MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 2988–2999 doi: 10.1111/mec.12786

Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities

FABIANA S. PAULA,*†‡ JORGE L. M. RODRIGUES,‡ JIZHONG ZHOU,§¶** LIYOU WU,§ REBECCA C. MUELLER,†† BABUR S. MIRZA,‡ BRENDAN J. M. BOHANNAN,†† KLAUS NÜSSLEIN,‡‡ YE DENG,§ JAMES M. TIEDJE§§ and VIVIAN H. PELLIZARI*

*Instituto Oceanografico, Universidade de Sao Paulo, 05508-120 Sao Paulo, Brazil, †Instituto de Ciencias Biomedicas, Universidade de Sao Paulo, 05508-900 Sao Paulo, Brazil, ‡Department of Biology, University of Texas, Arlington, TX 76019, USA, §Institute for Environmental Genomics, Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA, ¶Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, **State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, 100084 Beijing, China, ††Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA, ‡‡Department of Microbiology, University of Massachusetts, Amherst, MA 01003, USA, §§Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824, USA*

Abstract

Land use change in the Amazon rainforest alters the taxonomic structure of soil microbial communities, but whether it alters their functional gene composition is unknown. We used the highly parallel microarray technology GeoChip 4.0, which contains 83 992 probes specific for genes linked nutrient cycling and other processes, to evaluate how the diversity, abundance and similarity of the targeted genes responded to forest-to-pasture conversion. We also evaluated whether these parameters were reestablished with secondary forest growth. A spatially nested scheme was employed to sample a primary forest, two pastures (6 and 38 years old) and a secondary forest. Both pastures had significantly lower microbial functional genes richness and diversity when compared to the primary forest. Gene composition and turnover were also significantly modified with land use change. Edaphic traits associated with soil acidity, iron availability, soil texture and organic matter concentration were correlated with these gene changes. Although primary and secondary forests showed similar functional gene richness and diversity, there were differences in gene composition and turnover, suggesting that community recovery was not complete in the secondary forest. Gene association analysis revealed that response to ecosystem conversion varied significantly across functional gene groups, with genes linked to carbon and nitrogen cycling mostly altered. This study indicates that diversity and abundance of numerous environmentally important genes respond to forest-to-pasture conversion and hence have the potential to affect the related processes at an ecosystem scale.

Keywords: association index, functional gene arrays, GeoChip, soil microbes/tropical forest

Received 15 October 2013; revision received 30 April 2014; accepted 30 April 2014

Introduction

The Amazon rainforest, known for its rich biological diversity, has been under constant threat of deforestation as a consequence of intense agricultural development (Fearnside 1999; Feeley & Silman 2009). It is estimated that 20% of the original forest has already been lost due to anthropogenic activities (Barreto *et al.* 2005). This is of particular concern because tropical forests perform a number of essential global-scale functions, including the maintenance of freshwater supplies,

Correspondence: Fabiana S. Paula, Fax: +55-11-30916607; E-mail: fabianaspaula@gmail.com

carbon sequestration and the stabilization of biogeochemical processes (Fearnside 1999; Davidson *et al.* 2012). Land use has been identified as one of the main drivers of biodiversity alteration (Haines-Young 2009) and is predicted to be the most significant factor for biodiversity losses in tropical areas (Sala *et al.* 2000). Numerous studies have documented biodiversity losses for plants and animals in tropical forests (Fearnside 1999; Feeley & Silman 2009), and recent culture-independent community analyses based on ribosomal genes have found that changes in Amazon land use also alter soil bacterial (Borneman & Triplett 1997; Cenciani *et al.* 2009; Jesus *et al.* 2009; Rodrigues *et al.* 2013), fungal (Mueller *et al.* 2014) and archaeal (Navarrete *et al.* 2011) community composition.

An increasing number of studies have investigated how changes in microbial community structure and diversity affect soil biological processes following a sustained environmental impact (Loreau *et al.* 2001; Griffiths *et al.* 2004; Fitter *et al.* 2005). Changes in microbial community composition are frequently linked to alterations in the rates of ecosystem processes (Allison & Martiny 2008), but ecosystem process rates are more likely a consequence of the functional traits contained in a community than overall species richness and diversity (Griffiths & Philippot 2013). It has been suggested that the magnitude of some microbial processes may be affected by the abundance (He *et al.* 2007; Hallin *et al.* 2009) and the diversity (Philippot *et al.* 2013) of related functional genes, supporting the use of these parameters to assess community functional potential.

In the Amazon region, there is strong evidence for alterations in C and N cycling processes due to land use change (Steudler *et al.* 1996; Neill *et al.* 1997b, 2005), but studies on the response of microbial functional diversity to these changes have been limited to a few genes (Taketani & Tsai 2010; Navarrete *et al.* 2011; Mirza *et al.* 2014). A comprehensive evaluation of the microbial functional gene diversity harboured by Amazon forest soils is needed if we are to understand how land use change alters the metabolic potential of this ecosystem.

In this work, we evaluated the impact of land use change on the functional diversity of microbes in Amazon rainforest soils, using the diversity, abundance and similarity of environmentally important functional (i.e. protein coding) genes, including genes linked to nutrient cycling, as indicators of microbial functional potential. We used the GeoChip microarray (Hazen *et al.* 2010; Tu *et al.* 2014) to evaluate how functional gene groups respond to forest-to-pasture conversion and whether there is reestablishment of these groups with secondary forest growth. This functional microarray targets genes involved in different microbial processes, including biogeochemical cycles, organic compound degradation, antibiotic resistance and stress. It has been considered a comprehensive and effective tool for the evaluation of microbial functional diversity and responses to environmental impacts (Liang *et al.* 2009; Reeve *et al.* 2010). We sought to understand whether microbial functional gene diversity responses were comparable with those reported for microbial taxonomic diversity in Amazon soils under land use changes. Finally, we applied an ecological index for species–environment association (De Caceres *et al.* 2010) to identify functional groups significantly linked to land use type.

Material and methods

Site description and soil sampling

The study was performed at the Amazon Rainforest Microbial Observatory (ARMO; Rodrigues *et al.* 2013), located at the Fazenda Nova Vida (10°10′5″S and 62°49′27″W), a 22 000-ha cattle ranch in the State of Rondônia, Brazil (details regarding the sampling area can be found in the Supporting Information). Sampling occurred at the end of the rainy season, March 2010, for the following sites: a primary forest (F), two pastures that had been continuously managed for 6 (P6) and 38 (P38) years and a 13-year-old secondary forest (S), which was established in 1997 after pasture abandonment (Table S1 and Figure S1). All pastures were established following the same procedures (details of pasture establishment and management practices are described in the Supporting Information). A 100-m^2 transect was placed at each site, and nested transects of 10 m, 1 m and 0.1 m were made sharing the same point of origin, for a total of 12 sampling points, as described in Rodrigues *et al.* (2013). After the removal of the litter layer, a 5-cm diameter soil core was collected from 0 to 10 cm depth, homogenized and passed through a 2-mm mesh sieve. Samples for total DNA extraction were kept at −80°C, while samples for physicochemical analysis were stored at 4°C.

Total carbon and nitrogen were determined with auto analyzer LECO Truspec CN (St. Joseph, MI, USA) at the Centro de Energia Nuclear na Agricultura, University of Sao Paulo, Brazil. Elemental concentrations and soil fertility parameters were analysed according to the methods described by Van Raij *et al.* (2001), and soil granulometry was determined according to Camargo *et al.* (1986).

