The ISME Journal (2017) 11, 56–66

www.nature.com/ismej

ORIGINAL ARTICLE

Functional traits dominate the diversity-related selection of bacterial communities in the rhizosphere

Yan Yan[1,2], Eiko E Kuramae[1], Mattias de Hollander[1], Peter GL Klinkhamer[2] and Johannes A van Veen[1,2]

[1]*Department of Microbial Ecology, Netherlands Institute of Ecology NIOO-KNAW, Wageningen, The Netherlands and* [2]*Department of Plant Ecology and Phytochemistry, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands*

We studied the impact of community diversity on the selection of bacterial communities in the rhizosphere by comparing the composition and the functional traits of these communities in soil and rhizosphere. Differences in diversity were established by inoculating into sterilized soils diluted suspensions of the same soil. We used 16S ribosomal RNA amplicon sequencing to determine the taxonomical structure of the bacterial communities and a shotgun metagenomics approach to investigate the potential functional diversity of the communities. By comparing the bacterial communities in soil and rhizosphere, the selective power of the plant was observed both at the taxonomic and functional level, although the diversity indices of soil and rhizosphere samples showed a highly variable, irregular pattern. Lesser variation, that is, more homogenization, was found for both the taxonomic structure and the functional profile of the rhizosphere communities as compared to the communities of the bulk soil. Network analysis revealed stronger interactions among bacterial operational taxonomic units in the rhizosphere than in the soil. The enrichment processes in the rhizosphere selected microbes with particular functional genes related to transporters, the Embden–Meyerhof–Parnas pathway and hydrogen metabolism. This selection was not random across bacteria with these functional traits, but it was species specific. Overall, this suggests that functional traits are a key to the assembly of bacterial rhizosphere communities.
The ISME Journal (2017) **11,** 56–66; doi:10.1038/ismej.2016.108; published online 2 August 2016

Introduction

Loss of biodiversity can have significant consequences for ecosystem processes (Sala *et al.*, 2000; Magurran and Henderson, 2003; Butchart *et al.*, 2010), for example, the productivity and stability of ecosystems (Worm and Duffy, 2003; McGill *et al.*, 2007). However, whether or not this effect holds true for microbial communities, which are assumed to have a high degree of functional redundancy, is still a matter of debate. Soil microbes represent the majority of biodiversity in terrestrial ecosystems and are largely responsible for the maintenance of soil quality and functioning (Philippot *et al.*, 2013). Deeper knowledge of soil microbial biodiversity and the link with functionality could lead to a better understanding of the importance of biodiversity for the functioning of terrestrial ecosystems.

One of the most prominent hotspots of activity and diversity in soils is the rhizosphere. The composition of microbial communities and their activities in the rhizosphere have a large impact on the growth and health of plants (Mendes *et al.*, 2011; Berendsen *et al.*, 2012). The microbial community in the rhizosphere is mainly derived from the surrounding soil community. Therefore, changes in the soil community, for example, those brought about by manipulated disturbances, are expected to have significant effects on the assembly and final composition of the rhizosphere community.

Although there is an increasing amount of literature that deals with the influence of stochastic and deterministic factors, including soil and plant characteristics, on microbial community assemblage at various taxonomic levels (Langenheder and Szekely, 2011; Mendes *et al.*, 2011; Stegen *et al.*, 2012), the relative contribution of soil and plant characteristics to the process of microbial community assemblage at different functional levels is not yet known. Difficulties in experimental assessment constitute the major obstacle in understanding how microbial diversity is created and affected by factors such as soil and plant. A major limitation of current research on microbial diversity is the lack of sound

Correspondence: JA van Veen, Department of Microbial Ecology, Netherlands Institute of Ecology NIOO-KNAW, Droevendaalsesteeg 10, 6708 PB, Wageningen, The Netherlands.
E-mail: H.vanVeen@nioo.knaw.nl
Received 17 November 2015; revised 11 April 2016; accepted 24 June 2016; published online 2 August 2016

approaches to make directed and predictable changes in the diversity to address mechanisms underlying community assembly. An approach that is often used to assess the effects of diversity is the so-called dilution method. Yan *et al.* (2015) and several others provided evidence that the diversity of the microbial community in the soil was altered by inoculating diluted suspensions, although predictions on the outcome of the community assemblage processes were hard to make (Salonius, 1981; Garland and Lehman, 1999; Franklin *et al.*, 2001; Matos *et al.*, 2005; Franklin and Mills, 2006; Pedros-Alio, 2006; Hol *et al.*, 2010; Philippot *et al.*, 2013; Vivant *et al.*, 2013). Although this method has been used frequently in the past, until now, we have ignored the question of whether or not functional characteristics of the microbial community have a role in the selection of microbial species in soil and rhizosphere and if so, how. By comparing differences in soil and rhizosphere communities, we can get insights in the role of diversity in the selection processes operating in the rhizosphere; and by comparing taxonomic and functional traits, we can draw conclusions on whether the rhizosphere selection is based on taxonomy and/or functionality. Thus in combination with recent advances in high-throughput sequencing, which now allows for the assessment of both the taxonomic composition and function of the rhizosphere microbiome (Bulgarelli *et al.*, 2015), the dilution approach, may enable us to address this question.

The major aim of this study was to acquire a better understanding of the role of diversity in bacterial community selection at both the taxonomic and functional level in soil and rhizosphere. In order to obtain communities differing in diversity, we inoculated serial dilutions of suspensions into original sterilized bulk soil. After an established incubation period, plants were potted in the diverse soil samples. The plant species we used in this study, *Jacobaea vulgaris*, is one of the most common weeds in the Netherlands. We applied 16S ribosomal RNA gene amplicon sequencing to analyze the community structure in the diverse soil and rhizosphere samples, and a total DNA shotgun sequencing approach to assess their potential functions. In a previous study, we found that the soil has a strong selective impact on the assemblage of bacterial communities after incubation of various diluted inocula (Yan *et al.*, 2015) and that these strong selection processes operating in soil lead to a certain homogenization in the communities formed after regrowth of the diverse inocula. We hypothesized that plants will exert a further selection at both taxonomic and functional traits levels. In particular, we were interested whether or not selection processes in the rhizosphere exert a strong effect on functional traits and, if so, whether this selection is random across species with these traits or if it is species-specific.

