The ISME Journal (2014) 8, 1577–1587

www.nature.com/ismej

ORIGINAL ARTICLE

Taxonomical and functional microbial community selection in soybean rhizosphere

Lucas W Mendes[1,2,4], Eiko E Kuramae[2,4], Acácio A Navarrete[1,2], Johannes A van Veen[2,3,4] and Siu M Tsai[1]

[1]*Cell and Molecular Biology Laboratory, Center for Nuclear Energy in Agriculture (CENA), University of Sao Paulo (USP), Piracicaba, Brazil;* [2]*Department of Microbial Ecology, Netherlands Institute of Ecology NIOO-KNAW, Wageningen, The Netherlands and* [3]*Institute of Biology, Leiden University, Leiden, The Netherlands*

This study addressed the selection of the rhizospheric microbial community from the bulk soil reservoir under agricultural management of soybean in Amazon forest soils. We used a shotgun metagenomics approach to investigate the taxonomic and functional diversities of microbial communities in the bulk soil and in the rhizosphere of soybean plants and tested the validity of neutral and niche theories to explain the rhizosphere community assembly processes. Our results showed a clear selection at both taxonomic and functional levels operating in the assembly of the soybean rhizosphere community. The taxonomic analysis revealed that the rhizosphere community is a subset of the bulk soil community. Species abundance in rhizosphere fits the log-normal distribution model, which is an indicator of the occurrence of niche-based processes. In addition, the data indicate that the rhizosphere community is selected based on functional cores related to the metabolisms of nitrogen, iron, phosphorus and potassium, which are related to benefits to the plant, such as growth promotion and nutrition. The network analysis including bacterial groups and functions was less complex in rhizosphere, suggesting the specialization of some specific metabolic pathways. We conclude that the assembly of the microbial community in the rhizosphere is based on niche-based processes as a result of the selection power of the plant and other environmental factors.

The ISME Journal (2014) **8,** 1577–1587; doi:10.1038/ismej.2014.17; published online 20 February 2014
Subject Category: Microbial population and community ecology
Keywords: microbial ecology of tropical soils; amazonia; niche theory; neutral theory; soil metagenomics; land-use

Introduction

Within the soil system, the immediate surroundings of the plant root, that is, rhizosphere, is a microbial hot spot considered to be one of the most dynamic interfaces on earth (Philippot *et al.*, 2013). The microbial community of the rhizosphere constitutes part of a complex food web that utilizes the nutrients released by the plant (for example, exudates, border cells, mucilage), which are the major driving forces in the regulation of microbial diversity and activity in the immediate vicinity of plant roots (Mendes *et al.*, 2013). The main source of the species richness in the rhizosphere is the adjacent root-free soil, called the bulk soil, and so changes brought about in the communities of the bulk soil, for example, by land-use changes, will also have an effect on the assembly and the final composition of rhizosphere communities.

Several hypotheses have been raised regarding microbial community assembly, including the 'neutral theory' and the 'niche theory' (Dumbrell, 2010). The neutral theory predicts that the structure and composition of species communities is related to the geographic distance between samples as a result of dispersal limitation, as many species are functionally equivalent in their ability to exploit niches. Thus, their abundance will follow a zero-sum multinomial (ZSM) distribution (Hubbell, 2001; McGill *et al.*, 2006). The niche-based theory predicts that changes in species community composition are related to changes in environmental variables (Jongman *et al.*, 1995), as species have unique properties that allow them to exploit unique niches available. Their species abundances will follow pre-emption, broken stick, log-normal and Zipf–Mandelbrot models (Motomura, 1932; MacArthur, 1957; McGill *et al.*, 2007). Both theories are well connected with environmental factors, but neither

Correspondence: JA van Veen, Netherlands Institute of Ecology, NIOO-KNAW, Droevendaalsesteeg 10, Wageningen 6708 PB, The Netherlands.
E-mail: h.vanveen@nioo.knaw.nl
[4]These authors contributed equally to this work.
Received 23 October 2013; revised 10 January 2014; accepted 14 January 2014; published online 20 February 2014

suggest how the microbial community assembly in the main hot spot of life in soil, that is, the rhizosphere, is driven. We consider this knowledge to be vital for a more precise prediction and, thus, for the development of a better soil management strategy to cope with the negative effects of land-use changes and the development of more sustainable agricultural practices.

Several studies regarding microbial rhizosphere communities have shown the key role of plant species in shaping the microbial community in the rhizosphere. These studies included among others *Arabidopsis* (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012), rice (Knief *et al.*, 2011), oak (Uroz *et al.*, 2010), norway spuce (Calvaruso *et al.*, 2009), wild oats (DeAngelis *et al.*, 2008), potato (Rasch *et al.*, 2006), tobacco (Robin *et al.*, 2006) and soybean (Xu *et al.*, 2009). However, it is not well understood to what extent plants can select a constant rhizosphere community from highly contrasting reservoirs of bulk soil communities especially under humid tropical conditions.

Because of the increasing role of large-scale agriculture in Amazon and expansion of cropland (mainly soybean) into areas previously covered by forest, we would like to obtain a better insight into the ecological process of bacterial community selection and assembly in the soybean rhizosphere in Amazon soils recently converted into cultivation. For this purpose, we applied a DNA shotgun sequencing and integrating taxonomic and functional data to analyze the microbial community inhabiting the bulk soil and the rhizosphere of soybean cropland from an agricultural zone located in the Southeastern Brazilian Amazon, so to determine the extent to which a particular plant species, that is, soybean, is able to select a rhizospheric microbial community from the bulk soil reservoir. Because of the effects that land-use changes may have on the chemical and physical characteristics of former tropical forest soils, we hypothesize that niche-based mechanisms will explain the community assembly in the rhizosphere through a selective power of the plant in the rhizospheric environment. For this purpose, we compared the taxonomic and functional profiles among 24 independent samples from soils and rhizosphere after 1 and 5 years of soybean cultivation, respectively.

Although there are reports addressing the composition and structure of rhizospheric community from different plant species, this is the first study that used shotgun metagenome and integrating taxonomic and functional data to examine whether neutral or niche-based mechanisms would best explain the assembly of the microbial community of the rhizosphere.

Material and methods

Sampling sites

Bulk soil samples were collected in agricultural fields in two different sites located in the Southeastern Brazilian Amazon, in the state of Mato Grosso, Brazil, in the municipalities of Ipiranga do Norte (11°40′54.97″ S and 55°50′8.79″ W) and of Porto dos Gaúchos (11°44′29.62″ S and 56°15′44.52″ W). Oxisol is the predominant soil order of the sampling sites (Secretaria de Estado de Planejamento e Coordenação Geral, 2001), and the climate in the region is classified as Am (Koppen's classification), with annual average temperature of 28 °C and average precipitation of 2000 mm.

