SUPPLEMENTAL MATERIAL

Supplementary Discussion.

Soybean rhizosphere 1-year vs. 5-year

The phylum *Actinobacteria* was more abundant in the 1-year soybean rhizosphere. This group is well known for antimicrobial secondary metabolites production, playing an important role in the decomposition of organic materials, as organic matter turnover and carbon cycle (Ventura *et al.*, 2007). Studies have shown a higher abundance of *Actinobacteria* in desert soils (Fierer *et al.*, 2012) and deforested areas (Navarrete *et al.*, unpublished), while *Verrucomicrobia* decreased. In our experiments, areas of native forest were deforested through slash-and-burn and then prepared to agriculture. As a consequence of deforestation, a high amount of carbon from the ashes is deposited in the soil. Also, there is an increase of soil temperature. This event could explain the higher abundance of *Actinobacteria* in the 1-year samples, since they are important decomposers; also, when the soil dries out, they are capable to produce spores, which allow them to resist to perturbations events (Ventura *et al.*, 2007). Interestingly, in the 5-year rhizosphere samples the phylum *Verrucomicrobia* increased, indicating a recovery after the deforestation.

The phylum *Proteobacteria* was significantly abundant in 5-year soybean rhizosphere. Specific groups of *Proteobacteria* contain plant-growth-promoting members, as *Gammaproteobacteria*, which was abundant in rhizosphere samples, and respond chemotactically to exudates and are efficient in the utilization of plants exudate products (García-Salamanca *et al.*, 2012). The class *Alphaproteobacteria* presented a significant enrichment in the 5-year soybean rhizosphere. Within this class, groups related to the nitrogen cycle were abundant in the rhizosphere. Genera related to nitrogen fixing found were *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*, *Azorhizobium, and Mesorhizobium*. Therefore, the long-term land use had a directed effect towards enrichment of specific bacterial groups.

In relation to functional traits, in the 5-year rhizosphere samples the “protein degradation, processing and modification” were abundant, which indicates intense bacterial activity in the rhizosphere. Also, within “regulation and cell signaling”, the “cAMP signaling in bacteria” was abundant in 5-year samples. The cAMP signaling system regulates important metabolic pathways through signaling cascade used in cell communication (Agarwal and Bishai, 2009), which is part of a complex system that drives basic cellular activities and coordinates cell actions. The high abundance of sequences related to cAMP signaling in 5-year rhizosphere may indicate a mechanism in which some species perceive and correctly respond to the microenvironment of the rhizosphere.

References

Agarwal N, Bishai WR. (2009). cAMP signaling in Mycobacterium tuberculosis. *Indian J Exp Biol* 47: 393-400.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional atributes. *PNAS* **109**: 21390-21395.

García-Salamanca A, Molina-Henares MA, van Dillewijn P, Solano J, Pizarro-Tobías P, Roca A *et al.*, (2012). Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microb Biotech* **6:** 36-44.

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF *et al.*, (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71:** 495-548.

Supplementary Table 1. Soil chemical factors of sampling sites.

Soil Factors	1-year crop	5-year crop
pH	4.1 ± 0.3	5.0 ± 0.2
OM	25.8 ± 3.8	35.2 ± 2.4
Al	9.8 ± 3.6	2.0 ± 0.7
H+Al	42.6 ± 13.7	50.4 ± 8.6
m	52.4 ± 14.5	8.0 ± 4.41
P	11.2 ± 18.8	33.6 ± 20.9
K	0.8 ± 0.3	1.7 ± 0.7
Ca	5.2 ± 1.8	13.6 ± 3.0
Mg	2.8 ± 3.6	9.0 ± 2.1
SB	8.8 ± 3.1	24.3 ± 5.4
CEC	51.4 ± 13.9	74.7 ± 12.0
V	18.2 ± 7.7	32.4 ± 4.6
B	0.2 ± 0.03	0.1 ± 0.04
Cu	0.3 ± 0	0.3 ± 0.06
Fe	93.2 ± 49.5	47.6 ± 6.2
Mn	1.6 ± 0.5	0.5 ± 0.2
Zn	0.9 ± 1.3	2.6 ± 2.1
N	1520.0±117.7	1434.6±99.7
$NH4^+$	16.6 ± 1.69	15.0 ± 1.63
NO^-3	13.3 ± 3.39	7.33 ± 0.94
C	15.0 ± 1.63	16.0 ± 0.81
Sand	50.3±1.88	39.6±1.24
Silt	1.66±0.94	1.33±0.47
Clay	48.0±0.0	59.0±1.41

The values are averages based on quintuplicate sampling points in each site. Standard deviations are shown in the table.

Ca, Mg, K, Al, potential acidity (H+Al), sum of bases (SB) are expressed in $nmol.kg^{-1}$; OM and C is expressed in $g.Kg^{-1}$; P is expressed in $mg.kg^{-1}$; B, Fe, Mn, Zn, K, Cu, and cation exchange capacity in pH 7 (CEC) are expressed in $mg.dm^{-3}$; Sand, silt and clay are expressed in %; N (total), NH4+ and NO-3 are expressed in $mg.kg^{-1}$. P and K – Mehlich 1 extractor. Ca, Mg, and Al – Kcl 1N. H+Al – SMP extractor. N – Kjeldahl. NH4+ and NO-3 – Raney/Kjeldahl. OM – organic matter. m – Al saturation index. V – base saturation index.

Supplementary Table 2. Number of sequencing reads, base pairs, reads assigned to SEED Subsystems and percentages of predict proteins after quality control on MG-RAST pipeline.