Materials and methods

Soil sampling and plant selection

Thirty liters of bulk soil were collected at a depth of 15 cm from a dune soil in 2012 from Meijendel, the Netherlands, where *J. vulgaris* commonly grow naturally. The soil had a sandy texture, 9.11% organic matter content, pH 7.4, 30.43 mg kg^{-1} NO^{3-}, 2.23 mg kg^{-1} NH^{4+}, 15.16 mg kg^{-1} P. The soil was sieved (5 mm mesh size) and homogenized and stored in 500 g aliquots in plastic bags. One bag of soil was kept separately to prepare the inoculum. Soil was sterilized by γ-irradiation (>25 kGray, Isotron, Ede, The Netherlands). The sterility was tested by spreading 0.5 g of the soil from the inoculum-bag onto trypticase soy agar and potato dextrose agar media. No bacterial and fungal growth for the soils of the six replicates was observed on agar plates during 6 days of incubation. A subsample of the fresh soil was used to determine soil moisture (24 h, 105 °C). Soil suspension for inoculation was made by mixing 20 g of fresh soil in 190 ml of autoclaved demineralized water with a blender for 2 min. This was called the undiluted, 10^{-1} suspension. This suspension was sequentially diluted to obtain further dilutions of 10^{-6} and 10^{-9}, and these were added to the bags with the sterilized soil.

Microbial abundance was similar for all dilutions during 8 weeks of incubation as determined by quantitative real time PCR (Yan *et al.*, 2015). After 8 weeks of soil incubation, plants were potted in 0.5 l pots containing the incubated soil. We used *J. vulgaris* as the plant species. Seeds were collected in Meijendel (52° 9 N′, 4° 22′ E), the Netherlands. One seed was propagated by tissue culture (Joosten *et al.*, 2009). As tissue culture has often been defined as the 'sterile' plant, it was reasonable to use this 'clean' plant cloned for the further experiments. Plants were placed randomly in a climate room (relative humidity 70%, light 16 h at 20 °C, dark 8 h at 20 °C). Sterile demineralized water was given every 2 days with additions of 10 ml nutrient solution (Steiner, 1968) once every 2 weeks, in order to avoid nutrient limitation to plant growth. Water was added to the scale at the bottom of each pot where openings allowed water entering the pot. Samples were taken from the bulk soil at the moment of planting. After 6 weeks of plant growth, plants were collected and gently shaken to remove the loosely adhered soil after which rhizosphere soil samples were collected by removing the remnant soil with a fine sterile brush. Samples were stored at − 20 °C for further analysis.

The design of the experiment included three dilutions with six replicates each for both the incubated bulk soil and rhizosphere soil samples. One additional control group with three replicates, which was inoculated with sterilized demineralized water was included and incubated during the entire experimental period. Given that during plant growth the soil was only isolated by a layer of tin foil from

the atmosphere, we considered the possibility that this could constitute an unknown source of bacteria. However, we assumed that this would not have a major effect on our results as we know that the bulk soil had a full grown community of over 10^9 cells per gram of soil after the 8-week pre-incubation period following inoculation with the (un-) diluted suspensions. The impact of bacterial and extracellular DNA left in soil after sterilization before inoculation was accounted for by subtracting the operational taxonomic units (OTUs) found in the non-inoculated samples from those detected in the inoculated samples (Yan *et al.*, 2015).

DNA extraction, PCR reaction and 16S rDNA gene fragment sequencing

Total DNA was extracted from the suspensions, incubated bulk soil and rhizosphere soil to determine the composition of the respective microbial communities by 454-pyrosequencing of the 16S rDNA genetic marker (Yan *et al.*, 2015). The DNA was extracted using the MoBio Power Soil Extraction Kit according to the supplier's manual (MoBIO Laboratories, Carlsbad, CA, USA). Total DNA concentration was quantified on an ND-1000 spectrophotometer (Nanodrop Technology, Wilmington, DE, USA). PCRs were performed using 5 μM of each forward (515F) and reverse (806R) bar-coded primers (Bergmann *et al.*, 2011), 5 mM dNTPs (Invitrogen, Carlsbad, CA, USA), 1 unit of *Taq* polymerase (Roche, Indianapolis, IN, USA) and 5 ng μl^{-1} of sample DNA as the template in a total volume of 25 μl with a PCR program of 95 °C for 5 min, followed by 25 cycles each of 95 s for 30 s, 52 °C 1 min and 72 °C for 10 min. To detect any contamination during PCR preparation, negative controls (water in place of DNA) were included for all PCR reactions. PCR products of each subsample from the bar-coded primers were generated in six replicates and purified using the Wizard SV Gel and PCR Clean-Up System (Promega). Equimolar purified PCR products that were quantified by picogreen assays were mixed and sequenced using Roche Genome Sequencer FLX Titanium 454 sequencing platform (Macrogen Inc. Company, Seoul, Korea).

Amplicon sequence analysis

The raw data was processed using the QIIME v.1.6.0 pipeline (Caporaso *et al.*, 2010). Low quality sequences below 150 bp in length or with an average quality score below 25 were removed. After denoising the sequences using Denoiser 0.91 (Reeder and Knight, 2010), and testing for chimeras using USEARCH (Edgar *et al.*, 2011), OTUs were identified using the UCLUST 1.2.21 algorithm (Edgar, 2010) with a phylotype defined at the 97% sequence similarity level. The resulting OTUs were aligned against the Ribosomal Database Project database (Cole *et al.*, 2009).