Soil DNA extraction, purification and labelling

DNA was extracted from 250 mg of soil using the PowerLyzer PowerSoil DNA isolation kit (MO BIO Laboratories Inc., Carlsbad, CA, USA), according to the

manufacturer's instructions, but with a modified lysing time of eight minutes. Five DNA extractions were performed per soil sample, combined and concentrated with the Genomic DNA Clean and Concentrator kit (Zymo Research Corporation, Irvine, CA, USA). Concentrated DNA was re-suspended in 20 μl of PCR grade water, and purity was assessed using a spectrophotometer (Thermo Scientific, Wilmington, DE, USA) set for determining absorbance at the following wavelengths: 230, 260, 280 and 320 nm. DNA concentration was measured with the Quant-iT PicoGreen kit (Molecular Probes/Invitrogen, Carlsbad, CA, USA). The DNA extraction yield for forest and pasture samples had an average of 32.99 ng/μl (± 1.58) and 33.86 ng/μl (± 1.60), respectively. All 48 of the DNA samples (12 per sampling site) were stored at −20°C until use.

Purified DNA was labelled with the fluorescent dye Cy-3, according to Wu *et al.* (2006). Briefly, DNA (1.5 μg) was mixed with random primers (300 ng/μl), denatured at 99.9°C for 5 min and immediately chilled on ice. A solution containing 5 mM dAGC-TP, 2.5 mM dTTP, Klenow fragment (40 U) and Cy-3 dUTP (25 nM) was added to the denatured DNA, and the reaction volume was adjusted to 50 μl with H_2O. The labelling solution was incubated at 37°C for 6 h, followed by 3 min at 95°C. The labelled DNA was purified with the QIAquick Kit (Qiagen, Valencia, CA, USA), and dye incorporation was confirmed with a NanoDrop spectrophotometer. Labelled samples were vacuum-dried and stored at −20°C until hybridization.

Microarray hybridization, scanning and image processing

The GeoChip 4.0 (Tu *et al.* 2014) was used to evaluate the microbial functional diversity in the soil samples. This new generation functional gene array has been successfully used to characterize microbial communities in a variety of habitats (Hazen *et al.* 2010; Lu *et al.* 2012; Chan *et al.* 2013). It contains 83 992 gene probes targeting 410 gene families. Here, gene refers to each unique sequence targeted by GeoChip 4.0, and a gene family comprises all gene sequences assigned with the same name (e.g. named *nosZ*) and coding for the same class of proteins. The genes are related to the biogeochemical cycles of carbon, nitrogen, phosphorus and sulphur, along with metal resistance, organic contaminant degradation, antibiotic resistance and other microbial genes of known function. The probes were designed to target Bacteria, Archaea, fungi and viruses (bacteriophage) gene sequences. Additional information can be found in Table S2.

Prior to hybridization, labelled DNA was re-hydrated with 10 μl of hybridization solution master mix, containing 40% formamide (J.T. Baker, Philipsburg, NJ, USA), 5X SSC buffer (Ambion, Austin, TX, USA), 0.1% SDS, 2.68 μl of sample tracking controls (Cy3-labelled 48-mer oligonucleotide targeting control probes on GeoChip array, Roche NimbleGen, Madison, WI, USA), 3 μl of Cy5-labelled universal standard (designed at the Institute for Environmental Genomics for normalization procedures) and alignment oligonucleotide Cy3 (0.17 μl) and Cy5 (0.28 μl) (Roche NimbleGen, Madison, WI, USA). The mix was denatured at 95°C for 5 min and kept at 50°C until hybridization, which was performed at 42°C, for 16 h at a MAUI Hybridization station (BioMicro System, Madison, WI, USA). Microarray slides were washed with the NimbleGen Wash Buffer kit according to the manufacturer's instructions.

Slides were scanned with the NimbleGen MS200 (Roche NimbleGen, Madison, WI, USA). Spot alignment and signal intensity determination were performed with the NIMBLESCAN software v2.5. The raw data were submitted to the Microarray Data Manager at the Institute for Environmental Genomics website (http://ieg.ou.edu/microarray/) using the following parameters: (i) spots with signal intensity lower than 1000 or with a signal-to-noise ratio (SNR) less than 2.0 were removed as poor-quality spots (SNR = spot signal intensity − background signal average/SD background); (ii) normalization was performed with the signal intensity of each spot being divided by the average of universal standard spots and divided by the average of the signal intensities in each sample.

Statistical analyses

Total gene and gene family numbers were used to calculate soil microbial functional gene richness. Diversity of detected genes was calculated using the Shannon (H′) diversity index based on the frequency and abundance of all gene sequences detected (regardless of family). One-way ANOVA, followed by Tukey's test, was employed to verify the effect of land use on gene richness and diversity. The similarities in gene composition across the samples were visualized by nonmetric multidimensional scaling ordination and tested by analysis of similarity (ANOSIM) using Bray–Curtis index (Clarke 1993). The total gene abundance in each functional group was ordered in a heat map and normalized by functional group, with mean and standard deviation equal to one. The samples were then clustered using Euclidean distances with the complete linkage method, based on their colour intensity profiles (Kindt & Coe 2005; R Development Core Team 2010).

The distance decay of gene similarity, a measure of turnover, within each sampling site was evaluated by Mantel test (Legendre & Legendre 2012), employing the

Sorensen index, and the significance was calculated with 999 permutations. Multiple regression on distance matrices (MRM; Tuomisto *et al.* 2003; Lichstein 2007) was employed to calculate the contributions of geographical and environmental distances on gene similarity (detailed description of these methods can be found in the Supporting Information).

An association index (point-biserial correlation coefficient – r_{pb}) was used to evaluate the association of 49 carbon and nitrogen cycle-related gene families with the study sites. In this analysis, we ran two independent tests for each gene family, finding the single site (test 1) and site or site combination (test 2) with highest association index. The significances of the associations were calculated with 999 permutations (De Caceres & Legendre 2009; De Caceres *et al.* 2010; a detailed description of the methods can be found in the Supporting Information).

Results and discussion

Diversity of microbial functional genes in the Amazon rainforest and the impact of land use change

A high-throughput microarray approach was employed to assess the diversity of microbial functional genes in Amazon rainforest soils. A total of 65 676 out of 83 992 genes targeted by GeoChip 4.0 were detected in at least one of the study sites. We detected genes associated with 409 different gene families in the following functional groups: nutrient cycling, antibiotic resistance, metal resistance, organic compound remediation, bacteriophage, stress, virulence and energy processes. To assess the impact of forest-to-pasture conversion on functional gene diversity, soil samples from a primary forest, two pastures (6 and 38 years old), and a secondary forest were evaluated. Gene richness (total number of genes detected) was reduced by 31.8% in the 6-year-old pasture and 16.1% in the 38-year-old pasture relative to the primary forest (Figure 1A; ANOVA, F: 89.8; $P < 0.001$). Functional gene diversity, estimated using the Shannon (H′) diversity index, was also significantly lower in the pastures relative to the primary forest (Figure S2; ANOVA, F: 57.7; $P < 0.001$). Interestingly, these results are the opposite of those reported for taxonomic diversity estimated using the 16S rRNA gene, which has been observed to increase in Amazon soils following ecosystem conversion at our site (Cenciani *et al.* 2009; Rodrigues *et al.* 2013) and at another (western) Amazon site (Jesus *et al.* 2009). The secondary forest was not significantly different from the primary forest in gene richness or diversity (Figure 1A, Figure S2), suggesting that recovery of at least some aspects of microbial functional diversity is possible.