Sample ID	MG-RAST ID	Sample	N. of sequence reads	Mean sequence length	Total bp*	N. of predict Subsystems functions*	% of predicted proteins*
1LM	4477751.3	Riz_5a_1	141,552	523±46	61,451,072	98,237	74.7
2LM	4477749.3	Riz_5a_2	151,054	522±42	63,711,128	102,880	75.4
3LM	4477755.3	Riz_5a_3	140,546	520±45	58,624,166	94,250	75.1
4LM	4477757.3	Bulk_5a_1	144,908	520±46	62,571,569	94,952	71.0
5LM	4477789.3	Bulk_5a_2	137,109	523±42	58,459,177	87,173	69.9
6LM	4477790.3	Bulk_5a_3	144,587	521±41	57,485,200	87,555	71.4
7LM	4478030.3	Riz_1a_1	145,214	518±68	58,031,403	94,180	76.4
8LM	4478037.3	Riz_1a_2	124,076	527±56	51,095,519	80,186	73.8
9LM	4478934.3	Riz_1a_3	113,953	520±36	9,678,615	9,788	75.9
10LM	4478038.3	Bulk_1a_1	140,141	526±46	58,816,390	96,369	76.8
11LM	4478222.3	Bulk_1a_2	142,020	526±49	50,166,332	77,447	72.7
12LM	4478283.3	Bulk_1a_3	140,261	527±56	60,814,070	98,142	75.8
13LM	4478290.3	Riz_5b_1	147,028	525±45	65,025,641	108,156	77.4
14LM	4478292.3	Riz_5b_2	129,829	526±45	55,750,716	88,769	74.5
15LM	4478291.3	Riz_5b_3	112,056	527±50	36,967,527	56,999	72.9
16LM	4478294.3	Bulk_5b_1	138,556	525±47	55,060,604	83,310	71.2
17LM	4478936.3	Bulk_5b_2	137,939	523±41	58,658,766	92,450	73.6
18LM	4478937.3	Bulk_5b_3	146,385	524±43	55,611,272	85,403	71.9
19LM	4478938.3	Riz_1b_1	126,119	523±47	35,784,579	59,398	77.9
20LM	4478939.3	Riz_1b_2	133,146	525±47	51,791,233	89,065	80.1
21LM	4479311.3	Riz_1b_3	151,804	524±40	68,089,631	116,414	79.5
22LM	4478940.3	Bulk_1b_1	138,073	523±42	55,607,597	92,786	77.8
23LM	4478941.3	Bulk_1b_2	143,858	525±45	55,511,047	89,673	75.6
24LM	4478943.3	Bulk_1b_3	123,137	526±50	51,204,159	83,520	76.5

* Post Quality Control

Supplementary Table 3. Relative proportion of matches to a given subsystem hierarchy level 1.

Subsystem hierarchy level 1	Rhizosphere		Subsystem hierarchy level 1	Bulk Soil
Carbohydrates	0.133		Clustering-based subsystems	0.143
Clustering-based subsystems	0.112		Carbohydrates	0.141
Cofactors, Vitamins, Prosthetic Groups, Pigments	0.101		Amino Acids and Derivatives	0.110
Miscellaneous	0.091		Miscellaneous	0.078
Membrane Transport	0.069		Cofactors, Vitamins, Prosthetic Groups, Pigments	0.066
Amino Acids and Derivatives	0.058		Protein Metabolism	0.065
Iron acquisition and metabolism	0.045		RNA Metabolism	0.049
Cell Wall and Capsule	0.044		Fatty Acids, Lipids, and Isoprenoids	0.046
RNA Metabolism	0.042		Cell Wall and Capsule	0.042
Protein Metabolism	0.030		DNA Metabolism	0.033
Respiration	0.030		Respiration	0.025
Metabolism of Aromatic Compounds	0.028		Virulence, Disease and Defense	0.024
Stress Response	0.027		Membrane Transport	0.024
Fatty Acids, Lipids, and Isoprenoids	0.027		Nucleosides and Nucleotides	0.023
Virulence, Disease and Defense	0.026		Stress Response	0.023
Nucleosides and Nucleotides	0.023		Metabolism of Aromatic Compounds	0.021
Regulation and Cell signaling	0.020		Regulation and Cell signaling	0.013
Phosphorus Metabolism	0.018		Sulfur Metabolism	0.012
DNA Metabolism	0.017		Cell Division and Cell Cycle	0.011
Phages, Prophages, Transposable elements, Plasmids	0.015		Motility and Chemotaxis	0.010
Motility and Chemotaxis	0.012		Phosphorus Metabolism	0.010
Nitrogen Metabolism	0.010		Phages, Prophages, Transposable elements, Plasmids	0.010
Sulfur Metabolism	0.006		Nitrogen Metabolism	0.007
Cell Division and Cell Cycle	0.005		Iron acquisition and metabolism	0.004
Potassium metabolism	0.005		Potassium metabolism	0.003
Dormancy and Sporulation	0.004		Secondary Metabolism	0.003
Secondary Metabolism	0.003		Dormancy and Sporulation	0.001
Photosynthesis	0.000		Photosynthesis	0.001

Hits were generated by blasting sequences of 24 metagenomes (12 rhizosphere and 12 bulk soil) to the MG-RAST subsystem database with a minimum alignment length of 50 bp and an E value cut-off of $1x10^{-5}$. Relative representation in the metagenomes was calculated by dividing the number of hit to each category by the total number of hits to all categories.