Metagenomics library preparation for DNA shotgun sequencing

Shotgun metagenomic analyses were conducted on the soil DNA extracts following the illumine Pair-End Prep kit protocol with sequencing performed using 2 × 300 bp sequencing run on the Illumina MiSeq2000 (Macrogen Inc. Company). Paired-end reads were trimmed using Sickle (Joshi NA, 2011) with a minimum PHRED score of 30 and at least 150 bp in length. Next, a co-assembly of all data was made with Spades 3.1.1 (Bankevich *et al.*, 2012) at different k-mer lengths of 31, 91, 101 and 121. On the final assembly, genes were predicted using Prodigal 2.61 (Hyatt *et al.*, 2010) and converted from general feature format to general transfer format using cufflinks 2.1.1 (Trapnell *et al.*, 2010). Per sample reads were mapped to contigs using BamM 1.4.1 (Imelfort *et al.*, 2015) that uses BWA 0.7.12-r1039 (Li and Durbin, 2009) and samtools 1.2 (Li *et al.*, 2009). Next, the number of reads per sample mapping to genes was calculated using featureCounts (Liao *et al.*, 2014). To annotate the set of genes, hmmsearch 3.0 (Finn *et al.*, 2015) was used to screen the FOAM (Prestat *et al.*, 2014) set of Hidden Markov Models (release 1.0). Scripts provided by FOAM were used to select the best hit in the database. For each gene the best KO hits were added to the count matrix of featureCounts as a single column. Next, the KO column was aggregated using the Python Pandas library (McKinney, 2015). Hits to multiple KO terms were split equally. Finally, for each FOAM level a count matrix was made. The full analysis pipeline has been implemented in a Snakemake workflow (Koster and Rahmann, 2012).

Statistical analysis

Alpha diversity calculations were performed based on the rarefied OTU table to compare the diversity among samples at a given level of sampling effort (Hughes and Hellmann, 2005). The OTU table was rarefied to 1535 reads by 'single rarefaction' QIIME script as this number was the lowest number of reads for all samples. The average sequence reads from three sterilized controls were used as a baseline that was subtracted from the reads of all samples. The OTU table after this subtraction was used for further statistical analysis. We determined Chao1 richness, Simpson and Shannon diversity indices with the 'vegan' package (Dixon, 2003) in R (The R Foundation for Statistical Computing). The percentage coverage was calculated by the Good's method using the formula: % coverage = $[1 - (n/N)] \times 100$, where n is the number of phylotypes represented by singletons and N is the total number of sequences (Good, 1953). Good's method equation provides insight into the coverage of an entire sampled community by the data obtained from a limited data set. Two-way ANOVA was performed to test the effects of soil treatment (suspensions, soil and rhizosphere) and dilutions on the diversity indices.

Principal coordinates analysis (PCoA) matrices were used to visualize the community structure among samples, using the generated taxonomic and functional abundance matrices. The PCoA plots were generated from the Bray–Curtis similarity index matrices of all samples and created using the PAST software program (Hammer *et al.*, 2001). Differences in bacterial community composition among treatments were tested by analysis of similarities (ANOSIM). A two-way PERMANOVA analysis was performed to test the effects of soil treatment (soil and rhizosphere) and dilution on species composition and functional diversity. Analysis of the differences in relative abundance of taxa and functional genes were performed using the 'ggtern' package in R to rank taxa down to the genus level and level 2 of functional data (FOAM Database) per dilution group. The functions that were selected differed significantly between soil and rhizosphere for at least two dilutions, and the differences between soil and rhizosphere were in the same direction for all three dilutions. STAMP analysis was performed to get candidate species that were responsible for the separation within particular functional traits between soil and rhizosphere.

Network analyses were performed to gain a better understanding of the microorganism interactions in the soil and rhizosphere. Correlations among all OTUs were calculated with the Sparse Correlations for Compositional data algorithm (SparCC) (Friedman and Alm, 2012) implemented in mothur (Schloss *et al.*, 2009). The OTUs with less than three sequences were filtered as they were poorly represented. Only correlations with values above 0.5 or below −0.5 and a statistically significant *P* value lower than 0.05 were represented in the network using R, which were then visualized with the interactive platform Gephi (Bastian MHS, 2009).

All the analyses in this study were based on OTUs, except for diversity analysis within the phyla we considered here, which was based on the family level. Here, we only considered the most abundant, called dominant phyla, that is, the phyla of which the relative abundance is >2% of the total abundance in the 10^{-1} undiluted soil communities; we also included in our analyses classes of *Proteobacteria*.

Results

Diversity of the bacterial community in soil and rhizosphere

Dilution reduced the bacterial diversity in the suspension and in the soil after incubation and subsequently in the rhizosphere (Supplementary Table S1). The diversity indices of the rhizosphere samples were similar or in some cases significantly higher than the comparable indices of the soil samples. Both dilution and soil treatment (suspension, soil and rhizosphere) had significant effects on the species diversity with higher F values for the effect of dilution than of soil treatment for all parameters considered except for the Simpson index.

The dominant phyla considered here had contrasting reactions to the presence of plants; 14 significant differences were observed for the Shannon diversity indices between soil and rhizosphere samples, among which eight were significantly increased and six were significantly decreased (Table 1). More significant differences in the Shannon diversity indices for the various phyla in the soil versus the rhizosphere were observed in the 10^{-1}, undiluted samples than in the diluted ones. Good's estimator of coverage was above 97%.

Effects of dilution, soil and plant on bacterial community composition

After aligning OTUs with the RDP database, we identified the most dominant phyla in all samples, that is, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Acidobacteria*, *Verrucomicrobia*, *Planctomycetes* and *Firmicutes* (Supplementary Figure S1A). Information on the most relevant patterns in the relative

Table 1 Shannon diversity within abundant phyla in soil samples and rhizosphere samples

Phylum/family	*Soil* 10^{-1}	*Rhizosphere* 10^{-1}	P *value*	*Soil* 10^{-6}	*Rhizosphere* 10^{-6}	P *value*	*Soil* 10^{-9}	*Rhizosphere* 10^{-9}	P *value*
Acidobacteria	1.16 ± 0.04	1.39 ± 0.04	*	0.85 ± 0.07	0.75 ± 0.10	NS	0.54 ± 0.15	0.57 ± 0.11	NS
Actinobacteria	2.34 ± 0.03	1.75 ± 0.08	*	1.78 ± 0.07	1.19 ± 0.13	*	1.46 ± 0.16	1.38 ± 0.15	NS
Bacteroidetes	1.29 ± 0.04	1.14 ± 0.05	*	1.27 ± 0.06	1.08 ± 0.08	NS	1.16 ± 0.07	1.31 ± 0.05	NS
Firmicutes	1.04 ± 0.04	0.91 ± 0.04	NS	0.23 ± 0.12	0.90 ± 0.07	*	0.52 ± 0.11	0.92 ± 0.06	*
Verrucomicrobia	1.23 ± 0.03	1.34 ± 0.09	*	0.96 ± 0.06	0.98 ± 0.10	NS	0.81 ± 0.07	0.77 ± 0.11	*
Alphaproteobacteria	1.88 ± 0.02	2.14 ± 0.01	*	1.69 ± 0.04	1.95 ± 0.03	*	1.37 ± 0.12	1.99 ± 0.04	*
Betaproteobacteria	1.50 ± 0.03	1.25 ± 0.01	*	0.75 ± 0.14	0.91 ± 0.08	NS	0.91 ± 0.08	0.47 ± 0.10	*
Deltaproteobacteria	1.31 ± 0.08	1.16 ± 0.11	NS	0.78 ± 0.13	0.87 ± 0.08	NS	0.74 ± 0.09	0.87 ± 0.12	NS
Gammaproteobacteria	0.94 ± 0.04	1.11 ± 0.07	*	0.95 ± 0.07	0.72 ± 0.09	NS	0.47 ± 0.11	0.66 ± 0.12	NS

