Diversity and heritability of the maize rhizosphere microbiome under field conditions

Jason A. Peiffer[a,1,2], Aymé Spor[b,c,1,3], Omry Koren[b,c], Zhao Jin[b,c], Susannah Green Tringe[d], Jeffery L. Dangl[e,f,4], Edward S. Buckler[a], and Ruth E. Ley[b,c,4]

[a]Department of Plant Breeding and Genetics, [b]Department of Microbiology, and [c]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853; [d]Department of Energy, Joint Genome Institute, Walnut Creek, CA, 94598; and [e]Howard Hughes Medical Institute, Department of Biology, Curriculum in Genetics and Molecular Biology, and [f]Department of Microbiology and Immunology, Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC 27599

Contributed by Jeffery L. Dangl, February 15, 2013 (sent for review January 10, 2013)

The rhizosphere is a critical interface supporting the exchange of resources between plants and their associated soil environment. Rhizosphere microbial diversity is influenced by the physical and chemical properties of the rhizosphere, some of which are determined by the genetics of the host plant. However, within a plant species, the impact of genetic variation on the composition of the microbiota is poorly understood. Here, we characterized the rhizosphere bacterial diversity of 27 modern maize inbreds possessing exceptional genetic diversity grown under field conditions. Randomized and replicated plots of the inbreds were planted in five field environments in three states, each with unique soils and management conditions. Using pyrosequencing of bacterial 16S rRNA genes, we observed substantial variation in bacterial richness, diversity, and relative abundances of taxa between bulk soil and the maize rhizosphere, as well as between fields. The rhizospheres from maize inbreds exhibited both a small but significant proportion of heritable variation in total bacterial diversity across fields, and substantially more heritable variation between replicates of the inbreds within each field. The results of this study should facilitate expanded studies to identify robust heritable plant–microbe interactions at the level of individual polymorphisms by genome wide association, so that plant-microbiome interactions can ultimately be incorporated into plant breeding.

Most land plants grow in intimate association with complex microbiota. Microbes in both the phyllosphere and rhizosphere can be endophytic, epiphytic, or closely associated. Examples of close microbial associates in the phyllosphere include those inhabiting the fluid in pitcher plants (1), and close microbial associates in the rhizosphere include those not touching roots, but heavily influenced by root exudates in the nearby soil (2). The host plant often relies on the microbiome to provide it with critical nutrients, as minerals present in the soil may be in forms inaccessible to plants (2). In other cases, plant-associated microbes can act as protectants against phytopathogens (3), improve growth through production of phytohormones (4), help plants withstand heat (5), salt (6), and more. The plant, in turn, cultivates its microbiome by adjusting the soil pH, reducing competition for beneficial microbes, and providing an energy source, mostly in the form of carbon-rich rhizodeposits (2). Microbial community structure in the phyllosphere and rhizosphere often differ across plant species (7), as well as among genotypes within a single species (8, 9). Recent work with model systems (*Arabidopsis thaliana* cultivated under controlled conditions in natural soils) indicated that the host genotype has a small but measurable effect on the microbes inhabiting the endophyte compartment of the root (10, 11). Understanding interactions between microbiota and their host plants, and identifying the plant alleles controlling these interactions, could be transformational in plant breeding and biotechnology.

Maize is one of the most economically significant crops in the world, possesses exceptional phenotypic and molecular diversity (12), and is substantially influenced by environmentally conditional genetic variation (13). In addition, given its widespread planting in monoculture, maize may be viewed as an ecosystem engineer strongly responsible for shaping the agricultural environment for cohabitating species. Maize root exudates, such as sugars, organic acids, aromatics, and enzymes interact with soil traits, such as pH, water potential, texture, and nutrient availability, as well as existing microbial populations to promote growth and development of the plant. Recent work addressing the association between maize genotype and microbial diversity using 16S rRNA gene microarrays in a greenhouse setting or fingerprinting under field conditions support the notion that microbial diversity is related to plant genotype (8, 9) To date, however, the rhizosphere microbiota of mature plants growing under field conditions remain poorly characterized, and many of the roles and interactions of maize genetic diversity and the field environment (including the resident soil microbiota) remain to be elucidated.

Here, we characterize the rhizosphere microbiota across a genetically diverse collection of 27 modern maize inbreds (14) at their median flowering time in five agricultural field environments by pyrosequencing 16S rRNA gene amplicons. The fields were situated in two regions of the United States: two fields were located in the Midwest (Urbana, IL and Columbia, MO) and three fields in the Northeast (Aurora, Ithaca, and Lansing) in the state of New York. Each field included randomized and replicated plots of maize inbreds grown from seed obtained from a uniform stand in a single field environment. Roots of three to five randomly selected plants per genotype were gently processed for rhizosphere microbial DNA, for a total of ~100 samples per field. This design allowed us to evaluate the effects of the environment, sample type (bulk soil or rhizosphere), maize genotype, and genotype-field interactions on the microbial community diversity.

Results

16S rRNA Gene Sequence Region Impacts Recovered Maize Rhizosphere Microbial Diversity. Results from previous microbiota analyses in other systems (i.e., human gut microbiome) have suggested that

Author contributions: J.A.P., A.S., S.G.T., J.L.D., E.S.B., and R.E.L. designed research; J.A.P., A.S., O.K., Z.J., and S.G.T. performed research; S.G.T. and E.S.B. contributed new reagents/analytic tools; J.A.P., A.S., O.K., E.S.B., and R.E.L. analyzed data; and J.A.P., A.S., O.K., Z.J., J.L.D., and R.E.L. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the QIIME (Quantitative Insights into Microbial Ecology) database (study no. 1792).

[1]J.A.P. and A.S. contributed equally to this work.

[2]Present address: Department of Genetics, Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695.

[3]Present address: Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1347 Agroécologie, BP86510, F-21000 Dijon, France.

[4]To whom correspondence may be addressed. E-mail: dangl@email.unc.edu or rel222@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302837110/-/DCSupplemental.

the specific region of the 16S rRNA gene used to characterize microbial diversity can impact the diversity profile obtained (15). Furthermore, some commonly used primer pairs are prone to prime plastid genes (16). Thus, we first conducted a pilot study to test different 16S rRNA gene primer pairs on a subset of the samples, which included maize rhizosphere and bulk soil samples collected at median flowering time from one location (Columbia, MO) (Fig. 1 and Table S1). We tested four primer sets designed to amplify the V1–V2, V3–V4, V5–V8, and V6–V8 regions of the 16S rRNA genes (*Materials and Methods*).

The primer pairs tested herein resulted in different absolute counts of sequence reads, with the 804F-1392R (V5–V8) primer set yielding the fewest sequences, likely because of the longer amplicons compared with the other primers used. Importantly, the different primer sets yielded slightly different diversity profiles at the phylum level (Fig. 1*A*). Specifically, the 515F-806R primer pair (V3–V4 region) yielded the greatest diversity at the domain and bacterial phylum levels (Fig. 1*A*). In contrast, the commonly used 27F-338R (V1–V2), which is specific to the bacterial domain, resulted in low amplification of members of the phylum Verrucomicrobia, known to be one of the dominant phyla in soil (17). Our results showed that the 926F-1392R primers (V6–V8) amplified the highest fraction of plastid 16S rRNA genes ("unassigned" in Fig. 1*A*). The four primer pairs tested also varied in the proportion of classifiable sequences ($n \geq 14$, $P < 2.00\text{E-}04$) (Fig. S1).

We compared the β-diversity (between sample diversity) using the weighted UniFrac (18) distance metric (a phylogeny-based distance metric sensitive to sequence abundances). This analysis revealed differences in diversity based on the primer pair used (Fig. 1*B*). In particular, the 27F-338R and 515F-806R primed samples formed clear clusters, and a third cluster was formed by samples primed with the 804F-1392R and 926F-1392R primer pairs (Fig. 1*B*). However, within each of the three diversity clusters, the rhizosphere and bulk soil samples were consistently differentiable from the rhizosphere samples. Furthermore, within the 27F-338R and 515F-806R primed clusters, the three maize genotypes separated well, with the exception of three samples (Fig. 1*B*). In the third cluster (primers 804–1392R and 926F-1392R), the separation by genotype was less strong. Based on these results, we elected to amplify the V3–V4 region of the 16S rRNA gene (primers 515F-806R) in the full set of samples, consisting of maize inbred rhizosphere and matched bulk soil microbiome extractions across all five surveyed field environments.

Biogeography Effects on α-Diversity. To directly compare the α-diversity [within-sample diversity or estimate of species richness (19)] of the samples with differing sequence counts/sampling efforts, we rarefied the data (i.e., randomly picked an equal number of sequence across samples) using QIIME (Quantitative Insights into Microbial Ecology) (20). Rarefaction curves, which show the observed operational taxonomic unit (OTU) richness as a function of pyrosequencing effort, indicated that the sequencing depth was insufficient to wholly capture the diversity present (Fig. S2). In addition to observed OTU richness, we used Faith's PD [a phylogenetic measure of diversity based on total branch length of the bacterial 16S rRNA gene phylogeny captured by a sample (21)] and the Chao-1 estimator of total species richness (22), all of which yielded similar results (Table S2).

After controlling for factors, such as pyrosequencing run and PCR amplification batch, two important nongenetic factors accounted for a sizable fraction of the variation in OTU richness across samples: (*i*) the specific field [20.0% of variance; $P < 2.00\text{E-}04$; 95% confidence interval (CI) = 19.8%, 20.4%] (Fig. S2*A*), and (*ii*) the sample type (bulk soil or rhizosphere; 32.3% of variance; $P < 2.00\text{E-}04$; 95% CI = 32.0%, 32.6%) (Fig. S2*B*). The Columbia, MO field was the most OTU-rich field ($n \geq 258$ OTUs, $P < 8.00\text{E-}04$) (Fig. S2*A*). However, the remaining field environments located near Urbana, IL and the three fields sampled in New York did not significantly differ in α-diversity.

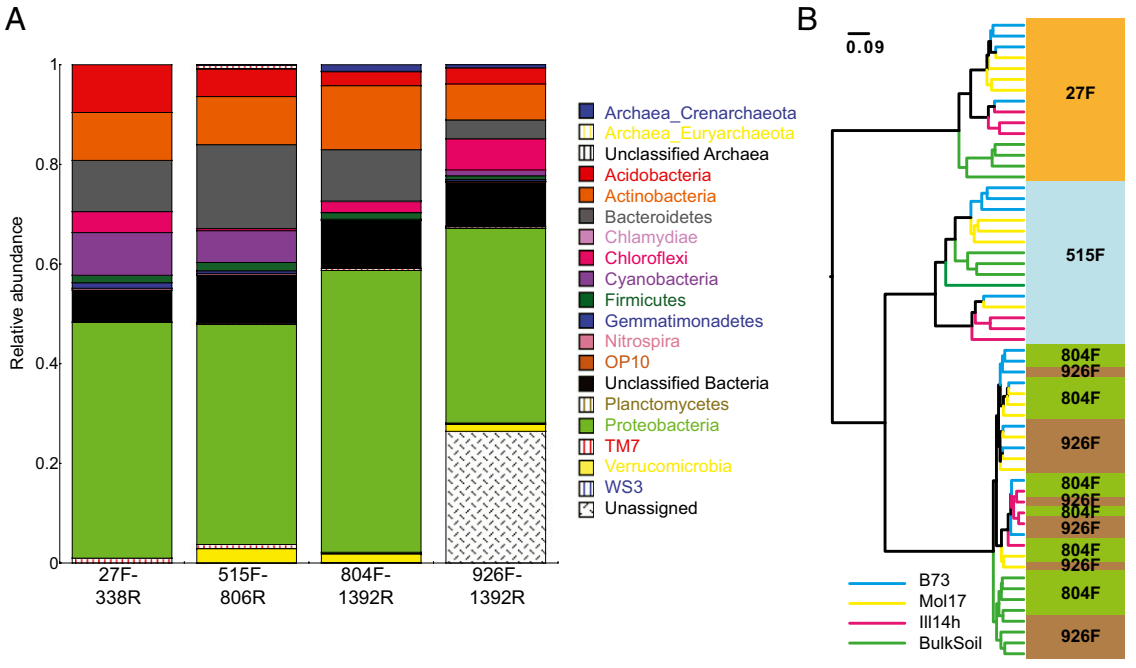

Fig. 1. Primer set influences the diversity profiles of the rhizosphere and bulk soil microbiota. (*A*) Relative abundances of the phyla and domains recovered using four different primer pairs (27F-338R, 515F-806R, 804F-1392R, 926F-1392R) applied to rhizosphere soils recovered from three maize genotypes (B73, Mol17, Ill14h) and four bulk soil samples. (*B*) Hierarchical clustering of the weighted UniFrac distance metric for the same sample set. To highlight patterns of clustering, branches are colored by sample origin (maize genotypes and bulk soil: light blue, B73; yellow, Mo17; pink, Ill14h; and green, Bulk Soil). Primer pairs are indicated on the branch labels (the forward primer is indicated) and are highlighted by the colored boxes (27F: orange; 515F: blue; 804F: green; 926F: brown).

