**Evolutionarily conserved core rhizosphere microbiota promotes host performance and fitness in heavy metal accumulating plants**

Jipeng Luo1\*, Shaohua Gu2\*, Xinyu Guo1, Yuankun Liu1, Qi Tao4, Yongchao Liang1, Samiran Banerjee3#, Tingqiang Li1#

1 Ministry of Education Key Laboratory of Environmental Remediation and Ecological Health, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China.

2 Center for Quantitative Biology, Peking University, Beijing, China, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

3 Department of Microbiological Sciences,North Dakota State University, Fargo, North Dakota, USA

4 College of Resources, Sichuan Agricultural University, Chengdu, 611130, China

\* These authors contributed equally to this work.

# **Corresponding author**:

Tingqiang Li, Ph.D., E-mail: [litq@zju.edu.cn](mailto:litq@zju.edu.cn)

Samiran Banerjee, Ph.D., E-mail: banerjee.soilmicro@gmail.com

Number of figures: 14

Number of tables: 5

Number of pages: 23

**Supplementary Results:**

**Plant biomass and Cd accumulation**

The five Cd-accumulating plants grown in the three diverse soils showed a consistent difference in root and shoot biomasses. *Solanum nigrum* had the largest root and shoot biomasses, and the two ecotypes of *Noccaea caerulescens* had a comparable plant biomass that was significantly (*p* < 0.05) lower than Indian mustard and *Sedum alfredii* (Fig. S5). Across the soils, Cd accumulated in the root and shoot tissuesare largest for *S. alfredii*. I. mustard and *Noccaea caerulescens* (Prayon) had a comparable root Cd accumulation that was significantly (*p* < 0.05) lower than that in *S. nigrum* and *N. caerulescens* (Ganges). I. mustard and *N. caerulescens* (Prayon) showed a similar shoot Cd accumulation that was significantly lower than that in *S. nigrum* and *N. caerulescens* (Ganges) (Fig. S6). More than 40 μg Cd pot-1 and 450 μg Cd pot-1 were accumulated in the root and shoot tissues of different Cd-accumulating plants, respectively.

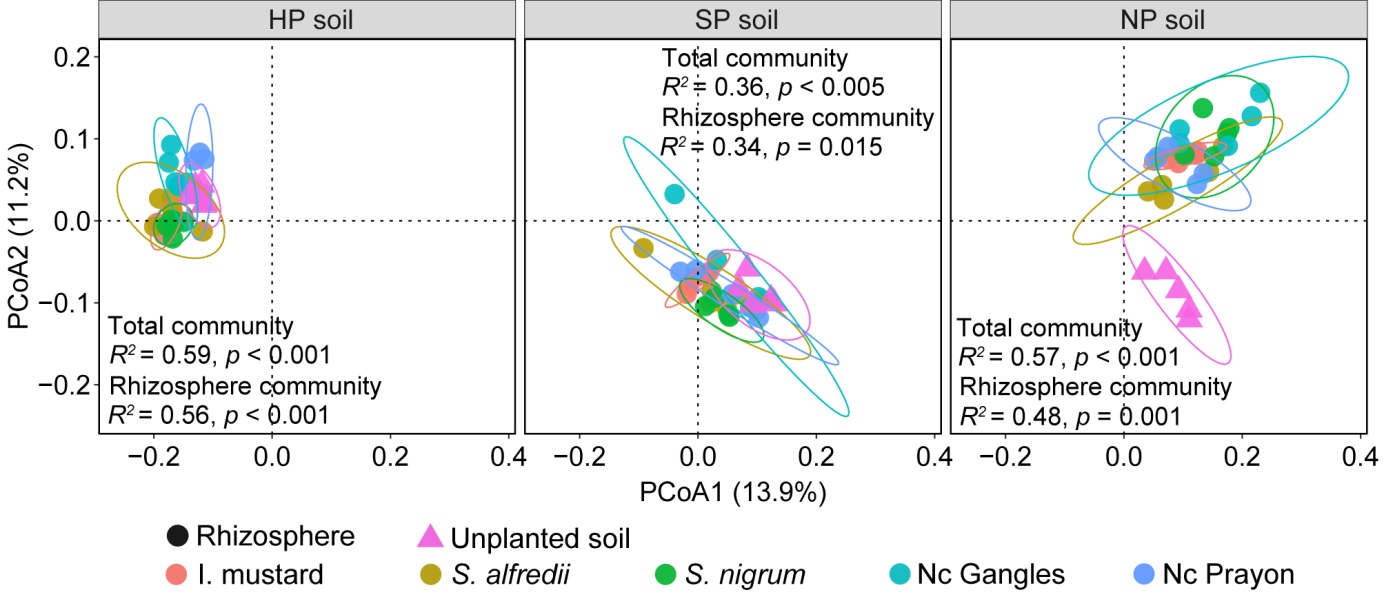
**Differentially abundant ASVs in soil across plant species and different soils**

As found in other studies on various plants1,2, Simpson’s D-1 rather than phylogenetic diversity (Faith’s pd) was strikingly higher in the unplanted soil than in the rhizosphere (Wilcoxon test, *p* < 0.05; Fig. S2c). This indicated that rhizocompartment is an important selective force that shapes the composition of root-associated microbiota.

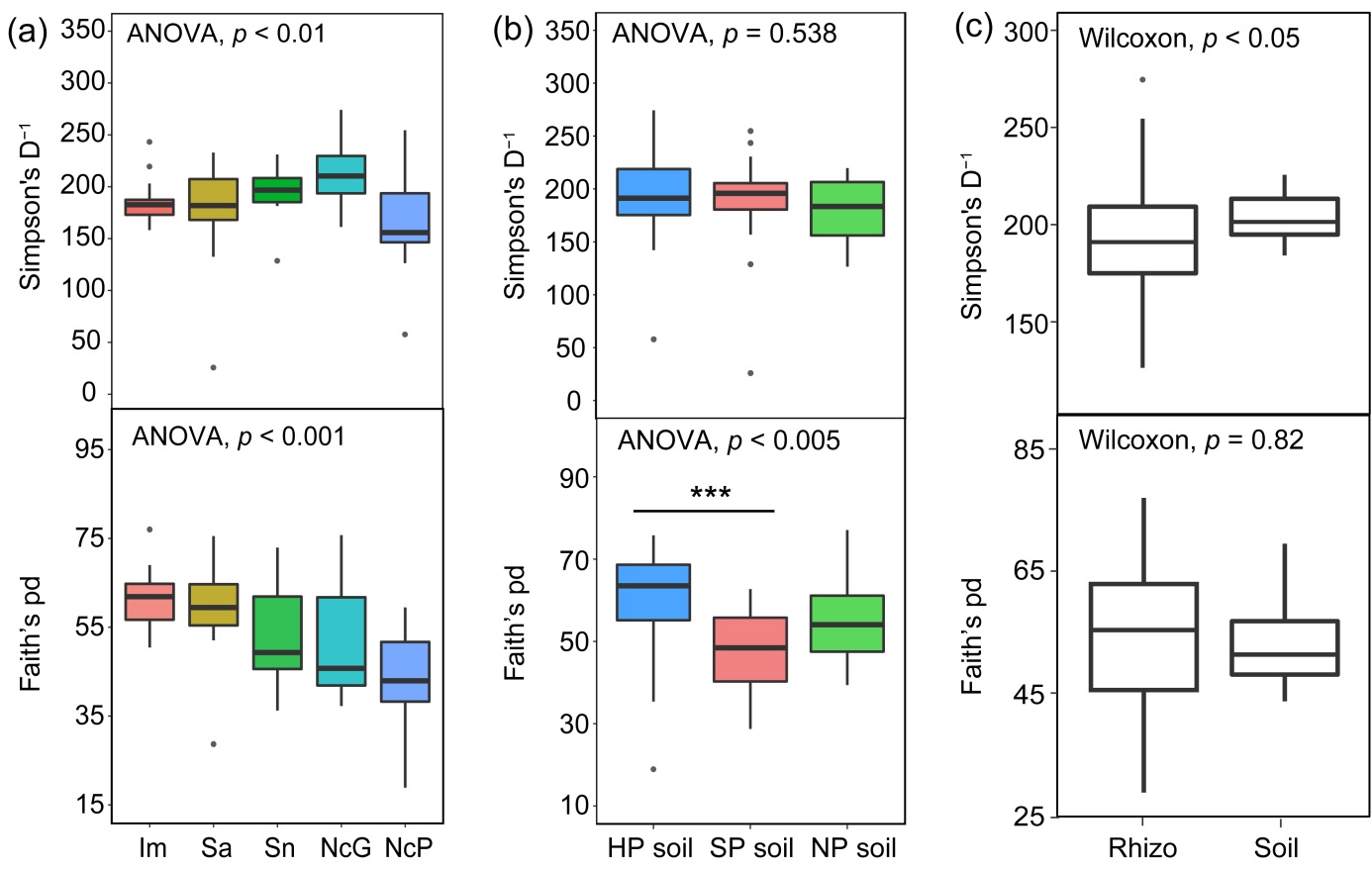
We then identified the community members that differed in relative abundance between rhizocompartment, and among plant species and soils using the DESeq2 R package 3. A multitude of ASVs (630) exhibited significant rhizosphere enrichment (defined as ‘RhizoASVs) and their relative abundances somewhat showed a soil-type-dependent distribution pattern (Fig. 3a). Rhizosphere bacterial community was represented by *Acidobacteriia*, α-proteobacteria, *Bacteroidia*, γ-proteobacteria (Fig. S3b). Specifically, three actinobacterial families, *Micrococcaceae*, *Streptomycetaceae* and *Mycobacteriaceae*, five alphaproteobacterial families, *Devosiaceae*, *Micropepsaceae*, *Rhizobiaceae*, *Sphingomonadaceae* and *Xanthobacteraceae*, four gammaproteobacterial families, *Burkholderiaceae*, *Nitrosomonadaceae*, *Rhodanobacteraceae* and SC-I-84 dominated the rhizosphere assemblages (Fig. S3c-e). Nearly 25% of bacterial phyla and 0.5% of bacterial ASVs were strongly affected by plant interspecies variation, and 50% of bacterial phyla and 0.9% of bacterial ASVs were significantly influenced by variation in soil’s geographic origin and 21% of bacterial phyla and 0.6% of bacterial ASVs were strongly affected by soil pollution level (Fig. S3b-d). Six bacterial phyla, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Proteobacteria* and *Verrucomicrobia*, were strongly affected by variation in plant interspecies (Fig. 1e). Eight bacterial phyla, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes*, *Proteobacteria* and *Verrucomicrobia*, were pronouncedly influenced by soil variation (Fig. 1e). These results demonstrated that soil type, plant species and rhizocompartment can affect the root-associated bacterial communities at different taxonomic ranks.