Abbreviation: NS, not significant; *, statistical significance ($P<0.05$).
Diversity was calculated for each dilution of incubated soil and rhizosphere samples ($n=5$–6) as well as significant comparisons (ANOVA; $P<0.05$) within abundant phyla among phylogenetic profile (family level).

abundances at the phylum and family level is provided in Supplementary Figures S1A and B.

To visualize differences in community structure between the six groups (three dilutions for the soil and the rhizosphere), taxonomic abundances were used to compute the Bray–Curtis similarity matrices (Figure 1a). Rhizosphere samples were clearly separately from the soil samples (ANOSIM, $R=0.36$, $P<0.001$). The individual effects of soil treatment (here, soil and rhizosphere) and dilution on the species composition were quantified by two-way PERMANOVA (Supplementary Table S2). On the basis of the associated F values, the soil treatment effect was slightly stronger than the dilution effect, and both main effects were larger than their interaction. A PCoA representing the taxonomic compositions of the soil samples showed a strong separation of the three dilutions (Figure 1c; $R=0.80$, $P<0.001$). In contrast, rhizosphere samples of the three dilutions were more clustered together although still distinct (Figure 1e; $R=0.49$, $P<0.001$).

There were marked differences in the network analysis of the soil and rhizosphere samples for all three dilutions (Figures 2a and b). In general, the number of correlations in the rhizosphere was larger than in the soil (Supplementary Table S3), and the number of positive correlations was higher than negative ones for both soil and rhizosphere samples. Betweenness Centrality (BC) of the rhizosphere community networks was much stronger than that of the soil communities, decreasing gradually on dilution (Figure 2c). In the 10^{-9} diluted samples of the rhizosphere communities, no potential keystone species were obtained (Figure 2c).

Effects of dilution, soil and plant on the functional potential of the bacterial community

The functional profiles of rhizosphere samples were separated from the incubated soil samples based on the Bray–Curtis similarity matrices (Figure 1b, $R=0.08$, $P=0.046$). Both soil treatment (soil and rhizosphere) and dilution as well as their interaction affected functional diversity (Supplementary Table S2). On the basis of the associated F value, soil treatment and dilution had similar effects on the functional traits. The PCoA plot of functional profiles of the different dilutions of rhizosphere samples showed a higher similarity than those of soil samples (Figures 1d and f; soil: $R=0.59$, $P=0.0001$; rhizosphere: $R=0.25$, $P=0.02$). The functional profiles of the soil samples differed significantly among the dilutions, but in rhizosphere the only significant difference in the functional profiles was between the undiluted (10^{-1}) and the most diluted samples (10^{-9}).

The functional profiles of the soil and rhizosphere communities overlapped more as compared to the species community structures (Figure 1). Similarly, the functional genes of all three dilutions of both soil and rhizosphere samples were more strongly centered in the ternary plot than were OTUs (Supplementary Figures. S2C and D). To compare the similarity among replicate samples of the six groups, we calculated the mean values of the Bray–Curtis similarity for both the taxonomic and functional data. Replicates of functional data within each dilution were highly similar (light gray bars in Figure 1g), whereas the taxonomic similarity decreased on dilutions for both soil and rhizosphere samples (dark gray bars in Figure 1g).

A higher number of significant differences in the functional traits of soil and rhizosphere were observed in the diluted communities than in the undiluted 10^{-1} communities (Figure 3). One of the

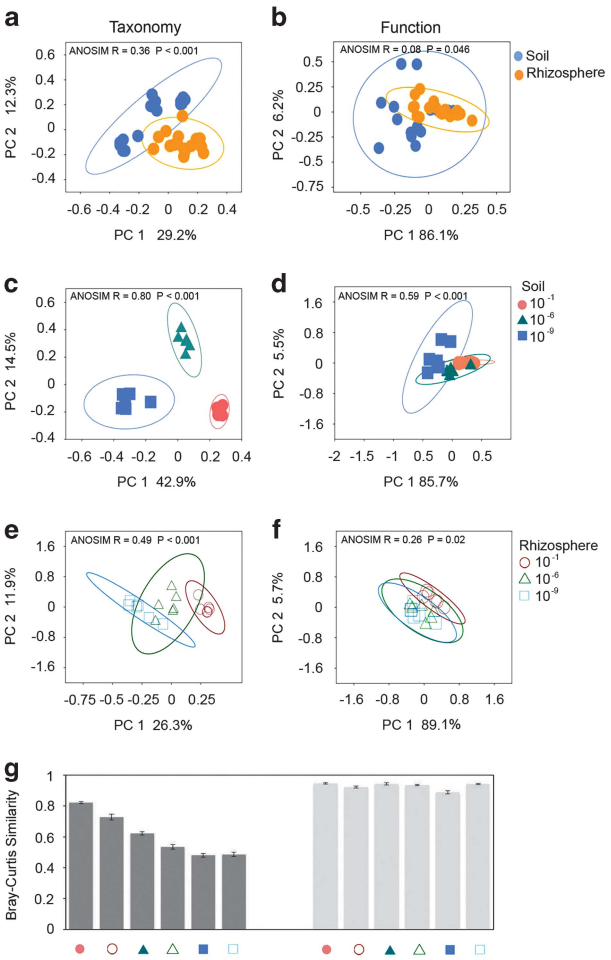

Figure 1 PCoA of the soil and rhizosphere bacteria community compositions and functional traits. (**a**) Variation between samples of soil and rhizosphere based on the Bray–Curtis similarity for taxonomical data and (**b**) functional traits using relative abundances based on FOAM level 2. Variation between dilutions of soil samples based on the Bray–Curtis similarity for taxonomical data (**c**) and functional traits (**d**). Variation between dilutions of rhizosphere samples based on the Bray–Curtis similarity for taxonomical data (**e**) and functional traits (**f**). Similarity values (analysis of similarity) are shown in the upper left of each plot. Similarities between replicates of each dilution are shown in (**g**) Dark gray bars indicate taxonomical data and light gray bars indicate functional traits. The error bars show standard errors of six replicates.