Interestingly, the microbiota of the organically managed field in Ithaca, NY did not differ significantly in species richness from the conventionally managed fields. In every comparison, the rhizosphere was less rich (lower bacterial α-diversity) than bulk soil ($P < 2.00E-03$) (Fig. S2*B*).

Biogeography Effects on β-Diversity. We used the unweighted (sensitive to rare taxa) and weighted (based on abundances of taxa) UniFrac distance metrics to estimate β-diversity. A bootstrapped partial canonical analysis of principal coordinates (CAP) (23) was performed to assess how β-diversity could be partitioned into variation attributable to fields, sample type, and maize inbreds. This analysis differs from the commonly used unconstrained principle coordinate (PC) analysis (PCoA) representation of UniFrac distances in two ways: first, the variation attributable to technical variance is controlled for, and second, the ordination is constrained such that only the variation attributable to the factors of interest is displayed; however, we also applied unconstrained PCoA of both UniFrac distances as a comparison. The results of these analyses revealed that the majority of the variation in microbial diversity across the samples could be attributed to the field (Fig. 2 and Figs. S3–S5). Indeed, field environments explained the largest fraction of the variation in β-diversity measured by both unweighted (overall variation explained: 13.6%; $P < 5.00E-03$; 95% CI = 12.7%, 14.4%) (Fig. 2) and weighted UniFrac distance metrics (18.3%; $P < 5.00E-03$; 95% CI = 14.6%, 18.5%) (Fig. S3) after accounting for the variation present among sample types, maize inbreds, amplification batch, and sequence run. Fig. 2*A* shows clear clustering of the New York fields relative to the Midwestern fields. However, no significant differences were observed between the organically managed field located in Ithaca, NY and conventionally managed field environments. The clustering of New York field environments is visible (Fig. 2*A*) despite differences in soil composition (Fig. 3 and Fig. S6), implying that other factors, such as a shared climate, may be driving this pattern. A similar pattern was observed in the unconstrained PCoA analysis of unweighted UniFrac distances (Fig. S5*A*).

After field, the most important source of variation in the microbial β-diversity was the sample type: maize rhizosphere versus bulk soil (Fig. 2*B*, and Figs. S3*B* and S4*B*). Sample type explained a substantial proportion of the total β-diversity across the field environments surveyed for unweighted (29.6%; $P < 2.00E-04$; 95% CI = 24.9%, 31.2%) and weighted UniFrac distance metrics (46.7%; $P < 5.00E-03$; 95% CI = 44.5%, 48.8%) after controlling for field, amplification batch, and sequencing run. Although significant, the proportion of variation captured by sample type within field environments remained small in both unweighted (3.7%; $P < 5.00E-02$; 95% CI = 0.8%, 4.7%) and weighted UniFrac measures (1.6%; $P < 5.00E-02$; 95% CI = 1.0%, 1.9%).

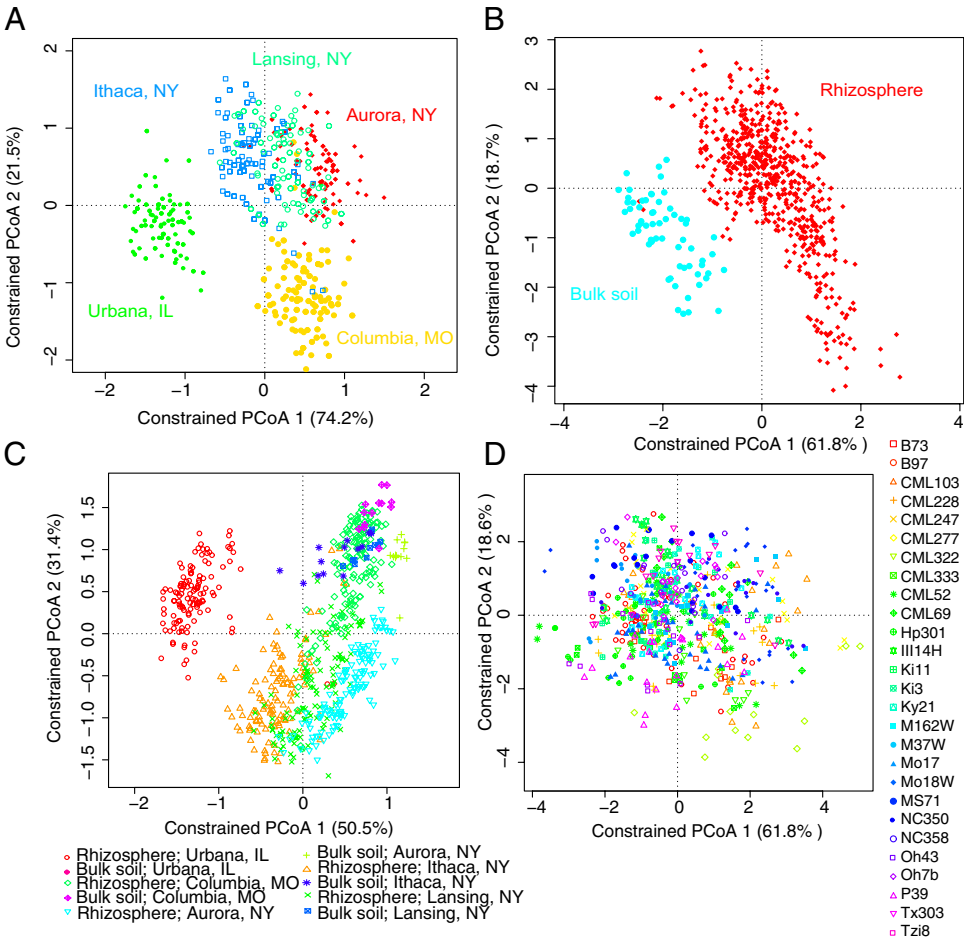

Fig. 2. Factors influencing rhizosphere and soil microbiome β-diversity. Variation in unweighted UniFrac dispersion by: (*A*) field environment (significance $P < 5.00E-03$); (*B*) bulk soil and the maize rhizosphere ($P < 5.00E-03$); (*C*) bulk soil and maize rhizosphere within all field environments ($P < 5.00E-02$); and (*D*) maize inbreds ($P < 5.00E-02$). Note that the percent variation explained by the PCs of this CAP analysis is indicated on the axes and refers to the fraction of the total variance (indicated in main text) explained by field (*A*), sample type (*B*), sample type by field (*C*), and maize inbred (*D*).

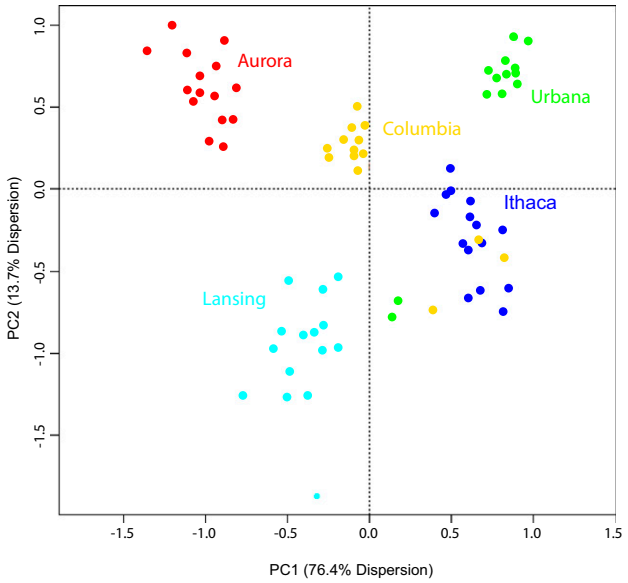

Fig. 3. Fields can be differentiated based on their physiochemical properties. Principal component analysis of the correlation matrix detailing soil relatedness among locations (Table S3). Soil samples are colored by field of origin. The percent variation explained by each PC is indicated on the axes.

Fig. 2*C* and Fig. S3*C*, which show the CAP analysis, indicate that for all fields, bulk and rhizosphere samples separate along PC2 in the same direction for each field and the PCoA results are similar (Figs. S4 and S5). In all fields, β-diversity was greater for samples derived from the maize rhizosphere than for those from bulk soil, indicating a potential for host genotype to induce these differences (Fig. 2 and Figs. S3–S5).

Variation in Specific OTU Abundances. Given the high levels of OTU richness, and limited sequencing depth (Table S2), only the most abundant OTUs and those of the highest taxonomic ranks could be quantified with a level of precision sufficient to compare them on an individual basis. The most marked contrast in the abundance of microbial taxa was observed between the maize rhizosphere and bulk soil samples. The primer set chosen to characterize the full dataset was selected because of its enrichment of classifiable sequences as well as its reduced amplification of maize plastid-related sequences. As an internal control, it is interesting to note that plastid sequences still remained a significantly enriched OTU in the rhizosphere microbiome extractions compared with bulk soil samples ($n \geq 120$, $P < 4.00\text{E-}04$).

Bacteria of the phylum Proteobacteria were enriched in the rhizosphere relative to bulk soil ($n \geq 120$, $P < 4.00\text{E-}04$). Within the Proteobacteria, taxonomic orders with confirmed enrichment in the rhizosphere microbiome relative to bulk soil included Burkholderiales ($n \geq 120$, $P < 4.00\text{E-}04$), Oceanospirillales ($n \geq 120$, $P < 4.00\text{E-}04$), and Sphingobacteriales ($n \geq 120$, $P < 4.00\text{E-}04$). Within this last group, *Sphingobium herbicidium* was especially enriched in rhizosphere samples ($n \geq 120$, $P < 4.00\text{E-}04$). In contrast to Proteobacteria, bacterial taxa of the phyla Acidobacteria ($n \geq 120$ $P < 4.00\text{E-}04$), Chloroflexi ($n \geq 120$, $P < 4.00\text{E-}04$), Planctomycetes ($n \geq 120$, $P < 4.00\text{E-}04$), and Verrucomicrobia ($n \geq 120$, $P < 4.00\text{E-}04$) were all significantly depleted within the rhizosphere compared with bulk soil samples across the surveyed field environments.

Effects of Soil Physiochemical Properties. Soil physicochemical properties of 15 randomly selected bulk soil samples were measured across each of the five field environments. These properties revealed substantial variation in the relative abundances of nitrate, phosphorous, potassium, and several other nutrients and minerals (Fig. S6 and Table S3). Significant pair-wise differences in all characteristics except moisture content were observed between field environments after Bonferroni multiple test correction. We correlated each of these characteristics with estimates of α- and β-diversity, as well as the rarefied abundances of the top 100 most common OTUs across the five field environments; however, no significant trends were observed. The raw unscaled physiochemical characteristics were plotted in scatter biplots to reveal their relationship with each other and with respect to field environment (Fig. S6). After centering and standardization to a common range, all soil nutrients were used to construct a correlation matrix detailing soil relatedness among the five environments. Principal component analysis of this metric showed that each soil presents a distinct chemical environment (Fig. 3). Of the Midwestern fields, the Columbia, MO field is most similar to the Northeastern fields in terms of physiochemical properties (Fig. 3). The most striking pattern, however, was that the rhizosphere communities of the three New York soils were quite similar, despite the fact that their chemical properties distinguished them quite strongly (Fig. 2*C* and Fig. S3*C*). This relatedness matrix of soil characteristics was compared with UniFrac estimates of β-diversity using a Mantel test; however, no significant trends were observed.