**Supplementary Figures**

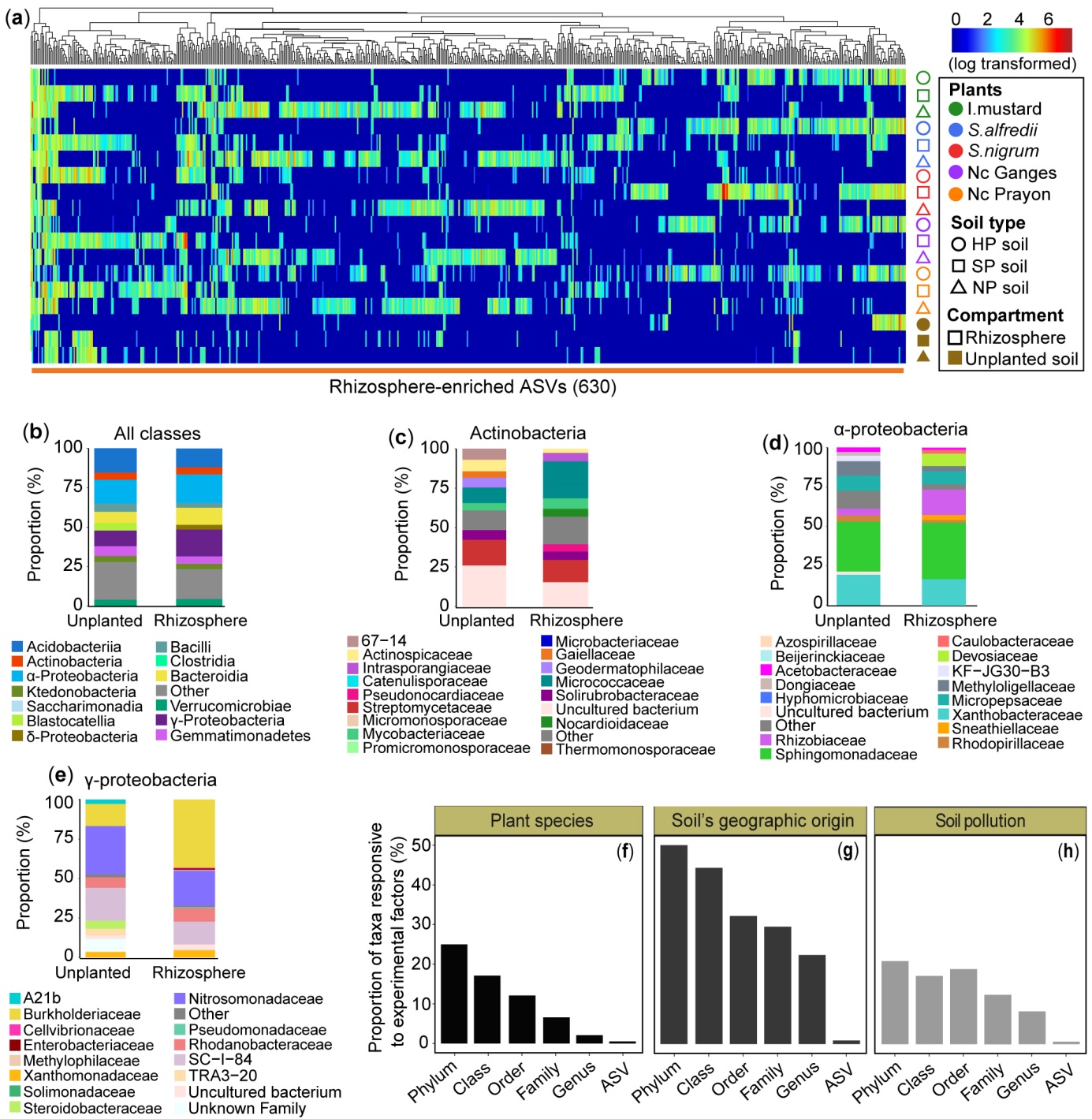
**Figure S1** The influence of plant species variation on the composition of total and rhizosphere bacterial communities across diverse soils. The R squared and *p*-value represent the strength and significance of plant species in community variation, as evaluated by PERMANOVA. Ellipses show the parametric smallest area around the mean that contains 90% of the probability mass for each plant species.

****

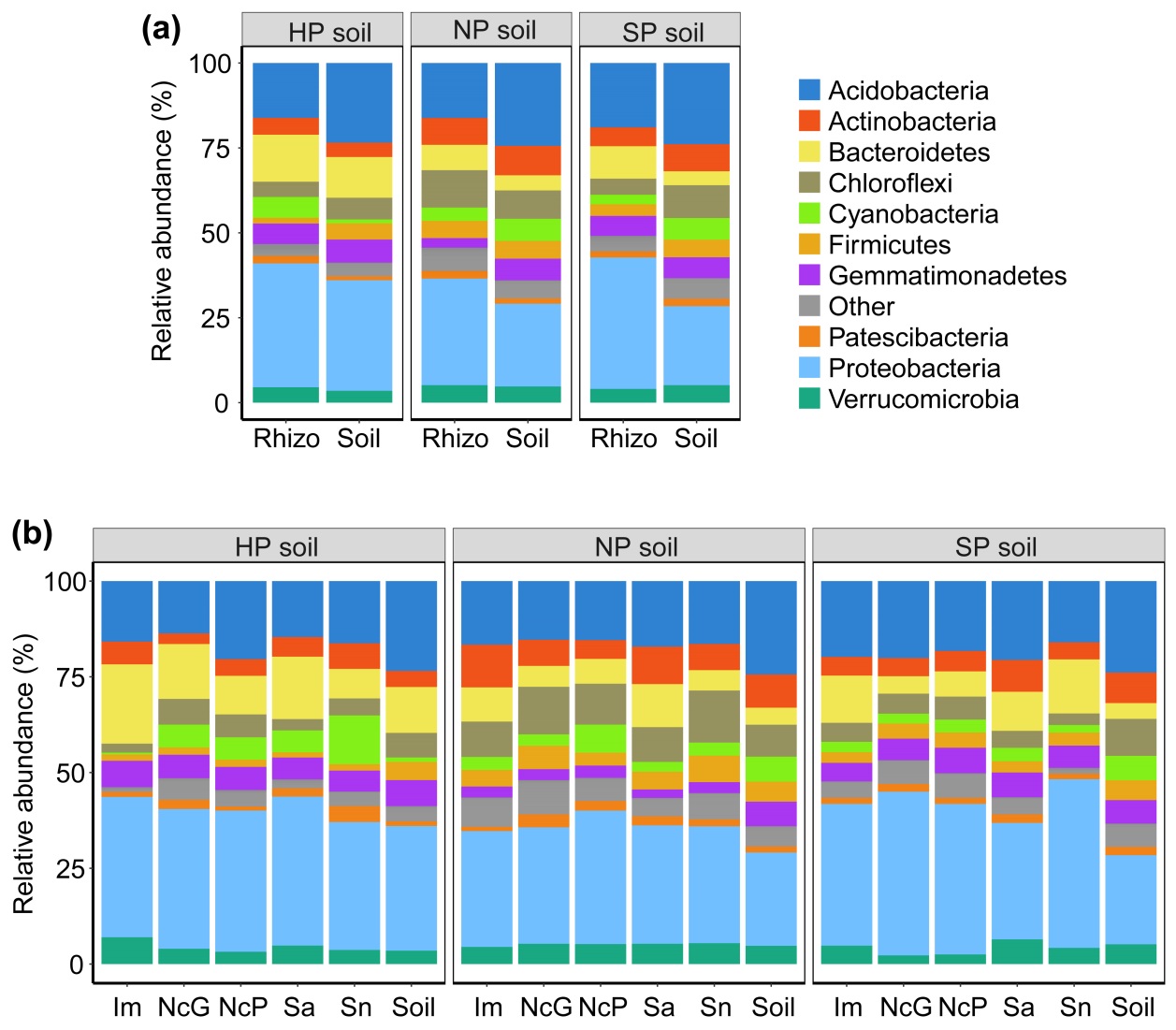
**Figure S2** Alpha diversity of the rhizosphere bacterial communities among plant species (a) and the three soils (b), and soil bacterial communities between rhizosphere and unplanted soil (c). Statistical comparisons of the pairs among soils and among plant species were performed by ANOVA, and between HP and SP soils and between rhizosphere and unplanted soil by Wilcoxon test, and the statistical results are indicated with asterisks, \*\*\* *p* < 0.001.

