scientific reports

OPEN

Compositional profiling of the rhizosphere microbiome of Canada thistle reveals consistent patterns across the United States northern Great Plains

Jed O. Eberly[1]✉, Asa Hurd[1], Dipiza Oli[2], Alan T. Dyer[2], Tim F. Seipel[3] & Patrick M. Carr[1]

Canada thistle is a pervasive perennial weed, causing challenges to agricultural and natural ecosystems globally. Although research has focused on the phenology, genetics, and control of Canada thistle, little is known about the rhizosphere microbiome and the role plant–microbe interactions play in invasion success. This study investigated the rhizosphere microbiome of Canada thistle across diverse climates, soils, and crops in the U.S. northern Great Plains. Soil and rhizosphere samples were collected and bacterial 16S and fungal ITS2 sequencing were performed to characterize the core microbiome and identify potential factors contributing to invasion success. Amplicon sequencing revealed a stable core microbiome that was detected in the Canada thistle rhizosphere across all locations. The core microbiome was dominated by the bacterial phyla Actinobacteriota and Proteobacteria and fungal phyla Ascomycota and Basidiomycota. Differential abundance analysis showed rhizosphere fungal communities were enriched in pathogen-containing genera with a 1.7-fold greater abundance of *Fusaria* and a 2.6-fold greater abundance of *Gibberella* compared to bulk soil. Predictive functional profiling showed rhizosphere communities were enriched ($p < 0.05$, FDR corrected) in plant pathogen fungal guilds which represented 19% of the fungal community. The rhizosphere microbiome was similar in composition across environments, highlighting the stable association between Canada thistle and specific microbial taxa. This study characterized the core microbiome of Canada thistle, and the findings highlight plant–microbe interactions shaping invasive behavior. These findings are important for understanding the ecological impacts of plant invasion and soil-microbe ecological processes.

Keywords Canada thistle, Rhizosphere microbiome, Invasive species, Weeds

Canada thistle (*Cirsium arvense* (L.) Scop.) is a ubiquitous perennial that is considered one of the most problematic invasive weed species in agricultural and natural systems[1]. Canada thistle can cause significant yield and economic loss in annual cropping systems and pastures[2]. It spreads both sexually via seeds and vegetatively by creeping roots and these life history traits coupled with high genetic diversity have hindered control efforts[3]. Control methods include tillage, herbicides, perennial crops, biocontrol agents, or combinations of these methods[4]. Despite over 100 years of research, control of Canada thistle remains challenging[1].

Previous work identified variations in Canada thistle based on morphology and phenology which are due to genotypic variation and contribute to management challenges[5]. Canada thistle differs markedly in response to herbicides such as 2,4-dichlorophenoxyacetic (2,4-D), 3-amino-1,2,4-triazole (amitrole), and 3,6-dichloro-2-methoxybenzoic (dicamba)[5,6]. Canada thistle genotypes also differ in vegetative traits and response to control methods such as defoliation[3]. The high genetic diversity of Canada thistle contributes to its success as an invasive species and may explain the limited success of control.

In addition to above-ground phenotypic traits, interactions between Canada thistle and soil microorganisms may contribute to invasion success. Microbiota inhabiting the rhizosphere play an important role in plant survival

[1]Central Agricultural Research Center, Montana State University, Moccasin, MT, USA. [2]Department of Plant Science and Plant Pathology, Montana State University, Bozeman, MT, USA. [3]Department of Land Resources and Environmental Science, Montana State University, Bozeman, MT, USA. ✉email: jed.eberly@montana.edu

by facilitating nutrient uptake and enhancing plant tolerance to biotic and abiotic stressors[7]. Plants selectively attract microorganisms through root exudates to enhance survival, and there is growing evidence that invasive plants alter microbial assemblages and functioning to promote successful invasion[8]. Invasive plants, such as Canada thistle, are thought to alter biotic and ecological functioning in the soil which threatens the natural balance of ecosystems[3,9].

Numerous interactions between plants and microbes have been implicated in invasive weed species establishment. The process of invasion is mediated by complex interactions at both the individual and community level[10,11]. Microorganisms can enhance host plant competition directly through increased nutrient availability and indirectly through decreased host pathogen pressure, enhanced growth of other pathogens, and loss of specialized microbes associated with native plants[10,12]. Recent studies have identified specific factors that contribute to invasive plant establishment. The rhizosphere of *Mikania micrantha*, a noxious invasive creeping perennial, is enriched in microorganisms involved in phosphate solubilization and pathogen suppression, which may enhance the competitive ability of this weed[13]. At an ecosystem scale, invasive grass dominance over native forbs has been correlated to changes in the rhizosphere bacterial families Methylophilaceae, Fibrobacteraceae, Clostridiaceae_1, Burkholderiaceae, Rhodocyclaceae, and Veillonellaceae[12]. At a global scale, studies of *Carpobrotus edul*is (L.) N. E. Br, an invasive succulent species, have identified significant relationships between the rhizosphere community composition linked to nutrient cycling and invasiveness[10]. To date, the plant–microbe interactions contributing to Canada thistle invasion success remain unknown.

Recent research has explored the functional diversity of rhizosphere microbial communities to better understand the mechanisms that enable successful invasion[11]. Metagenomic sequencing has increased understanding of microbial community functions associated with nutrient cycling that may provide a competitive advantage to invasive plant species[13]. Community function can also be inferred through computational approaches that approximate the functional potential from amplicon sequencing. This approach has been implemented through tools such as PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) which performs marker gene metagenome inference based on the bacterial 16S rRNA gene[14]. Similar approaches have been developed for fungal functional inference utilizing tools such as FUNGuild which uses taxonomic assignments to determine the ecological guild, or functional group, based on resource utilization[15]. These tools have enabled the interpretation of potential functions within the context of microbial community composition which aids understanding of the functional basis of invasion success[11].

Despite the economic and agronomic importance of Canada thistle and extensive research into methods of control, virtually no reports to date have characterized the rhizosphere microbiome or assessed the contribution of rhizosphere microorganisms to the invasiveness or competitive ability of this perennial weed species. The objective of this work was to characterize the rhizosphere microbiome of Canada thistle and determine the variability of the rhizosphere community across different climates and soils in the U.S. northern Great Plains (NGP). To our knowledge, this work is the first description of the rhizosphere microbiome and identification of a core microbiome in Canada thistle across the U.S. NGP.

Results

A total of 8,386,113 16S and 10,307,911 ITS raw sequence reads were generated from 163 rhizosphere and bulk soil samples. Subsequent quality filtering and ASV assignment resulted in a total of 16,779 bacterial and 1,379 Fungal ASVs. Rarefaction curves confirmed that the sequencing efforts were sufficient to achieve richness asymptotes (see Supplementary Fig. S1 online). Canada thistle rhizosphere bacterial community diversity determined by coverage, a measure of alpha diversity that standardizes for differences in sample size, and Hill-Simpson (inverse Simpson's), which is sensitive to differences in abundance of common species[16], was significantly decreased ($p < 0.001$ and $p < 0.01$, respectively) relative to bulk soil (Fig. 1a, b). Similar results were observed with the rhizosphere fungal community where richness and dominance were significantly reduced ($p < 0.001$) relative to bulk soils (Fig. 1c, d).

Rhizosphere bacterial communities in all locations were predominated by Actinobacteriota and Proteobacteria (Fig. 2a). Ascomycota followed by Basidiomycota were the most prevalent fungal phyla across all locations (Fig. 2b). Pseudomonadaceae, Ncardioidaceae, Chitinophagaceae, and Enterobacteriaceae were among the most abundant bacterial families (Fig. 2c) while Nectriaceae, Piskurozymaceae, and Mortierellaceae were among the most abundant fungal families (Fig. 2d). The NMDS ordination plot of the Aitchison distance and PERMANOVA results showed that the rhizosphere bacterial community composition was significantly different among locations in annual cropping systems ($F = 1.364$, $p = 0.004$) but not in perennial cropping systems ($F = 0.838$, $p = 0.999$). PERMANOVA results of the fungal community also indicated significant differences among locations ($F = 4.898$, $p < 0.001$). The NMDS ordination plot suggested these differences were likely due to distinct clustering that was observed for the Big Timber and Highwood, MT locations (Fig. 3).

Abundance-occupancy distributions of the microbial communities showed that most taxa were present in > 80% of the samples and taxa that were unique to a single location were relatively rare. A total of 21 and 24 bacterial taxa were detected in annual and perennial cropping systems respectively, and 15 fungal taxa were consistently detected in the Canada thistle rhizosphere across locations (Fig. 4d–f). Most bacterial taxa within the core microbiome belonged to the phyla Actinobacteriota and Proteobacteria (Fig. 4d, e) while most fungal taxa belonged to Ascomycota (Fig. 4f).