Host Genetic Influence on the Rhizosphere Microbiota. Across all replicated plots and field environments, we found that a significant proportion of variation in OTU richness within samples was explained by differences between maize inbred genotypes (19.1%; $P < 2.00\text{E-}04$; 95% CI = 18.9%, 19.2%) (Fig. S2*C*) after controlling for field and technical factors (*Materials and Methods*). The maize inbred possessing the rhizosphere with the highest OTU richness was the sweet corn Ill14h. However, an additional sweet corn, P39, possessed a lower OTU richness than 20 of the 27 maize inbreds surveyed.

We found that the OTU richness in a given maize inbred relative to the other inbreds largely depended on the particular field in which they were grown (Fig. S2*D*). In our model, the interaction term that captures this pattern (between diverse maize inbred genotypes and a given field environment) explained 48.7% of the total variation in OTU richness ($P < 2.00\text{E-}04$, 95% CI = 48.5%, 49.1%). These important results indicate that within any given field environment, genotypes affect OTU richness; however, the manner by which they do so varies considerably between field environments.

To discern if β-diversity had a heritable component across all replicated plots and field environments, we calculated the proportion of total β-diversity explained by maize inbred genotypes after accounting for field and technical factors using CAP. We found that maize genotype explained a small but significant fraction of total variation in β-diversity as measured using unweighted (5.0%; $P < 5.00\text{E-}02$; 95% CI = 4.8%, 5.6%) and weighted (7.7%; $P < 5.00\text{E-}02$; 95% CI = 7.1%, 15.4%) UniFrac distance metrics. Fig. 2*D* and Fig. S3*D* display the weak patterning of maize inbreds when PC1 and PC2 for the CAP are plotted.

We performed another canonical analysis of principal coordinates to assess the field-specific heritable component of β-diversity. This analysis revealed that a significant proportion of total variation within each field environment was explained by maize genotype using both unweighted (17.9%; $P < 5.00\text{E-}03$; 95% CI = 14.3%, 19.9%) and weighted (25.3%; $P < 5.00\text{E-}03$; 95% CI = 21.8%, 27.7%) UniFrac distance metrics, after accounting for field, maize inbred and technical factors. Figs. S7 and S8 show patterns of within-field separation by genotype; in several of the fields and for subsets of the genotypes, the replicate genotype samples cluster together. As in the case of OTU richness, the results indicate that β-diversity is weakly explained by maize inbred genotype across all field environments, and that the field in which a genotype is planted strongly influences this relationship.

As a final evaluation of the relationship between maize host genotype and bacterial diversity, we sought to determine if the diversification history of maize and the flow of total genetic diversity could explain the α- and β-diversity of the maize inbred rhizosphere microbiota. To infer the flow of total genetic diversity, we constructed a genetic relationship matrix [using identity by state (24)] between all 27 maize inbreds from the over 1.4 million polymorphisms composing the first-generation maize hapmap (12). Estimates of both α- (OTU richness) and β-diversity (weighted and unweighted UniFrac metrics) were tested for their correlation with total genetic relatedness among the lines by regression for α-diversity and a Mantel test for both β-diversity measures. Despite the significant differences noted between microbial community profiles of maize inbreds, kinship of the 27 maize inbreds was not significantly correlated with α- or β-diversity estimates. Similarly, rarefied abundances of the most common OTUs were also compared with estimates of kinship among the maize inbred rhizospheres, but no significant correlations were observed. Finally, we performed additional analyses fitting flowering time, as measured by days to pollen shed, and total plant height as covariates within our models both across and within field environments, but neither trait explained a significant fraction of variation in α- or β-diversity estimates of the rhizosphere microbiota.

Discussion

In this study we characterized the rhizosphere microbial community composition of 27 modern maize inbreds planted in randomized and replicated plots in each of five field environments at flowering time. We also profiled the microbial communities of 15 bulk soil samples in random locations within each of the five field environments to characterize bulk soil microbial community composition. Our design permitted us to test the influence of maize host genotype on its rhizosphere microbial community across the field environments. This process also allowed us to assess the degree to which these plant–microbe interactions depend upon the field environment in which they are measured. Although effects were small, we found that the maize genotype significantly influences α- and β-diversity across field environments.

Our fields were located in two distinct climatic regions in the United States, with three fields (Aurora, Ithaca, and Lansing) in the Northeastern state of New York and two in the Midwest (Columbia, MO and Urbana, IL). We observed that the microbiota of the five fields clustered grossly by geographic proximity: the three New York soils harbored more similar microbiota compared with the two Midwestern fields. However, the microbiota of the two Midwestern fields were not most similar to one another, as the Columbia, MO soil microbiota showed greater similarity to the New York microbiota rather than the Urbana, IL microbiota. If geographical distance results in isolation and diversification of microbiota, we might expect substantially different microbiota to be present in the Northeastern and Midwestern soils, but this was not the case, suggesting that the regional and climatic differences between the Midwestern and Northeastern soils are insufficient to explain the biogeographic patterns observed herein. Interestingly, we found that the organic management at our Ithaca, NY site did not result in a markedly different microbiota compared with conventional management performed in the other fields, which is consistent with previous work showing a long-term effect of cultivation, but not field management, on soil microbial diversity (25). Previous work on bacterial communities from the North and South America has shown that latitude or geographic distance did not significantly influence diversity (26); however, soil pH was later shown to be the most influential factor (27). These results indicate that environmental heterogeneity, such as pH and moisture content, as well as geographic patterns, likely interact to shape the spatial scaling of the maize rhizosphere microbiota (28, 29), and that the nature of these interactions may differ by environment.

After between-field variation, the contrast between bulk soil and rhizosphere was the next most significant source of variation in microbiota composition. We observed a reduction in community richness and greater β-diversity for the maize rhizosphere microbiota compared with the bulk soil microbiota in all five field environments. The fact that rhizosphere microbiota differ from bulk soil microbiota is well established. What our findings adds to this general view is that the subset of rhizosphere microbiota enriched across replicates of a given plant genotype within an environment can vary quite substantially across environments. We stress that even after accounting for genotypic differences and environmental differences their interaction appears highly important.

We observed a few taxa to be consistently enriched in the maize rhizosphere, for example the orders Burkholderiales, Oceanospirillales, and Sphingobacteriales of the Proteobacteria. Proteobacteria, and the order Burkholderiales in particular, have been shown previously to be enriched in the maize rhizosphere (8, 30, 31). Recent studies in Oak (32) and *A. thaliana* (10, 11) have also noted a rhizosphere (and endophyte in *Arabidopsis*) enrichment of Proteobacteria. Although in our study most bacterial taxa of lower rank were not sequenced at a high enough read depth to enable powerful comparisons, we did observe an enrichment of the genus *Sphingobium*, which is a well-known aerobic rhizosphere bacterium and has, for example, been isolated from the rhizosphere of *Populus deltoides* (33). Taken together, these results underscore the fact that Proteobacteria are adapted to the plant rhizosphere generally and across diverse plant species. This finding is not surprising, as Proteobacteria are well known to respond to labile carbon sources, and are generally considered to be r-selected, or weedy fast-growing microbiota whose populations fluctuate opportunistically (34). In contrast to the rhizosphere, the bulk soil is generally considered to be enriched in k-selected microbiota, or slower growing microbiota with more stable population sizes. In our study the bulk soil-enriched phyla included the Acidobacteria, Chloroflexi, Verrucomicrobia, and Planctomycetes, which have been previously described as soil oligotrophs (34, 35).

We found a small but significant fraction of variation in microbial diversity (both α- and β-diversity) that could be attributed to host genetics. Furthermore, we observed significant maize inbred by environment interaction that explained a substantial portion of the variation in diversity between rhizospheres. Bouffaud et al. have recently indicated that maize genotype can influence rhizosphere microbiota in 21-d-old seedlings grown in a greenhouse (9). Similarly, we have recently shown in a pyrosequencing-based study of eight *A. thaliana* inbred genotypes grown in two soil types under controlled conditions that genotype explains a similar fraction of variation in the endophyte microbiota (11). Greenhouse or well-regulated growth-chamber experiments can control variables such as climate and soil heterogeneity more tightly than field trials, and might be expected to yield a higher heritability. Field trials such as those performed here allow us to control variables, such as climate and soil heterogeneity, statistically through blocking and replication, and also enabled us to estimate environmental dependencies. Given the large proportion of microbial diversity between maize inbred genotypes that was field-specific, however, it is likely many relationships observed within the greenhouse will not be replicable in the field. Both field and greenhouse studies are needed and will benefit from greater sequencing depth, which may account in part for the low heritability estimates obtained (i.e., lack of power to detect effects).

Although we did detect a statistically significant but low level of heritability for the rhizosphere microbiota, we could not relate the kinship matrix for the maize inbreds to the microbiota diversity profiles in this study. Given such low heritability and limited effective population size, the most likely reason for our

inability to explain a substantial fraction of the heritable variance by kinship is a lack of statistical power. Nonetheless, the lack of a relationship between the maize kinship matrix and microbial diversity may suggest the microbial community is controlled by a few major alleles, rather than by many alleles of small effect located throughout the genome. It is also possible that effects on microbiota are very indirect. One such scenario could be that the microbiota diversity reflects resource distribution and the carbon source to sink relationships within the plant. For example, diverting carbohydrate to the kernels may allow for less carbon exudation from the roots. To begin addressing this possibility, we included covariates for flowering time and height in our analyses. However, neither trait had a significant effect on diversity. In the future, similar relationships with microbial diversity should be explored across many more phenotypes.

In summary, this study shows evidence of heritable variation in rhizosphere microbial community composition and considerable field-specific heritable variation. However, the following questions remain: What specific segregating maize alleles are responsible for this microbial variation? What phenotypic differences do they encode between the maize inbred rhizospheres? And finally, what specific elements of microbial diversity are being acted upon by these phenotypic differences? Larger effective maize populations and a deeper, more focused sequencing effort on the existing rhizosphere microbial diversity are necessary to characterize the symbioses that exist under natural environmental conditions. Focusing on functional groups of microbes rather than taxonomic relatedness of the microbial community may enhance future studies. In addition, surveying maize landraces and a pool of diversity capturing allelic variation that existed before breeding for adaptation to the heavily fertilized field environments of modern industrial agriculture may reveal functional alleles and symbiotic relationships not captured in this analysis.

Materials and Methods

Twenty-six maize genotypes planted in replicate plots in five fields in two regions were sampled at flowering time for rhizosphere soil microbiota. Microbial diversity was characterized by sequencing 16S rRNA gene sequences amplified by PCR from bulk DNA extractions, using the Roche 454 Titanium platform. Bulk soil samples were also profiled for chemical content. The 16S rRNA gene sequence analysis was performed with the software QIIME (20). Details for the methods stated here, as well as the statistical analyses conducted, are discussed further in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Dr. Sherry Flint-Garcia, Dr. Stephen Moose, and Dr. Margaret Smith, as well as their field assistants at the University of Missouri, University of Illinois, and Cornell University, respectively, for growth and management of the maize fields; undergraduate students Sarah Asman, Elle Glenny, and Eric Morris for their assistance in sample collection and processing; and members of the E.S.B. and R.E.L. laboratories for scientific insight and review of the manuscript. This research was supported by National Science Foundation Grants IOS-0958184 (to R.E.L.) and IOS-0958245 (to J.L.D.), and by the Howard Hughes Medical Institute-Gordon and Betty Moore Foundation. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. J.L.D. is a Howard Hughes Medical Institute-Gordon and Betty Moore Foundation Plant Science Investigator. R.E.L. is a Fellow of the David and Lucile Packard Foundation, the Arnold and Mabel Beckman Foundation, and The Hartwell Foundation.