****

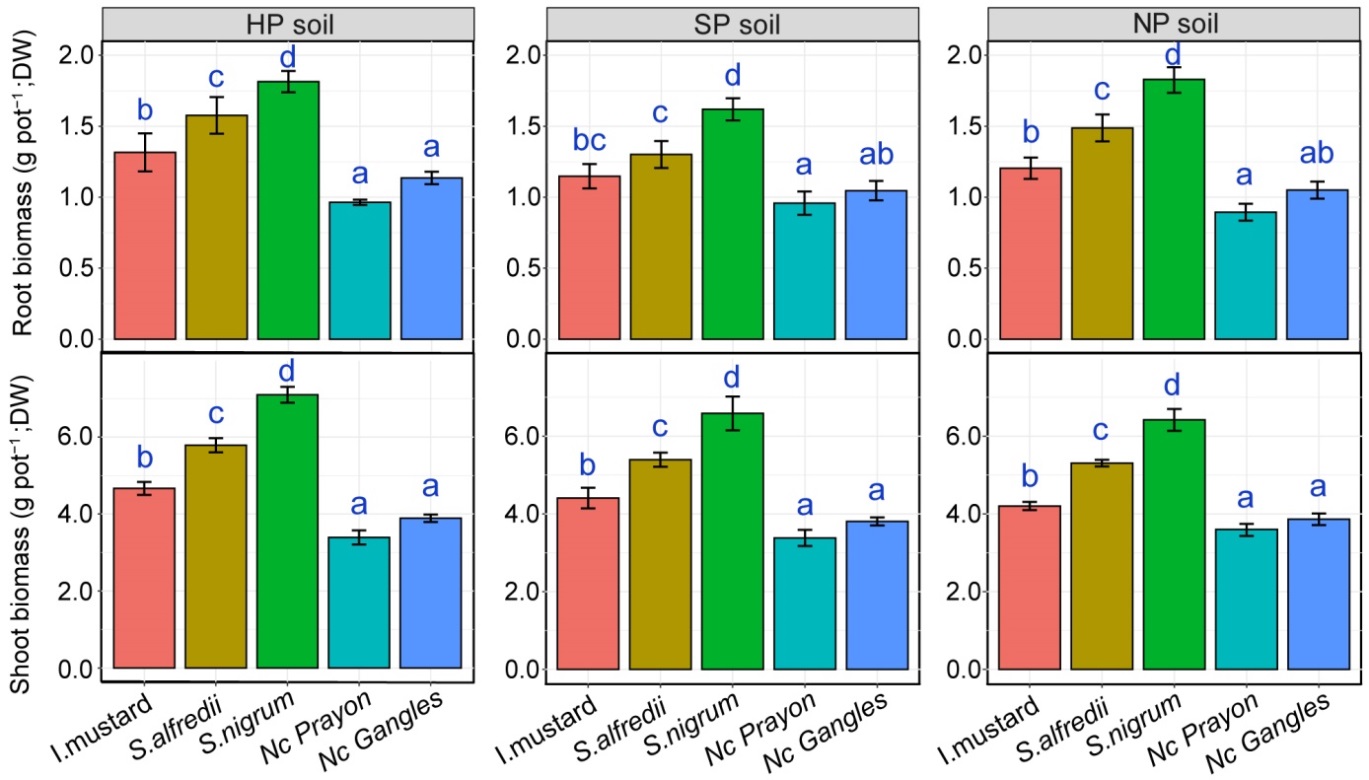
**Figure S3** Bacterial taxa that differed in their relative abundance according to rhizocompartment, plant species, soil’s geographic origin and soil pollution. (a) Heat map showing ASV counts from the rarefied ASV table (log2-transformed) for each of the rhizosphere-differentiating ASVs across three soils. ASVs were clustered on their Bray-Curtis dissimilarities. The key relates colors to the log-transformed read counts. (b) Histograms showing the proportion of bacterial classes present in unplanted soil and rhizosphere. The Proportion of bacterial families present among the ASVs from *Actinobacteria* (c), α-*proteobacteria* (d), γ-*proteobacteria* (e). The proportion of the rhizosphere bacterial taxa exhibiting differential abundance between plant species (f), soil’s geographic origin (g) and soil pollution (h) at different taxonomic levels.

****

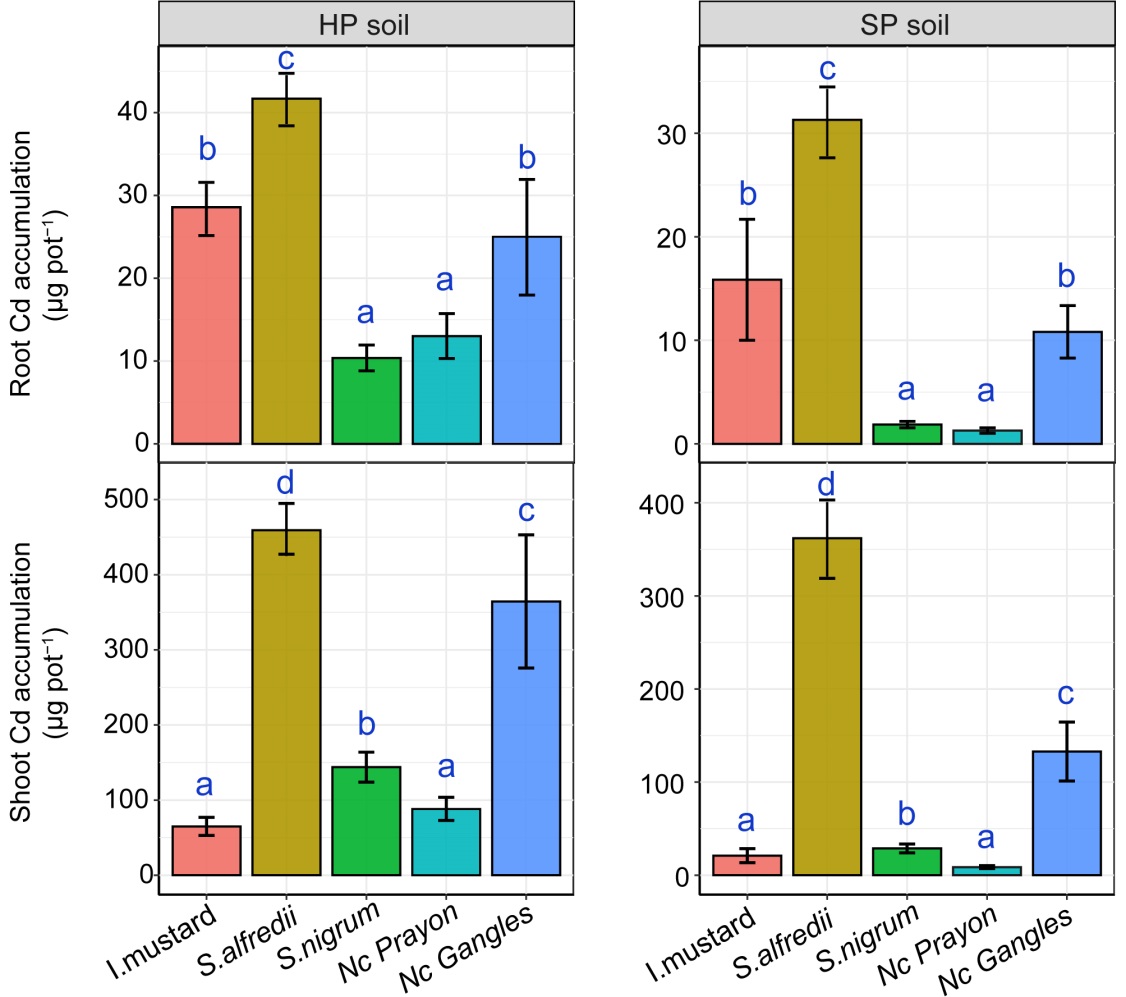
**Figure S4** Distribution of the top ten most abundant bacterial phyla between rhizosphere and unplanted soil (a), and among the five plants (b) across distinct soils.

