Assemblage and functioning of bacterial communities in soil and rhizosphere

Yan, Y.

Chapter 4

Functional traits determine rhizosphere selection of bacterial communities

Yan Yan, Eiko E. Kuramae, Mattias de Hollander, Peter G.L. Klinkhamer, Johannes A. van Veen

In revision

Abstract

We studied the selective effects of the rhizosphere on bacterial communities of different diversity 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 rDNA 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. We found clear differences between the soil and the rhizosphere bacterial community of each dilution at the OTU level. In many cases, the species diversity within a phylum differed significantly between soil and rhizosphere. Network analysis revealed stronger interactions among bacterial OTUs in the rhizosphere than in the soil. The enrichment processes in the rhizosphere selected microbes with particular functional genes related to transporters, Embden Meyerhof Parnas pathway and hydrogen metabolism. The species with particular functional traits that were over-represented in the rhizosphere samples differed between soil and rhizosphere samples. This selection was not random across bacteria with these functional traits. Overall this suggests selection on the bacterial community of the rhizosphere based on functional traits.

Keywords

Microbial diversity | Rhizosphere selection| Community structure| Microbial functions

4.1. 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). 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 fascinating 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 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 plants. In this study, as in many others (Salonius 1981, Garland and Lehman 1999, Franklin et al 2001, Matos et al 2005, Franklin and Mills 2006, Wertz et al 2006, Hol et al 2010, Philippot et al 2013, Vivant et al 2013), the experimental approach is based on the assumption that the diversity of the microbial community in the soil can be altered by inoculating diluted suspensions in a pre-sterilized soil. Although this method has been used frequently in the past, little is known about how the assembly of bacterial

communities in the soil and in the rhizosphere proceeds after inoculation. More specifically, until now, we have ignored the question of whether or not functional characteristics of the microbial community play a role in the selection of microbial species in soil and rhizosphere and if so, how. Recent advances in high-throughput sequencing now allow for the assessment of both the taxonomic composition and function of the rhizosphere microbiome (Bulgarelli et al 2015), which enables us to address this question.

The major aim of this study was to acquire a better understanding of microbial 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 soil. After an established incubation period, plants were potted in the various 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 rRNA 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 Chapters 2 and 3 we found that the soil has a strong impact on the assemblage of bacterial communities after incubation of various diluted inocula. We hypothesized that plants will exert a further selection at both taxonomic and functional trait levels. In particular, we studied whether species selection in the rhizosphere exerts an effect on functional traits of the microbes and, if so, whether this selection is random across species with these traits or species-specific.

4.2. Materials and Methods

4.2.1. Soil sampling and plant selection

Thirty liters of soil were collected at a depth of 15 cm from a dune soil in Meijendel, The Netherlands. Soil organic matter content (%) was 9.11 ± 0.36 (n=6), soil pH was 7.4 ± 0.005 (n=6), NO^{3-} content (mg/kg) was 30.43 ± 0.85 (n=6), NH^{4+} content (mg/kg) was 2.23 ± 0.25 (n=6), P content (mg/kg) was 15.16 ± 0.41 (n=6). The soil was sieved and homogenized and stored in 500 g aliquots in plastic bags. One bag of soil was kept separately to prepare the

inoculum. All the 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 TSA and PDA media. No bacterial and fungal growth was observed on agar plates with the sterilized soil after 6 days for 6 replicates. Three sterilized soil bags inoculated with sterilized water were used as a control for the community assemblage during the entire experimental period. A subsample of the fresh soil was used to determine soil moisture (24 h, 105 ºC). For the dilution treatments, a 10 % suspension of untreated soil in sterilized water (10^{-1}) was sequentially diluted to obtain further dilutions 10^{-6} and 10^{-9} and these were added to the bags with the sterilized soil. The 10^{-1} suspension was considered to be the undiluted treatment.