The neutral model fit of the abundance-occupancy distribution is shown by the solid line, with the dashed lines representing the 95% confidence intervals of the model (Fig. 4a–c). Many taxa had frequencies either above or below the 95% confidence intervals of the model.

A comparison of differentially enriched bacteria and fungi between rhizosphere and soil communities identified 226 bacterial and 17 fungal genera that were significantly ($p < 0.05$, FDR corrected) enriched. The top 20

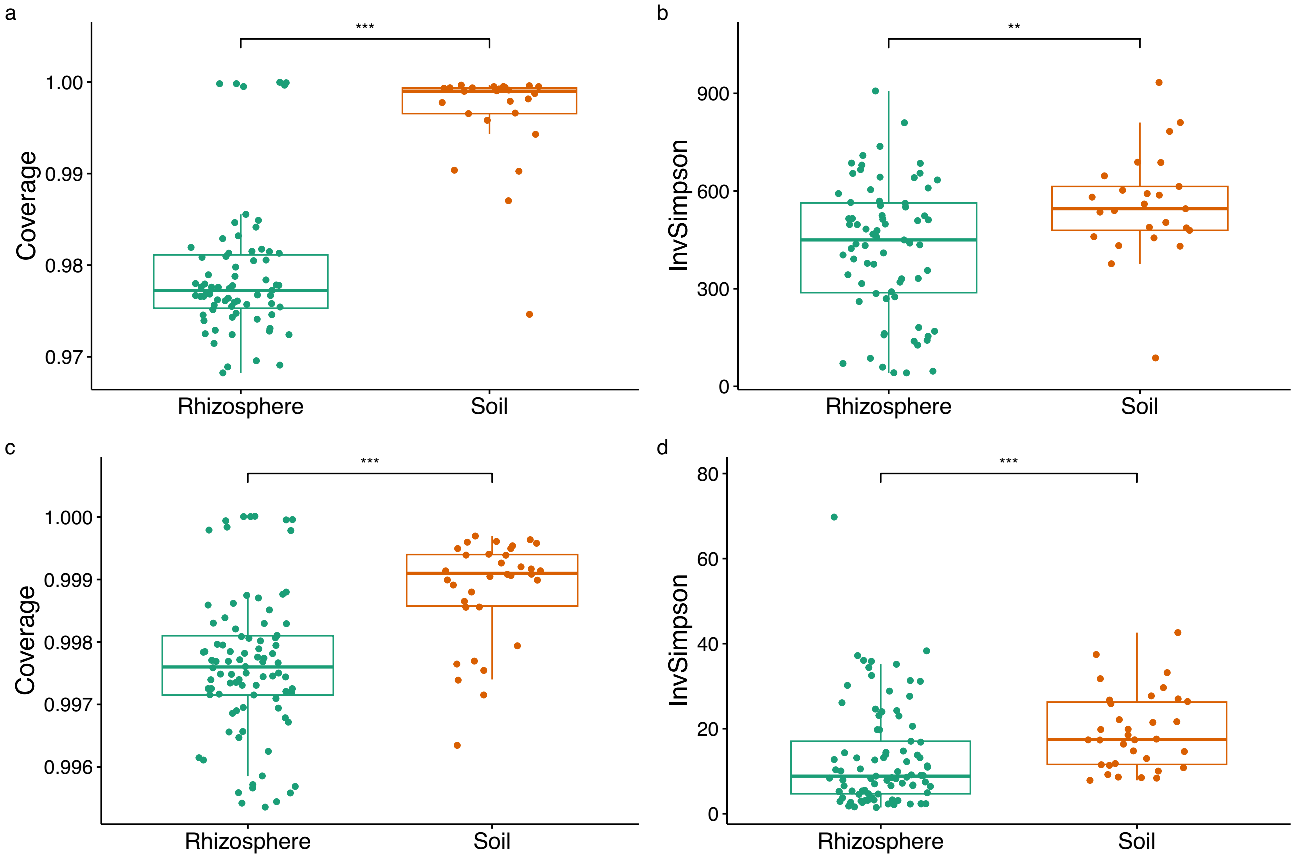

Figure 1. Bacterial and fungal alpha diversity comparison between Canada thistle rhizosphere and bulk soil across locations. Metrics used for microbial community diversity were coverage and species dominance (inverse Simpson's). (**a**) Bacterial coverage, (**b**) bacterial Simpson's dominance, (**c**) Fungal coverage, and (**d**) fungal Simpson's dominance.

most abundant bacterial genera and all 17 significant fungal genera were plotted. A total of 2 and 13 bacterial genera were enriched in the rhizosphere in annual and perennial cropping systems respectively, while 13 fungal genera were significantly enriched in the rhizosphere (Fig. 5).

Canada thistle rhizosphere communities contained a 1.7-fold greater abundance of Fusaria in the rhizosphere when compared to bulk soil (Fig. 5b). The genus *Gibberella* alone had a 2.6-fold increase in relative abundance (Fig. 5b). Bulk soils were significantly ($p<0.01$, FDR corrected) enriched in *Mortierella* and *Mycosphaerella* which represented around 5% relative abundance at the genus level. The genera *Mortierella* contains many saprotrophic and putative plant growth-promoting species[17,18]. This was consistent with the fungal functional analysis which showed bulk soils were significantly enriched in saprotrophs (Fig. 6).

Predictive functional profiling

Predictive functional profiling with PICRUSt2 using the MetaCyc database revealed 453 pathways (see Supplementary Table S1 online) predicted from 16S rRNA gene sequencing for the rhizosphere bacterial communities of Canada thistle and bulk soil. Ordination plots showed that the first two principal component axes explained 73% of the variability and predicted bacterial functions for rhizosphere and soil formed distinct clusters (Fig. 7a). In contrast, the first two axes of the PCA ordination plot of fungal trophic modes only explained 34.8% of the variation and no distinct clustering was observed between fungal rhizosphere and soil guilds (Fig. 7b).

Differential abundance analysis of the predicted pathways with Aldex2 indicated the rhizosphere was significantly ($p<0.001$, FDR corrected) enriched in multiple biosynthetic pathways including Fatty acid and lipid degradation, carboxylate degradation, cofactor and vitamin biosynthesis (Fig. 6a, b). Differential abundance analysis of predicted fungal ecological guilds showed that fungi associated with plant pathogens (19% relative abundance) and pathotrophs (22% relative abundance), including *Mycosphaerella tassiana*, *Gibberella tricincta*, *Gibberella intricans*, *Fusarium oxysporum*, and *Fusarium concentricum* were significantly ($p<0.001$, FDR corrected) enriched in the thistle rhizosphere compared to bulk soil (Fig. 6b).

Discussion

The composition of rhizosphere microbial communities is influenced by environmental factors, such as soil physical and chemical characteristics, climate conditions, and land use history[19,20]. Despite variation in these factors, plants consistently recruit specific taxa, collectively known as the core microbiome, that are important