1. Koopman MM, Fuselier DM, Hird S, Carstens BC (2010) The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities. *Appl Environ Microbiol* 76(6):1851–1860.
2. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006) The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu Rev Plant Biol* 57(1): 233–266.
3. Innerebner G, Knief C, Vorholt JA (2011) Protection of *Arabidopsis thaliana* against leaf-pathogenic *Pseudomonas syringae* by *Sphingomonas* strains in a controlled model system. *Appl Environ Microbiol* 77(10):3202–3210.
4. Ali B, Sabri AN, Ljung K, Hasnain S (2009) Auxin production by plant associated bacteria: Impact on endogenous IAA content and growth of *Triticum aestivum* L. *Lett Appl Microbiol* 48(5):542–547.
5. Castiglioni P, et al. (2008) Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. *Plant Physiol* 147(2):446–455.
6. Zhang H, et al. (2008) Soil bacteria confer plant salt tolerance by tissue-specific regulation of the sodium transporter HKT1. *Mol Plant Microbe Interact* 21(6):737–744.
7. Kuske CR, et al. (2002) Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Appl Environ Microbiol* 68(4):1854–1863.
8. Aira M, Gómez-Brandón M, Lazcano C, Bååth E, Domínguez J (2010) Plant genotype strongly modifies the structure and growth of maize rhizosphere microbial communities. *Soil Biol Biochem* 42(12):2276–2281.
9. Bouffaud M-L, et al. (2012) Is diversification history of maize influencing selection of soil bacteria by roots? *Mol Ecol* 21(1):195–206.
10. Bulgarelli D, et al. (2012) Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* 488(7409):91–95.
11. Lundberg DS, et al. (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488(7409):86–90.
12. Gore MA, et al. (2009) A first-generation haplotype map of maize. *Science* 326(5956): 1115–1117.
13. Buckler ES, et al. (2009) The genetic architecture of maize flowering time. *Science* 325(5941):714–718.
14. McMullen MD, et al. (2009) Genetic properties of the maize nested association mapping population. *Science* 325(5941):737–740.
15. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–214.
16. Chelius MK, Triplett EW (2001) The diversity of Archaea and Bacteria in association with the roots of *Zea mays* L. *Microb Ecol* 41(3):252–263.
17. Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* 72(3):1719–1728.
18. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12):8228–8235.
19. Jurasinski G, Retzer V, Beierkuhnlein C (2009) Inventory, differentiation, and proportional diversity: A consistent terminology for quantifying species diversity. *Oecologia* 159(1):15–26.
20. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
21. Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1): 1–10.
22. Chao A (1984) Non-parametric estimation of the number of classes in a population. *Scand J Stat* 11(4):265–270.
23. Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. *Ecology* 84(2):511–525.
24. Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 82(11):1420–1425.
25. Buckley DH, Schmidt TM (2001) The structure of microbial communities in soil and the lasting impact of cultivation. *Microb Ecol* 42(1):11–21.
26. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103(3):626–631.
27. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* 75(15):5111–5120.
28. Berg G, Smalla K (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol* 68(1): 1–13.
29. Green J, Bohannan BJ (2006) Spatial scaling of microbial biodiversity. *Trends Ecol Evol* 21(9):501–507.
30. Chauhan PS, Chaudhry V, Mishra S, Nautiyal CS (2011) Uncultured bacterial diversity in tropical maize (*Zea mays* L.) rhizosphere. *J Basic Microbiol* 51(1):15–32.
31. Salles JF, van Veen JA, van Elsas JD (2004) Multivariate analyses of Burkholderia species in soil: Effect of crop and land use history. *Appl Environ Microbiol* 70(7): 4012–4020.
32. Uroz S, Buee M, Murat C, Frey-Klett P, Martin F (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Env Microbiol Rep* 2(2):281–288.
33. Brown SD, et al. (2012) Twenty-one genome sequences from *Pseudomonas* species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of *Populus deltoides*. *J Bacteriol* 194(21):5991–5993.
34. Fierer N, Bradford MA, Jackson RB (2007) Toward an ecological classification of soil bacteria. *Ecology* 88(6):1354–1364.
35. Fuerst JA (1995) The planctomycetes: emerging models for microbial ecology, evolution and cell biology. *Microbiology* 141(Pt 7):1493–1506.

Supporting Information

Peiffer et al. 10.1073/pnas.1302837110

SI Materials and Methods

Maize Germplasm, Microbiome Sample Collection, and Soil Sample Analysis. Twenty-seven diverse maize inbreds, all founder genotypes of the Nested Association Mapping panel, were selected to maximize genetic dissimilarity using previously established genotypic data (1). Seeds for each of the inbreds were obtained from a uniform stand grown at Muskgrave Research Station in Aurora, NY in 2009. In 2010, these lines were hand planted in a randomized complete block design in five field environments located in three states [University of Illinois, Crop Sciences Research and Education Center near Champaign-Urbana, IL (Well-Drained Drummer silty-clay loam soil); University of Missouri, South Farm near Columbia, MO (Well-Drained Mexico silt loam soil); Cornell University, Muskgrave Research Station near Aurora, NY (Well-Drained Honeoye silt loam soil); Cornell University, Ketola Organic Research Farm near Ithaca, NY (Well-Drained Erie Channery silt loam soil); and Willet Dairy near Lansing, NY (Well-Drained Lyons silt loam soil)]. Conventional culture practices were used, including ammonium nitrate-based fertilization, weed, and pest control in all locations except Ketola Research Farm, where an organic management regime was implemented, including manure-based fertilization and no pesticide or chemical weed control. The rhizosphere microbiota of all maize inbred plots, as well as bulk soil samples, were collected at their mean pollen shed, approximately 12 wk after planting. The last significant precipitation event occurring in all field environments was at least 3 d before the date of sample collection.

Within each field environment, plants were carefully removed from the soil using a drain spade. Roots of three random plants were sampled from the middle of each plot composed of between 12 and 25 plants (varying by environment) to avoid border effects potentially attributable to increased nutrient availability in the end plant of a plot. For each plant, a root segment of ∼5 cm in length and 0.5–3 mm in diameter was collected near the base of the plant, along with any adherent soil particles. All sample collection was performed in well-drained soils. However, no efforts were taken to collect or model data detailing covariation in soil moisture content within each of the microbiome samples. Variation in moisture content was assumed independent of the randomized and replicated plot design for genotypes within an environment. However, differences in soil moisture content, and thus adhesion, were likely a contributing factor to the variation in microbial diversity observed between environments. All root-adherent soil particles were less than 0.01 mm in diameter; most of these particles were also less than 0.002 mm. Bulk soil samples across each of the fields were also taken midrange between maize plots using a soil core sampler of 4-cm diameter and 20-cm depth. All samples were chilled on ice immediately following collection and stored at −80 °C before DNA extraction.

Soil analyses (Table S3) were performed by the Cornell University Nutrient Analysis Laboratory using standard operational procedures for measures of moisture content, extractable phosphorus and nitrate (using the Morgan test method), as well as potassium, calcium, magnesium, iron, manganese, zinc, and aluminum by an inductively coupled plasma atomic emission spectrometer. Buffer pH was measured using the Modified Mehlich buffer test and organic matter was discerned by loss on ignition.

DNA Extraction and Amplicon Generation. Total genomic DNA was isolated from the maize root tip and its associated soil (∼0.25 g) using the PowerSoil High-Throughput DNA Isolation Kit (Mo Bio Laboratories). The root and its loosely associated soil were placed into a 2-mL well of a 96-well plate for bead beating. Samples were homogenized using a bead beater (BioSpec; 2 min on high; note that this procedure gently scoured the root but did not pulverize it). It is noted that this protocol allows for the introduction of a small fraction of endophytic microbial communities scoured from the root epidermis; however, this fraction is reduced compared with the entire rhizosphere microbiome sample. Fifteen samples from Columbia, MO were used for the preliminary primer testing experiments. 16S rRNA genes were amplified using four different primer sets (27F-338R: AGAGTTTGATCCTGGCTCAG–TGCTGCCTCCCGTAGGAGT; 515F-806R: GTGCCAGCMGCCGCGGTAA–GGACTACHVGGGTWTCTAAT; 804F-1392R: AGATTAGATACCCDRGTAGTC–ACGGGCGGTGTGTRC; and 926F-1392R: AAACTYAAAKGAATTGACGG–ACGGGCGGTGTGTRC) (2), including barcodes and titanium adapters. For the full study, we used the 515F-806R primer pair. The PCR primers were constructed as follow: forward primer = 454 Titanium Lib-l Primer A/5-base barcode/forward 16S primer and reverse primer = 454 Titanium Lib-l Primer B/reverse 16S primer. All PCR reactions were carried out in triplicate 50-μL reactions with 1× of Easy-A buffer, 1.25 U Easy-A Taq, 0.2 μM of forward and reverse primers, 3.5 mM $MgCl_2$, 0.2 mM of dNTPs, and about 50 ng template DNA. Thermal cycling consisted of initial denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 53 °C for 20 s, and elongation at 72 °C for 60 s. Negative control samples were treated similarly with the exclusion of template DNA; these negative controls failed to produce visible PCR products. Following PCR, DNA amplicons were purified with Ampure magnetic purification beads (Agencourt) and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Amplicons were then combined in equimolar ratios into a single tube with a final concentration of 12.5 ng/μL. Pyrosequencing was performed using Roche Titanium chemistry at the Department of Energy Joint Genome Institute.

Analysis of 16S rRNA Gene Sequence. Sequences were analyzed using the QIIME software package (Quantitative Insights into Microbial Ecology) using default parameters for each step (3). Sequences were removed if their lengths were shorter than 200 nt, their average quality score was <25, and they contained ambiguous bases, primer mismatches, homopolymer runs in excess of six bases or error in barcodes. Filtering of noisy sequences, chimera checking and operational taxonomic unit (OTU) picking was performed using the usearch series of scripts. De novo and reference-based chimera checking was performed and sequences that were characterized as chimeric by both methods were removed. More than 3.8-million quality-filtered reads were obtained for the samples, an average of 8,315 reads per sample (min = 2,225, max = 22,346). Sequences were chimera-checked and clustered into OTUs using Otupipe (4) and a minimum pair-wise identity of 97%. Each cluster was represented by its most abundant sequence. Representative OTUs sequences were then aligned to the Greengenes database (5) using the PyNAST algorithm (minimum percent identity was set at 80%) (6). A phylogenetic tree was built using FastTree (7). Taxonomy was subsequently assigned to each representative OTUs using the Greengenes database classifier with a minimum support threshold of 80% (5, 8).

Statistical Analyses. We used custom R scripts executed using R v2.13.2 (9) to calculate the percentage of classifiable reads. The

median proportion of Greengenes classifiable reads obtained from each primer set in the pilot experiment (Table S1) was calculated from 100 bootstrap samples of the surveyed microbiome extractions stratified by maize inbreds and bulk soil to maintain balance among these factors. Bootstrap sampling of microbiome extractions with replacement ensured equal representation of each inbred and bulk soil, and also provided a 95% confidence interval estimating the precision of estimates derived from the data. Given the lack of normality noted in the distributions of many populations tested, we used the function "aovp" from the R package lmPerm v1.1.2 (10) to discern variation in the proportion of classifiable reads between each primer set by permutation testing. Reported variances explained by each factor reflect the proportion of variance explained by that factor after accounting for the remaining factors and are calculated from the marginal sums of squares. The 95% confidence interval for variance explained was derived from the resulting distribution of variance estimates after fitting multiple regression models to each of the 100 bootstrap samplings of the data. A minimum of 5,000 permutations of the data were used to construct null distributions for each of the bootstrap samplings of the raw data in inferring significance. The reported significance values reflect the most conservative estimate obtained from the 100 bootstrap samplings. Significances for all pair-wise comparisons among the primer sets, soil, and maize inbreds in the pilot experiment were adjusted for multiple comparisons by Bonferroni correction.