Jacobaea vulgaris was selected 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). Since tissue culture has often been defined as the "sterile" plant, it was reasonable to use the "clean" cloned plants for the further experiments. After 8 weeks of incubation of the inoculated soils, tissue culture plants were potted in 0.5 L pots containing the incubated soil. Samples were taken from the bulk soil at the moment of planting. After 6 weeks of plant growth, plants were harvested 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 ºC for further analysis. The design of the experiment included 3 dilutions with 6 replicates each for both the incubated bulk soil and rhizosphere soil samples. 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 (Chapter 2). The impact of bacterial and extracellular DNA left in soil after sterilization prior to inoculation was accounted for by subtracting the OTUs found in the non-inoculated samples from those detected in the inoculated samples (Chapter 2).

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

Total DNA was extracted from the incubated bulk soil and rhizosphere soil to determine the composition of the respective microbial communities by 454-pyrosequencing of the 16S rDNA genetic marker. The DNA was extracted using the MoBio Power Soil Extraction Kit according to the supplier's manual (MO BIO Laboratories, Carlsbad, CA, USA). Total DNA concentration was quantified on an ND-1000 spectrophotometer (Nanodrop Technology, Wilmington, DE). 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), 1 unit of *Taq* polymerase (Roche, Indianapolis, IN), and 5 ng/µl of sample DNA as the template in a total volume of 25 µl with a PCR program of 95 ºC for 5 min, followed by 25 cycles each of 95 s for 30 s, 52 ºC 1 min and 72 ºC 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 barcoded 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, Seoul, Korea).

4.2.3. 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), Operational Taxonomic Units (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).

4.2.4. 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, South Korea). Paired end reads were trimmed using Sickle (Joshi and Fass, 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 GFF (General Feature Format) to GTF (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 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. 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).

4.2.5. 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 1,535 reads by "single rarefaction" QIIME script since this number was the lowest number of reads for all samples. The average sequence reads from 3 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 Good's method using the formula: % coverage = [1-(n/N)] × 100, where n is the number of phylotypes represented by singletons and N is the total number of sequences (Good 1953).

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 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). Differential 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) according to the contributions of the dilution groups. 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.

Network analyses were performed to gain a better understanding of the microbial interactions in the soil and rhizosphere. Correlations amongst 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 since 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 (R development Core Team, 2008), 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 particular phyla that were based on the family level.

4.3. Results

4.3.1. Diversity of the bacterial community in soil and rhizosphere

Remarkably, dilution had a stronger effect on the diversity indices than had the rhizosphere selection and in most cases the number of species detected and the diversity indices were similar or higher in the rhizosphere than in the bulk soil (Table 4.1).

Table 4.1. Estimators of microbial diversity and coverage in incubated soils and rhizosphere.

Treatment	Dilution	S.obs	S.chao-1	Shannon	Simpson	Good's estimator of coverage
Soil	10^{-1}	107.20±1.27	134.37±2.96	3.719±0.019	0.954±0.002	97.56±0.11
Rhizosphere	10^{-1}	113.88±2.34	141.85±6.99	3.747±0.049	0.952±0.004	97.77±0.14
Soil	10^{-6}	70.09±2.13	89.64±4.46	3.208±0.040	0.934±0.004	97.95±0.21
Rhizosphere	10^{-6}	85.25±1.28	110.78±4.38	3.334±0.046	0.928±0.005	98.24±0.10
Soil	10^{-9}	55.83±1.14	81.82±3.37	2.633±0.042	0.867±0.006	97.27±0.24
Rhizosphere	10^{-9}	76.36±2.45	95.83±3.92	3.209±0.097	0.916±0.012	98.23±0.16
Two-way ANOVA	Dilution	F=301.5 ***	F=65.26 ***	F=46.11 ***	F=104.8 ***	
	Soil/rhizosphere	F=82.49 ***	F=14.01 ***	F=7.032 **	F=28.85 ***	
	Interaction	F=7.613 **	F=1.332	F=10.9 ***	F=14.15 ***	

Estimators and statistical significance were calculated for each dilution treatment of soil and rhizosphere samples (n = 5-6) based on phylogenetic profiles at the species level. S.obs is the observed number of OTUs. NS means not significant. Results from two-way ANOVA comparisons of estimators diversity are given the *F*-statistic and *P* value (indicated by asterisk: *** $P < 0.001$; ** $P < 0.01$).