****

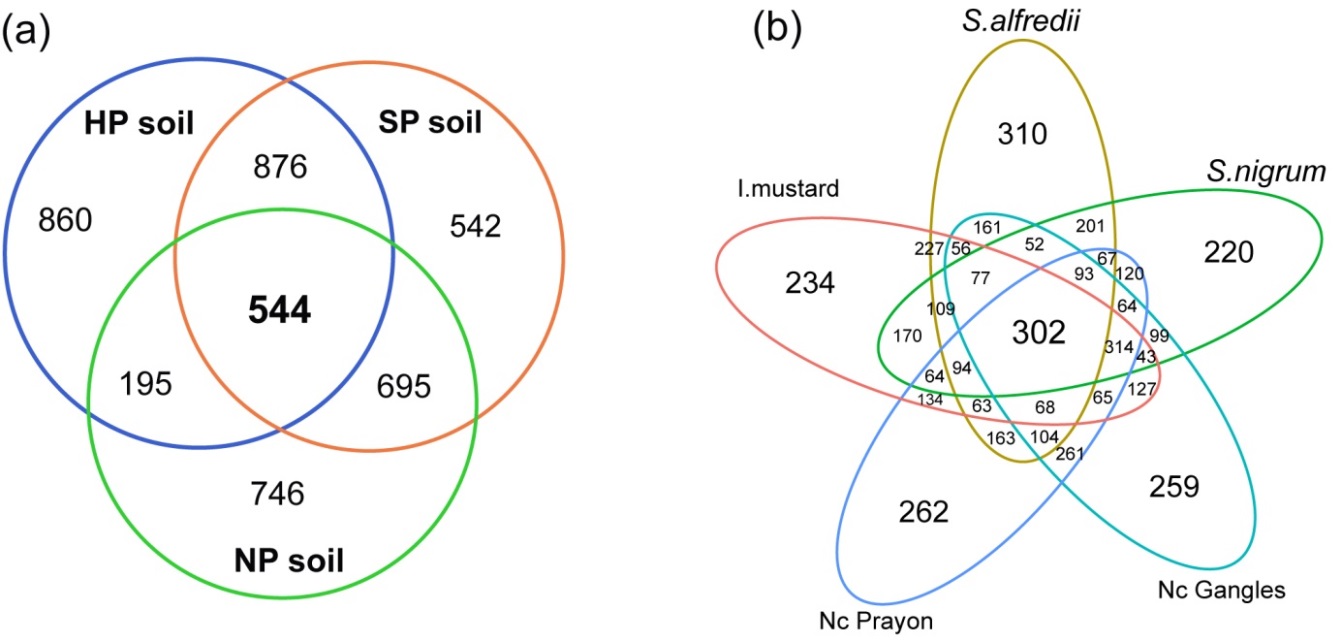
**Figure S5** Root biomass and shoot biomass of the five Cd-accumulating plants grown in three distinct soils. Values are mean (n = 5) ± standard deviations (s.d.). Letters above bars depict significant differences among different plants (*P* < 0.05). DW, dry weight. Abbreviations of Im, Sa, Sn, NcP and NcG are the same as those shown in Figure S4.

****

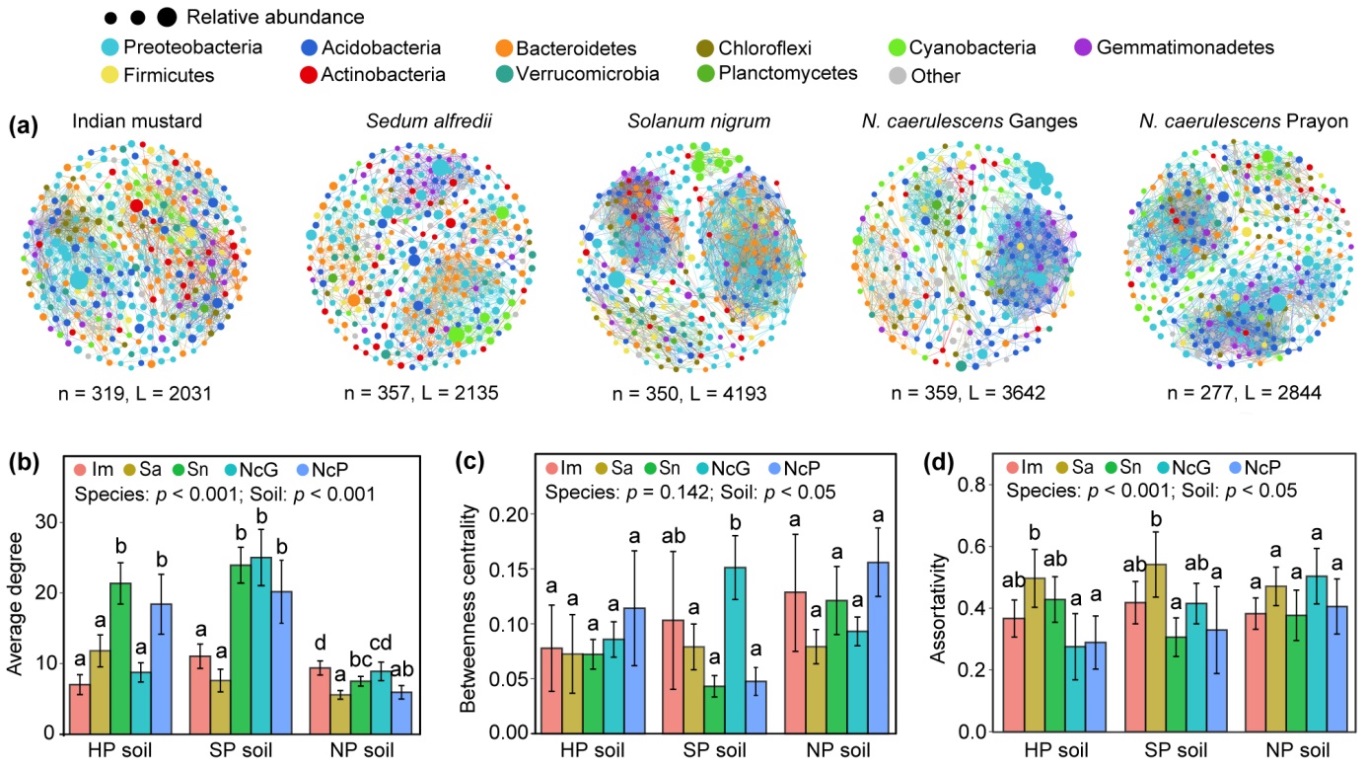
**Figure S6** Cd accumulation in the root and shoot of the five Cd-accumulating plants grown in highly polluted (HP) and slightly polluted (SP) soils. Values are mean (n = 5) ± s.d.. Blue letters above bars depict significant differences among different plants (*P* < 0.05).

****

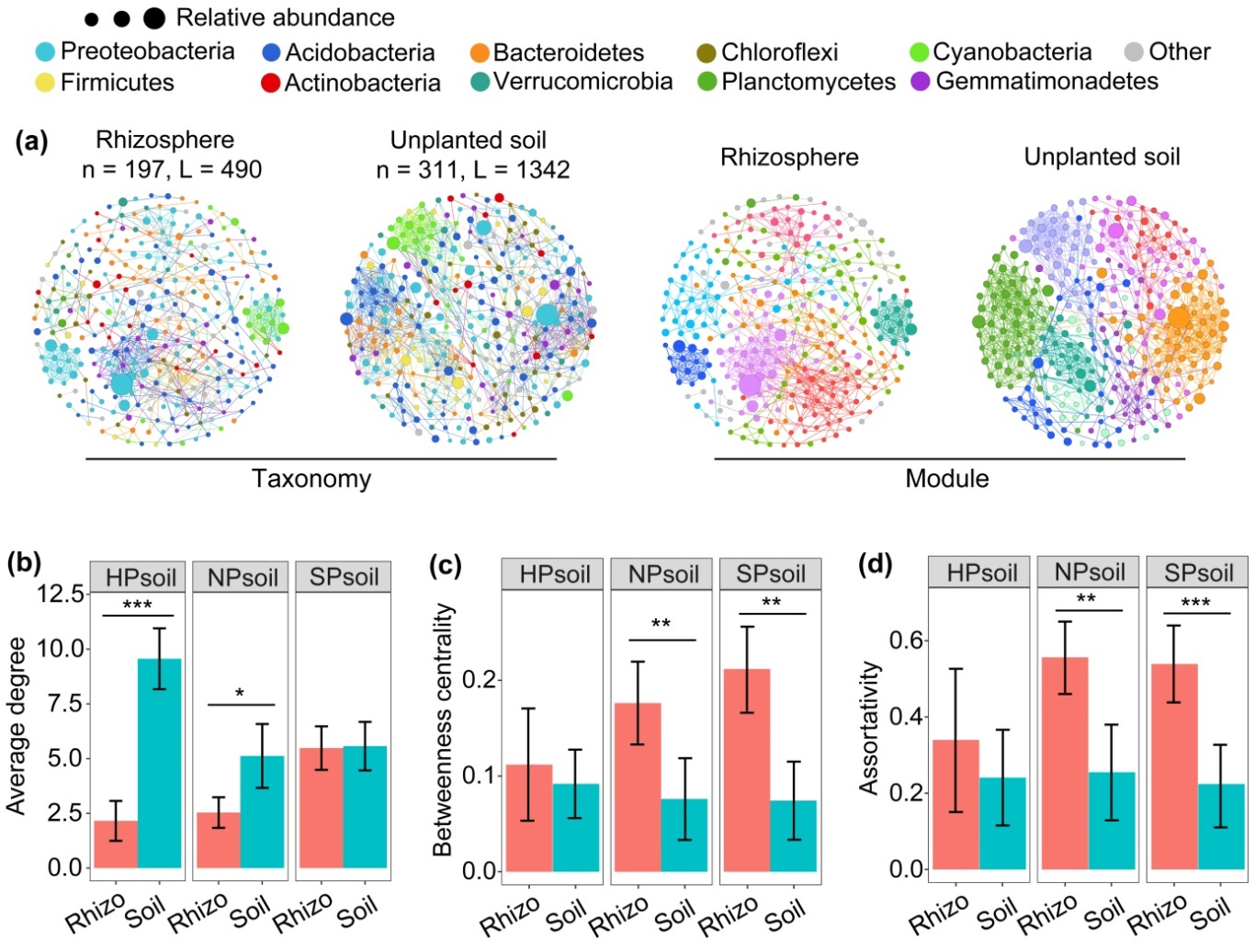
**Figure S7** Number of the shared phylotypes (ASVs) between the evaluated soils and between different Cd-accumulating plants. For Venn diagrams, ASV that found in more than two samples collected from a given soil type or a given plant species, and unplanted soil samples were removed.