Rarefaction was performed using QIIME to discern levels of OTU richness, Chao-1 diversity, and whole-tree phylogenetic diversity with respect to sequence depth (3). Following rarefaction, median abundances for each microbiome extraction were calculated at a level of 2,080 pyrosequence reads. Given an inability to accurately extrapolate OTU abundances beyond a microbiome extraction's maximum read depth, 2,080 reads was selected as a balance between removing microbiome extractions that did not possess this minimum and seeking to attain as many reads, and thus sensitivity, as possible in the included microbiome extractions. To address the unbalanced design resulting from removing extractions not possessing this minimum read depth, the microbiome extractions were bootstrapped for 100 samplings stratified by field environment, soil, and maize inbred. Permutation-based multiple regression analyses were performed in a manner similar to that implemented in discerning variation in the proportion of classifiable reads for partitioning variation in α-diversity, as measured by species richness among extractions. Reported variances in α-diversity explained by each biological factor, such as field environment, genotype, and genotype-within-field environment, reflect the proportion of variance explained after accounting for technical factors of amplification batch and pyrosequencing run and rarefying to the common depth of 2,080 reads. The reported 95% bootstrap confidence intervals were derived by reporting the distribution of variance explained upon sampling from the extractions.

To calculate β-diversity, unweighted and weighted UniFrac distance metrics were calculated and used to construct distance matrices using QIIME (3). Subsequently, the entries composing these matrices were bootstrapped for 100 samplings stratified by field environment, bulk soil, and maize inbred. The function "capscale" of the R package vegan v2.0.2 (11) was used in calculation of partial constrained principal coordinate analyses. The proportion of the total variance explained by each factor was calculated after conditioning on amplification batch, pyrosequencing run, and the remaining factors, and constraining variation to the factor of interest. The 95% confidence intervals for this variation explained were derived from the bootstrap samplings. Significances of factors within the model were calculated using vegan's permutation testing function "permutest" for constrained analysis of principal coordinates with 5,000 permutations (11). Comparisons of levels of within factor multivariate dispersion were performed using vegan's implementation of PERMDISP (12).

All comparisons of relative abundance of individual OTU as well as comparisons among soil characteristics were performed by permutation testing using the lmPerm package (10). Reported significance values are adjusted by Bonferroni correction. Normalization of the soil characteristics data and construction of the correlation matrix was performed using routines in the R base package (9). Estimations of the relatedness matrix among maize lines were performed using percent identity by state (12) as well as genotype data from the first-generation maize hapmap (13). Soil characteristic and maize kinship matrices were bootstrapped for 100 samplings stratified by field environment and maize inbred and performed using vegan's implementation of the Mantel test (11, 14).

1. Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178(1):539–551.
2. Kuczynski J, et al. (2012) Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* 13(1):47–58.
3. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
4. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194–2200.
5. McDonald D, et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6(3):610–618.
6. Caporaso JG, et al. (2010) PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26(2):266–267.
7. Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26(7):1641–1650.
8. Werner JJ, et al. (2012) Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *ISME J* 6(1):94–103.
9. R Development Core Team (2011) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
10. Wheeler RE (2010) lmPerm: Permutation tests for linear models. R package version 1.1-2. http://CRAN.R-project.org/package=lmPerm). Accessed August 5, 2012.
11. Oksanen J, et al. (2011) vegan: Community ecology package. R package version 2.0-2. http://CRAN.R-project.org/package=vegan). Accessed July 11, 2012.
12. Hardy OJ, Vekemans X (2002) spagedi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2(4): 618–620.
13. Gore MA, et al. (2009) A first-generation haplotype map of maize. *Science* 326(5956): 1115–1117.
14. Legendre P, Legendre L (1998) *Numerical Ecology* (Elsevier, Amsterdam), pp 853.

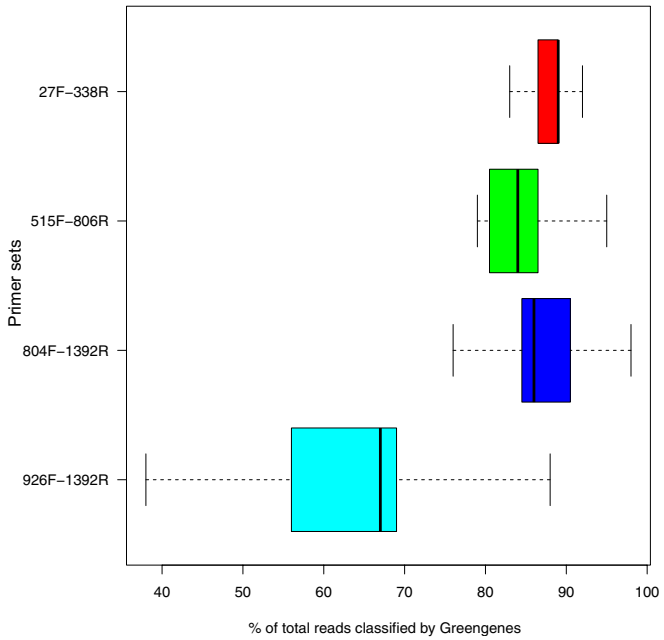

Fig. S1. Primer pair effects on proportion of classifiable sequence reads. The boxplots show the proportion of total sequence reads that matched the Greengenes database (*SI Materials and Methods*) and were obtained using the different primer sets (data are for samples shown in Fig. 1).

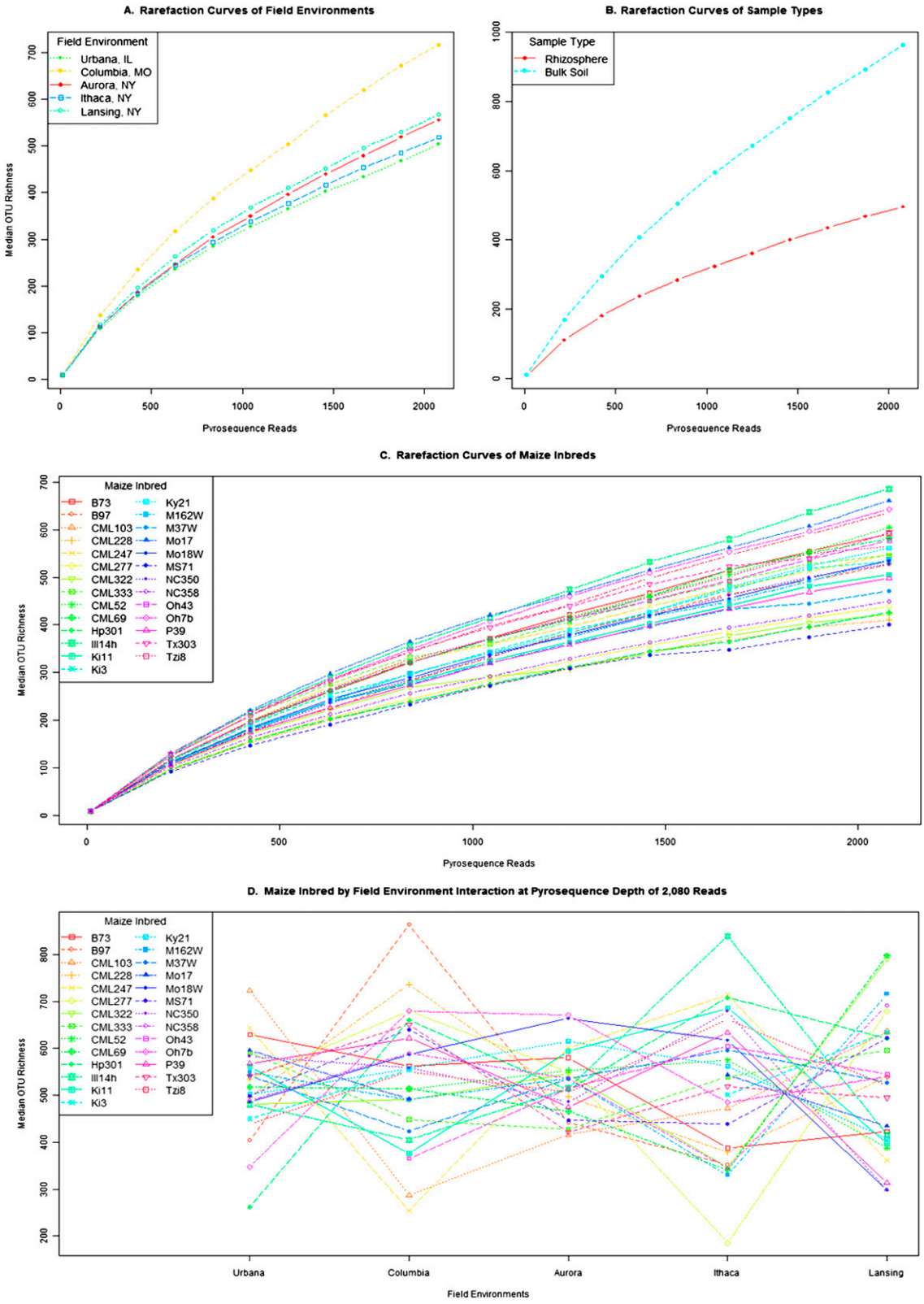

Fig. S2. Rarefaction curves for the full study. OTU counts as a function of sequence depth shown for (*A*) field environment, (*B*) sample type (bulk soil vs. the maize rhizosphere), (*C*) maize inbreds, and (*D*) maize inbreds by each soil.

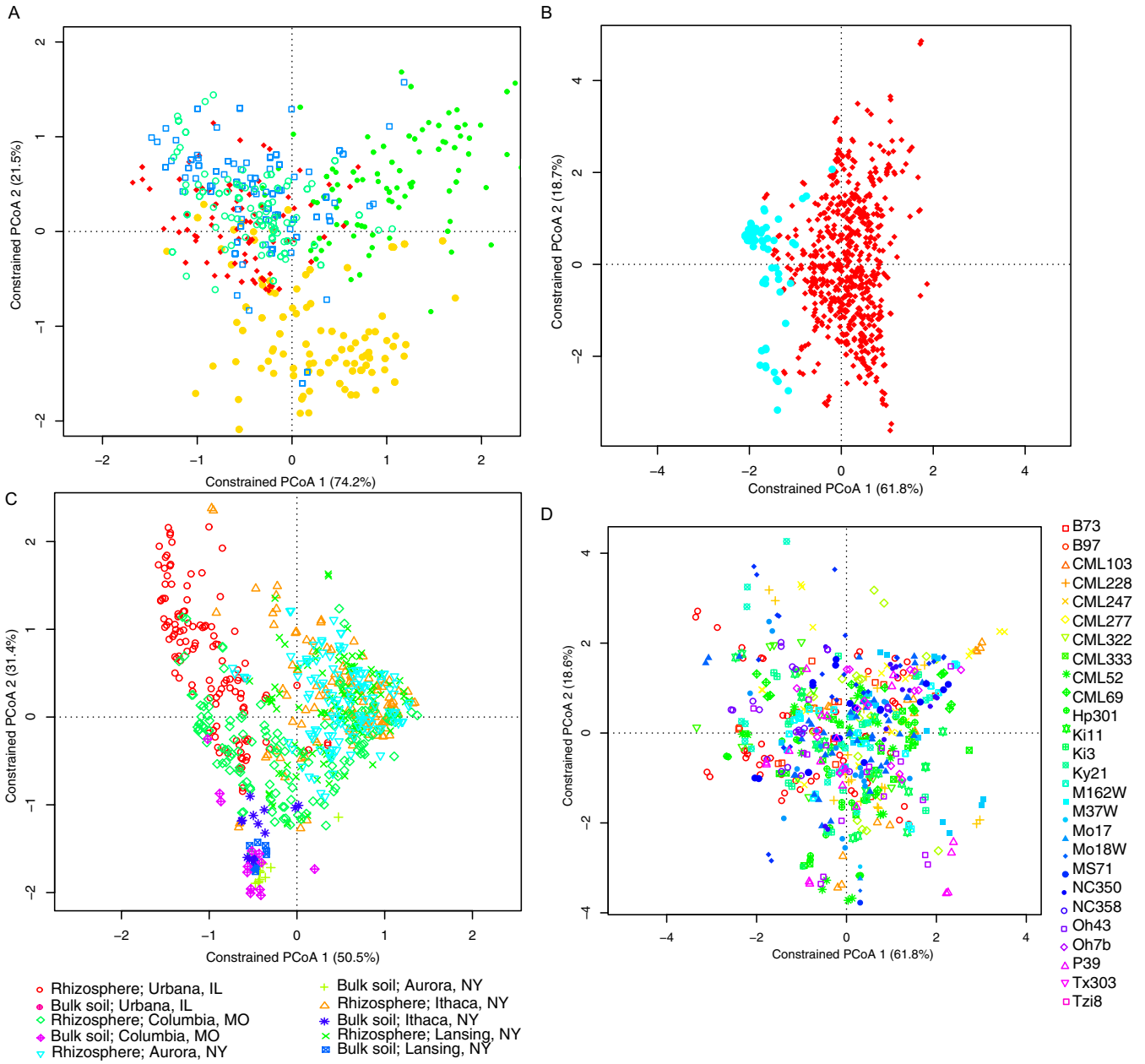

Fig. S3. Factors influencing rhizosphere and soil microbiome β-diversity. Variation in weighted UniFrac dispersion by (*A*) field environment ($P < 5.00E-03$); (*B*) sample type ($P < 5.00E-03$); (*C*) sample type within all field environments ($P < 5.00E-02$); (*D*) maize inbreds ($P < 5.00E-02$). Note that the percent variation explained by the principal coordinates (PCs) from this canonical analysis of principal coordinates (CAP) analysis is indicated on the axes and refers to the fraction of the total variance (indicated in the main text) explained by field (*A*), sample type (*B*), sample type by field (*C*), and maize inbred (*D*).