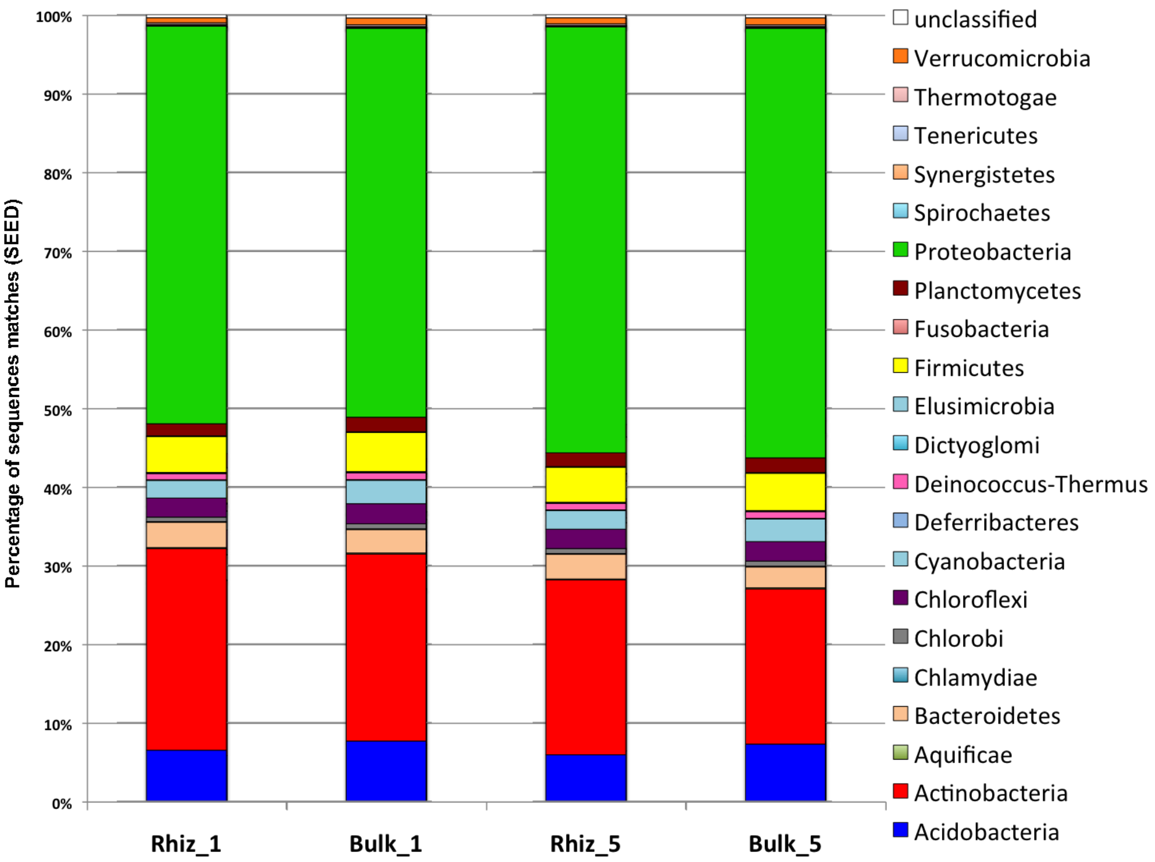
Supplementary Table 4. AIC values for 6 rank abundance distribution models. Lowest AIC value for each sample represents the best fit model.

				AIC[1]					
Sample	**Environment**	**Year**	**Exp.**	**Broken-stick**	**Pre-Emption**	**Log-Normal**	**Zipf**	**Zipf-Mandelbrot**	**ZSM**
10LM	bulk soil	1-year	A	16566.6	14433.6	3544.1	10722.1	*nd*	**3011.52***
11LM	bulk soil	1-year	A	15453.4	13136.4	3608.7	9253.8	*nd*	**3165.38***
12LM	bulk soil	1-year	A	18072.4	17130.7	3353.5	9213.3	4932.9	**3187.30***
22LM	bulk soil	1-year	B	12397.6	11266.7	3030.8	10558.3	*nd*	**2901.26***
23LM	bulk soil	1-year	B	12033.53	11631.04	2848.88	9408.63	4589.85	**2830.94***
24LM	bulk soil	1-year	B	9897.69	9333.39	**2878.37***	9506.42	4312.78	2905.78
4LM	bulk soil	5-years	A	23629.23	20288.77	**2845.67***	8706.34	4902.39	2931.52
5LM	bulk soil	5-years	A	20612.1	18608.4	3168.4	8213.9	4571.0	**2904.84***
6LM	bulk soil	5-years	A	20844.6	19224.9	3034.6	7691.1	4558.6	**2860.14***
16LM	bulk soil	5-years	B	12388.80	12005.43	**2482.06***	8162.28	4123.71	2960.74
17LM	bulk soil	5-years	B	15095.98	14049.54	**2599.84***	9048.28	4728.68	2914.10
18LM	bulk soil	5-years	B	13229.33	12020.41	**2763.61***	9433.87	4816.68	2952.98
7LM	rhizosphere	1-year	A	17797.50	14802.7	3453.7	10613.0	*nd*	**2984.68***
8LM	rhizosphere	1-year	A	17306.3	13818.2	3920.6	10024.1	*nd*	**2994.92***
9LM	rhizosphere	1-year	A	*nd*	*nd*	*nd*	*nd*	*nd*	*nd*
19LM	rhizosphere	1-year	B	7931.75	7375.81	**2433.63***	7143.04	2418.04	2882.70
20LM	rhizosphere	1-year	B	12679.91	11604.07	**2742.40***	9591.34	4776.32	2961.08
21LM	rhizosphere	1-year	B	16487.3	14630.0	3438.7	12509.2	*nd*	**2971.08***
1LM	rhizosphere	5-years	A	17254.95	15507.25	**2955.34***	10061.52	*nd*	2966.68
2LM	rhizosphere	5-years	A	20012.0	17093.2	3487.8	3495.4	*nd*	**3042.70***
3LM	rhizosphere	5-years	A	17708.09	15857.41	**2721.70***	9287.75	4968.54	3014.40
13LM	rhizosphere	5-years	B	16146.35	14662.35	**2865.83***	11147.48	*nd*	2938.24
14LM	rhizosphere	5-years	B	15409.74	13701.41	**2753.47***	9417.29	5076.29	2848.16
15LM	rhizosphere	5-years	B	9383.1	8949.1	**2271.80***	6210.9	3528.3	2762.82

