative abundance of sequences of the remaining OTUs was normalized using the subsampling-based method described in mothur (http://www.mothur.org/wiki/Normalize.shared), prior to comparative analyses.

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

The data set was deposited in the NCBI-SRA (Sequence Read Archive) with the submission Accession No. SRP029166.

Statistical data analyses

Data were analysed using R (R Core Team, 2013) and by Minitab 16 Statistical Software, 2010. Richness (Chao1, S_{Chao1} and ACE, S_{ACE}), Shannon diversity (H'), and rarefaction were calculated with the vegan package in R (Oksanen *et al.*, 2013) using the diversity, estimate, rarecurve and spec-number functions.

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

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

Variation in community composition across locations and treatments was graphically explored by pie charts created in Minitab. The pie charts were created by adding the corresponding subsets (locations or treatments). Samples were subsetted by location and the sum of six samples (three treatments and 2 years) was calculated at each location for each cluster. Similarly, samples were subsetted by treatment, and at each treatment, the sum of eight samples (four locations and 2 years) was calculated.

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

References

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

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

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

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

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

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

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

FAO (2012) Nitrogenous fertilizers. URL http://faostat3.fao.org/home/index.html.

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

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

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

Hill, S. (1992) Physiology of nitrogen fixation in free-living heterotrophs. In *Biological Nitrogen Fixation*. Stacey G., Burris R.H., and Evans J. (eds). New York: Chapman and Hall, pp. 87–129.

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

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

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

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

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

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

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

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

Minitab 16 Statistical Software (2010) Computer software. State College, PA: Minitab. URL http://www.minitab.com.

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

Nenadic, O., and Greenacre, M. (2007) Correspondence Analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Softw* **20:** 1–13.

Niederberger, T.D., Sohm, J.A., Tirindelli, J., Gunderson, T., Capone, D.G., Carpenter, E.J., and Cary, S.C. (2012) Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys. *FEMS Microbiol Ecol* **82:** 376–390.

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

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

Pinto, A.J., and Raskin, L. (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS ONE* **7:** e43093.

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

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

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: http://www.R-project.org.

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

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

Shu, W., Pablo, G.P., Jun, Y., and Danfeng, H. (2012) Abundance and diversity of nitrogen-fixing bacteria in rhizosphere and bulk paddy soil under different duration of organic management. *WM Microbiol Biotechnol* **28:** 493–503.

Souza, R.C., Cantão, M.E., Ribeiro Vasconcelos, A.T., Nogueira, M.A., and Hungria, M. (2013) Soil metagenomics reveals differences under conventional and no-tillage with crop rotation or succession. *Appl Soil Ecol* **72:** 49–61.

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

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

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

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

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

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

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

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Abundance of 1558 OTUs. Abundance is quantified as sequence count in each OTU. The hockey-stick curve indicates that only 19 OTUs are represented by more than 1000 sequences.

Fig. S2. Rarefaction curves grouped by location. Number of sequences (horizontal axis) plotted against number of OTUs (vertical axis). Each curve corresponds to a sample labelled by its location (B = Bengolea, M = Monte Buey, P = Pergamino, V = Viale), treatment (NE = natural environment, GAP = good no-till agricultural practices, PAP = poor no-till agricultural practices), and year (10 = 2010, 11 = 2011).

Table S1. Soil parameters that explain significant variation in the proportion of *nif*H subclusters.