****

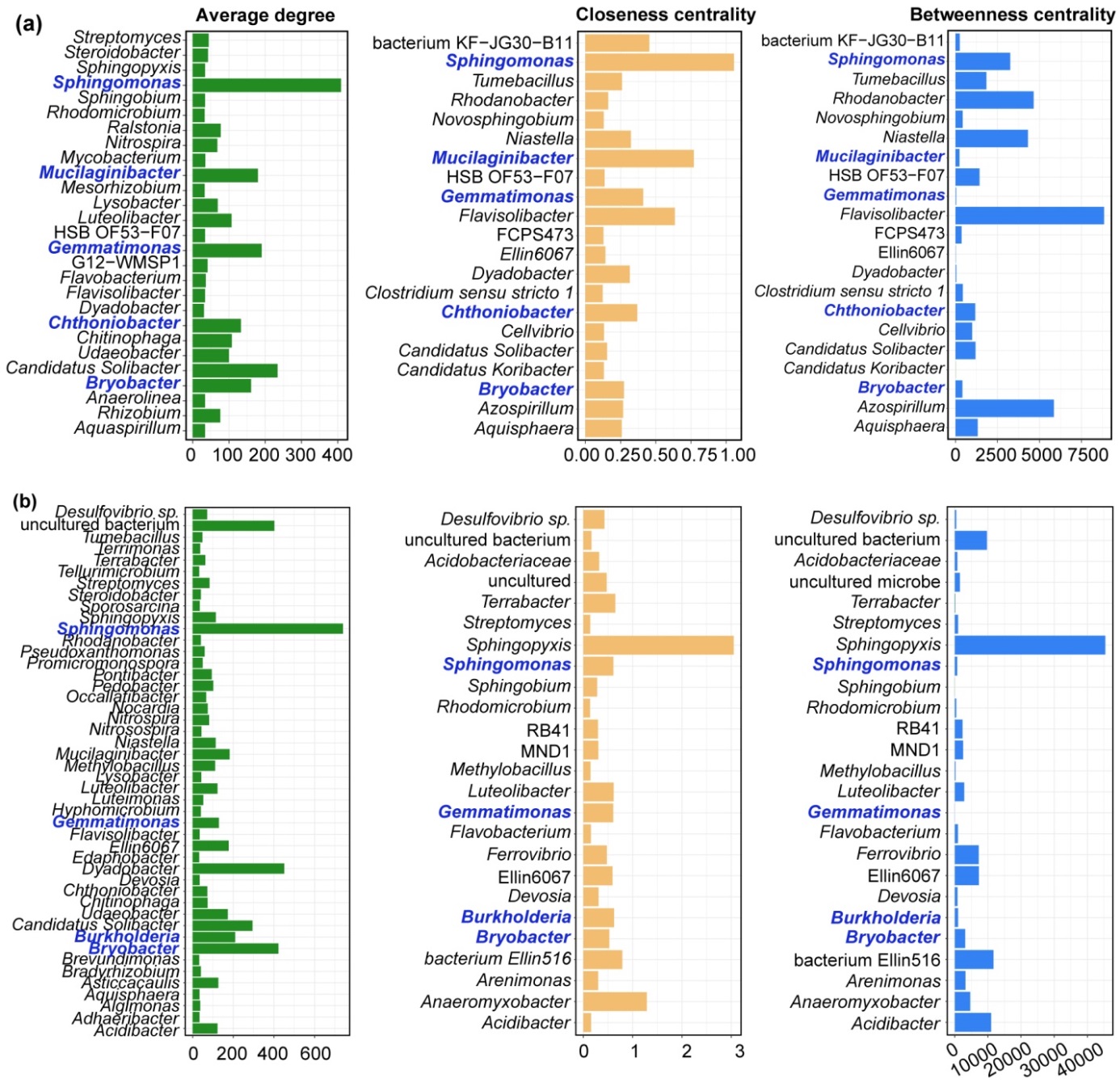
**Figure S8** Co-occurrence networks of the rhizosphere bacterial communities varied among plant species and soils evaluated. The rhizosphere co-occurrence networks varied among different Cd-accumulators (a). The size of each node is proportional to the relative abundance of ASV. A connection stands for a strong (Spearman’s ρ > 0.6) and significant correlation (*p*-value < 0.05, after Benjamini and Hochberg FDR adjust). Average degree (c), betweenness centrality (d), and assortativity (e) of the rhizosphere co-occurring networks of different Cd-accumulating plants grown in three diverse soils. Error bars are standard errors of the mean. The *p*-values represent the significance of plant species and soil types in the divergence of co-occurrence networks, as evaluated by Two-way ANOVA. Abbreviation, n and L represent the number of node and edge of each co-occurrence network.