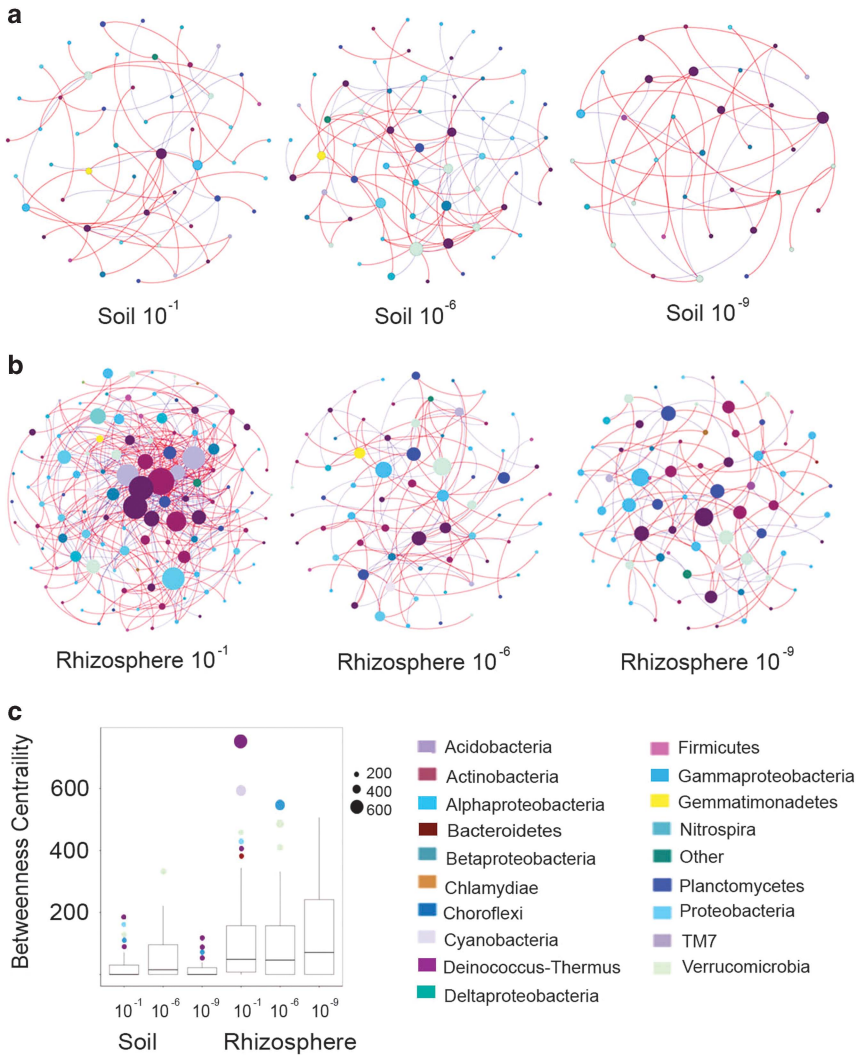

Figure 2 Co-occurrence patterns of bacteria in soil and rhizosphere. Correlations were presented in the soil samples (**a**) and in the rhizosphere samples of each dilution (**b**). Nodes indicate taxonomic affiliation at genus level. Red lines indicate positive correlations, and blue lines indicate negative correlations. The color of each node indicates the phylum shown below of the figures. The size of each node is proportional to the BC (**c**). The box-and-whiskers graphics show the median of BC as a line, the 25th and 27th percentiles of the data as the top and bottom of the box, and outlier dots to indicate the most extreme data point within 1.5*(75th–25th percentile) of the median. The size of outlier data points corresponds to the value of the BC.

most abundant types of genes, the 'transporters' genes, was significantly over-represented in the rhizosphere of all samples. This was also observed for the functions related to 'Embden–Meyerhof–Parnas (EMP) pathway' and 'hydrogen metabolism' in the rhizosphere of at least two dilutions. By contrast, the core functions related to 'cellular response to stress' and 'carbohydrate active enzymes' were more abundant in the soil than in the rhizosphere.

To further investigate differences in the functional traits of the soil and rhizosphere communities, we binned species within selected functions and then compared the species composition of the soil and the rhizosphere. The functions that were selected differed significantly ($P<0.05$) in soil and rhizosphere samples in at least two dilutions and these differences were in the same direction for all three dilutions. When testing the functions that were more abundant in the rhizosphere than in the soil, for example, 'transporters', 'EMP pathway' and 'hydrogen metabolism', we found that rhizosphere communities were clustered and significantly ($P<0.05$) separated from soil communities (Figure 3b). However, when testing functions that were more abundant in the soil than in the rhizosphere, for example, 'cellular response to stress' and 'carbohydrate active enzymes', we observed that soil and rhizosphere communities were not significantly separated (Figure 3b). Although we should be cautious with the interpretation of these results (the analysis is based on only 5 groups of functional traits), this seems to suggest that selection in the rhizosphere is directed towards, rather than against, species with particular functional traits.