The dominant phyla detected in this experiment had contrasting reactions to the presence of plants; the (Shannon) diversity indices for dominant phyla were higher in the rhizosphere than in the bulk soil, and vice versa (Table 4.2). The strongest differences between the diversity indices of the soil *versus* the rhizosphere samples were found in the undiluted 10^{-1} inocula. The rhizosphere samples showed more statistically significant differences within various phyla than did the soil samples for the diluted inocula. Good's estimator of coverage was above 97%.

Table 4.2. Shannon diversity within major phyla in incubated soil and rhizosphere samples.

Phylum/Family	Soil	Rhizosphere	*P*	Soil	Rhizosphere	*P*	Soil	Rhizosphere	*P*
	10^{-1}			10^{-6}			10^{-9}		
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

Diversity and statistical significance ($P < 0.05$) was calculated for each dilution of incubated soil and rhizosphere samples (n = 5-6) within the major phyla based on phylogenetic profiles at the family level. NS means not significant.

4.3.2. 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, *i.e., Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Acidobacteria*, *Verrucomicrobia*, *Planctomycetes* and *Firmicutes* (Fig. 4.1A). Information on the most relevant patterns in the relative abundances at the phylum and family level is provided in Figure 4.1 A and B.

To visualize differences in community structure between the six groups (three dilutions for the incubated soil and the rhizosphere), taxonomic abundances were used to compute the Bray-Curtis similarity matrices (Fig. 4.2A). Rhizosphere samples were clearly separately from the incubated soil samples (ANOSIM, R = 0.36, $P < 0.001$). A PCoA representing the taxonomic compositions of the soil samples showed a strong separation of the three dilutions (Fig. 4.2C; R= 0.80, $P < 0.001$). In contrast, rhizosphere samples of the three dilutions were more clustered together although still distinct (Fig. 4.2E; R = 0.49, $P < 0.001$).