The sampling sites were selected according to vegetation cover, soil use and management practices. In the Porto dos Gaúchos municipality, areas covered with native tropical rainforest were cleared in 2008 and subsequently converted into agricultural land. Since 2004, forest conversion to agricultural use occurred in areas located in Ipiranga do Norte. Both sites have the same classification and historical origin (slash-and-burn deforestation). In both fields, after forest conversion to arable land, the annual crop rotation was: millet, soybean, and maize, under no-tillage. After deforestation, fertilizers, pesticides and a liming treatment were applied to the cropland fields at both locations. The cropland fields received different amounts of lime to increase soil pH to 5 and 6.

At each sampling site, the soil samples were collected from five points before sowing in two different sampling periods (November 2009 and November 2010). Sampling was performed by selecting one central point and four other sampling points (at least 50 m apart from the central point) directed towards the north, south, east and west of the central points. Soil samples were taken from the 0- to 20-cm topsoil layer. First, the litter layer was removed, and, then the soil samples were collected. A total of 20 bulk soil samples were collected in field (2 field locations $\times$ 2 sampling periods $\times$ 5 samplings per site). Samples were transported to the laboratory within 72 h after sampling for the preparation of the mesocosms.

Mesocosm experiments

The soil samples collected at the field were used in mesocosm experiments, where soybean plants were grown in greenhouse at CENA—University of São Paulo (USP), Piracicaba, Brazil. The experiments were carried out in the greenhouse in order to normalize the influence of environmental parameters (such as moisture regime and temperature) on the growth conditions of the plants. To simulate field conditions, the same cultivar of soybean *Glycine max* (L.) Merril (Cultivar M-SOY 8866) was used in the experiments, and before sowing, the seeds were inoculated with *Bradyrhizobium japonicum*, in a concentration of 10^{10} viable cells per kg of seed, which is a common practice for soybean cultivation in Brazil.

The mesocosms consisted of ceramic pots (30 cm high $\times$ 20 cm diameter) with a stone layer of 5 cm on

the bottom. The pots were filled with approximately 8 kg of soil, and three soybean seeds were sowed in each pot. Each soil sample consisted of a composite sample by mixing five subsamples collected in each point from the 0- to 20-cm topsoil layer. Fifteen pots (three per sampling point) for each site were used in the experiment, including control pots without plants. The plants germinated at 28/19 °C (day/night) with a 12-h photoperiod. Temperature and moisture were regularly adjusted to create optimal growth conditions for the plants. Soil samples were collected after 80 days of plant growth, period which corresponds to the end of growing period of the plants. Plants were collected and the roots with attached soil were removed from the pots and transported on ice to the laboratory. The roots were shaken to remove the loose soil and the remaining attached soil, considered to be the rhizosphere soil, was collected by using sterile brushes. Soil samples from the control pots were considered as bulk soil. The experiments were carried out two times during the tropical summer period of November–January 2009/2010 and November–January 2010/2011.

DNA extraction and sequencing

DNA extraction from 250 mg of soil was carried out using PowerSoil DNA Isolation Kit (Mobio Laboratories, Carlsbad, CA, USA), according to the manufacturer's protocol. DNA quality and concentration were measured by 1% sodium boric acid (Brody and Kern, 2004) agarose gel electrophoresis and NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). In total, 24 DNA samples were sequenced (Macrogen Inc. Company, South Korea) on a Roche 454 automated sequencer GS-FLX system Titanium series reagent (454 Life Sciences, Brandford, CT, USA).

Annotation of metagenomic sequences and data analysis

Unassembled DNA sequences were annotated with the Metagenomics Rapid Annotation (MG-RAST) pipeline version 3.3 (Meyer *et al.*, 2008). Taxonomic and functional profiles were generated using the normalized abundance of sequence matches to the SEED database (Aziz *et al.*, 2008). A table of the frequency of hits to each individual taxa (taxonomy) or subsystem (function) for each metagenome was generated and normalized by dividing by the total number of hits to remove bias indifference in read lengths and sequencing efforts. To identify hits, BlastX was used with a minimum alignment length of 50 bp and an *E*-value cutoff of $E<1\times10^{-5}$ (Dinsdale *et al.*, 2008; Smith *et al.*, 2012). Nonmetric multidimensional scaling (NMDS) plots were used to visualize the structure among samples, using the taxonomic and functional abundance matrix generated as described above. The plots were generated from Bray–Curtis similarity index matrices of the 24 samples. Diversity measurements (Shannon's index) were calculated based on taxonomic abundance matrix at species level. NMDS and Shannon were calculated by using the PAST software (Hammer *et al.*, 2001).

All statistical analyses were performed for 2 years of the experiments (2009 and 2010), and when convenient, the data were presented together. To determine statistical differences between the bulk soil and the rhizosphere samples, the Statistical Analysis of Metagenomic Profiles (STAMP) software package was used (Parks and Beiko, 2010). For this, a table of the frequency of hits of taxa and functional subsystem (SEED database) for each metagenome was generated from MG-RAST and used as input. *P*-values were calculated using the two-sided Fischer's exact test (Fisher, 1958), while confidence intervals were calculated using the Newcombe–Wilson method (Newcombe, 1998) and correction was made using Benjamini–Hochberg false discovery rate (Benjamini and Hochberg, 1995).

Network analyses were performed to better understand the taxonomic and functional relations within the microbial communities. For analyzing the networks, we calculated all possible Spearman's rank correlation coefficients. In order to filter the data for reduced network complexity, we considered high correlations with cutoff at $r>0.7$ and statistically significant *P*-value <0.01 and <0.001 for taxonomy and function, respectively, taking into account all replicates. The nodes in the reconstructed network represent taxa and functional groups, and edges represent high and significant correlations between nodes. The network graphs were made based on a set of measures, as average node connectivity, average path length, diameter, and cumulative degree distribution. Statistical analyses were carried out in the R environment (http://www.r-project.org/) and networks visualization with the interactive platform Gephi (Bastian and Jacomy, 2009).

To test whether neutral or niche-based mechanisms best explain the assembly of the microbial community, we examined the rank abundance distribution. The neutral theory predicts that rank abundance distribution will be consistent with ZSM (Hubbel, 2001). On the other hand, the niche-based theory assumes that the rank abundance distribution would fit the pre-emption, broken stick, log-normal and Zipf–Mandlebrot models (Motomura, 1932; MacArthur, 1957; McGill *et al.*, 2007). The species rank abundance for each metagenomic sample were fit to broken stick, pre-emption, log-normal, Zipf and Zipf–Mandelbrot rank abundance models using the command 'radfit' found in the R package vegan (R Development Core Team, 2007; Oksanen, 2010) and the ZSM model using TeTame (Jabot *et al.*, 2008). The models were compared based on the Akaike Information Criterion (AIC), which is a measure of relative quality of a statistical model, providing a means for model selection. AIC values for generated models were calculated based on the

equation $AIC = -2\,\text{log-likehood} + 2 \times \text{npar}$, where npar represents the number of parameters in the fitted model (Feinstein and Blackwood, 2012). The AIC values were compared to determine which model provided the best fit to the empirical data, indicated by the lowest value (Dumbrell *et al.*, 2010).