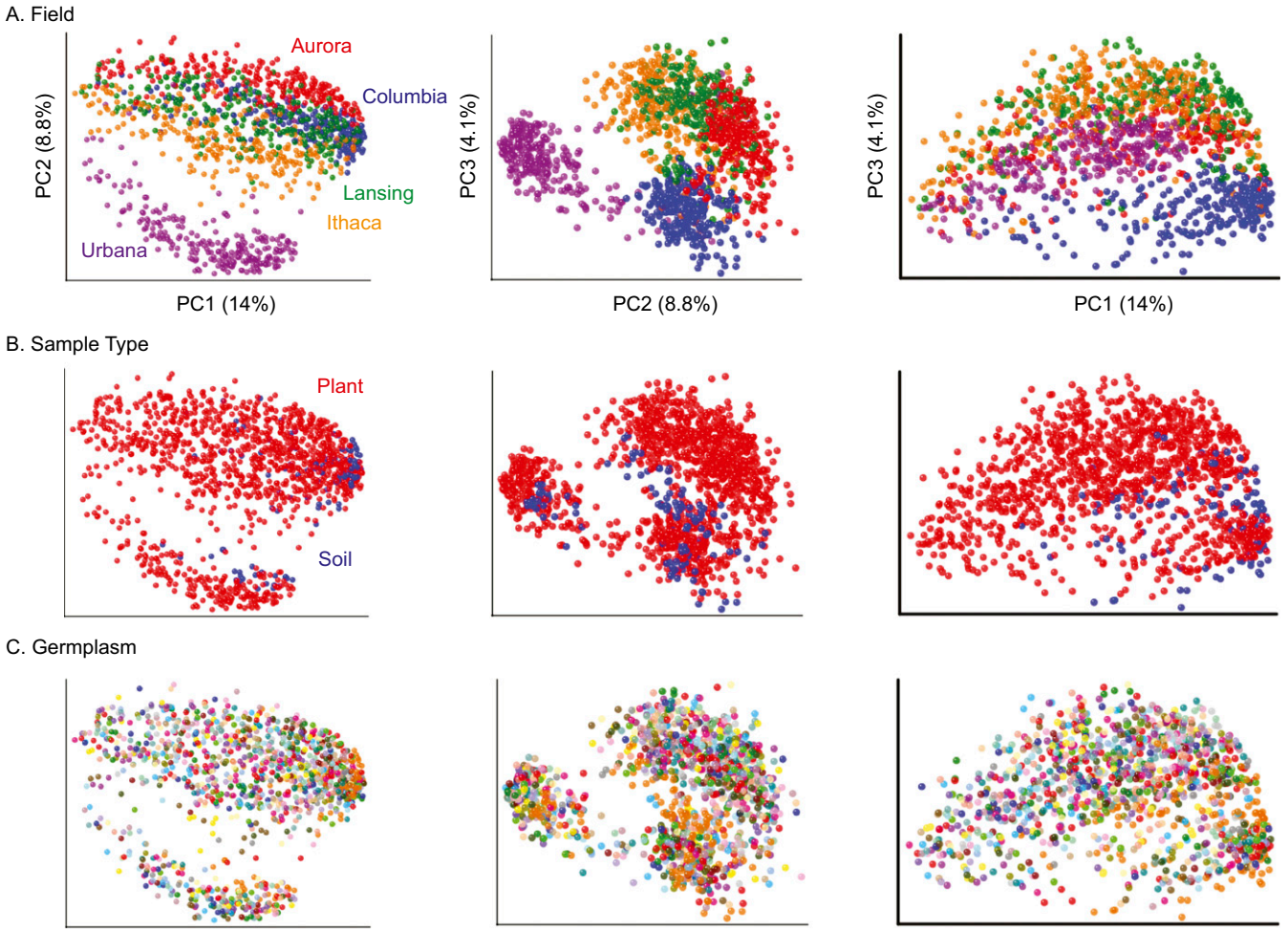

Fig. S4. Factors influencing rhizosphere and soil microbiome β-diversity. (*A*–*C*) Unconstrained principal coordinate analysis of weighted UniFrac distances. The percent variation explained by the PCs is indicated on the axes.

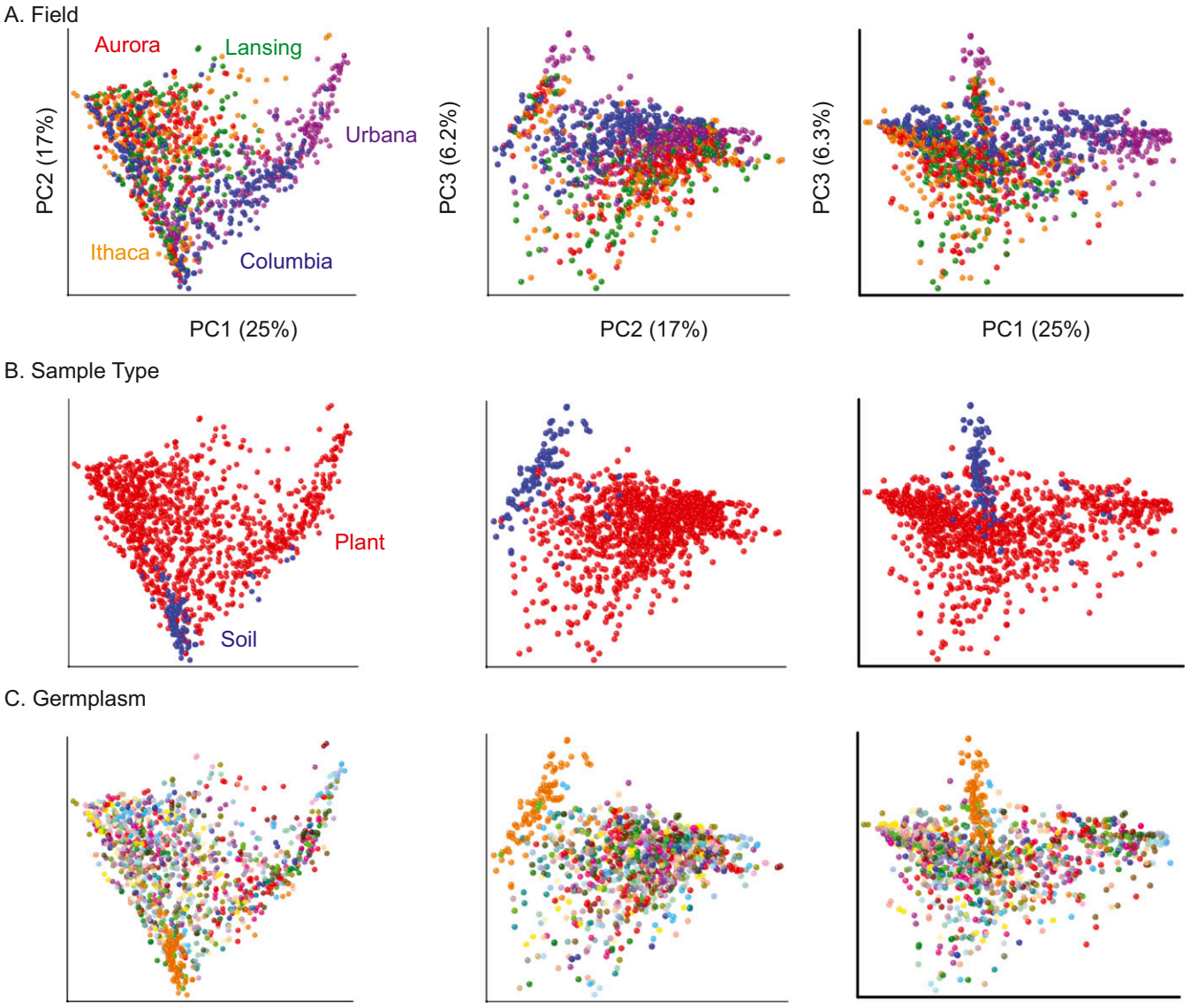

Fig. S5. Factors influencing rhizosphere and soil microbiome β-diversity. (*A*–*C*) Unconstrained principal coordinate analysis of unweighted UniFrac distances. The percent variation explained by the PCs is indicated on the axes.

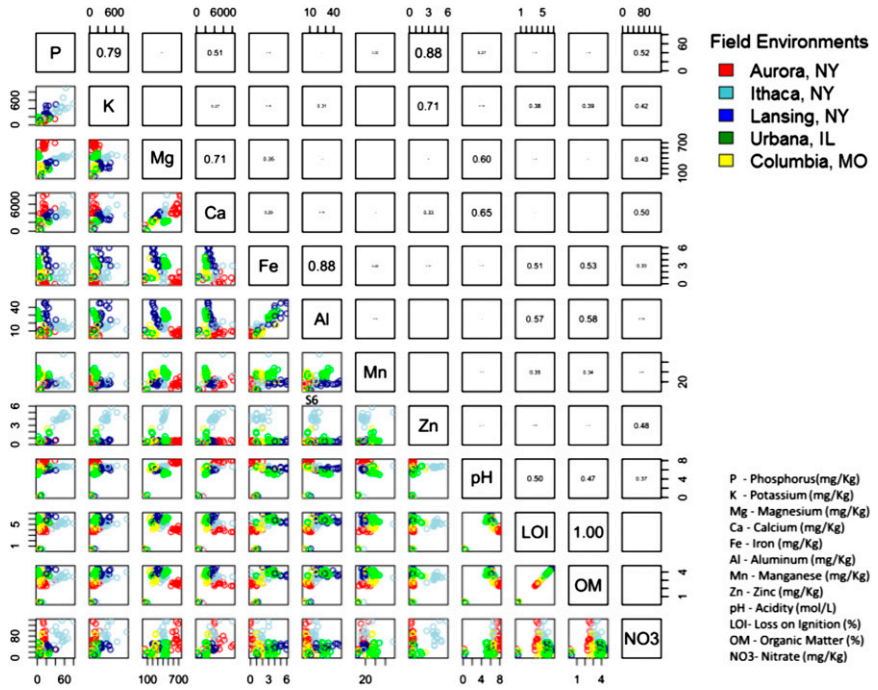

Fig. S6. Soil physiochemical properties. Scatterplots detailing relationships of physiochemical soil characteristics and field environment samples.