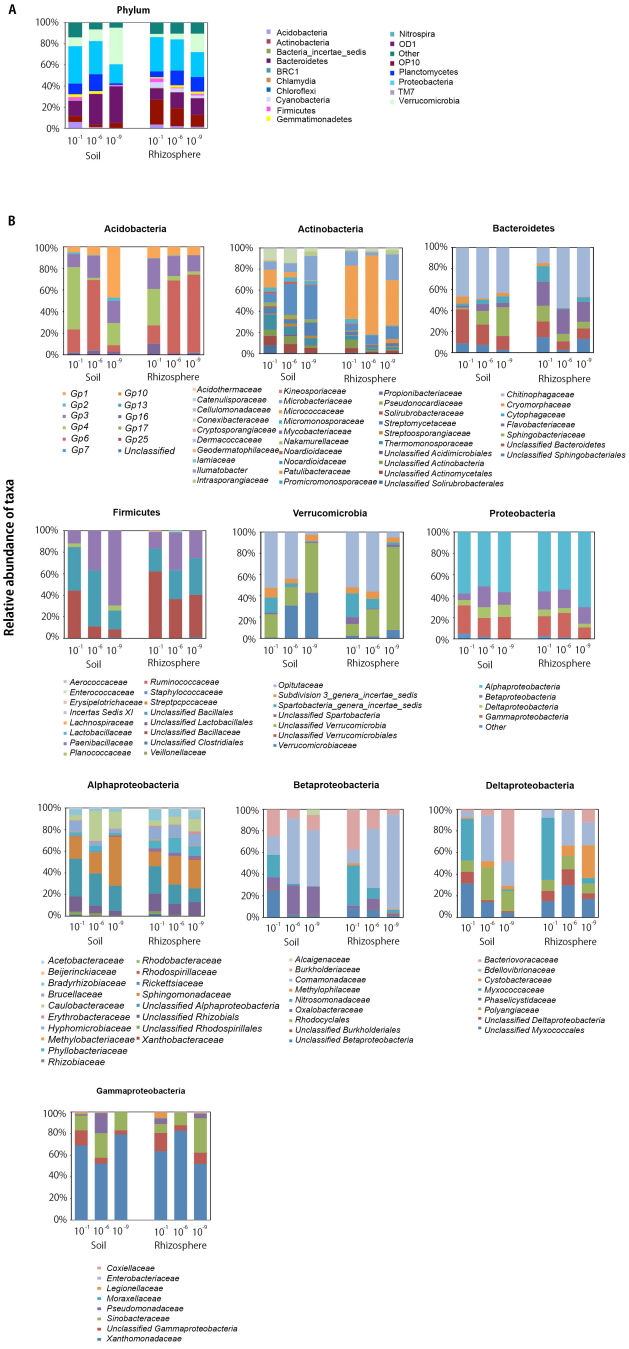

Figure 4.1. Profiles of soil and rhizosphere bacterial communities at the phylum (A) and family level (B) expressed as relative abundances.

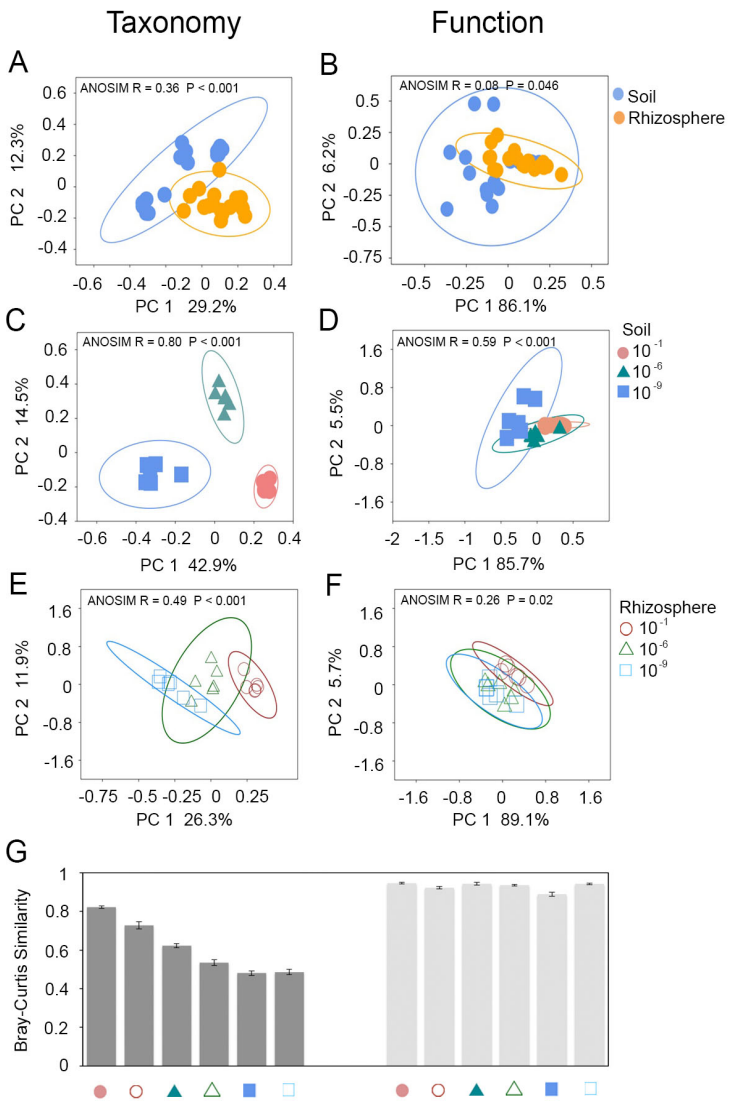

Figure 4.2. Principal Coordinate Analysis of the soil and rhizosphere bacteria community compositions and functional traits. (A) Variation between samples of soil and rhizosphere based on Bray-Curtis similarity for taxonomical data and (B) functional traits using relative abundances based on FOAM 'level 1'. Variation between dilutions of soil samples based on Bray-Curtis similarity for taxonomical data (C) and functional traits (D). Variation between dilutions of rhizosphere samples based on 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 grey bars represent taxonomical data, light grey bars refer to functional traits. The error bars show standard errors of six replicates.