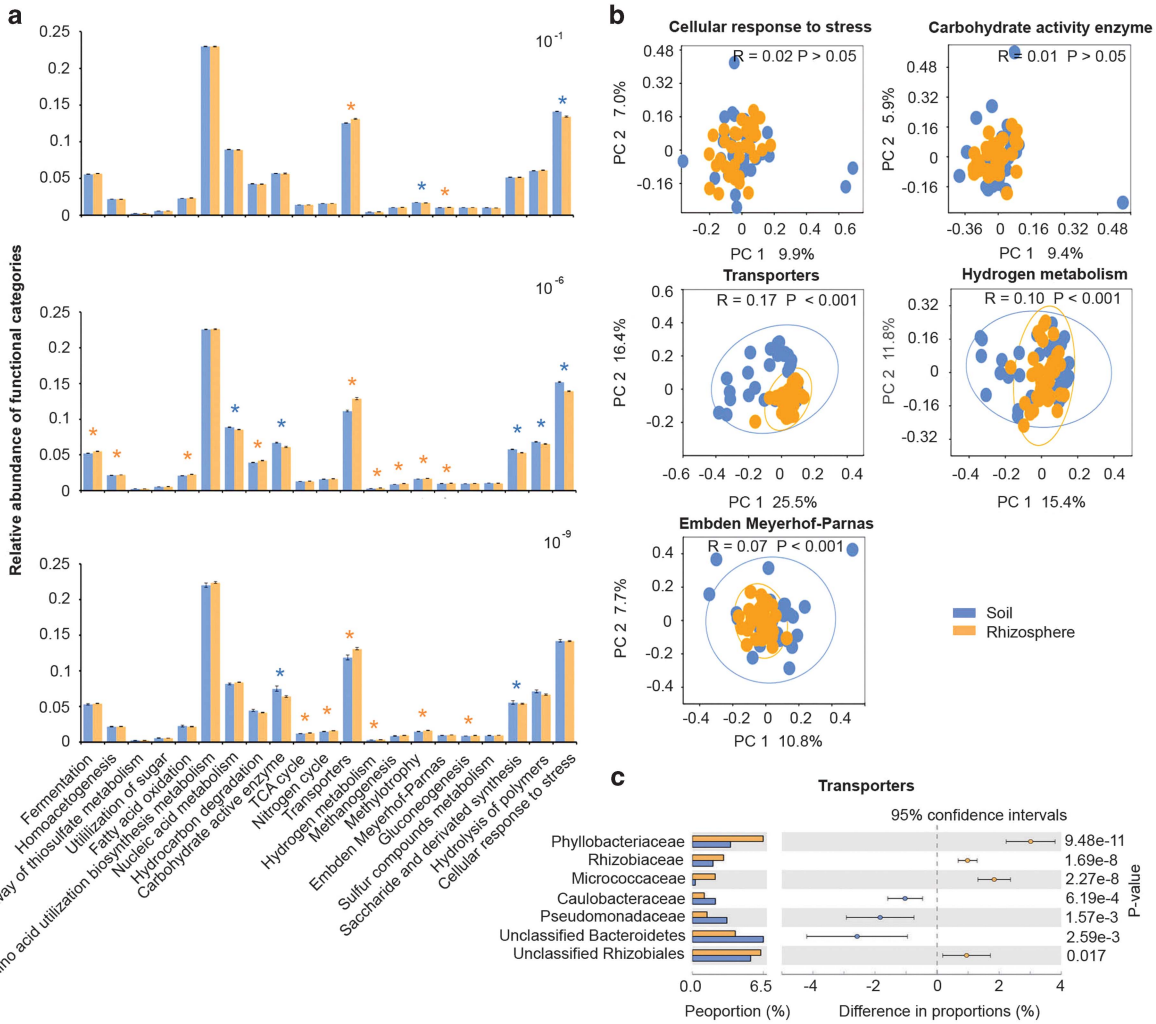

Figure 3 Profiles of soil and rhizosphere bacterial community functional traits. (**a**) The relative abundance of groups of functional genes in soil and rhizosphere for three dilutions. Relative abundance of functional genes (FOAM Database level 1) based on normalized shotgun metagenomics data in dilutions of 10^{-1}, 10^{-6} and 10^{-9}. The percentage of the total sequence reads in samples from soil and rhizosphere is presented for each dilution. The error bars show standard errors of six replicates and orange asterisks (*) indicate categories that are significantly more abundant in rhizosphere samples (P value <0.05) and blue asterisks (*) indicate categories that are significantly more abundant in soil samples. (**b**) PCoA plots of species with particular functional genes that were more abundant in the soil than in the rhizosphere (cellular response to stress and carbohydrate activity enzymes) and plots of species with particular functional genes that were more abundant in the rhizosphere than in the soil (transporter genes, EMP pathway and hydrogen metabolism). Similarity values are shown in the upper right corner of each plot. The circles represent the clustering of the soil and rhizosphere samples, respectively. (**c**) Differences in abundance of families with transporters between soil and rhizosphere samples (Welch's t-test; $P<0.05$).

As an illustration of the changes in the composition of the communities involved in these functions in soil and rhizosphere, we identified the species as detected by metagenomic shotgun data analysis that were involved in the 'transporters' function, which differed in abundance between soil and rhizosphere samples. STAMP analysis showed that *Phyllobacteriaceae*, *Rhizobiaceae*, unclassified *Rhizobiales* and *Micrococcaceae* were the major families based on PC1 score (with abundance above 1%) responsible for the PCA separations in the rhizosphere (Figure 3c). In contrast, *Caulobacteraceae*, unclassified *Bacteroidetes*, and surprisingly, *Pseudomonadaceae* were over-represented in the soil.

Discussion

The dilution approach applied here is one of the few available methods to manipulate microbial biodiversity of complex natural ecosystems such as the soil (Pedros-Alio, 2006; Hol *et al.*, 2010; Philippot *et al.*, 2013; Yan *et al.*, 2015). And, indeed, our results show that dilution reduces the microbial biodiversity in the soil suspension, and the soil and rhizosphere after incubation of more or less diluted suspensions as we showed earlier (Yan *et al.*, 2015) (Supplementary Table S1). The present results indicate that the dilution procedure leads to reduction of bacterial diversity, but there is less clear differentiation between soil and rhizosphere in terms

of community diversity. In some cases the diversity indices measured here are larger in the rhizosphere than in the soil. Also, the number of species detected in the rhizosphere was, sometimes, larger than in the bulk soil. Considering that we used sterile plants, it is fair to assume that the plants did not add a substantial inoculum to the community. Presumably, the depth of sequencing is still not sufficient to encompass the entire microbial community in soil, although Good's estimator of coverage was always above 0.97.