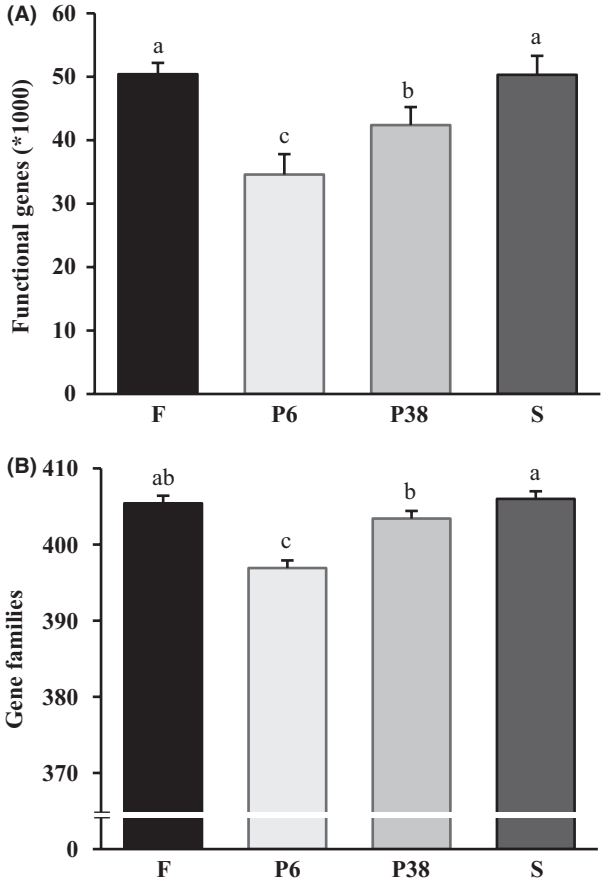

Fig. 1 Functional gene richness in Amazon soils under different land uses. Values are presented as mean ± standard deviation of 12 samples per sampling site. (A) Total gene sequences detected. (B) Total gene families detected. [a,b,c] sites not sharing the same letter are significantly different from each other (ANOVA and Tukey's *post hoc* test, $P < 0.05$). F, primary forest; P6, 6-year-old pasture; P38, 38-year-old pasture; S, secondary forest.

The number of gene families detected also varied across sites. Pastures had significantly fewer gene families relative to the primary forest (P6, $P < 0.001$; P38, $P < 0.05$), while secondary forest was not significantly different from primary forest (Fig. 1B). Even though the loss is statistically significant, the reduction is only 2.1% and 0.5% of gene families in the P6 and P38, respectively. Most of the affected genes are associated with organic contaminant degradation and metal resistance (Table S3), which could be attributed to the high number of probes associated with these functions present on the GeoChip microarray slide. Nevertheless, these gene groups might be important genetic reservoirs for novel biotechnological products with potential applications in biodegradation and bioremediation. Although there was only slight difference between forests and pastures regarding the number of gene families detected (Fig. 1B), the forest sites harbour a higher number of

different gene sequences for the same functions (Fig. 1A), which suggests higher functional redundancy in these sites. This finding has important implications for ecosystem maintenance as the magnitude and stability of a function have been shown to be dependent on trait redundancy across members of the microbial community (Levine *et al.* 2011; Griffiths & Philippot 2013). Therefore, one would expect forest microbial communities to maintain a more stable functional performance in response to environmental oscillations than communities in soils under pasture.

Microbial functional gene community structure similarity and land use change

Hierarchical cluster analysis showed that, in general, primary and secondary forests were clustered together in the upper portion of the dendrogram, suggesting similar gene group abundances (Fig. 2). Interestingly, some secondary forest samples were also clustered with the 38-year-old pasture samples. We hypothesize that secondary forest communities are in an intermediate stage between primary forest and pasture, suggesting a progressive reestablishment of 'forest-like' functional diversity following pasture abandonment. The 6-year-old pasture samples all clustered together, with only two 38-year-old pasture samples associated with that specific cluster. Analysis of similarity (ANOSIM) indicated that differences in functional gene similarity among soil microbial communities were associated with differences in land use ($R = 0.387$; $P = 0.001$). Pairwise comparisons (Table 1) showed no significant difference between the primary and the secondary forests, while in all other site comparisons, they were significantly different from one another ($P < 0.05$). Hence, the results suggest that the microbial communities in the soils from the pasture and forest sites harbour different functional gene contents, indicating that land use may affect not only taxonomic (Jesus *et al.* 2009; Rodrigues *et al.* 2013), but also functional composition in Amazon soils.

Our results suggest that environmental variables associated with pasture establishment and age might also be important determinants of microbial functional gene similarity across sites (Fig. 3). We sought to identify environmental factors suggestive of functional gene similarity by quantifying 24 edaphic variables (Table S4) and performing a correlational analysis through ordination (Oksanen *et al.* 2011). Primary and secondary forest functional gene composition were correlated negatively with pH and positively with Al^{+3}. This result is in agreement with what has been reported in taxonomic studies performed with Amazon forest soils (Jesus *et al.* 2009; Rodrigues *et al.* 2013). In our study, not only soil pH, which has long been regarded as an important factor driving microbial community structure and function in soils (Fierer & Jackson 2006; Nicol *et al.* 2008; Wakelin *et al.* 2008; Jesus *et al.* 2009), but also other variables, such as boron, phosphorus, manganese concentrations and sand content (texture), were positively correlated with functional gene similarity in forest samples and therefore are probably to influence these communities. Functional gene similarity in pasture samples was correlated with increased concentrations of iron, organic matter and clay content (Fig. 3). We attribute the differences in functional gene similarity observed in the 6-year-old pasture (Figures 2 and 3) as a consequence of a still recent slash and burn deforestation, characterized by an increased deposition of nutrients, including alkaline ash, from burned vegetation. The 6-year-old pasture had the lowest values for acidity, Al^{+3}, B, P, Mn and sand and the highest values for pH, C/N ratio, Fe and clay content (Table S4).

To measure the turnover in functional gene composition within each site, we used distance decay analysis (Soininen *et al.* 2007). We observed significant decay of functional gene similarity with geographical distance in the primary forest (Mantel Coefficient: 0.605; $P < 0.05$), but not in the other sites (Table S5). Rodrigues *et al.* (2013) also reported reduction in the slope of the distance decay relationship for taxonomic and phylogenetic similarity of bacterial communities after pasture establishment in the same area. The present results provide evidence that the functional gene similarity follows a similar pattern and further shows that functional gene turnover in the secondary forest differs from the primary forest.

To determine whether the changes in the functional community were a consequence of spatial distance or environmental variability among the sites, we evaluated how geographical and environmental distances affected variation in gene similarity across the sites using multiple regression on distance matrices (MRM; Tuomisto *et al.* 2003; Lichstein 2007). In this analysis, we separated the environmental measures into two matrices based on the measures of soil nutrients and soil texture. We observed significant contribution of soil texture to the variation in gene composition similarity throughout the sites ($P < 0.05$; Table S6). However, geographical distance did not have a significant effect (Table S6). The contribution of soil texture is in agreement with the variation in clay content across sites, mainly in P6 (Table S4). Although major changes in clay content in response to land use are not common (Holmes *et al.* 2004), studies performed in the southwestern Amazon have reported such changes, with variations in clay content of 13 to 106% at some sites, primarily young pastures (De Moraes *et al.* 1996; Neill *et al.* 1997a, 2005).