Results

Soil characteristics

The soil pH of the 1-year soybean site was 4.1 ± 0.3 and the texture of 50:3:47 (sand:silt:clay) while the soil pH of the 5-year soybean site was 5.0 ± 0.2 and the texture of 40:3:57. Among all sites and replicates, the organic matter content ranged from 34 to 38 g kg^{-1}, total nitrogen from 1.43 to 1.82 g kg^{-1} and organic carbon from 16 to 24 g kg^{-1} (Supplementary Table S1; for more detailed soil characteristics discussion, see Navarrete *et al.*, 2013).

Taxonomic profiling of metagenomes

Overall, sequencing yielded >3.2 million reads. After quality trimming, a total of 2 472 359 sequences, with average length of 523 bp, were obtained for 24 samples (Supplementary Table S2). Using a cutoff of $E < 1 \times 10^{-5}$ and 50-bp minimal align length on MG-RAST server, an average of 75% of sequences were predicted as protein. Metagenomic libraries were dominated by Bacteria (96% of hits to SEED) with sequences also matching Eukaryote (3%) and Archaea and Virus (1%). *Proteobacteria* represented the highest percentage of matches to the SEED database for all samples with average of 47% of all sequences, followed by *Actinobacteria* (23%), *Firmicutes* (6%) and *Acidobacteria* (5%) (Supplementary Figure S1A). When focusing on the most abundant phylum *Proteobacteria*, the numbers of sequences affiliated with distinct classes showed subtle differences among bulk soil and rhizosphere samples (Supplementary Figure S1B). Rarefaction curve analysis shows that sequencing has detected much, if not most, of the diversity present in all samples for the 2 years of the experiment (Supplementary Figure S2). In general, 28% of total sequences could not be assigned to known sequences in the database.

Bulk soil vs soybean rhizosphere

In order to visualize the differences in community structure and function between bulk soil and soybean rhizosphere samples, the taxonomic abundance profiles were used to compute a Bray–Curtis similarity matrix, coordinated into two dimensions by using NMDS (Figure 1). Samples were grouped according to site and age of harvest (analysis of similarity $Rglobal = 0.811$, $P < 0.01$); rhizosphere samples were more similar to each other than to bulk soil samples. This analysis revealed clear differences in the microbial community structure between bulk soil and rhizosphere. When samples were compared using the STAMP software, there was an over-representation of the phyla *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Cyanobacteria*, *Chlamydiae*, *Tenericutes*, *Deferribacteres*, *Chlorobi*, *Verrucomicrobia* and *Aquificae* ($P < 0.01$) in the rhizosphere (Figure 2). At the class level, the rhizosphere presented an over-representation of *Bacilli*, *Mollicutes*, *Clostridia*, *Gammaproteobacteria*, *Epsilonproteobacteria*, *Chlamydiae* and *Thermomicrobia* ($P < 6.9e^{-4}$) (Supplementary Figure S3).

To predict soil microbial functions, the functional profiles of bulk soil and soybean rhizosphere samples were analyzed according to the SEED database, and the most prevalent core of functions for all samples was 'carbohydrates', while a high abundance of sequences were matching 'membrane transport', 'iron acquisition and metabolism' in the rhizosphere samples (Supplementary Table S3). The comparison of the functional profiles by STAMP revealed an over-representation of membrane transport ($P < 1e^{-15}$) and functional cores related to metabolism of nitrogen, phosphorus, potassium and iron (Figure 3 and Supplementary Figure S4).

The network analysis was markedly different for bulk soil and rhizosphere samples (Figure 4 and Table 1). In general, the number of correlations in the rhizosphere was less than in the bulk soil. In the bulk soil, the network had 181 nodes and 194 edges (positive correlations), and the modularity was 0.828 with 55 communities while, for the rhizosphere, the network presented 182 nodes and 143 edges, that is, positive correlations, and the modularity was 0.914 with 59 communities. All positive and negative correlations were measured: between bacterial groups (bact–bact), between functional groups (func–func), and between bacterial and functional groups (bact–func) (Table 1). The number of positive correlations was higher than the negative ones for both bulk soil and rhizosphere. In general, the number of correlations was higher in the bulk soil than in the rhizosphere, except for function–function negative correlations, which showed the highest number in the rhizosphere. For bulk soil samples, the five bacterial groups that presented more correlations were *Deltaproteobacteria*, *Bacteroidia*, *Chloroflexi*, *Planctomycetacia* and *Sphingobacteria*, while *Chloroflexi*, *Deltaproteobacteria*, *Solibacteres*, *Sphingobacteria* and *Gammaproteobacteria* were the groups with most correlations in the rhizosphere (as indicated in Figure 2 and Supplementary Figure S5).

The data of all samples were fitted to theoretical species abundance distribution models to test whether neutral or niche-based mechanisms could best explain the composition and structure of microbial communities in both the bulk soil and the rhizosphere. The comparison of different rank abundance distribution models based on AIC values indicated that the ZSM gave the closest fit for the

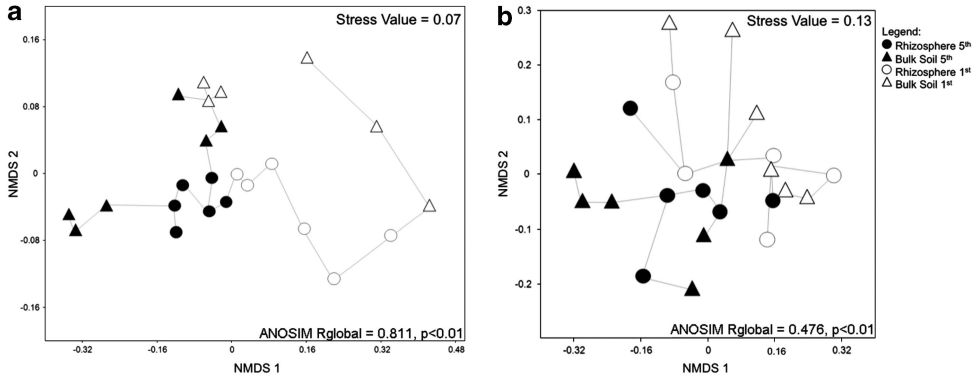

Figure 1 NMDS of Bray–Curtis similarity matrix among 24 samples from mesocosms experiments with soybean. Taxonomic (**a**) and functional (**b**) analyses using relative abundance based on SEED bacterial matches at the phylum level and subsystem level 2. The lines between dots represent the minimal spanning tree, which connects all points with minimal total length, based on similarity index. Stress values are shown in the upper right of the graphs. Similarity values (analysis of similarity) are shown in the down right of each plot.