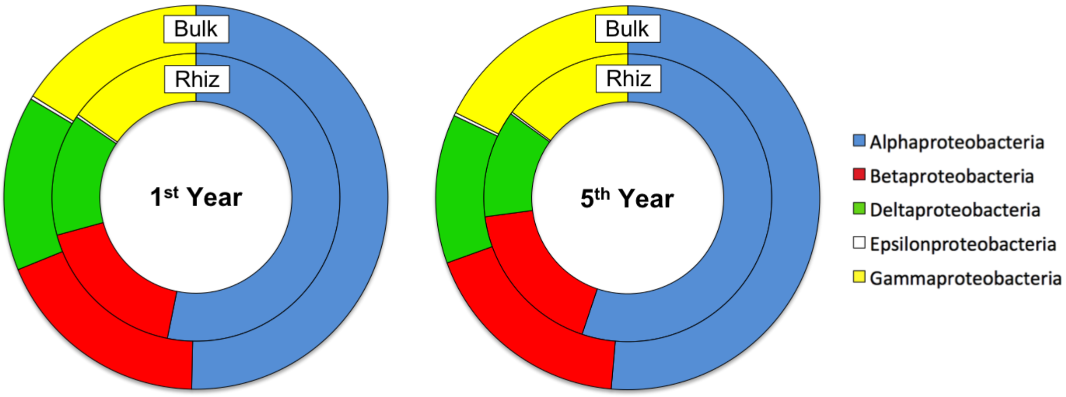
[1] AIC for radfit-generated models calculated the equation AIC = -2log-likehood + 2*npar. AIC was calculated in the same way for the zero-sum model, from the minimum of log-likehood reported by TeTame, then multiplied by -1 to obtain the maximum log-likehood value. * The best fit model with lowest AIC value. *nd* not determined

Supplementary Figure 1.