There was more similarity between the different dilutions of the rhizosphere samples than between different dilutions of soil samples (Figure 1). This shows that convergence took place in the rhizosphere as a direct or indirect selective effect of the roots. This is especially true for the functional traits. Yet, there were significant effects between soil and rhizosphere communities and between dilutions both on the basis of species composition (Figure 1a; Supplementary Table S2) and to a lesser extent of functional traits (Figure 1b; Supplementary Table S2). A selective change in the microbial community structure of the rhizosphere has also been reported in many other studies (Duineveld *et al.*, 1998; Mendes *et al.*, 2011; Mendes *et al.*, 2014) and plant hosts (Ofek-Lalzar *et al.*, 2014; Bulgarelli *et al.*, 2015), and soil characteristics (Kuramae *et al.*, 2012) may contribute to this. The variance among samples appeared to be smaller in the rhizosphere than in the bulk soil suggesting a homogenizing effect of the plant on the structure of the communities. This effect was larger in the functional profile of the community than in its taxonomic structure. Moreover, there were less significant differences in the Shannon diversity of the dominant phyla between soil and rhizosphere samples in the diluted samples than the 10^{-1} undiluted samples (Table 1). These results point to the selective power of the plant in the structuring of the bacterial rhizosphere community. The only differences in functionality between dilutions observed in the rhizosphere were between 10^{-1} un-diluted and 10^{-9} diluted communities. These observations suggested considerable overlap in species' functional capabilities in communities with different diversity. This is consistent with the results of a large number of studies, which provide evidence of the existence of a large degree of functional redundancy in soil (Nannipieri *et al.*, 2003; Mendes *et al.*, 2015; Sunagawa *et al.*, 2015).

Our results regarding both species composition and functional traits indicate that the plant exerts selection on the microbial community in the rhizosphere based on particular functional traits, which may occur directly or through changes in abiotic environmental factors. The enrichment processes in the rhizosphere selected microbes with specific functional genes in particular related to 'transporters', 'EMP pathway' and 'hydrogen metabolism'. These three functional cores that were over-represented in the rhizosphere suggest that the rhizosphere selects specific species based on functional traits. These functions appeared to be relevant for interactions with the plant. Some of these features have also been shown by others to be important in rhizosphere communities (Mendes *et al.*, 2014; Ofek-Lalzar *et al.*, 2014; Bulgarelli *et al.*, 2015). Consistently with our study, in particular transporter systems were found to be of great importance in the rhizosphere. Therefore, we focused on 'transporters' genes to get a better understanding of the functional selection process in the rhizosphere by the STAMP analysis.

A clear separation between soil and rhizosphere samples was found for species with particular functional traits only if these were over-represented in the rhizosphere samples (Figure 3b). The latter suggests that the above mentioned rhizosphere selection process across species was not random. As an example we showed that few specific species containing the 'transporters' functions were selected in the rhizosphere. The species found belonged to the families of *Phyllobacteriaceae*, *Rhizobiaceae*, unclassified *Rhizobiales* and *Micrococcaceae*. At least some species in these families have been reported earlier as beneficial to plants (Sanguin *et al.*, 2009; Hayat *et al.*, 2010). Remarkably, genes of species belonging to the family of the *Pseudomonadaceae,* which are considered generally as typical rhizosphere organisms (Mendes *et al.*, 2011), were found to be more abundant in soil. This may question the role of this family in plant–microbe interactions. We only focused on species involved in transporters functions as an illustration of the details of the taxonomic analysis that is possible on the basis of the metagenomics shotgun data. However, these analyses of the composition of species community involved in the functional traits must be taken with caution: in our experience usually only 25–30% of the reads of the assembled shotgun data can be annotated. Thus, in our opinion, the used approach does not allow for more detailed considerations, because of the weak coverage of the sequence data.

Nevertheless, the conclusion is justified that the core functional genes selected in the rhizosphere are not restricted to one particular taxonomic group. This is consistent with a report on the *Ulva australis* (marine alga) that showed that they selected functional genes, rather than taxonomic relatedness (Burke *et al.*, 2011). If, indeed, the selection process in the rhizosphere is also based on functional traits, and these specific functional traits are not randomly distributed over all bacterial phyla detected here, this may be an explanation for the variation in the taxonomic diversity of the different phyla as presented in Table 1.

The network analysis revealed many more correlations and potential keystone species in the rhizosphere than in the soil (Figure 3). This indicates that the network architecture was more stable and had

more complex connections in the rhizosphere than in the soil. This is what we expected given the stronger selection observed on the bacterial community in the rhizosphere than in the soil. We based our network analysis on 16S ribosomal RNA amplicon data and not on the binned shotgun data because of the above mentioned low annotation rate of the sequences.

In conclusion, our results indicate that the dilution procedure leads to reduction of bacterial diversity, but the outcome of the community assembly processes during incubation in soil and rhizosphere cannot be predicted on the basis of the composition of the inoculum. Soils have a strong selective power in shaping the bacterial community, which leads to more uniform structures of the communities even after inoculation of variate suspensions. Further selection takes place in the rhizosphere. This rhizosphere selection seems to be dominated by particular functional traits. To what extent this selection is controlled by the plants or is caused by indirect factors remains to be investigated. At this point, the categorization of the functional genes is too broad to relate these genes to potential effects on plant fitness. This conclusion is based on the findings that the relative abundance of some particular functional genes in the rhizosphere was higher than in soil, suggesting that the rhizosphere selects for these functional traits rather than against them. Although the mechanisms and consequences of the functional selection in the rhizosphere for plant fitness remain unclear, the present results add valuable information to better understand the highly complex processes of microbial community assemblage in both soil and rhizosphere.

Data availability

Raw 454-pyrosequencing data have been deposited to EMBL database (European Nucleotide Archive): https://www.ebi.ac.uk/ena/data/view/PRJEB12988.
Shotgun metagenomics data have been deposited to EMBL database (European Nucleotide Archive): https://www.ebi.ac.uk/metagenomics/projects/ERP009103.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank A Pijl for her technical assistance and N Cassman for English editing of the manuscript. Publication number 6123 of the NIOO-KNAW, Netherlands Institute of Ecology.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

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

Berendsen RL, Pieterse CMJ, Bakker PAHM. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci* **17**: 478–486.

Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA *et al.* (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol Biochem* **43**: 1450–1455.