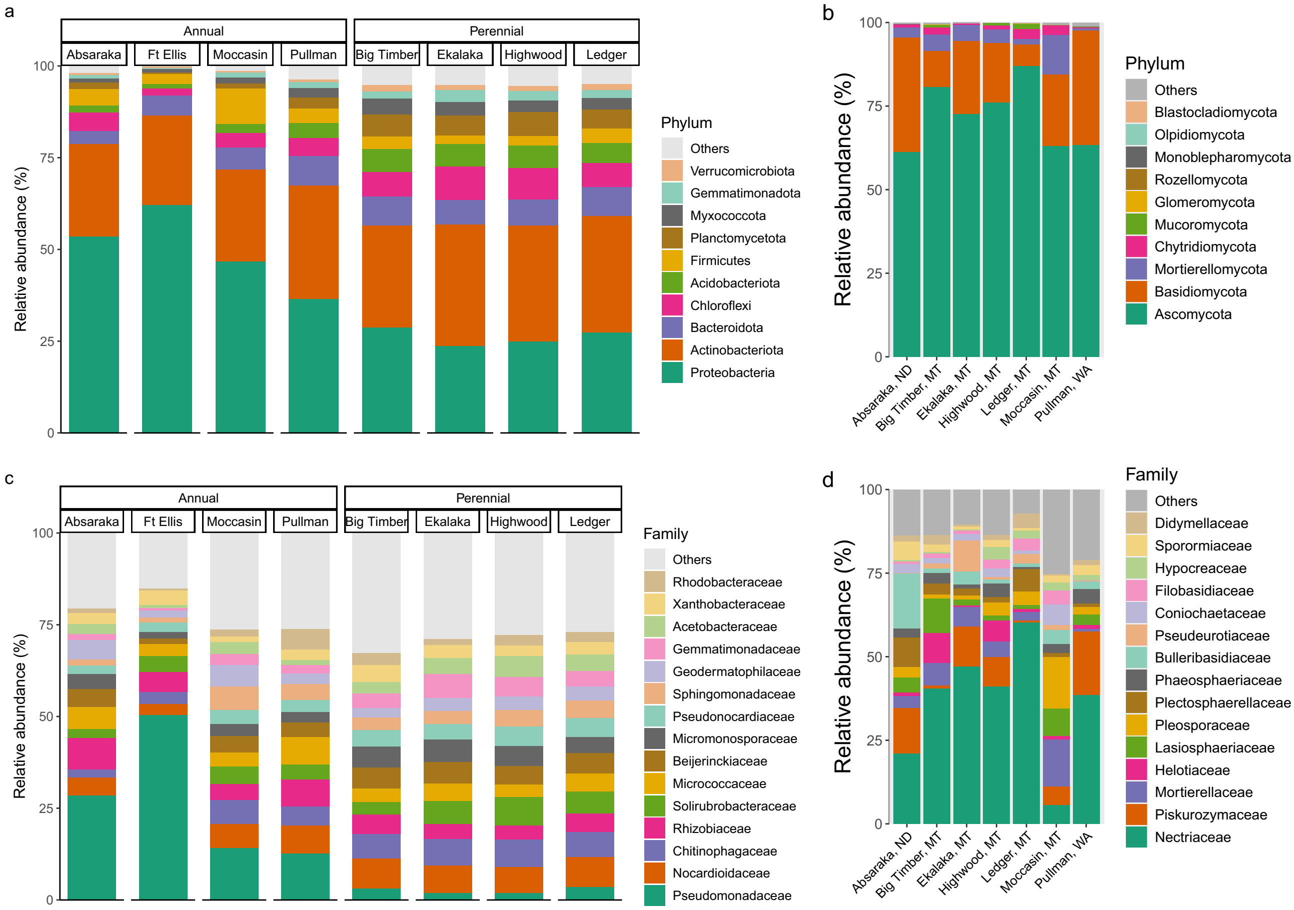

Figure 2. Relative abundance of rhizosphere bacterial phyla (**a**) and fungal phyla (**b**) and bacterial families (**c**) and fungal families (**d**) across locations. Bacterial taxa are faceted by cropping system (annual or perennial) where thistle rhizosphere samples were collected.

for plant survival[21]. Taxa that were unique to a location were relatively rare in the current study which indicates that Canada thistle has a stable core microbiome across environments. This consistently occurred across different annual cropping systems and perennial forages, in contrast to the variation that was observed in bulk soil community composition across environments. Few taxa were shared between only a few locations sites but when grouped by state, MT and WA had over twice as many bacterial taxa in common compared with either of those states and ND (see Supplementary Fig. S2 online).

The neutral model fit of the abundance-occupancy distribution showed many taxa had frequencies either above or below the 95% confidence interval of the model which suggests the rhizosphere community was more strongly influenced by deterministic factors, such as the rhizosphere environment, than by stochastic factors, like drift and dispersal[22]. Our results revealed that Canada thistle can recruit a common and consistent core rhizosphere microbiome across different environments. This contrasts with other studies that found rhizosphere communities associated with other weed species were strongly driven by regional factors such as latitude, rainfall, soil type, and agricultural practices[10,23] and suggests that Canada thistle roots exert a strong influence on the local soil microbial population which shapes the rhizosphere community and likely contributes to invasion success.

Successful establishment of an invasive plant species requires adaptation to the physical environment, enhanced nutrient acquisition, and decreased host pathogen pressure relative to native plant species[10,13]. Successful invasions can be mediated by indirect mechanisms such as enhanced growth of pathogens of native plants[10]. This apparent competition is the basis of the accumulation of pathogens hypothesis which posits that an invasive plant can accumulate local soil pathogens that can negatively affect native plant species and enhance the dominance of the invader[9,24]. This has been observed with other invasive plant species such as *Carpobrotus edulis* (L.) L. Bolus[11] and *Chromolaena odorata* (L.)[25] where the rhizosphere of the invader was significantly enriched relative to native plants. We have also found in separate studies in similar dryland environments, the rhizosphere of barley is not enriched in *Gibberella* or *Fusarium* (unpublished data). Results of our work showed fungal guilds associated with plant pathogens were significantly enriched in the rhizosphere relative to bulk soils and represented 19% of the Canada thistle rhizosphere community. This functional enrichment in pathogens was due to increased abundance of *Gibberella* and *Fusarium*. Functional annotation of fungal guilds must be interpreted carefully given the limitations of short read sequencing and available reference genomes[26,27], thus taxa associated with predicted guilds were compared to taxonomic assignments (Fig. 5).

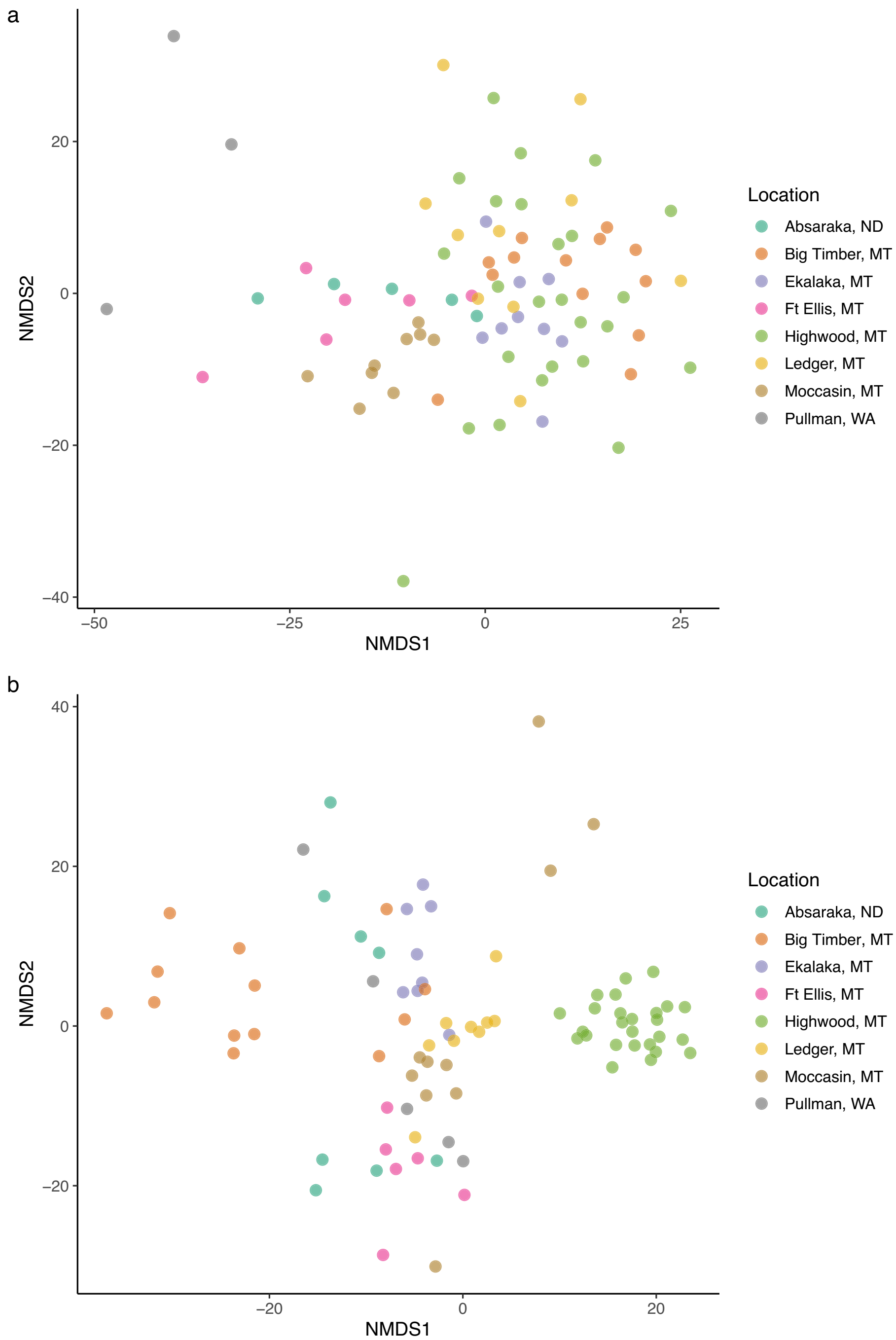

Figure 3. NMDS ordination plot of the Aitchison dissimilarity matrix at the genus level with differences in microbial community structure colored by location. (**a**) Bacterial community, (**b**) fungal community.

Nomenclature of some taxa remains inconsistent, and teleomorphs of some *Fusarium* species are listed as *Gibberella* in most classification schemes[28,29]. *Gibberella* is one of three teleomorphic genera that comprise the anamorphic genus *Fusaria*[30]. It represents *Fusarium* clades V, VI, and VII7 which include many of the most notorious *Fusarium spp.* including *F. oxysporum, F. graminearum*, and *F. sporotrichoides*[31,32]. As there are both accomplished generalists as well as specialist pathogens among them, distinguishing *Gibberella* at the species level will be important for understanding their potential role in conferring a competitive advantage to Canada thistle.