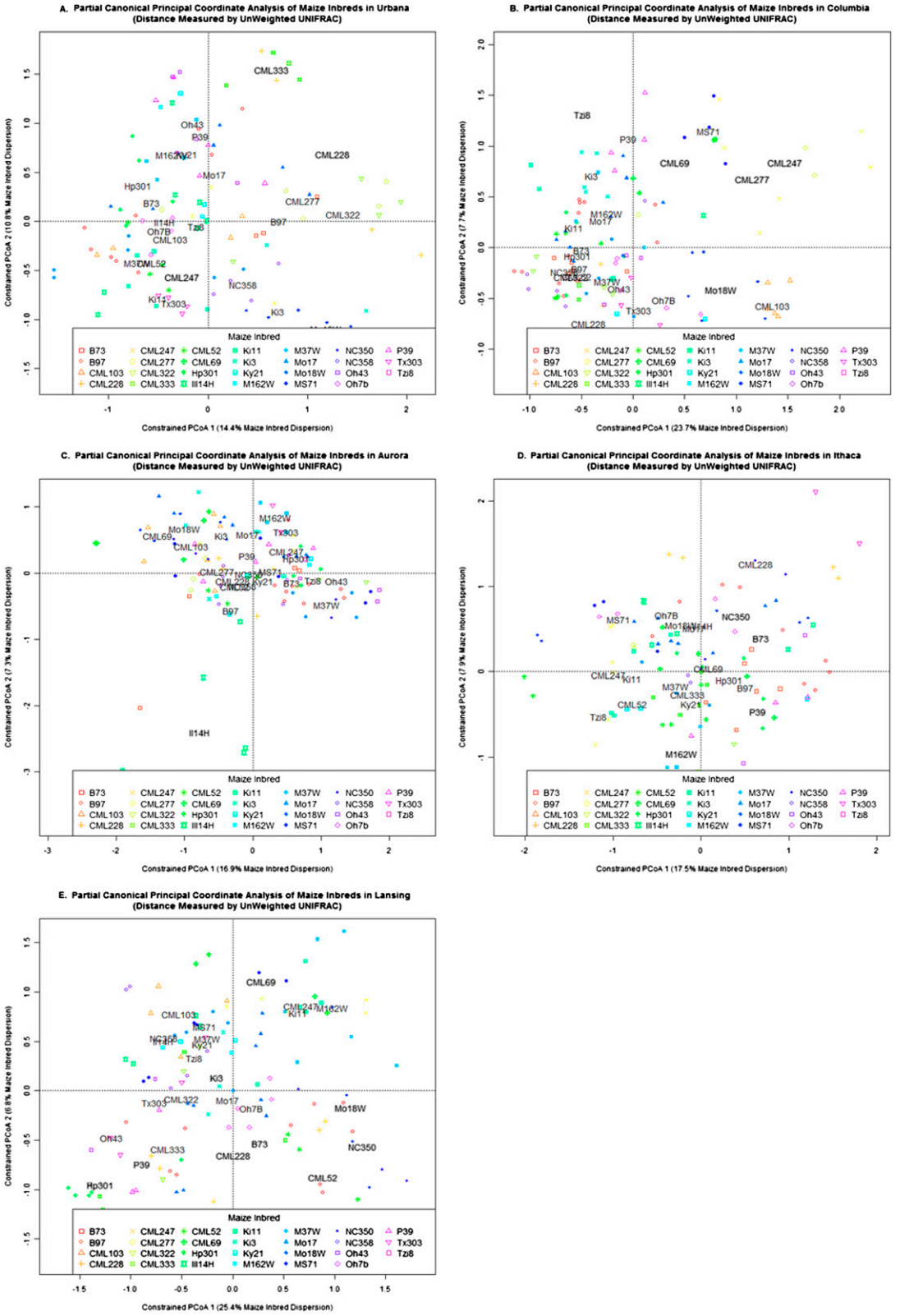

Fig. S7. Constrained principal coordinate analysis of unweighted UniFrac by field environment. Differences among maize inbreds within the different fields: (*A*) Urbana, IL; (*B*) Columbia, MO; (*C*) Aurora, NY; (*D*) Ithaca, NY; (*E*) Lansing, NY.

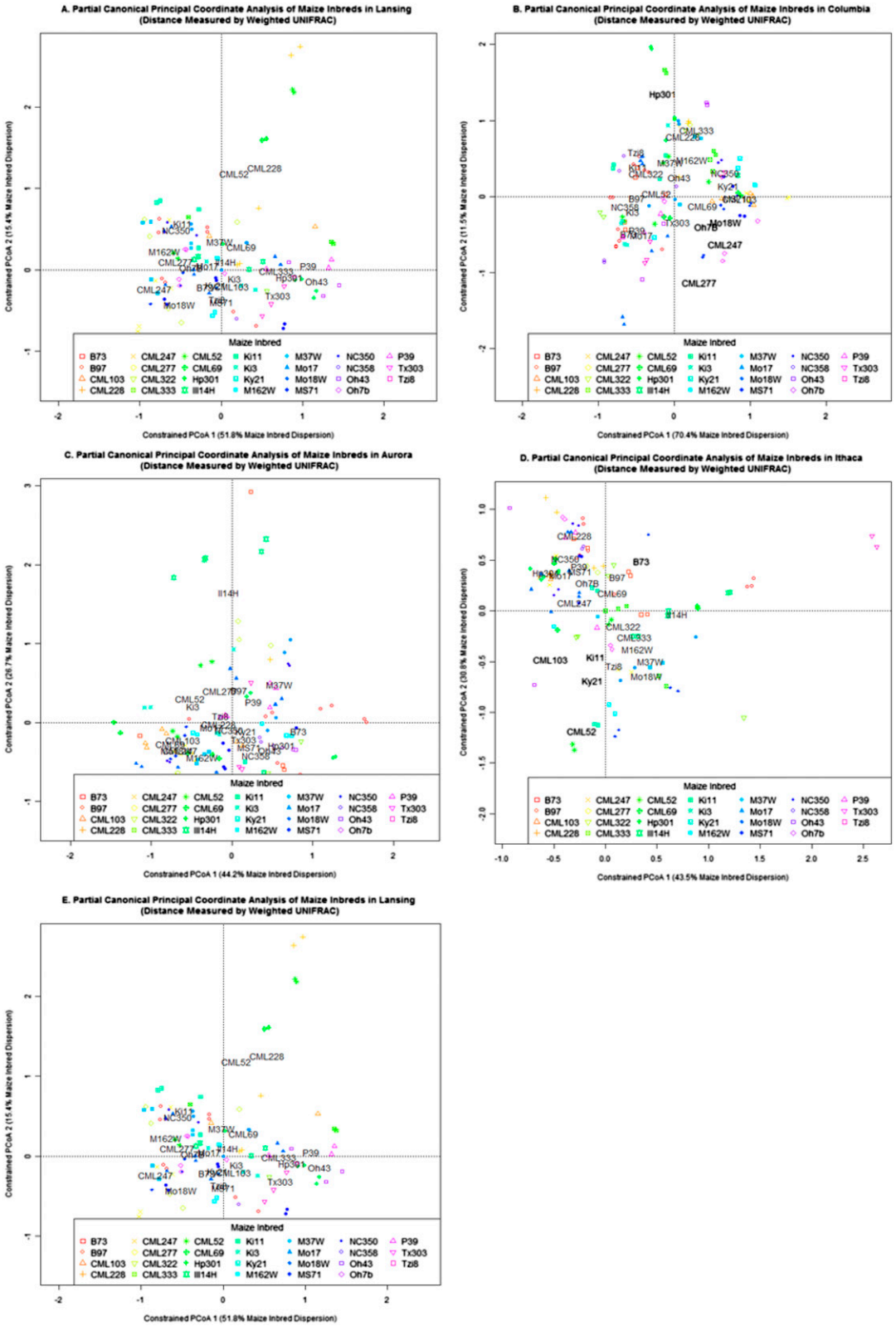

Fig. S8. Constrained principal coordinate analysis of weighted UniFrac by field environment. Differences among maize inbreds within the different fields: (*A*) Urbana, IL; (*B*) Columbia, MO; (*C*) Aurora, NY; (*D*) Ithaca, NY; (*E*) Lansing, NY.

Table S1. Summary of the 16S rRNA gene sequences obtained from the primer test experiment

Primer	Maize inbred	Sample type	Total pyrosequence reads	Pyrosequence reads (without singletons)	Greengenes classifiable reads
27F-338R (V1-V2)	B73	Rhizosphere	12,311	10,324	9,108
	B73	Rhizosphere	8,923	7,453	6,665
	B73	Rhizosphere	10,582	9,123	8,074
	B73	Rhizosphere	9,606	7,672	6,850
	Bulk soil	Bulk soil	13,859	9,804	8,183
	Bulk soil	Bulk soil	5,748	4,034	3,349
	Bulk soil	Bulk soil	12,957	10,143	9,039
	Bulk soil	Bulk soil	12,247	9,248	7,676
	Ill14h	Rhizosphere	10,501	9,035	8,532
	Ill14h	Rhizosphere	10,068	9,040	8,727
	Ill14h	Rhizosphere	12,845	10,401	9,608
	Mo17	Rhizosphere	10,411	8,154	7,077
	Mo17	Rhizosphere	11,704	8,819	7,549
	Mo17	Rhizosphere	17,217	13,941	12,419
	Mo17	Rhizosphere	9,436	7,340	6,380
515F-806R (V3-V4)	B73	Rhizosphere	17,113	15,227	12,586
	B73	Rhizosphere	37,551	34,684	29,009
	B73	Rhizosphere	31,550	29,991	25,256
	B73	Rhizosphere	29,148	24,828	21,489
	Bulk soil	Bulk soil	44,370	34,949	28,481
	Bulk soil	Bulk soil	13,809	10,880	8,693
	Bulk soil	Bulk soil	17,589	14,636	12,555
	Bulk soil	Bulk soil	16,239	13,637	10,774
	Ill14h	Rhizosphere	18,655	16,987	15,769
	Ill14h	Rhizosphere	13,499	12,648	12,012
	Ill14h	Rhizosphere	20,298	17,686	15,904
	Mo17	Rhizosphere	12,134	10,792	8,674
	Mo17	Rhizosphere	26,637	22,771	17,972
	Mo17	Rhizosphere	12,311	10,324	9,108
	Mo17	Rhizosphere	8,923	7,453	6,665
926F-1392R (V5-V8)	B73	Rhizosphere	4,791	3,640	2,034
	B73	Rhizosphere	5,220	4,148	1,944
	B73	Rhizosphere	4,552	3,856	1,451
	B73	Rhizosphere	5,693	4,471	3,085
	Bulk soil	Bulk soil	4,394	2,629	1,764
	Bulk soil	Bulk soil	4,492	3,081	2,092
	Bulk soil	Bulk soil	4,657	2,780	1,895
	Bulk soil	Bulk soil	4,493	2,956	2,124
	Ill14h	Rhizosphere	3,741	3,202	2,814
	Ill14h	Rhizosphere	4,489	4,121	3,900
	Mo17	Rhizosphere	5,386	3,883	2,520
	Mo17	Rhizosphere	4,780	3,337	1,952
	Mo17	Rhizosphere	6,732	5,403	2,861
804F-1392R (V6-V8)	B73	Rhizosphere	1,463	1,144	984
	B73	Rhizosphere	481	384	336
	B73	Rhizosphere	458	370	329
	B73	Rhizosphere	513	423	390
	Bulk soil	Bulk soil	625	429	305
	Bulk soil	Bulk soil	556	395	299
	Bulk soil	Bulk soil	1,378	1,031	879
	Bulk soil	Bulk soil	1,815	1,239	937
	Ill14h	Rhizosphere	664	583	556
	Ill14h	Rhizosphere	509	481	469
	Ill14h	Rhizosphere	405	342	320
	Mo17	Rhizosphere	426	324	277
	Mo17	Rhizosphere	572	428	358
	Mo17	Rhizosphere	558	440	380
	Mo17	Rhizosphere	455	339	290