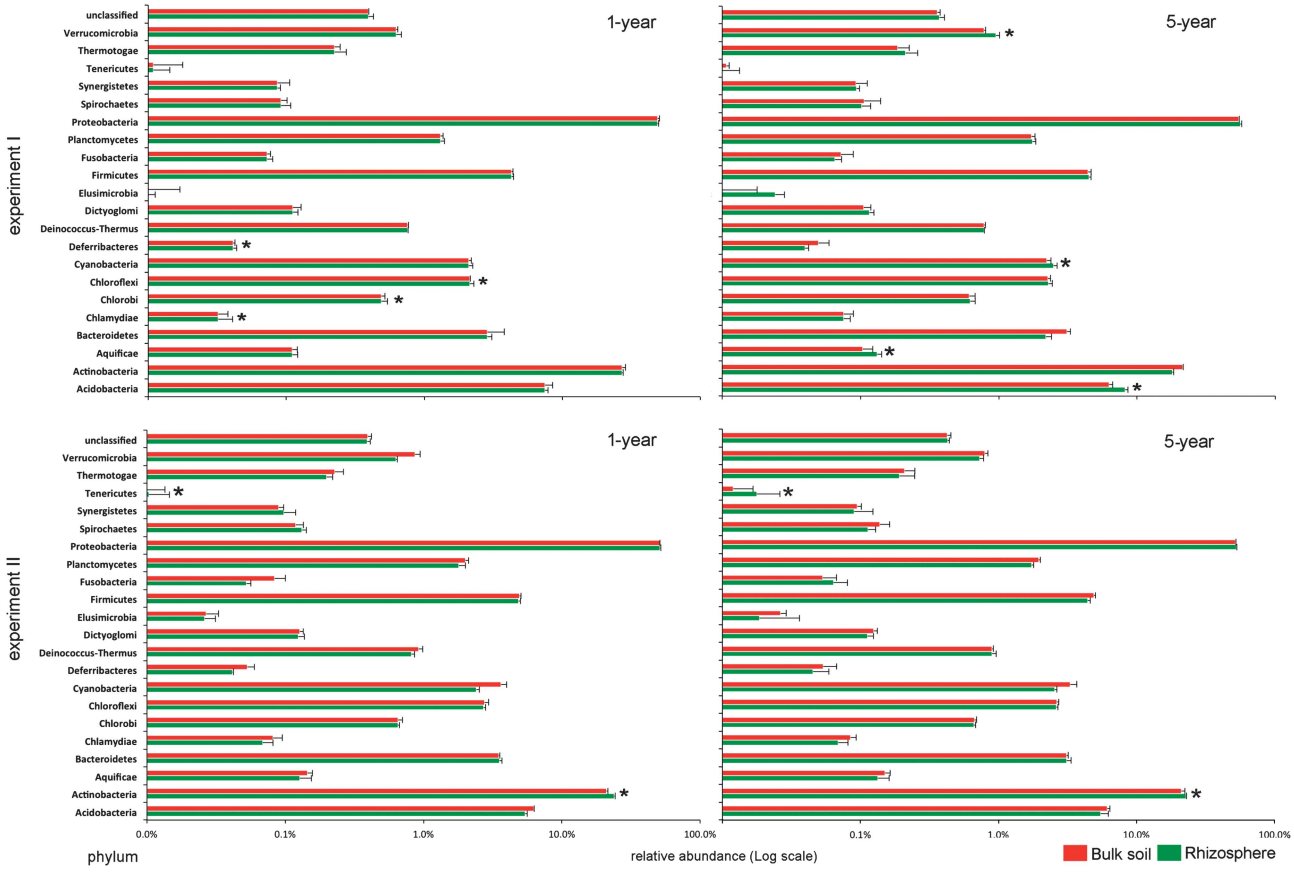

Figure 2 Relative abundance of bacteria at the phylum level based on shotgun metagenomics data. Percentage of total sequence reads in samples from bulk soil and rhizosphere of 1-year and 5-year of soybean harvesting is presented here for the 2 years of experiments (I and II). The error bars show calculated standard variation of triplicate samples, and asterisks (*) indicate more abundant phyla in rhizosphere (P-value <0.05). Corrected P-values were calculated using the Benjamini–Hochberg false discovery rate approach ($P<0.05$).

data of the bulk soil, which is consistent with neutral theory dynamics (Table 2). On the other hand, the rhizosphere data showed the best fit with the log-normal model, indicating dynamics according to the niche-based theory (Table 2). When we analyzed the data based on the time of harvesting, 1-year samples fitted neutral model, while 5-year samples fitted the niche-based model (Supplementary Table S4).

Soybean rhizosphere 1-year vs 5-year

The taxonomical and functional profiles of the rhizosphere were compared for the 1 year and 5