The genus *Mycosphaerella* is historically one of the largest genera within the Ascomycota. While ITS sequencing data has previously suggested the genus is monophylytic, additional sequence comparisons support a *Mycosphaerella* complex comprising at least four teleomorphic families[33,34]. Reliance on ITS sequences suggests our annotations likely reflect the more historic *Mycosphaerella* designations. While there are saprophytes within the *Mycosphaerella*, the genus is heavily populated with foliar pathogens, the most locally relevant being *M.*

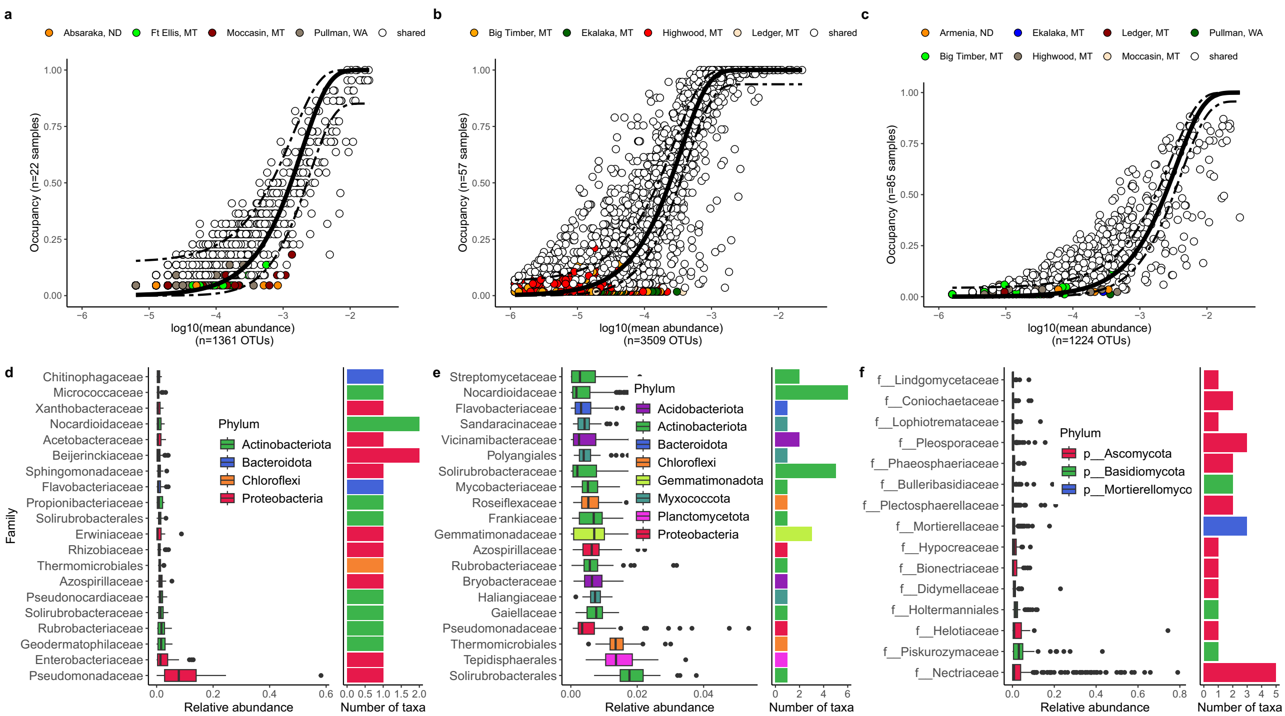

Figure 4. A high number of microbes were consistently detected in the Canada thistle rhizosphere across the US northern Great Plains. Abundance-occupancy distributions were used to identify core rhizosphere microbiome members for bacteria in annual (**a**) and perennial (**b**) cropping systems and fungi (**c**). Taxa unique to a location are colored by location and taxa shared across locations are white. The solid line represents the fit of the neutral model, and the dashed line is 95% confidence around the model prediction. Taxa detected in at least 80% of samples were considered members of the core. Relative abundance of core taxa are shown in box plots, color coded by phylum, and grouped by family. (**d**) rhizosphere bacteria from thistle sampled in annual cropping systems, (**e**) rhizosphere bacteria from thistle sampled in perennial cropping systems, and (**f**) fungi. Bar plots show the number of taxa within each family in the core microbiome.

pinodes and *M. graminicola*, important foliar pathogens of peas (*Pisum sativum* L.).and wheat (*Triticum aestivum* L.), respectively[35,36]. As such, the greater abundance of *Mycosphaerelle* in bulk soil samples is likely the result of foliar residues being incorporated into soil samples, and enrichment of this genera perhaps reflects local foliar disease dynamics.

Our results showed that *Nocardioides* and *Marmoicola*, both belonging to the family Nocardioidaceae, were among the significantly enriched bacterial genera in the Canada thistle rhizosphere. Members of the genus *Nocardioides* are metabolically diverse and capable of degrading nitrophenols, triazine containing herbicides, and the mycotoxin deoxynivalenol[37,38]. Deoxynivalenol is produced by *Fusarium* and other pathogens and plays an important role in fungal development and colonization[39]. Members of the fungal genus Fusarium were also enriched which suggests the increased rhizosphere abundance of *Nocardioides* could be in response to the greater abundance of Fusarium. Other work has reported an increased abundance of *Nocardioides* in soils enriched with vermicompost where they were hypothesized to play a role in inhibiting the tomato (*Solanum lycopersicum* L) wilt pathogen, *Fusarium oxysporum* f. sp. lycopersici[40]. Another study reported a strain belonging to the genus *Nocardioides* was effective as a biocontrol agent for tomato fungal pathogens by inhibition of spore germination[41]. Our work is consistent with these findings and suggests enlisting bacteria to suppress fungal pathogens may be one mechanism by which Canada thistle gains a competitive advantage. More work is needed to test this hypothesis.

Functional profiling of the fungal community identified fungal guilds associated with plant pathogens and pathotrophs that were significantly more abundant in the Canada thistle rhizosphere community compared to bulk soils. These findings support the fungal community profiling where the fungal genera *Fusarium* and *Gibberella* were significantly more abundant. A higher abundance of pathotrophs and saprotrophs has been reported in other invasive plants[11]. The greater abundance of plant pathogens could be advantageous for the invasive plants since these pathogens are likely more harmful to native and crop plants and may inhibit their growth thus providing a competitive advantage to Canada thistle.

Conclusion

Our work showed that the invasive perennial weed *Cirsium arvense* (L.) Scop. has a distinct and consistent rhizosphere microbial community across diverse climates and soils. Specific taxa within this community likely play a role in invasion success. To our knowledge, this is the first work characterizing the rhizosphere microbiome of Canada thistle. These findings are important in the context of Canada thistle control. Specifically, enrichment