A

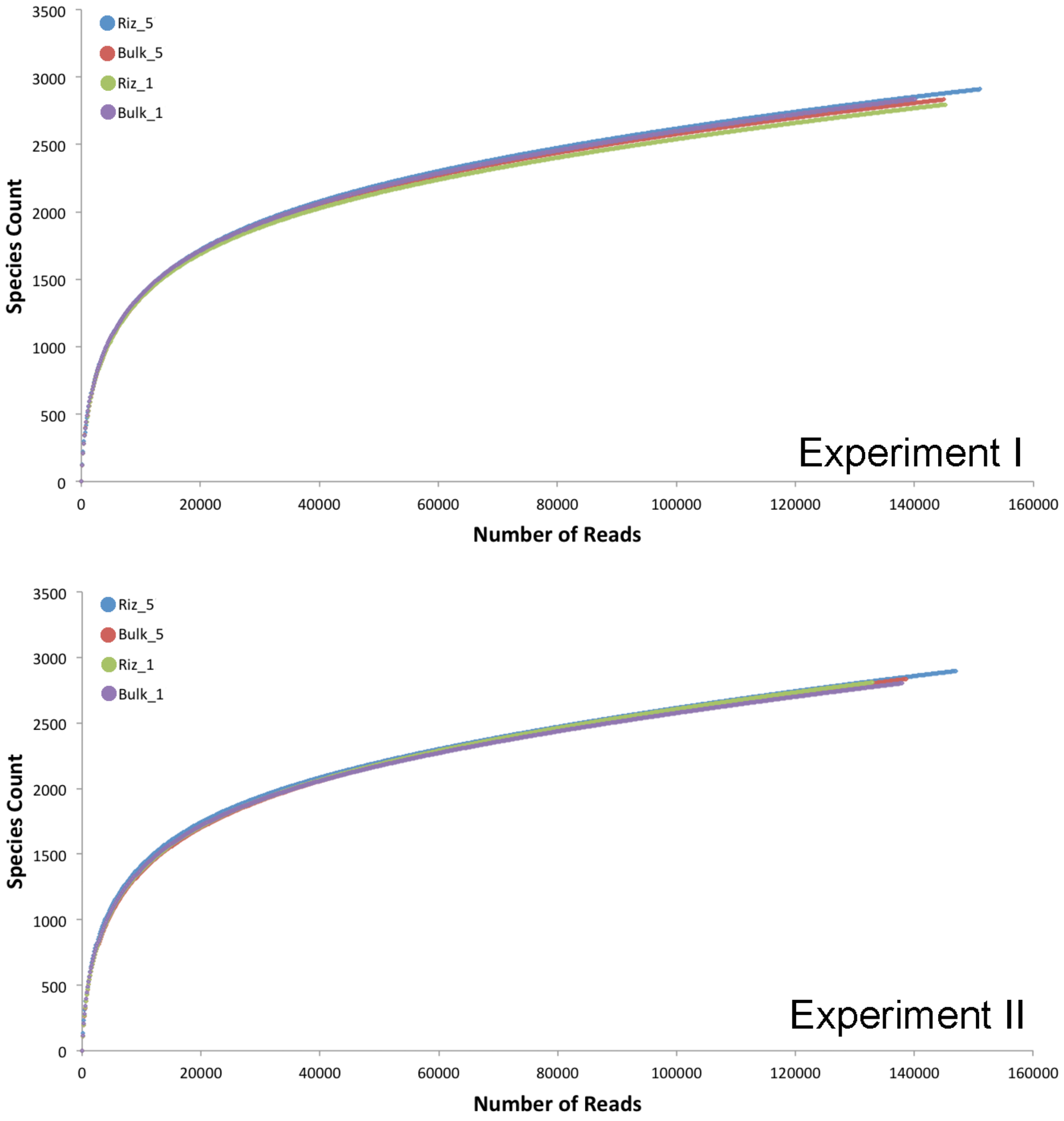


B



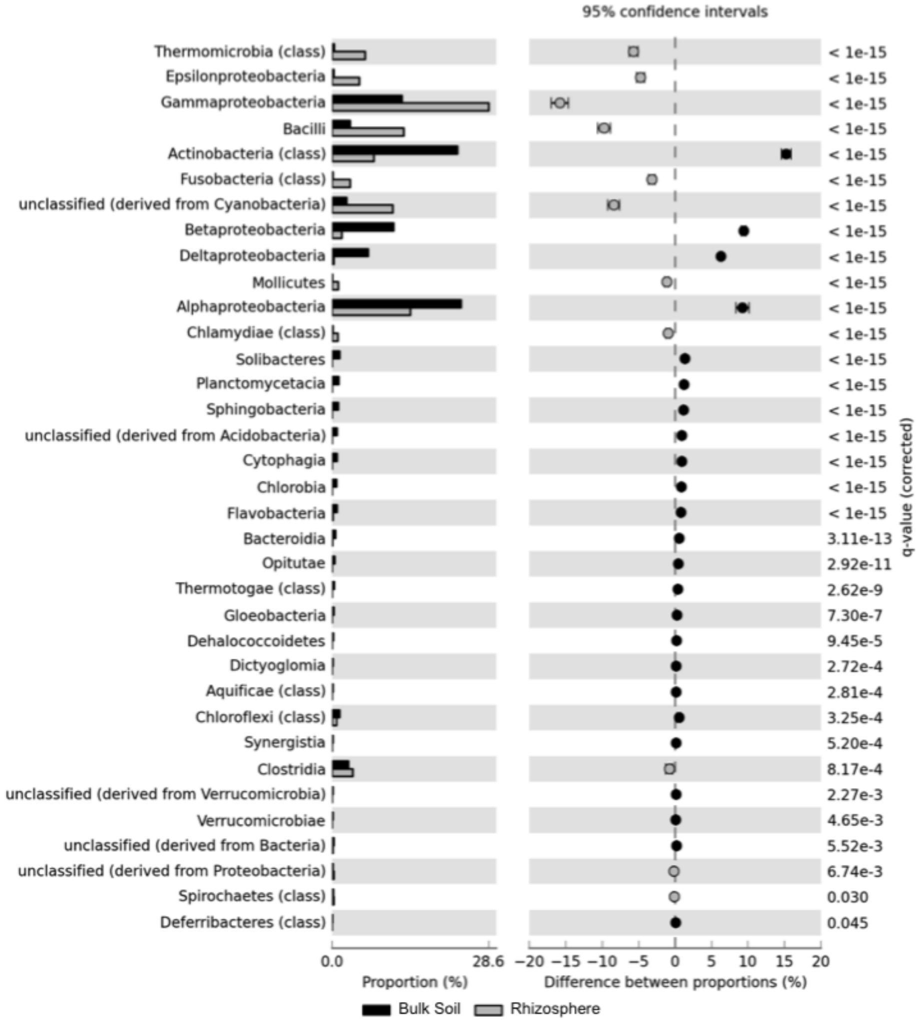
SF1. Taxonomic affiliation of metagenomic reads. (A) Results for complete datasets evaluated by BLASTX analysis against the SEED database using MG-RAST v 3.3 software. (B) Differential proportion of sequences assigned within the phylum *Proteobacteria* for rhizosphere and bulk soil of first and fifth years of crop.

Supplementary Figure 2.



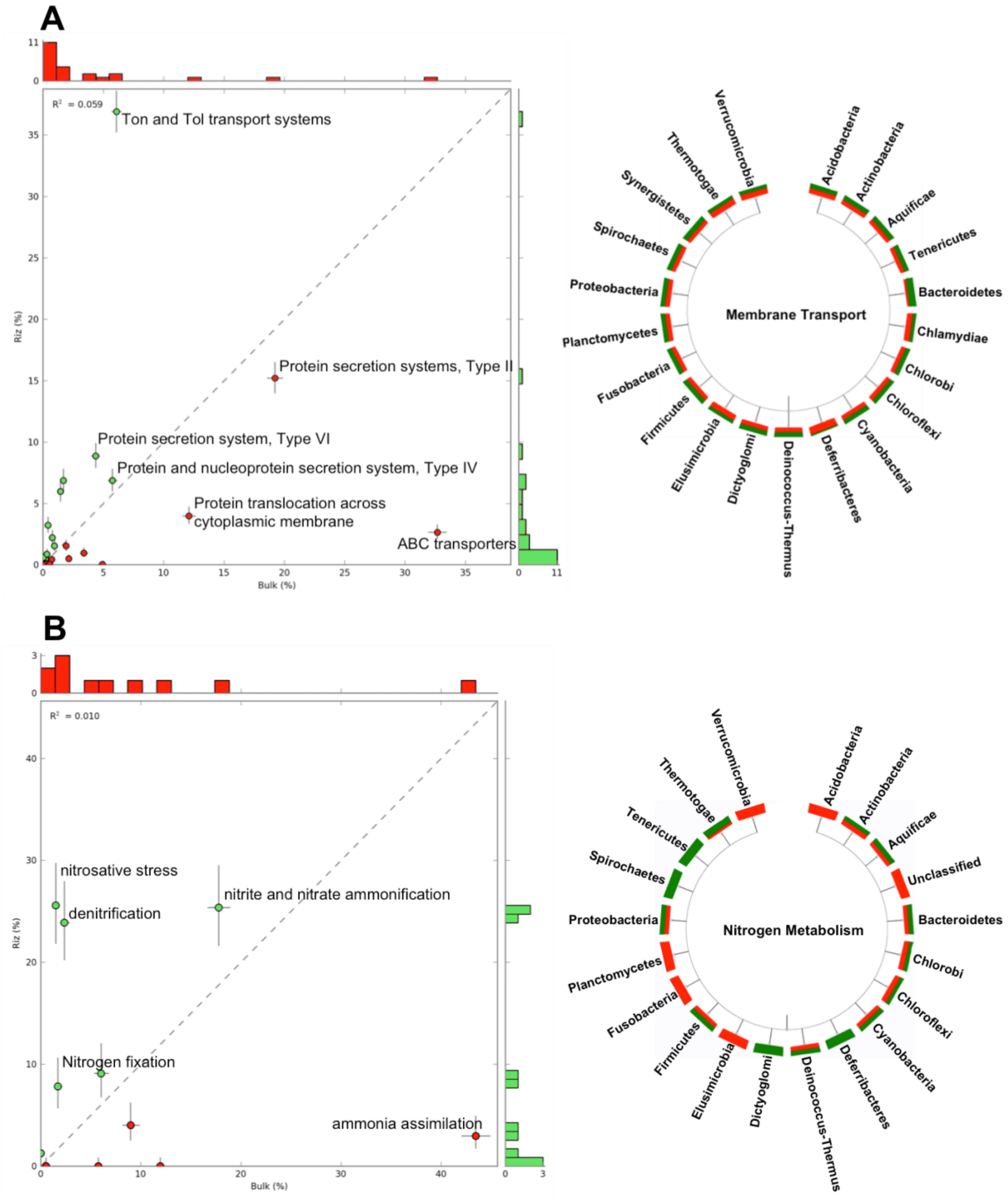
SF2. Rarefaction curves of annotated species richness generated for samples from bulk soil and rhizosphere of first and fifth of soybean cultivation. The curves represent the average number of different species annotations for subsamples of the complete dataset.