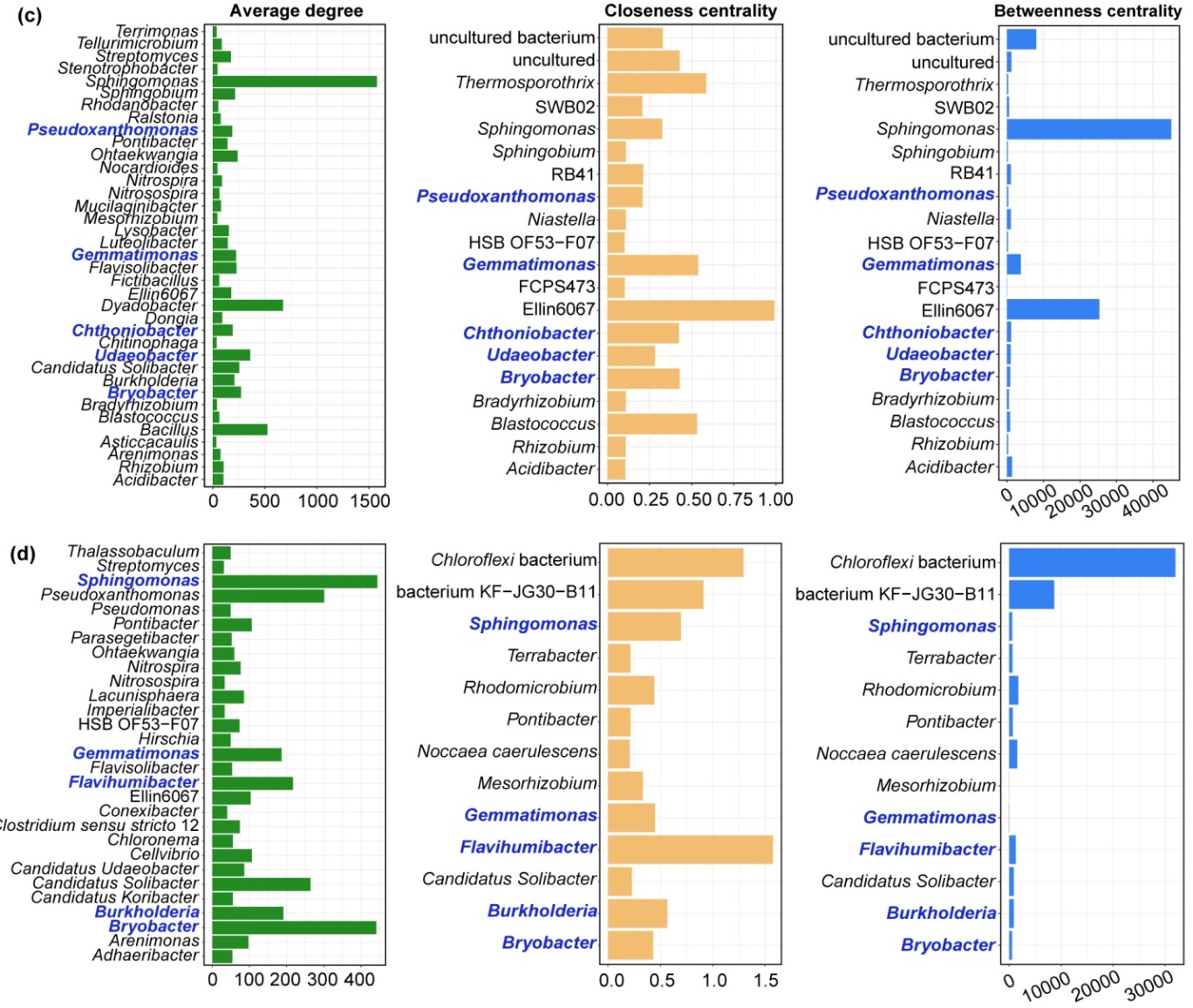
****

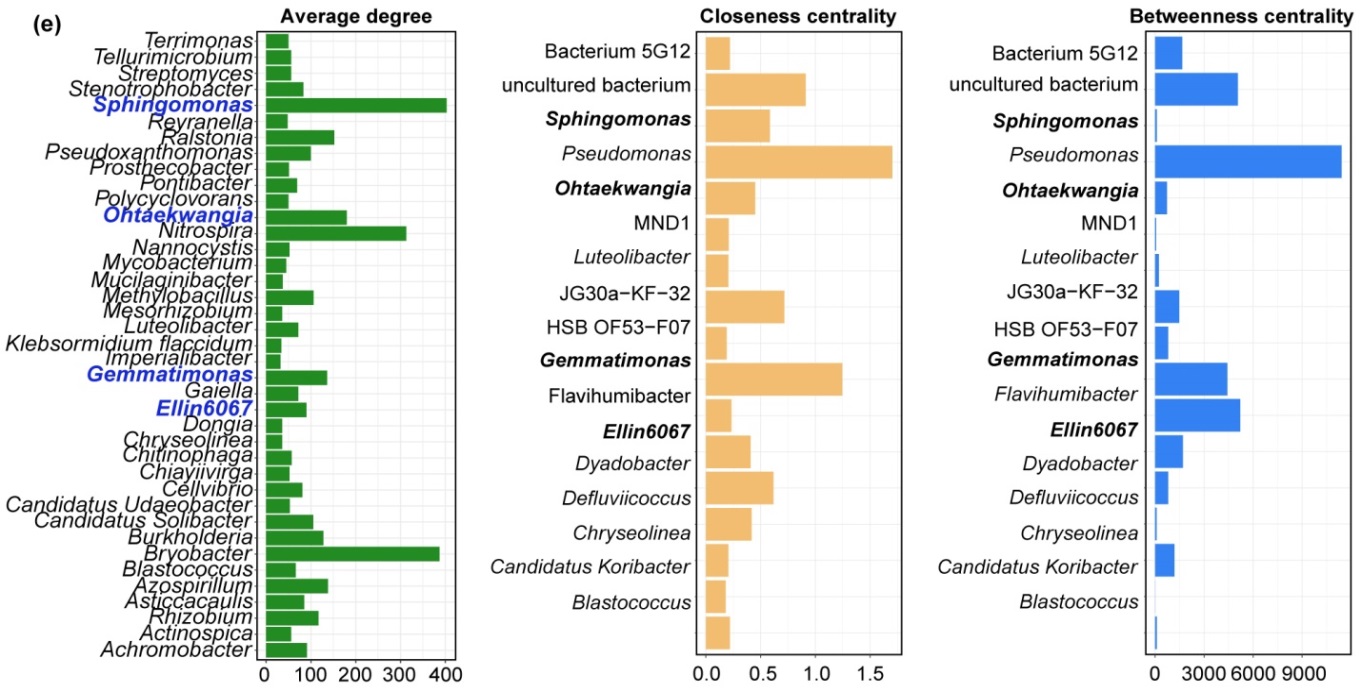
**Figure S9** Bacterial co-occurrence networks differed between the rhizosphere and unplanted soil. Bacterial co-occurrence networks of the overall rhizosphere and unplanted soil samples (a). The Average degree (b), betweenness centrality (c), and assortativity (d) of bacterial co-occurrences networks between the unplanted soil and rhizosphere of different plants grown in three diverse soils. Significant differences in these topological parameters between the rhizosphere and bulk soil were determined using the independent T test,\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001. Abbreviation, n and L represent the number of node and edge of each co-occurrence network; Rhizo, rhizosphere; Soil, unplanted soil.

****

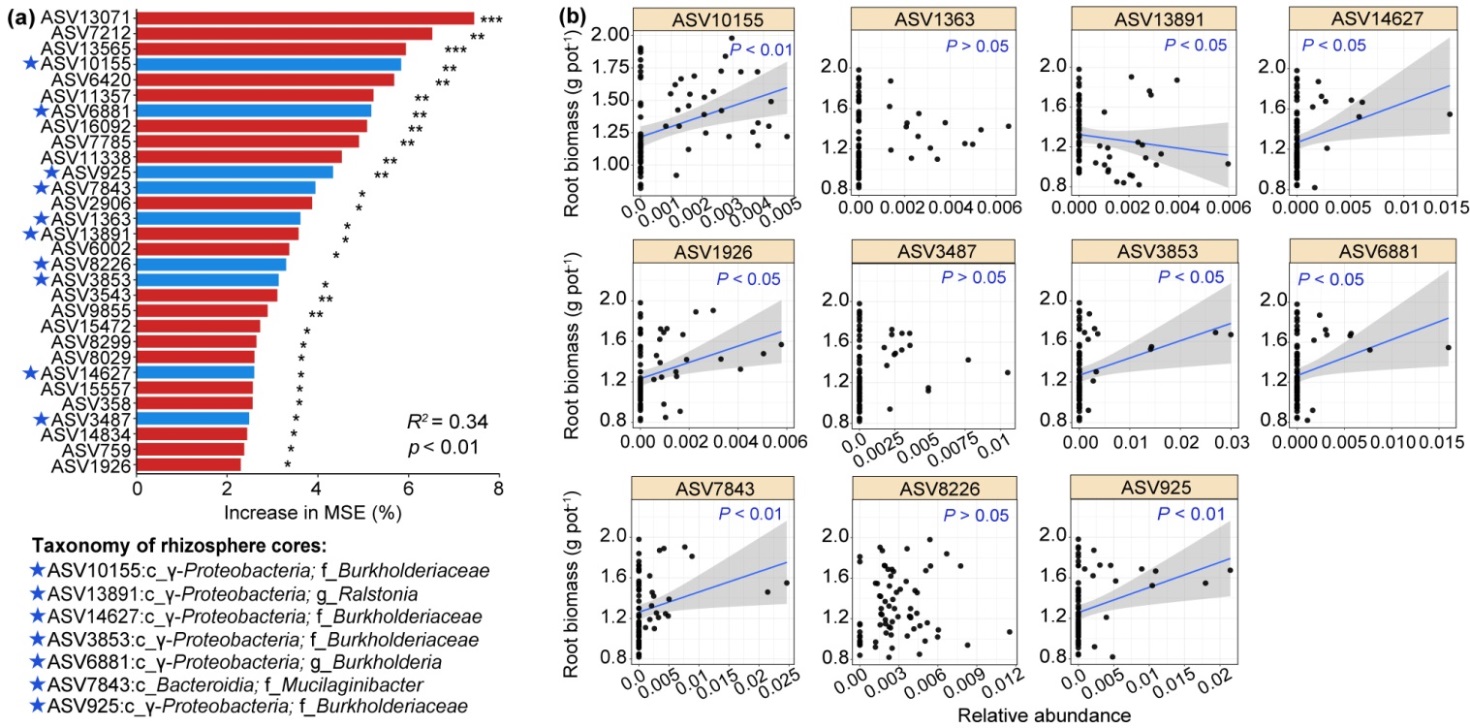
**Figure S10** Average degree, closeness centrality and betweenness centrality of the bacterial co-occurrence networks associated with the rhizosphere compartment of Indianmustard (a), *Sedum alfredii* (b), *Solanum nigrum* (c), *Noccaea caerulescens* (Ganges) (d), and *Noccaea caerulescens* (Prayon) (e). Keystone species is defined as a node with high average degree (> 50), high closeness centrality (> 0.2) and low betweenness centrality (< 5000). Taxa shown in blue bold indicate potential keystone species based on this threshold.