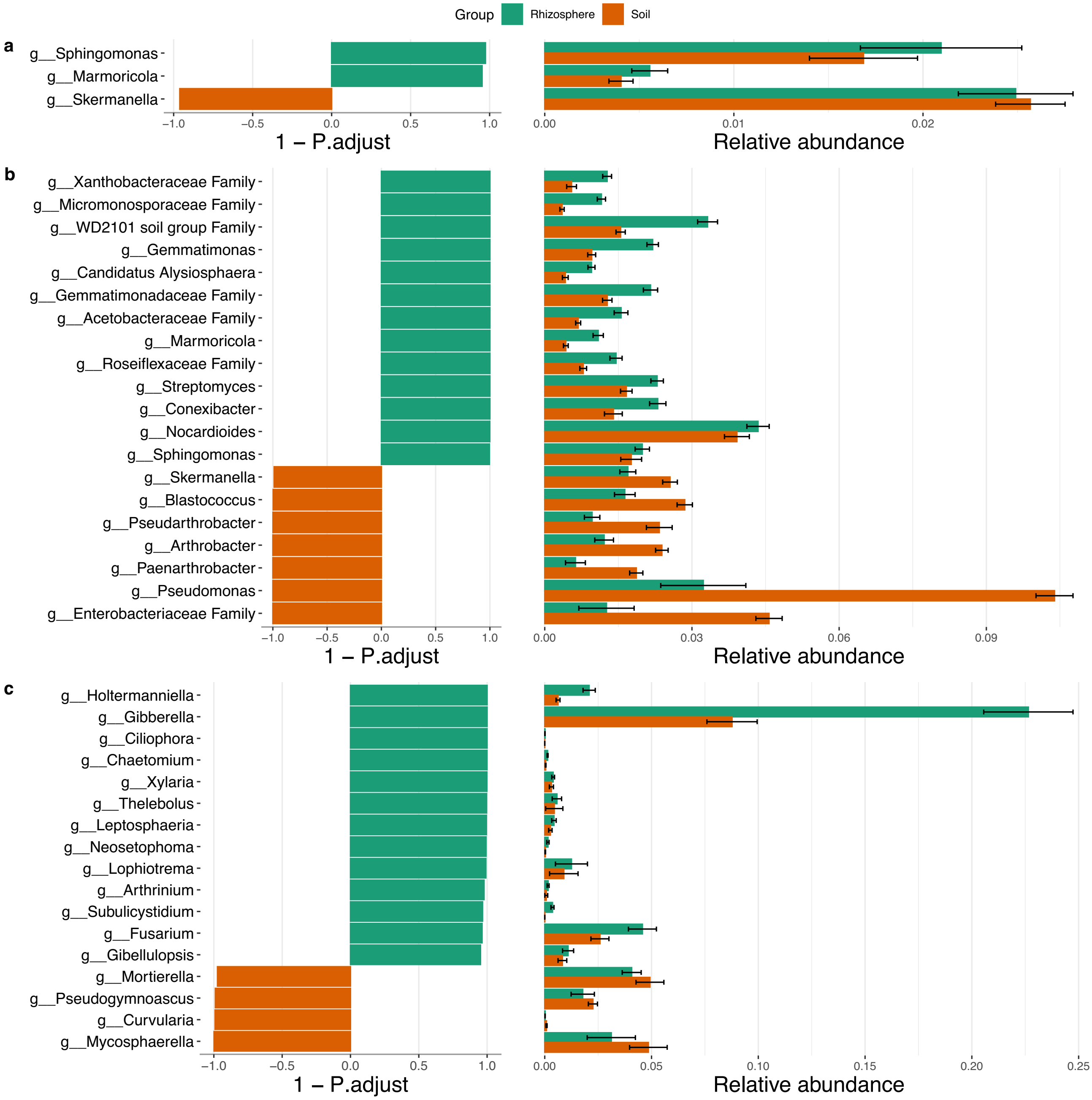

Figure 5. Differential abundance between rhizosphere and bulk soil at the genus level based on ALDEx2 center log ratio transformed data with Kruskal–Wallis rank sum test and p values FDR corrected using Holm's method. Only the 20 most abundant significant taxa ($p < 0.05$) are shown. (**a**) bacterial taxa from thistle sampled in annual cropping systems, (**b**) bacterial taxa from thistle sampled in perennial cropping systems, and (**c**) Fungi. Bacterial communities were analyzed separately from annual and perennial cropping systems due to differences in sequencing approaches.

of pathogenic genera within the rhizosphere suggests Canada thistle might be resistant to biocontrol efforts that utilize fungal pathogens.

While this work provides insight into the core rhizosphere microbiome of Canada thistle, more work is needed. This study has several limitations, which decrease the robustness of our findings. First, compared with whole-genome shotgun sequencing, amplicon sequencing has limited resolution and cannot reliably annotate to the strain level which is particularly important for determining pathogenicity. Also, functional assignment with tools such as PICRUSt and FUNGuild are useful for hypothesis generation but cannot reliably or conclusively determine function in place of metagenomic sequencing. Given differences in sequencing technologies that were used in this survey, more work is needed to identify differences in community structure between locations. Future work should also characterize the rhizosphere communities of native plants undergoing Canada thistle invasion. Predictive functional profiling suggested pathogen tolerance and specific biosynthetic pathways may

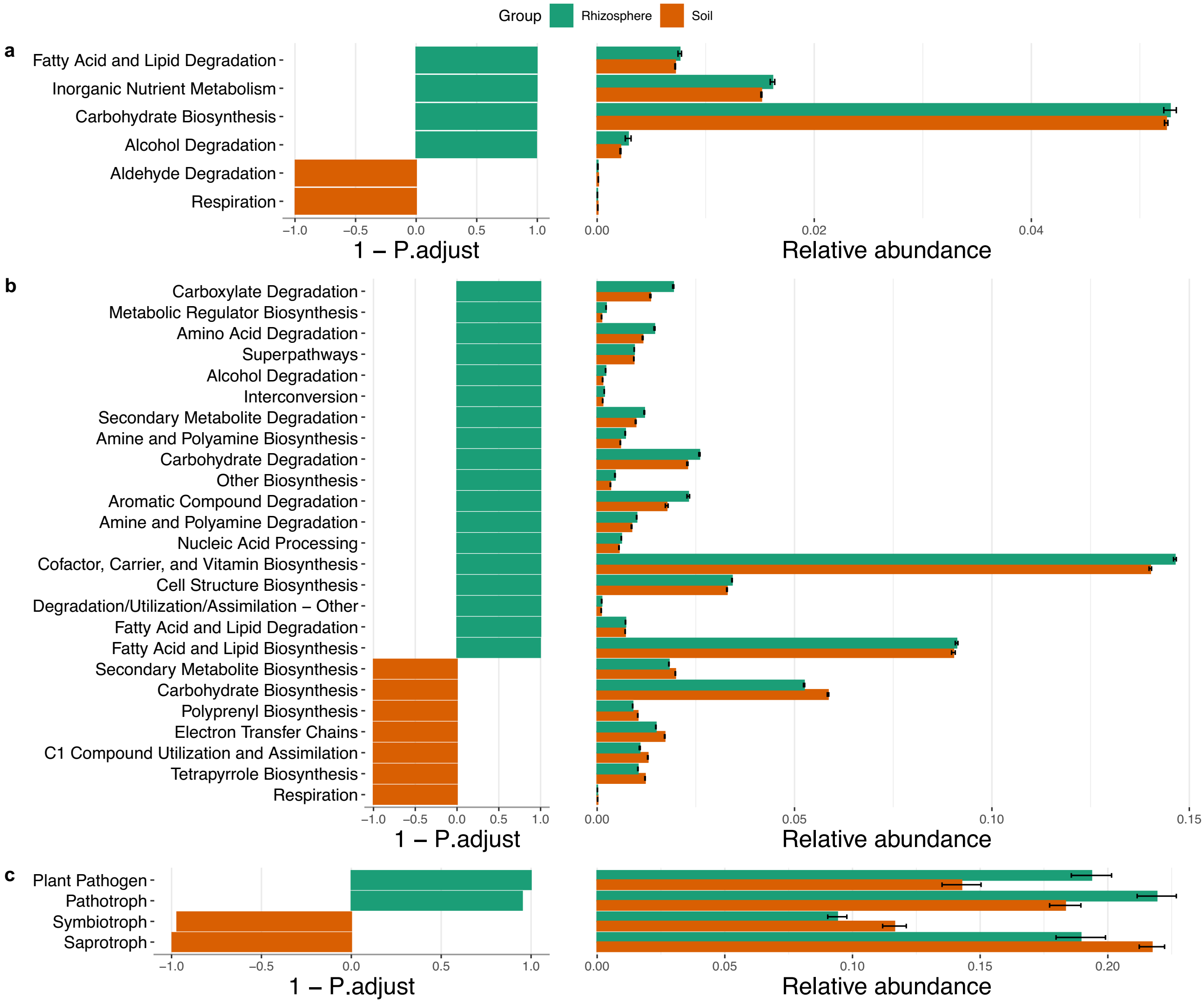

Figure 6. Differential abundance analysis by ALDEx2 with center log ratio transformed data with Kruskal–Wallis rank sum test and *p* values FDR corrected using Holm's method. Bacterial MetaCyc superclass 2 functions predicted by PICRUSt2 that were different between thistle rhizosphere and bulk soil from (**a**) thistle sampled in annual cropping systems, (**b**) thistle sampled in perennial cropping systems. (**c**) Fungal ecological guild predictions by FUNGuild. Only taxa with significant ($p<0.05$, FDR corrected) differences are shown.

provide Canada thistle a competitive advantage but cannot conclusively establish connections between putative functions and invasion success. These findings provide a basis for future hypothesis-driven experiments to empirically determine mechanisms for Canada thistle rhizosphere microbiome recruitment and the role this plays in invasion success. This study was also limited to annual cropping systems and perennial grasslands across the U.S. NGP. Our findings highlight the need for more work to characterize the rhizosphere microbiome in other regions and ecosystems (e.g., forests and grasslands) given the global importance of this invasive weed species.