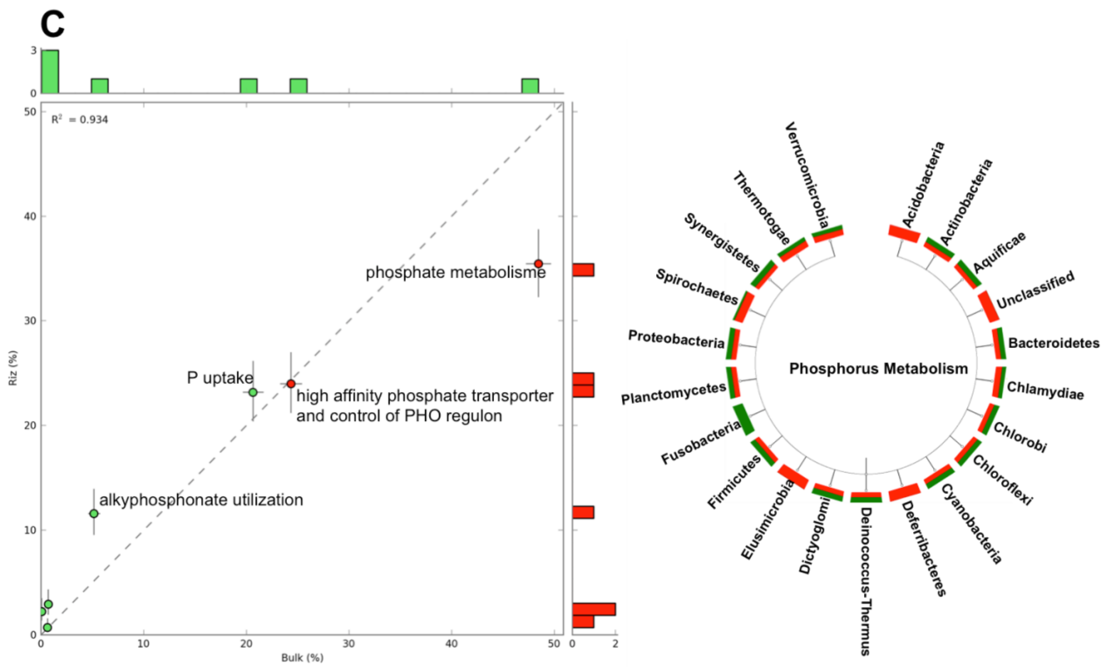
Supplementary Figure 3.