Table S2. Summary of three measures of α-diversity

Factor	Level	Chao-1 (10 reads)	PD (10 reads)	Species richness (10 reads)	Chao-1 (838 reads)	PD (838 reads)	Species richness (838 reads)	Chao-1 (1,459 reads)	PD (1,459 reads)	Species richness (1,459 reads)	Chao-1 (2,080 reads)	PD (2080 reads)	Species richness (2,080 reads)
Type	Bulk Soil	48	3	9	1,310	47	483	1,687	62	726	1,973	73	929
	Rhizosphere	33	2	8	810	28	299	1,032	36	425	1,209	42	532
Field	Aurora	33	2	8	846	30	304	1,088	38	439	1,271	45	555
environment	Columbia	39	2	9	1,090	39	387	1,405	50	565	1,646	59	716
	Ithaca	33	2	8	716	28	294	920	35	416	1,075	41	517
	Lansing	35	2	8	887	29	318	1,110	37	451	1,297	44	567
	Urbana	32	2	8	757	25	285	958	32	402	1,134	38	502
Maize Inbred	B73	39	2	9	857	32	343	1,121	40	485	1,289	46	593
	B97	35	2	8	943	33	342	1,221	43	498	1,445	51	634
	CML103	25	2	8	643	22	232	850	29	336	937	32	399
	CML228	34	2	8	832	29	295	1,052	37	424	1,260	44	533
	CML247	27	2	8	633	22	238	816	29	343	917	33	409
	CML277	31	2	8	845	28	298	1,038	36	416	1,242	43	537
	CML322	30	2	8	781	27	274	998	35	398	1,179	41	498
	CML333	33	2	8	903	31	327	1,120	39	458	1,277	45	564
	CML52	36	2	9	848	30	320	1,062	38	460	1,317	47	605
	CML69	35	2	8	871	30	320	1,136	40	467	1,333	47	591
	Hp301	32	2	8	792	28	289	1,039	37	420	1,233	44	535
	Il14H	36	2	9	950	33	356	1,268	45	532	1,527	54	685
	Ki11	30	2	8	856	28	284	1,077	37	418	1,296	43	527
	Ki3	32	2	8	732	25	279	969	33	403	1,116	39	504
	Ky21	34	2	8	886	30	320	1,131	38	451	1,367	46	577
	M162W	32	2	8	669	25	268	771	29	341	922	34	423
	M37W	35	2	8	876	30	324	1,067	36	438	1,266	43	548
	Mo17	39	2	9	987	34	364	1,233	43	515	1,446	51	661
	Mo18W	32	2	8	618	22	238	812	29	344	955	34	426
	MS71	29	2	8	686	24	256	878	31	363	932	32	396
	NC350	34	2	8	747	26	275	966	34	396	1,104	38	470
	NC358	36	2	8	886	32	332	1,059	39	449	1,176	44	546
	Oh43	34	2	9	885	30	320	1,131	39	458	1,328	46	581
	Oh7B	37	2	8	908	32	348	1,178	43	509	1,392	50	642
	P39	27	2	8	682	24	245	897	32	357	1,023	36	437
	Tx303	31	2	8	798	29	297	1,045	37	427	1,274	46	560
	Tzi8	31	2	8	743	26	282	1,005	36	423	1,183	42	526

Chao-1 is an estimator of total species richness. It infers the abundance of unsampled diversity present within the community as a function of the abundance of singleton and doubleton species. This estimate is then added to the observed species richness. PD or phylogenetic diversity is a measure of biodiversity that incorporates phylogenetic differences between species. In this approach related individuals increase estimates of biodiversity less than unrelated individuals. Species richness is a measure of the observed number of unique OTU characterized at a given rarefaction level of reads.

Table S3. Physiochemical properties of the soils from the five field environments

Location	Sample no.	Moisture (%)	P (mg/kg)	K (mg/kg)	Mg (mg/kg)	Ca (mg/kg)	Fe (mg/kg)	Al (mg/kg)	Mn (mg/kg)	Zn (mg/kg)	pH (mg/kg)	LOI (%)	OM (%)	NO_3 (mg/kg)
Aurora, NY	1	2.21	18	95	715	5,230	2	12	18	0.3	7.8	4.0	2.6	56
	2	2.23	14	145	690	4,030	<det	5	11	0.3	7.4	4.2	2.7	126
	3	2.15	17	210	585	4,720	1	7	12	0.2	7.6	4.1	2.7	53
	4	2.12	18	235	650	3,670	<det	4	14	0.5	7.3	4.1	2.6	99
	5	2.17	17	155	660	5,940	<det	6	18	0.6	7.6	4.2	2.7	80
	6	1.86	13	80	730	7,770	1	10	30	0.7	8.0	3.6	2.3	40
	7	2.05	13	120	585	3,990	<det	7	16	0.6	7.4	4.2	2.7	96
	8	1.92	39	145	695	5,100	<det	9	15	0.6	8.0	4.7	3.0	31
	9	2.39	14	125	665	4,120	<det	6	14	0.3	7.4	4.1	2.6	68
	10	1.89	12	65	675	3,990	<det	5	16	0.5	7.5	3.9	2.5	66
	11	2.04	6	45	555	3,990	1	12	7	0.6	7.7	3.6	2.3	27
	12	2.14	12	95	675	6,230	1	9	14	0.3	7.8	3.8	2.4	65
	13	2.01	20	130	725	8,230	<det	9	23	0.5	7.9	3.6	2.3	48
	14	1.86	13	80	655	4,000	<det	9	26	0.3	7.4	3.9	2.5	134
Lansing, NY	1	2.27	36	295	390	3,410	1	19	13	3.9	6.6	4.9	3.2	47
	2	2.28	48	330	405	3,280	1	16	12	4.1	6.4	5.0	3.2	86
	3	2.02	41	430	395	2,930	3	20	16	3.7	6.4	4.7	3.1	115
	4	1.99	33	270	345	3,430	2	17	18	2.7	6.7	4.2	2.7	99
	5	2.26	54	410	430	4,560	2	17	24	4.2	6.7	5.5	3.6	75
	6	2.19	80	520	475	8,040	3	18	33	5.8	6.7	5.2	3.4	135
	7	2.19	48	305	390	3,590	<det	12	13	3.9	6.6	5.3	3.5	44
	8	2.31	54	450	440	4,980	1	15	25	4.7	6.6	5.3	3.5	138
	9	1.94	42	365	390	4,040	1	14	22	3.7	6.8	4.8	3.1	95
	10	2.30	54	420	415	4,140	2	17	12	3.9	6.6	5.6	3.7	61
	11	2.12	67	415	430	4,160	2	14	12	4.4	6.7	5.4	3.5	70
	12	2.24	63	900	425	3,600	<det	12	17	4.3	6.6	6.0	4.0	95
	13	2.12	57	660	450	5,870	3	22	76	5.0	6.9	5.1	3.3	59
	14	1.76	44	375	390	3,180	1	13	17	3.8	6.6	4.6	3.0	85
	15	2.04	61	385	395	4,340	1	15	14	4.9	6.7	5.0	3.3	65
Ithaca, NY	1	1.70	19	270	215	2,840	5	33	18	0.3	6.6	5.5	3.6	44
	2	1.75	15	270	195	2,460	5	45	16	0.5	6.4	5.3	3.5	51
	3	1.71	20	405	215	2,620	4	42	15	0.4	6.3	5.7	3.8	52
	4	1.76	25	465	240	2,660	4	39	15	0.5	6.2	6.3	4.2	43
	5	1.65	19	470	205	2,490	6	48	22	0.8	6.5	5.6	3.7	45
	6	1.68	17	230	200	2,680	6	45	19	0.4	6.1	5.5	3.6	52
	7	2.17	23	215	290	3,720	3	24	12	0.6	6.2	6.6	4.4	26
	8	1.84	21	295	245	3,000	4	31	15	0.3	6.1	5.7	3.8	36
	9	1.79	13	165	230	2,870	6	32	15	0.5	6.3	5.2	3.4	30
	10	1.76	16	265	245	3,080	5	26	24	0.4	5.7	5.5	3.6	41
	11	2.33	26	235	325	4,120	3	18	19	0.5	5.9	6.4	4.3	37
	12	2.24	41	500	310	3,270	3	24	17	0.7	5.9	6.4	4.3	30
	13	2.01	19	255	310	3,460	4	18	19	0.6	6.1	5.8	3.8	53
	14	2.05	19	170	290	3,460	4	19	15	0.4	5.8	5.6	3.7	28
	15	2.48	29	200	340	4,360	2	15	22	0.3	6.0	6.9	4.6	34
Columbia, MO	1	1.75	7	104	203	2,187	2	6	36	0.5	6.8	4.1	2.6	5
	2	1.73	9	108	217	2,322	2	5	38	0.6	6.5	4.4	2.9	37
	3	2.1	7	116	266	2,484	1	6	35	0.5	6.5	4.7	3.0	23
	4	1.96	9	104	233	2,490	1	5	27	0.5	6.9	4.5	2.9	10
	5	1.66	11	123	155	2,089	2	12	21	0.4	6.2	6.3	4.2	10
	6	2.21	7	122	270	2,462	1	6	41	0.5	6.4	4.5	2.9	39
	7	1.61	14	183	143	1,810	4	18	20	0.5	6.1	6.0	4.0	12
	8	2.07	10	131	256	2,482	1	5	32	0.5	6.5	4.5	2.9	85
	9	2.14	10	127	258	2,514	1	5	34	0.6	6.5	4.5	2.9	89
	10	2.03	8	131	247	2,482	1	6	38	0.5	6.6	4.6	3.0	53
	11	2.02	8	130	245	2,518	1	6	35	0.5	6.7	4.5	2.9	30
	12	2.13	7	115	267	2,407	1	6	35	0.4	6.6	4.5	2.9	52
	13	1.95	8	133	244	2,471	2	6	34	0.5	6.7	4.4	2.9	37
	14	2.07	9	154	254	2,468	1	6	36	0.5	6.7	4.3	2.8	49
	15	1.32	21	164	180	1,523	2	12	33	2.5	6.2	5.4	3.5	19

Table S3. Cont.

Location	Sample no.	Moisture (%)	P (mg/kg)	K (mg/kg)	Mg (mg/kg)	Ca (mg/kg)	Fe (mg/kg)	Al (mg/kg)	Mn (mg/kg)	Zn (mg/kg)	pH (mg/kg)	LOI (%)	OM (%)	NO_3 (mg/kg)
Urbana, IL	1	2.19	1	103	416	1,924	4	31	41	0.6	5.2	5.5	0.6	21
	2	2.17	1	115	414	1,892	4	33	48	0.6	4.9	5.5	0.6	55
	3	2.22	1	78	413	1,929	5	36	44	0.6	5.1	5.6	0.6	22
	4	2.39	1	167	455	2,247	4	29	30	0.7	5.2	6.3	0.7	27
	5	2.42	2	195	469	2,275	3	27	37	0.7	5.3	6.3	0.7	41
	6	2.44	1	195	500	2,332	3	27	36	0.7	5.2	6.2	0.7	54
	7	2.39	1	169	452	2,211	3	26	29	0.5	5.5	6.0	0.5	3
	8	2.3	2	170	449	2,176	4	29	39	0.7	5.2	6.0	0.7	39
	9	2.25	2	193	458	2,166	3	25	37	0.6	5.2	6.1	0.6	56
	10	2.37	2	188	439	2,117	3	28	34	1.3	5.4	6.1	1.3	13
	11	2.41	1	148	478	2,333	3	31	29	0.5	5.3	6.3	0.5	20
	12	2.37	1	119	466	2,146	4	33	41	0.6	5.3	5.9	0.6	9
	13	2.16	1	127	411	1,960	4	32	38	0.6	5.3	5.7	0.6	5
	14	1.35	27	155	196	2,092	1	7	31	1.8	6.7	5.4	1.8	21
	15	1.05	26	145	178	2,279	1	7	36	2.0	6.9	5.3	2.0	13

Al, aluminum measured in milligrams per kilogram; Ca, calcium measured in milligrams per kilogram; Fe, iron measured in milligrams per kilogram; K, potassium measured in milligrams per kilogram; LOI, percentage of mass lost on ignition; Mg, magnesium measured in milligrams per kilogram; Mn, manganese measured in milligrams per kilogram; Moisture, percentage of moisture present within sample; NO_3, nitrate measured in milligrams per kilogram; OM, percentage of organic matter; P, phosphorus measured in milligrams per kilogram; pH, acidity hydrogen ion concentration; Zn, zinc measured in milligrams per kilogram; <det, indicates quantity below detectable limit of instruments used.