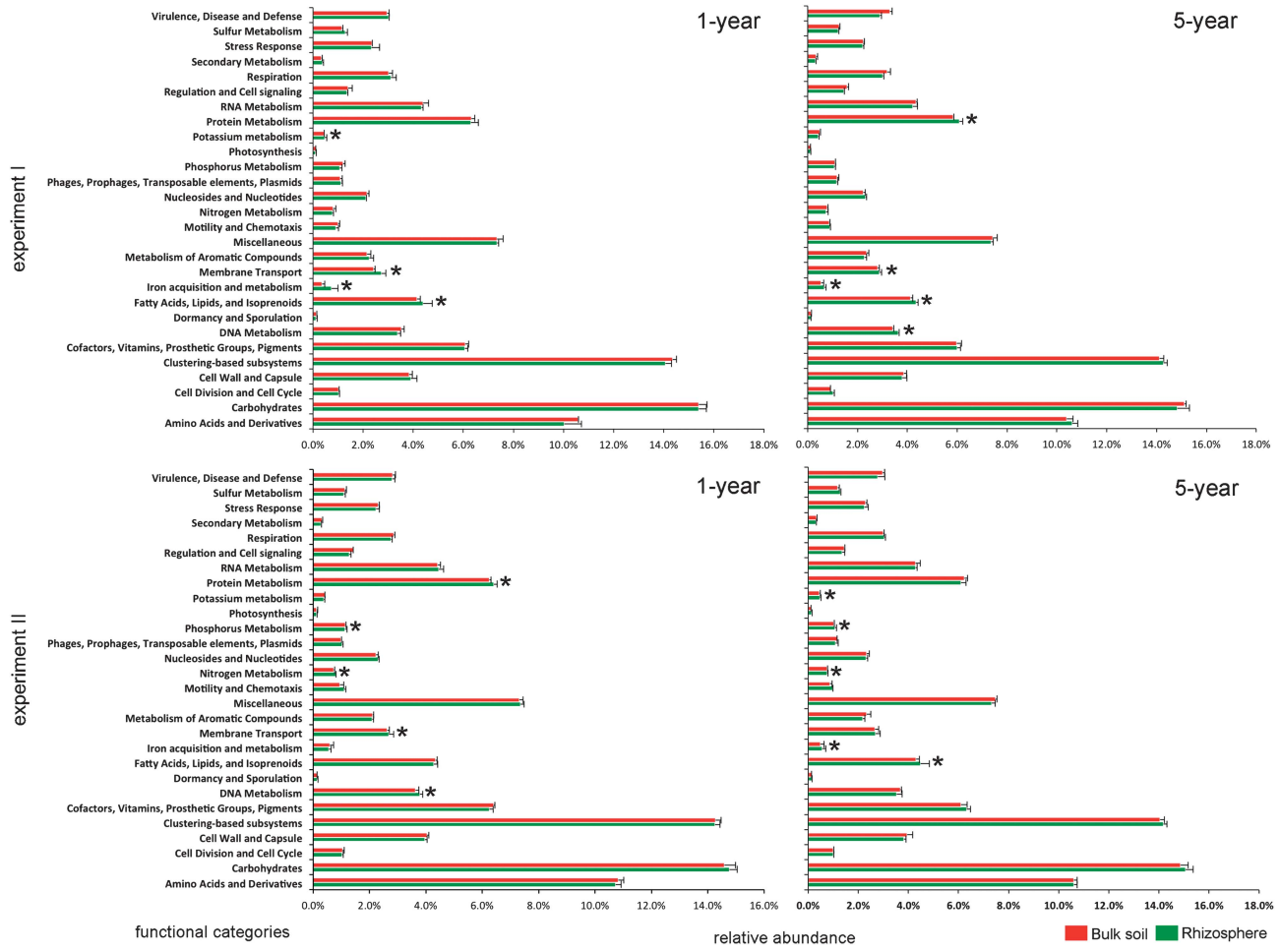

Figure 3 Relative abundance of functional categories (SEED subsystem level 1) based on shotgun metagenomics data. Percentage of total sequence reads in samples from bulk soil and rhizosphere of 1-year and 5-year of soybean harvesting is presented here for the 2 years of experiments (I and II). The error bars show calculated standard variation of triplicate samples and asterisks (*) indicate categories more abundant in rhizosphere (P-value<0.05). Corrected P-values were calculated using the Benjamini–Hochberg false discovery rate approach ($P<0.05$).

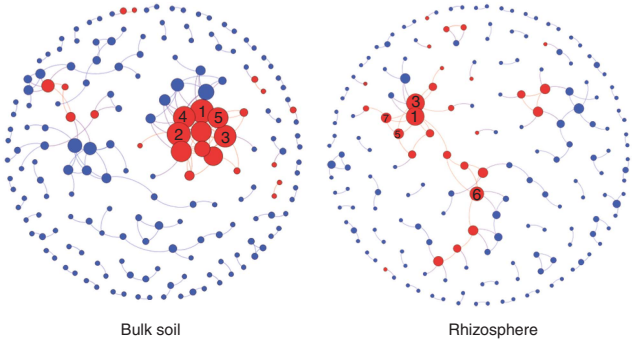

Figure 4 Network of bulk soil and rhizosphere based on correlation analysis from taxonomic and functional profiles. A connection stands for strong (Spearman's $r>0.7$) and significant ($P<0.01$) correlation. Red nodes indicate taxonomic affiliation at class level, and blue nodes indicate functional categories based on subsystems at level 2 (SEED Database). The size of each node is proportional to the number of connections. The numbers inside red nodes indicate the phyla with more correlations, as follows: (1) *Deltaproteobacteria*; (2) *Bacteroidetes*; (3) *Chloroflexi*; (4) *Planctomycetacia*; (5) *Sphingobacteria*; (6) *Solibacteres*; and (7) *Gammaproteobacteria*.

years of soybean cultivation. The NMDS analysis applied to visualize the differences among samples showed a separation according to time, revealing separated communities but with some overlap at both levels, phylum (analysis of similarity $R=0.525$, $P<0.01$) and class ($R=0.52$, $P<0.01$) (Supplementary Figures S6A and S3B). On the other hand, the functional profile did not present a clear separation between 1 year and 5 years for both Subsystem level 1 ($R=0.12$, $P>0.05$) and Subsystem level 2 ($R=0.24$, $P>0.05$) (Supplementary Figures S6C and D).

Although the diversity was not statistically different ($P>0.05$) between 1 year (Shannon–Wiener $H'=2.49$) and 5 years ($H'=2.27$), the community structure became more homogeneous in the 5-year samples, as seen in the NMDS plot with points more disperse for the 1-year samples (Figure 1a) and indicated by the nearest neighbors' analysis (mean distance: 1-year = 0.056; 5-year = 0.013).

STAMP analysis showed an enrichment of the phyla *Proteobacteria* ($P<0.001$), *Verrucomicrobia* ($P<0.001$) and *Planctomycetes* ($P<0.01$) in the

Table 1 Number of correlations as inferred by Spearman

	Phylogenetic	*Functional*	*Phylogenetic × functional*	*Total*
Bulk soil				
Pairwise correlations	732	98,503	17 256	116 491
Significant correlations	121 (62)	5871 (184)	1333 (64)	7325 (310)
Significant positive correlations	86 (52)	3044 (109)	623 (33)	3753 (194)
Significant negative correlations	35 (10)	2827 (75)	710 (31)	3572 (116)
Rhizosphere				
Pairwise correlations	739	97,957	17,278	115,974
Significant correlations	109 (45)	5355 (140)	1240 (51)	6704 (236)
Significant positive correlations	79 (28)	2792 (84)	675 (31)	3546 (143)
Significant negative correlations	30 (17)	2563 (56)	565 (20)	3158 (93)

The total numbers of pairwise correlations as well as significant correlations ($P<0.05$) among phylogenetic profile (class level), among functional profile (subsystem level 2) and between phylogenetic and functional profiles are shown. The numbers in parentheses indicate higher correlation coefficient for phylogenetic ($r>0.7$ and $P<0.01$), functional ($r>0.7$ and $P<0.001$) and phylogenetic × functional ($r>0.7$ and $P<0.001$) data.

Table 2 AIC values for six rank abundance distribution models

Sample	*Environment*	*Years*	*Experiment*	*AIC*[a]					
				Broken stick	*Pre-emption*	*Log-normal*	*Zipf*	*Zipf–Mandelbrot*	*ZSM*
Bulk_1a	Bulk soil	1 year	A	17235.7	14449.1	4571.4	14198.6	10029.4	**4549.86**
Bulk_1b	Bulk soil	1 year	B	13225.1	11380.2	4822.3	14634.8	9054.6	**4819.66**
Bulk_5a	Bulk soil	5 years	A	23718.1	20727.1	4943.7	13135.4	10583.7	**4896.90**
Bulk_5b	Bulk soil	5 years	B	13340.0	12351.5	4992.**7**	12244.8	8659.2	**4933.22**
Riz_1a	Rhizosphere	1 year	A	16629.4	14618.7	**4558.3**	14262.4	10062.1	5029.24
Riz_1b	Rhizosphere	1 year	B	15827.6	12981.1	**4735.0**	16420.1	12984.1	4758.06
Riz_5a	Rhizosphere	5 years	A	15740.9	12506.4	**4458.9**	14415.8	12509.7	4900.18
Riz_5b	Rhizosphere	5 years	B	7953.25	7096.45	**3620.14**	10011.28	5710.07	4853.28

Abbreviations: AIC, Akaike Information Criterion; ZSM, zero-sum multinomial.
Lowest AIC value for each sample represents the best fit model. AIC was calculated in the same way for the zero-sum model, from the minimum of log-likehood reported by TeTame, then multiplied by -1 to obtain the maximum log-likehood value. Bold values indicate the best fit model.
[a]AIC for radfit-generated models was calculated from the equation AIC $= -2\,\log\text{-likehood} + 2 \times \text{npar}$.