There were marked differences in the network analysis of the soil and rhizosphere samples for all three dilutions (Fig. 4.3A and B). In general, the number of correlations in the rhizosphere was larger than in the soil (Table 4.3), and the number of positive correlations was higher than negative ones for both soil and rhizosphere samples. Between-ness Centrality (BC) of the rhizosphere community networks was much stronger than that of the soil communities, decreasing gradually upon dilution (Fig. 4.3C). In the 10^{-9} diluted samples of the rhizosphere communities, no potential keystone species were obtained (Fig. 4.3C).

Table 4.3. Number of network correlations as inferred by sparCC.

Treatment	Number of nodes	Total number of significant correlations	Number of significant positive correlations	Number of significant negative correlations
Soil 10^{-1}	52	54	40	14
Soil 10^{-6}	59	84	54	30
Soil 10^{-9}	36	38	28	10
Rhizosphere 10^{-1}	119	471	309	162
Rhizosphere 10^{-6}	63	100	59	41
Rhizosphere 10^{-9}	73	102	69	33

4.3.3. 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 Bray-Curtis similarity matrices (Fig. 4.2B, R= 0.08, $P = 0.046$). The PCoA plot of functional profiles of the different dilutions of rhizosphere samples showed a higher similarity than those of soil samples (Fig. 4.2D 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 (Fig. 4.2). 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 (Fig. 4.4C and D). To compare the similarity among replicate samples of the six groups, we calculated the mean values of Bray-Curtis similarity for both the

taxonomic and functional data. Replicates of functional data within each dilution were highly similar (light gray bars in Fig. 4.2G), whereas the taxonomic similarity decreased upon dilutions for both soil and rhizosphere samples (dark gray bars in Fig. 4.2G).