Methods
Study locations and sampling
Bulk soil and Canada thistle rhizosphere samples were collected from seven locations across Montana, North Dakota, and Washington to assess the eco-geographic variation of the rhizosphere community. A total of 163 separate Canada thistle patches were sampled across all locations in late June through early July during the 2022 and 2023 field seasons. In 2022, rhizosphere samples were collected from patches in organic annual cropping systems, while 2023 sampling was performed in perennial forages. At each location, from 3 to 25 separate patches were sampled (Table 1). Rhizosphere samples were collected by using a shovel to excavate Canada thistle plants with attached roots from the upper 30 cm of soil within a 0.25 m^{-2} quadrant at 5-, 10-, and 15-m points along a transect established in each Canada thistle patch. A composite sample was prepared by pooling roots from 5 to 10 plants from each transect. Where patches were too small to establish a transect, plants were randomly selected throughout the patch. Bulk soil samples (0–15 cm depth) were collected with a 2.54-cm soil probe from areas that were free of Canada thistle to avoid possible rhizosphere effects. Samples were transported on ice to the lab

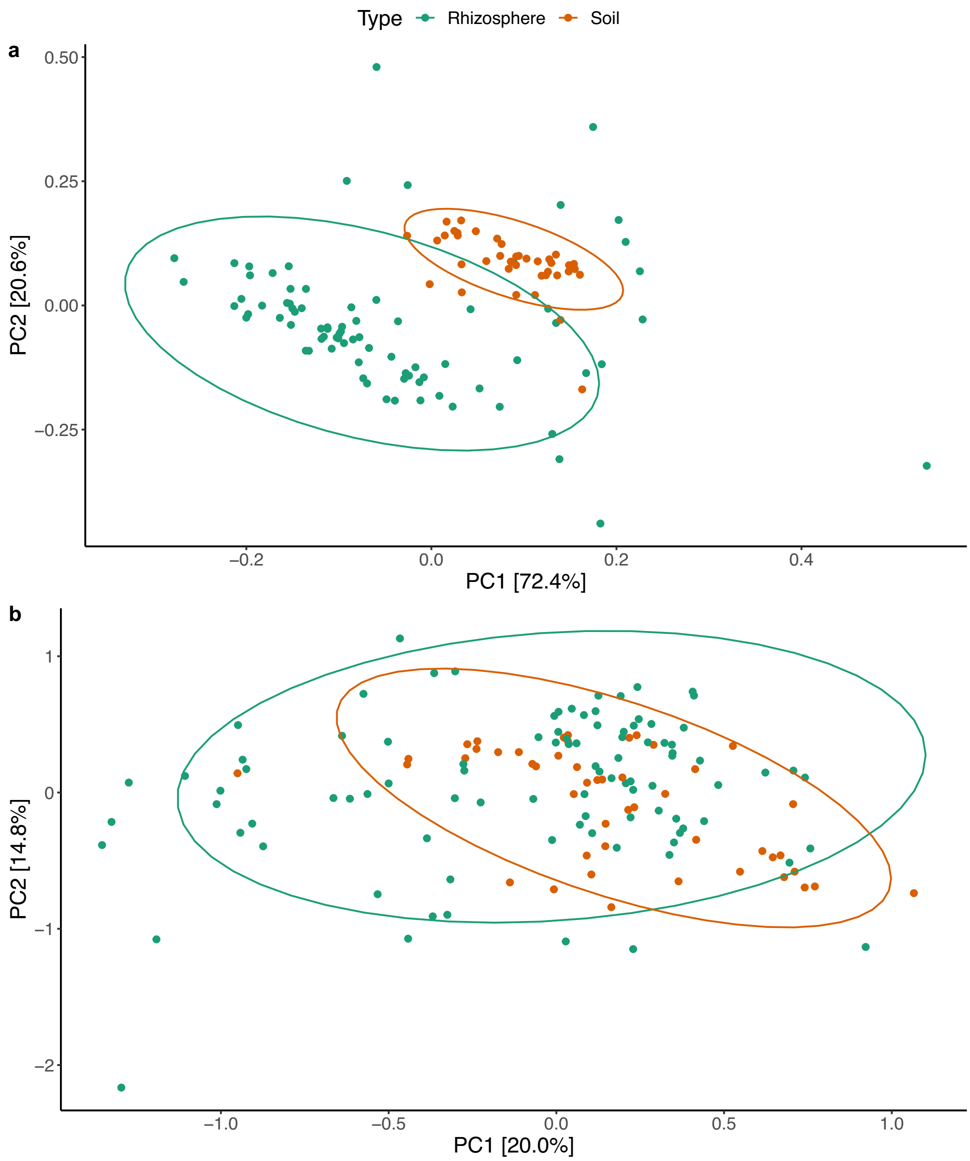

Figure 7. PCA ordination plots of predicted functions. (**a**) Center log ratio transformed bacterial PICRUSt2 predicted Metacyc functions at a rank of Superpathway1. (**b**) Center log ratio transformed FUNguild predicted fungal functional guilds. Ellipses show the 95% confidence intervals.

where bulk soil cores from each site were homogenized, passed through a 2-mm sieve, and stored at – 80 °C for DNA extraction.

Root samples were processed as previously described[42]. Briefly, loose soil was removed from roots and roots were then placed in 50-ml falcon tubes with 30 ml 0.9% NaCl. Tubes were placed in a sonicating water bath for 1 min. Roots were removed and the remaining soil solution was centrifuged for 2 min at 2000 × g, decanted and the remaining soil pellet was resuspended in sterile DI H_2O and stored at – 80 °C until DNA extractions were performed. DNA was extracted from a 250-mg soil subsample or rhizosphere extract using a Qiagen DNeasy PowerSoil Pro DNA Isolation Kit (Qiagen Inc., Germantown, MD, USA). In 2022, purified DNA was submitted to Integrated Microbiome Resource (Dalhousie University, Halifax, NS) for amplicon sequencing of the bacterial 16S V6-V8 region and fungal ITS2 region. Sequencing was performed on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) using a 2 × 300 PE kit. In 2023, DNA was submitted to Novogene (Sacramento, CA) for amplicon sequencing of the bacterial 16S V4-V5 region and fungal ITS2 region. Sequencing was performed on the Illumina NovoSeq platform (Illumina, San Diego, CA, USA) using a 2 × 250 PE kit.

Data processing

Quality control and processing of sequence reads was performed in R 4.1.3[43]. Demultiplexed sequences were trimmed by 20 bases from the start position to remove primer sequences and truncated using a Q score threshold

Location	Coordinates	Crop	Mean annual rainfall (mm)	Mean annual temperature (°C)	pH	Organic matter (%LOI)	CEC (cmolc kg^{-1})	Soil classification
Big Timber, MT	45.995863, – 109.91025	Perennial grasses	412.2	8.3	6.8	3.0	20.0	Verson loam, Mollisol
Ekalaka, MT	45.662814, – 104.391726	Annual forage	426.7	6.9	7.9	2.0	27.5	Harlake silty clay loam, Entisol
Highwood, MT	47.65152, – 110.62299	Mixed grass hay	662.2	7.5	6.9	3.5	25.9	Bearpaw-Waltham complex, clay loam, Mollisol
Ledger, MT	48.225276, – 111.144619	Lentil	302.5	6.1	7.1	2.0	20.1	Joplin-Hillon loams, Mollisol
Moccasin, MT	47.06242, – 109.94765	Wheat	388.6	6.1	7.3	3.9	27.3	Danvers-Judith clay loam, Mollisol
Absaraka, ND	46.988319, – 97.352284	Spring wheat, barley	573.5	5.8	7.8	3.6	24.5	Warsing sandy loam complex,frigid Oxyaquic Hapludolls
Pullman, WA	46.76242, – 117.19837	Wheat, pasture	516.9	8.6	5.0	4.4	20.6	Palouse silt loam, Mollisol

Table 1. Description of geographical, climate, and soil properties at each sampling location.

of 30. Fungal barcode sequences were removed using cutadapt[44] and filtered to exclude reads shorter than 50 bp. Quality filtering, chimera removal, and assembly of reads into error-corrected amplicon sequence variants (ASVs) were performed using DADA2[45]. Taxonomic assignment was performed using a naive Bayes classifier pre-trained on the weighted Silva 138.1 database with a 99% identity threshold[46] for bacterial 16S amplicons and the Unite version 8.3 fungal taxonomy database for ITS amplicons[47]. Taxa were filtered with a 0.25% minimum relative abundance threshold to remove spurious sequences[48].

Statistical analysis

Statistical analyses and data visualization were performed using phyloseq, microeco, microViz, and ggplot2, ggpubr packages[49–52]. Since amplicon sequencing was performed using bacterial 16S V6–V8 primers in 2022 and V4–V5 primers in 2023, each year was analyzed separately. Samples were rarefied to 20,000 sequences per sample to account for differences in library size among samples[53]. Species diversity was estimated using coverage, a measure of alpha diversity that standardizes for differences in sample size, and species dominance was estimated with Hill-Simpson (inverse Simpson's)[16,54]. Significant differences in alpha diversity between bulk soils and rhizosphere were determined using a Wilcoxon rank-sum test as recommended for nonparametric comparisons[55]. Relative abundances of major taxa were visualized at the phylum and family level using a 1% average relative abundance threshold. To gain insight into differences between rhizosphere communities across different ecoregions, non-metric multidimensional scaling (NMDS) based on the Aitchison dissimilarity matrix was performed using the "ord_plot" function of the microViz R package to visually compare the taxonomic composition across locations. The effect of location on rhizosphere community composition was also assessed on the same Aitchison dissimilarity matrix using permutational multivariate analysis (PERMANOVA) with 9999 permutations.