5-year samples and a decrease of *Actinobacteria* ($P<0.001$) (Figure 5). Regarding the functional profile, the subsystem 'protein metabolism' ($P<0.01$) was more abundant in the samples of the 1-year crop site, while the subsystem 'regulation and cell signaling' ($P<0.05$) was more abundant in the 5-year sites (Figure 6; these results are discussed in detail in the Supplementary Material section).

Discussion

The ordination of the taxonomic profiles revealed a clear separation between bulk soil and rhizosphere samples (analysis of similarity $Rglobal = 0.811$, $P<0.01$), which indicates a selective change in the bacterial community structure in the rhizosphere as compared with the composition of the bulk soil community as has been shown by others (Marschner *et al.*, 2001; Berg and Smalla, 2009; Uroz *et al.*, 2010). This influence is related to plant species and soil characteristics (Marschner *et al.*, 2001; Kowalchuk *et al.*, 2002; Berg and Smalla, 2009). Remarkably, the NMDS analysis of the functional profiles did not show a clear separation between rhizosphere and bulk soil samples.

A network analysis was accomplished in order to gain a more integrated understanding of the microbial community composition and functional traits and to compare the complexity of networks operating in the bulk soil and the rhizosphere. Our analysis was constructed using all positive and negative correlations between taxonomic and functional groups. In general, the number of correlations was lower in the rhizosphere than in the bulk soil samples. Considering that the rhizosphere community is a subset of the bulk soil community, a less complex network would be expected.

From the top five bacterial groups that presented most correlations with other bacterial groups and functional categories, *Deltaproteobacteria*, *Chloroflexi* and *Sphingobacteria* are representatives in both bulk soil and soybean rhizosphere. Exclusively in the soybean rhizosphere, next to *Gammaproteobacteria*, *Solibacteres* also presented a high number of correlations, which interestingly does not belong

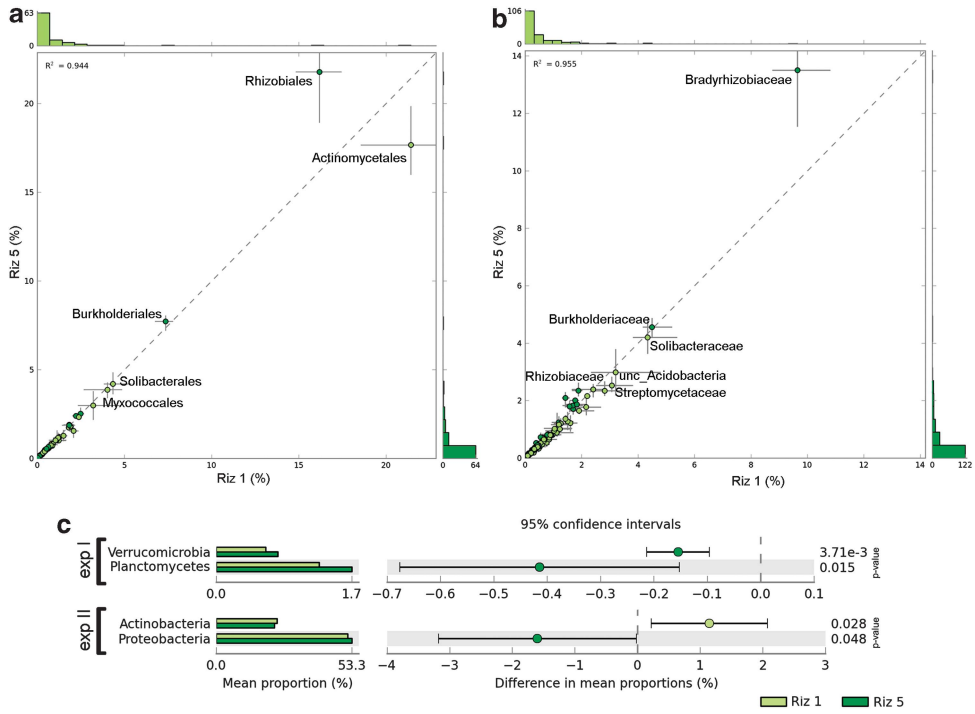

Figure 5 Comparison of taxonomic profile between 1-year (light green) and 5-year (dark green) soybean rhizosphere samples. (**a**) Scatter plot showing differences at order level ($P<0.05$). (**b**) Scatter plot showing differences at the family level ($P<0.05$). (**c**) Differences in phylum abundance between 1-year and 5-year rhizosphere samples for the 2 years of experiments (I and II). Corrected *P*-values were calculated using the Benjamini–Hochberg false discovery rate approach ($P<0.05$).

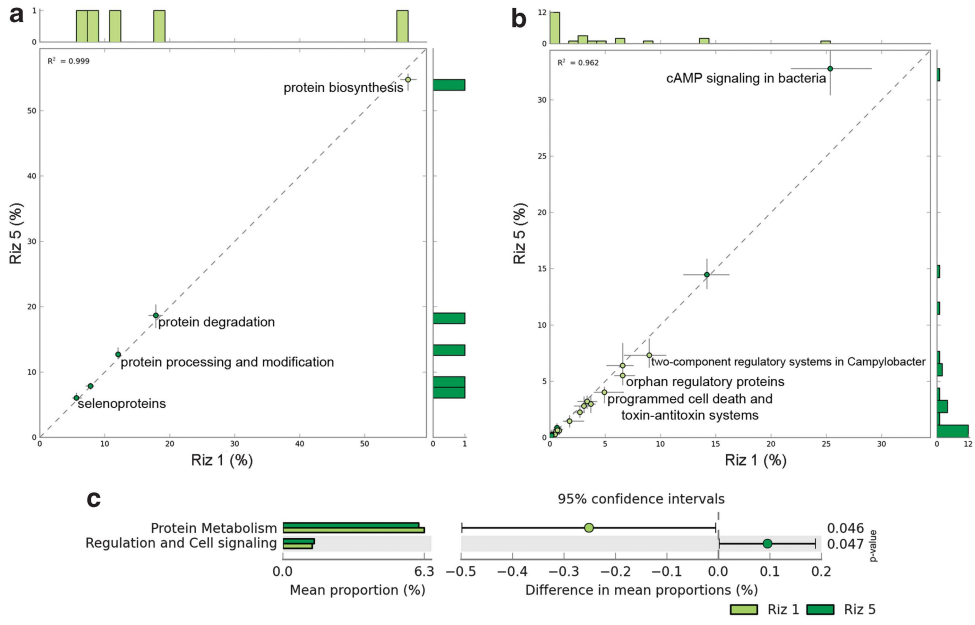

Figure 6 Comparison of functional profile between 1-year (light green) and 5-year (dark green) soybean rhizosphere samples. (**a**) Scatter plot showing differences for protein metabolism (SEED level 2). (**b**) Scatter plot showing differences for regulation and cell signaling (SEED level 2). (**c**) Functional groups (SEED level 1) statistically different between 1-year and 5-year rhizosphere. Corrected *P*-values were calculated using the Benjamini–Hochberg false discovery rate approach ($P<0.05$).

to the most abundant groups. These data point to an important role of relatively rare groups in the community, by keeping important connections on a larger scale with other groups and displaying important functional traits. Within the class of *Gammaproteobacteria*, the two most abundant

orders were *Enterobacteriales* and *Pseudomonadales*, which were shown earlier to preferentially inhabit the rhizosphere of several plants and are known for their beneficial effects on plant growth and/or protection against pathogens (Haichar *et al.*, 2008).