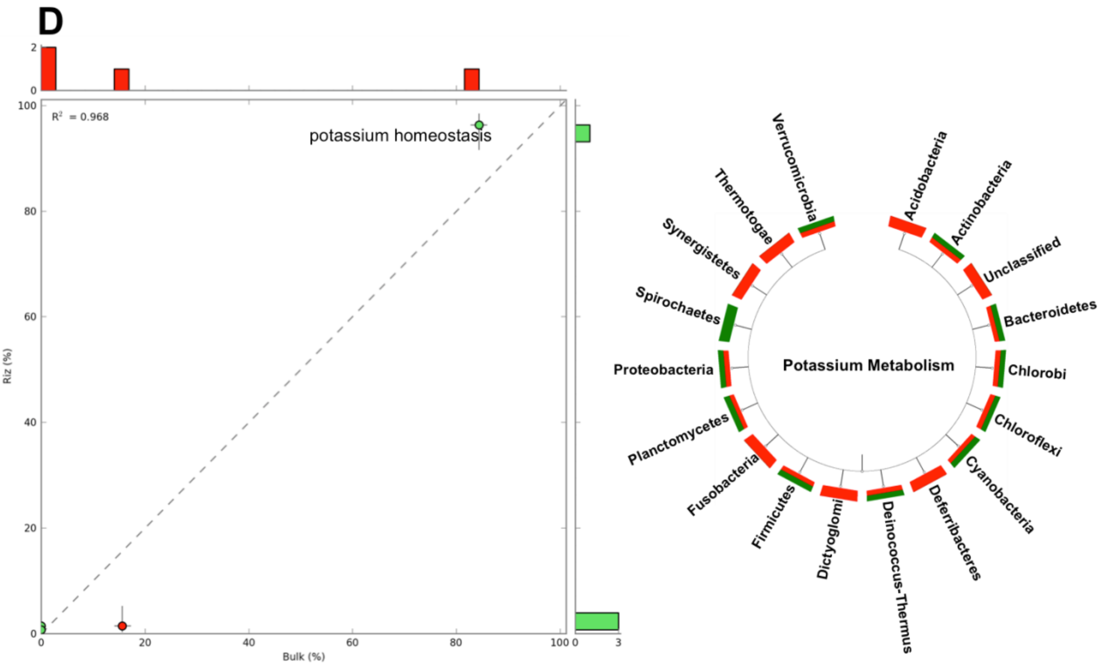


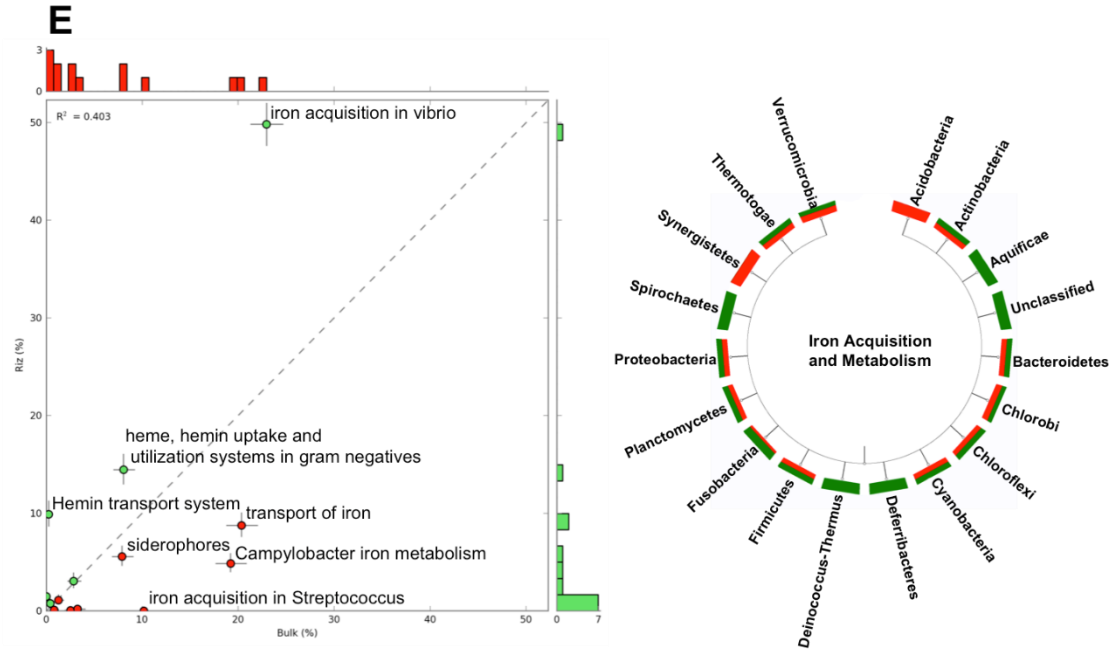
SF3. STAMP analysis of taxonomic profiles at class level between rhizosphere and bulk soil samples. Groups overrepresented in the bulk soil (black) correspond to positive differences between proportions and groups overrepresented rhizosphere (gray) correspond to negative differences between proportions. Corrected *P*-values (q-values) were calculated using Benjamini-Hochberg FDR approach (p<0.05).

Supplementary Figure 4.



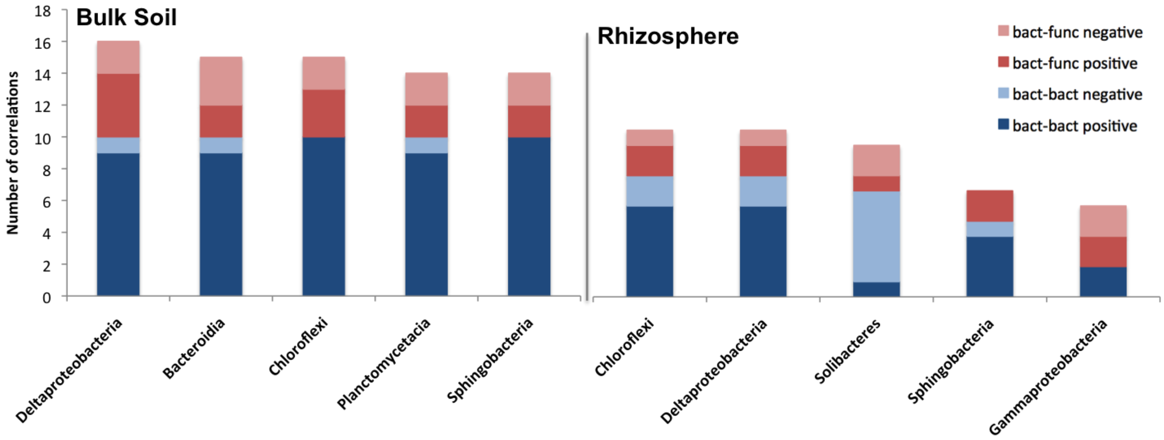

C
R² = 0.934
phosphate metabolisme
P uptake
high affinity phosphate transporter and control of PHO regulon
alkyphosphonate utilization
Riz (%)
Bulk (%)
Phosphorus Metabolism
Verrucomicrobia
Acidobacteria
Actinobacteria
Aquificae
Unclassified
Bacteroidetes
Chlamydiae
Chlorobi
Chloroflexi
Cyanobacteria
Deferribacteres
Deinococcus-Thermus
Dictyoglomi
Elusimicrobia
Firmicutes
Fusobacteria
Planctomycetes
Proteobacteria
Spirochaetes
Synergistetes
Thermotogae