****

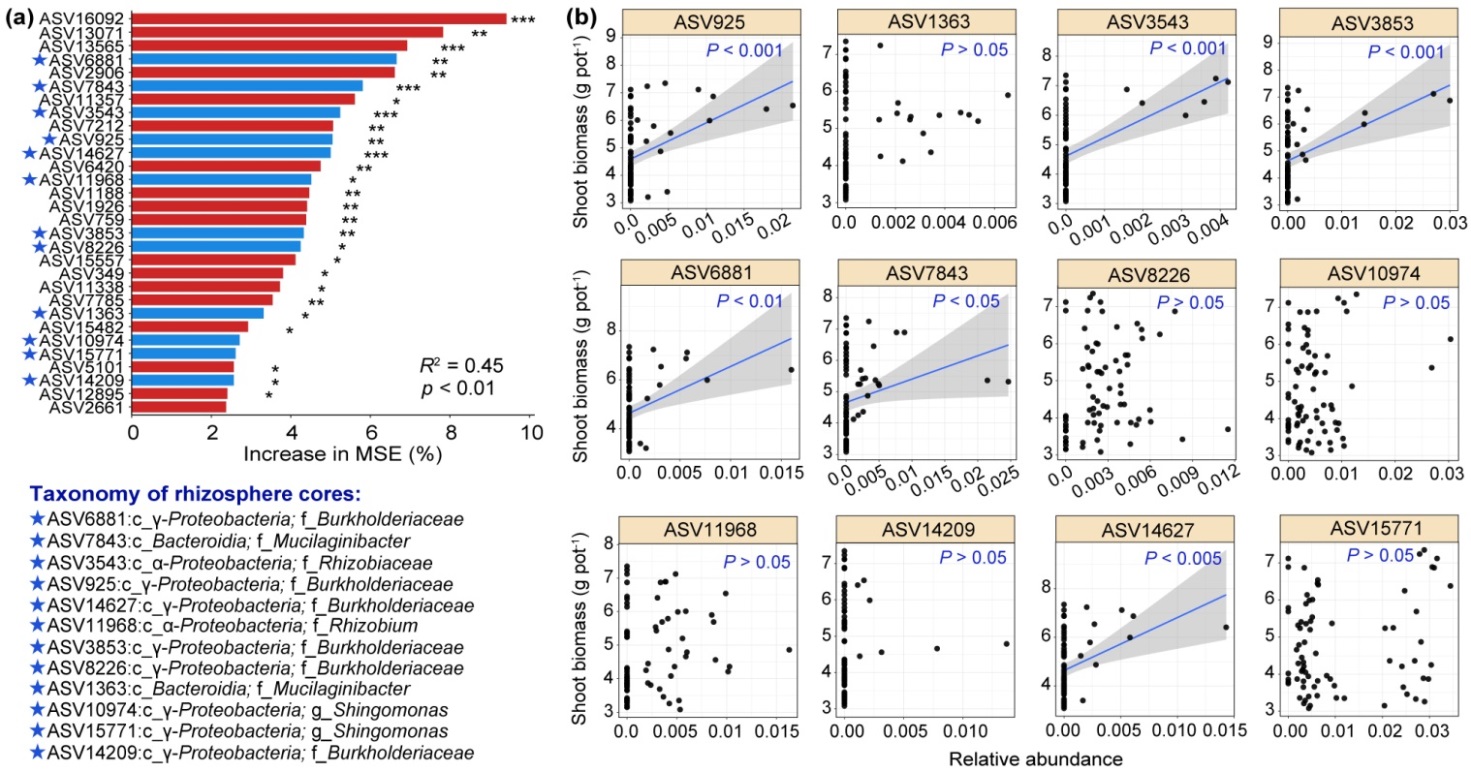
****



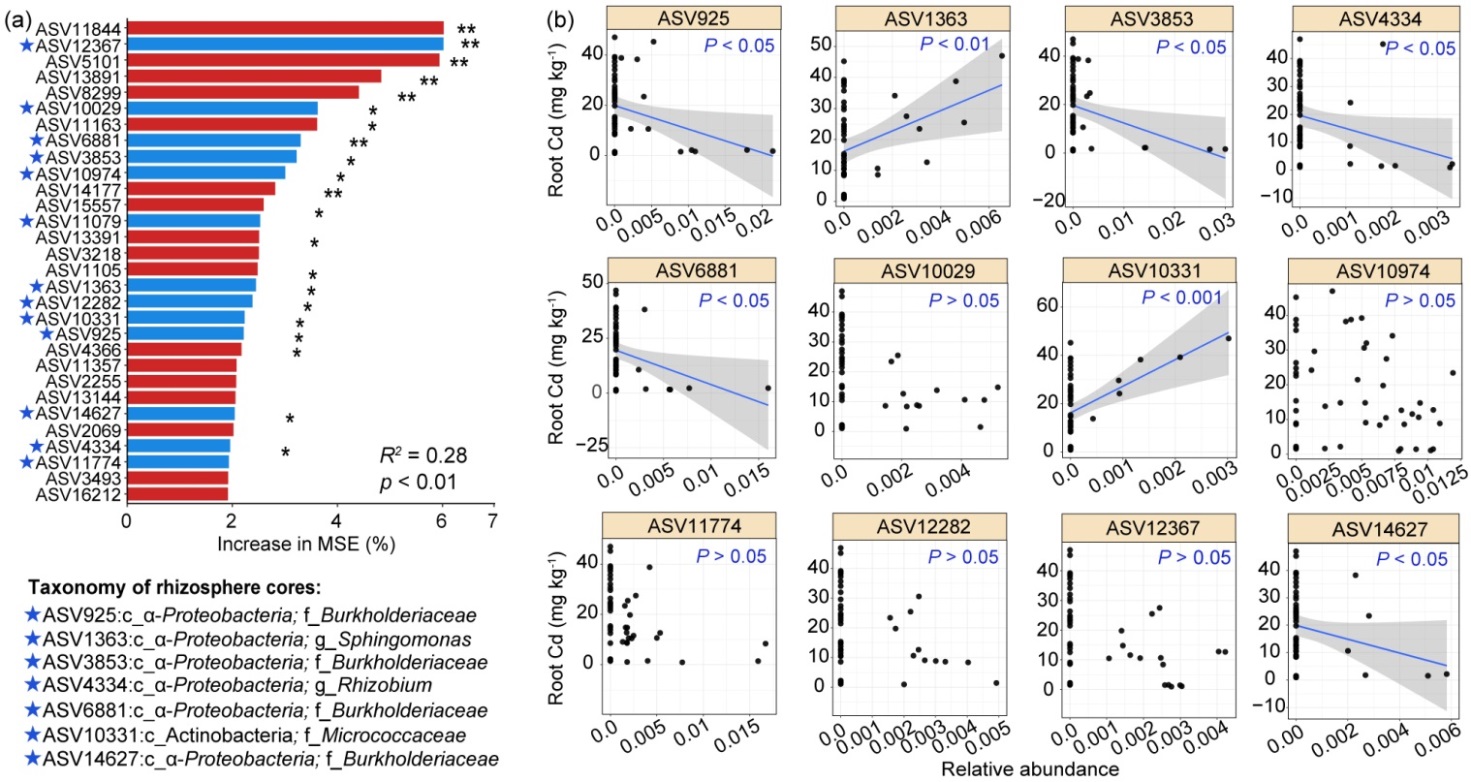
**Figure** **S11** Relationships between the core rhizosphere ASVs and root biomass of Cd-accumulating plants. Random Forest mean predictor importance (% of increase of MSE) of the top 30 most important ASVs in the rhizosphere (a) on root biomass. Linear regression relationship between root biomass and relative abundance of the core rhizosphere ASVs present in the top 30 most important predictors (b). The shaded area represents the 95% confidence interval (CI), which was displayed when the regression model was statistically significant at *P* < 0.05. Shown are core rhizosphere ASVs signified by pentacle and their taxonomic affiliations.



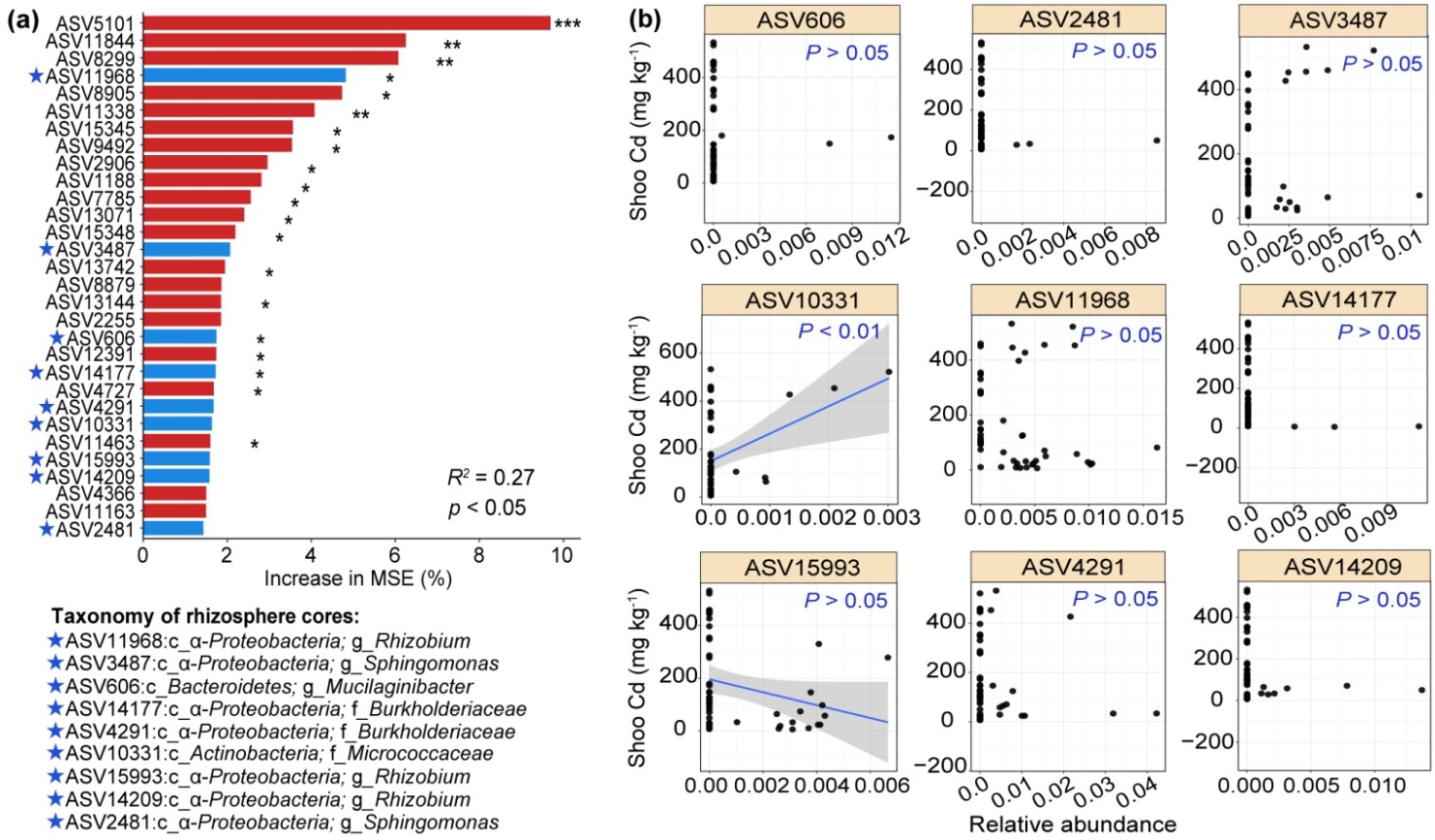
**Figure** **S12** Relationships between the core rhizosphere ASVs and shoot biomass of Cd-accumulating plants. Random Forest mean predictor importance (% of increase of MSE) of the top 30 most important ASVs in the rhizosphere (a) on shoot biomass. Linear regression relationship between shoot biomass and relative abundance of the core rhizosphere ASVs present in the top 30 most important predictors (b). The shaded area represents the 95% confidence interval (CI), which was displayed when the regression model was statistically significant at *P* < 0.05. Shown are core rhizosphere ASVs signified by pentacle and their taxonomic affiliations.