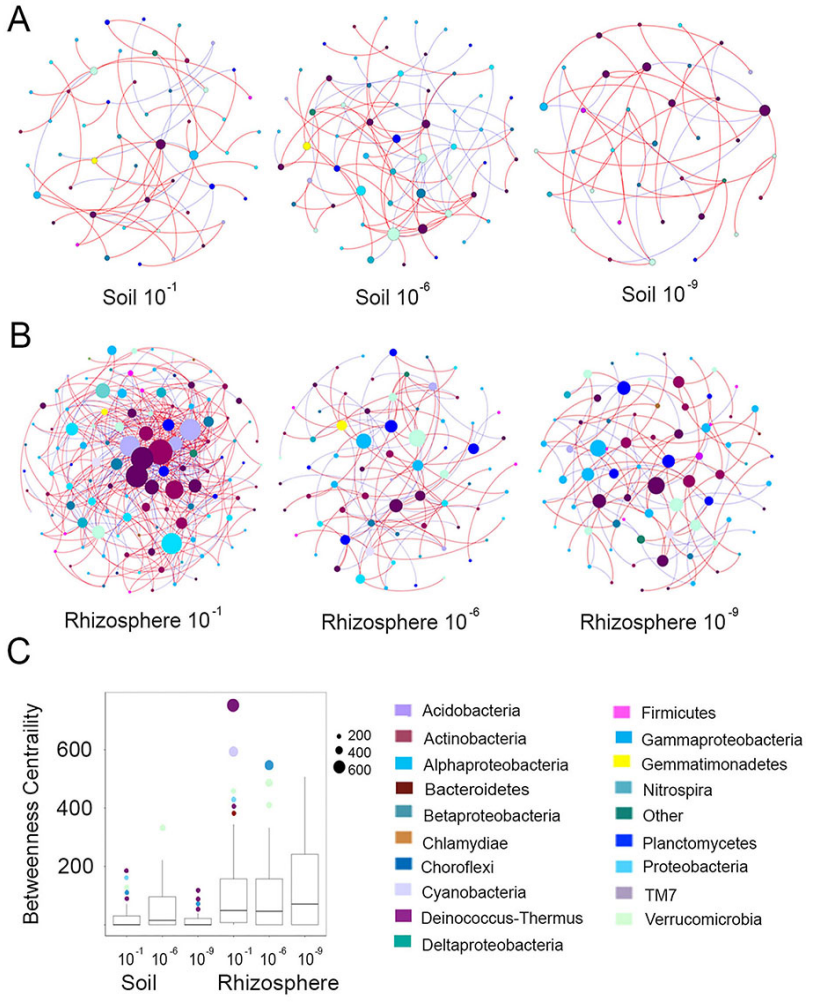

Figure 4.3. Co-occurrence patterns of bacteria in soil and rhizosphere. Correlations were presented in the soil samples (A) and in the rhizosphere samples of each dilution (B). Nodes indicate taxonomic affiliation at genus level. Red lines indicate positive correlations, and blue lines indicate negative correlations. The color of each node indicates the phylum shown below of the figures. The size of each node is proportional to the Betweenness Centrality (C). The box-and-whiskers graphics show the median of betweenness centrality 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 to 25th percentile) of the median. The size of outlier data points corresponds to the value of the Betweenness Centrality.

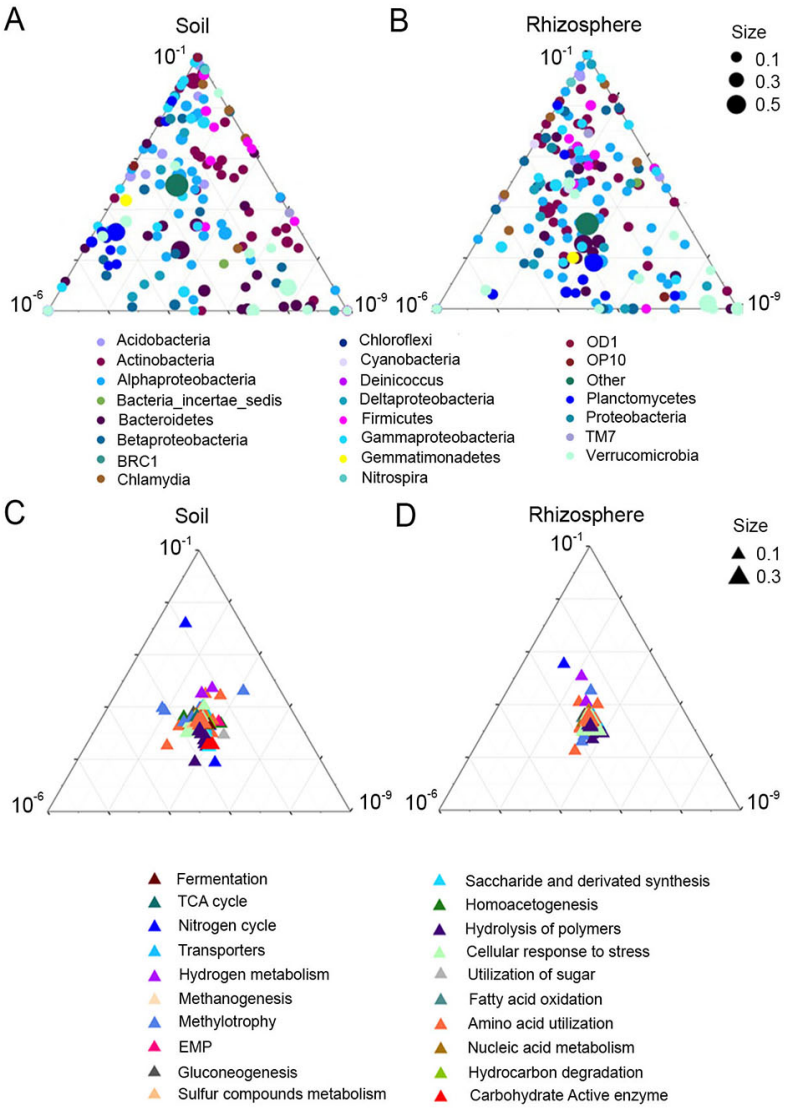

Figure 4.4. Distribution of species and functional traits in each dilution of the soil and rhizosphere samples. Ternary plots of OTUs associated with each dilution in soil (A) and rhizosphere (B) and of functional cores at FOAM 'level 1' associated with each dilution in soil (C) and rhizosphere (D). The position of each point is determined by the contribution of the indicated compartments to the total relative abundance. The size of the dots represents its relative abundance (weighted average). Colors indicate phyla (A and B) and functions at 'level 1' (C and D).