In our study, certain specific functional groups were more representative in the rhizosphere than in the bulk soil. The affiliation of the sequences in the SEED database and further analysis in STAMP indicated five functional cores over-represented in the rhizosphere, that is, 'membrane transport', 'nitrogen metabolism', 'phosphorus metabolism', 'potassium metabolism' and 'iron acquisition and metabolism'. These data suggest that these traits may be related to benefits to the plant and thus it suggests that the soybean plant selects a specific microbial community in the rhizosphere based on functional traits beneficial to its own performance. Specifically, we have listed the ecological relevance of these traits below.

Membrane transport
We found a high abundance of sequences affiliated with 'secretion system type IV'. The 'secretion system type IV' is associated with symbiotic interactions between bacterial community and other organisms (Burke *et al.*, 2011). Interestingly, we found 'secretion system type IV' binned to bacterial phyla such as *Acidobacteria*, *Actinobacteria*, *Bacteroidetes* and *Proteobacteria*, besides the inoculated *Bradyrhizobium* (*Proteobacteria*).

Nitrogen metabolism
We found a high abundance of genes involved with nitrogen fixation, denitrification and nitrite and nitrate ammonification in the rhizosphere. The presence of a large contingent of N_2 fixation genes is logical considering the leguminous environment. However, we also found the 'nitrosative stress' function in the rhizosphere. This function is related to a de-regulated synthesis or overproduction of NO and NO-derived products that can have toxic physiological consequences to the plant (Corpas *et al.*, 2011).

Phosphorus metabolism
The annotation of sequences related to phosphorus metabolism in rhizosphere revealed a predominance of genes related to P uptake and alkylphosphonate utilization, indicating the enhancement of plant P availability by P solubilization and mineralization and decreasing the pH (Marschner *et al.*, 2011).

Potassium metabolism
Potassium is one of the essential macronutrient needed for sufficient crop growth. Microorganisms have a key role in the K cycle, with specific groups, that is, *Pseudomonas*, *Bacillus*, *Rhizobium* and *Flavobacterium*, capable of mobilizing potassium in accessible form in soils (Hu *et al.*, 2006).

Iron acquisition and metabolism
Genes related to 'iron acquisition *in vibrio*' and 'heme and hemin uptake and utilization system in Gram negatives' were observed, which might be related to an increase of iron acquisition by the plant. Iron is poorly available in most soils and, in the rhizosphere, Fe is mobilized by plant- or microbe-derived compounds, and there is an intense competition for Fe uptake (Marschner *et al.*, 2011).

The composition of the rhizosphere community after 5-year soya cultivation showed some degree of homogenization as compared with the 1-year soya cultivation. This process is a common result of ecosystem conversion and a subsequent increase of community similarity in the course of time (Olden and Poff, 2003). Also, Rodrigues *et al* (2013) showed that the conversion of the Amazon rainforest to agricultural land results in biotic homogenization of soil bacterial communities. One of the reasons of the homogenization is the alteration of the ranges of existing species, as found in the 5-year samples, which showed an enrichment of *Verrucomicrobia* ($P<0.001$), *Proteobacteria* and *Planctomycetes* ($P<0.01$) and a decrease in abundance of *Actinobacteria* ($P<0.01$) (Figure 5c).

Community assembly is highly dependent on a multitude of trophic influences, which depend on the environmental biological diversity (Caruso *et al.*, 2011). The two theories that may be used to explain best the microbial community assembly are the neutral theory, which focuses on stochastic process, and the niche theory, which considers the importance of deterministic processes (Leibold and McPeek, 2006). Both theories, albeit viewed as contradictory, are not mutually exclusive. In some systems, both deterministic (niche theory) and stochastic (neutral theory) processes are responsible for structuring ecological communities (Chave, 2004). This holds for the assembly of microbial communities as well as some studies have shown that both deterministic and neutral processes are operational in structuring microbial communities (Dumbrell *et al.*, 2010; Caruso *et al.*, 2011; Ferrenberg *et al.*, 2013). Our data point to a microbial community selection in the rhizosphere via niche filtering, while the bulk soil composition and structure seemed to be regulated by neutral processes. Also the selection at the functional level in the rhizosphere seems to be based on processes according to the niche-based theory. This trend is more evident for the 5-year samples, which indicate that the power of selection increases during prolonged soybean cultivation. As the assembly based on niche mechanisms is determined by the niche requirements and local habitat conditions (Chase and Myers, 2011), we presume that the observed

niche-based selection of the rhizosphere microbiome is largely influenced by the interaction of soil physical–chemical characteristics and root-derived products, which shape the niches and exerts niche forces in the community assembly.

Although most of the data from the rhizosphere samples fitted the niche-based model, some samples fitted the ZSM, indicating the relevance of neutral-based processes as well, mainly in the site of 1 year of soybean cultivation. Considering that the microbial community of the rhizosphere is a subset of the bulk soil, both assembly theories can be important. Nevertheless, our data clearly indicate the importance of niche-based processes in structuring the rhizospheric community.

Conclusion

Our results suggest that soybean selects a specific microbial community inhabiting the rhizosphere based on functional traits, which may be related to benefits to the plant, as growth promotion and nutrition. This selection follows largely the niche-based theory, indicating the selection power of the plant and other environmental variables in shaping the microbial community both at the taxonomic and functional level. Long-term cultivation strengthens the selective power of the crop to communities and functions that are beneficial to the plant. Further analysis are needed to better understand the mechanisms by which the plant selects the rhizospheric community, whereby the study of the role of rhizodeposits in shaping microbial communities is of prime importance.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This study was supported by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2010/20353-7, 2008/58114-3, 2011/51749-6), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Wageningen 1240/11-0), and Conselho Nacional de Desenvolvimento Científico (CNPq 485801/2011-6). We thank R Mendes for critical review of the manuscript and W van der Putten for collaboration and support. Publication 5570 of the Netherlands Institute of Ecology (NIOO-KNAW).

References

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al.* (2008). The RAST server: Rapid Annotation using Subsystems Technology. *BMC Genomics* **9**: 75.

Bastian MHS, Jacomy M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Weblogs and Social Media: San Jose, CA, USA.

Benjamini Y, Hochberg Y. (1995). Controling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc* **57**: 289–300.