**Figure** **S13** Relationships between the core rhizosphere ASVs and root Cd content of Cd-accumulating plants. Random Forest mean predictor importance (% of increase of MSE) of the top 30 most important ASVs in the rhizosphere (a) on root Cd content. Linear regression relationship between root Cd content and relative abundance of the core rhizosphere ASVs present in the top 30 most important predictors (b). The shaded area represents the 95% confidence interval (CI), which was displayed when the regression model was statistically significant at *P* < 0.05. Shown are core rhizosphere ASVs signified by pentacle and their taxonomic affiliations.



**Figure** **S14** Relationships between the core rhizosphere ASVs and shoot Cd content of Cd-accumulating plants. Random Forest mean predictor importance (% of increase of MSE) of the top 30 most important ASVs in the rhizosphere (a) on shoot Cd content. Linear regression relationship between shoot Cd content and relative abundance of the core rhizosphere ASVs present in the top 30 most important predictors (b). The shaded area represents the 95% confidence interval (CI), which was displayed when the regression model was statistically significant at *P* < 0.05. Shown are core rhizosphere ASVs signified by pentacle and their taxonomic affiliations.



**Supplementary Tables**

**Table S1** Determining drivers of soil bacterial community using PERMVNOVA. Bold values indicate significance at *p* < 0.05. Factors soil type indicates variation in the three evaluated soils, and pollution indicates the differences between HP and SP soils.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Factors | Sum of sqs | F. Model | *R2* | *p*-value |
| **Total experiment** |  |  |  |  |
| Compartment | 0.239 | 17.00 | 6.46 | **0.001** |
| Species | 0.436 | 7.75 | 11.79 | **0.001** |
| Soil type | 1.519 | 54.00 | 41.05 | **0.001** |
| Compartment × Soil type | 0.153 | 5.44 | 4.14 | **0.001** |
| Species × Soil type | 0.354 | 3.15 | 9.57 | **0.001** |
| Residuals | 0.998 |  | 26.99 |  |
| **Soil pollution experiment** |  |  |  |  |
| Compartment | 0.187 | 9.33 | 7.40 | **0.001** |
| Species | 0.371 | 4.62 | 14.65 | **0.001** |
| Pollution | 0.790 | 19.65 | 31.16 | **0.001** |
| Pollution × Species | 0.222 | 2.76 | 8.75 | **0.001** |
| Residuals | 0.964 |  | 38.05 |  |

**Table S2** Rhizosphere bacterial indicator ASVs in distinct Cd-accumulating plants

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phylum or Class | ASV (Genus or higher) | ASVs represented | Average relative abundance (%) | *Indian mustard* | *Sedum alfredii* | *Solanum*  *nigrum* | *Nc.*  *Ganges* | *Nc.*  *Prayon* |
| β-*Proteobacteria* | *Burkholderiaceae* | 9 | 5.17 | Y (3) | Y (3) | Y (3) | Y (1) | Y (1) |
| β-*Proteobacteria* | *Ralstonia* | 1 | 0.20 |  | Y (1) |  | Y (1) |  |
| α-*Proteobacteria* | *Sphingomonas* | 17 | 2.89 | Y (6) | Y (6) | Y (5) | Y (3) | Y (4) |
| δ-*Proteobacteria* | *Myxococcales* | 1 |  |  |  |  |  |  |
| α-*Proteobacteria* | *Bradyrhizobium* | 1 | 0.36 |  |  |  |  |  |
| *Bacteroidetes* | *Flavisolibacter* | 5 | 0.44 | Y (2) | Y (2) | Y (1) |  |  |
| *Bacteroidetes* | *Mucilaginibacter* | 1 | 0.12 |  | Y (1) |  |  |  |
| *Bacteroidetes* | *Microscillaceae* | 1 | 1.87 |  |  | Y (1) |  |  |
| *Firmicutes* | *Sporosarcina* | 8 | 0.06 | Y (3) | Y (2) | Y (3) |  | Y (1) |
| *Verrucomicrobia* | *Udaeobacter* | 2 | 1.18 |  | Y (1) | Y (1) |  |  |
| *Gemmatimonadetes* | *Gemmatimonadaceae* | 1 | 4.88 |  |  | Y (1) |  |  |
| *Planctomycetes* | *Aquisphaera* | 1 | 0.06 |  |  | Y (1) |  |  |
| *Cyanobacteria* | *Coleofasciculaceae* | 1 | 0.32 |  |  |  |  | Y (1) |

Data table shows results for the analysis where rare ASVs (< 0.1% relative abundance) were excluded, and an ASV with *Indicator* value > 0.6 and *P* < 0.05 are strong indicators for a species. Average relative abundance of an ASV is the total amount of its representative sequences within rhizosphere samples. Value in the bracket is the number of the ASV representing the indicator species.

**Table S3** Correlations and topological properties of the total rhizosphere and unplanted soil microbiome networks, and rhizosphere network of individual Cd-accumulating plant

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Plant | Node | Positive  edge | Negative edge | Average  degreea | Modularityb |
| Unplanted soil | 311 | 1325 | 17 | 8.63 | 0.751 |
| Rhizosphere | 307 | 687 | 64 | 6.97 | 0.803 |
| *I.mustard* | 319 | 2014 | 17 | 12.73 | 0.674 |
| *S. alfredii* | 357 | 2120 | 15 | 11.96 | 0.703 |
| *S. nigrum* | 350 | 4180 | 13 | 23.96 | 0.603 |
| Nc. Prayon | 359 | 3613 | 29 | 20.29 | 0.638 |
| Nc. Ganges | 277 | 2822 | 22 | 20.53 | 0.532 |

a The average number of connections per node in the network, that is, the node connectivity.

b The capability of the nodes to form highly connected communities, that is, a structure with high density of between nodes connections.

Abbreviations: Nc. Prayon, Prayon ecotype of *Noccaea caerulescens*; Nc. Ganges, Ganges ecotype of *Noccaea caerulescens*.

**Table S4** Comparison of phyla differential abundance between 13C-labelled and 12C-labelled rhizosphere samples. Fold change (FC) was determined from the contrasts of the ratio of relative abundance in13C-labelled samples to 12C-labelled samples using the DESeq function in DESeq2 package in R. lfcSE, standard deviation of log2 fold change. *p*-adj, the *p*-values adjusted using the false discovery rate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Phylum | log2FC | lfcSE | stat | *p*-value | *p*-adj |
| Acidobacteria | -11.69 | 0.86 | 13.61 | 3.44E-42 | 4.13E-41 |
| Chloroflexi | -10.96 | 0.83 | 13.20 | 8.69E-40 | 5.22E-39 |
| Planctomycetes | -10.42 | 0.83 | 12.62 | 1.67E-36 | 6.69E-36 |
| Gemmatimonadetes | -10.10 | 0.87 | 11.61 | 3.48E-31 | 1.04E-30 |
| Firmicutes | 11.35 | 1.37 | 8.28 | 1.23E-16 | 2.96E-16 |
| Bacteroidetes | 10.02 | 1.33 | 7.53 | 5.02E-14 | 1.00E-13 |
| Nitrospirae | -7.43 | 1.07 | 6.94 | 3.99E-12 | 6.83E-12 |
| Verrucomicrobia | -3.65 | 1.76 | 6.62 | 3.65E-05 | 5.47E-05 |
| Armatimonadetes | -6.75 | 1.67 | 4.05 | 5.23E-05 | 6.97E-05 |
| Proteobacteria | 2.63 | 0.83 | 3.16 | 0.0016 | 0.0019 |
| Patescibacteria | -3.91 | 2.96 | 1.32 | 0.187 | 0.1969 |
| Actinobacteria | -2.92 | 1.67 | 1.75 | NA | NA |

**Table S5** Physiochemical parameters of the three tested soils used in this study

|  |  |  |  |
| --- | --- | --- | --- |
| Soil parameters | Highly polluted soil | Slightly polluted soil | Non-polluted soils |
| pH (soil:water=1:2.5) | 6.43 | 6.57 | 4.69 |
| Organic matter (g kg-1) | 13.5 | 9.54 | 10.24 |
| NH4+-N (mg kg-1) | 14.8 | 13.9 | 18.7 |
| Total N (g kg-1) | 0.75 | 0.57 | 0.95 |
| Available P (mg kg-1) | 41.66 | 98.45 | 23.5 |
| Available K (mg kg-1) | 72.66 | 76.28 | 52.54 |
| CEC (cmol kg-1) | 9.87 | 9.12 | 14.3 |
| Total Cd (mg kg-1) | 12.54 | 0.98 | ND |
| Total Zn (mg kg-1) | 824.5 | 242.7 | 35.62 |
| DTPA-Cd (mg kg-1) | 4.86 | 0.46 | ND |
| DTPA-Zn (mg kg-1) | 52.57 | 20.75 | 2.15 |
| Water-Cd (mg kg-1) | 82.14 | 7.86 | ND |
| Water-Zn (mg kg-1) | 378.4 | 122.54 | 8.63 |

Total N, total nitrogen; P, phosphorus; K, potassium; CEC, cation exchange capacity; DTPA-Cd, DTPA-extractable Cd; DTPA-Zn, DTPA-extractable Zn; Water-Cd, water-soluble Cd; Water-Zn, water-soluble Zn; ND, not detected.

**References**

1. Yeoh YK, Dennis PG, Paungfoo-Lonhienne C, Weber L, Brackin R, Ragan MA, et al. Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. Nat Commun. 2017;8:215.

2. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ, et al. Assembly and ecological function of the root microbiome across angiosperm plant species. Proc Natl Acad Sci USA. 2018;115:E1157-E1165.

3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.