D
R² = 0.968
potassium homeostasis
Riz (%)
Bulk (%)
Potassium Metabolism
Verrucomicrobia
Acidobacteria
Actinobacteria
Unclassified
Bacteroidetes
Chlorobi
Chloroflexi
Cyanobacteria
Deferribacteres
Deinococcus-Thermus
Dictyoglomi
Firmicutes
Fusobacteria
Planctomycetes
Proteobacteria
Spirochaetes
Synergistetes
Thermotogae



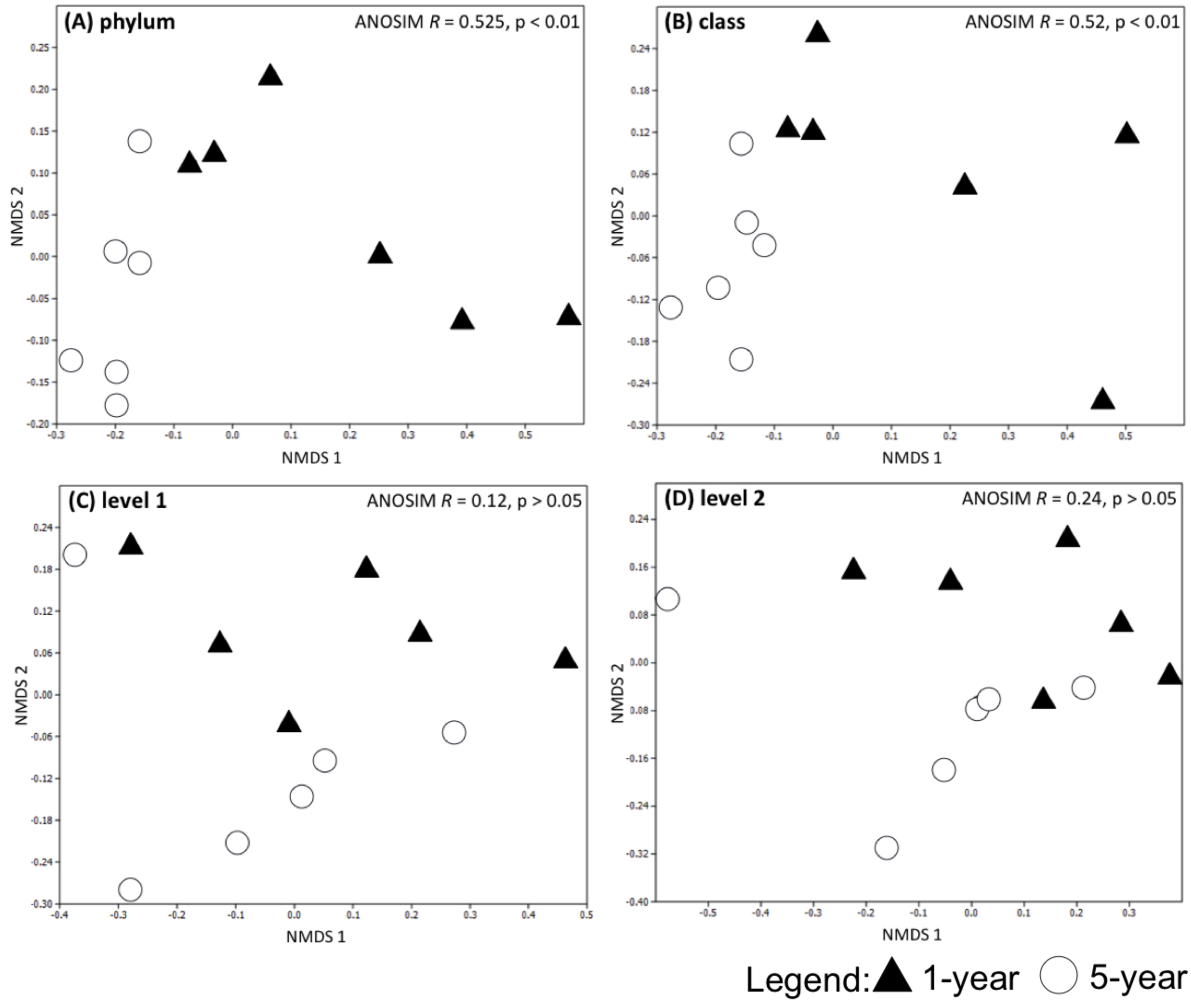
SF4. Functional and taxonomic profiles of bulk soil samples (red) and rhizosphere (green) of (A) membrane transport, (B) nitrogen metabolism, (C) phosphorus metabolism, (D) potassium metabolism and (E) iron acquisition and metabolism. The data was calculated for metagenomes from bulk soil and rhizosphere and compared to SEED database using a maximum *E*-value of $1e^{-5}$ and a minimum alignment length of 50 bp.

Supplementary Figure 5.



SF5. Top five bacterial groups with higher positive and negative correlations in bulk soil and rhizosphere samples. Numbers of correlations (positives and negatives) between bacteria and bacteria (bact-bact), bacteria and function (bact-func), and function and function (func-func) in bulk soil and in rhizosphere samples.

Supplementary Figure 6.



SF6. NMDS of Bray-Curtis similarity matrix performed on taxonomic and functional profiles between 1-year and 5-year soybean rhizosphere samples. (A) Taxonomic plot at phylum level. (B) Taxonomic plot at class level. (C) Functional Plot at level 1. (D) Functional plot at level 2. Similarity values (ANOSIM) are shown in the upper right of each plot.