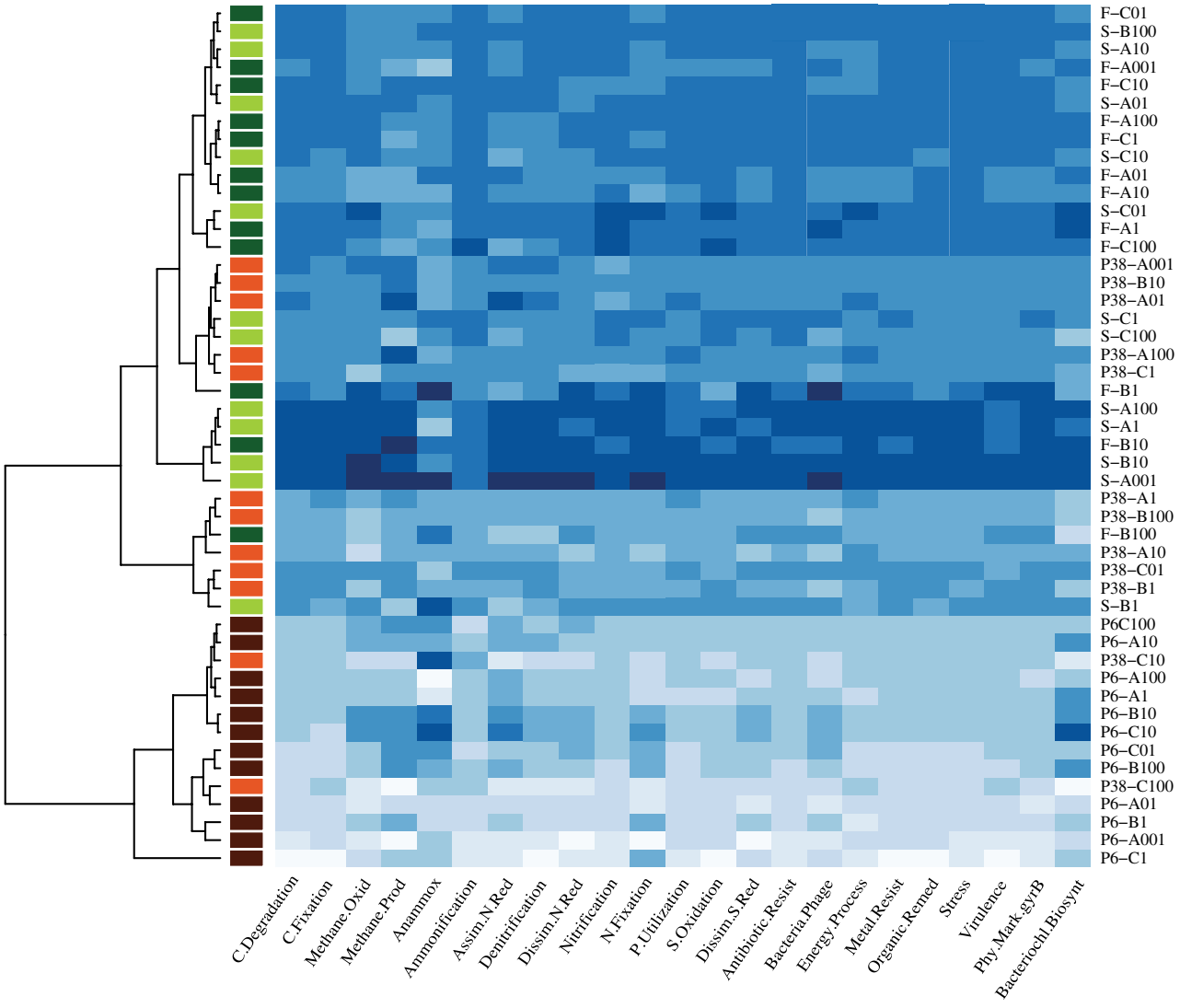

Fig. 2 Cluster analysis of site-specific microarray abundance data for genes related to different functional processes. Higher abundance is indicated by increased blue colour intensity. On the *x*-axis, genes were grouped by functional process and their abundance summed. On the *y*-axis, samples were reordered and clustered according to their colour intensity profile. The colour bar on the left side denotes the sampling site, as follows: dark green: primary forest; brown: 6-year-old pasture; orange: 38-year-old pasture; light green: secondary forest.

Table 1 Analysis of similarity (ANOSIM). Pairwise comparison between functional structures of Amazon soil microbial communities under different land use systems

	F	P6	P38	S
F	-	0.888	0.457	0.014
P6	0.006*	-	0.640	0.897
P38	0.006*	0.006*	-	0.471
S	0.999	0.006*	0.006*	-

Upper triangle shows ANOSIM statistic based on Bray–Curtis dissimilarity between two sites. Lower triangle shows *p* values of the test, after Bonferroni correction. *$P < 0.05$. F, primary forest; P6, 6-year-old pasture; P38, 38-year-old pasture; S, secondary forest.

Our study utilized a 'chronosequence', a series of sites distributed in space that have different histories. Because each treatment was studied in a single site, spatial variability may be of concern as environmental characteristics unrelated to chronosequence history could affect our results. However, there are two strong lines of evidence that spatial variability is not a primary driver of the results in our study. First, our chronosequence has been extensively studied for more than 20 years, and previous studies have shown that underlying spatial variability (in microclimate, soil type, etc.) is small relative to the variation caused by land use histories (e.g. De Moraes *et al.* 1996; Steudler *et al.* 1996; Neill *et al.* 1997a,b). Second, the geographical locations of our

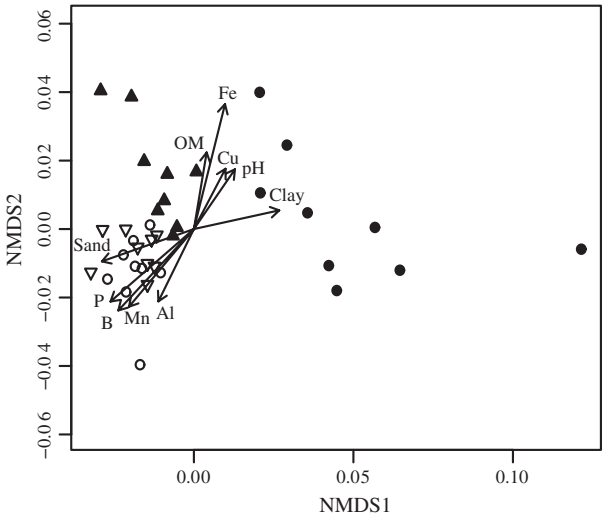

Fig. 3 Nonmetric multidimensional scaling of functional genes correlated with environmental factors in Amazon rainforest soils. Only environmental factors that share significant correlation ($P < 0.05$) with functional genes are displayed with vectors: OM, organic matter; Fe, iron; P, phosphorus, B; boron; Mn, manganese; Al, aluminium; pH, potential activity of hydrogen. The stress value was 0.08. Empty circles: primary forest (F); solid circles: 6-year-old pasture (P6); solid triangles: 38-year-old pasture (P38); empty triangles: secondary forest (S). Only samples with available environmental (9 per site) data were considered in this analysis.

sites support our conclusion that our results are due to land use history, not spatial variability (Figure S1). The young pasture (P6) is spatially closer to the forests than it is to the old pasture (P38), and the secondary forest is geographically much closer to the young pasture than to the primary forest. If spatial variability was the driver of our results, then we should have observed that the young pasture and primary forest were more similar than primary forest and secondary forest, and secondary forest and old pasture should be more similar than old pasture and young pasture. However, we observed that the pastures and forest sites each grouped together, a strong indicator that our results are driven by factors related to land use, rather than spatial variability.