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 (Fig. 4.5). One of the most abundant types of genes, the transporter 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, e.g. 'transporters', 'EMP pathway' and 'hydrogen metabolism', we found that rhizosphere communities were clustered and significantly ($P < 0.05$) separated from soil communities (Fig. 4.5B). However, when testing functions that were more abundant in the soil than in the rhizosphere, e.g. cellular response to stress and carbohydrate active enzymes, we observed that soil and rhizosphere communities were not significantly separated (Fig. 4.5B). 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 for, rather than against species with particular functional traits.

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 (Fig. 4.5C). In contrast, *Caulobacteraceae*, unclassified *Bacteroidetes,* and, surprisingly, *Pseudomonadaceae* were over-represented in the soil.

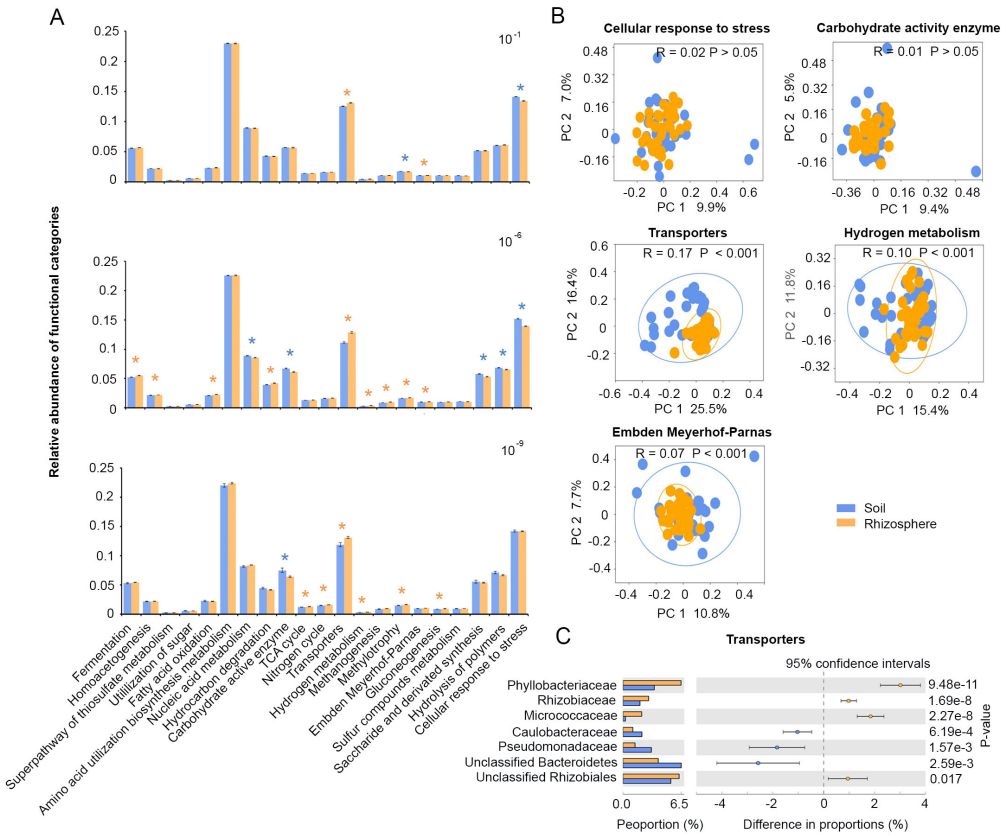

Figure 4.5. Profiles of soil and rhizosphere bacterial functional traits. (A) The relative abundance of groups of functional genes in soil and rhizosphere for three dilutions. Relative abundance of functional genes (FOAM 'level 1') based on normalized shotgun metagenomics data of dilutions of 10^{-1}, 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 more abundant in rhizosphere samples ($P < 0.05$) and blue asterisks (*) indicate categories that are more abundant in soil samples ($P < 0.05$). (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, Embden Meyerhof-Parnas 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 transporter genes between soil and rhizosphere samples (Welch's t-test; $P < 0.05$).

4.4. Discussion

There was a clear separation between soil and rhizosphere samples on the basis of species composition (Fig. 4.2A). 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 number of species detected in the rhizosphere was, however, 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 suspensions and 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. 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 (Fig. 4.2).