Berg G, Smalla K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol* **68**: 1–13.

Brody JR, Kern SE. (2004). Sodium boric acid: Atriz-less, cooler conductive medium for DNA electrophoresis. *Biotechniques* **36**: 214–216.

Bulgarelli D, Rott M, Schlaeppi K, van Themaat EVL, Ahmadinejad N, Assenza F *et al.* (2012). Revealing structure and assembly cues for Arabidopsis roo-inhabiting bacterial microbiota. *Nature* **488**: 91–95.

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. (2011). Bacterial community assembly based on functional genes rather than species. *PNAS* **108**: 14288–14293.

Calvaruso C, Mareschal L, Turpault MP, Leclerc E. (2009). Rapid clay weathering in the rhizosphere of Norway spruce and oak in an acid forest ecosystem. *SSSAJ* **73**: 331–338.

Caruso T, Chan Y, Lacap DC, Lau MCY, McKay CP, Pointing SB. (2011). Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *ISME J* **5**: 1406–1413.

Chase JM, Myers JA. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos Trans R Soc Lond B Biol Sci* **366**: 2351–2363.

Chave J. (2004). Neutral theory and community ecology. *Ecol Lett* **7**: 241–253.

Corpas FJ, Leterrier M, Valderrama R, Airaki M, Chaki M, Palma JM *et al.* (2011). Nitric oxide imbalance provokes a nitrosative response in plants under abiotic stress. *Plant Sci* **181**: 604–611.

DeAngelis KM, Brodie EL, DeSantis TZ, Andersen GL, Lindow SE, Firestone MK. (2008). Selective progressive response of soil microbial community to wild oat roots. *ISME J* **3**: 168–178.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM *et al.* (2008). Functional metagenomic profiling of nine biomes. *Nature* **452**: 629–U628.

Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. (2010). Relative roles of niche and neutral process in structuring a soil microbial community. *ISME J* **4**: 337–345.

Feinstein LM, Blackwood CB. (2012). Taxa-area relationship and neutral dynamics influence the diversity of fungal communities on senesced tree leaves. *Environ Microbiol* **14**: 1488–1499.

Ferrenberg S, O'Neill SP, Knelman JE, Todd B, Duggan S, Bradley D *et al.* (2013). Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J* **7**: 1102–1111.

Fisher WD. (1958). On grouping for maximum homogeneity. *J Am Stat Assoc* **53**: 789–798.

Haichar FZ, Marol C, Berge O, Rangel-Castro JI, Prosser JI, Balesdent J *et al.* (2008). Plant host habitat and root exudates shape soil bacterial community structure. *ISME J* **2**: 1221–1230.

Hammer Ø, Harper DAT, Ryan PD. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**: 9. pp.

Hu X, Chen J, Guo J. (2006). Two phosphate- and potassium-solubilizing bacteria isolated from Tianmu Mountain, Zhejiang, China. *World J Microb Biotechnol* **22**: 983–990.

Hubbell SP. (2001). The unified neutral theory of biodiversity and biogeography. *Monographs in Population Biology 32Princeton University Press: Princeton, NJ, USA.*

Jabot F, Etienne RF, Chave J. (2008). Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* **117**: 1308–1320.

Jongman RHG, Ter Braak CFJ, Van Tongeren OFR. (1995). *Data Analysis in Community and Landscape Ecology*. Cambridge University Press: Cambridge, UK.

Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R *et al.* (2011). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* **6**: 1378–1390.

Kowalchuk GA, Buma DS, de Boer W, Klinkhamer PGL, van Veen JA. (2002). Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. *Antonie Van Leeuwenhoek* **81**: 509–520.

Leibold MA, McPeek MA. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology* **87**: 1399–1410.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S *et al.* (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**: 86–90.

MacArthur R. (1957). On the relative abundance of bird species. *Proc Natl Acad Sci USA* **43**: 293–295.

Marschner P, Yang CH, Lieberei R, Crowley DE. (2001). Soil and plant specific effects on bacterial community composition in the rhizosphere. *Soil Biol Biochem* **33**: 1437–1445.

Marschner P, Crowley D, Rengel Z. (2011). Rhizosphere interactions between microorganisms and plants govern iron and phosphorus acquisition along the root axis—model and research methods. *Soil Biol Biochem* **43**: 883–894.

McGill BJ, Rampal S, Etienne JS, Gray DA, Anderson MJ, Benecha HK *et al.* (2007). Species abundance distribution: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett* **10**: 995–1015.

McGill BJ, Maurer BA, Weiser MD. (2006). Empirical evaluation of neutral theory. *Ecology* **87**: 1411–1423.

Mendes R, Garbeva P, Raaijmakers JM. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev* **37**: 634–663.

Meyer F, Paarman D, D'Souza M, Olson R, Glass EM, Kubal M *et al.* (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Motomura I. (1932). On the statistical treatment of communities. *Zool Mag* **44**: 379–383.

Navarrete AA, Kuramae EE, de Hollander M, Pijl AS, van Veen JA, Tsai SM. (2013). Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. *FEMS Microbiol Ecol* **83**: 607–621.

Newcombe RG. (1998). Improved confidence intervals for the difference between binomial proportions based on paired data. *Stat Med* **17**: 2635–2650.

Oksanen P. (2010). Vegan 1.17-0 in R version 2.10.1 (2009-12-14). URL http://cc.oulu.fi/~jarioksa/softhelp/vegan.html.

Olden JD, Poff NL. (2003). Toward a mechanism understanding and prediction of biotic homogenization. *Am Nat* **162**: 442–460.

Parks DH, Beiko RG. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**: 715–721.

Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* **11**: 789–799.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation For Statistical Computing: Vienna, Austria, http://www.R-project.org).

Rasch F, Hödl V, Poll C, Kandeler E, Gerzabek MH, van Elsas JD *et al.* (2006). Rhizosphere bacteria affected by transgenic potatoes with antibacterial activities compared with the effects of soil, wild-type potatoes, vegetation stage and pathogen exposure. *FEMS Microbiol Ecol* **56**: 219–235.

Robin A, Mougel C, Siblot S, Vansuyt G, Mazurier S, Lemanceau P. (2006). Effect of ferritin overexpression in tobacco on the structure of bacterial and pseudomonad communities associated with the roots. *FEMS Microbiol Ecol* **58**: 492–502.

Rodrigues JLM, Pellizari VH, Mueller R, Baek K, Jesus ED, Paula FS *et al.* (2013). Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *PNAS* **110**: 988–993.

Smith RJ, Jeffries TC, Roudnew B, Fitch AJ, Seymour JR, Delpin MW *et al.* (2012). Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. *Environ Microbiol* **14**: 240–253.

Uroz S, Buee M, Murat C, Frey-Klett P, Martin F. (2010). Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep* **2**: 281–288.

Xu Y, Wang G, Jin J, Liu J, Zhang Q, Liu X. (2009). Bacterial communities in soybean rhizosphere in response to soil type, soybean genotype, and their growth stage. *Soil Biol Biochem* **41**: 919–925.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)