Specific responses of genes related to the carbon and nitrogen cycles

Investigating how specific functional groups respond to land use change may help to predict potential impacts on processes performed by those groups. Microbial groups related to carbon and nitrogen cycling are of major interest, given their importance for ecosystem maintenance. We borrowed the concept of species association (also known as indicator species) traditionally employed in ecology (Dufrene & Legendre 1997) and used this index to identify C and N cycling genes associated with specific land use types, considering their frequency and abundance across sites. This analysis not only provided information on how land use may impact the distribution of ecologically important genes, but also identified genes that might be functionally relevant at particular sites. Point-biserial correlation coefficients (r_{bp}; De Caceres & Legendre 2009) were used to evaluate the association between a gene family and a single site (test 1) or a gene family and combinations of different sites (test 2; De Caceres *et al.* 2010). The purpose of the first test was to identify a unique habitat preferred by a particular functional group. Only associations with a single site are identified with this approach. The purpose of the second test was to provide a more realistic framework of how land use impacts the distribution of functional groups. This test allowed associations to single site or site combinations.

Of 49 gene families tested, 43 had significant associations to a site or site combination (Table 2). In test 1, the values of the significant ($P < 0.05$) correlation coefficients (r_{bp}) ranged from 0.330 to 0.542. We detected significant associations of 41 gene families to the two forest sites, especially the secondary forest (35 families). These included genes linked to C fixation, degradation of cellulose, hemicellulose, lignin and chitin, methane oxidation, ammonification, anammox, assimilatory N reduction, denitrification, dissimilatory nitrate reduction to ammonium (DNRA) nitrification and N fixation (Table 2). Interestingly, two gene families, pullulanase (*amyX*; $P < 0.01$) and isopullulanase ($P < 0.05$), showed an association with the 6-year-old pasture site. These genes are related to the hydrolysis of alpha-1,4-glucosidic linkages of starch and pullulan (Hii *et al.* 2012), two very labile carbon compounds. These results may reflect an increase in labile carbon compounds in pastures relative to forests, an observation consistent with the previous studies (Jensen & Nybroe 1999). The low number of gene families associated with the pastures is probably related to the reduction in detected genes in these sites (Fig. 1A). Overall, our results from this analysis suggest that factors linked to land use change may affect the frequency and abundance of these important microbial functional groups in Amazon soils.

When association analyses were performed using a combination of sites (test 2), r_{pb} values increased (ranging from 0.465 to 0.888), indicating stronger association between gene families and site combinations, with two potential explanations: (i) two or more sites may offer similar conditions that allow broad colonization by the same functional group of microbes, or (ii) microorganisms carrying the same gene families may colonize different habitats. These are nonmutually exclusive possibilities, and our study was not designed to partition

Table 2 Association of C and N cycling related gene families with the study sites. For each gene family, the site with the highest correlation, the value of the correlation index (r_{pb}) and the statistical significance of the association are indicated. The association with combined sites for each gene is also shown

	Association to single sites		Association to site combinations	
Functional processes/ gene families	Site	Index (r_{pb})	Site combination	Index (r_{pb})
Carbon fixation				
aclB	S	0.480**	F+P38+S	0.856**
CODH	S	0.439**	F+P38+S	0.747**
pcc	S	0.401**	F+P38+S	0.888**
rubisco	S	0.498**	F+S	0.671**
FTHFS	F	0.419**	F+S	0.683**
C Degradation – cellulose				
CDH	S	0.449**	F+P38+S	0.874**
cellobiase	F	0.411**	F+P38+S	0.815**
endoglucanase	S	0.450**	F+P38+S	0.852**
exoglucanase	S	0.330*	F+P38+S	0.755**
C Degradation – chitin				
acetylglucosaminidase	S	0.443**	F+P38+S	0.749**
endochitinase	S	0.416**	F+P38+S	0.819**
exochitinase	P38	0.154	F+P38+S	0.360
C Degradation – hemicellulose				
ara	S	0.453**	F+P38+S	0.777**
ara_fungi	S	0.518**	F+S	0.778**
mannanase	F	0.503**	F+P38+S	0.805**
xylA	S	0.420**	F+P38+S	0.785**
xylanase	S	0.392**	F+P38+S	0.689**
C Degradation – lignin				
glx	S	0.441**	F+P38+S	0.775**
lip	S	0.350*	F+P38+S	0.517*
mnp	S	0.239	S	0.239
phenol oxidase	S	0.405**	F+P38+S	0.638**
C Degradation – pectin				
pectinase	S	0.195	F+S	0.213
C Degradation – starch				
amyA	S	0.430**	F+P38+S	0.813**
amyX	P6	0.492**	P6	0.492*
apu	S	0.430**	F+P38+S	0.824**
cda	F	0.433**	F+S	0.743**
glucoamylase	S	0.399**	F+P38+S	0.647**
isopullulanase	P6	0.385*	P6+P38	0.491*
nplT	S	0.375*	F+S	0.598**
pulA	S	0.426**	F+P38+S	0.818**
Methane oxidation				
mmoX	S	0.363*	F+P38+	0.713**
pmoA	S	0.460**	F+S	0.633**
Methane production				
mcrA	S	0.277	F+P38+S	0.380
Ammonification				
gdh	S	0.431**	F+S	0.743**
ureC	S	0.464**	F+P38+S	0.866**
Anammox				
hzo	S	0.279	F+S	0.465*
Assimilatory N reduction				
nasA	S	0.404**	F+P38+S	0.633**
nirA	P38	0.275	P6+P38+S	0.329
nirB	S	0.361*	S	0.361
Denitrification				
narG[a]	S	0.392**	F+S	0.562**
nirK	S	0.464**	F+P38+S	0.787**
nirS	S	0.424**	F+S	0.626**
norB	S	0.532**	F+S	0.763**
nosZ	S	0.452**	F+P38+S	0.749**
Dissimilatory N reduction (DNRA)				
napAa	S	0.458**	F+S	0.747**
nrfA	F	0.429**	F+S	0.739**
Nitrification				
amoA	S	0.542**	F+S	0.825**
hao	F	0.355*	F+S	0.471*
Nitrogen fixation				
nifH	S	0.494**	F+S	0.677**

F: forest; P6: 6-year-old pasture; P38: 38-year-old pasture; S: secondary forest. *$P < 0.05$; **$P < 0.01$.