Abundance-occupancy distributions were used to identify members of the rhizosphere core microbiome across locations using a prevalence threshold of 80% as previously described[56]. To assess factors influencing assembly, a Sloan neutral model was used to determine if the rhizosphere community was more strongly influenced by deterministic factors or neutral assembly processes[22,57,58]. The neutral theory of abundance-occupancy distributions suggests that rare taxa will occupy few sites while very abundant taxa will have high occupancy[22,59,60]. Differential abundance testing was performed with ALDEx2 using center log ratio (clr) transformed data and the Kruskal–Wallis rank sum test with p-values FDR corrected using Holm's method were used to identify taxa that were differentially abundant between rhizosphere and bulk soil communities[61,62].

Predictive functional profiling

Predictive functional profiling of the bacterial communities was performed using PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States)[14]. This gene-based computational approach has been used for assessing ecological functions in various environments[63,64]. ASV sequences were aligned with reference sequences using the HMMER implementation of profile hidden Markov models[65], and phylogenetic placement of reads was done with EPA-NG[66]. Functional profile analysis and visualization were performed in R with microeco[52] and Metacyc pathways were compared at a superclass 2 level[67]. The accuracy of PICRUSt prediction was evaluated using the weighted nearest sequenced taxon index (NSTI) that measures the relatedness of ASVs in a given sample to reference genomes. The NSTI score of the samples was 0.19 ± 0.001 (s.e.) which was near the range reported for soils[26]. Ecological functions of the fugal ASVs were predicted with FUNGuild as implemented in the microeco R package[15,52]. Fungal functional groups (guilds) were classified according to trophic modes (e.g., pathotrophs, saprotrophs, and symbiotrophs).

To gain insight into differences between predicted functional profiles in the Canada thistle rhizosphere and bulk soil, principal component analysis (PCA) was performed on the clr transformed Metacyc functions at the superclass1 level using the "ord_plot" function of the microViz R package. Differential abundance analysis was performed by ALDEx2 as described above to identify significant differences in the relative abundance of predicted functions in the Canada thistle rhizosphere and bulk soil.

Data availability

The raw sequence data are deposited in the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1080620). All read processing steps, bioinformatic workflows, R code, and custom scripts are available on GitHub (https://github.com/Eberly-Lab/thistle).

Received: 14 March 2024; Accepted: 31 July 2024
Published online: 04 August 2024

References

1. Cripps, M. G. *et al.* Classical biological control of *Cirsium arvense*: Lessons from the past. *Biol. Control* **57**, 165–174. https://doi.org/10.1016/j.biocontrol.2011.03.011 (2011).
2. Kentjens, W., Casonato, S. & Kaiser, C. Californian thistle (*Cirsium arvense*): Endophytes and *Puccinia punctiformis*. *Pest. Manag. Sci.* https://doi.org/10.1002/ps.7387 (2023).
3. Hodgins, K. A., Guggisberg, A., Nurkowski, K. & Rieseberg, L. H. Genetically based trait differentiation but lack of trade-offs between stress tolerance and performance in introduced Canada thistle. *Plant Commun.* **1**, 100116. https://doi.org/10.1016/j.xplc.2020.100116 (2020).
4. Katovich, E., Becker, R., Chandler, M. & Marek-Spartz, M. Biological control of Canada thistle (*Cirsium arvense*) revisited: Host range of *Hadroplontus litura* on *Cirsium* species native to the Upper Midwest, USA. *Biocontrol Sci. Technol.* **32**, 1050–1064. https://doi.org/10.1080/09583157.2022.2085245 (2022).
5. Hodgson, J. M. The response of Canada thistle ecotypes to 2,4-D, amitrole, and intensive cultivation. *Weed Sci.* **18**, 253–255. https://doi.org/10.1017/S0043174500079686 (1970).
6. Frank, J. R. & Tworkoski, T. J. Response of Canada thistle (*Cirsium arvense*) and leafy spurge (*Euphorbia esula*) clones to chlorsulfuron, clopyralid, and glyphosate. *Weed Technol.* **8**, 565–571. https://doi.org/10.1017/s0890037x00039695 (2017).
7. Terrazas, R. A. *et al.* A footprint of plant eco-geographic adaptation on the composition of the barley rhizosphere bacterial microbiota. *Sci. Rep.* https://doi.org/10.1038/s41598-020-69672-x (2020).
8. Zhang, J., Liu, W., Bu, J., Lin, Y. & Bai, Y. Host genetics regulate the plant microbiome. *Curr. Opin. Microbiol.* **72**, 102268. https://doi.org/10.1016/j.mib.2023.102268 (2023).
9. Li, C. *et al.* Reshaping of the soil microbiome by the expansion of invasive plants: Shifts in structure, diversity, co-occurrence, niche breadth, and assembly processes. *Plant Soil* **477**, 629–646. https://doi.org/10.1007/s11104-022-05445-x (2022).
10. Rodriguez-Caballero, G., Caravaca, F., Diaz, G., Torres, P. & Roldan, A. The invader *Carpobrotus edulis* promotes a specific rhizosphere microbiome across globally distributed coastal ecosystems. *Sci. Total Environ.* **719**, 137347. https://doi.org/10.1016/j.scitotenv.2020.137347 (2020).
11. Caravaca, F., Torres, P., Diaz, G. & Roldan, A. Elevated functional versatility of the soil microbial community associated with the invader *Carpobrotus edulis* across a broad geographical scale. *Sci. Total Environ.* **813**, 152627. https://doi.org/10.1016/j.scitotenv.2021.152627 (2022).
12. LaForgia, M. L., Kang, H. & Ettinger, C. L. Invasive grass dominance over native forbs is linked to shifts in the bacterial rhizosphere microbiome. *Microb. Ecol.* **84**, 496–508. https://doi.org/10.1007/s00248-021-01853-1 (2022).
13. Yin, L. *et al.* The rhizosphere microbiome of *Mikania micrantha* provides insight into adaptation and invasion. *Front. Microbiol.* **11**, 1462. https://doi.org/10.3389/fmicb.2020.01462 (2020).
14. Douglas, G. M. *et al.* PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **38**, 685–688. https://doi.org/10.1038/s41587-020-0548-6 (2020).
15. Nguyen, N. H. *et al.* FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006 (2016).
16. Roswell, M., Dushoff, J. & Winfree, R. A conceptual guide to measuring species diversity. *Oikos* **130**, 321–338. https://doi.org/10.1111/oik.07202 (2021).
17. Lindahl, B. D. *et al.* Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytol.* **173**, 611–620. https://doi.org/10.1111/j.1469-8137.2006.01936.x (2007).
18. Li, F. *et al.* *Mortierella elongata*'s roles in organic agriculture and crop growth promotion in a mineral soil. *Land Degrad. Dev.* **29**, 1642–1651. https://doi.org/10.1002/ldr.2965 (2018).
19. Shi, S. *et al.* The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol. Lett.* **19**, 926–936. https://doi.org/10.1111/ele.12630 (2016).
20. Li, X. *et al.* Legacy of land use history determines reprogramming of plant physiology by soil microbiome. *ISME J.* **13**, 738–751. https://doi.org/10.1038/s41396-018-0300-0 (2018).
21. Busby, P. E. *et al.* Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol.* **15**, e2001793. https://doi.org/10.1371/journal.pbio.2001793 (2017).
22. Stopnisek, N. & Shade, A. Persistent microbiome members in the common bean rhizosphere: An integrated analysis of space, time, and plant genotype. *ISME J.* **15**, 2708–2722. https://doi.org/10.1038/s41396-021-00955-5 (2021).
23. Simonin, M. *et al.* Influence of plant genotype and soil on the wheat rhizosphere microbiome: Evidences for a core microbiome across eight African and European soils. *FEMS Microbiol. Ecol.* https://doi.org/10.1093/femsec/fiaa067 (2020).
24. Dawson, W., Schrama, M. & Austin, A. Identifying the role of soil microbes in plant invasions. *J. Ecol.* **104**, 1211–1218. https://doi.org/10.1111/1365-2745.12619 (2016).
25. Mangla, S., Inderjit, I. & Callaway, R. M. Exotic invasive plant accumulates native soil pathogens which inhibit native plants. *J. Ecol.* **96**, 58–67. https://doi.org/10.1111/j.1365-2745.2007.01312.x (2007).
26. Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821. https://doi.org/10.1038/nbt.2676 (2013).
27. Djemiel, C. *et al.* Inferring microbiota functions from taxonomic genes: A review. *GigaScience* https://doi.org/10.1093/gigascience/giab090 (2022).
28. Xue, C. *et al.* Effect of LSU and ITS genetic markers and reference databases on analyses of fungal communities. *Biol. Fertil. Soils* **55**, 79–88. https://doi.org/10.1007/s00374-018-1331-4 (2018).
29. Lin, H.-A. & Mideros, S. X. The effect of *Septoria glycinesand* fungicide application on the soybean phyllosphere mycobiome. *Phytobiomes J.* **7**, 220–232. https://doi.org/10.1094/pbiomes-12-21-0075-r (2023).
30. Watanabe, M. *et al.* Molecular phylogeny of the higher and lower taxonomy of the *Fusarium* genus and differences in the evolutionary histories of multiple genes. *BMC Evol. Biol.* **11**, 322. https://doi.org/10.1186/1471-2148-11-322 (2011).