Bulgarelli D, Garrido-Oter R, Munch PC, Weiman A, Droge J, Pan Y *et al.* (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* **17**: 392–403.

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

Butchart SHM, Walpole M, Collen B, van Strien A, Scharlemann JPW, Almond REA *et al.* (2010). Global biodiversity: indicators of recent declines. *Science* **328**: 1164–1168.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Dixon P. (2003). VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**: 927–930.

Duineveld BM, Rosado AS, van Elsas JD, van Veen JA. (1998). Analysis of the dynamics of bacterial communities in the rhizosphere of the chrysanthemum via denaturing gradient gel electrophoresis and substrate utilization patterns. *Appl Environ Microbiol* **64**: 4950–4957.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

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

Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F *et al.* (2015). HMMER web server: 2015 update. *Nucleic Acids Res* **43**: W30–W38.

Franklin RB, Garland JL, Bolster CH, Mills AL. (2001). Impact of dilution on microbial community structure and functional potential: Comparison of numerical simulations and batch culture experiments. *Appl Environ Microbiol* **67**: 702–712.

Franklin RB, Mills AL. (2006). Structural and functional responses of a sewage microbial community to dilution-induced reductions in diversity. *Microb Ecol* **52**: 280–288.

Friedman J, Alm EJ. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput Biol* **8**: e1002687.

Garland JL, Lehman RM. (1999). Dilution/extinction of community phenotypic characters to estimate relative structural diversity in mixed communities. *FEMS Microbiol Ecol* **30**: 333–343.

Good IJ. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237–264.

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

Hayat R, Ali S, Amara U, Khalid R, Ahmed I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann Microbiol* **60**: 579–598.

Hol WHG, de Boer W, Termorshuizen AJ, Meyer KM, Schneider JHM, van Dam NM *et al.* (2010). Reduction of rare soil microbes modifies plant-herbivore interactions. *Ecol Lett* **13**: 292–301.

Hughes JB, Hellmann JJ. (2005). The application of rarefaction techniques to molecular inventories of microbial diversity. *Environ Microbiol* **397**: 292–308.

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Imelfort M, Woodcroft B, Parks D. (2015). Ecogenomics/BamM [WWW Document]. GitHub. Available at: https://github.com/Ecogenomics/BamM (accessed on 21 July 2015).

Joosten L, Mulder PPJ, Klinkhamer PGL, van Veen JA. (2009). Soil-borne microorganisms and soil-type affect pyrrolizidine alkaloids in *Jacobaea vulgaris*. *Plant Soil* **325**: 133–143.

Joshi NA FJ. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at: https://github.com/najoshi/sickle (accessed on 21 July 2015).

Koster J, Rahmann S. (2012). Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **28**: 2520–2522.

Kuramae EE, Yergeau E, Wong LC, Pijl AS, van Veen JA, Kowalchuk GA. (2012). Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol Ecol* **79**: 12–24.

Langenheder S, Szekely AJ. (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J* **5**: 1086–1094.

Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Liao Y, Smyth GK, Shi W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Magurran AE, Henderson PA. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714–716.

Matos A, Kerkhof L, Garland JL. (2005). Effects of microbial community diversity on the survival of Pseudomonas aeruginosa in the wheat rhizosphere. *Microb Ecol* **49**: 257–264.

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

McKinney W. (2015), Python Data Analysis Library—pandas: Python Data Analysis Library [WWW Document]. Available at: http://pandas.pydata.org/ (accessed on 22 July 2015).

Mendes LW, Kuramae EE, Navarrete AA, van Veen JA, Tsai SM. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J* **8**: 1577–1587.

Mendes LW, Tsai SM, Navarrete AA, de Hollander M, van Veen JA, Kuramae EE. (2015). Soil-borne microbiome: linking diversity to function. *Microb Ecol* **70**: 255–265.

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM *et al.* (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**: 1097–1100.

Nannipieri P, Ascher J, Ceccherini MT, Landi L, Pietramellara G, Renella G. (2003). Microbial diversity and soil functions. *Eur J Soil Sci* **54**: 655–670.

Ofek-Lalzar M, Sela N, Goldman-Voronov M, Green SJ, Hadar Y, Minz D. (2014). Niche and host-associated functional signatures of the root surface microbiome. *Nat Commun* **5**: 4950.

Pedros-Alio C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.

Philippot L, Spor A, Henault C, Bru D, Bizouard F, Jones CM *et al.* (2013). Loss in microbial diversity affects nitrogen cycling in soil. *ISME J* **7**: 1609–1619.

Prestat E, David MM, Hultman J, Tas N, Lamendella R, Dvornik J *et al.* (2014). FOAM (functional ontology assignments for metagenomes): a Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res* **42**: e145.

Reeder J, Knight R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* **7**: 668–669.

Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R *et al.* (2000). Biodiversity—global biodiversity scenarios for the year 2100. *Science* **287**: 1770–1774.

Salonius PO. (1981). Metabolic capabilities of forest soil microbial-populations with reduced species-diversity. *Soil Biol Biochem* **13**: 1–10.

Sanguin H, Sarniguet A, Gazengel K, Moenne-Loccoz Y, Grundmann GL. (2009). Rhizosphere bacterial communities associated with disease suppressiveness stages of take-all decline in wheat monoculture. *New Phytol* **184**: 694–707.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Stegen JC, Lin XJ, Konopka AE, Fredrickson JK. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653–1664.

Steiner. (1968). *Soiless Culture. Proceedings of the 6th Colloquium of the International Potash Institute,*

Florence, Italy. International Potash Institute: Bern, Switzerland, pp 324–341.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G *et al.* (2015). Structure and function of the global ocean microbiome. *Science* **348**: 1261359.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ *et al.* (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–U174.

Vivant AL, Garmyn D, Maron PA, Nowak V, Piveteau P. (2013). Microbial diversity and structure are drivers of the biological barrier effect against *Listeria monocytogenes* in soil. *PLoS One* **8**: e76991.

Worm B, Duffy JE. (2003). Biodiversity, productivity and stability in real food webs. *Trends Ecol Evol* **18**: 628–632.

Yan Y, Kuramae EE, Klinkhamer PG, van Veen JA. (2015). Revisiting the dilution procedure used to manipulate microbial biodiversity in terrestrial systems. *Appl Environ Microbiol* **81**: 4246–4252.

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