Enzymes coded by the genes:
aclB: succinyl-coA synthetase; *amoA*: ammonia monooxygenase; *amyA*: alpha-amylase; *amyX*: pullulanase; *apu*: amylopullulanase; *ara*: alpha-L-arabinosidase; *cda*: glycosyl hydrolase; CDH: carveol dehydrogenase; CODH: carbon monoxide dehydrogenase; FTHFS: formayltetrahydrofolate ligase; *gdh*: glutamate dehydrogenase; *glx*: glyoxal oxidase; *hao*: hydroxylamine oxidoreductase; *hzo*: hydrazine oxidoreductase; *lip*: ligninase; *mcrA*: methyl coenzyme M reductase; *mmoX*: soluble methane monooxygenase; *mnp*: lignin peroxidase; *napA*: nitrate reductase; *narG*: nitrate reductase; *nasA*: nitrate reductase; *nifH*: nitrogenase reductase; *nirA, nirB*: nitrite reductase; *nirK, nirS*: nitrite reductase; *norB*: nitric oxide reductase; *nosZ*: nitrous oxide reductase; *nplT*: neopullulanase; *nrfA*: nitrite reductase; *pcc*: propionyl-CoA carboxylase; *pmoA*: particulate methane monooxygenase; *pulA*: glycosyl hydrolase; *ureC*: urease; *xylA*: xylose isomerase.

[a]*narG* and *napA* are linked to both denitrification and DNRA.

the above outcomes; nonetheless, it indicates that specific gene families are preferentially associated with some habitat combinations, which may be a result of long-term disturbances, and hence, ecosystem functions may be altered. A total of 16 gene families were significantly associated with a combination of the two forests (F+S). Genes related to C and N cycles, such as nitrification, C and N fixation, C degradation, methane oxidation and nitrate reduction, were present in both environments in similar abundance/frequency.

Specifically, the genes responsible for the synthesis of the RuBisCo and formyltetrahydrofolate ligase (FTHFS) were associated with both forests. The former is a key enzyme in the Calvin–Benson–Bassham cycle and is involved in the first step of carbon fixation (Tabita *et al.* 2007), while the latter is associated with CO_2 uptake via acetogenesis, an anaerobic process known to occur in forest soils (Kusel & Drake 1995; Wagner *et al.* 1996). It is noteworthy that CO_2 reduction through acetogenesis is a thermodynamically less favourable process (−105 kJ/mol) than methanogenesis (−135 kJ/mol) under most conditions, yet Amazon forest soils have been observed to have a negative flux of CH_4 (Fernandes *et al.* 2002).

We also found that *pmoA,* a gene that encodes for the α subunit of particulate methane monooxygenase (Horz *et al.* 2001), was significantly associated with forest soils. The richness and abundance of the *pmoA* gene has been correlated with methane oxidation rates (Rahalkar *et al.* 2009; Tuomivirta *et al.* 2009) and process stability (Levine *et al.* 2011) in several environments. Previous biogeochemical studies at our study site observed a shift from sink to source in the annual CH_4 flux after forest-to-pasture conversion (Steudler *et al.* 1996; Fernandes *et al.* 2002). This change in CH_4 flux was attributed to alterations in physicochemical soil properties, such as a decrease in porosity as a result of soil compaction, and increased organic matter and pH, leading the authors to postulate that those changes might influence the soil microbiota. The limited oxygen diffusion in pasture soils may cause negative impacts on the local methanotrophic communities, while favouring methanogenic activity. Recently, Davidson *et al.* (2012) suggested that the Amazon rainforest is an ecosystem in transition, from carbon sink to source, due to anthropogenic activities. Understanding how land use changes impact microbial groups with a key role in carbon flux may provide valuable information on the causes of this ecosystem transition.

The impact of land use change on the nitrogen cycle is also of great concern because deforestation causes changes in several nitrogen cycling processes, including N mineralization and nitrification rates, as well as on the stocks of inorganic N (Neill *et al.* 1995). Changes in the abundance of N-cycling genes following forest-to-pasture conversion are summarized in Figure S3. We evaluated 16 different gene families directly linked to the N cycle and observed that 10 of them were significantly associated with primary and secondary forests (Table 2). Genes linked to nitrification, a key process controlling N cycling and loss in forest soils (Isobe *et al.* 2011), were significantly more abundant in forest sites. Both *amoA* and *hao* genes, encoding ammonia monooxygenase and hydroxylamine oxidoreductase, respectively, were associated with the forests. These results are consistent with those obtained by Neill *et al.* (1995), who found higher net nitrification rates in soils under native vegetation in comparison with pastures. Previous studies based on clone libraries also observed differences in community composition, with a decrease in diversity of *amoA* genes following conversion (Taketani & Tsai 2010; Navarrete *et al.* 2011). Concurrently, denitrification-related genes, *nirS* and *norB,* were linked to forest sites. Denitrification has been regarded as the most important source of the potent greenhouse gas N_2O in soils (Bollmann & Conrad 1998; Isobe *et al.* 2011), although Neill *et al.* (2005) have suggested that the nitrification is an important source of N_2O in Amazon forest soils. N_2O flux has been extensively measured in Amazon soils, and it is widely accepted that primary forest soils are sources of the gas (Neill *et al.* 2005; Davidson *et al.* 2012). The association of nitrification and denitrification genes with forests indicates that these land use types may have a higher potential for N_2O production, compared with pastures. Measurements previously made at our study site revealed a peak of gas emission after pasture establishment (up to 3 years), which was followed by a progressive reduction with pasture ageing, reaching levels lower than the forests (Neill *et al.* 2005).

We also observed the association of *napA* and *narG* with forests. These genes are present in microorganisms capable of denitrification and dissimilatory nitrate reduction to ammonium (DNRA). The DNRA exclusive marker *nrfA,* which encodes for the enzyme responsible for converting nitrite to ammonium (Kraft *et al.* 2011), was also associated with the forests, with a higher correlation index with primary forest. Although the DNRA process is still not fully understood, it plays an important role in N retention in soils, as well as in NO_3^- turnover in humid tropical forests, limiting substrate availability for denitrification (Silver *et al.* 2001).

Two ammonification-related gene families were analysed in our study, *ureC* and *ghd*, encoding for urease and glutamate dehydrogenase, respectively. While the former was associated with forest samples, which is consistent with high rates of N mineralization observed in forest soils at our study site (Neill *et al.* 1995), the latter did not show any associations with land use type. The lack of correlation between mineralization rates and the presence of ammonification genes, measured by GeoChip, has been reported previously, and it may be related to the wide distribution of these genes over different microbial taxa (Yergeau *et al.* 2007). Similarly, genes related to assimilatory nitrate reduction, which are known to be phylogenetically widespread (Luque-Almagro *et al.* 2011), did not show site selectivity in our association test.

Finally, the *nifH* gene, which encodes nitrogenase reductase, was significantly associated with both forest sites. Although the potential for N fixation has been suggested as important for secondary forest growth in the Amazon (Gehring *et al.* 2005) and other tropical forests (Erickson *et al.* 2001), very limited information is available for the Amazon forest regarding this process (Davidson *et al.* 2007; Mirza *et al.* 2014). We do not yet know whether the increased abundance of *nifH* in forest soils is associated with higher rates of nitrogen fixation in these soils.

Concluding remarks

The Amazon rainforest maintains an extraordinary biological diversity, with important implications for carbon sequestration and biogeochemical cycling of nutrients. The area of this tropical ecosystem has been shrinking at an alarming rate, with consequences for the phylogenetic and taxonomic diversity of both macro- and microorganisms. Our study demonstrates that Amazon deforestation can also alter microbial functional diversity. The establishment of a secondary forest resulted in the recovery of some, but not all, aspects of microbial functional gene diversity, suggesting that biogeochemical processes may recover as forests reestablish. Finally, we identified several gene families with significant association to one or more habitats, which may be a response to land use change. These gene families are linked to C and N cycling, raising concerns about impacts of land use change on ecosystem services.