Our results, regarding both species composition and functional traits, clearly indicated 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, transporter systems were found to be of great importance in the rhizosphere. This was not reported on EMP pathway and hydrogen metabolism.

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 (Fig. 4.5B). 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*, in particular the genus *Arthrobacter*. These species have been reported earlier as beneficial to plants (Sanguin et al 2009, Hayat et al 2010). Remarkably, 'transports' 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 transport 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 4.2.

The network analysis revealed many more correlations and potential keystone species in the rhizosphere than in the soil (Fig. 4.3). 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 rRNA amplicon data and not on the binned shotgun data because of the above mentioned low annotation rate of the sequences.

In conclusion, we have shown here that the rhizosphere exerts selection on the microbial community also based on particular functional traits. However, 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. We found that the relative abundance of some particular functional genes in the rhizosphere was generally higher than in soil, suggesting that the rhizosphere selects for these functional traits rather than against them. The case in which the relative abundance was clearly higher in soil than in the rhizosphere was for functional traits related to cellular response to stress. This may indicate that the environment in the rhizosphere is less stressful for the bacterial community. On the other hand, the relative abundance of functional transporter genes was significantly higher in the rhizosphere than in soil, clearly showing that selective processes operated on these genes in the rhizosphere. 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.

4.5. 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. *Journal of Computational Biology* **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 in Plant Science* **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 Biology & Biochemistry* **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. *Proceedings of the National Academy of Sciences of the United States of America* **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 Meth* **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 Research* **37:** D141-D145.

Dixon P (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **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:** W30-38.

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. *Applied and Environmental Microbiology* **67:** 702-712.

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

Friedman J, Alm EJ (2012). Inferring Correlation Networks from Genomic Survey Data. *Plos Computational Biology* **8**.

Garland JL, Lehman RM (1999). Dilution/extinction of community phenotypic characters to estimate relative structural diversity in mixed communities. *Fems Microbiology Ecology* **30:** 333-343.

Good IJ (1953). The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika* **40:** 237-264.

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

Hayat R, Ali S, Amara U, Khalid R, Ahmed I (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Annals of Microbiology* **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. *Ecology Letters* **13:** 292-301.

Hughes JB, Hellmann JJ (2005). The application of rarefaction techniques to molecular inventories of microbial diversity. *Environmental Microbiology* **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**.

Imelfort M, Woodcroft, B., Parks, D. (2015). Ecogenomics/BamM [WWW Document]. GitHub. URL https://github.com/Ecogenomics/BamM (accessed 7.21.15).

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

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

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 Microbiology Ecology* **79:** 12-24.

Langenheder S, Szekely AJ (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISMEJ* **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. *Microbial Ecology* **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. *Ecology Letters* **10:** 995-1015.

McKinney W (2015). Python Data Analysis Library — pandas: Python Data Analysis Library [WWW Document]. URL http://pandas.pydata.org/ (accessed 7.22.15).

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

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.

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. *Nature Communications* **5**.

Philippot L, Spor A, Henault C, Bru D, Bizouard F, Jones CM *et al* (2013). Loss in microbial diversity affects nitrogen cycling in soil. *ISMEJ* **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. *Nature 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 Biology & Biochemistry* **13:** 1-&.

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 Phytologist* **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. *Applied and Environmental Microbiology* **75:** 7537-7541.

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

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. *Nature Biotechnology* **28:** 511-U174.

Verginer M, Siegmund B, Cardinale M, Muller H, Choi Y, Miguez CB *et al* (2010). Monitoring the plant epiphyte Methylobacterium extorquens DSM 21961 by real-time PCR and its influence on the strawberry flavor. *Fems Microbiology Ecology* **74:** 136-145.

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**.

Wertz S, Degrange V, Prosser JI, Poly F, Commeaux C, Freitag T et al (2006). Maintenance of soil functioning following erosion of microbial diversity. *Environ Microbiol* 8: 2162-2169.

Worm B, Duffy JE (2003). Biodiversity, productivity and stability in real food webs. *Trends in Ecology & Evolution* **18:** 628-632.