31. Gordon, T. R. *Fusarium oxysporum* and the *Fusarium* wilt syndrome. *Annu. Rev. Phytopathol.* **55**, 23–39. https://doi.org/10.1146/annurev-phyto-080615-095919 (2017).
32. Windels, C. E. Economic and social impacts of fusarium head blight: Changing farms and rural communities in the northern great plains. *Phytopathology* **90**, 17–21. https://doi.org/10.1094/PHYTO.2000.90.1.17 (2000).
33. Crous, P. W., Braun, U. & Groenewald, J. Z. Mycosphaerella is polyphyletic. *Stud. Mycol.* **58**, 1–32. https://doi.org/10.3114/sim.2007.58.01 (2007).
34. Quaedvlieg, W. *Re-evaluating Mycosphaerella and Allied Genera* (Utrecht University, 2014).
35. Kema, G. H., Yu, D., Rijkenberg, F. H., Shaw, M. W. & Baayen, R. P. Histology of the pathogenesis of *Mycosphaerella graminicola* in wheat. *Phytopathology* **86**, 777–786 (1996).
36. Roger, C., Tivoli, B. & Huber, L. Effects of temperature and moisture on disease and fruit body development of *Mycosphaerella pinodes* on pea (*Pisum sativum*). *Plant Pathol.* **48**, 1–9. https://doi.org/10.1046/j.1365-3059.1999.00312.x (1999).
37. Ikunaga, Y. *et al. Nocardioides* sp. strain WSN05-2, isolated from a wheat field, degrades deoxynivalenol, producing the novel intermediate 3-epi-deoxynivalenol. *Appl. Microbiol. Biotechnol.* **89**, 419–427. https://doi.org/10.1007/s00253-010-2857-z (2011).
38. Fida, T. T., Palamuru, S., Pandey, G. & Spain, J. C. Aerobic biodegradation of 2,4-dinitroanisole by *Nocardioides* sp. strain JS1661. *Appl. Environ. Microbiol.* **80**, 7725–7731. https://doi.org/10.1128/AEM.02752-14 (2014).
39. Mudge, A. M. *et al.* A role for the mycotoxin deoxynivalenol in stem colonisation during crown rot disease of wheat caused by *Fusarium graminearum* and *Fusarium pseudograminearum*. *Physiol. Mol. Plant Pathol.* **69**, 73–85. https://doi.org/10.1016/j.pmpp.2007.01.003 (2006).
40. Zhao, F. *et al.* Vermicompost can suppress *Fusarium oxysporum* f. sp. lycopersici via generation of beneficial bacteria in a long-term tomato monoculture soil. *Plant Soil* **440**, 491–505. https://doi.org/10.1007/s11104-019-04104-y (2019).
41. Carrer Filho, R., Romeiro, R. S. & Garcia, F. A. O. Biocontrol of foliar disease of tomato plants by *Nocardioides thermolilacinus*. *Trop. Plant Pathol.* **33**, 457–460 (2008).
42. Mavrodi, O. V. *et al.* Rhizosphere microbial communities of *Spartina alterniflora* and *Juncus roemerianus* from restored and natural tidal marshes on Deer Island, Mississippi. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2018.03049 (2018).
43. R: A language and environment for statistical computing. R Foundation for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2023).
44. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* https://doi.org/10.14806/ej.17.1.200 (2011).
45. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from *Illumina amplicon* data. *Nat. Methods* **13**, 581–583. https://doi.org/10.1038/nmeth.3869 (2016).
46. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-596. https://doi.org/10.1093/nar/gks1219 (2013).
47. Nilsson, R. H. *et al.* The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**, D259–D264. https://doi.org/10.1093/nar/gky1022 (2019).
48. Reitmeier, S. *et al.* Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. *ISME Commun.* https://doi.org/10.1038/s43705-021-00033-z (2021).
49. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217. https://doi.org/10.1371/journal.pone.0061217 (2013).
50. Barnett, D., Arts, I. & Penders, J. microViz: an R package for microbiome data visualization and statistics. *J. Open Source Softw.* **6**, 3201. https://doi.org/10.21105/joss.03201 (2021).
51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
52. Liu, C., Cui, Y., Li, X. & Yao, M. microeco: An R package for data mining in microbial community ecology. *FEMS Microbiol. Ecol.* https://doi.org/10.1093/femsec/fiaa255 (2021).
53. Mcmurdie, P. J. & Holmes, S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531. https://doi.org/10.1371/journal.pcbi.1003531 (2014).
54. Morris, E. K. *et al.* Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* **4**, 3514–3524. https://doi.org/10.1002/ece3.1155 (2014).
55. Midway, S., Robertson, M., Flinn, S. & Kaller, M. Comparing multiple comparisons: Practical guidance for choosing the best multiple comparisons test. *PeerJ* **8**, e10387. https://doi.org/10.7717/peerj.10387 (2020).
56. Edwards, J. A. *et al.* Genetic determinants of switchgrass-root-associated microbiota in field sites spanning its natural range. *Curr. Biol.* **33**, 1926–1938. https://doi.org/10.1016/j.cub.2023.03.078 (2023).
57. Sloan, W. T., Woodcock, S., Lunn, M., Head, I. M. & Curtis, T. P. Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microb. Ecol.* **53**, 443–455. https://doi.org/10.1007/s00248-006-9141-x (2007).
58. Shade, A. & Stopnisek, N. Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr. Opin. Microbiol.* **49**, 50–58. https://doi.org/10.1016/j.mib.2019.09.008 (2019).
59. Gaston, K. J. *et al.* Abundance–occupancy relationships. *J. Appl. Ecol.* **37**, 39–59. https://doi.org/10.1046/j.1365-2664.2000.00485.x (2002).
60. Woodcock, S. *et al.* Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* **62**, 171–180. https://doi.org/10.1111/j.1574-6941.2007.00379.x (2007).
61. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15. https://doi.org/10.1186/2049-2618-2-15 (2014).
62. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
63. Chen, H., Xia, Q., Yang, T. & Shi, W. Eighteen-year farming management moderately shapes the soil microbial community structure but promotes habitat-specific taxa. *Front. Microbiol.* **9**, 1776. https://doi.org/10.3389/fmicb.2018.01776 (2018).
64. Zarezadeh, S. *et al.* Microalgae and phototrophic purple bacteria for nutrient recovery from agri-industrial effluents: Influences on plant growth, rhizosphere bacteria, and putative carbon- and nitrogen-cycling genes. *Front. Plant Sci.* **10**, 1193. https://doi.org/10.3389/fpls.2019.01193 (2019).
65. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195. https://doi.org/10.1371/journal.pcbi.1002195 (2011).
66. Barbera, P. *et al.* EPA-ng: Massively parallel evolutionary placement of genetic sequences. *Syst. Biol.* **68**, 365–369. https://doi.org/10.1093/sysbio/syy054 (2018).
67. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes—A 2019 update. *Nucleic Acids Res.* **48**, D445–D453. https://doi.org/10.1093/nar/gkz862 (2019).

Acknowledgements

This research was supported in part by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, Organic Research and Extension Initiative, Project 2018-51300-28432, Accession No. 1016580. The findings and conclusions in this preliminary presentation have not been formally disseminated by the U.S. Department of Agriculture and should not be construed to represent any agency determination or policy. This research was also supported in part by the Montana Specialty Crop Block Grant (SCBG) Project Number 21SC07004 - 031. Special thanks to John Gaskins and Kim Mann (USDA, ARS) for their assistance with sample collection. Permission was obtained from growers prior to on farm collection of thistle root samples.

Author contributions

J.E.: Conceptualization, formal analysis, visualization, writing—original draft, validation, resources, data curation, supervision. A.H.: Investigation, writing—review and editing. D.O.: Investigation, writing—review and editing. A.D.: Funding acquisition, investigation, writing—review and editing. T.S.: Investigation, writing—review and editing. P.C.: Funding acquisition, project administration, conceptualization, writing—review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-69082-3.

Correspondence and requests for materials should be addressed to J.O.E.