Acknowledgements

We thank the owners and staff of Agropecuária Nova Vida for logistical support and permission to work on their property; Dr. Kyunghwa Baek and Wagner Piccinini for field sampling; Drs. Siu M. Tsai and Brigitte Feigl for soil analyses; and Dr. Jomar Barbosa for the photograph of the study area.

This project was supported by Agriculture and Food Research Initiative Competitive Grant 2009-35319-05186 from the US Department of Agriculture National Institute of Food and Agriculture. FSP was supported by the CAPES Foundation, Ministry of Education of Brazil (Grant 0246-10-7). This study is also partially supported by the United States Department of Agriculture (Project 2007-35319-18305) through NSF-USDA Microbial Observatories Program (JZ). The development of the GeoChip and associated computational pipelines used in this study was supported by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) through the Office of Science, Office of Biological and Environmental Research, the U. S. Department of Energy under Contract No.DE-AC02-05CH11231, by the OBER Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program (DE-SC0004601) and by the U.S. National Science Foundation MacroSystems Biology Program under the contract (NSF EF-1065844).

References

Allison SD, Martiny JBH (2008) Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences USA*, **105**, 11512–11519.

Barreto P, Souza-Jr C, Nogueron R *et al.* (2005) *Pressão humana na floresta Amazônica brasileira*. Imazon, Belém.

Bollmann A, Conrad R (1998) Influence of O_2 availability on NO and N_2O release by nitrification and denitrification in soils. *Global Change Biology*, **4**, 387–396.

Borneman J, Triplett EW (1997) Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Applied and Environmental Microbiology*, **63**, 2647–2653.

Camargo OA, Moniz AC, Jorge JA *et al.* (1986) *Métodos de análise química, mineralógica e física de solos do Instituto Agronômico de Campinas*. Instituto Adronômico, Campinas.

Cenciani K, Lambais MR, Cerri CC *et al.* (2009) Bacteria diversity and microbial biomass in forest, pasture and fallow in the southwestern Amazon Basin. *Revista Brasileira de Ciência do Solo*, **33**, 907–916.

Chan Y, Van Nostrand JD, Zhou J *et al.* (2013) Functional ecology of an Antarctic dry valley. *Proceedings of the National Academy of Sciences USA*, **110**, 8990–8995.

Clarke KR (1993) Nonmetric multivariate analysis of changes in community structure. *Australian Journal Ecology*, **18**, 117–143.

Davidson EA, de Carvalho CJR, Figueira AM *et al.* (2007) Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. *Nature*, **447**, 995–U996.

Davidson EA, de Araujo AC, Artaxo P *et al.* (2012) The Amazon basin in transition. *Nature*, **481**, 321–328.

De Caceres M, Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology*, **90**, 3566–3574.

De Caceres M, Legendre P, Moretti M (2010) Improving indicator species analysis by combining groups of sites. *Oikos*, **119**, 1674–1684.

De Moraes JFL, Volkoff B, Cerri CC *et al.* (1996) Soil properties under Amazon forest and changes due to pasture installation in Rondonia, Brazil. *Geoderma*, **70**, 63–81.

Development Core Team R (2010) *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing.

Dufrene M, Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*, **67**, 345–366.

Erickson H, Keller M, Davidson EA (2001) Nitrogen oxide fluxes and nitrogen cycling during postagricultural succession and forest fertilization in the humid tropics. *Ecosystems*, **4**, 67–84.

Fearnside PM (1999) Biodiversity as an environmental service in Brazil's Amazonian forests: risks, value and conservation. *Environmental Conservation*, **26**, 305–321.

Feeley KJ, Silman MR (2009) Extinction risks of Amazonian plant species. *Proceedings of the National Academy of Sciences USA*, **106**, 12382–12387.

Fernandes SAP, Bernoux M, Cerri CC *et al.* (2002) Seasonal variation of soil chemical properties and CO_2 and CH_4 fluxes in unfertilized and P-fertilized pastures in an Ultisol of the Brazilian Amazon. *Geoderma*, **107**, 227–241.

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences USA*, **103**, 626–631.

Fitter AH, Gilligan CA, Hollingworth K *et al.* (2005) Biodiversity and ecosystem function in soil. *Functional Ecology*, **19**, 369–377.

Gehring C, Vlek PLG, de Souza LAG *et al.* (2005) Biological nitrogen fixation in secondary regrowth and mature rainforest of central Amazonia. *Agriculture, Ecosystems and Environment*, **111**, 237–252.

Griffiths BS, Philippot L (2013) Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiology Reviews*, **37**, 112–129.

Griffiths BS, Kuan HL, Ritz K *et al.* (2004) The relationship between microbial community structure and functional stability, tested experimentally in an upland pasture soil. *Microbial Ecology*, **47**, 104–113.

Haines-Young R (2009) Land use and biodiversity relationships. *Land Use Policy*, **26**, S178–S186.

Hallin S, Jones CM, Schloter M *et al.* (2009) Relationship between N-cycling communities and ecosystem functioning in a 50-year-old fertilization experiment. *The ISME Journal*, **3**, 597–605.

Hazen TC, Dubinsky EA, DeSantis TZ *et al.* (2010) Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science*, **330**, 204–208.

He JZ, Sheng JP, Zhang LM *et al.* (2007) Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea of a Chinese upland red soil under long-term fertilization practices. *Environmental Microbiology*, **9**, 2364–2374.

Hii SL, Tan JS, Ling TC *et al.* (2012) Pullulanase: role in starch hydrolysis and potential indurstrial applications. *Enzyme Research*, **2012**, Article ID 921362.

Holmes KW, Roberts DA, Sweeney S *et al.* (2004) *Global Change Biology*, **10**, 796–814.

Horz HP, Yimga MT, Liesack W (2001) Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of pmoA, mmoX, mxaF, and 16S rRNA and ribosomal DNA, including pmoA-based terminal restriction fragment length polymorphism profiling. *Applied and Environmental Microbiology*, **67**, 4177–4185.

Isobe K, Koba K, Otsuka S *et al.* (2011) Nitrification and nitrifying microbial communities in forest soils. *Journal of Forest Research*, **16**, 351–362.

Jensen LE, Nybroe O (1999) Nitrogen availability to Pseudomonas fluorescens DF57 is limited during decomposition of barley straw in bulk soil and in the barley rhizosphere. *Applied and Environmental Microbiology*, **65**, 4320–4328.

Jesus EdC, Marsh TL, Tiedje JM *et al.* (2009) Changes in land use alter the structure of bacterial communities in Western Amazon soils. *The ISME Journal*, **3**, 1004–1011.

Kindt R, Coe R (2005) *Tree diversity analysis*. World Agroforestry Centre, Nairobi.

Kraft B, Strous M, Tegetmeyer HE (2011) Microbial nitrate respiration - Genes, enzymes and environmental distribution. *Journal of Biotechnology*, **155**, 104–117.

Kusel K, Drake H (1995) Effects of environmental parameters on the formation and turnover of acetate by forest soils. *Applied and Environmental Microbiology*, **61**, 3667–3675.

Legendre P, Legendre L (2012) *Numerical Ecology*. 3rd English Edition. Elsevier, Oxford.

Levine UY, Teal TK, Robertson GP, Schmidt TM (2011) Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *The ISME Journal*, **5**, 1683–1691.

Liang Y, Li G, Van Nostrand JD *et al.* (2009) Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field. *FEMS Microbiology Ecology*, **70**, 324–333.

Lichstein JW (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, **188**, 117–131.

Loreau M, Naeem S, Inchausti P *et al.* (2001) Ecology - Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, **294**, 804–808.

Lu Z, Deng Y, Van Nostrand JD *et al.* (2012) Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *The ISME Journal*, **6**, 451–460.

Luque-Almagro VM, Gates AJ, Moreno-Vivian C (2011) Bacterial nitrate assimilation: gene distribution and regulation. *Biochemical Society Transactions*, **39**, 1838–1843.

Mirza BS, Potisap C, Nusslein K *et al.* (2014) Response of free-living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. *Applied and Environmental Microbiology*, **80**, 281–288.

Mueller RC, Paula FS, Mirza BS *et al.* (2014) Links between plant and fungal communities across a deforestation chronosequence in Amazon rainforest. *The ISME Journal*, doi:10.1038/ismej.2013.253. [Epub ahead of print].

Navarrete AA, Taketani RG, Mendes LW *et al.* (2011) Land-use systems affect archaeal community structure and functional diversity in western Amazon soils. *Revista Brasileira de Ciência do Solo*, **35**, 1527–1540.

Neill C, Piccolo MC, Steudler PA *et al.* (1995) Nitrogen dynamics in soils of forest and active pastures in the western Brazilian Amazon Basin. *Soil Biology and Biochemistry*, **27**, 1167–1175.

Neill C, Melillo JM, Steudler PA *et al.* (1997a) Soil carbon and nitrogen stocks following forest clearing for pasture in Southwestern Brazilian Amazon. *Ecological Applications*, **7**, 1216–1225.

Neill C, Piccolo MC, Cerri CC *et al.* (1997b) Net nitrogen mineralization and net nitrification rates in soils following deforestation for pasture across the southwestern Brazilian Amazon Basin landscape. *Oecologia*, **110**, 243–252.

Neill C, Steudler PA, Garcia-Montiel DC *et al.* (2005) Rates and controls of nitrous oxide and nitric oxide emissions following conversion of forest to pasture in Rondonia. *Nutrient Cycling in Agroecosystems*, **71**, 1–15.

Nicol GW, Leininger S, Schleper C *et al.* (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environmental Microbiology*, **10**, 2966–2978.

Oksanen J, Blanchet FG, Kindt R *et al.* (2011) Vegan: community ecology package version 1.17-11.

Philippot L, Spor A, Hénault C *et al.* (2013) Loss in microbial diversity affects nitrogen cycling in soil. *The ISME Journal*, **8**, 1609–1619.

Rahalkar M, Deutzmann J, Schink B *et al.* (2009) Abundance and activity of methanotrophic bacteria in littoral and profundal sediments of lake Constance (Germany). *Applied and Environmental Microbiology*, **75**, 119–126.

Reeve JR, Schadt CW, Carpenter-Boggs L *et al.* (2010) Effects of soil type and farm management on soil ecological functional genes and microbial activities. *The ISME Journal*, **4**, 1099–1107.

Rodrigues JLM, Pellizari VH, Mueller R *et al.* (2013) Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proceedings of the National Academy of Sciences USA*, **110**, 988–993.

Sala OE, Chapin FS, Armesto JJ *et al.* (2000) Biodiversity - Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Silver WL, Herman DJ, Firestone MK (2001) Dissimilatory nitrate reduction to ammonium in upland tropical forest soils. *Ecology*, **82**, 2410–2416.

Soininen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological communities. *Ecography*, **30**, 3–12.

Steudler PA, Melillo JM, Feigl BJ *et al.* (1996) Consequence of forest-to-pasture conversion on CH4 fluxes in the Brazilian Amazon Basin. *Journal of Geophys Research*, **101**, 18547–18554.

Tabita FR, Hanson TE, Li H *et al.* (2007) Function, structure, and evolution of the Rubisco-like proteins and their Rubisco homologs. *Microbiology and Molecular Biology Reviews*, **71**, 576–599.

Taketani RG, Tsai SM (2010) The Influence of Different Land Uses on the Structure of Archaeal Communities in Amazonian Anthrosols Based on 16S rRNA and amoA Genes. *Microbial Ecology*, **59**, 734–743.

Tu Q, Yu H, He Z *et al.* (2014) GeoChip 4: a functional gene array-based high throughput environmental technology for microbial community analysis. *Molecular Ecology Resources*, doi:10.1111/1755-0998.12239. [Epub ahead of print].

Tuomisto H, Ruokolainen K, Yli-Halla M (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, **299**, 241–244.

Tuomivirta TT, Yrjala K, Fritze H (2009) Quantitative PCR of pmoA using a novel reverse primer correlates with potential methane oxidation in Finnish fen. *Research in Microbiology*, **160**, 751–756.

Van Raij B, Andrade JC, Cantarella H *et al.* (2001) *Análise Química Para Avaliação da Fertilidade de Solos Tropicais*. Instituto Agronômico, Campinas.

Wagner C, Griebhammer A, Drake HL (1996) Acetogenic capacities and the anaerobic turnover of carbon in a Kansas prairie soil. *Applied and Environmental Microbiology*, **62**, 494–500.

Wakelin SA, Macdonald LM, Rogers SL *et al.* (2008) Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. *Soil Biology and Biochemistry*, **40**, 803–813.

Wu L, Liu X, Schadt CW *et al.* (2006) Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Applied and Environmental Microbiology*, **72**, 4931–4941.

Yergeau E, Kang S, He Z *et al.* (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *The ISME Journal*, **1**, 163–179.

J.L.M.R., B.J.M.B, K.N., J.M.T. and V.H.P designed research; F.S.P, J.L.M.R., R.M., B.S.M., B.J.M.B and K.N. collected samples and performed experiments; J.L.M.R., J.Z. and V.H.P. contributed with reagents, materials and analytic tools; F.S.P., J.L.M.R., J.Z., L.W., R.M., B.J.M.B. and Y.D. analysed the data; F.S.P. and J.L.M.R. wrote the article. All authors critically revised the manuscript.

Data accessibility

Microarray data available: http://ieg.ou.edu/4download/. Sampling locations, GeoChip 4.0 slide information and soil physical and chemical attributes: online Supporting Information.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Material and methods.

Fig. S1 Location of the study area and aerial photograph with sampling sites.

Fig. S2 Functional gene diversity in Amazon soils under different land use systems, assessed by Shannon index.

Fig. S3 Changes in abundance of N-cycling gene families following forest-to-pasture conversion.

Table S1 GPS coordinates of each sampling location.

Table S2 Functional groups targeted by GeoChip 4.0.

Table S3 Gene families detected in Amazon rainforest soils.

Table S4 Soil physicochemical attributes and analysis of correlation between each variable and the axes of the nonmetric multidimensional scaling ordination.

Table S5 Mantel test evaluating distance decay of similarity in gene composition within sites.

Table S6 Multiple regression on distance matrices. Contribution of geographical and environmental distances to the variation in gene similarity across sites.